

Cheikh M. F. Kebe
Assane Gueye
Ababacar Ndiaye
Aminata Garba (Eds.)



249

Innovations and Interdisciplinary Solutions for Underserved Areas

LNICST

Second International Conference, InterSol 2018
Kigali, Rwanda, March 24–25, 2018
Proceedings



Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering

249

Editorial Board

Ozgur Akan

Middle East Technical University, Ankara, Turkey

Paolo Bellavista

University of Bologna, Bologna, Italy

Jiannong Cao

Hong Kong Polytechnic University, Hong Kong, Hong Kong

Geoffrey Coulson

Lancaster University, Lancaster, UK

Falko Dressler

University of Erlangen, Erlangen, Germany

Domenico Ferrari

Università Cattolica Piacenza, Piacenza, Italy

Mario Gerla

UCLA, Los Angeles, USA

Hisashi Kobayashi

Princeton University, Princeton, USA

Sergio Palazzo

University of Catania, Catania, Italy

Sartaj Sahni

University of Florida, Florida, USA

Xuemin Sherman Shen

University of Waterloo, Waterloo, Canada

Mircea Stan

University of Virginia, Charlottesville, USA

Jia Xiaohua

City University of Hong Kong, Kowloon, Hong Kong

Albert Y. Zomaya

University of Sydney, Sydney, Australia

More information about this series at <http://www.springer.com/series/8197>

Cheikh M. F. Kebe · Assane Gueye
Ababacar Ndiaye · Aminata Garba (Eds.)

Innovations and Interdisciplinary Solutions for Underserved Areas

Second International Conference, InterSol 2018
Kigali, Rwanda, March 24–25, 2018
Proceedings

Editors

Cheikh M. F. Kebe
Cheikh Anta Diop University
Dakar
Senegal

Assane Gueye
University System of Maryland
Bambey
Senegal

Ababacar Ndiaye
Assane University
Seck de Ziguinchor
Senegal

Aminata Garba
Carnegie Mellon University
Kigali
Rwanda

ISSN 1867-8211 ISSN 1867-822X (electronic)
Lecture Notes of the Institute for Computer Sciences, Social Informatics
and Telecommunications Engineering
ISBN 978-3-319-98877-1 ISBN 978-3-319-98878-8 (eBook)
<https://doi.org/10.1007/978-3-319-98878-8>

Library of Congress Control Number: 2018950769

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

InterSol is an international conference that meets annually in Africa and brings together scientists from around the world. It serves as a framework where participants present the results of research projects that address issues faced by people in underserved areas. In fact, populations living in underserved areas face problems in almost all sectors, particularly in energy, water, communication, climate, food, health, education, transportation, social development, and economic growth, just to name a few. Moreover, experience has shown that none of these issues can be solved in isolation to the others: they are all interdependent. Therefore, any sustainable and effective solution must integrate an interdisciplinary dimension, involving expertise from several disciplines, while taking into account certain geographical, social, economic, and environmental specificities. This is the main motivation behind Interdisciplinary Solutions (InterSol); an international conference dedicated to the advancement of interdisciplinary research works that address people's needs in underserved areas.

The first edition was held in Dakar, Senegal (April 11–12, 2017) and validated our choice, given the number, the diversity, and the quality of submitted and accepted papers. This second edition, held in Kigali, Rwanda, during March 24–25, 2018, witnessed the participation of researchers (academia and industry), graduate students, NGO professionals (from Africa, Europe, America, and Asia), as well government officials, who discussed for two days results in the themes defined in the conference main track as well as during the workshop on climate change, co-located with the conference. Finally, like the previous edition, InterSol 2018 hosted the CNRIA (Colloque National sur la Recherche en Informatique et ses Applications) workshop, held during April 18–21, 2018, in Ziguinchor, Senegal.

This volume presents papers accepted for publication and presented during InterSol2018 and CNRIA 2018. It includes 23 articles, eight of which were part of CNRIA.

The papers presented in the main track address issues in the following themes:

- Energy: with papers dealing with the development of the performance of solar systems in Africa, the wind potential in Chad, and building energy efficiency in Africa
- Socio-economic: with papers addressing the issues of public health, communication, education, and access to energy
- Telecom: with papers studying the business model of telecommunication in the Rwandan context and the phenomenon of boomerang routing in intra-Africa communications
- Math models: in which the papers present mathematical models of the climatic phenomenon and prediction models for solar radiation in Africa
- Health: with papers dealing with developing medical devices that are suitable to underserved areas

The workshop on climate change (Beyond the Debate: Climate Change as an Economic Opportunity for Africa) offered a unique opportunity to discuss the impacts of climate in Africa, the solutions that need to be developed, as well as the most appropriate means for climate change control and adaptation in underserved areas. A round table on the issue of climate change in Africa concluded this session.

This volume, like the previous one, will serve as a channel for disseminating the solutions proposed during this two-day conference.

June 2018

Cheikh M. F. Kebe
Assane Gueye
Ababacar Ndiaye
Aminata A. Garba

Organization

Steering Committee

Imrich Chlamtac	EAI/Bruno Kessler Professor, University of Trento, Italy
-----------------	---

Organizing Committee

General Chair

Cheikh Mouhamed F. Kebe	ESP/UCAD, Senegal
-------------------------	-------------------

General Co-chairs

Assane Gueye	UADB, Senegal
Aminata Garba	UMD-USA CMU, Rwanda

TPC Chair and Co-chairs

Kara Nelson	UCB, USA
Wilfred Ndifon	AIMS, Rwanda
Mohamed Mejri	Université de Laval, Canada
Melissa Densmore	University of Cape Town, South Africa

Web Chair

Marie Helene Mballo	UADB, Senegal
---------------------	---------------

Publicity and Social Media Chair and Co-chairs

Ghada Bassioni	Ain Shams University, Egypt
Moustapha Diop	UMD, USA

Workshops Chair

Francois Kabore	Institut Universitaire Jésuite, Ivory Coast/Georgetown University, USA
-----------------	---

Sponsorship and Exhibits Chair

Paulette Mpano	Rwanda Online, Rwanda
----------------	-----------------------

Publications Chair

Ababacar Ndiaye	UASZ, Senegal
-----------------	---------------

Tutorials Chair

Alassane Diop UVS, Senegal/Université de Laval, Canada

Local Chair

Youssef Travalay NEF, Rwanda

Conference Manager

Lenka Biliska European Alliance for Innovation

Technical Program Committee

Cole-Rhodos Arlene	Morgan State University, Baltimore, USA
Kane Cheikhou	Ecole Supérieure Polytechnique, Dakar, Senegal
Senghane Mbodj	Université Alioune Diop de Bambey, Senegal
Wilfred Ndifon	African Institute of Mathematical Sciences, Kigali, Rwanda
Moustapha Diop	University of Maryland, College Park, USA
Youssef Travalay	Next Einstein Forum, Kigali, Rwanda
Gaoussou Camara	Université Alioune Diop de Bambey, Senegal
Mouhammadou Lamine Ba	Université Alioune Diop de Bambey, Senegal
Jessica C. Rivas	WindAid Institute, Texas, USA
Bamba Gueye	Université Cheikh Anta Diop, Dakar, Senegal
Kara Nelson	University of Berkeley, California, USA
Ghada Bassioni	Ain Shams University, Egypt
Hamidou Tembine	NYU, Abu Dhabi, UAE
Mohamed Mejri	Université de Laval, Quebec, Canada
Aminata Garba	Carnegie Mellon University, Kigali, Rwanda
Assane Gueye	Université Cheikh Anta Diop, Dakar, Senegal/University of Maryland, College Park, USA
Francois Pazisnewende Kabore	Institut Universitaire Jésuite, Ivory Coast/Georgetown University USA
Melissa Densmore	University of Cape Town, South Africa
Joyojeet Pal	University of Michigan, USA
Ababacar Ndiaye	Université Assane Seck de Ziguinchor, Senegal
Charif Mahmoudi	National Institute of Standards and Technology (NIST), USA

Contents

INTERSOL Main Track

On the Prevalence of Boomerang Routing in Africa: Analysis and Potential Solutions	3
<i>Assane Gueye and Babacar Mbaye</i>	
Ego-Community Evolution Tracking in Instant Messaging Networks	13
<i>Ahmed Ould Mohamed Moctar and Idrissa Sarr</i>	
Investigation of Degradation in Crystalline Silicon Photovoltaic Modules After 10 Years Exposition in Senegal by Infrared (IR) and Electroluminescence (EL).	23
<i>Ababacar Ndiaye, Cheikh M. F. Kébé, Boudy Ould Bilal, Abdérafi Charki, Vincent Sambou, and Papa A. Ndiaye</i>	
Building Energy Audit in Nigeria: Some Guides for Energy Efficiency Building Regulations	41
<i>Abdulhameed Danjuma Mambo and Cheikh Mouhamed Fadel Kebe</i>	
ICT Performance Indicators in Formal Education at the Secondary School Level in Rwanda	52
<i>Florien Gumyusenge, Robert Mugisha, Aminata A. Garba, and Martin Saint</i>	
The Impact of over the Top Service Providers in the Rwandan Telecommunications Market: An Analysis of Business Models	62
<i>Emelyne Kwizera, Darius Mico, Michael Nayebare, Aminata A. Garba, Martin Saint, and Linda G. Deen</i>	
Challenges and Opportunities in Instrumentation and Use of High-Density EEG for Underserved Regions	72
<i>Ashwati Krishnan, Ritesh Kumar, Arnelle Etienne, Amanda Robinson, Shawn K. Kelly, Marlene Behrmann, Michael J. Tarr, and Pulkit Grover</i>	
E-Medicine: A Secure Transmission of Electrocardiograms Using Chaotic Oscillators Synchronization	83
<i>Alain Tiedeu, Yannick Abanda, and Gutenbert Kenfack</i>	
Bringing Life Where There Is No Light: A Low-Cost Movie Projection and Data Collection Solution for Underserved Areas	90
<i>Alain Shema and Michel Bézy</i>	

A Study of the Wind Potential in Climatic Zones of Chad	100
<i>Zoutene Pabame, Cheikh Mouhamed F. Kebe, Boudy Ould Bilal, Ababacar Ndiaye, Assane Gueye, and Pape Alioune Ndiaye</i>	

Climate Change Workshop

Beyond Participation: Welfare Effects of Gender-Differentiated Group-Based Approaches Under Climate Change in Kenya	111
<i>Marther W. Ngigi and Ulrike Mueller</i>	

Climate Change Signals Over Senegal River Basin Using Regional Climate Models of the CORDEX Africa Simulations.	123
<i>Mamadou Lamine Mbaye, Samo Diatta, and Amadou Thierno Gaye</i>	

Climate Change Mitigation Potential in Agricultural and Forestry Sector: The Impact of Expanded Woody Biomass Co-firing on Global Climate Stabilization	133
<i>Aklesso Y. G. Egbendewe</i>	

Climate Change May Result in More Water Availability in Parts of the African Sahel.	143
<i>Ousmane Seidou</i>	

Evaluation and Update of Two Regional Methods (ORSTOM and CIEH) for Estimations of Flow Used in Structural Design in West Africa	153
<i>Sehouevi M. D. Agoungbome, Ousmane Seidou, and Moussa Thiam</i>	

Statistical Downscaling of Global Climate Model MIROC_4h Outputs to Precipitation in Rwanda	163
<i>Nkusi Pearl and Mutabazi Alphonse</i>	

CNRIA Workshop

An Internet of Things Infrastructure for Rainfall Monitoring in Dakar	175
<i>Abdoulaye Kama, Cheikh Modou Noreyni Fall, Mamadou Simina Drame, and Moussa Diallo</i>	

A Parallelized Spark Based Version of mRMR.	187
<i>Reine Marie Ndéla Marone, Fodé Camara, and Samba Ndiaye</i>	

Multi-scenario Modelling of Learning	199
<i>Guy Merlin Mbatchou, François Bouchet, and Thibault Carron</i>	

RailMon: Distance, Temperature and Location Railway Monitoring Using IoT Technologies.	212
<i>Pa Saffiong Kebbeh, Madoune R. Seye, Bassirou Ngom, Bamba Gueye, and Moussa Diallo</i>	
An Encoding for the Theta Model of Elliptic Curves.	224
<i>Nafissatou Diarra and Emmanuel Fouotsa</i>	
BEDWE: A Decentralized Workflow Engine for Best-Effort Infrastructures	236
<i>Palakiyem Wallah, Cédric Tedeschi, and Jean-Louis Pazat</i>	
A Robust Process to Identify Pivots Inside Sub-communities in Social Networks	248
<i>Joseph Ndong and Ibrahima Gueye</i>	
Author Index	261

INTERSOL Main Track



On the Prevalence of Boomerang Routing in Africa: Analysis and Potential Solutions

Assane Gueye^{1,2(✉)} and Babacar Mbaye¹

¹ Universite Alioune Diop de Bambey, Bambey, Senegal
{assanel.gueye, babacar.mbaye}@uadb.edu.sn

² University of Maryland, College Park, College Park, USA

Abstract. When an African Internet user sends a message to a friend in another country in the continent, the data travels around the world (mostly Europe and USA) before coming back to the continent. This phenomenon is called *boomerang routing* at the continent level. The implications of boomerang routing include: higher cost, increased delay and increased information exposure. In this paper, we use active measurement data (ICMP traceroute) from 2015 and 2016 to empirically study the prevalence of boomerang routing in Africa by focusing on the implications on cost and delay. We also discussed possible improvements of the current African connectivity map to reduce boomerang routing, which will eventually translate into reduced cost and lower delays for end-users.

Keywords: Africa · Boomerang · ICT · e-Trading · Internet access cost
Delay

1 Introduction

The 21st century has been dubbed by many as the African century. In a recent printout, McKinsey reports that, following a decade of economic expansion, Africa is going digital. Although only 16% of the people in the continent are online today, that share is rising rapidly, thanks to recent growth in mobile networks and falling cost of Internet-capable devices. In 2016, there were more than 720 million mobile phone users in Africa, some 167 million African already used the Internet, and 52 million were on Facebook [1].

This follows a global trend of information and communication technologies (ICT) being an enabler for increased socio-economic growth. Not only ICTs help increase productivity, they also contribute to the overall GDP of an economy. ICTs can also help connect remote populations to markets, promote citizens' access to government and social services, expand educational opportunities, create platforms for innovation and increase people's freedoms [2].

This strong impact on socio-economic development is likely to carry-on in a much larger scale in the coming decade. This is significant for Africa because, according to the World Bank [3], mobile and broadband have more impact in developing economies than in developed economies. This has already been observed in the mobile communication sector. In 2014 the impact of mobile phones in term of GDP was 5.7% in Africa, which was more than any other continent [3].

Today, the Internet’s contribution to Africa’s GDP is relatively low, at 1.1%—just over half the levels seen in other emerging markets and well below the average of 3.7% in developed economies [4]. However, studies [1] have found that a 10% increase in broadband penetration in developing countries is correlated with a 1.35% increase in GDP, suggesting that Africa has even more to gain from the digital revolution. By 2025, the digital contribution to African GDP is expected to catchup with Sweden and Taiwan [1]. If the Internet matches or exceeds that level of impact, the result could be a leap forward in Africa’s economic growth and development.

However, to meet those expectations, a certain number of barriers need to be removed. Today there is growing consensus that the primary barriers to connectivity in Africa are availability, affordability, relevance and readiness [5]. Relevance refers to information and service available to people’s in their native language, while readiness reflects to people having the necessary basic skills, awareness, understanding, and cultural and social acceptance to the Internet.

In this paper, we are primarily interested in the availability of Internet infrastructures and the affordability to the end-users. More specifically, we are interested in understanding how the lack of a continent-wide connectivity infrastructures is impacting the cost to the end-users. For that, we analyze data (gathered between January 1, 2015 and December 31, 2016) to quantify the prevalence of *boomerang* routing in intra-Africa Internet communication.

Boomerang routing (or *tromboning* [7]) was originally used [6] to designate the situation whereby an Internet path starts and ends in a country, yet transits in an intermediate country. In this paper, we consider boomerang routing at the continent level and focus our attention to Africa. More precisely, we are interested in the implications of boomerang routing in cost and delay. This is important because today Africa is the region with the most expensive fixed-broadband prices, with an average price of more than 100% of GNI per capita [2]. Also, anyone who has used the Internet in Africa, might have experienced very long delays when trying to download a relatively large file.

For our study, we use publicly available data from the Center for Applied Internet Data Analysis (CAIDA) [8], which consists of router level communication traces for the entire 2015–2016 period. We then use the Max-Mind service [9] to geolocate each router within a country and we convert the router-level paths into country level paths. The dataset includes paths from the entire world. In this study, we only exploit paths for which both the origin and the destination are in Africa.

We first count the number of paths for which the traffic leaves the continent before coming back (i.e., there is a boomerang). Our analysis shows that in 2015, 70% of intra-Africa communication exited the continent before reaching the destination. In 2016, the fraction of boomerang traffic was 60%. When it exits, most of the traffic passes through the EU. US comes in second position. We also compute the number of hops a boomerang traffic goes through outside the continent. We found that in average, when exited, the traffic traverses 3.5 hops before returning to the continent.

We then focus on the communication delay (mainly round-trip-time) of the traffic. Our analysis suggests that, despite the fact that boomerang traffic roams in other continents before coming back to Africa, their RRT delay is smaller than the RTT delay of none-boomerang traffic that entirely stays within the continent.

The time that the traffic spends during the boomerang period is a key parameter for the cost of communication seen by end-users. In fact, when using international Internet connection, the African Internet Service Provider leases connectivity infrastructures and pays a fee known as transit cost. This fee is eventually transferred to the end-user who is then paying an additional (international) cost even when communicating with a friend in the same or neighboring country. A similar phenomenon is happening in the airline industry, where to go from one African country to another, travelers go through Europe, leading to additional cost.

A solution to this situation that has gained a large consensus is to build new communication infrastructures. This is certainly the reason why it is well considered in the African Internet Exchange System (AXIS) Project [10]. The main recommendation in this project is to build new Internet Exchange Points (IXPs) in the continent. Our study shows that, based on the data we gathered and the current connectivity between the existing IXPs, although it is important to build new IXPs, it is more pressing to connect the existing ones and to improve the bandwidth within the continent.

The rest of this paper is organized as follows. In Sect. 2, we give an overview of the current international Internet Peering principle and we talk about the boomerang routing phenomenon and its consequences in cost, delay and privacy. We then describe the dataset used in our study in Sect. 3 and perform our analysis in Sect. 4. In Sect. 5 we explore ways to improve the current African connectivity map to reduce boomerang routing. Concluding remarks and future work are presented in Sect. 6.

2 International Internet Connectivity: Peering, Transit and Boomerang

For international telephone routing, the provider in the country that originates the call makes a compensation payment to the provider in the destination country. If traffic is balanced in both directions, these payments usually cancel each other out. In the 1990s, traffic was higher in the direction of developed countries to developing countries and the ITU estimated that the net flow of settlement payments from developed countries to developing ones amounted to some USD 40 billion between 1993 and 1998 [11]. By the late 1990s, these payments have declined and even reversed as more traffic has shifted to the Internet.

Interconnectivity in the global Internet is essentially done through *peering* and *transit*. Peering refers to a relationship between two or more ISPs of similar size, in which the ISPs create direct links between each other and agree to forward each other's packets directly across this link. Transit refers to a bilateral interconnection where the customer pays the provider for connectivity to the global Internet. It is mainly the service provided by larger ISPs to smaller ISPs (such as those in Africa).

With the current international Internet charging principles (called full-circuit model), ISPs based in countries remote from Internet backbones, particularly in the developing countries, must pay the full (transit) cost of the international circuits. For example when a user in Kenya sends an email to a user in the USA, it is the Kenyan ISP who is bearing the full cost of the international connectivity. Conversely, when a user in the USA sends an email to a user in Kenya, it is still the Kenyan ISP

who bears the international cost (different from the telephone network where payments goes in both directions) [12]. This cost is eventually transfer to the end user in Kenya.

Because most Africa ISPs rely on transit via ISPs in Europe or in the US, this very phenomenon is still experienced when an African Internet user sends a message to a friend in another country in the continent. In this case, the data travels around the world (mostly Europe and USA) before coming back to the continent. This phenomenon is dubbed *boomerang routing* at the continent level. In this case, ISPs also must pay the international transit cost. This represents an additional cost for intra-Africa communication that is due to the boomerang routing.

Other consequences of boomerang (that are not detailed in this paper) are increased delay and privacy or information exposure. Indeed, since communication is bound by the speed of light, when the traffic roams around the world before coming back to Africa, this creates additional delays, which are commonly experienced by users in Africa. Also, data sent over the Internet can be monitored and manipulated by the entities that transmit that data from the original source to the destination. For unencrypted communications (and some encrypted communications with known weaknesses), eavesdropping and man-in-the-middle attacks are possible. For encrypted communication, the identification of the communicating endpoints is still revealed. In addition, encrypted communications may be stored until newly discovered weaknesses in the encryption algorithm or advances in computer hardware render them readable by attackers. Thus, when two African countries are communicating, the information is exposed to the rest of the world because of the boomerang routing phenomenon. In this paper, we study the prevalence of boomerang routing in intra-Africa communication. For that, we use real-life communication data gathered during the period of 2015–2016. In the next section we describe the data utilized in this study.

3 Data Acquisition and Description

3.1 Data Acquisition

We obtained our data from the public datasets provided by the Center for Applied Internet Data Analysis (CAIDA) [8] covering the years 2015 and 2016. We collected traceroute type data from a worldwide set of monitors. Every 2–3 min, each monitor probes a random prefix using ICMP traceroute and stores the router-level path information. We convert these paths into country paths by using the MaxMind [9] geolocation database. While geolocation data of routers can be inaccurate, previous work has found that it is more accurate at a country level of abstraction [13]. We then filter out the traces for which both the source and the destination are within Africa and keep them for our analysis. Overall, we collected a total of 140K traces for 2015 and 350K traces for 2016.

3.2 Data Limitation

The main limitation of the data is the low number of sources countries of the collected traces. In fact, the Ark infrastructure only has a limited number of monitors within

Africa (only in five African countries in 2015 and in 13 countries in 2016). Since the source of a given trace is the country where it is geo-located, this limits the number of source countries in our analysis. However, previous studies have shown that these countries are very representative when it comes to Internet communication in the continent. Therefore, we recognize that our study might not be complete but believe that it generates well to the whole continent.

Another limitation of the data is that the Ark infrastructure is only capable of discovering preferred paths to and from subnets containing a monitor. Routes between subnets that do not lie on a preferred route from a monitor to a target subnet will not be discovered. However, given that each subnet will be probed from a random monitor every two to three days, the traceroutes to each subnet will come from many different directions. Thus, we have confidence that we are discovering the major pathways from each source countries to the rest of the continent.

Despite these limitations, we believe that our study can be used to have a very good insight about Internet communication in Africa. Furthermore, since data is in general very rarely available in Africa, we consider this study as one step towards the general effort to understanding the evolution and implications of Internet communication within the continent.

4 Data Analysis

With the data collected for each year, we perform a certain number of experiments. First, we count the number of boomerang traces. Those are traces that contain routers that were geo-located outside the continent. Then, for each boomerang trace, we compute the number of countries that are traversed during the “*exit period*”. Finally, we record the round-trip-time (RTT) for each trace. For non-boomerang traces, we only record the RTT to the destination. For boomerang traces, we record the time for the three portions of the communication: the first portion before exiting the continent, the exit portion (when the traffic roams outside the continent), and the return portion when the traffic comes back to the continent. It is worth to notice that we did not find any trace that boomerang more than once.

4.1 Number of Boomerang Routes

Table 1 summarizes the results of our study. Among the 138505 traces collected in 2015, 97157 exited the continent before arriving to the destination country. This represents 70% of the traffic. This percentage improved slightly in 2016 where 60% of the traffic was found to boomerang.

These numbers are consistent with previous studies [14, 15]. In [14], Fanou *et al.* collected and analyzed traceroutes data from November 2013 to April 2014, while Chavula *et al.* analyzed data collected during a two-week period between April 6th and April 20th 2014. Both studies revealed a percentage of more than 70% of boomerang routes. Our study (which can be considered as a complement and update to these previous studies) shows that although there is a slight improvement in 2016, most of the intra-Africa communication traffic still leaves the continent before coming back.

This translates into high costs of intra-Africa communication that are mainly paid to ISPs in Europe and the US. According a study of the African Union [16], Africa spends between US \$400 millions and \$600 million per year in transit fees for intra-African traffic [14].

Table 1. Results summary

	Year 2015	Year 2016
Number of traces	138505	354575
Number of boomerang	97157	215259
% boomerang	70.14	60.07

In our analysis, we were also interested in two other questions: (1) where does the “exit traffic” go? And (2) how many countries are traversed outside Africa?

The data shows that most of the boomerang traffic goes to Europe (58%). We could not perform a more granular study because there are many routers for which the geo-location database just returns “EU” as country of residence, without further information about which European country it is. The US comes in second position receiving 36% of intra-Africa traffic. In very rare occasion the traffic roams through Asia and Latin America. When the traffic gets outside of Africa, we found several patterns: some traces remain entirely within Europe, some within the US and some traces go both through Europe and the US. Finally, the data shows that a boomerang traffic will traverse in average 3.5 countries. This indicates a high degree of information exposure when African countries are communicating over the Internet.

4.2 Delay Analysis

In this section, we want to understand the causes of the large delays experienced in Africa-to-Africa communication. For that, we record the round-trip-time (RTT) for each trace. For non-boomerang traces, we only record the RTT to the final destination. For boomerang traces, we record the RTT delay for the three portions of the communication: (2) the first portion before exiting the continent. We approximate this delay as the RTT between the source of the trace and the last route of the trace that is still within Africa. (b) the exit portion (when the traffic roams outside the continent). We estimate this delay as the RTT between the last router just before the traffic exits the continent and the first router just after coming back. (c) the return portion when the traffic comes back to the continent. We approximate this delay as the RTT between the first router when the trace comes back to the continent and the destination. We then derive the average delay of each of these portions (computed using all traces with boomerang). The result is shown in the three bottom columns of Table 2. Note that our 2015 data and 2016 data produced similar results (for all portions) and thus we present only the results for 2016 in this paper.

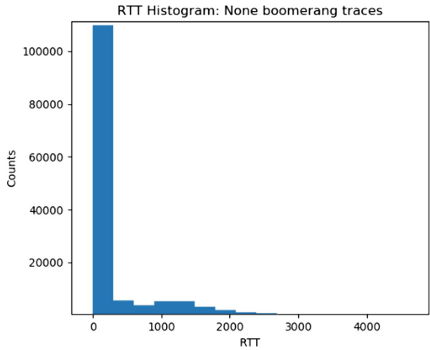


Fig. 1. Histogram of RTT (milli-second) for none boomerang traces

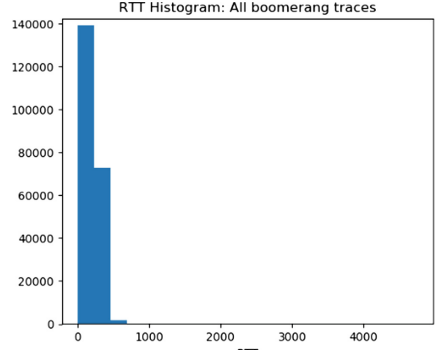


Fig. 2. Histogram of RTT (milli-second) for boomerang traces

Figure 1 shows the histogram of the end-to-end RTT for traces that stayed within Africa, while Fig. 2 shows a similar histogram for traces that exited the continent. We have observed (not shown here) that both types of traces experience large delays (up to 30 s). However, the figures show that, despite the additional distance travelled by boomerang traffic, none-boomerang traces (i.e., traces that stay within the continent), tend to have larger delays in general. Table 2 shows the average delay for each type of traffic: the average round trip time delay for none-boomerang traces is 407 ms (top column) while the average RTT delay for boomerang traces is 240 ms.

We then zoom into the boomerang traces to analyze the delays in the three different portions: (a) Portion 1: when the traffic is still within Africa, (b) Portion 2: when the traffic exists the continent, and (c) Portion 3: when the traffic returns back to the continent. The average RTT delay for these portions are given in the three bottom columns of the table.

Table 2. Average delay analysis

Traffic type		Average RTT delay (ms)
None-boomerang traffic (entire trace within Africa)		407
Boomerang traffic	End-2-End trace	240
	Portion 1: within Africa	146
	Portion 2: out of Africa	31
	Portion 3: within Africa	63

As can be seen in the table, the delay experienced by the traffic that stays within the continent is much larger. The first portion (from the source to the last router within the continent that is traversed before the boomerang) has the largest delay. It has an average RTT delay of 146 ms, while the last portion (from the first router within the continent when the traffic comes back, to the destination) has an average RTT delay of 63 ms. The boomerang delay is just 30 ms.

The reasons of the differences in these delays are worth investigating. One plausible explanation for the smaller delay experienced by the boomerang traffic (portion 2) might be the larger bandwidth that is available in Europe and in the US. We believe that more data is needed in order to understand why the first portion of the traffic has a considerably larger delay than the last portion (despite the fact that they both stay within the continent).

5 Network Improvements

To correct intra-Africa boomerang and eventually improve cost and delay for the African end-users, many have suggested building new infrastructures and mainly Internet Exchange Points (IXPs). This has led the African Union to adopt in 2012 the Program for Infrastructure Development In Africa (PIDA) [17]. In its priority action plan the PIDA has highlighted the importance of establishing Internet Exchange Points. It is in this context that the African Union Commission initiated the African Internet Exchange System (AXIS) project to promote keeping of intra-Africa's internet traffic within the continent by supporting the establishment of National and Regional IXPs in Africa [10]. Through the support of the African Internet Exchange System (AXIS) project, AU Member States with internet exchange points (IXPs) have increased from eighteen in 2013 to thirty-eight in 2017, in 29 countries [18].

We collected the names/urls of all the thirty-eight African IXPs and used reversed DNS lookup to compute their IP addresses. We then use their IP addresses to check the appearance of the IXPs in our router-level traces. We also checked the appearance of IXP links in the traces, where *a link is considered to exist between two IXPs if their network address (subnet mask/16) appeared together in one trace and are adjacent*. From these links, we constructed an IXP-level connectivity graph for Africa. Since most of the IXPs were built after 2015, we only consider 2016 traces.

Although many of the IXPs were observed in the traces (thirty-two among the thirty-eight), occurrences of links were very rare. We only found 40 links that were mostly between IXPs in coastal countries and specially in the south side of the continent. As a consequence, the IXP-level connectivity graph was disconnected with many isolated nodes. This suggests that the additional IXPs built in the continent are still not connected. This is mainly due to the lack of cross-border links, specially for landlocked countries.

Since the IXPs are not connected among themselves, there is still the need to go through transit ISPs for communication between most of African countries. In other terms, despite the construction of new IXPs inside the continent, boomerang routing is still present. In our opinion, the optimal solution for the AXIS project is to combine the building of IXPs with that of links between them. To illustrate that, we use the IXP connectivity graph built earlier and consider the following optimization problem: given that n (cross-border) links are to be added between neighboring countries in Africa, where should we add them to maximally reduce the fraction of boomerang routes? How much this should reduce the fraction of boomerang? Notice that n is a parameter that takes different values. In this paper, we consider $n = 1, 2, 3, 4$.

As a first step, we use a naïve approach in which we consider that all links have the same cost (which is certainly not the case). Our analysis shows that by adding only one links, we reduce 0.7% of the boomerang. With two links, we are able to reduce 5% of the boomerang, while with 3 links, 15% of the boomerang are removed. With four additional links, 27% of the boomerang are corrected.

The analysis above shows that adding only IXPs without links between them will not solve the boomerang problem in the continent. We believe that the effort of constructing new IXPs should be backed with an effort to building high speed links between them.

6 Conclusion

In this paper, we use publicly available traceroute data to study the prevalence of boomerang routing in intra-African Internet communication. The analysis has shown that in 2016, still 60% of traffic between two African countries has to go through Europe or the US before reaching their final destination. This raises issues with regards to cost, delay, as well as information exposure. We have also analyzed the delay profile of the communication paths and have discovered that, despite the boomerang, the traffic that stays within the continent experiences larger delays. Finally, we have used our traces to build a connectivity graph between the existing Internet Exchange Points (IXPs) in the continent. Our study has shown that the African IXPs are not connected. This indicates that despite the effort spent in building these IXPs, the boomerang phenomenon remains current in the continent. Consequently, we believe that efforts by the African Union (AU) to support individual member states to build IXPs should be backed with parallel effort to build high speed cross-border links between the countries. Related to this, we plan to follow up this paper with a study on how to optimally determine where to build those links by considering cost, regulation and geographical constraints.

Acknowledgements. This work was partially accomplished under NIST Cooperative Agreement No. 70NANB16H024 with the University of Maryland. The authors thank Dr. Peter Mell (CSD/NIST) for his support (time and advice). The authors would like to thank Christopher Schanzle (ACMD/NIST) for his help in gathering the data.

References

1. McKinsey Global Institute: Lions Go Digital. The Internet's Transformative Potential for Africa, 2013 Annual Report. <https://www.mckinsey.com/industries/high-tech/our-insights/lions-go-digital-the-internets-transformative-potential-in-africa>. Accessed 03 Dec 2017
2. Guerriero, M.: The impact of internet connectivity on economic development in Sub-Saharan Africa. EPS Speaks, January 2015
3. World Bank (WB): Information and communications for development 2009: extending reach and increasing impact. <https://issuu.com/world.bank.publications/docs/9780821376058>. 15 Dec 2017

4. ITU: White Paper on Broadband Regulation and Policy in Asia-Pacific Region: Facilitating faster Broadband Deployment, November 2016
5. Internet.org by Facebook: The State of Connectivity 2015–2016: A Report on Global Internet Access <https://fbnewsroomus.files.wordpress.com/2016/02/state-of-connectivity-2015-2016-02-21-final.pdf>. Accessed 03 Dec 2017
6. Edmundson, A., Ensafi, R., Feamster, N., Rexford, J.: Characterizing and avoiding routing detours through surveillance states. ArXiv. <http://arxiv.org/abs/1605.07685>. Accessed 31 Dec 17
7. Obar, J.A., Clement, A.: Internet surveillance and boomerang routing: a call for canadian network sovereignty. In: Ross, P., Shtern, J. (eds.) Proceedings of the Technology and Emerging Media, Track—Annual Conference of the Canadian Communication Association, TEM2013, Victoria, 5–7 June 2012 (2013). <http://www.tem.fl.ulaval.ca/fr/victoriaP2013>
8. CAIDA: Macroscopic Internet Topology Data Kit. <http://www.caida.org/data/internet-topology-data-kit/>. Accessed 13 Dec 2017
9. Maxmind geolocation service, December 2017. <https://www.maxmind.com/en/home>
10. The African Internet Exchange System (AXIS) Project. <https://au.int/en/african-internet-exchange-system-axis-project-overview>. Accessed 24 Dec 2017
11. ITU: Are poor countries subsidizing the rich? The Issues of International Internet Connectivity. <https://www.itu.int/itunews/manager/display.asp?lang=en&year=2005&issue=03&ipage=internerconnectiv&ext=html>. Accessed 03 Dec 2017
12. The Halfway Proposition, Background Paper on Reverse Subsidy of G8 Countries by African ISPs. In: Conference of African Ministers of Finance, Planning and Economic Development, Johannesburg, South Africa, 19 October 2002. www.afrispa.org/HalfwayDocs/HalfwayProposition_Draft4.pdf
13. Huffaker, B., Fomenkov, M., Claffy, K.: Geocompare: a comparison of public and commercial geolocation databases. In: Proceedings of NMMC, pp. 1–12 (2011)
14. Fanou, R., Francois, P., Aben, E.: On the diversity of interdomain routing in Africa. In: Mirkovic, J., Liu, Y. (eds.) PAM 2015. LNCS, vol. 8995, pp. 41–54. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15509-8_4
15. Chavula, J., Feamster, N., Bagula, A., Suleman, H.: Quantifying the effects of circuitous routes on the latency of intra-Africa internet traffic: a study of research and education networks. In: Nungu, A., Pehrson, B., Sansa-Otim, J. (eds.) AFRICOMM 2014. LNICST, vol. 147, pp. 64–73. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16886-9_7
16. African Union: Study on Harmonisation of Telecommunication, Information and Communication Technologies Policies and Regulation in Africa, March 2008. http://www.itu.int/ITU-D/projects/ITU_EC_ACP/hipssa/docs/2_Draft_Report_Study_on_Telecom_ICT_Policy_31_March_08.pdf
17. Program for Infrastructure Development in Africa. <http://www.au-pida.org/>. Accessed 02 Jan 2018
18. The African Peering & Interconnection Forum (AFPIF) 2017 Conference Report. <http://www.afpif.org/afpif2017/>. Accessed 02 Jan 2018



Ego-Community Evolution Tracking in Instant Messaging Networks

Ahmed Ould Mohamed Moctar^(✉) and Idrissa Sarr

Department of Mathematics and Computer Science, Cheikh Anta Diop University,
Dakar - Fann, BP 5005, Dakar, Senegal
`amed.mohameden@gmail.com`, `idrissa.sarr@ucad.edu.sn`

Abstract. Community detection is one of the most topics that are covered by social network analysis researchers. Early works focused mainly on partitioning networks into several global communities before they target communities evolution over time. The main drawback of such approach is the difficulty to obtain the entire network on which we may process a set of algorithms to track evolution. As a result, many researchers focused on detecting and tracking local dynamic communities.

Basically, detecting and tracking local dynamic communities is a process that moves from a set of nodes to build community structure followed by a structure evolution tracking. A particular type of local communities is called “ego-communities”. Unlike local communities, ego-centered ones have the advantage of being able to expand the community according to the neighborhood step of interest node. They allow, among others, to better identify and track the network elements activities.

Existing ego-community detection algorithms do not support directed networks and are designed for static networks. This failure led us to propose a new solution allowing to detect and track ego-community evolution in directed, weighted and dynamic social networks. We present some illustrative examples to explain the working principle of our solution.

Keywords: Ego-community · Dynamic networks · Instant messaging

1 Introduction

The development of on-line social media has created many opportunities to communicate, access, and share information from anywhere and at anytime. The kind of application such as *Viber*, *WhatsApp*, *Imo*, *Line*, as well as *Facebook* affords plenty of possibilities for getting in touch with friends, colleagues, and relatives at every moment with real-time messages, photos, videos, etc. Data collected from those applications integrate the time from which users send and/or receive data. Therefore, it is worthwhile to build a social network based on instant messaging content to find out “who talk to whom” and/or “who is closed to whom”. Such a network that we call Instant Messaging Network (IMN) can be represented as a graph where individuals are the nodes and their interactions the links. In

such a network, finding out the sub-groups of individuals, *a.k.a* communities, is worthwhile. In fact, detecting communities within a network affords opportunities to place network elements in classes associated with specific needs or characteristics. For example, identifying the group of people who had a contact with an Ebola patient helps to identify the target either to be monitored or to be informed about the exposed danger. However, in a dynamic social network, such a detection may be challenging at many points. For instance, how to detect changes within a community structure and when they occurs is not trivial.

In this study, we rely on IMN which is an illustration of dynamic social networks. To detect dynamic communities in IMN, we use the snapshots approach which relies on capturing a series of snapshots whose each one is aggregated over a time window. For each snapshot, one may apply a set of processes and finally compare the resulting communities with the ones obtained in the other snapshots. This strategy is widely used in dynamic community studies in order to track changes over time. In short, we plan to face three challenges, namely:

1. Data gathering and interpretation;
2. Time window size management;
3. Monitoring the evolution of dynamic communities.

The first challenge is the extraction and interpretation of data from instant messaging platforms, which provide huge and various information with a high velocity. Actually, in many studies, data contents are extracted and analyzed for unveiling knowledge. Important parts of these approaches rely on semantic rules, and/or more simply, on keywords to identify the network structures (nodes and their links). However, when it comes to deal with dynamic networks, it is worthwhile to integrate the fact that links can either vanish or intensify over time. That is, everything (nodes as well as links) must be set as dynamic. Moreover, based on the exchange flows, a link direction may switch over time and leads to many changes in the community structure. Therefore, instant messaging networks should be modeled as a directed and weighted dynamic graph in order to deal with the frequency and the direction of interactions over time.

The second challenge is how to manage the size of time window which has a real impact on a community structure evolution since it determines the data that belong to each snapshot. In other words, a bad time windows setting can lead to miss interesting structure changes. To manage the time window size, we define a quality function in order to quantify the changes that the network undergoes during a given period. If the quality difference between two times exceeds a given threshold, this means that the network has undergone a considerable changes, and then so, we make a new snapshot.

The third challenge is the monitoring of dynamic community evolution. To handle this aspect, we propose a solution, based on snapshots approach, for detecting and tracking the evolution of dynamic ego-community over time. It consists in selecting some nodes of interest among the most important nodes in network. Then, it seeks to track changes in the ego-communities over time.

In brief, our solution proceeds in 3 phases. First, it decomposes the network evolution into several snapshots. Next, it detects in each snapshot the static

ego-communities based on the algorithm we proposed in [5]. Finally, it applies a dynamic tracking algorithm to interpret the communities evolution over time. In our best knowledge, there is a glaring lack of studies that target the challenges we pointed out while we count hundreds of works in community detection topic.

The remainder of this paper is organized as follows. First, we present, in Sect. 2, the challenges and problem statement as well as the data gathering and interpretation. Next, we discuss, respectively, in Sects. 3 and 4 time window size management and monitoring the evolution of dynamic communities. Then, we present, in Sect. 5, some illustrative examples in order to illustrate the working principle of our solution before concluding in Sect. 6.

2 Network Set-Up

To model the instant messaging network, we use a set of nodes and links attributes explained below. In this perspective, a set of nodes constitutes the network initial composition, and when an existing node communicates with a new element, the latter it's added as a new node in the network, provided with the corresponding link. The links weights indicate the communications number between nodes. An interaction between two existing nodes implies the appearance of a new link. This is how the network grows over time.

2.1 Challenges and Problem Statement

The design of a community detection method unveils several questions that one can summarize in two categories, namely, supported networks and detected community properties. Several works have been proposed in order to find a method able to face the challenges pointed out above [1–4, 6]. However, in our best knowledge, there is no method allowing to handle all these issues.

Our goal is to propose a solution that supports dynamic, directed and weighted networks and allows to detect and track the evolution of overlapping local ego-communities.

2.2 Nodes Attributes

Each node is identified by the following information:

- Identifier: this attribute uniquely identifies each user (*e.g.*, phone number, ID of the user profile, etc.);
- Instant of appearance: this attribute indicates the moment where a node comes into the network.

2.3 Links Attributes

Each link is characterized by:

- Direction: It shows who initiates the interaction and who reacts;

- Duration: It represents the time that lasts each interaction. If the communication is instant, we set a short-time threshold beyond which we consider that the interaction is over;
- Weight: We compute weight depending on a set of durations so that the link weight during a time window is equal to interactions number multiplied by the sum of interactions durations.

2.4 Operations of Network Evolution

Related works analyze the networks evolution according to a macroscopic vision. That's to say, they focus only on community changes. In fact, existing works suffer from two weaknesses: *(i)* they neglect the communication intensity and are limited to the topological changes; *(ii)* they interpret the community evolution according to a global scope, in other words, they can't interpret the microscopic changes of network (nodes and links changes).

The approach we propose overcomes these two weaknesses as we introduce the communication intensity aspect into the definition of network evolution operations. In addition, we define 5 nodes/links evolution operations, namely:

- Incoming node (IN): a node comes into the network if it receives a communication from a node of network;
- Vanishing node (VN): a node is considered as vanishing if its communication rate is equal to zero during a given time window;
- Incoming link (IL): it reflects the first time two nodes start exchanging;
- Vanishing link (VL): a link is considered as vanishing if the communication between nodes ends or also if one of the link endpoints vanishes;
- Growing link (GL): a link with an increasing weight, thanks to the higher communication rate of the corresponding nodes at a given time window.

3 Time Window Size Management

In this section, we propose a new method allowing to decompose the dynamic network evolution into several snapshots, each of which represents the state of the network during a given period. Our method works periodically as it consists of capturing snapshots by period. Note that the period is not fixed, it depends on degree of changes network over time.

The method we propose consists of two parts:

1. a variant, whose scope is global, of the quality function that we proposed in [5]. This variant serves to measure the state of network during a given period taking into account the links cohesion and the communication intensity between nodes;
2. a strategy, based on the quality function, to decompose the dynamic network evolution into several snapshots.

In the following, we present, in Sect. 3.1, our global quality function. Then, we explain, in Sect. 3.2, how we decompose the evolution of dynamic networks.

3.1 Network Quality

Let \mathcal{N} a dynamic network and \mathcal{N}_t a network capture at time t . To get an idea of the communication intensity between nodes within \mathcal{N}_t , we compute the inverse of the weights sum of all links $\frac{1}{\sum w_{\mathcal{N}_t}}$. The intuition behind is that the more the communication is intense within \mathcal{N}_t (high value of weights), the more $\frac{1}{\sum w_{\mathcal{N}_t}}$ tends to 0.

Regarding to the internal cohesion of \mathcal{N}_t , we take again the idea of the local quality function: the more the topological structure of the network at instant t approaches a clique, the more the snapshot is considered dense. Therefore, the proportion $\frac{|V_{\mathcal{N}_t}|}{|E_{\mathcal{N}_t}|}$ allows to evaluate at what level the snapshot \mathcal{N}_t is cohesive. $|V_{\mathcal{N}_t}|$ designate the nodes number of \mathcal{N}_t and $|E_{\mathcal{N}_t}|$ the links number.

This, our global quality function is defined as follows:

$$\psi(\mathcal{N}_t) = \frac{1}{\sum w_{\mathcal{N}_t}} \times \frac{|V_{\mathcal{N}_t}|}{|E_{\mathcal{N}_t}|} \quad (1)$$

Note that we choose the multiplication between tow parts of the quality function for weighting the communication intensity to the cohesion between nodes. The overall complexity of $\psi(\mathcal{N}_t)$ is approximately equal to $O(n + 2m)$.

3.2 Dynamic Network Evolution

The optimal size of the time window should enable us to decompose the network evolution into several snapshots, each of which includes an interesting number of changes. To this end, we propose a strategy that works in two phases:

1. First, we rely on the previous global quality function to measure the changes that network has undergone during a given period;
2. Next, we model the changes rate across a threshold ε . If the quality difference of the network between two instants exceeds the threshold value, it means that the network has been considerably changed.

Formally, we consider that the network has undergone a considerable change if the difference between its quality at time t and that at instant $t + 1$ becomes greater than a given threshold ε :

$$|\psi(\mathcal{N}_t) - \psi(\mathcal{N}_{t+1})| \geq \varepsilon \quad (2)$$

Note that increasing the threshold value implies the abundance of changes between two instants. In this paper, we are not interested in the way that the right threshold should be determined. The choice of threshold value will be addressed in our future work.

To predict the next instant $t + 1$, we use a time series (built from the quality scores over time) that we expose in another paper.

4 Tracking the Changes of Dynamic Ego-Community

The question we have to address first is how do we track the changes: should we do it in a continuous manner or a discontinuous one? In our case, we decide to use the discontinuous fashion since we do not aim to have the trace of every single change but the overall change in some particular points. The first problem raised is to define these particular points. After creating the snapshots, we detect, on each one, the community structure using a static ego-community detection algorithm that we detail in another paper. Its working principle consists in optimizing a quality function that allows to evaluate the relevance of an ego-community according to its internal cohesion, on the one hand, and the communication intensity between its nodes, on the other hand.

Basically, we take a snapshot for each time point and see how the ego-communities evolve over time. To this end, we apply our static ego-community detection algorithm in a sequential manner on each snapshot. The outcome of the snapshot at t is compared with the one of $t + 1$ in order to identify changes if ever.

As in the literature, the operations indicating the evolution of dynamic communities are categorized in seven classes: *growth*, *contraction*, *continuing*, *division*, *merging*, *birth* and *death*. In opposite with the existing solutions, we do not just limit our analysis on finding the class of the evolution, but also we clarify what kind of changes in terms of nodes (IN or VN), links (IL and VL), as well as communication intensity (GL). The reason of doing so is to be able to explain what factors underline the evolution. In our context, this strategy is very useful since an evolution of a community is not always conducted by the topological aspects but with also the behavior of a node regarding to an ego node during a time window.

Moreover, we found interesting to introduce vanishing node (VN) in order to identify nodes that stay inactive for a long time. However, an ego node cannot be considered as a VN since it can stop communicating with its neighbors that keep exchanging between them. However, a community disappears when the number of its nodes is lesser than a fixed number. For example, if we consider that a community is composed at least of n nodes, then an ego community will disappear if we have less than $n - 1$ nodes.

5 Application Case

In this section, we present, first of all, an illustrative example in order to clarify the role of our static ego-community detection algorithm. Next, we show an example of dynamic ego-community tracking over time. In these examples, we detect 2-steps ego-communities. Let $C_{\{u,2\}}$ the 2-steps ego-community of interest node u and $S_{\{u,2\}}$ the seed used to build the ego-community.

5.1 Static Ego-Community Detection

On Fig. 1, we present the initial structure of a dynamic, oriented and weighted network, consisting of 33 nodes and 61 links. The weights values are shown on

links. As for the interest nodes, we choose 4 thanks to their central positions, namely, the nodes 1, 10, 17 et 27, denoted, respectively, by a , b , c et d .

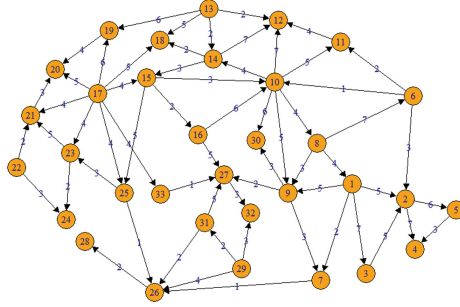


Fig. 1. Initial structure of a dynamic network, oriented and weighted.

Table 1 shows the 2-steps a 's ego-community detection procedure. In this procedure, we considered that the seed of a 's ego-community includes the whole 2-steps neighborhood. Then, at each iteration, we calculated the affinity degree¹ of the ego's neighbors and we chose the neighbor having the smallest value of affinity degree. Such node is removed from the community if its removing decreases the quality score; otherwise, we keep it and move on to the next neighbor. Following this process, we noticed that the removing of nodes 26, 27 and 10 implies a decrease in the quality score unlike the other nodes. In Fig. 1, we see that node 26 has been removed since it is loosely linked to the nodes of a 's community. We also note that nodes 10 and 27 are more connected to the outside as to nodes of a 's community, hence the reason for their removing.

After building the 2-steps a 's ego-community, the algorithm repeats the same process to detect the ego-communities of nodes 10, 17 and 27. Figure. 2 shows the detected ego-communities. Note that a node can belong to several ego-communities as our algorithm supports overlap.

5.2 Tracking Ego-Communities Evolution

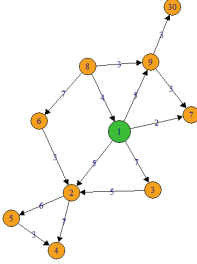
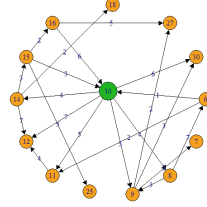
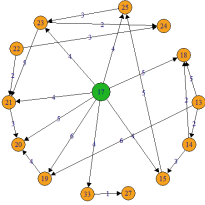
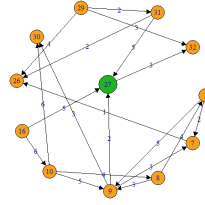
In this part, we illustrate our method of tracking the evolution of dynamic ego-communities. To this end, we consider two snapshots captured, respectively, at the time t and $t + 1$, represented on Fig. 3. The snapshot t represents the initial state of the evaluation network already shown in Fig. 1 and whose communities are presented individually on Fig. 2. We present on Fig. 3a all detected ego-communities in snapshot t to highlight the overlap aspect.

On Fig. 3b, we find that the microscopic changes (relative to nodes or links) that took place at time $t + 1$ are:

¹ To select the ego's neighbors, we proposed a new metric allowing to classify the nodes based on the number of their neighbors in ego-community as well as on their weights of adjacent links. This metric is detailed in another paper.

Table 1. Illustration of the procedure of 2-steps a 's ego-community detection.

Community current nodes	Selected node	Affinity degree	Quality score	Decision
$S_{\{a,2\}} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 26, 27, 30\}$			0.930	—
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 26, 27, 30\}$	26	0.02	0.903	Removed
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 27, 30\}$	27	0.025	0.876	Removed
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 30\}$	10	0.173	0.809	Removed
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	30	0.166	0.842	Kept
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	9	0.261	0.144	Kept
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	7	0.111	0.922	Kept
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	8	0.305	0.940	Kept
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	6	0.057	0.938	Kept
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	3	0.208	1.004	Kept
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	2	0.2	1.513	Kept
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	4	0.15	1.051	Kept
$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$	5	0	0.972	Kept
Detected ego-community: $C_{\{a,2\}} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 30\}$				

(a) 2-steps a 's ego-community.(b) 2-steps b 's ego-community.(c) 2-steps c 's ego-community.(d) 2-steps d 's ego-community.**Fig. 2.** 2-steps ego-communities detected by our algorithm on the network evaluation.

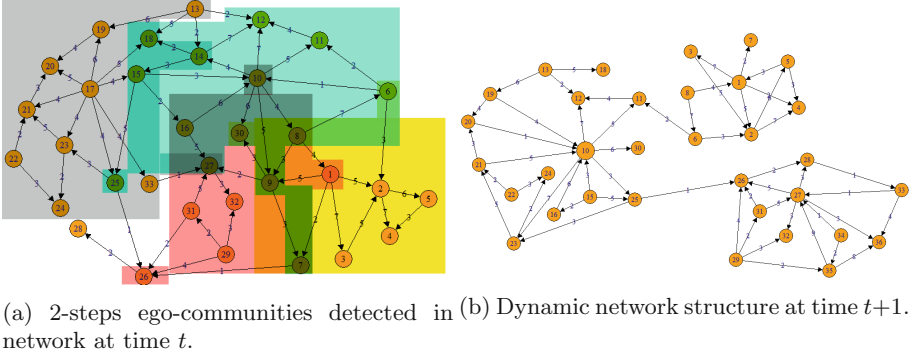


Fig. 3. Two successive snapshots of our dynamic evaluation network.

- 03 vanishing nodes: 9, 14 et 17;
- vanishing of links (16, 27) and (10, 8) as well as all links that were adjacent to vanishing nodes;
- 03 incoming nodes: 34, 35 et 36;
- 16 incoming links: (5, 1); (1, 4); (27, 26); (29, 35); (27, 28); (28, 33); (33, 36); (35, 36); (34, 35); (27, 36); (35, 27); (34, 27); (20, 10); (10, 23); (21, 10); (19, 10).

As for the macroscopic changes, after having applied our dynamic ego-community tracking algorithm, we found 3 ego-communities and 5 kinds of evolution:

1. A contraction with $C_{\{a,2\}}$ loosing two nodes: 9 et 30;
2. A growth with $C_{\{a,2\}}$ gaining two new links (5, 1) et (1, 4). Thus, $C_{\{a,2\}} = \{1, 2, 3, 4, 5, 6, 7, 8\}$;
3. A death of $C_{\{c,2\}}$;
4. A growth with $C_{\{b,2\}}$ after gaining the nodes that were in the community of $C_{\{b,2\}}$. Therefore, $C_{\{a,2\}} = \{10, 11, 12, 13, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 30\}$;

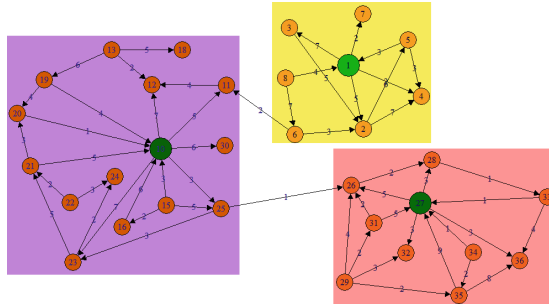


Fig. 4. 2-steps ego-communities detected by our algorithm in snapshot $t+1$.

5. A growth with $C_{\{d,2\}}$ thanks to the arrival of new nodes and links. Thereby $C_{\{d,2\}} = \{26, 27, 28, 29, 31, 32, 33, 34, 35, 36\}$.

In Fig. 4, we portray the new community structures of the snapshot $t + 1$.

6 Conclusion

The objective of this paper is to propose a new method able to track the evolution of dynamic ego-communities in instant messaging networks. To this end, we consider data from an instant messaging platform such as Facebook and build a social network based on “who talk to whom” and/or “who is closed to whom”. We aim therefore at finding out communities centered on special nodes due to their characteristics or social positions. This is very useful since identifying individuals that are closely exposed to a disease is a starting point for controlling an epidemic. With this insight, finding out nodes that share a community with an infected individual may help to point out who is exposed or not and where it is more relevant to make specific actions to break down the disease spread.

References

1. Danisch, M.: Mesures de proximité appliquées à la détection de communautés dans les grands graphes de terrain. Ph.D. thesis, Paris 6 (2015)
2. Hou, J., Yu, Z., Hong, X., Wang, L.: Social circle detection based on micro-blogging topic and user follow relationship. In: 2016 International Conference on Asian Language Processing (IALP), pp. 222–227. IEEE (2016)
3. Lan, C., Yang, Y., Li, X., Luo, B., Huan, J.: Learning social circles in ego-networks based on multi-view network structure. *IEEE Trans. Knowl. Data Eng.* **29**(8), 1681–1694 (2017)
4. McAuley, J., Leskovec, J.: Discovering social circles in ego networks. *ACM Trans. Knowl. Discov. Data (TKDD)* **8**(1), 4 (2014)
5. Moctar, A.O.M., Sarr, I.: Ego-centered community detection in directed and weighted networks. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM 2017, New York, USA, pp. 1201–1208. ACM (2017)
6. Rees, B.S., Gallagher, K.B.: EgoClustering: overlapping community detection via merged friendship-groups. In: Özyer, T., Rokne, J., Wagner, G., Reuser, A. (eds.) *The Influence of Technology on Social Network Analysis and Mining. LNSN*, vol. 6, pp. 1–20. Springer, Vienna (2013). https://doi.org/10.1007/978-3-7091-1346-2_1

Investigation of Degradation in Crystalline Silicon Photovoltaic Modules After 10 Years Exposition in Senegal by Infrared (IR) and Electroluminescence (EL)

Ababacar Ndiaye^{1,2(✉)}, Cheikh M. F. Kébé², Boudy Ould Bilal^{2,4},
Abdérafi Charki³, Vincent Sambou², and Papa A. Ndiaye²

¹ UFR – Sciences et Technologies, Département de Physique,
Université Assane Seck de Ziguinchor, BP 523, Ziguinchor, Senegal
ab.ndiaye@univ-zig.sn, ababacar.ndiay@gmail.com

² Centre International de Formation et de Recherche en Energie Solaire (CIFRES),
Ecole Supérieure Polytechnique – UCAD, BP 5085, Dakar-Fann, Senegal

³ Université d'Angers–ISTIA–LARIS,
62 Avenue Notre Dame du Lac, 49000 Angers, France

⁴ Ecole Supérieure Polytechnique, BP 5259, Nouakchott, Mauritania

Abstract. The effects of a sub-Saharan coastal climate on PV modules degradation was studied in this paper. A Mono and a polycrystalline-silicone solar PV module exposed in Dakar, dry and coastal climate, at the extreme West of Senegal was studied. As first inspection in this region, the electrical parameters of two PV modules A and B operated during about 10 years, are measured under standard testing condition (STC) and their I-V characteristics were fitted. The initial I-V characteristics was performed under real conditions and translated to STC and compared to the measured I-V characteristics at standard test conditions (STC) obtained in PV Laboratory after exposition. After the operating years, the main important parameters of the studied PV modules: short-circuit current I_{SC} , open circuit voltage U_{OC} , maximum power P_{MPP} , nominal current I_{MPP} and Voltage $VMPP$ are evaluated and then compared to the initial parameters obtained during initial exposition to estimate their degradation.

Moreover, the defects that affected the PV module are explored by visual inspection, Electroluminescence (EL) and thermography (IR) imaging methods. The results show absolute degradation of maximum power (ΔP_{MPP}) nearing -5.35% and -2.92% for the mono and polycrystalline silicon operating about 10 years. The inspection reveals many degradations in both modules. Most of the degradations due to the climate are found in the mono and a very advanced I_{SC} mismatch is found with the IR image. The polycrystalline has many mechanical defects that's does not too much affect the performance characteristics.

Keywords: Photovoltaic · Degradation · Performance
Electroluminescence imaging · Thermography imaging

1 Introduction

The use of Renewable Energies becomes more and more necessary for the satisfaction of the world Energy demand, with less negative effects on the environment. Global warming and fossil fuels depletion force the world to seek for solutions to ensure our survival and those of future generations. Africa in particular is facing a critical energy scenario. According to the World Bank, 25 countries in sub-Saharan Africa are in energetic crisis. Only 32% [1] of the population has access to electricity in 2012 with a very high average price of US \$ 0.13 [2] per kilowatt hour.

However, the continent can rely on its significant renewable energy resources to improve its situation. Several renewable energy sources are available in different localities. East Africa is known for its large geothermal deposit. The extremities are marked by the presence of remarkable wind potential with an estimated power of 1,300 GW. Hydropower with an available capacity of 238 GW is the most exploited clean energy in the continent. But solar energy remains the most abundant source on the continent. Africa is the sunniest continent of the world. An average radiation of 2650 kWh/m²/year, with an estimated sunshine duration of 3500 h/year could enable the continent to satisfy its energy needs.

Several energy projects based on solar photovoltaic are planned or executed on the continent. Nevertheless, for better exploitation of solar resources, adaptation of technologies to the African environment is essential. The technical characteristics of the existing solar panels are given under the standard testing conditions (STC) corresponding to a mild climate of 25 °C and a sunshine of 1000 W/m².

These conditions are totally different from the ambient exposure conditions of panels in Africa. In this context, it is pertinent to identify, quantify and compare the major defects or expected modes of failure for the various climates that can be encountered in sub-Saharan Africa (arid, continental, wet or coastal monsoon) and which are Very different from the climate of central Europe.

The study is about degradation of the performance of mono and polycrystalline solar panels after about ten (10) years of exposure in a coastal climate precisely at the Polytechnic superior school of Dakar (ESP).

The analysis focuses on data collected using electroluminescence (EL) and infrared imaging methods.

2 Location and Platform Test Presentation

2.1 Presentation of the Experimental Environment

2.1.1 Temperature

The photovoltaic platform shown in Fig. 2 is used in this study. It is installed at Dakar in Senegal. Dakar is located on the extreme western Africa with geographic coordinates 14.61° North latitude and 17.37° West longitude. In Senegal, the climate is of sub-desert tropical type punctuated by damp summers and dry winters. On the other hand, the DAKAR region, which has an advanced position in the Atlantic Ocean, is characterized by a coastal microclimate. This is strongly influenced by the trade winds and the monsoon coming from the sea. On average, the temperatures are always high.

the average daily maximum temperature is 24 °C from January to March and between 25 and 27 °C in April, May and December. From June to October, temperatures reach 30 °C (Fig. 1) [3].

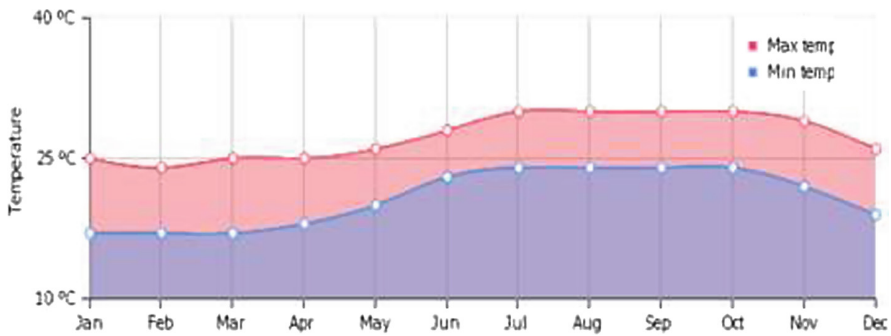


Fig. 1. Average min & max temperature in Dakar.

2.1.2 Radiation and Insolation

Radiation and insolation are key parameters, among others, in the quantification of the producible energy but also of the effects of radiation on the photovoltaic material. The radiation is expressed in kWh/m² while the insolation is expressed in hours. High radiation and insolation values correspond to very high temperatures and low values at low temperatures and vice versa. The average sunshine varies from 7.3 h/d during the rainy season when the sky is always cloudy at 9 h/d during the dry season when the sky is clear [4]. The highest average monthly value for radiation is 6.92 kWh/m²/d and is in the period from March to June, while the lowest is 4.57 kWh/m²/d corresponding to the months from July to February.

2.1.3 Humidity

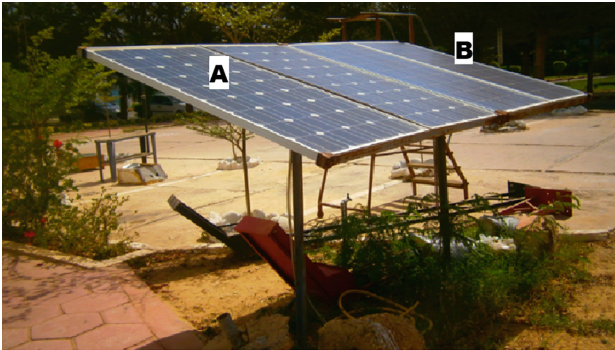
Variations in relative humidity depend in part on the air temperature and the hygro-metric characteristics of the air masses. The annual evolution of the relative humidity of the air is also tempered by the maritime influence and the annual average is around 70%. The highest values coincide with the heart of the rainy and low season in the months of April-May and October-December-January. Main climatic characteries are presented by Table 1 [5].

2.2 Platform and Modules Presentation

The platform used is composed of four (4) modules. Two monocrystalline manufactured by WAAREE in the left. The Two polycrystalline modules in the left are BP Solar products. One module on each technology have been chosen (A & B) to evaluate and analyze the degradation affecter about 10 years of operation. The two modules A & B fielded have been installed in same period 2007. The PV modules initial characteristics have been measured using the analyzer IV-400. The obtained IV curve initially designed and the characteristics obtained in exposition location and translated to STC are detailed for each module.

Table 1. Main climatic characteristics of Dakar.

Months	Temperature	Relative humidity	Daily solar radiation-horizontal
	°C	%	kWh/m ² /d
January	20.7	70.2%	4.89
February	20.7	74.9%	5.80
March	21.0	78.5%	6.57
April	21.4	83.0%	6.92
May	22.8	82.9%	6.71
June	25.6	82.3%	6.21
July	27.1	79.7%	5.60
August	27.4	83.0%	5.34
September	27.6	84.7%	5.34
October	27.6	81.8%	5.53
November	25.8	73.8%	4.98
December	23.4	68.6%	4.57
Annual	24.3	78.6%	5.70

**Fig. 2.** Photovoltaic platform installed in Dakar University.

A monocrystalline solar panel of 106 W_P (given by manufacturer) was used as a reference model for experimentation and simulation. However, the data used for degradation assessment are those measured on-site. The measurement carried out just after the installation of the system on the site, have been done under local conditions and automatically “translated” to the standard testing conditions (STC). The main performance characteristics are extracted are gathered in following table. The maximum current and voltage have been used to obtain the P_{MPP}. The maximum power obtained initially is equal to 114 W_P.

The second module used (B) is a polycrystalline also of 110 W_P according to the data given by the constructor. Like the module A, the initial measurements were established using the IV-400 analyzer. These measurements on site and “translated” to

the STCs determine the different performance parameters grouped in Table 2. With its 36 polycrystalline cells connected in series, the measured power on site for the poly is equal to 114 W_p.

Table 2. Technical characteristics of the PV modules.

Modules	Technology	Manufacturers	References	Characteristics	Values
Module A	Monocrystalline silicon	WAAREE	WS-110	Maximum power (P _{MPP})	114 W
				Nominal voltage (V _{MPP})	16.22 V
				Nominal current (I _{MPP})	7.04 A
				Open circuit voltage (U _{OC})	21.41 V
				Short-circuit current (I _{SC})	8.56 A
				Fill factor (FF)	62.20%
Module B	Polycrystalline silicon	BP Solar	WS-110	Maximum power (P _{max})	114 W
				Maximum voltage (V _{max})	17.31 V
				Maximum current (I _{max})	6.59 A
				Open circuit voltage (V _{OC})	21.76 V
				Short-circuit current (I _{SC})	7.43 A
				Fill factor (FF)	70.50%

3 Data Acquisition and Analysis Methodology

3.1 Data Acquisition

The main objective of this work is to evaluate and analyze the degradation on two crystalline modules exposed for about 10 years. But the main element in all that being to have the effects of the environment on the panel and the consequences generated on the electrical characteristics. For this, the main data used are: the electrical characteristics initially obtained and after the years of exposure, the EL and IR images.

- The initial Electrical performance characteristics are obtained using the analyzer IV-400. Photovoltaic module analyzer I-V 400 carries out the field measurement of the I-V characteristic and of the main characteristic both of a single module and of module strings. The acquired data are then processed to extrapolate the I-V characteristic at standard test conditions (STC). Output current or voltage from the module is measured with the 4-terminal method, which allows extending the measurement cables without requiring any compensation for their resistance, thus always providing accurate measures.
- IV curve and electrical characteristics obtained after exposition, have been measured using a lab solar simulator. The PASAN solar simulator is equipped with 4 Xenon flash tubes that generate a pulsed, calibrated and time-steady light. The light travels through a black tunnel and illuminates the module, which is positioned 8 m away on a uniformly illuminated 3 × 3 m surface. Different irradiance levels can be reproduced, however in our case the teste was carried out under STSs. A tracer records the electrical response of the module measuring up to 4000 points of the I-V curve, along with other electrical parameters (Fig. 3).

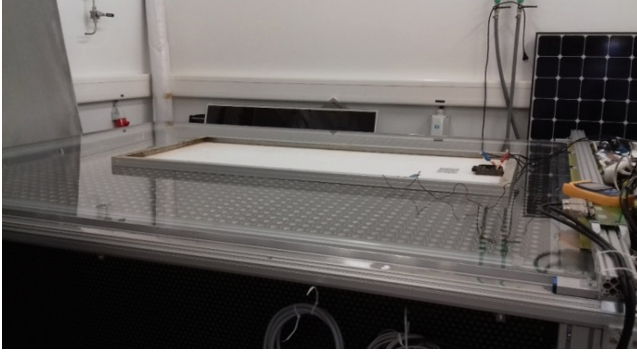


Fig. 3. IV measurements on the module.

- The set-up for PV EL Imaging is simple, consisting of: shortwave infrared (SWIR) camera with lens darkened enclosure solar cell/module Power Supply. An SWIR camera with In GaAs sensor and a spectral range from 0.9 to 1.7 μm was used in this study for capturing the EL image. Two different EL images have been made using different current. The first with a current equal to I_{SC} , the second was made with just $I = 10\% I_{SC}$ which allows to clearly identify the most affected surfaces (Fig. 4).

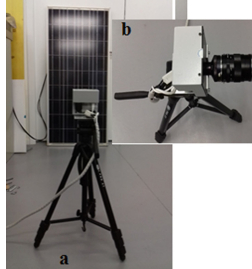


Fig. 4. (a) Module during EL imaging - (b) Close-up of the EL camera.

- For IR Imaging, the measuring device is a thermal imaging camera with a surface temperature measurement range of $-20\text{ }^{\circ}\text{C}$ to $250\text{ }^{\circ}\text{C}$ and a $31^{\circ} \times 22.5^{\circ}$ field of view, the user can view and save a quick snapshot of temperature patterns in any given area, thus quickly identifying unusual hot or cold spots. With $\leq 80\text{ mK}$ thermal sensitivity, temperature differences of just $0.8\text{ }^{\circ}\text{C}$ are visualized on the 3.5" cooler screen via the 160×120 -pixel thermal sensor. Three different color palettes give added functionality whilst hot and cold spots can be activated to immediately highlight hot and cold spots in the field of view. A digital read out of the surface temperature at the measurement point is shown alongside both the thermal image and the visible-light picture from the camera (Fig. 5).

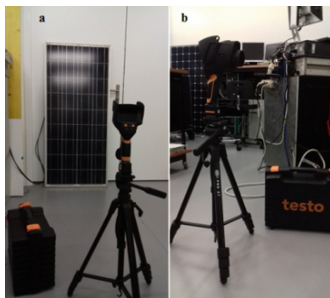


Fig. 5. (a) Module during IR imaging - (b) Close-up of the IR camera.

- The percentage of the surface affected by the different types of degradation is obtained with the GeoGebra software. GeoGebra is a dynamic mathematics software that combines geometry, algebra, spreadsheet, graph, statistics and infinitesimal calculus into a single easy-to-use software. It allows, according to a chosen scale, to quantify the selected surfaces and to put them in color. Thus, these surfaces in comparison with the total surface gives the percentage of the surface affected by the degradation.

3.2 Analysis Approaches

In this study, degradation phenomena observed in PV modules operating in the field for about 10 years are presented. Nondestructive diagnostic techniques including I-V curve analysis, infrared (IR) thermography and electroluminescence (EL) are employed here to assess PV performance and identify the defects. And the software Geogebra with the EL image give important information on the percentage of affected surface. Two Modules that have undergone different stages of degradation, from mechanical and manufacturing defects to severe visual degradation phenomena due to environmental effects, are analyzed, and performance characteristics degradation estimates are given for the 2 different technologies tested, revealing the need for a deeper understanding of PV degradation phenomena that occur under real conditions of operation. The diagnostic methodology of degradation is presented in Fig. 6.

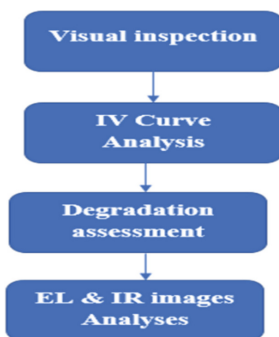


Fig. 6. Degradation diagnostic methodology.

4 Analysis of Electrical Characteristics

In this section, the aim is to assess, quantify and analyze the degradation of the two PV modules that operated during nearly ten (10) years in the tropical climate of Dakar located at the extreme west of Senegal. This locality, as described in Sect. 3 is marked by a tropical and semi-arid climate. We investigated the degradation of the main important electrical parameters (P_{MPP} , I_{SC} , U_{OC} , I_{MPP} , U_{MPP} and the FF) of two PV modules installed in the University of Dakar.

For further details in addition to the visual inspection, both modules were analyzed by Electroluminescence (EL) and Infrared (IR) imagery. This inspection allows to identify the different defects in the module, quantify the affected surface and analyze the temperature distribution to understand the different performance characteristics degradation that affected the two PV modules.

4.1 Photovoltaic Modules Degradation Assessment

4.1.1 Monocrystalline Module (A)

4.1.1.1 Characteristics After 10 Years of Exposure

After using the module for about 10 years, the performance characteristics were again measured. These measurements obtained at PV LAB in Switzerland are carried out directly in the STC with a simulator. The PASAN tester TC 1.1.3 used, is a pulsed solar simulator, for current-voltage (IV) characterization of photovoltaic modules. The results obtained are plotted in the curve IV in blue in Fig. 7 (Table 3).

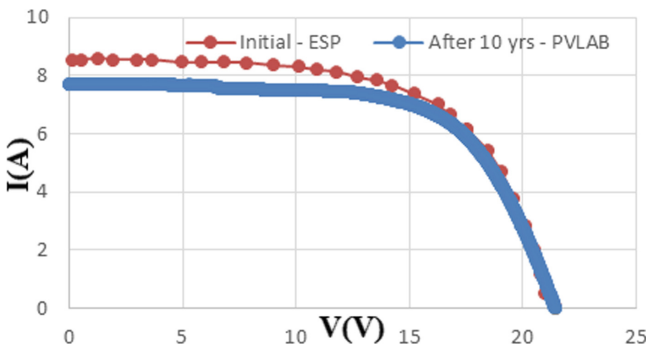


Fig. 7. I-V characteristics of monocrystalline module (A). (Color figure online)

Table 3. Performance characteristics comparison after about 10 years.

Measurements	P_{MPP} [W]	U_{OC} [V]	U_{MPP} [V]	I_{SC} [A]	I_{MPP} [A]	FF [%]
Initial - Measured at ESP	114	21.405	16.24	8.56	7.020	62.2
After 10 yrs - Measured at PVLAB	107.905	21.44	16.178	7.750	6.670	64.9

The realization of this curve IV in comparison with the initial curve already shows degradation of certain performance characteristics. The short-circuit current I_{SC} from 8.56 A to 7.75 A is the most affected characteristic. Therefore, the maximum power current I_{MPP} is affected. Initially the maximum power of 114 W obtained with I_{MPP} of 7.02 A decreases by 6.095 W. The maximum power current thus becomes 6.67 A.

4.1.1.2 Degradation Assessment

The measurements initially carried out with the natural sunlight and translated back to the STC with the IV400 analyzer; and those using a simulator in the STC conditions at PV LAB made are used to evaluate the different variations that occurs during the module exposition. The nominal (I_{MPP}) and short circuit (I_{SC}) current, the nominal (U_{MPP}) and the open-circuit (U_{oc}) voltage, and the PV module maximum power (P_{MPP}) degradation are evaluated by comparing each measured value after 10 years with the reference value obtained with the initial measurements given in Table 2. The degradation of these different parameters is expressed in percentage as a function of the difference between the initial normalized values and the obtained after exploitation [3].

The results obtained from the degradation calculations are summarized in the following graph (Fig. 8).

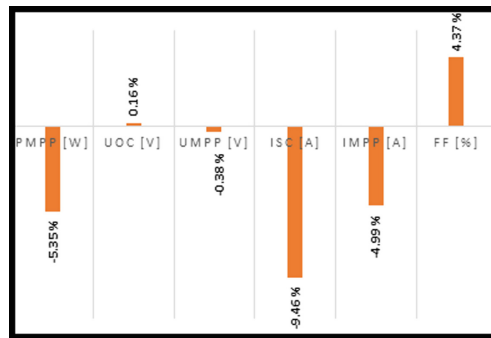


Fig. 8. Performance degradation of module A.

From Fig. 8, it can be seen that the open-circuit voltage is not affected because the calculated degradation is positive. The U_{MPP} also expresses a very slight degradation of just -0.38% . However, the nominal and short circuit current degradation is striking and equal to -4.99% and -9.46% respectively, which is generally caused by mismatched cells and increase sometimes series resistance increase. Such kind of IV distribution and I_{SC} reduction refer to irradiation decreases or shaded cells. In general, this is the consequences of operating area reduction or cells properties inhomogeneity. Also, a non-homogeneous temperature distribution could be expected by thermal imaging to confirm the aging of this module. The considerable degradation of I_{MPP} particularly affects the maximum power which is degraded by -5.35% . The fill factor degradation rate is also positive, meaning that the theoretical power degradation is more important than the maximum power diminution.

4.1.2 Polycrystalline Module (B)

4.1.2.1 Characteristics After 10 Years of Exposure

After 10 years of operation in hot and semi-arid climatic conditions, the PV module I-V curve were measured using pulsed solar simulator the comparison between measurements made initially and after exposition (Fig. 9) showed that, the short circuit current (I_{SC}) and the open circuit voltage (U_{OC}) did not change too much. However, the trend of the IV curve obtained after the operation years shows a slight decrease in the maximum power. In the literature [6, 7], such a change on curve IV is due to an increase in the series resistance R_S . This variation in resistance mainly affects the PV module operating voltage.

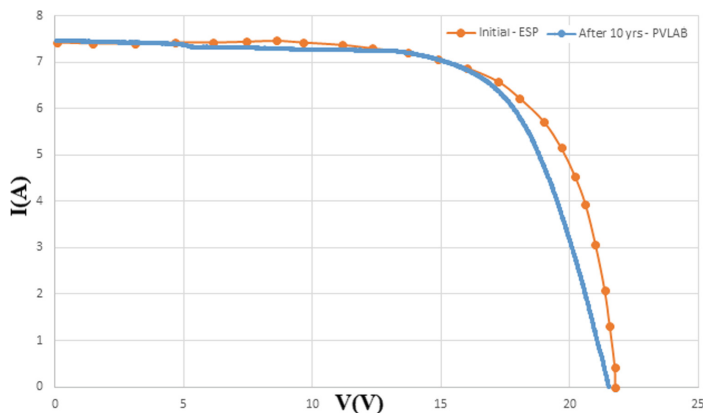


Fig. 9. I-V characteristics of monocrystalline module (B).

The comparison between the performance characteristics initially measured and that after the years of operation clearly shows that the nominal (I_{MPP}) and short-circuit (I_{SC}) current have not been affected. However, even if the variation is not too great, among all the parameters, voltage was the only electrical characteristic that deteriorated. The open-circuit voltage goes from 21.765 V to 21.5 V. The maximum power voltage has experienced the greatest variation, with an initial value equal to 17.3 V, it becomes 16.73 V, i.e. an absolute reduction of 0.6 V (Table 4).

Table 4. Performance characteristics comparison after about 10 years.

Measurements	P_{MPP} [W]	U_{OC} [V]	U_{MPP} [V]	I_{SC} [A]	I_{MPP} [A]	FF [%]
Initial - Measured at ESP	114.01	21.765	17.311	7.429	6.586	70.5
After 10 yrs - Measured at PVLAB	110.686	21.502	16.726	7.480	6.618	68.8

4.1.2.2 Degradation Assessment

The results obtained are summarized in Fig. 10. As expected with series resistance increase visible in the IV curve (Fig. 10), the maximum power voltage is the most affected characteristic, its absolute deterioration is -3.38% . Indeed, for a PV module exposed for almost 10 years, a degradation of only -2.92% of the maximum power is a good compatibility sign of this technology within the climate. However, it is always necessary to evaluate the different types of defects that affected the module to verify the material defects caused by the tropical climate. The Fill factor (FF) degradation is mainly due to the degradation nominal voltage degradation.

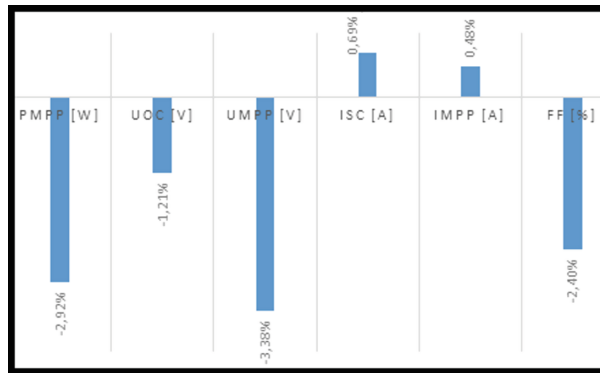


Fig. 10. Performance degradation of module B.

4.1.3 Comparative Study

The comparison between calculated degradation of the different characteristics for the two PV modules are summarized in Table 5. For the monocrystalline PV module (A), the short circuit current is the most affected characteristic with a degradation equal to -9.46% . The polycrystalline module (B) characteristics are slightly affected, but it is important to mention that the nominal voltage is the most affected parameter. This is to say that the drop in the maximum power output in the two PV modules does not have the same causes. The two modules than have different defects leading to such reduction of the maximum power.

The P_{MPP} degradation is equal to -5.35% for module A, and just -2.92% for module B. Thus, electrically the monocrystalline is more affected under such a climate.

Table 5. Performance characteristics degradation for the two PV Modules.

	P_{MPP} [W]	U_{OC} [V]	U_{MPP} [V]	I_{SC} [A]	I_{MPP} [A]	FF [%]
PV module A	-5.35%	0.16%	-0.38%	-9.46%	-4.99%	4.37%
PV module B	-2.92%	-1.21%	-3.38%	0.69%	0.48%	-2.40%

5 Analysis by Imaging Techniques

5.1 Module A: Monocrystalline

5.1.1 Electroluminescence Analysis (EL)

5.1.1.1 Degradation Types Identification

Defects in the material of the cell, micro-cracks, broken metallization, shunts, inactive regions are detected via Electroluminescence imaging in Fig. 11. This EL used to identify the defects is made under I_{SC} current. The fault types present in the modules are identified with different colors.

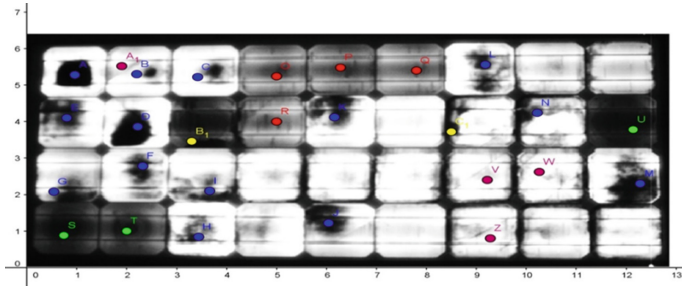


Fig. 11. Module A EL image with $I = I_{SC}$. (Color figure online)

- 15 cells (blue) have visual impairments such as discoloration and corrosion identified during visual inspection. These types of defects are directly related to climatic conditions mainly humidity.
- 2 cells with breakages which often occur during transport, packaging or installation of the modules. These types of mechanical damage are caused by insufficient packaging and vibration/shock.
- 4 cells with casting defects better known by the Contact finger interruptions. These types of defects are usually caused by insufficient screen printing during manufacturing. Because of their significant presence and non-uniform distribution, their remarkable effects on cells productivity are clearly evident.
- 4 other cells have microcracks, these types of defects such as breaks often occur during transport of the modules. The variation of the intrinsic manufacturing process can also be the cause of cells microcracks. But they are less serious and mostly invisible during visual inspection. However, they may cause deterioration in performance after several years of operation.
- 3 cells have unspecified dark areas. Several assumptions can be made about the origin of the dark shade: it may come from a local modification of the efficiency favoring non-radiative recombination of the charges, modification of optical properties of this zone and therefore the number of photons collected by the detector, or a modification of the contact resistance.

These assumptions mean that it is not possible to identify all the types of degradation present in an EL image on the simple visual criterion [5, 8].

In summary (Table 6), most of the defects identified on the module are due to the effects of the exposure climate conditions. 14 cells suffer from material degradation due mainly to contact of water with the cells. This water often present in moist air causes discoloration and corrosion. Given the climatic conditions of the operating site, it is evident that the high humidity of the hot location has effects on the cells. Indeed, Dakar is a very wet place with an annual average of 78.6%.

Table 6. Defects observed in module A.

Degradation type	Number of effected cells	Percentage relative to the total number of cell
Micro cracks	2	5.56%
Visual degradation	14	33.89%
Contact finger interruption	5	13.89%
Dark area	3	8.33%
Breakages	5	13.89%

The rest of the defects are for the most part mechanical. 6 cells have either microcracks or breaks due to poor handling during transport or laying. Manufacturing defects are fairly present. 4 cells with metallic hits severely affected are identified. 3 other cells have unidentified defects but often come from manufacturing.

5.1.1.2 Affected Area Evaluation

The use of the GeoGebra dynamic mathematics software allowed to quantify the total area affected by the different degradation types. The EL image used to determine the surfaces is the one supplied with a current equal to 10% of I_{SC} . An EL image taken at about 10% of the rated current of the photovoltaic module is more suitable for isolated cell parts quantification.

The maximum power (P_{MPP}) that the PV module can produce depends strongly on its productive surface as described in Sect. 1. Thus, the evaluation of the affected surface by the different types of degradation makes gives the opportunity to analyze the effects of productive surface decrease on the characteristics.

According to the chosen scale on the Fig. 12, the panel's total area is 75 cm². The fifteen (15) degraded areas evaluation gives a total dark surface of 20.97 cm² affected by the different types of defects cited above. Thus, 27.96% of the module's surface has been deteriorated. This materiel degradation of the surface leaded to a degradation of -5.35% in the maximum power (P_{MPP}). Then, the degradation of 1% of the surface causes a reduction of about 2% in the maximum power. Nevertheless, the black areas should not be interpreted as a totally zero emission, a black zone emits very little compared to the rest of the cell. And this is what causes electrical mismatches and reduce the I_{SC} .

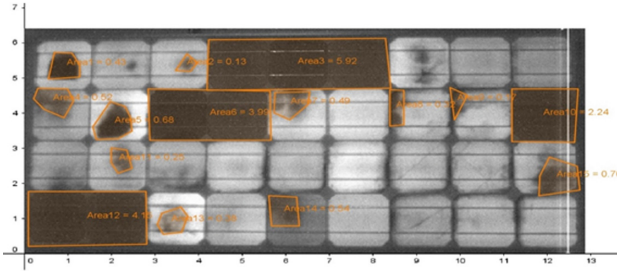


Fig. 12. Module A EL image with $I = 10\% I_{SC}$.

5.1.2 Analysis by Infrared Imaging (IR)

In a normal PV module, all the cells are considered identical. For a defect less module, the temperature distribution is homogeneous because all the cells have the same characteristics. Thus, PV module thermography image visualization is a very useful tool to state a PV module. Infrared imagery specifically identifies hot spots and cold spots. Hot spots indicate a decline in productivity in these areas, which dissipates the surplus current from other parts as heat. This occurs when the cell is totally or partially shaded, fractured or electrically defective. The severity of this degradation is directly related the cell's the temperature. The cold zones highlight the visual degradations. The areas affected by these types of degradation act as insulators. The monocrystalline module IR image in Fig. 13 shows a striking thermal disparity. Temperatures range from 0 °C to 62 °C. Several hot points or surfaces are visible; Which proves that the different types of degradations identified previously decrease the electrical current. The cold parts identified represent the visual defects highlighted in the EL image. This photovoltaic module exposed for about ten (10) years shows a great thermal inhomogeneity, meaning an advanced electrical mismatch. This justifies the significant I_{SC} degradation and the decrease of the I_{MPP} .

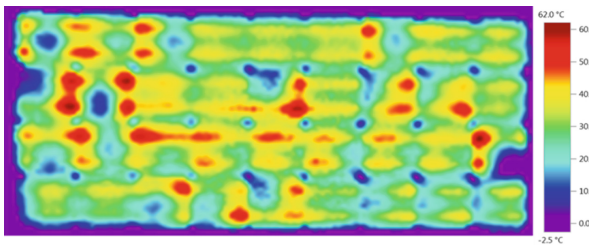


Fig. 13. Module A IR Image

5.2 Module B: Polycrystalline

5.2.1 Electroluminescence Analysis (EL)

5.2.1.1 Degradation Types Identification

The EL image for polycrystalline module was realized under the same conditions as the previous analyzed monocrystalline. For defects identification, the EL image used is realized under the short circuit current (I_{SC}) with different colors pointing fingers at the types of degradation present in the module (Fig. 14).

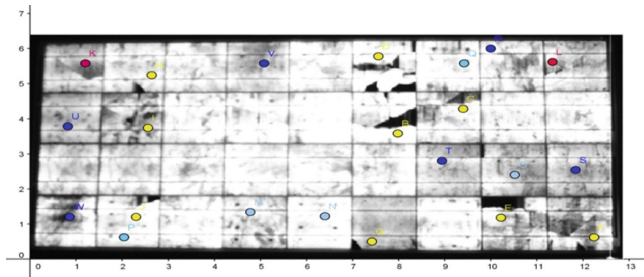


Fig. 14. Module B EL image with $I = I_{SC}$.

- 9 cells of the polycrystalline module experienced fairly extensive breaks which isolated the broken cell parts.
- 6 cells have visual defects due to the effects of the environment. During visual inspection, delamination of PV module's encapsulat has been identified. This probably justifies the presence of these dark surfaces indicating visual degradation.
- 5 cells have localized shunts. This type of defects is generally due to problems during cell's manufacturing (a particle, an anomaly in growth, etc.). It may also be a consequence of an external event such as electrostatic discharge or mechanical stress.
- 2 cells have micro cracks which elongated the list of mechanical defects.

For the polycrystalline module, mechanical defects due to physical constraints are the most present defects. Nine (9) cells in total have breakages often very advanced, and 2 others have micro cracks (Table 7).

Table 7. Defects observed in module B.

Degradation type	Number of effected cells	Percentage relative to the total number of cell
Micro cracks	2	5.56%
Visual degradation	6	16.67%
Localised shunt	5	13.89%
Broken cells	9	25.00%

Concerning the material or visual defects often due to the effect of the environment, they are present on 6 cells which are not nevertheless very affected. Indeed, since a polycrystalline module is not homogeneous in nature, it is sometimes difficult to identify the photoelectric effect. Manufacturing defects also are still present with many localized shunts.

5.2.1.2 Affected Area Evaluation

Using the same software, and with the EL image of the polycrystalline at $I = I_{SC}$, the results obtained show that the isolated surface is not very important. Seven (7) zones are isolated, representing a total area of 6.38 cm^2 . So only 8.5% of the module's surface was severely affected, causing a degradation of -2.92% of the maximum power (P_{MPP}). Thus 3.5% of the power decreased to 1% of affected area.

However, it has always been pointed out that these isolated areas represent the consequences of all the defects identified. Similarly, they are not totally eliminated from the module, even if their participation is weak they slightly participate in the functioning of the module (Fig. 15).

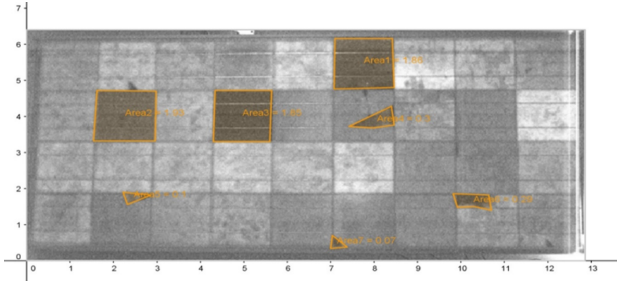


Fig. 15. Module B EL image with $I = 10\% I_{SC}$.

5.2.2 Analysis by Infrared Imaging (IR)

The IR image of the polycrystalline module shows a more homogeneous thermal distribution compared to the mono. Temperatures range from 29 to about 39 °C. The maximum temperature is not then so critical because it is less than the NOCT of 45 °C. the biggest part of the surface has a temperature of about 30 °C. Apart from 3 hot spots identified and some cold areas, the thermal disparity is not significant enough. Thus, the passage of the electric current is not too disturbed hence the low degradation of the current compared to the initial characteristics (Fig. 16).

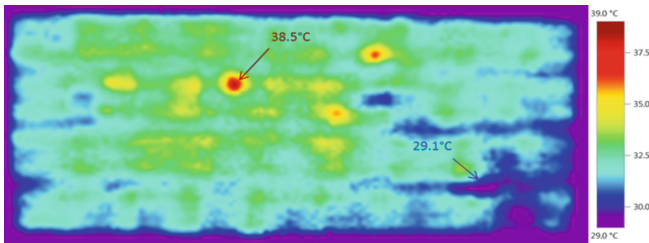


Fig. 16. Module B IR Image

6 Conclusion

The evaluation of the degradation already proves that the modules were affected after the years of exposure. The P_{MPP} degradation is equal to -5.35% for module A, and just -2.92% for module B. But it should be noted that the modules did not have the same degradation behavior. Module A experienced a significant drop in its short-circuit current (I_{SC}), while module B had a slight decrease in its nominal voltage U_{MPP} .

In this way, the visual inspection, EL and IR imaging allow to identify the different defects leading to such kind of degradation for the two modules. Even if there is not enough information on the degradation effects of each type of climate on the different crystalline technologies; defects due to the environmental conditions found in this study are much more present in the monocrystalline module than on the polycrystalline. The mechanical defects on the polycrystalline module are too advanced and will obviously be at the origin of most of the consequences on performance characteristics. The surface monocrystalline module isolated surface by the identified defects is much more important (27.96% of total area), compared to the poly (only 8.5% total area). Finally, EL images of the two PV modules in this study show that with the technology used, the environmental effects have been much more felt on the single-crystalline one. However, the affected surface analysis shows that the effect of percentage of affected surface on the P_{MPP} are more important for the polycrystalline than the mono.

The IR image for the monocrystalline module well justifies extensive degradation of the short circuit current and maximum power. The Module I_{SC} mismatch clear shown in the thermography with many hot spots means that those parts can no more produce the short circuit current. So, the damage previously identified as corrosion and discoloration decreased the electrical output of several cells. Indeed, in the IR image analysis with a current equal to I_{SC} shows a large rise in temperature which reached $60\text{ }^{\circ}\text{C}$. This means that a big part of the module no longer produces enough power and transforms the surplus current coming from the healthy cells into heat. For the poly, the IR image shows that productivity is not too affected. The temperature distribution is almost homogeneous. We can then say that; the important mechanicals defects identified with the EL image do not affect too much the cells' production. Thus, by comparing the analyzes of the IR and EL images of the two modules it can be concluded that the defects caused by the climate on the mono module affect much more the production than the numerous mechanical defects on the poly.

References

1. AIE: African energy outlook: focus on Sub-saharan Africa countries. International Energy Agency, Paris (2014)
2. The World Bank and Energy in Africa, World Bank. <http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/AFRICAEXT/0,,contentMDK:21935594~pagePK:146736~piPK:146830~theSitePK:258644,00.html>
3. Ndiaye, A., Kebe, C.M.F., Charki, A., Ndiaye, P.A., Sambou, V., Kobi, A.: Degradation evaluation of crystalline-silicon photovoltaic modules after a few operation years in a tropical environment. *Sol. Energy* **103**, 70–77 (2014)

4. ANAMS: National Agency for Meteorology of Senegal (2012). <http://www.meteo-senegal.net>
5. Beate, R., Jan, S., Michael, K.: Fluorescence imaging for analysis of the degradation of PV modules. *IEEE-J. Photovolt.* **3** (2013)
6. Freire, F.D.S.: Performance and Degradation Analysis of Operating PV Systems. Rochester Institute of Technology (2016)
7. Virtuani, A., Müllejans, H., Dunlop, E.D.: Comparison of indoor and outdoor performance measurements of recent commercially available solar modules. *Prog. Photovolt.: Res. Appl.* **19**, 11–20. <https://doi.org/10.1002/pip.977>
8. Kaplani, E.: Degradation in field-aged crystalline silicon photovoltaic modules and diagnosis using electroluminescence imaging. In: 8th International Workshop on Teaching in Photovoltaics, Prague (2016)



Building Energy Audit in Nigeria: Some Guides for Energy Efficiency Building Regulations

Abdulhameed Danjuma Mambo¹(✉)
and Cheikh Mouhamed Fadel Kebe²

¹ Department of Civil Engineering, Nile University of Nigeria, Plot 681,
Cadastral Zone C-OO, Research & Institution Area, Jabi Airport Bypass,
Abuja 900001, Nigeria

abdulhameed.mambo@nileuniversity.edu.ng

² Centre International de Formation et de Recherche en Energie Solaire
(CIFRES/ESP) Ecole Supérieure Polytechnique, BP: 5085, Dakar-Fann, Senegal

Abstract. Buildings account for a large proportion of total energy use in the world and that has surpassed all the other sector of the economy of most countries. There is a great need to gather better data on building energy performance and to be able to compare performance across locations and building typologies so that best practices can be more rapidly evolved. The question of how much energy the average Nigerian building consume is still an open ended one. This paper presents a synthesis of energy supplied and utilized in 105 buildings based on building energy audit, physical examination, onsite measurements and discussions with relevant stakeholders, conducted in three cities from different climatic belts of Nigeria. The aim was to draw a basic energy consumption profile of residential, commercial and institutional buildings. This survey found the preponderance of the use of energy inefficient products, inadequate utilization of daylighting, complete absence of building energy management systems, extremely low adoption of renewable energy systems and orienting building without due consideration to bioclimatic factors. This paper recommends policy changes that can be used to realise a large and feasible energy saving in new as well as existing buildings in Nigeria.

Keywords: Building energy audit · Nigeria building code
Building energy consumption

1 Introduction

The precarious energy situation in Nigeria is not just that of shortage of power but also that of gross end-use inefficiency and wastage. It was estimated by International Energy Agency that this wastage is as high as 40% of the power generated [1].

The main concern here is that, the current Nigeria standards [2] for new conventional construction result in buildings that are not energy efficient [3] and therefore use significant amounts of energy to cool the buildings, which results in increased greenhouse gas emissions nationwide. There is the need therefore, to evolve a new

standard that set the minimum level that will take into account energy influential factors in the life cycle of buildings [4].

One of the benefit of energy efficiency standard for buildings is that it will help to stretch the availability of non-renewable energy resources to meet current demands while also providing the time to develop renewable energy sources. Other benefit is that it will reduce water and air pollution thereby improving health conditions of the people.

The objective of this paper is to provide information about energy consuming factors in the selected buildings, develop an energy profile of various building types in some selected towns and to make recommendation on the inputs for energy efficiency in the propose amendment of the National Building Code to incorporate energy efficiency.

2 Energy Audit of Buildings in Nigeria

Energy audit is the first step in increasing the energy efficiency of buildings [5]. It is often said; if you cannot measure it, you cannot manage/control it. There are very few energy efficiency studies in Nigeria. Community Research and Development Centre conducted an energy efficiency survey to elicit information that will guide the development of energy efficiency policy in Nigeria [6]. This study conducted in 3 cities of Benin, Abuja and Lagos is an opinion survey rather than energy audit. Olayinka and Oladele [7] conducted energy audit of food processing industry and Distillation and Bottling Company in Ota, Nigeria. The study observed that electric motors were the major consumer of electrical energy, accounting for 40–47% of total electric energy. In the study by Oyedepo *et al.* [8], energy audit was conducted at Covenant University, Ota to assess the pattern of electricity consumption in order to improve energy consumption efficiency in the University. The paper by Babangida *et al.* [9] presents a complete energy auditing for a student hostel in Ahmadu Bello University, Zaria to ascertain the effectiveness of energy management in the building. Other energy audit of educational buildings can be found in [10, 11]. Previous researches are made mostly on educational buildings. The aim has always been to develop energy efficiency strategy for a particular building. This energy audit involves visiting building facilities for a walkthrough survey, undertaking measurement of the orientation and building extents, material, size and orientation of openings, collection of electricity bills and interacting with building owners to know the duration of use of appliances, to determine the fuel type and the pattern of fuel use.

3 Materials and Methods

This work used the ASHRAE Level 1 and 2 energy audit method [12]. This entails brief selection interviews with building operating staff or occupants to have an overview of the facility's utility bills and additional data, and an abbreviated walk-through of the building. The audit was carried out in the March and April of 2016. These months are the hottest and driest months of the year with highest cooling demand as shown in the temperature and CDD chart of Fig. 2. The audit focused on the

identification of the potential for energy efficiency improvements, understanding the overall building configuration, and defining the type and nature of energy systems. Specifically, during the course of this work, data were collected from two major sources which included; the use of Energy audit forms and interviews. The forms consist of tables with the variables of interest alongside spaces for respondent's answers as well as enumerators' measurements and observations. The enumerators carried out measurements to determine orientation of the buildings, size of façade and other building element of interest. The audit forms were subjected to thorough review and testing before being administered to willing respondents. 105 pieces of detailed audit forms were distributed to three towns in the Northwest representing the Semi-desert climate, Southwest rain forest region and Northcentral savanna climatic zones of Nigeria with slightly differing climate characteristics (see Figs. 1 and 2). Delineation of Nigeria's climate for architectural design purpose is still a work-in-progress.

The buildings selected were in the same area of each city so as to eliminate the factor of climate. The buildings were either of residential, commercial and institutional functions.

The enumerators ask questions from owners and/or occupants. The respondents were asked these questions to extract information on energy use in their respective buildings. The Interviews took care of questions that cannot be obtained from physical measurement and observation. For example, asking about for the actual duration of use of each appliance in the building. Other information such as year and nature of refurbishment or conversion were also obtained.

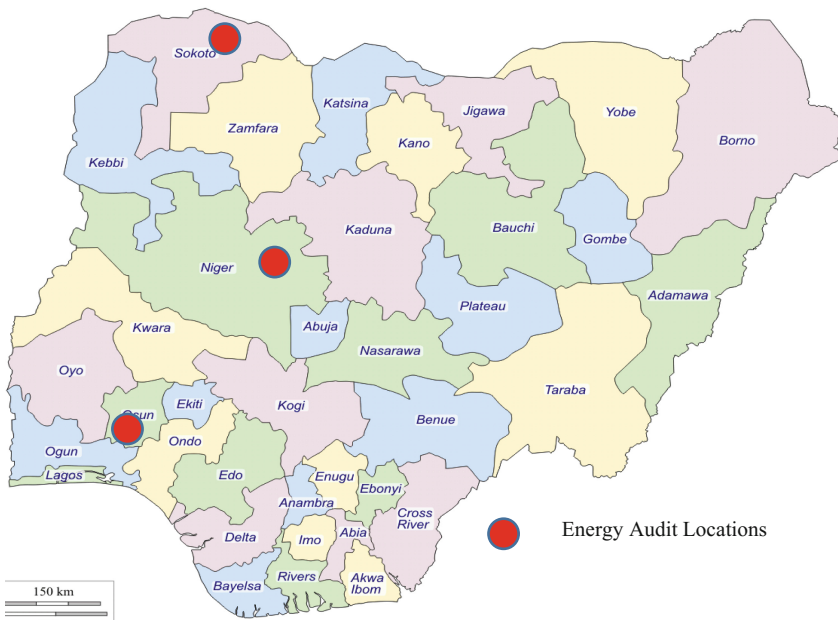


Fig. 1. Map of Nigeria showing building of audit location. Source: (<http://www.d-maps.com>)

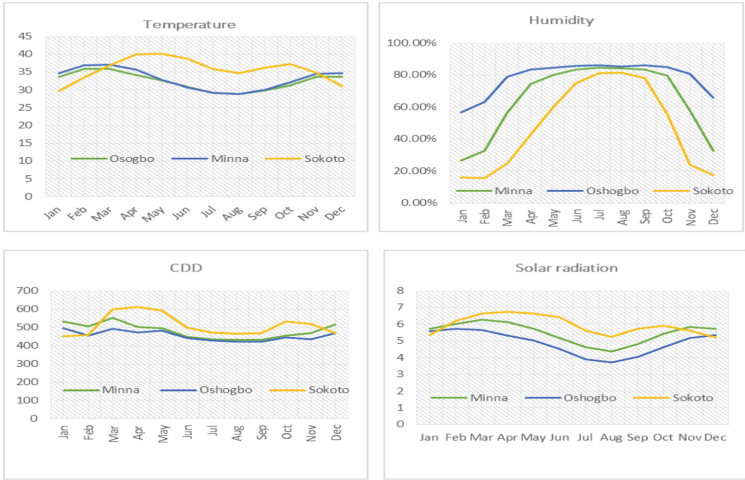


Fig. 2. Comparing the climate variables of the audited cities

Spreadsheets were developed and used to calculate and analyse the data gathered from the energy audit. The spreadsheet results of the survey were then used to develop the energy consumption pattern. The profile shows variations across building types as well as across the region. Charts and graphs were used where appropriate to provide pictorial information.

4 Results

4.1 Sources of Electricity

In all the buildings audited, sources of electricity were mainly from the national grid which is inadequate and epileptic in all the cities considered. The supply of electricity to households, commercial and institutional buildings averaged about 8 h per day across the zones. The supply from the grid was often augmented by private generators. The generators used petrol or diesel to produce power. Very few buildings have storage batteries and inverters used to store power whenever is available from the grid for use during the periods of outages. None of the building surveyed uses any form of renewable energy although about 30% of Nigeria’s electricity comes from hydropower stations.

4.2 Energy Supply Profile

About 45 residential buildings were audited in the three ecological zones of the South-West, North-Central and the North-West. This number is a mixture of different types of residential accommodation such as low income, medium income and high income. The profile of energy consumption obtained from electricity bill and estimated fuel consumption from generators in Kwh/m²/year is shown in Fig. 3 below.

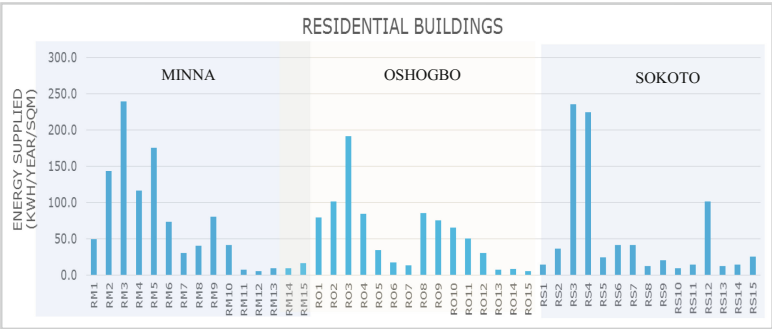


Fig. 3. Energy consumption profile of residential buildings

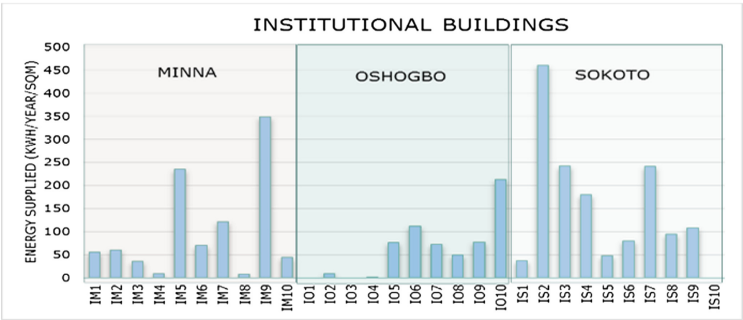


Fig. 4. Energy consumption profile of institutional buildings

A total of 30 commercial buildings were considered, 10 from each zone. The consumption pattern was similar to that of residential and institutional buildings.

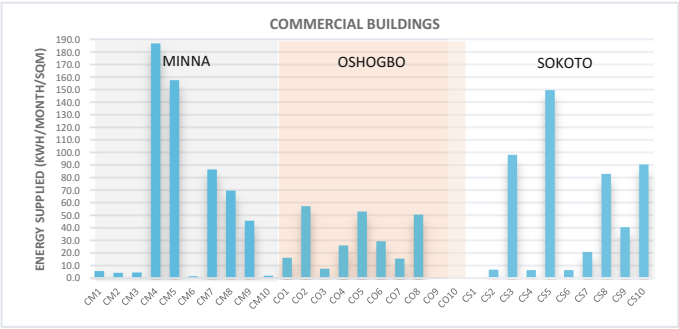


Fig. 5. Energy consumption profile of commercial buildings

4.3 Energy Use by Sector

In order to analyse the energy use pattern for the buildings, energy consuming equipment were categorized based on their functions as in Table 1. The number of such equipment, their power ratings and hour of use per day was used to compute energy use in kWh/m²/a.

Table 1. Equipment categories

Category	Appliances
Ventilation & cooling	Air conditioners Fans
Hot water & cooking	Water heater Boiling rings/kettles Oven/microwave Cooker
Appliances	Fridge/freezer Washing machine Dish washer Television/computer VCR/DVD players Iron/vacuum cleaner Photocopiers Printer/scanner Water dispenser Stabilizers decoder etc.
Lighting	Energy bulbs Conventional bulbs

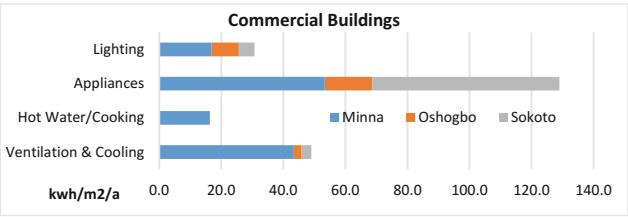


Fig. 6. Energy use pattern of commercial buildings

From Fig. 6, energy consumption of appliances dominates. It also appears that lighting contribution increases as one progresses southward of the country. Sokoto (North West) 10% is least followed by Minna (North Central) and the highest contribution from lighting (33%) comes from Oshogbo (South-West). Cooling energy

consumption per floor area of Minna is higher (33%) because the type of commercial buildings surveyed comprise mainly of those with prestige air conditioning such as banks, telecom companies and big restaurants (explain the reason for 13% for hot water and cooking in Minna). In Sokoto and Oshogbo, the buildings comprise mainly of small scale retail shops where cooling is mainly achieved by natural ventilation and the use of fans.

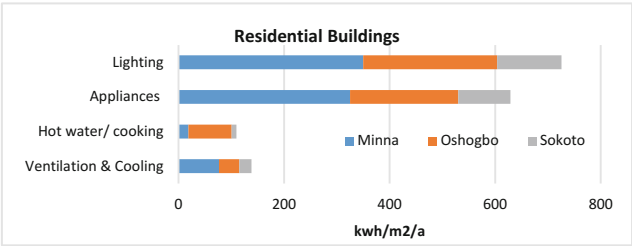


Fig. 7. Energy use pattern of residential buildings

From Fig. 7 lighting dominates the consumptions in residential buildings because of the duration of use. In commercial and institutional buildings, lighting is smaller because activities in those buildings occur mostly in the day time while occupancy for residential buildings occur throughout the day.

From Fig. 8 below, appliances energy consumptions dominate in Institutional buildings followed by ventilation and cooling. Lighting was smaller because period of use of most institutional buildings was during the day when little or no daylight will be required. Ventilation and cooling energy consumption was significant because appreciable number of these buildings air-conditioning and all the buildings uses fans. Expectedly hot water/cooking was the least energy consumer for most of the buildings in this category.

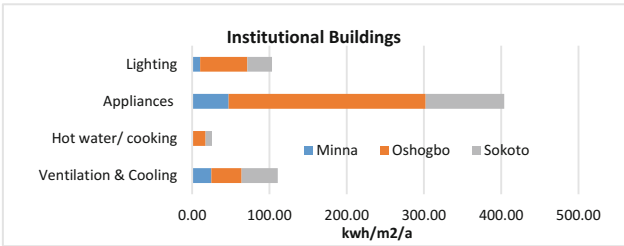


Fig. 8. Energy use pattern of Institutional buildings

4.4 Orientation of Building Major Façade

The orientation of the building's major façade was considered. Selecting the most optimal building orientation is one of the critical energy efficient design decisions that could have impact on building envelope energy performance, as it can be used to minimize the direct sun radiation into the buildings through windows, building openings as well as external opaque walls. It will be most affected for full glazed building. Buildings that are long should be oriented in the north to south direction. This allows the avoidance of direct solar gain in to a greater path of the building. It also helps to reduce areas subject to frequent energy fluctuations due to the rising and setting of the sun. Solar heat gains from the east may be a least nuisance since it often occurs in the morning after the cooler night compare to the gains from the west which usually occurs after an already warm day (Fig. 9).

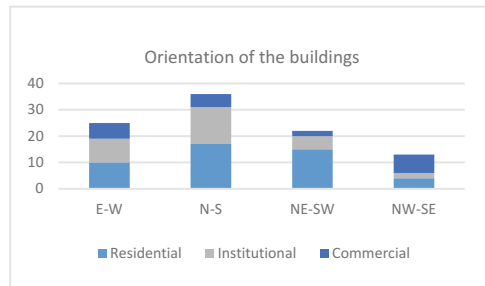


Fig. 9. Orientation of Institutional buildings

Nine (9) out of the 30 institutional buildings surveyed have the orientation of their long axis facing East-West, and 14 out of the 30 buildings have their orientations in the North -South direction. The remaining buildings surveyed have their orientations in the North East-South West or North West-South West direction.

Similarly, except for 5 building with their long axis in the North -South direction, the orientation of the remaining 25 buildings ranges from East-West, Northeast-Southwest and Northwest-Southeast. Overall, most of the buildings long axis were not oriented to offer protection to the façade against solar radiation.

4.5 Building Fabrics: Window, Roofs and Floor

The wall fabrics of all the building surveyed were made up of (from external to internal) paint, plaster, hollow concrete block (6 or 9 inches thick), plaster and paint resulting in very similar u-values, varies only according to the expertise of the builder. The roofs materials range from galvanized roofing sheets, alloyed zinc roof sheets or aluminum sheets and concrete slabs.

Window sizes are very similar; however, openings may differ from using sliding windows (allowing 50% opening) to louvres (allowing over 80% opening). This is significant because appropriate selection for windows orientation, optimal size of the

glass and applying natural ventilation system, can reduce the negative effect of solar radiation in increasing indoor air temperature.

It was observed in the survey that majority of the rooms in the building types surveyed had very poor ventilation, most of them being single sided ventilation. Most buildings adopt the sliding window types which offers only 50% of the window area for natural ventilation. Also, the glazing of most of the windows, even of those in the major building façade were of single glazing and they had poor sealing quality. This could increase energy consumptions of air-conditioning system.

The lack of appropriate natural ventilation and presence of fairly good sealing quality of windows, could increase the use of energy in these commercial buildings.

5 Discussion of Results

Consumption level differs remarkably from house to house. From the chart in figure, it may be concluded that Nigerian residential buildings are generally energy efficient since more than half of the building falls below the 50 kWh/m²/a and none of the buildings exceeded 750 kWh/m²/a. This assumption can be erroneous because the case with Nigeria is that of suppress demand. At the time of this survey, maximum number of hours of electricity supplied from the grid to consumers ranges from 8 to 12 h per day. The consumption for those that augment grid supply with private generation was higher than those that rely exclusively on the national grid.

The same can be said for commercial and institutional buildings as shown in Figs. 4 and 5. Institutional buildings comprise school, hospitals, libraries and government offices. 30 buildings were analysed. The results show that consumption ranges from zero to 450 kWh/m²/a. The zero consumption comes from some public schools that do have electricity at all and the highest consumers augment their electricity supply from the grid with private generators.

6 Conclusion and Recommendations

This report is a basic building energy audit of some 105 buildings in three cities of Nigeria from three different climatic zones earlier identified in the report. The report studied energy supply and use pattern, orientation of major building façade, types of wall, roof and window fabrics and their influence on energy consumption.

From the observations, physical measurements, and interview with occupants and owners of buildings surveyed, the following conclusions were reached: It was obvious that buildings are being constructed without deliberate recourse to energy efficiency issues. Orientation of majority of the building does not seek to harness bioclimatic factors of the building site.

Nigeria's electricity is still predominantly sourced from gas thermal stations and from petrol or diesel generators. Renewable energy uptake has remained remarkably low in spite of the huge potentials. A policy encouraging the increase use of renewable energy to power buildings will improve the environment, stretch the existing energy capacity, introduce stability in supply of power and provide gainful employment.

Although the buildings vary in size and outlook, they are about the same in terms of the fabric layers. Across the region, there was no clearly discernible difference in the features of the buildings to suggest that the differences in climate of the three cities has any influence on the design of the building fabric, window, floor or roofing. A policy encouraging climate specific design and construction could be helpful.

Appliances energy consumption in almost all the building types audited across the three zones was high. A policy encouraging the use of energy efficient appliances, switch offs of appliance when not in use and discouraging the use of second hand electronics will have tremendous impact.

Although energy consumption of cooling and ventilation appears comparatively low, it is a component with perhaps greater future implication. There is likely to be an upsurge in the use of air-conditioning and refrigeration equipment as is the pattern in other tropical countries. A policy laying emphasis on passive cooling, use of solar air-conditioning, energy efficiency labelling and reduction in the importation of old fridges and air-conditioners will be in the right direction. Discouraging the use of sliding windows that leaves only 50% of the widow size for natural ventilation should be implemented. Households should be encouraged to plant trees on the eastern and western sides of their building to provide solar shading against morning and evening sun heat as the sun rises and fall in the horizon respectively.

There are significant opportunities for energy savings in the area of energy management systems and renewable energy as none of the buildings surveyed used EMS or has usable renewable energy system install.

Federal government should review the existing building code to incorporate enforceable energy and resource efficiency provision. This is necessary to ensure that energy efficiency is introduced right from the design stage and also in the event of any refurbishment work.

Dissemination of best practice in energy and resource efficiency should be supported by constructing demonstration buildings such low or zero energy buildings and passive house so that people can see that building energy efficiency is a verifiable reality and not just some impractical concepts.

Energy and resource efficiency should be targeted in all major refurbishment across all building typologies. Government should lead the way in ensuring that public buildings during construction, refurbishment and operation are energy efficiency compliant.

Energy efficiency should be made attractive and visible in the market place such that building owners would see it as a viable alternative to conventional built forms. These could be achieved through labelling, certification of building and appliances as well as subsidies and tax rebate.

The huge advantage in traditional architecture in thermal mass, use of local material, courtyard and shaded veranda system could be harnessed and improved upon to evolve a new built form that is durable, functional and beautiful. The potentials that lie in mud bricks, stone and straw bale may also need to be combined with modern technology so as to produce a new generation of homes that require no mechanical air-conditioning during hot days and nights.

References

1. U.S. Energy Information Administration: International Energy Statistic: Total Electricity Net Consumption (2012)
2. Ministry of Environment, Government of Finland: National Building Code. NBD 2013 (2013)
3. Nduka, D.O.: Stakeholders Perception of Factors Determining the Adoptability of Green Building Practices in Construction Projects in Nigeria, vol. 5, no. 2, pp. 188–197 (2015)
4. Dahiru, D., Abdulazeez, A.D., Abubakar, M.: An Evaluation of the Adequacy of the National Building Code for Achieving a Sustainable Built Environment in Nigeria, vol. 4, no. March 2010, pp. 857–865 (2012)
5. Etiosa, U.: The concept of energy efficiency. In: The Report of a National Dialogue to Promote Renewable Energy and Energy Efficiency in Nigeria, p. 16 (2008)
6. Etiosa, U., Agho, M., Edevbaro, A., Godfrey, O.O., Uyigue, O.P., Okungbowa, O.G.: Energy Efficiency Survey in Nigeria: A Guide for Developing Policy and Legislation. Benin, Nigeria (2009)
7. Olayinka, S., Oladele, A.T.: Energy Audit of Manufacturing and Processing Industries in Nigeria : A Case Study of Food Processing Industry and Distillation & Bottling Company, vol. 1, no. 3, pp. 36–44 (2013)
8. Oyedepo, S.O., et al.: A study on energy demand and consumption in. In: International Conference on African Development Issues (CU-ICADI) 2015 (2015)
9. Babangida, M., Abubakar, M., Abba, L.B., Abubakar, K.A.: Energy auditing and management. *J. Multidiscip. Eng. Sci. Technol.* **2**(7), 1807–1813 (2015)
10. Itabor, N.A., Damisah, L.E., Olatunji, J.O., Eyakwanor, T.O., Akingba, O.O., Oerome, R. O.: Energy efficiency in buildings: case of post graduate hostels at the University of Benin, Nigeria. *Phys. Sci. Res. Int.* **3**(November), 51–58 (2015)
11. Maina, M.B., Shuwa, M.: Electricity use characteristics of tertiary institutions in Nigeria: a case study of Ramat Polytechnic Maiduguri, Borno State. *Int. J. Eng. Res. Technol.* **4**(12), 596–601 (2015)
12. Deru, M.P., Jim, K.P.E., ASHRAE: Procedures for Commercial Building Energy Audits. American Society of Heating, Refrigerating, and Air-Conditioning Engineers (2011)



ICT Performance Indicators in Formal Education at the Secondary School Level in Rwanda

Florien Gumyusenge, Robert Mugisha^(✉), Aminata A. Garba,
and Martin Saint

Department of Electrical and Computer Engineering,
Carnegie Mellon University, Kigali, Rwanda
{fgumyuse, rmugisha, aminata, msaint}@andrew.cmu.edu

Abstract. The objective of this study is to evaluate the performance of ICT indicators used to support education in formal secondary schools in Rwanda. By examining related research papers, Rwandan ministerial reports, and relevant journals, we were able to obtain secondary data for analysis. Additionally, primary data were collected from 46 secondary schools through a questionnaire. Overall, Rwanda has demonstrated significant improvement over the course of just one year. The ratio of students to computers increased from 27:1 in 2016 to 7:1 in 2017, and access to electricity increased from 25% in 2016 to 95.7% in 2017. Additionally, the number of ICT-trained teachers increased from 2.5% in 2016 to 43.7% in 2017. However, there are still improvements to be made; only 43.5% of schools have an Internet connection, and, of those, only 37.1% of students actually use it.

Keywords: Information and communication technology · ICT Policy · Indicators · Education · Rwanda

1 Introduction

The purpose of this project is to examine the procedures used in Rwandan schools to measure and evaluate information and communications technology (ICT) indicators in education. According to the Rwanda Ministry of Youth and ICT (MYICT), the Rwandan government is following an ambitious twenty-year plan to enhance its ICT services. “The revolution started when ICT was regarded as a necessity, not a choice,” said the Minister of MYICT, Jean Philbert Nsenigimana, when Rwanda started structuring the 2020 National Information Communication Infrastructure (NICI) plan. The plan aims to strengthen education that eradicates the current ICT gap, among other goals.

Given the growing role of ICT in education, there is an increasing need for monitoring the development of this integration and analyzing the reliability and validity of the indicators used to determine this change. This will allow policymakers to monitor the progress of the country relative to the national

targets and other adequate benchmarks. This national focus generated a need for analyzing the performance of ICT indicators in education in Rwanda, as it is the main pillar for the development of the country. Thus, the objectives of this analysis are as follows:

- Measure the accuracy of ICT indicators for the use of ICT in education in Rwanda.
- Evaluate International Telecommunication Union (ITU) performance indicators used in Rwanda to support the use of ICT in education.
- Identify and analyze any trends in the results and make recommendations.

1.1 Variables

The scope of this project is limited to educational institutions and their students [1]. Both public and private educational institutions at International Standard Classification of Education (ISCED) [2] Levels 2 and 3 are included, and the students are limited to those entitled to use school computers. This study focuses on five of the nine ICT indicators in education based upon data available for those five [3]:

- ED4: Learners-to-computer ratio in schools with computer-assisted instruction.
- ED5: Proportion of schools with Internet access, by type of access.
- ED6: Proportion of learners who have access to the Internet at school.
- ED8: Proportion of ICT-qualified teachers in schools.
- EDR1: Reference indicator: Proportion of schools with electricity.

1.2 Measurement of Performance Indicators by ITU and UNESCO

According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), there are five main categories to focus on in performance measurement and analysis of ICT indicators in education.

- Policy
- Technology Infrastructure and Access
- ICT Curriculum
- Teaching and Teaching Support Staff
- Learning Process and Outcomes

These indicators can be used at the international, regional, national, sub-national, and individual school levels for monitoring and supporting policy formulation, monitoring and evaluation, and decision making. The analysis also assists in comparing ICT infrastructure, access, usage, and outcomes. Most of the quantitative data required are collected from the educational institutions during school censuses or surveys, and analysis is completed by calculation. The following section outlines five of the indicators and how they are analyzed mathematically [3].

ED4: *Learner-to-computer ratio in schools with computer-assisted instruction.*

$$\frac{\sum_{h=2}^3 LC_h^t}{\sum_{h=2}^3 CP_h^t} \quad (1)$$

Where LC_h^t is the number of learners entitled to use computer laboratories at school as pedagogical aid at level of education h in school year t . CP_h^t is the number of computers available for pedagogical use at level of education h in school year t .

In the absence of nationally defined norms by pedagogues, a ratio of one learner to one computer indicates a perfect adequacy in the provision of a computer to all learners officially entitled to benefit from it in schools that offer computer-assisted instruction (CAI) [4].

ED5: *Proportion of schools with Internet access by type of access.*

This indicator is utilized to compare the number of schools with Internet access with the total number of schools in a given academic year [4]. The following expression is used to calculate the proportion of schools with Internet access:

$$\frac{\sum_{h=2}^3 EII_{h,s}^t}{\sum_{h=2}^3 EI_h^t} * 100 \quad (2)$$

Where $EII_{h,s}^t$ is the number of educational institutions with access to the Internet at level of education h by type of Internet access in school year t and EI_h^t is the number of educational institutions at level of education h in school year t . The indicator is expressed in percentages and the high expectation is 100% [4].

ED6: *Proportion of learners who have access to the Internet at school.*

The purpose of evaluating this indicator is to measure Internet accessibility among learners for educational purposes. The expression used is:

$$\frac{\sum_{h=2}^3 LI_{h,s}^t}{\sum_{h=2}^3 L_h^t} * 100 \quad (3)$$

Where $LI_{h,s}^t$ is the number of learners entitled to use Internet at school as a pedagogical aid at education level h in school year t and L_h^t is the number of learners enrolled at education level h in school year t . The indicator is expressed in percentages and the high expectation is 100%.

ED8: *Proportion of ICT-qualified teachers in schools.*

This indicator aims to measure the extent to which secondary school teachers have the required ICT training to teach basic computer skills or computing classes [4]. The expression used is:

$$\frac{\sum_{h=2}^3 TTB_{h,s}^t}{\sum_{h=2}^3 T_h^t} * 100 \quad (4)$$

Where $TTB_{h,s}^t$ is the number of teachers trained to teach basic computer skills or computing at level of education h in school year t and T_h^t is the number of teachers at level of education h in school-year t . The indicator is expressed in percentages and the high expectation is 100% [4].

EDR1: *Reference indicator: Proportion of schools with electricity.*

This indicator measures the availability of electricity as a minimum, necessary condition for ICT to be introduced to schools, although not all ICT tools used in education require a permanent and sustainable supply of electricity. The indicator is measured as the percentage of schools with electricity compared to the total number of schools, with a high expectation of 100% [4].

$$\frac{\sum_{h=2}^3 EIE_{h,s}^t}{\sum_{h=2}^3 EI_h^t} * 100 \quad (5)$$

Where $EIE_{h,s}^t$ is the number of educational institutions with electricity at level of education h in school-year t and EI_h^t is the number of educational institutions at level of education h in school-year t .

2 Design and Methodology

The following steps were taken to provide appropriate context for the work and collect and analyze data.

1. Review literature: Relevant research papers, governments, ITU and UNESCO reports, newspapers, and journals.
2. Identify indicators: Collected from annual, country-level school censuses.
3. Collect secondary data: Collected from websites of the Ministry of Education (MINEDUC) and Ministry of Youth and ICT (MYICT), newspapers, and journals. MINEDUC provided important information about policies focused on using ICT in education and updated reports and statistics, while MYICT provided crucial information about ICT integration in education.
4. Collect primary data: Of the 2108 schools who received the questionnaire, only 31 secondary school administrators participated. Administrators, specifically, were targeted as they have a better understanding of ICT tools than most students. The survey was based on one used by ITU and comprises of both open and closed questions. The survey was also anonymous to encourage honesty and depth of answers.

3 Results

The following section details the results from the secondary data concerning the ICT performance indicators collected.

3.1 The Role of ICT in Education

The Rwandan government outlined the specific aims of ICT use in the education sector as follows [1]:

- Improve administrative efficiency.
- Disseminate teaching and learning materials to teachers and students.
- Improve the ICT skills of teachers and students.
- Allow teachers and students access to sources of information from around the world.
- Allow educators to share ideas on education and learning.
- Encourage teachers to collaborate on joint projects and conduct lessons from a remote location.

3.2 Access to Electricity, Telephone, and Internet

Selected ICT indicators in education in Rwanda between 2006 and 2010 are represented in Fig. 1. As shown, between 2006 and 2010 Rwanda had accessibility to electricity, computers, and the Internet and use of websites in secondary schools. These elements are used as indicators of ICT integration in education, and the status and change of each is analyzed.

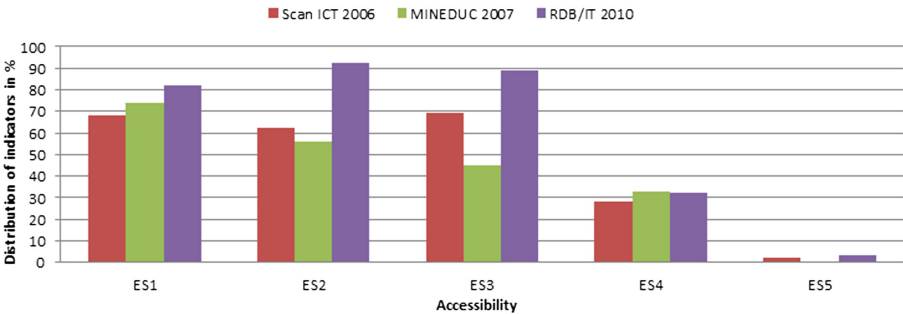


Fig. 1. Accessibility to electricity, telephone and Internet by secondary schools in Rwanda.

Figure 1 compares the percentage of schools with access to electricity (ES1), the percentage of schools with telephone access (ES2), the percentage of schools using computers (ES3), the percentage of schools with access to the Internet (ES4), and the percentage of schools with websites in 2006, 2007, and 2010 (ES5). There is a significant increase in electricity penetration in the country between 2006 and 2010, and there was adequate usage of telephones and computers. However, this does not clearly highlight if computers were accessed by school administrators or students. More sensitization on the use of websites is needed in the country where less than 5% of schools had websites in 2010.

3.3 Relevant Policies

One Laptop per Child. The One Laptop per Child program and other ministry initiatives have registered significant successes to date, with more than 200,000 laptops distributed in over 400 schools in 2015. However, with over 3 million students in Rwanda's education system, only a small proportion of students received the benefits of the use of technology in education. As a solution, the government decided to partner with POSITIVO-BGH to assemble electronic devices in Rwanda, including desktops, laptops, and tablets. This partnership strives to make such devices readily available on the market at low prices to serve both students and teachers.

One Laptop per Teacher. After the introduction of One Laptop per Child, it became clear that most teachers who are expected to instruct the students on ICT did not have computers themselves, or the capacity to use them. As a solution, the government devised a plan to train all teachers in ICT and allocate one laptop per teacher.

GiraICT. In addition to laptops, the GiraICT organization also gives teachers Internet access for six months to enhance their laptop usage and skills. The program is not only focused on supplying laptops, but it also ensures that the beneficiaries use them to improve the quality of teaching.

Investment Implication. The implementation of ICT integration in education is planned to be achieved through four phases from 2015 through 2019, with specific investment levels for each phase as shown in Fig. 2.

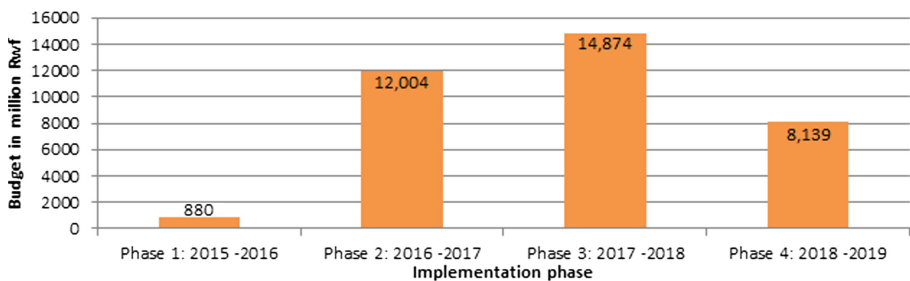


Fig. 2. Budget for ICT integration in education in Rwanda.

Challenges of ICT Integration in Education

- Poor infrastructure: Only 47% electricity coverage in public schools.
- Only 6% of primary and 18% of secondary schools connected to the Internet.
- Limited ICT resources for education.
- Turnover in expertise at learning institutions.

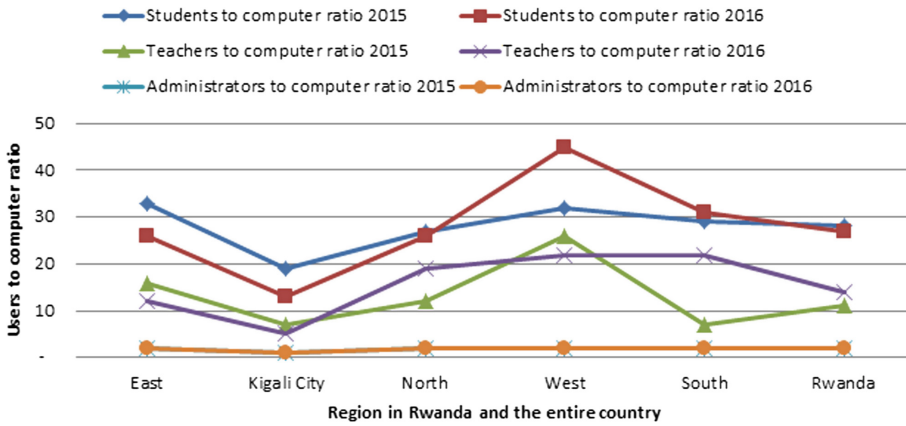


Fig. 3. Secondary school users to computer ratio in 2015 and 2016 in Rwanda.

The ratio of learners to computers and teachers to computers is high compared to the ratio of administrators to computers as shown in Fig. 3.

The ratio of learners to computers increased in the Eastern Province from 35:1 to 25:1. The ratio also increased in Kigali from 18:1 to 15:1. There was no change in the Northern Province, whereas in the Western and Southern Provinces, the ratios decrease. The reports do not comment on the cause of this decrease. There was a small change in computer accessibility in 2016 as compared to the 2015 ratios. The ratio of the total number of teachers to the number of teachers with access to computers in the country changed from 11:1 to 14:1 and the learner to computer ratio changed from 28:1 to 27:1. The change in the ratio of teachers with access to computers might have been caused by the recruitment of new teachers, while the change in learner to computer ratios might have been impacted by the One Laptop per Child program.

According to the Rwanda's national goals, the 2016/2017 target of electrifying 65% of the country has not been achieved. As shown in Fig. 4, only 46% of secondary schools have access to electricity through the national grid [5].

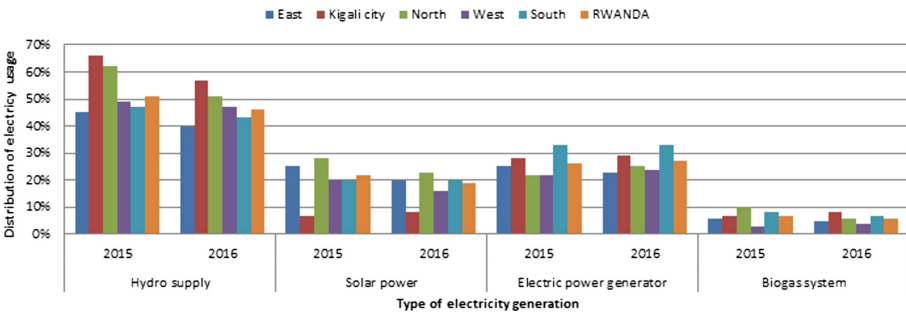


Fig. 4. Secondary schools with electricity in Rwanda in 2016.

Primary Data. The analysis of data in this section is done mathematically and discussion is given per indicator. The data collected are summarized in Table 1.

Table 1. Primary data collected

Total number of learners	22791
Number of computers	5719
Number of computers accessed by students	3730
Number of computers connected to Internet	3428
Number of learners owning computers	1357
Girls enrolled in ICT related fields	811
Learners entitled to use computers at school	22791
Learners entitled to use Internet at school	22791
Number of learners who have access to Internet	8450
Total number of teachers	817
Teachers trained to teach basic computer skills	357
Teachers who use ICT facilities	366
Teachers who teach basic computer skills	124
Number of teachers trained to teach using ICT facilities	464
Total number of schools	46
Schools with Internet	20
Schools with electricity	44

ED4: *Learners-to-computer ratio in schools with computer-assisted instruction.*

$$\frac{\sum_{h=2}^3 LC_h^t}{\sum_{h=2}^3 CP_h^t} = \frac{22791}{3730} = 7 \quad (6)$$

In secondary schools in Rwanda, one computer is shared by seven students. Although there is improvement compared to the year 2016, when 27 students were sharing one computer, the number leaves room for further improvement. The government of Rwanda has a vision of having one laptop per child.

ED5: *Proportion of schools with Internet access by type of access.*

$$\frac{\sum_{h=2}^3 EII_{h,s}^t}{\sum_{h=2}^3 EI_h^t} * 100 = \frac{20}{46} * 100 = 43.5\% \quad (7)$$

ED6: *Proportion of learners who have access to the Internet at school.*

$$\frac{\sum_{h=2}^3 LI_{h,s}^t}{\sum_{h=2}^3 L_h^t} * 100 = \frac{8450}{22791} = 37.1\% \quad (8)$$

Although about 44% of secondary schools have access to the Internet, it is seen that only 37.1% of students use Internet as a pedagogical aid.

ED8: *Proportion of ICT-qualified teachers in schools.*

$$\frac{\sum_{h=2}^3 TTB_{h,s}^t}{\sum_{h=2}^3 T_h^t} * 100 = \frac{357}{817} = 43.7\% \quad (9)$$

According to Fig 5, only 43.7% of all teachers are trained in ICT, including educators who teach computer science. This indicates a large gap in the education potential of ICT facilities.

EDR1: *Reference indicator: Proportion of schools with electricity.*

$$\frac{\sum_{h=2}^3 EIE_{h,s}^t}{\sum_{h=2}^3 EI_h^t} * 100 = \frac{44}{46} * 100 = 95.7\% \quad (10)$$

Of the secondary schools in Rwanda, 95.7% are electrified. This is a small step toward implementing ICT curricula and facilities, since all ICT tools function using electricity. The above performance indicators of 2017 are summarized in Fig. 5.

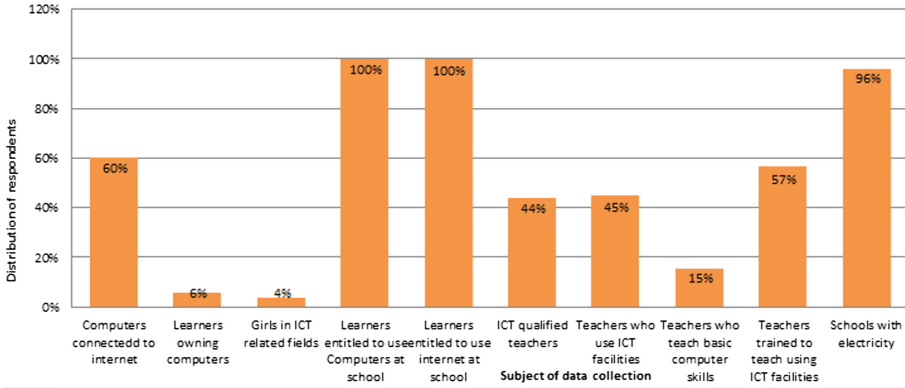


Fig. 5. Summary of all questions represented on a histogram.

4 Recommendations

Based on the successes of Rwanda's advancement of ICT, all East African countries should define and specify ICT integration in education and use ITU performance indicators to evaluate their achievements. The significance of ICT warrants its own sector, budget, and policy. Rwanda should focus on Internet access

and its usage in all schools in rural areas as well as in cities so that all schools can progress at the same pace and provide better quality education. The government of Rwanda should also focus on infrastructure and capacity building as well as developing teacher awareness of the importance of the role of ICT in educational development. In addition to this, the government should assist private schools in obtaining ICT equipment. The government should focus on expanding off-grid solutions to assist schools in rural areas to obtain electricity so that they can make essential technological advancements.

5 Conclusion

By implementing clear ICT policies for integration in the education sector, Rwanda has made significant strides toward technological advancement compared to its neighboring countries, even though it still has a number of areas in which to improve. All indicators measured in this paper show improvement compared to the results of previous years as.

References

1. UNESCO Institute for Statistics: ICTs and education indicators: suggested core indicators based on meta-analysis of selected international school surveys, Report, United Nations Educational, Scientific, and Cultural Organization (UNESCO), Quebec, Canada (2006)
2. UNESCO Institute for Statistics: International standard classification of education ISCED 2011, Report, United Nations Educational, Scientific, and Cultural Organization (UNESCO) Institute for Statistics, Quebec, Canada (2012)
3. UNESCO Institute for Statistics: Guide to measuring information and communication technologies (ICT) in education, Report, United Nations Educational, Scientific, and Cultural Organization (UNESCO), Quebec, Canada (2009)
4. Ngeze, L.V. (ed.): ICT integration in teaching and learning in secondary schools in Tanzania: readiness and way forward, vol. 7, no. 6, Dodoma, Tanzania, June 2017. University of Dodoma, International Journal of Information and Education Technology (IJIET)
5. Rwanda Ministry of Education: 2016 education statistical yearbook, Report, Rwanda Ministry of Education (MINEDUC), Kigali, Rwanda, February 2016



The Impact of over the Top Service Providers in the Rwandan Telecommunications Market: An Analysis of Business Models

Emelyne Kwizera^(✉), Darius Mico, Michael Nayebare, Aminata A. Garba, Martin Saint, and Linda G. Deen

Department of Information and Communications Technology,
Carnegie Mellon University, Kigali, Rwanda
ekwizera@alumni.cmu.edu,
{dmico,mnayebare,aminata,msaint,linda}@andrew.cmu.edu

Abstract. The telecommunication industry is facing constant changes in its operations and business environment. One of the major causes of these changes is the rise of Over The Top (OTT) service providers that have developed and implemented innovative new business models that are disrupting the traditional telecommunication industry. In this paper we investigate the impact of OTT business models on Communication Service Providers (CSPs) in the Rwandan telecommunication market. We seek to understand the business models and strategies CSPs and OTTs are employing to stay relevant in the market, the advantages and disadvantages of each model, and to assess whether they are applicable in the Rwandan telecommunications market. The findings and recommendations highlight some of the major trends and business models in the telecommunication markets, and can be used to provide insight for CSPs and regulators to build strong business strategies and relevant regulations.

Keywords: Telecommunications · Communication service providers
CSP · Business models · Over the top · OTT · Rwanda

1 Introduction

Companies operating in the global telecommunications market face vast challenges due to the pace of innovation. Companies in this sector traditionally provide data, Internet, and other value-added services, often delivered over mobile networks. However, telecommunications business models are changing as consumers of their products change their behavior. Telecommunications providers are also being challenged by the arrival of over the top service providers (OTTs) that provide similar services in a cheaper, easier and more innovative way. In some cases, new service providers threaten the business models of traditional communications service providers.

The telecommunications industry has long been dominated by communication service providers (CSPs) offering services including voice, data and SMS (short message service). Traditionally, CSPs like AT&T and Verizon relied on voice and short message services (SMS) as a primary source of revenue. Mobile phone development and mobile Internet are at the forefront of an innovation drive that is transforming the market. Market transformation is also accelerated by the decline in smartphone prices, growing access to the Internet, and an explosive increase in user-generated content especially from social media platforms, search engines, and other content service providers [1].

Over The Top (OTTs) players are service providers that offer content such as voice, video, data, instant messaging, and other media to consumers. Providers of OTT services include companies like Google, Netflix, and Facebook. They are referred to as OTTs because they do not possess telecommunications infrastructure to deliver their content to the final consumer and they do not need licenses to operate. OTT service providers rely on the CSPs existing infrastructure, including tangible assets like masts, ducts, and transmission components, and intangible assets like spectrum licenses and rights of way [2]. The emergence of OTTs has triggered a wave of transitions, causing many CSPs to adapt their business models and strategies to remain competitive and profitable in the market, mostly because OTTs offer similar products and services that include voice and instant messaging [3]. In addition, there is a growing demand for OTT services such as video and audio because of the high value placed on effective communication, and improvements in service quality and pricing. However, this kind of competition creates significant pressure on CSPs to provide new and improved value-added services.

The objective of the paper is to analyze the existing models of OTTs and CSPs in the market, then propose a business model that CSPs might adopt in order to design strategies that will help them mitigate competition from similar services being offered by OTTs. In addition, we cover emerging trends in the technology industry and strategies that OTTs are developing that could pose a threat to the stability of the telecommunication industry and CSP services like mobile data and SMS. Further, we analyze the business models that are hybrid (i.e. shared by OTTs and CSPs or by CSPs and other stakeholders) in the telecommunication market and their impact on all stakeholders in the market.

2 Background

2.1 The Telecommunications Market in Rwanda

The telecommunication market in Rwanda is occupied by three major CSPs: MTN, Airtel, and Tigo. They offer fixed (landline) telephone, mobile telephone (mainly mobile calling and SMS), and Internet data. The market is regulated by the Rwanda Utilities Regulatory Authority (RURA), which is responsible for oversight countrywide [4], and the East Africa Communications Organization (EACO), which is a regional organization that connects ICT stakeholders in countries of the East African Community (EAC) [5].

In the Rwandan telecommunication market, CSPs face pressure to provide high quality services because of the increase in demand for data and broadband. This rapid increase in demand is due to the growing penetration of mobile devices and increasing demand for multimedia content and applications [3]. According to a report by RURA, mobile telephone subscriptions increased from 7.7 million in the fourth quarter 2014 to about 7.9 million in the first quarter of 2015 [4]. Growth in subscribers and demand is causing major shifts in the industry.

2.2 Over the Top Players (OTTs)

OTT growth is facilitated by advancements in technology, especially smartphones, fast Internet that can reach speeds up to 1 Gbps, new services like WhatsApp, and consumer affinity for services like online gaming and video streaming [6]. Services offered by OTTs are often free, or at least less expensive compared to similar services from CSPs. OTTs have flexible business models that have made them successful players in the industry. For example, OTTs like Facebook, Google and Skype have managed to create successful business models without having to build their own infrastructure; instead, they are able to take advantage of existing CSP infrastructure. By riding on the existing infrastructure, OTTs are decreasing the capital expenditures (CAPEX) necessary to profitably offer new services. For example, in 2014, WhatsApp reached over 600 million global subscribers and generated over \$1,289,000 in revenue. In the same year, Netflix had 50 million subscribers with \$5.5 billion in revenue [5].

2.3 Zero-Rating

There are currently more than one billion people who cannot afford access to the Internet [7]. In many developing countries, CSPs work in partnership with OTTs to provide their subscribers with free access to select websites or applications. Zero-rating occurs when communication service providers waive data charges to these select services [7]. Consumers typically have access to a Free Basics plan that allows them access to websites such as Facebook Zero, Google, and Wikipedia alongside localized content and information ranging from Ebola health advisories to women's rights applications and job postings [7]. Proponents of zero-rating argue that the practice of zero-rating increases innovation and development and gives Internet access to people that would otherwise find it unaffordable. Customers gain access to a number of free services including health, education, communication, sports, travel, jobs and local information [8]. Zero-rating is offered under different plans; for example, under the "free website plan," users can access one website with limited access, e.g. Facebook Zero users only access Facebook, but with no images [9].

In Africa, zero-rating is widespread. For example, MTN South Africa zero-rates Twitter, in Kenya, Airtel zero-rates Facebook Free Basics, Safaricom zero-rates Wikipedia, and MTN Nigeria partnered with Opera Mini and naij.com for "1 million days of free Internet" [10]. In 2015, during the Transform Africa

Summit in Kigali, Facebook launched internet.org, a free basic platform in Kinyarwanda that offers free health care information, educational content, and information about jobs and government [10]. Facebook has also partnered with some cellphone operators like Airtel in Rwanda to offer zero-rated services to areas that are hard to reach or underserved [8].

3 Analysis of Business Models

We adopted a qualitative approach wherein we analyzed secondary data from academic papers, online articles, and journals in order to understand the existing OTT and CSP business models, their importance, and the advantages and disadvantages in order to give an in-depth recommendation concerning adoption in the Rwandan Telecom market.

After reviewing the gathered information, we formed conclusions on the relevance of recommendations in the telecommunications market. Procedurally, we first determined which of the data collected to associate with the project. Second, we interpreted the relevant data to better understand the business models of both OTTs and CSPs and the possible impact on the market. Third, we cross-analyzed the data to evaluate the impact of OTTs on affordability and adoption of telecommunication services, and to determine if the existing model is positive or negative in relation to the competition. Lastly, we provide recommendations and ideas for the evolution of OTT models in the Rwandan market.

OTTs and CSPs have traditionally been linked to different kinds of business models. CSPs in Africa are well known for their “pay as you go” business model, while most OTTs are known for their “freemium” and “advertising” business models. The business models covered in this paper are grouped into two categories: OTT business models, and business models between CSPs, users, and OTTs (hybrid business models).

3.1 Over the Top Service Provider Business Models

These are some of the business models that were traditionally exclusive to over the top players. They include: freemium, cloud services, and advertising business models.

Freemium Business Model. As the usage of smartphones and the Internet continues to increase, OTTs are now focusing on services that are offered on mobile platforms. Many OTTs build on the freemium business model. In the freemium model, companies offer a basic service for free and charge a subscription fee for premium services with advanced features [5]. For OTTs to profit from the freemium business model, they need to have a large customer base where even a small percentage of customers paying for the premium service will generate enough revenue to sustain the OTT. Well known examples of companies who have leveraged this model include Skype and LinkedIn. Skype and LinkedIn have over 600 million and 500 million users respectively, most of whom are using their

services for free. However, the small percentage of users who pay for premium services is enough to generate around a billion dollars in revenue for each of these companies [11]. The cost-free tier attracts many users, which in turn places a huge burden on telecom operator networks. Many users switch from using CSP services like SMS and voice calling to OTT services like WhatsApp and Skype, mainly due to their low cost. CSP data traffic has increased due to the increase in OTT service users. In 2015, Skype had around 250 million users per month [12]. This has a negative impact on CSP voice revenue, which is estimated to continue decreasing in the next few years. Studies show that telecom operators can expect fierce competition in the future from freemium-model services [1].

While CSPs in developing countries face issues of infrastructure upgrades and reduced revenue in voice calls, some CSPs are also benefiting from the freemium business model. According to a 2015 ITU report, four billion people (more than 66% of the world population) from developing countries do not have access to the Internet [7] and only 89 million people in the least developed countries use the Internet, representing a 9.5% Internet penetration rate. These statistics demonstrate a potential market for CSPs to target underserved customers. In Rwanda, for instance, MTN, Tigo and Airtel use the OTTs' freemium business model to reach out and gain new customers. OTTs also enable new capabilities for CSPs. For example, customer comments or complaints are passed on to CSPs cheaply and efficiently through OTT services like Facebook posts. CSPs are then able to get instant feedback from customers through Facebook or Twitter, increasing the efficiency and responsiveness of CSPs. There is also improvement in service due to customer rating and reviews on OTT services. Besides CSPs using OTT services, some have adopted similar business models. In Rwanda, for example, Airtel gave customers who bought a new SIM card 500 MB of data for free, which enabled Airtel to acquire more subscribers. Although this is not a typical freemium business model, it was inspired by the popularity of freemium services. The freemium model has proven to be popular in the Rwanda telecommunication market, and CSPs have built on it to start other business models such as "bundling," explained in Sect. 3.2.

Advertisement Business Model. OTTs like Google and Facebook provide their services for free and then sell advertising space to businesses that want to target the same audience [5]. Since the services are free, these OTTs usually have many users who feed them critical information from which they can profit. A typical example is Google search, where users of the Google search engine type in their key search words and businesses interested in the search terms then place a bid with Google to have their products displayed to customers on the first page. Google then uses an algorithm to determine which advertisement to post to the first page of user search results. Advertisements, therefore, are tailored to the exact interests of the user at the time they are displayed. Facebook's business model is also built around this concept, with over 85% of revenue coming from advertisement.

The rise of services like Facebook, Twitter, and Google has had a significant impact in the Rwandan telecommunications industry. First, CSPs are able to market their products and services and reach more customers through OTT services at a very low cost. In Rwanda, CSPs formerly had to put up billboards from district to district and province to province to advertise their products, but through Facebook, WhatsApp, Twitter, YouTube, and other OTT platforms, CSPs reach customers with ease. Secondly, customer satisfaction has increased because there are more ways for customers to communicate with CSPs. Customer care lines are no longer the sole means to file a complaint or place inquiries, as consumers can also use OTT services. When a customer asks a question on MTNs Facebook page, there is no need to wait for MTN to answer the question; any user on the page who knows the answer can respond.

Free OTT services that use the advertising business model have increased CSP traffic as more people use various forms of OTT services. The increase in traffic can have two major impacts on the CSP network. One, CSPs may profit from the increase in data traffic, seen particularly in Rwanda where there is still low market penetration and CSPs are looking for users. Or, two, the traffic may place a burden on the CSPs infrastructure. As there are no reports of CSPs complaining about the burden of data traffic in Rwanda, this suggests that CSPs actually benefit from their customer's use of OTT services. An increase in data traffic increases CSP revenue under metered or pay-as-you-go plans.

Cloud Services. OTTs like Google and Amazon are providing cloud services to both businesses and individuals all over the world. They provide services like storage, applications like enterprise resource planning (ERP) that include accounting and customer management software, computer networks, and server space. These services are helping businesses reduce expenditures on initial installations, such as paying for technical labor and expensive equipment, thereby saving cost and deployment time. Business employees have the luxury of working from anywhere in the world and can collaborate easily with coworkers. Businesses need not bother with equipment maintenance, as this service is provided by CSPs. The cloud model has helped CSPs increase revenue by providing cloud services to many businesses in Africa. The ability to reduce the cost of initial investments and scale easily to match growth has led to the popularity of cloud services, particularly for small and medium enterprises.

In Rwanda, CSPs like MTN are offering cloud services to small and medium businesses. Cloud solutions include voice and data services. CSPs offering cloud services are able to increase their own revenue by reaching more businesses with extended service offerings. Customers have the opportunity to work with local service providers, and gain benefits including services that are tailored to their needs and the ability to pay in their own country currency to access these services. In Rwanda, the success of the cloud model has been facilitated by development of an optical fiber backbone network that serves around 90% of the country.

3.2 Hybrid Business Models

CSP Business Models After the Rise of OTTs. With the rise of OTTs, some of which offer similar services as CSPs, telecom operators have had to change the way they operate, such as changing the structure of their loyalty programs, and even partnering with OTTs. CSPs partner with OTTs to provide services to consumers, and allow them to win prizes or earn rewards for using a particular service [13]. For example, a content provider offering e-commerce services might award consumers different points for using the service. The loyalty points can then be converted into data credit for the consumer to use, or converted into M-Pesa, a mobile phone based money transfer service that allows users to deposit, withdraw, send, and pay for goods through a mobile device. The intent is that consumers continue to use these services to make online purchases even after the promotion ends [13]. The CSP continues to provide reliable data for the consumer, which creates an engagement, and both the CSP and the OTT benefit from the consumer purchasing more data and buying more products [14].

Bundling: CSP Bundling Before OTTs. Bundling is a marketing strategy whereby entities (CSPs) providing several products/services separately offer users a discount for using the same products grouped as a single, combined package [15]. Bundling can include products offered by a single company or services from different companies put in a single package [16]. There are variations of bundling: pure bundling, mixed bundling, and intermediate bundling. Pure bundling occurs when a user buys the full bundle. Mixed bundling occurs when a consumer purchases separate parts of the product; either one, two, or all of the products included in the bundle. The intermediate bundle allows users to benefit from the second product only if the first product is purchased [15].

Over the years, CSPs have bundled services in the same package in order to satisfy their customers. However, they typically only bundled services that they offered directly like voice and SMS, voice and data, or voice, SMS, and data. With the rise of OTT services like WhatsApp and Skype, which have billions of users all over the world, CSPs began to lose revenue from voice and SMS. According to Dorflinger, revenue development growth in relation to traffic development growth shows that data traffic is increasing exponentially while voice revenue is decreasing [13]. However, the data revenue is not enough to cover the investment in infrastructure needed by CSPs. Most CSPs have had to adopt bundling as a business model to accommodate the impact of OTTs. CSPs are not only bundling their traditional services, but also adding OTT services in their bundles to attract more customers. Some CSPs have added OTTs as a marketing strategy for their services.

Bundling CSP Services with OTT Services, the Rwandan Context. The bundling of services has become an essential component of CSP business strategies. These bundling services are usually targeted towards specific groups of customers: low, medium, or high income classes, or corporate or residential customers. Bundled services help to bridge the rising gap between the usage of

OTT services and traditional telecom services like voice and SMS. Traditional telecom companies have usually offered voice, SMS, and data separately, but have now adopted a bundling model whereby voice, SMS, and data are clustered into a single package and sold to subscribers at a lower price than what they would sell for as individual packages.

Tigo-Rwanda packages different services in a single service bundle named the “Cyizere pack,” which costs 100 Rwandan francs (0.11 USD). With this pack, a customer receives 800 s of Tigo-to-Tigo voice calls, 13 SMS messages to other Tigo customers, and 3 MB of data. Customers only have 24 h to use these services. This type of service bundle costs less than the individual sum of each of the services when purchased separately.

Changes in Interactions and Targets. Under this business model, telecommunication operators have changed the way they interact with their customers. Traditionally, the telecommunications operator would install the infrastructure and wait for clients to join by subscribing for service. Using this strategy, operators have a significant amount of investment in infrastructure and other services like technical support and customer service, without a rapid return on investment. A telecommunication operator in Norway, Lyse, piloted a project to provide discounts to users who physically dug trenches for the final connections in fiber to the home projects [17]. The project improved connectivity for the local area, created a bond between Lyse and its customers—who felt that they were participating in the growth of the company—and created financial benefits for both.

In Rwanda, another alternative business model was used. The government of Rwanda partnered with Korea Telecom (KT) to provide fiber optic and 4G wireless network coverage across the whole country. Other telecom operators act as mobile virtual network operators (MVNOs) to KT, who provides 4G wholesale service. This model has resulted in improved coverage and quality of service, and a reduction in consumer prices. Incumbent operators such as MTN, Tigo, and Airtel now focus on the different value added services that they can offer their subscribers, rather than the quality of their infrastructure. The telecommunication operators are able to take advantage of the popularity of content providers to provide services to the final consumers. In return, the content providers benefit from the fact that subscribers are using their services and contribute user-generated content.

4 Discussion and Recommendations

We observe that telecommunications companies in Africa need to place innovation and collaboration with OTTs at the core of their business strategy. CSPs need to understand the needs of different customer segments and offer unique and affordable solutions. As the penetration rate of mobile phones continues to grow in Africa, CSPs must pursue continued partnerships with OTT players to provide consumers with services that are relevant to their needs. This partnership is likely to increase the growth of local content and contribute to the growth of small and medium enterprises, private content producers, and even machine-generated sensor information.

4.1 Business Model Recommendations

The global population now stands at over 7 billion people. Of these, there are 3.2 billion Internet users in the developed world and 2 billion users in developing countries [7]. About 2 billion people do not have access to the Internet [7]. Many OTTs understand that there is a large community of people who are not yet connected to the Internet, and that for them to expand their services to these people, they need to increase accessibility.

We propose a business model where CSPs collaboratively build infrastructure with OTTs as a service. Under this model, CSPs and OTTs form strategic partnerships to share resources and infrastructure in order to reduce the cost of deployment. This can create low CAPEX potential for CSPs who are now able to offer relatively cheaper services or affordable data bundles to customers. For example, in the US, Google has partnered with Sprint, T-Mobile, and US Cellular to create a Mobile Virtual Network called *Project Fi* that provides phone, messaging, and data services to consumers [9] using 3G/4G technology. The mobile subscriber's devices are able to select the carrier with a stronger signal. It is important that CSPs partner to build and share infrastructure because of the reduction in required initial capital expenditures [2].

CSPs should also consider developing mobile applications that are built around the freemium business model. These applications can provide current services such as airtime top-up and mobile money, and can include extra features which are relevant to consumers. For example, they can create an event announcement application where users can check for free what events, such as concerts and games, are happening in their area, while at the same time allow them to buy tickets using the CSPs online mobile money payment system. If this model was implemented, it has the potential to enable service providers to reach new classes of customers, increase mobile subscriptions by offering unique services, and drive demand for existing and new services.

5 Conclusion

The telecommunications market is changing quickly due to the emergence of OTTs, who are able to offer services without the bounds of regulation and infrastructure investment. They are providing services sought by customers and are improving connectivity, particularly among the underserved. CSPs have three choices: (1) to do nothing and proceed with declining traditional business models, (2) to move to block OTT services from using their infrastructure without a direct tariff, or (3) to adopt the strategies of OTTs to better serve market demand and ensure their competitiveness in the market. CSPs can also learn by benchmarking against the success of OTTs to assess and optimize the best strategies. There are opportunities for CSPs, OTTs, and consumers as long as service providers are willing to predict, innovate, reduce cost, and partner with other stakeholders.

References

1. Stork, C., Esselaar, S., Chair, C., Kahn, S.: OTT—threat or opportunity for African Telcos? Working paper, Research ICT Africa, Cape Town, South Africa, March 2016
2. Marino Garcia, J., Kelly, T.: The economics and policy implications of infrastructure sharing and mutualisation in Africa. World Development, November 2015
3. Barclay, C.: Is regulation the answer to the rise of over the top (OTT) services? An exploratory study of the Caribbean market. In: ITU Kaleidoscope: Trust in the Information Society (K-2015), Barcelona, Spain, pp. 1–8. ITU, December 2015
4. Rwanda Utilities Regulatory Authority: Statistics report RURA for the first quarter, 2015, Statistics report, Rwanda Utilities Regulatory Authority (RURA), Kigali, Rwanda, March 2015
5. Dörflinger, T.: The OTT butterfly effect—small companies with big influence. Presentation, Detecon, Bonn, Germany, February 2015
6. Sujata, J., Sohag, S., Tanu, D., Chintan, D., Shubham, P., Sumit, G.: Impact of over the top (OTT) services on telecom service providers. Indian J. Sci. Technol. **8**(S4), 145–160 (2015)
7. Ard, B.J.: Beyond neutrality: how zero rating can (sometimes) advance user choice, innovation, and democratic participation. Maryland Law Rev. **75**(4), 984–1028 (2016)
8. Niyitegeka, T.: Rwandans to benefit from free web access as Airtel partners with Facebook, July 2016. <http://www.en.igihe.com/business/rwandans-to-benefit-from-free-web-access-as.html>. Accessed 15 Mar 2017
9. Springham, J.: Mozilla and Tigo zone in on zero-rating, July 2016. <https://www.mobileworldlive.com/m360-2016-africa-article/mozilla-and-tigo-zone-in-on-zero-rating/>. Accessed 15 Mar 2017
10. Chair, C: Africa supply side assessment of zero-rating. Presentation, Research ICT Africa, Cape Town, South Africa, November 2015
11. McGrath, R.G.: Business models: a discovery driven approach. Long Range Plan. **43**(2), 247–261 (2010)
12. Barot, T., Oren, E.: Guide to chat apps. Tow Center for Digital Journalism, November 2015
13. Joshi, S., Dalal, R., Egbert, R., Chaudhary, A.: Telecom-OTT partnership: generating new revenue sharing models. Telecom Bus. Rev. **9**, 21–31 (2016)
14. National Institute of Statistics of Rwanda (NISR): Statistical yearbook: edition (SYB2016), Gov Report 1-216, National Institute of Statistics of Rwanda (NISR), Kigali, Rwanda, p. 2016, November 2016
15. Garcia-Mariñoso, B., Martinez-Giralt, X., Olivella, P.: Bundling in telecommunications. Working Paper 356, Universitat Autònoma de Barcelona, Bellaterra, Spain, January 2008
16. Australian Competition & Commission Consumer: Bundling in telecommunications markets. Information paper, Australian Competition & Consumer Commission, Melbourne, Australia August 2003
17. Schön, O., Zimmermann, P., Subrahmanyam, K.V.J.: Innovating the telecom business model: drivers and emerging trends. White paper. Capgemini, Mumbai, India, July 2011



Challenges and Opportunities in Instrumentation and Use of High-Density EEG for Underserved Regions

Ashwati Krishnan¹(✉), Ritesh Kumar³, Arnelle Etienne¹,
Amanda Robinson⁴, Shawn K. Kelly², Marlene Behrmann⁴,
Michael J. Tarr⁴, and Pulkit Grover¹

¹ Department of Electrical and Computer Engineering,
Carnegie Mellon University, Pittsburgh, PA 15213, USA
{ashwatik, pgrover}@andrew.cmu.edu

² Engineering Research Accelerator, Carnegie Mellon University,
Pittsburgh, PA 15213, USA

³ Department of Biomedical Engineering, Carnegie Mellon University,
Pittsburgh, PA 15213, USA

⁴ Department of Psychology, Carnegie Mellon University,
Pittsburgh, PA 15213, USA

Abstract. Electroencephalography (EEG) is a non-invasive method of measuring electrical signals from the brain. However, traditional clinical EEG uses only 10–40 electrodes for diagnosis which limits its potential as an imaging modality. High-density (HD) EEG, as well as the more recent Ultra-High-Density (UHD) EEG, are imaging platforms that can be used to image the brain using various techniques to solve inverse problems. These platforms comprise a measurement device and algorithms for data analysis. Recent studies have provided promising evidence that increasing the density of electrodes can improve resolution up to at least approximately 1,000 electrodes for whole-scalp coverage. Both HD and UHD-EEG can be made inexpensive and portable; therefore, perhaps most importantly, accessible to many parts of the world. However, there are remaining challenges that can hinder HD- and UHD-EEG development and use. Here, we discuss these challenges and present the approaches our research program has developed to overcome them.

Keywords: High-resolution brain imaging · Neural inference
High-density EEG

1 Introduction

Electroencephalography, or EEG, is a technique used to measure electrical signals produced by activity in the brain. EEG measurements are made by a technician, and have classically been used to diagnose neurological disorders (e.g. epilepsy, brain injuries, stroke, sleep disorders, etc.), interface machines with the brain, as well as study the brain activity noninvasively. The miniaturization of electronics has allowed EEG systems to use more electrodes, and has opened the possibility of using EEG as a

portable imaging technology, even if the obtained spatial resolution is lower than that of less portable technologies (e.g. Magnetoencephalography, MEG, and functional Magnetic Resonance Imaging, fMRI). Roughly, a system that packs 64–256 electrodes on the scalp is referred to as a **high density** (HD)-EEG, and even higher-density systems, reaching as many as 1,000 electrodes, are possible and considered Ultra-High-Density (UHD)-EEG. Recently, our research has established, both theoretically [1, 2] and experimentally [3], that a UHD-EEG system provides higher spatial resolution of neural signals than its lower density counterparts.

HD-EEG systems are non-invasive, and require only inexpensive sensors and electronic circuits, making them a low-cost solution for diagnosis of neural disorders, interfacing with the brain, and neuroscientific studies. On the spectrum of available imaging technologies, shown in Table 1, there are three main parameters that we consider for comparison: temporal resolution, spatial resolution and cost. For instance, fMRI provides very high spatial resolution images, but these are obtained very slowly (>1 s), making conditions such as seizures hard to detect and/or localize. On the other hand, MEG has the same temporal resolution as EEG and a high spatial resolution, but MEG systems cost millions of US dollars; and only a few hundred machines are in use across the world. Moreover, both fMRI and MEG systems are non-portable because they require bulky shielding from earth’s magnetic field and, typically, superconducting subsystems that necessitate coolants such as liquid helium. In contrast, HD-EEG can be made relatively inexpensively *and* portable, and thus can be accessible to many more communities around the world.

While HD-EEG demonstrates a great deal of promise as a viable brain imaging modality, there are a number of emerging obstacles that limit the scope of its use in clinical settings, particularly in developing countries with limited resources. In this paper, we do not intend to provide a comprehensive survey of how HD/UHD-EEG has developed. Instead, it is a biased perspective on how our team has been pursuing

Table 1. Available imaging technologies versus high density EEG (HD-EEG)

Imaging Modality	Cost of the Machine (USD)	Temporal Resolution (seconds)	Spatial Resolution (mm)
Magnetic Resonance Imaging (MRI)	\$0.5M - \$3M	200 - 1000	0.5 - 2.8
Functional MRI (fMRI)	\$0.5M - \$3M	0.05 - 5	2 - 4
Computed Tomography (CT)	\$50,000 - \$0.5M	0.1 - 1	0 - 1.5
Positron Emission Tomography (PET)	\$100,000 - \$0.5M	10 - 1000	2.5 - 5
Functional Near Infra-Red Spectroscopy (fNIRS)	\$15,000 - \$0.2M	0.1 – 1	5 - 10
Magnetoencephalography (MEG)	\$1M - \$2M	0.5m – 1m	5 - 28
Electroencephalography (EEG)	\$3000 - \$10,000	0.5m – 1m	55 - 120
High Density-EEG? (anticipated)	\$3000 - \$10,000	0.5m – 1m	5 - 20

research and advancement of high-resolution EEG inference, and understanding its implications in clinical science. We provide an overview of some of the challenges we observed in instrumentation and platform development of HD/UHD-EEG, and briefly discuss how we have addressed or are trying to address these issues. More detailed literature review on HD/UHD-EEG can be found in [1, 3].

2 Background

2.1 Origin, Measurement and Uses of EEG

EEG signals primarily originate from synchronized synaptic activity of neurons in the brain. The signal that is measured by an electrode placed on the scalp is the sum of positive and negative charges within the brain. A typical clinical EEG measurement system is non-invasive and uses anywhere from 10 to 40 electrodes spread across the scalp. Commonly, these systems use “wet” electrodes (i.e., where electrode-skin contact is gel-based) to reduce impedance of the electrode skin contact. EEG recording electrodes in clinical settings often involve abrasion of the skin to remove the epidermal layer because it has a higher impedance than the underlying tissue. Disinfection as well as sterilization is required for EEG electrodes because of the abrasion under CDC guidelines.

EEG systems are used for observing different types of brain signals such as temporal EEG signatures, frequency-domain signatures, event related potentials (ERP) [4], and steady-state visually evoked potentials (SSVEPs) [5]. A common use-case of EEG is in the detection of epileptic seizures which appear as rapidly oscillating waves; indeed, EEG is the gold standard in epilepsy diagnosis. Further, source localization techniques using EEG are used to identify the foci of seizures, so that they can be surgically removed. In neuroscience, EEG is used to study the neural mechanism underlying cognitive processes such as attention, learning, reading, memory, etc. ERPs and SSVEPs are recorded from the continuous stream of EEG data which allows characterization of brain processes triggered by the specific events. More recently, the use of EEG has become prevalent in Brain Machine Interfaces (BMI) [6]. In this section, we briefly describe existing methods of EEG measurement and the applications of high-density EEG.

2.2 High-Density EEG

EEG signals propagate through cerebrospinal fluid, skull and finally the scalp, where they are measured. Low electrode count EEG systems do not measure high spatial frequency signals from the brain because they get severely attenuated through these layers, which act as spatial low-pass filters [1, 7] (Fig. 1).

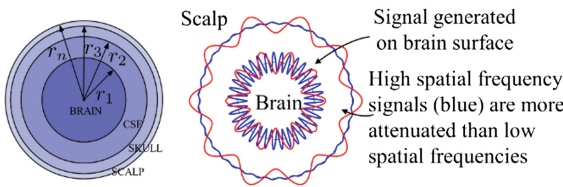


Fig. 1. Computer simulations of a spherical head model depicting the low pass filtering of spatial high frequency signals that originate in the brain. High-density EEG increases the number of electrodes over scalp to reliably reconstruct high spatial frequency signals from within the brain.

The main advantage of EEG over other imaging systems such as fMRI and fNIRS (functional Near InfraRed Spectroscopy) is that EEG has a much higher temporal resolution. By increasing the number of electrodes, we aim to improve the spatial resolution as well, thereby creating a new neuroimaging modality in HD/UHD-EEG. The combination of high temporal and spatial resolution allows a myriad of new applications. For instance, HD-EEG may allow improved reliability of detecting a phenomenon known as Cortical Spreading Depolarization (CSD). CSDs are waves of neural silencing that spread slowly across the brain surface and manifest in the presence of a traumatic brain injury, stroke, hemorrhages, and even migraines [8].

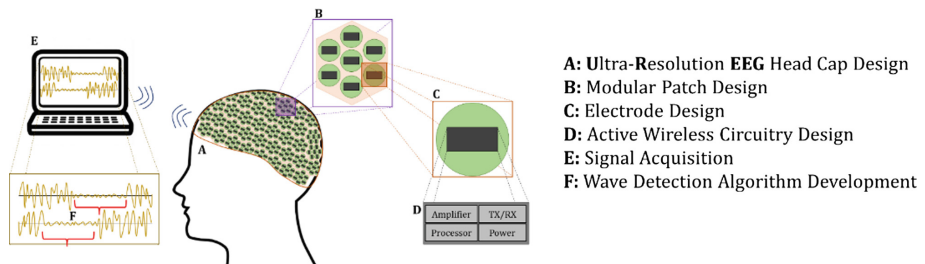


Fig. 2. High Density Electroencephalography platform design: the EEG signal is measured from a high density cap (A) consisting of more than 500 electrodes. These electrodes are modularized into patches (B). Each electrode needs to have a low impedance interface to the scalp (C). The signal-to-noise ratio of the measurement increases if the electrodes are active (D). Once the signal has been acquired, anomalies are detected using novel algorithms (F).

A typical HD-EEG system is shown in Fig. 2. There is considerable effort involved in designing electrodes, the high-density head cap, scaling the active electrodes to a wireless portable system, and designing effective algorithms for studying the measured data. In Sect. 3, we briefly discuss some of the impediments that limit the development of HD-EEG, as well as a description of our approach to address each of these challenges.

3 HD-EEG: Implementation Challenges

Designing HD-EEG systems should be holistic, invariably involving an interdisciplinary approach. Our research simultaneously approaches issues fundamental understanding, algorithms, neuroscientific validation, instrumentation, and scalability.

3.1 Electrodes and Gel Developments

In order to measure electrical signals from the brain, electrodes should have good contact with the skin [9]. The range of EEG voltage is small (order of 10 μV), and for a high signal-to-noise (SNR) ratio, the electrode-skin impedance must be low (order of 5 k Ω). Because the skin is a hydrated material, a low impedance is often obtained by a combination of abrading the skin, and using a conducting gel between the electrode and the scalp. Dry metal electrodes that are often used in many consumer EEG systems have high impedance and hence low signal-to-noise ratio.

Wet Electrodes Dry Up for Chronic Use. Conditions such as epilepsy require long term EEG monitoring to detect the onset of a seizure. One of the main challenges here is the frequent drying up of the electrode gel, resulting in a non-conducting crystalline residue at the electrode-skin interface. Because there is an increase in impedance, the measured signal is unreliable. With HD-EEG, it is impractical to repeatedly apply gel to 500+ electrode locations.

Our Approach: We believe this is one of the most underappreciated research challenges for practical, long-term use of low and high density EEG systems, and requires bringing together expertise from several disciplines. Our research has explored the development of new materials that can remain hydrated for extended time.

Different Hair Types Prevent Reliable Measurement. The goal of HD-EEG is to have an easily accessible diagnostic imaging modality. Therefore, it is important to consider the social aspects of getting a reliable EEG measurement. While patients in critical care might have their heads shaved in order to get access to the scalp, patients who have come for a preliminary EEG measurement may not wish to do so. Existing EEG systems do not work reliably with thick, coarse, curly, or long hair. Moreover, strands of hair between two electrodes can enable gel flow between them through the capillary effect, thereby shorting two electrodes. A high-density EEG makes the hair issue more of a challenge.

Our Approach: Our research is trying to look at the problem of hair under the EEG electrode from a different perspective. Instead of only designing an EEG device that works with the hair, we can also modify the hair to work with the device. The key idea is to maximize the contact made by the EEG electrodes with the scalp by altering the position of the hair (Fig. 3).



Fig. 3. Traditional EEG gel based EEG methods do not work satisfactorily with different hair types (*left*). Modification of hair structure can provide reliable methods of obtaining stable EEG measurements.

Identification of Bridged Electrodes for High Densities. One major problem with high-density electrode arrays and caps is the bridging that occurs between adjacent electrodes, implying that these two electrodes electrically behave as one large electrode. It is pertinent to identify which electrodes are bridged during measurement so that there is no loss of spatial information from the EEG signal.

Our Approach: We have been able to successfully prototype arrays with pre-load our high-density arrays with high viscosity conductive gels (Ten20 Conductive paste). These instrumentation methods are supplemented with algorithms to identify bridged electrodes after data acquisition through correlation studies [10].

Electrode Configuration and Caps for high Density Arrays. Traditional EEG locations are based on the 10–20, 10–10 or 5–10 electrode placement protocols [11]. With high-density EEG, individually preparing 300+ electrodes with these protocols is impractical. Moreover, injured parts of the head cannot be covered up by an electrode cap. There also needs to be a way to apply pressure on the scalp so that the electrodes do not move. If there is movement, that could mean shorting of adjacent electrodes, lowering spatial resolution.

Our Approach: Away to approach this is through modularization over existing electrode locations. Designing high-density electrode patches will ensure that areas of the scalp that do not need to be measured can be left out, as in the case of a brain injury. An interlocked grid pattern allows us to fasten different portions of the cap quickly, and these modules can be pre-loaded with high viscosity electrode gels. Moreover, our research has shown that using hierarchical referencing techniques within these electrode modules [12] enables low power active circuitry at the electrode site (Fig. 4).

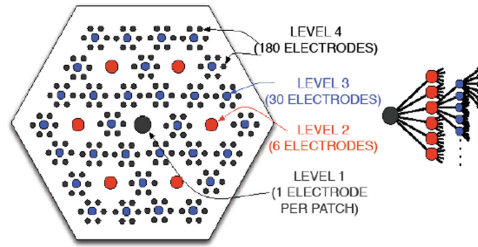


Fig. 4. Hierarchical referencing for HD-EEG electrode arrays: modularization of EEG electrode arrays enables the use of hierarchical referencing [9]. The method decreases the length of wires routing from the electrode to the back-end circuit, thereby increasing the signal-to-noise ratio of the measurement.

3.2 Algorithm and Neuroscientific Development

We have thus far highlighted the most relevant obstacles that stands in the way of *instrumenting* a high-density EEG measurement device. The device by itself is limited in its use without the accompanying algorithms that process the data obtained from the EEG sensors, using validation from neuroscientific studies.

Design of Detection Algorithms with HD-EEG. Traditional EEG and MEG imaging systems use a combination of algorithms to analyze data. It is often useful to project data in brain space (instead of sensor space) by use of one of many “source localization” or imaging algorithms. Use of “surface Laplacian,” a spatial filter that approximately projects scalp signal on to the dura has been shown to be useful [7]. However, the precision of these techniques is limited, particularly for lower density systems.

Our Approach: For high-density EEG, our simulation-based results [1] demonstrate that increasing sensor count will increase resolution and narrow the region of uncertainty, i.e., the width of the point-spread-function in reconstructing a single dipole [2]. It is plausible that to best exploit the information provided by HD-EEG, novel algorithms would be needed, and this warrants further study.

Neuroscientific Validation of HD-EEG. Traditionally, EEG has been used for low-spatial precision measures such as event-related potentials (ERPs), frequency-domain power (e.g., power in alpha, beta, gamma, delta bands and their ratios), coherence measures, steady-state visually evoked potentials (SSVEPs), etc. While source-localization and imaging techniques have been around for many decades, only recently have they been adopted in practice, and their use continues to be limited. Scientifically validated research with high-density EEG and ultra high-density EEG are limited (Fig. 5).

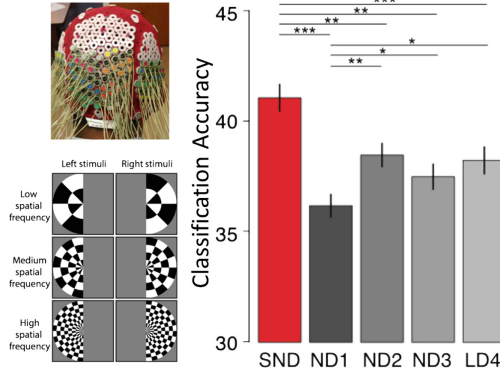


Fig. 5. High-density EEG improves classification of high spatial resolution images beyond different configurations of Nyquist rate spatial sampling [3]. Here, SND denotes Super Nyquist Density implying a higher density of electrodes when compared to Nyquist Density (ND).

Our Approach: Our recent work [3] has established that HD-EEG, with sub-cm spacing of electrodes, offers higher spatial resolution than existing systems, superseding even those with 128 or 256 electrodes in 10–20 arrangement. Our simulations also reveal that HD-EEG’s localization accuracy will increase with greater electrode counts. Further experimental validation is needed to quantify limits of HD-EEG in practice.

Clinical Adaptation of HD-EEG in Hospitals. Although research studies have used higher densities (72–128 electrodes), there are few validated studies for the clinical use of HD-EEG.

Our Approach: Our aim here is to identify disorders that can benefit from higher spatiotemporal resolution monitoring. Disorders such as epilepsy, migraine, stroke, subarachnoid hemorrhage, and traumatic brain injury, can benefit from such monitoring. For example, for epilepsy, improved source localization, detect multiple seizure foci, or even improved depth inference, can alter treatment approaches [13]. For other disorders as mentioned above, detecting wave of Cortical Spreading Depolarizations (CSDs) can help understand the severity of the disorder, and also help manage and treat it [8]. CSDs are narrow waves of neural silencing that spread slowly on the cortical surface, and only recently was it shown that they can be detected using low-density EEGs, although the reliability of this detection was low [14], especially for narrower waves. In a recent study, our team demonstrated through rigorous simulations that even narrower CSDs can be detected reliably using HD-EEGs by stitching together data across time and space [15]. We expect other such biomarkers to be discovered in near future where HD-EEG is applicable, which can lead to novel HD-EEG designs and solutions for diagnoses and treatments.

3.3 Scalability and Accessibility

A successful technology platform needs to be scalable to be useful to society. One of the most attractive features of HD-EEG is that it can be made low-cost and portable. Miniaturization of electronics has made scalability complex, but not impossible.

Design of a Wireless HD-EEG Device. Traditional EEG devices attach long shielded lead wires to each scalp electrode, and the wire bundle extends back to instrumentation with amplifiers and analog-to-digital converters (ADCs). Some newer EEG systems utilize small pre-amplifier circuits on top of each electrode to boost the EEG signals before sending them through the wires, reducing the effects of noise [16]. But all of these systems require the subject to be tethered to the measurement system, and restrict movement.

Our Approach: We are developing a wearable EEG system with wireless data transfer capabilities. Wearability requires miniaturized electrodes, amplifier and ADC circuitry, wireless radio, power, and packaging, while maintaining costs that are competitive with existing EEG systems. We are using off-the-shelf amplifier, ADC chips, off-the-shelf wireless radios, and utilizing our group's experience in ultra-small circuit boards and assembly techniques to make the overall EEG system lightweight, wearable, and cost-competitive.

FDA Device Classification. Although the HD-EEG portable platform is non-invasive, it can be used to diagnose significant brain illnesses, and introduces an element of criticality to its device classification.

Our Approach: Because it is non-invasive, it will likely be a Class II device according to the US FDA. Our initial discussions with regulatory consultants regarding this device suggest that it may qualify for a 510(k) clearance, i.e., it is substantially equivalent to a similar legally marketed device. However, it remains to be understood what clinical trials may be necessary.

4 Discussion and Conclusions

The development of a new medical technology is exciting because of the potential it has to improve the quality of life to the people who use it. Electroencephalography (EEG) is a classical and well accepted modality used for brain imaging. For example, in the case of epilepsy, which affects 1 in 17 people in developing countries [17], EEG is one of the three pillars that guide its treatment (along with structural MRI and patient semiology). Our research has shown that traditional EEG is limited in its scope because only 10–40 electrodes are regularly used in clinics. Increasing the spatial resolution to about 500+ electrodes, while utilizing the high temporal resolution of EEG is the reason to advance to High Density EEG (HD-EEG).

HD-EEG is a non-invasive, low-cost and portable imaging technology that can be used globally to diagnose disorders such as epilepsy, stroke and traumatic brain disorders. These conditions often go unnoticed early on due to lack of resources, only to become more severe later. Other imaging techniques such as (functional) Magnetic

Resonance Imaging (fMRI), Magnetoencephalography (MEG) are very equipment intensive and expensive and are not easily accessible to a large section of world population. HD-EEG can be made portable and inexpensive because the accompanying electronics is a well-matured field, and has been applied judiciously in several biomedical devices such as pacemakers and fitness monitors. Scaling a 40 electrode system to a 500+ one, while maintaining the same surface area (i.e., the human scalp) is a challenge we are addressing through both engineering and algorithmic improvements. This paper highlighted a few pertinent points of interest that arise when designing a HD-EEG system. We described factors in electrode and gel design - which included issues such as having a portable HD-EEG device work with different types of hair without bridging. We also emphasized a critical need for simultaneous development of algorithms and neuroscientific verification to process the data that is acquired.

The study of EEG measurements for diagnostics is embedded in clinical training all around the world. Subsequently, HD-EEG is conducive to massive adoption as compared to other modalities because of familiarity, training, immense historical knowledge and data, while delivering resolutions comparable to or even superseding competing modalities. Development of HD-EEG is an immense interdisciplinary effort by engineers, clinicians and neuroscientists. The challenges outlined in this article are not fundamentally insurmountable, and ongoing research continues to make significant strides in overcoming them to achieve unparalleled efficiency for low-cost imaging platforms.

References

1. Grover, P., Venkatesh, P.: An information-theoretic view of EEG sensing. *Proc. IEEE* **105**, 367–384 (2017)
2. Venkatesh, P., Grover, P.: Lower bounds on the minimax risk for the source localization problem. In: *IEEE International Symposium on Information Theory*, pp. 3080–3084 (2017)
3. Robinson, A.K., Venkatesh, P., Boring, M.J., Tarr, M.J., Grover, P., Behrmann, M.: Very high density EEG elucidates spatiotemporal aspects of early visual processing. *Sci. Rep.* **7**, 16248 (2017)
4. Luck, S.J.: *An Introduction to the Event-Related Potential Technique*. MIT Press, Cambridge (2005)
5. Norcia, A.M., Appelbaum, L.G., Ales, J.M., Cottareau, B.R., Rossion, B.: The steady-state visual evoked potential in vision research: a review. *J. Vis.* **15**, 4 (2015)
6. Nicolas-Alonso, L.F., Gomez-Gil, J.: Brain computer interfaces, a review. *Sensors (Basel)* **12**, 1211–1279 (2012)
7. Nunez, P.L., Srinivasan, R.: *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, Oxford (2006)
8. Lauritzen, M., Dreier, J.P., Fabricius, M., Hartings, J.A., Graf, R., Strong, A.J.: Clinical relevance of cortical spreading depression in neurological disorders: migraine, malignant stroke, subarachnoid and intracranial hemorrhage, and traumatic brain injury. *J. Cereb. Blood Flow Metab.* **31**, 17–35 (2011)
9. Tronstad, C., Johnsen, G.K., Grimnes, S., Martinsen, Ø.G.: A study on electrode gels for skin conductance measurements. *Physiol. Meas.* **31**, 1395–1410 (2010)

10. Alschuler, D.M., Tenke, C.E., Bruder, G.E., Kayser, J.: Identifying electrode bridging from electrical distance distributions: a survey of publicly-available EEG data using a new method. *Clin. Neurophysiol.* **125**, 484–490 (2014)
11. Jurcak, V., Tsuzuki, D., Dan, I.: 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage* **34**, 1600–1611 (2007)
12. Grover, P., Weldon, J.A., Kelly, S.K., Venkatesh, P., Jeong, H.: An information theoretic technique for harnessing attenuation of high spatial frequencies to design ultra-high-density EEG. In: 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 901–908. IEEE (2015)
13. Kumar, R., et al.: Ultra-high-density scalp EEG outperforms localized invasive ECoG grids in inferring depth of seizure foci. In: 31st International Congress of Clinical Neurophysiology (2017)
14. Hartings, J.A., et al.: Spreading depression in continuous electroencephalography of brain trauma. *Ann. Neurol.* **76**, 681–694 (2014)
15. Chamanzar, A., George, S., Venkatesh, P., Ding, W., Grover, P.: Systematic and automated algorithms for detecting cortical spreading depolarizations using EEG and ECoG to improve TBI diagnosis and treatment. *Brain Inj.* **31**, 990 (2017)
16. Metting Van Rijn, A.C., Kuiper, A.P., Dankers, T.E., Grimbergen, C.A.: Low-cost active electrode improves the resolution in biopotential recordings. In: 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 101–102. IEEE (1996)
17. Senanayake, N., Román, G.C.: Epidemiology of epilepsy in developing countries. *Bull. World Health Org.* **71**, 247–258 (1993)



E-Medicine: A Secure Transmission of Electrocardiograms Using Chaotic Oscillators Synchronization

Alain Tiedeu^(✉), Yannick Abanda, and Gutenbert Kenfack

LAGEMES, National Advanced School of Engineering,
University of Yaoundé 1, Yaoundé, Cameroon
alain.tiedeu@polytechnique.cm

Abstract. Telemedicine is developing at high speed. In this context, patient's privacy and security is of great importance. Therefore any physiological signal, needs to be encrypted before their transmission over any channel. In this paper, we have developed an encryption system using chaotic synchronization to encrypt and decrypt information. The system was used for secure transmission of electrocardiograms signals as example.

Keywords: Electrocardiograms · Chaos synchronization · Secure transmission

1 Introduction

E-medicine uses information technologies to deliver health services. Its goal is to widen the access to medical services. As communication costs are growing cheap, e-medicine is becoming widely affordable (Moore 2002). Recently, advances in telecommunications networks have led to many successful e-medicine experiences around the world. Kontaxakis et al. (2000) developed an e-medicine workstation to acquire process and transmit ultrasonic images while Sachpazidis and Hohlfeld (2005) proposed a communication system for medical applications. The vast indian subcontinent has witnessed telemedicine success stories (Ayyaga et al. 2003; Pal et al. 2005; Deodhar 2001). Remote monitoring of patients is more and more common due either to an aging patient population, long distances to cover to find well equipped health centers or the need to decrease healthcare costs. These informations often transit through public channels with risk of being hijacked, intercepted, etc. This makes the need for encryption or other protection techniques crucial. Unfortunately, the works mentioned above and others in the field of telemedicine have not addressed the concern over protection of patient health information. We intend to do it in this paper.

Electrocardiograms (ECG) have a dual nature in the fact that they are used for both medical and identification purposes (Sufi et al. 2011; Almehmadi and Chatterjee 2015). A literature review on ECG encryption reveals methods including permutation encoding, wavelet anonymization, and noise-based obfuscation, just to name a few. Chaos-based encryption has an advantage over the other schemes because it is applicable to continuous signals, possesses a highly unpredictable nature, is sensitive to initial conditions and other key parameters.

To the best of our knowledge, secured ECG signal transmission with chaotic oscillators of different natures at emission and reception ends has not been studied. In this paper, we bridge this gap. Firstly, we develop an active control based strategy to synchronize a Colpitts and a Hartley oscillator. Secondly, we carry out encryption of ECG signal by the Colpitts oscillator, send it through the channel and decrypt it using the Hartley oscillator. The block diagram of the proposed system is given in Fig. 1.

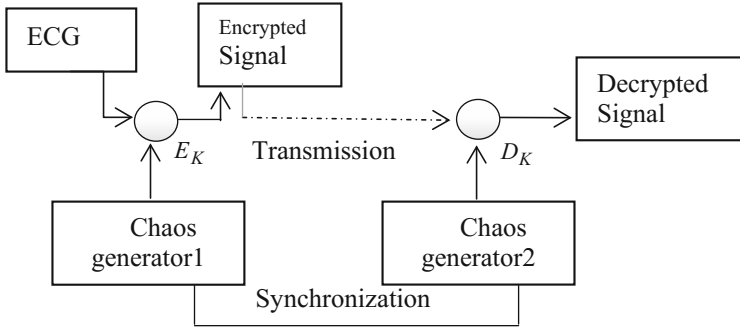


Fig. 1. ECG secured transmission system

1.1 Presentation of the Encryption System

The encryption/Decryption system is made of the ECG source, two chaotic generators and a transmission system (line). The signal from the ECG source is multiplied by the output of chaotic generator 1, modulated then transmitted. At the reception end, the output of chaotic generator 2 which is synchronized with chaotic generator 1 is used to decrypt the signal by simple division of received signal after it has undergone detection. The different elements of the system are described below.

1.2 Chaotic Oscillators

In this work, we use two different chaotic oscillators, namely, Colpitts (generator 1) and Hartley (generator 2) oscillators. These are drawn in Figs. 2 and 3 below.

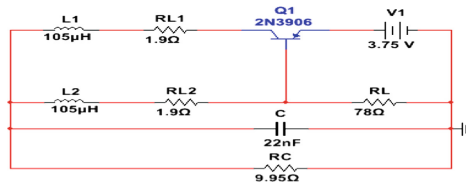


Fig. 2. Hartley oscillator

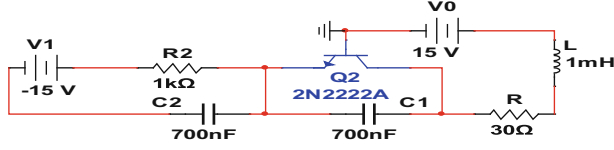


Fig. 3. Colpitts oscillator

1.3 Oscillators Dynamics

Applying Kirchoff voltage law to the circuit in Fig. 2 and changing variables, we obtain dimensionless equations:

$$\begin{cases} \dot{x}_1 = y_1 - z_1 - ax_1, \\ \dot{y}_1 = q - x_1 - by_1 - F(z_1), \\ \dot{z}_1 = dx_1 - ez_1 + F^*(z_1). \end{cases} \quad (1)$$

with

$$F(z_1) = \begin{cases} -\frac{1}{V_{TH}}(h + m \times z_1), & z_1 \geq lamda \\ s, & z_1 < lamda \end{cases} \quad \text{and} \quad F^*(z_1) = \begin{cases} 0, & z_1 \geq lamda \\ f(g + a_1 z_1), & z_1 < lamda \end{cases}$$

Then, when we apply the Kirchhoff voltage law to the circuit in Fig. 3 and carry out a change of variables, we obtain the following:

$$\begin{cases} \dot{x} = y - a_2 \vartheta(z), \\ \dot{y} = c - x - z - b_1 y, \\ \dot{z} = \varepsilon(y - d_1). \end{cases} \quad (2)$$

where

$$\vartheta(z) = \begin{cases} -(1 + z) & z < -1, \\ 0 & z \geq -1. \end{cases}$$

Solving (1) and (2) numerically using 4th order Runge-Kutta, yields the dynamics of the oscillators. The phase portraits obtained are shown in Figs. 4 and 5.

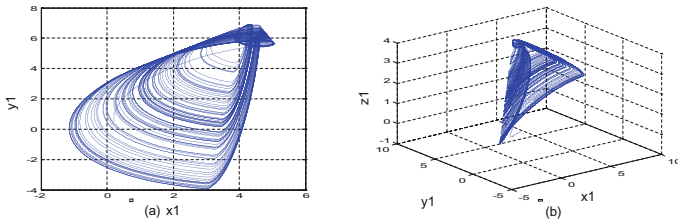


Fig. 4. (a) 2D phase portrait for Hartley oscillator. (b) 3D phase portrait for Hartley oscillator.

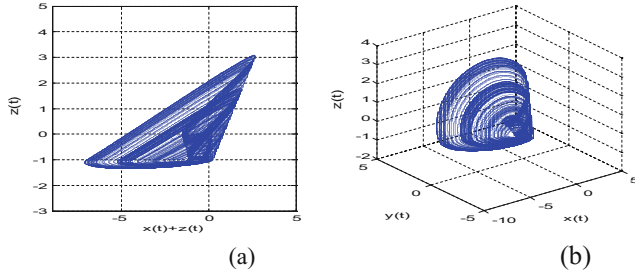


Fig. 5. (a) 2D phase portrait for Colpitts oscillator. (b) 3D phase portrait for Colpitts oscillator.

The phase portraits of Figs. 4 and 5 are strange attractors and indicate the possibility of chaotic behavior. A common method to confirm chaotic dynamics is to compute Maximum Lyapunov Exponent (MLE). The dynamics of MLE below (Figs. 6 and 7) confirm the chaotic nature of the oscillators.

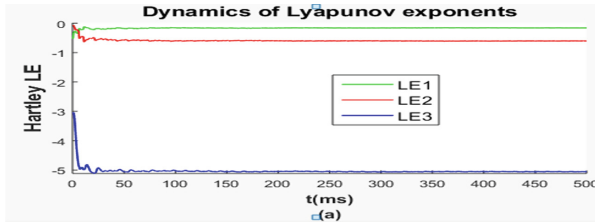


Fig. 6. Dynamics of Lyapunov exponents for Hartley oscillator

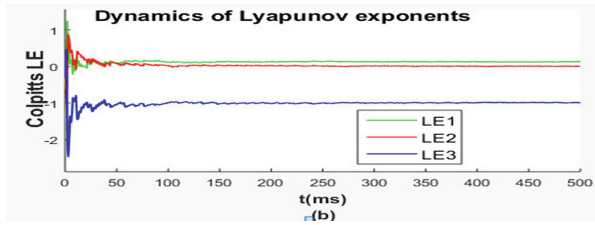


Fig. 7. Dynamics of Lyapunov exponents for Colpitts oscillator

2 Synchronization

Simply put, synchronizing the two oscillators is making sure that their output have the same values with time. Ideally, the difference should be zero. But practically, a “very small” error is enough. The first oscillator is the drive, while the second is the response. The controller U is the system that ensures the synchronization while the synchronization error is e .

$$\text{For the drive system: } \begin{cases} \dot{x} = y - a_2 \vartheta(z), \\ \dot{y} = c - x - z - b_1 y, \\ \dot{z} = \varepsilon(y - d_1). \end{cases} \quad (3)$$

$$\text{And the response system: } \begin{cases} \dot{x}_1 = y_1 - z_1 - ax_1 + U_1(t), \\ \dot{y}_1 = q - x_1 - by_1 - F(z_1) + U_2(t), \\ \dot{z}_1 = dx_1 - ez_1 + F^*(z_1) + U_3(t). \end{cases} \quad (4)$$

$U(t) = [U_1(t), U_2(t), U_3(t)]^T$ being the controller.

The synchronization error is: $e_1 = x_1 - x, e_2 = y_1 - y, e_3 = z_1 - z$.

Let's choose the controller described by Eq. (5)

$$\begin{cases} U_1(t) = -e_2 + z_1 + ax - a_2 \vartheta(z) \\ U_2(t) = -y(b_1 - b) + e_1 - z + F(z_1) \\ U_3(t) = -dx_1 + ez + \varepsilon(y - d_1) + F^*(z_1) \end{cases} \quad (5)$$

Simulations were then carried out based on this controller and the state variables for drive and response respectively. The error was evaluated and the system was finally used to encrypt then decrypt the ECG signal, when the computed error converged towards zero. The following section presents some results yielded by the system.

3 Results and Discussion

In this section we shall present results from the synchronization, encryption and finally decryption.

3.1 Synchronization

The decrypted signal will be as close to the original one as far as the synchronization is accurate.

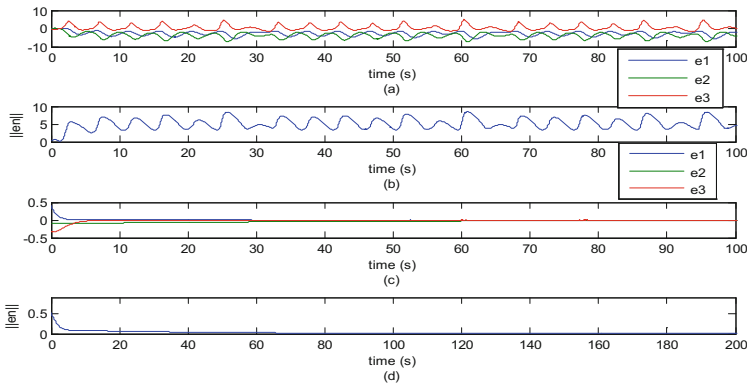


Fig. 8. (a) Synchronization error e_1, e_2, e_3 when the controller U is not activated, (b) the error's norm $\|e_n\|$ when the controller U is not activated (c) synchronization error e_1, e_2, e_3 with the controller U activated, (d) the error's norm $\|e_n\|$ with the controller U activated.

Figure 8 is a plot of the variable, the error when there is no controller and then the error when the oscillators are synchronized. We can see the error converging towards zero in event of synchronization.

3.2 Encrypted and Decrypted Signals

Figures 9 displays a visual example of encrypted, then decrypted signals.

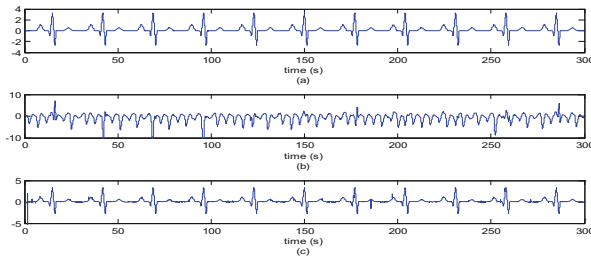


Fig. 9. (a) Original ECG signal (b) encrypted ECG signal, (c) decrypted ECG

We can see from Fig. 9 that the ECG signals are first encrypted, then decrypted correctly by the proposed system. Visually, these first results are satisfactory but will need in future works to be confirmed by some metrics like signal over noise ratio or/and mean square error.

4 Conclusion

In this work, we have designed and proposed an encryption and decryption system based on synchronization of chaotic oscillators. This was applied to the secured transmission of ECG signal. Results yielded by our system are encouraging and we hope to implement the experimental version in our future works.

References

- Moore, S.K.: Extending health care reach. *IEEE Spectr.* **39**(1), 66–71 (2002)
- Kontaxakis, G., Walter, S., Sakas, G.: EU-TeleInViVo: an integrated portable telemedicine workstation featuring acquisition, processing and transmission over low-bandwidth lines of 3D ultrasound volume images. In: *Proceedings of IEEE EMBS International Conference on Information Technology Applications in Biomedicine*, pp. 158–163 (2000)
- Sachpazidis, I., Hohlfeld, O.: Instant messaging communication gateway for medical applications. In: *IATED International Conference on Telehealth*, Banff, Canada, 19–21 July 2005, pp. 12–16 (2005)
- Ayyagari, A., et al.: Use of telemedicine in evading cholera outbreak in Mahakumbh Mela, Prayag UP India: an encouraging experience. *Telemed. J. E. Health* **9**, 89–94 (2003)

- Pal, A., Mbarika, V.W., Cobb-Payton, F., Datta, P., McCoy, S.: Telemedicine diffusion in a developing country: the case of India (March 2004). *IEEE Trans. Inf Technol. Biomed.* **9**(1), 59–65 (2005)
- Deodhar, J.: Telemedicine by email—experience in neonatal care at a primary care facility in rural India. *J. Telemed. Telecare* **8**, 20–21 (2001)
- Sufi, F., Han, F., Khalil, I., Hu, J.: A chaos-based encryption technique to protect ECG packets for time critical telecardiology applications. *J. Netw. Comput. Appl.* **4**, 515–524 (2011)
- Almehmadi, F.S., Chatterjee, M.R.: Secure chaotic transmission of electrocardiography signals with acousto-optic modulation under profiled beam propagation. *Appl. Opt.* **54**(2), 195–203 (2015)



Bringing Life Where There Is No Light: A Low-Cost Movie Projection and Data Collection Solution for Underserved Areas

Alain Shema^{1(✉)} and Michel Bézy²

¹ Syracuse University, Syracuse, NY 13244, USA
sralain@syr.edu

² Carnegie Mellon University Africa, Kigali, Rwanda
bezy@andrew.cmu.edu

Abstract. Limited access to electricity in rural Africa impedes the human and economic development in these areas by restricting access to information. In addition, data collection in this region remains a costly and time consuming endeavor. However, recent developments in solar energy and the fast adoption of mobile phones present new opportunities to reach people living in African villages. In this paper, we present an innovative solution that combines a relatively inexpensive, solar-powered movie projection kit with a simple USSD based mobile application to provide entertainment in villages, while collecting data in a cost effective, real-time manner. The results from a pilot project in 8 villages where we deployed this solution highlight the possibility of obtaining detailed and instantaneous market data for advertisers, whilst simultaneously entertaining audiences with high quality content. We also discuss additional possible uses for the solution in other sectors such as health and education.

Keywords: ICT innovation · USSD · ICTD · M4D
Rural entertainment

1 Introduction

According to the World Bank, more than 82% of the populations living in rural Sub-Saharan Africa lack access to electricity [1]. While access to electricity is critical for many basic human activities, such as lighting, refrigeration, operating equipment, etc., the absence of electricity also impedes another important human need, namely access to information. This information is today perceived as crucial to human and economic development, and as having the ability to increase empowerment in rural areas through greater levels of community connectivity [3].

In spite of this challenge, countries in Sub-Saharan Africa have witnessed rapid growth in mobile phone subscriptions, which have climbed from a paltry 0.5% of the population in 2000 to 71.3% in 2016 according to data from the

International Telecommunication Union (ITU). This mass adoption of mobile phones presents a major opportunity for access to Digital Media. At the same time, the development of a Digital Media industry in Africa, and in Rwanda in particular, is constrained by the limited number of distribution outlets. Indeed, according to the Demographic and Health Survey conducted by the National Institute of Statistics of Rwanda (NISR) in 2014 and 2015, only 10% of households in Rwanda owned TV sets. This number is even lower in rural areas, where only approximately 4% of households owned TV sets [6].

We attempted to address these challenges in two ways. First, we built a video projection kit that is robust, relatively inexpensive and powered by solar energy. Second, we developed a mobile application to collect feedback from viewers. This mobile application used the Unstructured Supplementary Service Data (USSD) communication technology in order to be accessible by virtually every mobile phone, including feature phones. In this paper, we present the results of a pilot study, called the “Village Mobile moVie (VMV)”, that we conducted in 8 villages in Rwanda. Here, we projected movies using the video kit and collected feedback from viewers through the mobile application. We discuss the implications of these results and avenues for future work below.

2 Background

2.1 Media Distribution in Rural Rwanda

Rural Rwanda remains relatively disconnected from the mainstream audiovisual media distribution channels. For example, the government reported that in 2014 and 2015, only 4% of rural households owned a television set, whereas 52% owned a radio receiver [6]. This limited adoption of media receivers contributes to the isolation of rural dwellers as well as impeding the development of a local Digital Media industry. To address the low rate penetration of television sets, a small number of entrepreneurial people in rural communities may purchase television sets that they use as a sort of cinema, where people can watch movies and shows for a fee. However, the small sizes of the television sets offer poor viewing conditions to the relatively large audience, who require a big screen.

Other efforts to project movies in rural areas include the annual Rwanda Film Festival held in July, which uses a large inflatable screen and a professional projector powered by a diesel generator. Nonetheless, the high cost of the equipment used (around \$30,000 per kit) severely restricts the spread of this type of initiative in the country. Therefore, there is a need for a projection kit that is relatively inexpensive in order for it to be affordable to entrepreneurs in villages.

2.2 Unstructured Supplementary Service Data

Unstructured Supplementary Service Data (USSD) is a GSM communication method that enables fast exchange of text between a mobile phone and an application on the network. Unlike the Short Message Service (SMS), USSD is a

session-oriented service that provides “minimal delay between the sending and receiving of queries” [9]. These characteristics allow the creation of menu-based, interactive applications that can enrich the user’s experience.

Perrier et al. argue that USSD is a better suited method for data collection in regards to ICTD [7]. Like voice and SMS, USSD is a universal service and is accessible from any phone as it is part of the GSM specifications. Data collection through SMS presents a number of constraints such as the requirement to send text in a specific format in order for it to be understood by an SMS application. This creates a need for external systems, such as paper-based systems, that support the creation of well-formatted SMS messages [10]). However, USSD improves the user’s experience during data collection in a number of key ways. Firstly, it provides forms that guide the user. In addition, it removes the need for multiple messages when providing multiple answers. Finally, USSD allows for the immediate validation of data during the collection process.

These advantages have made USSD an effective platform for building mobile applications that can reach virtually every mobile phone user. They have also enabled it to be used in a number of different domains such as the health sector [5], mobile money services [4], and even in mobile crowdsourcing where workers are paid for completing tasks [2].

3 The VMV Projection Kit and USSD App

The setting of this study was 8 villages spread across three rural districts in Rwanda. The low electricity penetration and relatively high ownership of mobile phones (mostly feature phones) informed the requirements for the design of our system. Our proposed system combined a relatively inexpensive video projection kit with a USSD application for data collection (Fig. 3).

3.1 Movie Projection Kit

A key specification of the movie projection kit was its suitability for use in rural areas. To achieve this, we aimed to build a system that would (1) rely on solar energy for power; (2) be robust, small in size, and transportable; (3) be easy to assemble, without requiring significant training; and (4) be relatively inexpensive, with a total cost of \$1,000 or less for all required hardware. We used an iterative design method, building prototypes, testing them and thereafter making changes.

The first prototype of the projection kit used solar energy and batteries that could support 2 to 3 h of operation (Fig. 1a). The entire system, including the solar panel and battery, one speaker with built-in amplifier, and a large custom-made screen met our initial target cost of \$1,000, could fit in a standard travel bag and weighed approximately 35 Kg (Fig. 1b). However, this prototype was complex to build since we had to connect, and in some cases solder, all the components together. Hence, a significant amount of specialized labor was required to assemble the system, resulting in the prototype being neither robust

nor easy to set up. Moreover, the power of the single speaker was not enough for outside projections with a large audience.



(a) First kit prototype



(b) First prototype being transported

Fig. 1. First prototype of the projection kit

To build a more robust and easy to assemble system, we decided to use off-the-shelf components, particularly in terms of the solar energy equipment. Thus, our second prototype included a “pico” portable projector, two 800 W speakers, a complete solar power unit with solar panels and a battery, a microphone, and a custom made portable screen (7 m by 3 m) for a total cost of about \$2,000. The battery could provide up to 3 h of continuous projection and could be charged in a day by the solar panels. Though this prototype went beyond our cost requirement, it was nonetheless more robust and required less effort to assemble.

3.2 USSD Application

We built a USSD application for data collection, choosing to use USSD over SMS for two main reasons. Firstly, USSD allowed us to keep the overall cost lower for our viewers as, unlike SMS, USSD sessions are free for users. Secondly, it allowed us to reduce incorrect and invalid data by using the interactive forms that USSD provides. This application also meant that viewers were not required to remember the short codes to dial as these were projected on the screen with the adverts. Additionally, prepaid customers of telecommunication providers were able to load airtime and check their available balance through USSD. Crucially, since more than 98% of the Rwandan mobile service subscribers are prepaid

customers [8], the vast majority of mobile phone users are familiar with using USSD applications.

Finally, our USSD application allowed the person in charge of a projection to register viewers by collecting the viewers' location (village name), age, gender, phone number and the household size. These data points allowed us to obtain basic demographics regarding our audience. We could then share this information with advertisers to give them a clearer understanding about their customers.

4 Movie Projection and USSD App Deployment

In March 2015, we projected movies in 8 villages in Rwanda, spread across three districts. We had initially planned to target 9 villages, however we had to cancel a show in one village due to heavy rains and a lack of access to an indoor facility. Similarly, we faced various other issues that impeded our data collection in two more villages. Therefore, we present the data obtained from the projections in six villages.

We selected these villages based on the presence of a market that attracts people from nearby villages. Indeed, we set our projection dates to coincide with market days so as to attract the maximum number of viewers. The movie projections were completed using our projection kit. We also built our own USSD application for data collection.

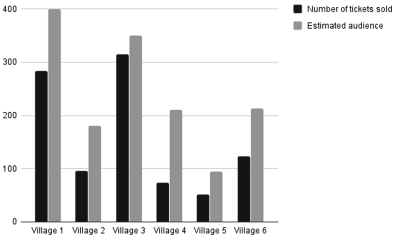


Fig. 2. Number of tickets sold vs estimated audience

We projected the movies in the evenings as our projector performed poorly in the daylight. During the day, our operatives visited the markets informing people of the upcoming movie projection. To pay the wages of the operatives, we charged RWF100 per ticket (about \$0.12). However, since the projections were made in open areas, many more people watched the movies than had initially bought the tickets (see Fig. 2).

Prior to starting the movie projection, we asked the viewers to use our app to signal their presence in the audience. We projected the code they could dial (e.g., *868*1#) to achieve this. This enabled us to know the approximate number of viewers who had access to a mobile phone, allowing us to compute the effectiveness of our data collection method. During the movie projection, we had a number of advert breaks, similar to a number of local TV stations. However, we limited the number of advert breaks to three in order to minimize interruptions. At the end of each advert, we asked questions which were projected on the screen regarding the viewers' feelings towards the advert and/or product advertised. For example, we projected messages such as: *dial *868*1*1# if you liked this advert or if you would buy this product. If you do not like this advert or product, dial *868*1*2#*. The first part of the code (868) represented our USSD

short code. The second part of the code identified the number of the question being asked. For example, 1 was used to indicate the first question, whereas 2 indicated the second question. The third and last part of the code identified the answer of the viewer. For example, 1 was used by the viewer to indicate approval of the advert or product. Those answers were sent via the cellular network to our server, where they could be monitored and analyzed in real-time using a web application (see Fig. 3). To encourage people to respond with their mobile phones, we informed them that there would be a random drawing of phone numbers that had responded to questions. The owners of these phones could then win various prizes including vouchers for phone airtime, various products from advertisers, local food and drinks.

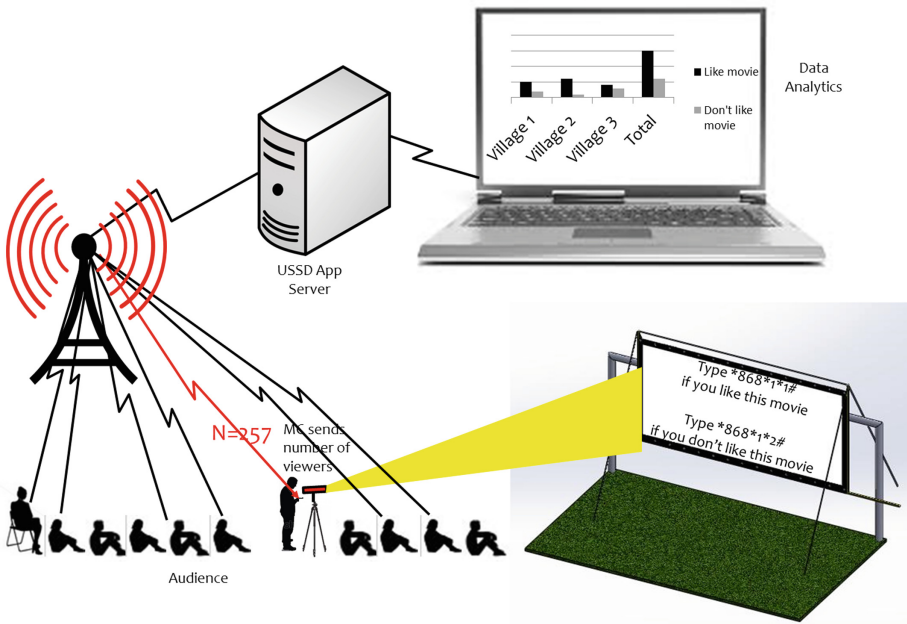


Fig. 3. System design, with video projection and USSD app

Figure 4 shows the average number of responses by projection to the question relating to audience presence and the three advert questions. This figure shows that the number of people answering questions gradually reduces over time. We posit that this may be due to a fatigue in answering questions. This reduction in the number of answers could also indicate a decrease in the number of people present, as we noticed that a number of women would leave the site early to go attend to house chores. However, we believe further research would need to be conducted in order to determine the ideal number of adverts. On one hand, possible viewer fatigue may reduce the optimum advertisement number,

on the other hand, it is likely that advertisers would like to ask viewers as many questions as possible.

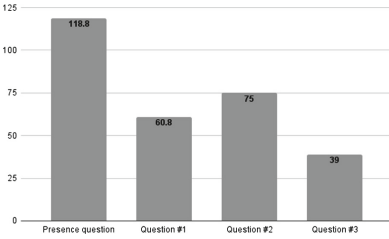


Fig. 4. Average number of answers per questions

in general, the audience was more receptive of the second advert (Pub 2) than of the first advert (Pub 1). Women were more receptive than men; people aged 1–20 and 31–40 were less receptive; and people from smaller households (less than 9 people) were more receptive. More in-depth analysis could be obtained by combining our collected data with public economic and social data from the census for those particular villages, emphasizing the effectiveness and efficiency of our system.

Our system also enabled us to conduct simple experiments with adverts to determine which was best received by the audience. For example, one company had created two adverts for the same product and wanted to test their popularity with a real audience. We projected both adverts and asked the audience to text the number of their preferred advert. Figure 5 shows the result of this test.

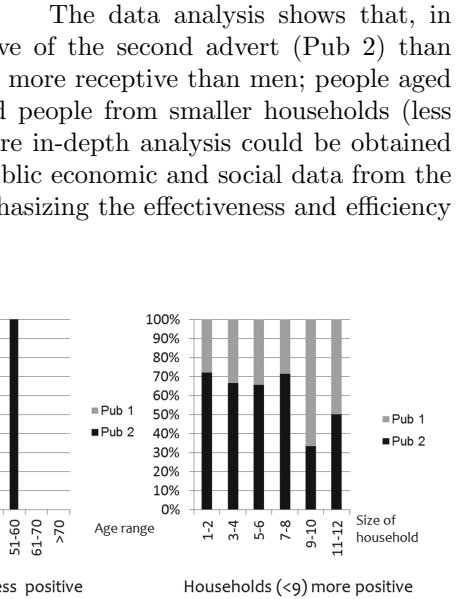


Fig. 5. Experimenting with two adverts

5 Discussion

In this paper, we have demonstrated how we can offer entertainment in remote parts of a developing country and collect data by using simple, low-cost technologies. We described how we built a robust and relatively inexpensive movie projection kit using off-the-shelf equipment. More importantly, we have highlighted how the use of solar energy to power the projection kit increased usability in remote Rwandan locations that do not have access to electricity. Additionally,

we have described how we created a custom-built application for data collection. The application used the Unstructured Supplementary Service Data (USSD) technology, making it accessible to virtually every mobile phone.

By using simple code displayed on the projection screen, our USSD application enabled the viewers to provide feedback to advertisers in an easy and speedy manner at a minimal cost. Therefore, the design of the USSD app enabled virtually any viewer with a mobile phone to provide quick feedback to advertisers allowing the latter to obtain an almost instantaneous picture of their target market's feelings towards their products.

The simplicity of the USSD app design makes it portable to other media as well. For example, advertisers could display similar short code on their TV ads, billboards and posters to receive real-time market insight regarding their products.

Additionally, the projection kit together with the USSD application provide a powerful method for reaching people who are otherwise outside the domain of mainstream audiovisual media. For example, it may be possible for governmental or non-governmental organizations to use the projection kit to educate people on relevant local issues, such as best practices in the health, or new agricultural techniques. After the training through video projection, the organizing entity could then ask questions to the audience and collect answers through the USSD app. In this way, the organization can get real-time feedback on the effectiveness of their training material.

6 Challenges and Conclusion

During the implementation of this project, we met with a number of challenges that we discuss below:

– Obtaining locally-made films:

In order to make our viewers relate more with the movies projected and watch them longer, we sought locally-made movies. Indeed, our project showed the average number of responses to advert questions sent through the USSD application remained constant when locally-made movies were projected, but reduced when other foreign movies were projected. For this study, we secured two movies: a locally-made and a foreign-made movie that was translated into the local language. When the local movie was projected, the average number of responses to the last advert question was 61.6% of the responses to the presence question. This number dropped to only 35.8% when we projected the foreign movie, indicating that more people stayed longer to watch the locally-made movie and they remained engaged throughout the projection.

However, we faced significant challenges when trying to procure locally-made movies as the Rwanda's movie industry is still in its infancy and the cost of films still prohibitive. With a large scale deployment of the movie kits across many villages, we could help foster the movie industry by providing a bigger audience. This would help bring down the movie cost.

- **Working with mobile service providers:**

To implement the USSD application, we had to work with mobile service providers to leverage their network infrastructure. Partnering with these companies proved difficult as this study did not have the ability to generate significant revenue for them. Despite the fact that only one of the three telecommunication service providers agreed to work with us, it still took approximately five months to connect our application to their network. Fortunately, this provider was the dominant market player with more than 50% of mobile subscribers at the time. The other two companies did not officially refuse us the connection, but had pricing terms that were out of our budget.

- **Finding suitable rooms for projection:**

Though we built our projection kit for the outdoor environment, sometimes strong weather conditions, such as rains and powerful winds required that we conduct our movie projection indoors. However, this proved difficult in a number of villages where there was not a room large enough to host all the viewers. Thus, we were at times forced to cancel some projections.

- **Local authorities were sometimes not supportive:**

We learned from experience how important it was to contact local authorities in advance in order to explain the benefit of the project for their communities and obtain their buy-in.

Despite these challenges, this project was successful and demonstrated that a low-cost video projection kit, powered by solar energy, can be an effective media distribution channel. If distributed across thousands of villages and combined with an easy to use and simple mobile data collection application, this solution could revolutionize data collection and analysis in rural areas for advertisers in Africa.

Compared with traditional surveys that are costly, labor intensive, and time consuming, our solution offers an effective and more efficient approach to data collection with the following two main advantages:

- During in-person surveys, interviewees are sometimes influenced by the interviewer. By using our USSD application, this potential bias is removed as the user directly provides answers through their mobile phone, without revealing them to an interviewer.
- The answers are collected in real-time and can be analyzed automatically for quick information provision to the advertisers.

Our USSD application is more effective at collecting answers for closed-ended questions that give respondents a fixed set of options to choose from. Closed-ended response choices can be simple yes/no options, multiple choice options and rating scales options.

Finally, as the projection kit was to be used only in the evening on market days, the concept was to have the projection kit managed by the village school allowing them to use it to project educational videos during class time, thereby enabling the use of multimedia educational content in rural classrooms.

References

1. World bank open data (2017). <https://data.worldbank.org/>
2. Eagle, N.: txt eagle: mobile crowdsourcing. In: Aykin, N. (ed.) IDGD 2009. LNCS, vol. 5623, pp. 447–456. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02767-3_50
3. Grunfeld, H., Hak, S., Pin, T.: Understanding benefits realisation of ireach from a capability approach perspective. *Ethics Inf. Technol.* **13**(2), 151–172 (2011)
4. Hughes, N., Lonie, S.: M-PESA: mobile money for the “unbanked” turning cell-phones into 24-hour tellers in Kenya. *Innovations* **2**(1–2), 63–81 (2007)
5. Merrill, J., Hershow, R., Gannett, K., Barkley, C.: Pretesting an mHealth intervention for at-risk adolescent girls in Soweto, South Africa: studying the additive effects of SMSs on improving sexual reproductive health & rights outcomes. In: *Proceedings of the Sixth International Conference on Information and Communications Technologies and Development: Notes-Volume 2*, pp 96–99. ACM (2013)
6. NISR: Demographic and Health Survey (2014/15). National Institute of Statistics of Rwanda (NISR) (2016). <http://www.statistics.gov.rw/data-portals/>
7. Perrier, T., DeRenzi, B., Anderson, R.: USSD: the third universal app. In: *Proceedings of the 2015 Annual Symposium on Computing for Development*, pp 13–21. ACM (2015)
8. RURA: Active Mobile Telephone Subscriptions as of 31st October 2017. Rwanda Utilities Regulatory Authority (RURA) (2017). <http://www.rura.rw/index.php?id=60>
9. Suddul, G., Soobul, A., Bahadoor, U., Ramdoyal, A., Doolhur, N., Richomme, M.: An open USSD enabler to simplify access to mobile services in emerging countries. In: *2011 4th International Conference on Emerging Trends in Engineering and Technology (ICETET)*, pp 323–326. IEEE (2011)
10. di Tada, N.: It without software (2010). <http://ndt.instedd.org/2010/05/it-without-software.html>



A Study of the Wind Potential in Climatic Zones of Chad

Zoutene Pabame¹, Cheikh Mouhamed F. Kebe^{1(✉)},
Boudy Ould Bilal², Ababacar Ndiaye^{1,4}, Assane Gueye³,
and Pape Alioune Ndiaye¹

¹ Centre Internationale de Formation et de Recherche en Énergie Solaire
(C.I.F.R.E.S), Dakar Fann, Senegal
cmkebe@gmail.com

² Ecole Des Mines de Mauritanie (EMIM), Nouakchott, Mauritania

³ Université Alioune Diop de Bambey, Bambey, Senegal
assanel.gueye@uadb.edu.sn

⁴ Université Assane Seck de Ziguinchor, Ziguinchor, Senegal

Abstract. This paper focuses on the assessment of wind energy potential in the three climatic zones of Chad: the Saharan (north), the Sahel (center) and the Sudan (south) zones. For each zone, three representative meteorological locations were chosen and assessed based on satellite data provided by NASA. The data comes from MERRA (Modern-Era Retrospective Analysis for Research and Applications) and covers the period 2005–2014. The wind speed frequency distribution of locations was found by using Weibull distribution function. From this statistical data analysis, we found that the wind regime is different in the three regions. It is higher in the Saharan region (with annual mean wind speed of 5.78 m/s) followed by the Sahel (4.32 m/s) and Sudanian (3.7 m/s) regions. There are two distinct seasons in Chad: the dry and the rainy seasons with varying periods, with respect to the regions (2 months of rain in the Saharan zone vs 7 months in the Sudanian zone). For all regions the mean wind speed is higher in the dry season. Diurnal variations of mean wind speed show two regimes characterized respectively by high values in the early morning and the night and low values during the day. The corresponding power density was 193 w/m², 76.15 w/m² and 29.0 w/m², resp. for the Saharan, Sahelian and Sudanian regions. The wind regimes are globally stable with dominant directions North-East (for Saharan region), East (Sahelian), South-Southwest (Sudanian).

Keywords: Wind energy · Wind potential · Chad climatic zones
Weibull · Power density · Turbulence index

1 Introduction

Chad's ability to achieve increased energy access and poverty reduction is constrained by significant challenges in the power sector. It currently only has about 125 MW of installed generation capacity to serve a population of 14.5 million people. As a result, Chad's

government is working to expand its electricity supply and encourage investment in the energy sector to stimulate the economy. Chad is endowed with the tenth-largest oil reserves in Africa, as well as wind and solar resource potential. The majority of its existing capacity comes from diesel and HFO generation [1]. Usage of fossil energy resources can create new problems environment and natural resources (NR). In addition, high oil prices and the growing depletion of existing sources, will require new and renewable energy sources to replace fossil energy. One of new and renewable energy resources is wind energy. Wind is a solar energy formation which occurs when the sun heats the air and causing air to rise and form a vacuum, then vacuum down to the cooler air form the wind. Wind occurs because of uneven heating by the sun. Experts estimate that the energy of sunlight received by the earth may be converted into wind kinetic energy of nearly 2% [2]. So, wind energy appears as a clean and good solution to cope with 11 a great part of this energy demand [3]. Meanwhile exploitation of Wind energy need precious evaluation of potential in the desired location. This means to have good database for minimal one entire year with a good time resolution. Recently, many researchers [4–10] have studied the wind energy resources in the sites all over the world. Government of Chad strategy involved the use of Renewable energy for rural electrification and income activities. So, the contribution of this paper is to evaluate the wind power potential in the three climatic zones of Chad: Sudanian, Sahelian and Saharan zones to undertake it, the paper is arranged in three sections; in Sect. 2, we present the material (data) and mathematical models of the underlying theory and statistically analyze the data set. Section 3 is dedicated to discussing the results of our findings and we end with a conclusion in Sect. 4.

2 Material and Method

2.1 Description of the Sites and Used Data

Chad is located between the 7th and 24th degrees North and the 13th and 24th degrees East. It is divided into three (3) major bioclimatic zones, namely the Saharan, Sahelian and Sudanian zones.

The Saharan zone covers an area of $600'350 \text{ km}^2$ or 48% of the land area. Its climate ranges from isohyets 0–200 mm and is characterized by low annual rainfall (less than 200 mm).

The Sahelian zone covers an area of $490'570 \text{ km}^2$. It extends between isohyets 200 to 600 mm.

The Sudanian region, is the wettest part of the country and is characterized by a rainfall of 600 to 1200 mm.

For our study we selected nine (9) cities; three (3) in each climatic zone. The geographical positions of these cities are shown in Table 1 below.

Due to the lack of data from meteorological stations during period and with good time resolution, we use the MERRA (Modern-Era Retrospective Analysis for Research and Applications) satellite data from NASA. Data have tested and validate using field

Table 1. Geographical position of the selected sites

Zones	Cities	Latitude	Longitude	Altitude (m)
Saharan	Bardaï	21.3544	17.001	1088
	Fada	17.1667	21.55	588
	Faya	18	19.1667	235
Sahelian	Abéché	13.85	20.85	628
	Bol	13.4667	14.7167	281
	Ndjamena	12.1333	15.0335	296
Soudanian	Moundou	8.5667	16.0853	423
	Pala	9.3667	14.9667	460
	Sarh	9.1333	18.3833	363

truth. Data have been sampled hourly that includes wind speed, relative humidity, pressure, and temperature and wind directions. It covers the period 2005 to 2014 that is ten year, with 87648 records for each site. Data are available at 10 m of altitude above Sea level.

The duration of the rainy season in Chad decreases from South to North. There are seven (7) months of rainfall in the Sudanian zone, around 4–6 months in the Sahelian zone and 0–2 months in the Saharan area.

2.2 Approach and Theoretical Models

The estimation of the wind potential is based essentially on the determination of the mean wind speed, the power density, the turbulence index, the shear coefficient, allowing the extrapolation of the wind speed at altitude, as well as the determination of the wind rose to indicate the dominant wind direction.

As part of this paper, the average of these parameters was determined for each year, each month and for each season. This work was done for each of the three regions of Chad. The analysis of the influence of the hourly variation of these parameters on the quality of the potential was also carried out. The model to calculate the last parameters is given the following seven equations.

2.2.1 Mean Wind Speed

In the present study, the wind speeds data measured every hour for ten years in each site were used to calculate the wind potential. The annual, monthly and seasonal mean wind speed values were calculated by using Eq. (1), [11].

$$v_m = \frac{1}{n} \cdot \sum_{i=1}^n v_i \quad (1)$$

Where n is the observation number and v_i is the wind speed in time stage i .

2.2.2 Weibull Distribution

Weibull distribution has been commonly used in literature to express the wind speed distribution and to estimate the wind power density. The Weibull distribution is a good match with the experimental data. The probability density function of Weibull is given by Eq. (2) [16]. In the case of this study the Weibull distribution was determined and for every year, month and season. The Weill parameter (C and K) was calculated.

$$f(v) = \frac{K}{C} \cdot \left(\frac{v}{C}\right)^{K-1} \exp\left(-\left(\frac{v}{C}\right)^K\right) \quad (2)$$

Where C and K are respectively the scale and the shape parameters of Weibull, which can be determined by using the Maximum likelihood method Eqs. (3) and (4) [17].

$$K = \left(\left(\frac{\sum_{i=1}^n v_i^K \ln(v_i)}{\sum_i v_i^K} \right) - \left(\frac{\sum_{i=1}^n \ln(v_i)}{n} \right) \right)^{-1} \quad (3)$$

$$C = \left(\frac{\sum_{i=1}^n v_i^K}{n} \right)^{\frac{1}{K}} \quad (4)$$

Smaller k values correspond to more variable (more gusty) winds and the higher A values correspond to a good potential.

2.2.3 Wind Power Density

The long-term wind speed distribution $f(v)$ is combined with the available wind power to give the average wind power density, which can be expressed as follows [18]:

$$\bar{P} = \frac{1}{2} \cdot \rho \cdot C^3 \cdot \Gamma\left(1 + \frac{3}{K}\right) \quad (5)$$

Where $\Gamma(x)$ is the gamma function of (x) given by:

$$\Gamma(x) = \int_a^{\infty} t^{x-1} e^{-t} dt \quad (6)$$

2.2.4 Turbulence Index

The turbulence index is defined as the ratio of the standard deviation to the wind-speed (Eq. 1) [38]. It provides information on the turbulence of the site and can be used to explain the effects of the tower on the quality of the measured wind-speed data.

$$I_n = \frac{\sigma}{v_n} \quad (7)$$

σ : the standard deviation of the wind-speed (*m/s*)

v_m : mean wind-speed (*m/s*)

3 Results

It was observed that between 01 h am and 05 and 08 pm to -11 Pm, the mean wind speed quite the highest. The values of mean wind speed become the lowest between 10 am and 03 Pm. This implies that all the nine, the wind potential was higher during the nighttime compared to the daytime. Notice that the curves of the three sites of the Sudanian zone are in the bottom, followed by the curves of the Sahelian sites in the middle and the curves of the Saharain sites that are at the top of the figures. It should also be noted that the wind speeds of the Bardaï are similar to the speeds of the sites located in the Sahelian zone. The mean wind speed collect was also used to calculate the level of turbulence site by determine the turbulence index given using Eq. 7 this parameters was determined for the all data, seasonal period and every hour on during 2005 to 2014. Figure 2 illustrates results for the hourly variation of turbulence index. It can be seen that the curves are decreasing from 01 am to 05 am and then between 04 pm and 12 am. It was the higher between 01 pm and 03 pm. While, the site of Bardaï has a pseudo periodic variation within the period of the day and the turbulence index reaches the higher value between 10 am and 11 am. This shows that on the 9 locations, the turbulence index was the lowest during the night time and the highest during the day-time.

A wind rose diagram is a representation of the frequencies of directions from which the wind blowing, on a given site and for a given observation period. The wind direction collected on the sites (every hour, for the period of 10 years) was used to determine the wind rose on the located site. Results for the all sites are illustrated by Fig. 1. It can be noted that the sites of Bol, Ndjamen, and Faya-Largeau have North-East dominant direction. The dominant direction was Nort-Northwestern for Abéché, Fada and Bardaï, Est for Moundou, It is Sourth-sourthwertern for Pala and South-western for the site of Sarh (Fig. 3).

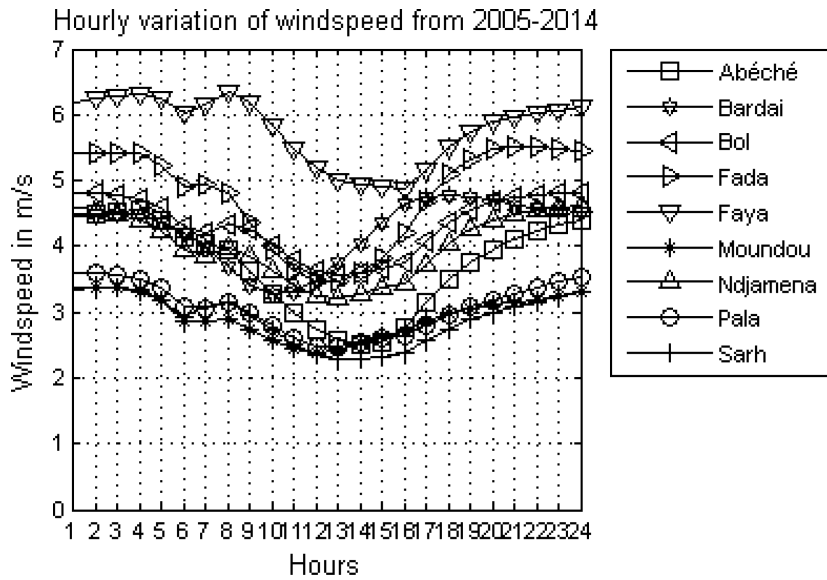


Fig. 1. Hourly variation of wind speed

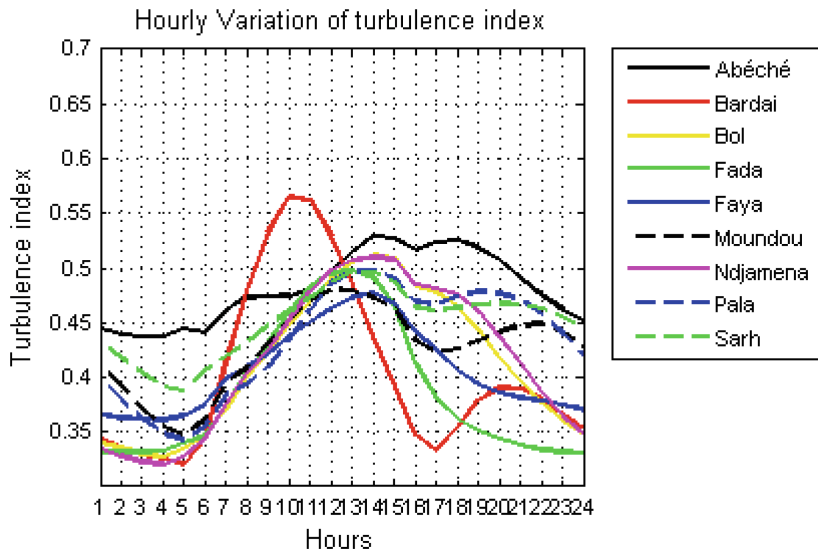


Fig. 2. Variations times of turbulence index.

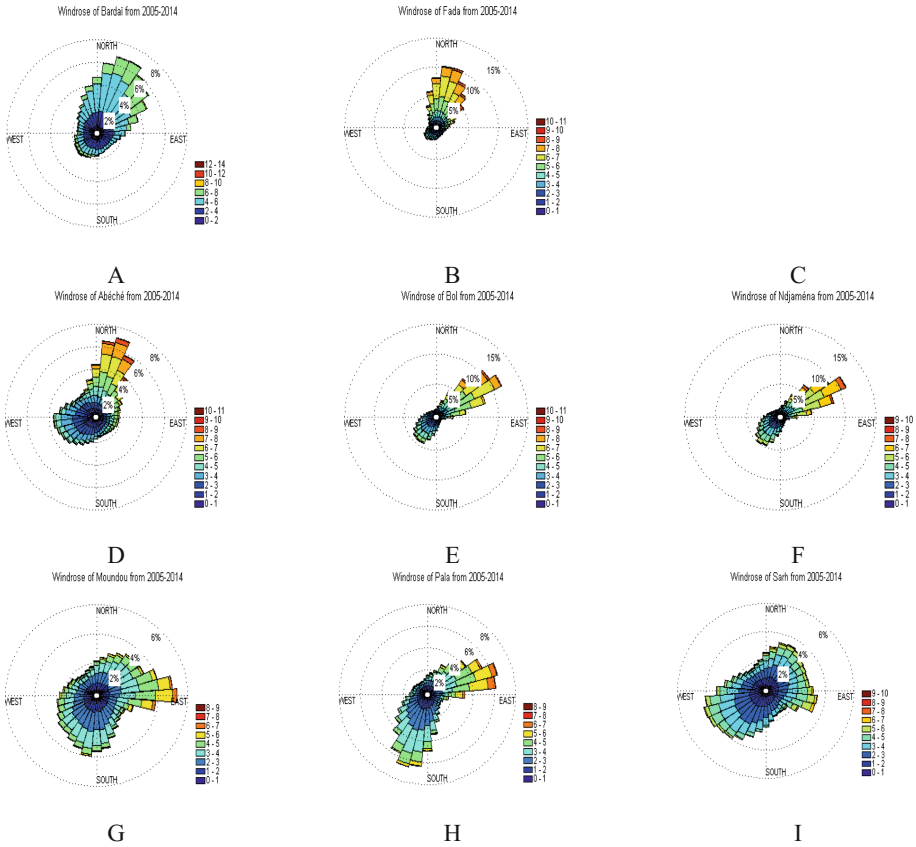


Fig. 3. The wind rose diagrams at an altitude of 10 m over 10 year period; A-Bardai, B-Fada, C-Faya, D-Abéché, E- Bol, F-Ndjaména, G-Moundou, H-Pala and I-Sarh

4 Conclusion

This study has allowed us to have a better understand of the wind potential in Chad. It has shown that wind speeds on the nine selected sites are more important during the dry season than during the rainy season. We have also seen that during the 12 months of the year, there is a favorable month and an unfavorable month for each site. In addition it was found that the wind speeds are higher during the night than during the day. As for the wind roses, the most dominant directions oscillate between north and east during the dry season, and between south and west during the rainy season. However, we have noticed that the sites of Bardai and Faya-Largeau are not influenced by the rise of the monsoon; their dominant directions oscillate throughout the year between north and east. This means they often remain above the ITF (Intertropical Front) and under the domination of the Harmattan. This informs us on why this Saharan region remains suffers from scarcity of rain for the entire year. Still, the site of Bardai deserves special attention. In fact, despite being in the Sahara, its wind speeds are similar to those of the

Sahel zone. As such, this site requires a deep study with data measured on the ground, to understand its peculiarity vis-à-vis to other sites. Apart from this site, wind speeds are increasing going from the Sudan zone to the Sahara region. Finally, it is to be noted that the site of Faya-Largeau is windier than all the other sites, with an average speed of 5.78 m/s at altitude 10 m, a scale factor (C) of 6.5 m/s, and a form factor (K) 2.62.

References

1. USAID: <https://www.usaid.gov/powerafrica/chad>. Accessed 25 Friday
2. Al-Abbadi, N.M.: Wind energy resource assessment for five locations in Saudi Arabia. *Renew. Energy* **30**, 1489–1499 (2005)
3. Gokcek, M., Bayulken, A., Bekdemir, S.: Investigation of wind characteristics and wind energy potential in Kırklareli, Turkey. *Renew. Energy* **32**, 1739–1752 (2007)
4. Celik, A.N.: A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. *Renew. Energy* **29**, 593–604 (2003)
5. Seguro, J.V., Lambert, T.W.: Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *J. Wind Eng. Industr. Aerodyn.* **85**, 75–84 (2000)
6. Zhou, W., Yang, H., Fang, Z.: Wind power potential and characteristics analysis of the Pearl river delta Region. *Renew. Energy* **31**, 739–753 (2006)
7. Irfan, U., Qamar-uz-Zaman, C., Andrew, J.C.: An evaluation of wind energy potential at Kati Bandar, Pakistan. *Renew. Sust. Energy Rev.* **14**, 856–861 (2010)
8. Ahmed, O., Hanane, D., Roberto, S., Abdelaziz, M.: Monthly and seasonal assessment of wind energy characteristics at four monitored locations in Liguria region (Italy). *Renew. Sust. Energy Rev.* **14**, 1959–1968 (2010)
9. Raichle, B.W., Carson, W.R.: Wind resource assessment of the Southern Appalachian Ridges in the Southeastern 70 United States. *Renew. Sust. Energy Rev.* **13**, 1104–1110 (2009)
10. Ali, M.: Feasibility study of harnessing wind energy for turbine installation in province of Yazd in Iran. *Renew. Sust. Energy Rev.* **14**, 93–111 (2010)
11. Serdari, E., Berberi, P., Muda, V., Buzra, U., Mitrushi, D., Halili, D.: Wind profile characteristics and energy potential assessment for electricity generation at the Karaburun Peninsula, Albania. *J. Clean Energy Technol.* **5**(4) (2015)
12. Hacène1, F.B., Merzouk, N.K., Loukarfi, L., Abdelbaki, C.: Contribution A L'etude Des Caracteristiques Eoliennes De La Vallee Du Cheliff
13. Waewsak, J., Chancham, C., Landry, M., Gagnon, Y.: An analysis of wind speed distribution at Thasala, Nakhon Si Thammarat, Thailand. *J. Sust. Energy Environ.* **2**, 51–55 (2011)

Climate Change Workshop



Beyond Participation: Welfare Effects of Gender-Differentiated Group-Based Approaches Under Climate Change in Kenya

Marther W. Ngigi¹(✉) and Ulrike Mueller²

¹ Machakos University, P.O. BOX 136-90100, Machakos County, Kenya
marthernngigi@gmail.com, m_ngigi@mksu.ac.ke

² GFA Consulting Group, Wallstr. 15, 10179 Berlin, Germany
Ulrike.MuellerPM@gfa-group.de

Abstract. A gender-differentiated data collection approach is an essential step toward understanding gendered perspectives in climate change research. Innovative institutions like group-based approaches provide opportunities to improve socio-economic, political or environmental situations with positive outcomes in the midst of climate change. However, little is known on the potential of gender-differentiated group-based approaches in the context of improving men's and women's welfare outcomes under climate risk. The study shows that husbands and wives associate in different groups, hence acquire different gendered benefits. Econometric analysis shows that participation in group-based approaches is influenced by both gender-specific factors such as level of education, perception of climate change and institutional factors, which in turn influence welfare outcomes of participating in social groups. In the wake of climate change, innovative institutions present important pathways to strengthen the ability of men and women to manage risks and improve their welfare. Hence, there is a need for enabling policies that nurture social capital and group-based approaches at the local level.

Keywords: Gender-differentiated group-based approaches
Intra-household analysis · Welfare outcomes · Climate change
Kenya

1 Introduction

Climate change is a global challenge that threatens livelihoods and undermines efforts for overcoming hunger, poverty reduction, gender equality, and environmental sustainability. Sub-Saharan Africa (SSA) is extremely susceptible to adverse impacts from climate change and variability, attributable to low adaptive capacity, low investment in infrastructure, low levels of physical and human capital, high rates of poverty, over-reliance on rain-fed agriculture and lack of a coherent climate policy (IPCC 2014; African Union 2014). However, climate change may also present an economic opportunity for Africa that can be exploited through institutional and technological innovations. In this paper, we argue that in the wake of climate change, social capital created through group-based approaches (GBA) and community-based organizations

present important pathways to strengthen the ability of men and women to manage risks under climate change. Ngigi (2017) shows that social groups often divert from their main mandate to address challenges of climate change and risks through sharing of adaptation information and options and facilitating non-traditional livelihoods that build assets and abilities especially for women.

An institutional innovation is a process of changing norms or generating social change in order to improve a challenging situation like climate change with a positive welfare outcome. Group-based approaches consist of participation in social and political groups that create risk awareness and offer risk managing strategies. Group-based approaches imply a forum for people or communities to participate in decision-making processes in a collective ruling for solutions to difficulties, risks, and shocks facing them (Ngigi 2017). Recent studies suggest that borrowing through group-based approaches is crucial for tackling shocks such as health shocks and market shocks and enhancing food security for the households (Ngigi 2017; Bonfrer and Gustafsson-Wright 2016). Social capital is a valuable post-shock recovery tool that empowers households to rebuild assets (Mawejje and Holden 2014), builds resilience of rural communities or individuals against extreme events (Bernier and Meinzen-Dick 2014; IFAD 2016) and fosters adaptation to climate change (Ngigi et al. 2017). Hence, social capital promotes economic development, poverty reduction and rural transformation even in the face of accelerating climate change.

In spite of vast literature on social capital, there has been little attention to gender-differentiated roles regarding group-based approaches in the context of improving men's and women's welfare outcomes even in the wake of challenges like climate change. There is inadequate quantitative and qualitative information about the actual benefits individuals derive from gender-differentiated groups and about the determinants of such benefits. There is even less information about how group participation or even how social capital is accumulated and how the derived benefits differ across gender. A research gap exists on the potential of gender-differentiated group-based approaches in protecting and accumulating welfare or empowering men and women in the face of fast-track climate change. Men and women are likely to accumulate different forms of social capital that would apparently have different impacts on adaptation to climate change and their well-being. To bridge this knowledge gap, the study examines determinants for formulation of gender-differentiated group-based approaches and how they in turn influence adaptation and welfare outcomes for husbands and wives. The study contributes to the existing literature on the role of social capital in risk management by applying a gender-disaggregated data set that allows for a more nuanced gender analysis in order to shed light on intra-household decision-making on participation in group-based approaches. This paper hence represents a valuable contribution to the debate concerning strategies for coping and adapting with and benefiting economically from climate change.

2 Data and Sampling Procedure

Data for this study was collected from three agro-ecological zones (AEZs) of rural Kenya. These include semi-arid regions (Mbeere South and Nakuru districts), sub-humid regions (Gem and Siaya districts) and humid regions (Mukurweini and Othaya districts). The study included 75 villages. A mix of qualitative and quantitative data collection techniques was used. The survey involved individual- and intra-household level data through interviewing husbands and wives separately. Intra-household interviews were carried out on parallel time, whereby couples were not allowed to consult or communicate with each other. Overall, a random sample of 156 pairs of spouses was interviewed, resulting in 312 respondents in total. Qualitative research involving gender-disaggregated focus group discussion where seven women focus groups and eight men focus groups were conducted, resulting in 15 focus group discussions in total. Narratives from qualitative data were used to supplement quantitative information as well as to interpret and discuss selected results of the quantitative analysis.

3 Descriptive Results of Intra-household Analysis

3.1 Intersection of Gender and Group-Based Approaches

Table 1 presents summary statistics of husbands and wives differentiated by group membership. The cross-tabulation analysis of gender and membership in social groups shows that husbands and wives who belong to different kinds of social groups have more access to extension services, farmer field schools, early warning information, credit facilities and bargaining power than non-group members.

Table 1. Relations between gender, group membership and key variables (mean)

Key variables	Wives			Husbands			Households		
	Non-group members	Group members	Diff. in mean (t-test)	Non-group member	Group members	Diff. in mean (t-test)	Non-group members	Group members	Diff. in mean (t-test)
Adaptation crop [†]	0.71	0.83	-0.12*	0.53	0.76	-0.23**	0.59	0.80	-0.21**
Adaptation livestock [†]	0.57	0.51	0.06	0.43	0.56	-0.13	0.48	0.54	-0.06
Adaptation decision [†]	0.71	0.86	-0.15*	0.63	0.79	-0.16*	0.66	0.83	-0.17**
Intensity of adaptation	2.57	2.42	0.15*	1.43	2.48	-1.04**	1.80	2.45	-0.65*
Year of schooling	4.14	6.39	-2.25*	6.57	8.30	-1.74*	5.80	7.29	-1.49*
Farming experience	41.29	30.91	10.38**	30.40	32.09	-1.69	33.86	31.46	2.40
Entrepreneurial experience	0.43	3.16	-2.73*	3.97	2.38	1.59	2.84	2.79	0.05
Credit access [†]	0.21	0.49	-0.27*	0.30	0.56	-0.26**	0.27	0.52	-0.25**

(continued)

Table 1. (continued)

Key variables	Wives			Husbands			Households		
	Non-group members	Group members	Diff. in mean (t-test)	Non-group member	Group members	Diff. in mean (t-test)	Non-group members	Group members	Diff. in mean (t-test)
Information sources	1.36	1.90	−0.54*	1.60	1.98	−0.38*	1.52	1.94	−0.42*
Information trust index	0.76	0.70	0.06	0.60	0.66	−0.06*	0.65	0.68	−0.03
Extension services [†]	0.14	0.41	−0.27*	0.33	0.57	0.24*	0.27	0.49	0.21**
FFS [†]	0.29	0.44	−0.15*	0.23	0.21	0.03	0.25	0.33	−0.08
Early warning [†]	0.07	0.28	−0.21*	0.23	0.42	−0.19*	0.18	0.35	−0.17*
Weather forecast [†]	0.71	0.63	0.09	0.60	0.41	0.19*	0.64	0.53	0.11
TLU	3.01	4.61	−1.59	5.91	4.45	1.46*	4.99	4.53	0.45
Consumer durable assets	0.22	0.29	−0.08*	0.28	0.32	−0.04	0.26	0.30	−0.05*
Agricultural durable assets	0.47	0.51	−0.04	0.58	0.52	0.06	0.54	0.52	0.03
Bargaining power [†]	0.29	0.35	−0.06	0.10	0.26	−0.16*	0.16	0.31	−0.15*
N	14	142		30	126		44	268	

Note: Superscripts [†] present variables in binary format. ***p < 0.01, **p < 0.05, *p < 0.1

Wives belonging to social groups have more access to consumer durable assets and higher entrepreneurial experience than non-group members. Results further show that social groups improve wives' access to consumer durable asset through collective purchase of household appliances, cooking stoves and pots that augment their asset portfolios. Husbands and wives belonging to social groups are more likely to adjust crop production system as well as to make a decision to adopt several strategies to climate change influenced by their roles and responsibilities, social norms, risk perceptions and access to resources as elaborated in Ngigi et al. (2017). Intensity of adoption of climate-smart technologies was considered as the number of adopted practices/strategies aggregated at the household level, where groups influence intensity of adoption of these strategies.

3.2 Formulation and Accumulation of Group-Based Approaches by Husbands and Wives

Table 2 shows how husbands and wives form and accrue their social capital by being involved in various group-based activities. The study applied Principal Component Analysis (PCA) to compute a group-based approaches index that consisted of variables on trust, reciprocity, group participation and social support. Husbands have a higher social capital index (0.71) as compared to the wives (0.68), a difference that is statistically significant at 10% (t-test P value < 0.10). However, a higher percentage of wives (91%) belong to social groups than husbands (81%). This could be explained by

the fact that husbands are more willing to participate in community activities, mostly belong to community-based organizations (CBOs) hence having wider networks and political capital than wives. In contrary, wives are more active in women's groups and micro finance groups.

Table 2. Formulation of group-based approaches for husbands and wives

Proxy of group-based approaches	Wives (% Yes)	Husbands (% Yes)	Difference in % point	Significance χ^2 (P-value)	Agreement (%)	Kappa	Significance Kappa (P-value)
Group-based approaches index (mean)	0.68	0.71	-0.03	[†] 0.060*			
Belong to any social group	91.17	80.81	10.36	0.018**			
At least one group is a mixed-gender group	48.08	75.64	-27.56	0.000***			
Duration of group membership in years	10.12	11.91	-1.79	[†] 0.285			
Number of groups belonging to (mean)	1.26	1.15	0.11	[†] 0.087*			
Willing to participate in disaster management	91.67	98.08	-6.41	0.010**	91.03	0.10	0.056*
Willing to contribute labor	89.10	97.43	-8.33	0.003***	89.10	0.16	0.005*
Willing to contribute funds for community work	78.85	93.59	-14.74	0.000***	75.00	-0.01	0.536
Involvement in group activities	90.38	83.33	7.05	0.065*	80.13	0.14	0.034*
Work with others in community work	35.90	67.31	-31.41	0.000***	49.36	0.08	0.119
Witnessed sanction	64.10	66.03	-1.93	0.722	62.18	0.17	0.017*
Support from neighbours	36.54	35.90	0.64	0.906	53.21	-0.01	0.563
Support from friends	29.49	17.31	12.18	0.011**	59.62	-0.10	0.915
Trust neighbours with your kids	74.36	78.21	-3.85	0.525	64.10	0.01	0.450
Most people in the community are trustworthy	46.15	50.00	-3.85	0.497	56.41	0.13	0.054*
N	156		156				

Notes: Superscript * presents significance at the 10% level, ** at the 5% level, *** at the 1% level. [†] indicates t-test estimates of population-level mean comparisons.

Husbands and wives affirm that they are willing to participate in disaster management activities, contribute both time and labor in group activities as well as are willing to participate in other group activities. Besides, a higher percentage of wives than husbands are willing to participate in group-based activities and have received

support from members of social groups in the occurrence of extreme events (Pearson $\chi^2 < 0.001$). Further, husbands and wives slightly agree that most people in the community are trustworthy (56% in agreement) (Kappa P-value < 0.05)¹ and they affirm to have witnessed sanctions to those community members who are not willing to participate in group-based approaches and community activities (62% in agreement) (Kappa P-value < 0.05). The findings suggest that, compared to wives, husbands are more willing to cooperate in community activities (67% and 36%, respectively) (Pearson $\chi^2 < 0.05$).

3.3 Gendered Benefits of Group-Based Approaches Under Climate Change

Gender-differentiated group-based approaches benefit husbands and wives in different ways as shown in Table 3. Husbands mainly acquire climate information, adaptation ideas, and access to farm inputs through social groups. For example, farmers belonging to social groups are more likely to change crop variety and types, supported by group-based seed acquisition.

Table 3. Gender-differentiated linkages of group-based approaches to risks under climate change

Benefits acquired through group-based approaches	Wives (% Yes)	Husbands (% Yes)	Difference in % point	Significance χ^2 (P-value)
Access to climate information	22.44	38.46	-16.03	0.002***
Advice on adaptation options	32.05	46.79	-14.74	0.008***
Access to agricultural inputs	32.05	49.36	-17.31	0.002***
Diversify sources of livelihood	73.72	64.74	8.97	0.086*
Manage risks	80.77	68.59	12.18	0.013**
N	156	156		

Notes: Superscript * presents significance at the 10%, ** at the 5%, and ***at the 1% level.

On the other hand, women's groups often assist women to diversify their sources of livelihood and managing climate (as well as non-climate) risks. These benefits are often not the main mandate of groups but diversified to address various challenges arising from climate change. Women's groups often help them diversify livelihoods through innovative pathways such as group-based savings, micro-credit and income generating activities. Groups also help women build their welfare and assets and manage sudden events such as illness. Women collectively hire land, use it for production thus increasing income and food security for the household. They also pull resources and buy food in bulk and sub-divide among themselves hence reducing cost of food and

¹ The Kappa statistics are often used to examine the significance in inter-rater agreement of two or more groups.

promote food and nutritional security. These findings suggest that climate change presents an opportunity for growth and strengthening of innovative institutions such as group-based approaches that promote non-tradition livelihoods and foster ability of men and women to manage risks.

4 Econometric Results

4.1 Empirical Strategy

The study aims to examine contributing factors to gender-differentiated group-based approaches index for the households and individually for husbands and wives. This study goes beyond membership and considers several factors that contribute to group-based approaches besides group membership. Hence, the study applied ordinary least squares (OLS) to assess factors that influence formation of group-based approaches as follows

$$GBA_i = X_i\beta + \varepsilon_i \quad (1)$$

Where GBA_i presents GBA index for household i , husband i , and for wife i , and X_i is a vector of explanatory variables, including household and individual characteristics, socioeconomics, climate variables and institutional factors, where beta represents the coefficients to be estimated. The regression ε_i refers to unobservable errors.

The study goes beyond participation to examine welfare effects of group-based approaches towards assets accumulation for households and individually for husbands and their spouses. GBA index is likely to be endogenous because participation in social groups faces self-selection or is not likely to be a random participation. (As such the study adopts a control function approach where first OLS residue of GBA and inverse Mill's ratio (Heckman 1979; Wooldridge 2010) are used in the OLS model of welfare effects of group-based approaches such as

$$A_i = \beta_1 GBA_i + X_i\beta_2 + \lambda_i + \rho\hat{\varepsilon}_i \quad (2)$$

Where A_i presents asset or asset indices for household i , husband i , and for wife i , GBA_i is the index for group-based approaches, X_i is a vector of explanatory variables, where betas represent the coefficients. The regression λ_i corrects for selection bias in the model and $\hat{\varepsilon}_i$ yields consistent estimates. We applied assets as a measure of well-being because different kinds of assets are built up over time and provide a better proxy for welfare than income or expenditure measures (Johnston and Abreu 2016).

4.2 Factors Affecting Formulation of Group-Based Approaches

The findings show that variables associated with a spouse are likely to influence an individual decision to participate in group-based approaches. Participation of a wife in a social group is likely to influence the husband's participation in group-based approaches and vice versa. Years of schooling of wives positively influence formulation of GBA index, while this is not the case for formulation of GBA index for the

husbands or for the households. This suggests that wives with higher education levels are more knowledgeable on benefits of group-based approaches that boosts their social capital and social engagement. Access to farm visits and farmer field schools (FFS) as a mode of extension service influence the GBA index of husbands. A notable finding is the influence of trust in information acquired from various sources on formation of GBA for wives and for the households, which implies that reliable information can strengthen group-based approaches. Access to credit influences positively the GBA index of husbands and their spouses as well as of that of household levels. Interestingly, perception of climate change - especially worry about the impacts of climate change - influences positively husband's participation in GBA. The findings also show that assets held by the households are likely to influence the GBA index. Household's tropical livestock units (TLU) and consumer durables are likely to influence wives' GBA index, while land size influences the GBA index for wives and that of the households (Table 4).

Table 4. Results of ordinary least squares on determinants of formulation of group-based approaches

	Determinants of individual GBA		Determinants of household GBA	
	<i>Wives</i>	<i>Husbands</i>	<i>Wives</i>	<i>Husbands</i>
GBA of wives		0.161** (0.064)		1.161*** (0.644)
GBA of husbands	0.171* (0.091)		1.171*** (0.084)	
Years of schooling of husbands	–	0.003 (0.003)	–	0.002 (0.005)
Years of schooling of wives	0.041* (0.005)	–	0.004 (0.004)	–
Number of information sources of husbands	–	0.021** (0.011)	–	0.022* (0.022)
Number of information sources of wives	0.014* (0.015)	–	0.032* (0.020)	–
Trust index- information of husbands	–	0.058 (0.057)	–	0.123 (0.011)
Trust index- information of wives	0.074 (0.071)	–	0.055** (0.073)	–
Worried about Climate change of husbands	–	0.047* (0.025)	–	0.054** (0.024)
Worried about Climate change of wives	0.05 (0.032)	–	–0.040 (0.048)	–
Farm visits of husbands	–	0.042** (0.024)	–	0.042** (0.024)
Farm visits of wives	0.025 (0.036)		0.047 (0.051)	

(continued)

Table 4. (continued)

	Determinants of individual GBA		Determinants of household GBA	
	<i>Wives</i>	<i>Husbands</i>	<i>Wives</i>	<i>Husbands</i>
FFS of husbands	–	0.059** (0.027)	–	0.142** (0.025)
FFS of wives	-0.005 (0.037)	–	-0.008 (0.050)	–
Access to credit of husbands	–	0.050** (0.025)	–	0.103** (0.026)
Access to credit of wives	0.047* (0.029)	–	0.032* (0.045)	–
Household's decision on land use	0.039 (0.030)	0.049** (0.029)	0.058* (0.045)	0.042* (0.027)
Household land size (acres)	0.034** (0.001)	0.019 (0.016)	0.050*** (0.00)	0.004* (0.002)
Household's agricultural asset index	0.067* (0.037)	0.019* (0.035)	-0.132 (0.053)	-0.066 (0.060)
Household's consumer durable assets	0.023* (0.037)	-0.425 (0.032)	-0.008 (0.055)	0.049 (0.056)
Household's TLU	0.002 (0.002)	-0.009** (0.003)	0.013* (0.003)	-0.007 (0.006)
Household's rainfall*temperature	0.459* (0.310)	0.144 (0.328)	0.543 (0.485)	0.982** (0.05)
Constant	-3.960* (3.043)	0.267 (3.268)	-3.960 (3.063)	0.267 (3.268)
R-squared	0.257	0.314	0.680	0.761
Prob > F	0.006	0.000	0.000	0.000
Number of observations	156	156	156	156

Notes: Corrected and robust standard errors in parentheses ***p < 0.01, **p < 0.05, *p < 0.1. Selected variables presented.

4.3 Beyond Participation: Welfare Effects of Gender-Differentiated Group-Based Approaches

The findings show that group-based approaches are likely to influence accumulation of assets owned by wives and by the households. Group-based approaches are likely to increase consumer durables of wives, but this is not the case for the husbands. These findings are supported by cross-tabulation analysis, which shows that wives and households belonging to social groups are more likely to own more assets than non-group members. This implies that even under climate change households or individuals can protect their assets through participation in group-based activities especially dealing with risk management and diversification of non-traditional livelihoods. Other factors that affect welfare of households include access to credit, access to numerous sources of agricultural and climate information, livestock and land size. Level of

education of wives, access to credit, land size, collective decision on use of land are likely to influence positively and significantly assets owned by wives. Further, perception of climate change, tropical livestock unit (TLU), access to extension services and land size positively influence assets owned by husbands (Table 5).

Table 5. Two-step estimation of welfare effects of GBA under climate change

	<i>Individual welfare effects</i>		<i>Household welfare effects</i>	
	<i>Wives</i>	<i>Husbands</i>	<i>Wives</i>	<i>Husbands</i>
GBA index of husbands	–	0.142 (0.178)	–	0.192 (0.278)
GBA index of wives	0.037* (0.116)	–	0.023* (0.225)	–
Residue (husbands/wives)	–0.115 (0.213)	–0.105 (0.216)	–0.464 (0.457)	–0.103 (0.342)
Mills ratio (husbands/wives)	0.119 (0.159)	0.268 (0.218)	–0.037 (0.279)	0.279 (0.326)
Constant	0.755 (3.327)	8.194 (3.792)	11.018* (5.495)	7.281 (6.486)
R-squared	0.331	0.322	0.312	0.329
Prof > F	0.000	0.000	0.000	0.000
Number of observations	156	156	156	156

Notes: Corrected and robust standard errors in parentheses ***p < 0.01, **p < 0.05, *p < 0.1. We only present the variables that are of interest for the study.

5 Discussion

While most studies on participation in social groups tend to neglect gender aspects, this study applied a gender-differentiated analysis and examined factors that influence formulation of group-based approaches and how these in turn influence welfare outcomes of husbands and their spouses and of the households. Husbands and their spouses belong to different social groups and accrue different benefits. These findings could be explained on the basis of pre-existing gender and social norms determining women's roles in the household, including cooking and taking care of kids, which is limiting their mobility and discourages them from joining inter-village social groups and CBOs. Similar studies in Sub Saharan Africa (SSA) show that men possess more social and political capital, hence having greater access to supporting interventions than women (see Katungi et al. 2008; Pérez et al. 2014; Ngigi et al. 2017).

The econometric findings show that gender-specific factors influence the formulation of group-based approaches and social capital. Interestingly, participation of one spouse in a social group is likely to influence the other spouse's participation in group-based approaches. Years of schooling of wives boost their social engagement unlike

men. Contrary, Rakib (2015) shows that the education of household heads, who are often men, influences participation in social groups in Bangladesh. Further, assets owned by individuals or households are likely to influence accumulation of social capital and participation in group-based approaches. Group-based approaches are likely to help individuals building and protecting their assets under climate change. Evidence shows that social groups are a crucial pathway particularly for women and asset poor to protect their assets and smoothen consumption in occurrence of climate and non-climate shocks²¹. Men mostly benefit from social groups by acquiring agricultural information, inputs and climate adaptation, while women diversify their non-traditional livelihoods and manage climate and non-climate risks as well as other shocks that they face. Further, social groups help households to enhance food security, rebuild assets and deal with post-shock recovery as well as build resilience against extreme events (IFAD 2016). This study contributes to the literature that gender-differentiated group-based approaches can build and protect assets for wives and of the households. Recent studies support our finding that social and political capital created by group-based approaches contribute to households' assets and consumption expenditure (See Yusuf 2008; Rakib 2015). However, while some literature shows that extreme events such as drought can weaken social capital and networks (Fuente 2007; Ngigi 2017), a certain degree of shocks could strengthen association in social groups (Gebremedhin et al. 2010).

6 Conclusions and Recommendations

This study presents unique intra-household findings of the potential of gender-differentiated group-based approaches in enhancing welfare under climate change. It suggests that gender-differentiated social groups benefit men and women differently based on differential gender needs, opportunities, roles and responsibilities. The findings suggest that climate change presents an opportunity for growth and strengthening of innovative institutions such as group-based approaches that promote non-traditional livelihoods and foster ability of men and women to manage risks. The study found that gender-specific factors such as level of education, access to extension services, access to credit and trust in information influence the formulation and accumulation of group-based approaches, while household factors such as land size and decision-making influence wives' and husbands' social capital. This in turn influences accumulation of assets of husbands and wives and of the households. Gender-differentiated group-based approaches are essential for enhancing welfare outcomes through accumulating essential individual and household assets. They are also essential pathways towards protecting assets in occurrence of shocks through offering alternative risk managing tools and alternative consumption smoothing strategies under climate change. This suggests a need for policies that nurture social capital and group-based approaches for men and women at both household and community levels. Possible pathways towards this goal include capacity building programs and training in basic entrepreneurship and in risk management skills as well as in effective measures for coping and adapting to climate risks for both men and women. Further research should focus on understanding dynamics of gendered innovative institutions through use of panel data and the

opportunities they present in fostering resilience to climate and non-climate risks, while empowering men and women. It could be interesting to examine the economics of group formulation and participation in innovative institutions in terms of costs and benefits that are accrued by either men or women in attempts to foster their livelihood resilience.

References

- IPCC: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Summary for Policymakers. IPCC WGII AR5 Phase 1 44 (2014)
- African-Union: Common Africa Position (CAP) on the Post- 2015 Development Agenda. In: 22nd Ordinary Session of the Assembly of the Union on 31 January 2014 21 (2014)
- Ngigi, M.W.: Managing Risk Under Climate Change in Kenya Multiple Shocks, Poverty, Gender and Potential for Group-Based Approaches. University of Bonn, Bonn (2017)
- Bonfrer, I., Gustafsson-Wright, E.: Health shocks, coping strategies and foregone healthcare among agricultural households in Kenya. *Glob. Public Health* 1–22 (2016)
- Mawejje, J., Holden, S.T.: Social capital, shocks and livestock investments: evidence from Masaka District, Uganda. *Int. J. Dev. Issues* **13**, 98–112 (2014)
- Bernier, Q., Meinzen-Dick, R.: Resilience and social capital. In: 2020 Conference Building Resilience for Food and Nutrition Security, p. 4 (2014)
- IFAD: Rural Development Report 2016. Fostering inclusive rural transformation. International Fund for Agricultural Development (2016)
- Ngigi, M.W., Mueller, U., Birner, R.: Gender differences in climate change adaptation strategies and participation in group-based approaches: an intra-household analysis from rural Kenya. *Ecol. Econ.* **138**, 99–108 (2017)
- Heckman, J.: Sample selection bias as a specification error. *Econometrica* **47**, 153–161 (1979)
- Wooldridge, J.M.: *Econometric Analysis of Cross Section and Panel Data*. MIT Press, London (2010)
- Johnston, D., Abreu, A.: The asset debates: how (not) to use asset indices to measure well-being and the middle class in Africa. *Afr. Aff. (Lond.)* **115**, 399–418 (2016)
- Katungi, E., Edmeades, S., Smale, M.: Gender, social capital and information exchange in rural Uganda. *J. Int. Dev.* **20**, 35–52 (2008)
- Pérez, C., et al.: How resilient are farming households, communities, men and women to a changing climate in Africa? CCAFS Working Paper (2014)
- Rakib, M.: Gender- Differentiated Asset Dynamics in Bangladesh: Individual Adaptation and the Potential for Group Based Approaches in the Context of Climate Change. University of Bonn, Bonn (2015)
- Yusuf, S.A.: Social capital and household welfare in Kwara State Nigeria. *J. Hum. Ecol.* **23**(3), 219–229 (2008)
- Fuente, A.: Climate Shocks and their Impact on Assets. Human Development Report Office Occasional Paper, Human Development Report 2007/2008 (2007)
- Gebremedhin, L., Bezabih, M., Gebreegziabher, Z.: Local Social Networks, Shocks and Rainfall Patterns in the Highlands of Ethiopia. Efd Ethiopian Working Paper 28 (2010)



Climate Change Signals Over Senegal River Basin Using Regional Climate Models of the CORDEX Africa Simulations

Mamadou Lamine Mbaye^{1(✉)}, Samo Diatta¹,
and Amadou Thierno Gaye²

¹ Laboratoire d'Océanographie,
des Sciences de l'Environnement et du Climat (LOSEC),
Université Assane SECK de Ziguinchor, BP 523, Ziguinchor, Sénégal
mlmbaye@univ-zig.sn

² Laboratoire Physique de l'Atmosphère et de L'Océan (LPAO),
ESP/UCAD/Sénégal, Dakar, Sénégal

Abstract. This study provides an overview of the impact of a statistical bias correction based on histogram equalization functions on a set of high resolution climate simulations over the Senegal River Basin. Regarding the future changes of extreme precipitation (greater than 50 mm), the models diverge in predicting heavy rainfall events in the majority of the basin. However, an increase of extreme precipitation is found around the Guinean Highlands. The results show also an increase of dry day's length and a decrease of wet days spells by all the RCMs, except one model that shows an opposite change of these climate indices. The bias correction affects mainly the magnitude of the climate change signals of extreme precipitation. Changes under the Representative Concentration Pathways (RCP8.5) are the most pronounced with uncorrected data. Bias corrected RCMs data are potentially useful for climate change impact studies over the Senegal River Basin. This study highlights a convergence of all RCMs (except for RCA RCM) in projecting a decrease of wet days and an increase of dry days over the Senegal River Basin.

Keywords: Climate change · Bias correction · Signal · Regional climate model Senegal Basin

1 Introduction

Climate change over rural areas has led to considerable impacts on critical sectors such as water resources, agriculture, health, etc. Due to its weak capabilities, Africa is highly affected by these impacts, and this is likely to be exacerbated in the future by extreme events (such as flood, drought) and the high population growth which increases its exposure and vulnerability. Moreover, the changes in extreme precipitation and temperature are predicted by many global climate models as a response to greenhouse gas increase and such changes will have significant environmental and social impacts [1]. The well-known droughts during the three last decades of the 20th century over West Africa were the most severe consequences ever seen in Africa. Nowadays, climate and

hydrological models are the most powerful tools available at our level to address the issues of climate change on water resources. Among these climate models, GCMs output are not suitable for impact studies due to their coarse resolution and systematic biases [2]. Therefore multiple efforts in improving the quality of climate models output made by the international scientific community has come to an international project known as the COordinated Regional climate Downscaling Experiment [3]. The CORDEX purpose is to provide higher regional climate simulations for climate change impact studies and decision making at the regional level [4]. The finer spatial resolution of RCMs when compared to GCMs can be an added value while simulating regional climate features such as precipitation over mountainous areas [5]. However, it is pointed out that RCMs output are associated with systematic biases that can affect hydrological simulations [6]. Previous studies over West Africa, have shown an increase in the frequency of extreme rainfall events from the end of 20th century to early 21st century over the Sahel by using observed stations data [7]. In addition, a probable intensification of droughts events is found by [8] through an increase of dry days and the frequency of their occurrences. [9] by analyzing trends in extreme rainfall indices over Senegal from 1950 to 2007, suggested that modest to significant increases in daily wet days intensity. Projected changes of four RCMs indicated a decline in mean precipitation except for one RCM over one region in Senegal [10]. Furthermore, [11, 12] found that despite the projected decreases in total rainfall, the proportion of total annual rainfall that falls in heavy events tends towards increases in the ensemble projections. Moreover, an assessment of plausible regional trends associated to the return period, from the hazard maps of annual mean of daily rainfall by [13], showed a general rise, owing to an increase in the mean and the variability of extreme precipitation over the Senegal River basin. The future evolution of extreme precipitation dwell unclear at the basin scale where more detail information are needed by water resources managers and users. Hence, to our knowledge there is a lack studies and understanding that account for the removal of RCMs biases and the effect of the statistical bias correction on the climate change signals of extreme precipitation at the basin scale. For this reason, post processing of climate model output known as bias correction is often used for bias removal. However, bias correction cannot improve the misrepresentation of physical processes and transient response to greenhouse gas emissions [14]. In recent impact studies, quantile mapping or histogram equalization methodology is commonly used due to its skills, effectiveness and less parameters to fit, and also because it improves both mean and variance of precipitation fields (as seen in [14–18]). These authors have modified/used the methodology developed by [19] that is also applied in the present study. This study aims to assess the potential changes of climate change signals of extreme precipitation at the basin scale by using bias corrected data and uncorrected data from CORDEX RCMs.

The paper is structured as follows. The data and the methods are briefly given in Sect. 2. Section 3 presents the results of the projections of precipitation characteristics. The discussion of the results is detailed in Sect. 4. Finally, Sect. 5 delivers the conclusion.

2 Data and Methods

2.1 Senegal River Basin

The Senegal River Basin (SRB) is situated in West Africa (Fig. 1). The basin's catchment is about 300,000 km² and it is shared by four countries such as Guinea, Mali, Mauritania and Senegal [20]. The basin is subject to a large south–north rainfall gradient with maximum in the south (up to 1800 mm/year) and minimum in the north (150 mm/year) as like most West African countries. In addition, the northern basin (sahelian part) has the hottest temperatures.

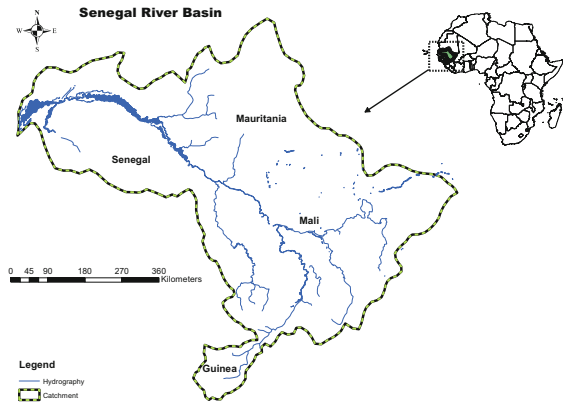


Fig. 1. Senegal River Basin

2.2 Data

Rainfall simulations from five regional climate models (REMO, HIRHAM5, RACMO22T, CCLM4, and RCA4) that are involved in the COordinated Regional climate Downscaling Experiment [3] under the Representative Concentration Pathways (RCPs 4.5 and 8.5) are analyzed. Some characteristics of the models are described briefly in Table 1. Additionally, gridded precipitation data from the EMBRACE forcing data based on ERA-interim known as WFDEI [21] are used for observational datasets. Due to the well-known biases of RCMs output, a statistical bias correction is applied to correct models biases following the method of [19]. The WFDEI data are used to bias correct RCMs simulations during the training period (1979–2005) where transfer functions are derived (details in [22]).

2.3 Bias Correction

The bias correction is based on fitted histogram equalization where the corrected variable is a function of the modeled counterpart ($V_{\text{cor}} = F(V_{\text{mod}})$). The statistical bias correction technique is described in detail in [19], and the method is widely used in several studies [14–16, 18, 22], etc. Therefore, we give only a short summary of the

method as follows. Transfer functions (TFs) of daily rainfall data are generated. To obtain the TFs, observed and corresponding simulated time series of the same length are sorted according to their magnitude, from smallest to largest. A transfer function maps the cumulative probability distribution function of the modeled data onto that of the observed. The bias correction transfer functions are derived for the training period 1979–2005 between RCMs and WFDEI data and applied to 1971–2000 for the historical and to 2071–2100 RCMs simulations for both scenarios.

2.3.1 Precipitation Bias Correction

The following transfer functions (TFs) were used:

$$P_{\text{cor}} = a + bP \quad (1)$$

$$P_{\text{cor}} = (a + bP) \left(1 - e^{-(P-P_0)/\tau} \right) \quad (2)$$

Where P_{cor} represents the bias corrected precipitation, P is a given value to be corrected and, a , b , P_0 and τ are fit parameters. In the linear Eq. (1), the coefficients a , and b are respectively additive and multiplicative correction factors. P_0 is the value of precipitation below which modeled precipitation is set to zero. Equation (2) is composed by an exponential that tends to a linear asymptote ($a + bP$); τ is the rate at which the asymptote is approached and P_0 is the dry day correction term. For high intensities the TF of Eq. (2) tends to the linear term as Eq. (1), and it presents a systematic change of slope at the lowest intensities [18].

Table 1. Some characteristics of the RCMs

	MPI-REMO	DMI-HIRHAM5	CLMcom-CCM4-8-17	KNMI-RACMO22T	SMHI-RCA4
Institution	Climate Service Center	Danish Meteorological Institute	Climate Limited_area Modelling Community (CLM-Community)	Royal Netherlands Meteorological Institute	Swedish Meteorological and Hydrological Institute, Rossby Centre
Short name	REMO	HIRHAM5	CCLM4	RACMO22T	RCA4
Driving model	MPI-ESM-LR	ICHEC-EC-EARTH	MPI-M-MPI-ESM-LR	ICHEC-EC-EARTH	NOAA-GFDL-GFDL-ES2 M
Resolution/projection	0.44° Rotated pole	0.44° Rotated pole	0.44° Rotated pole	0.44° Rotated pole	0.44° Rotated pole
Advection scheme	semi-lagrangien	semi-lagrangien	Fith order upwind [23]	semi-lagrangien	eulerian
Convection scheme	[24]	[24]	[24]	[24]	[25, 26]
Vertical coordinates/levels	Hybrid/27	Hybrid3.1	Terrain following/3.5	Hybrid/40	Hybrid/40
References	www.remo-rcm.de	www.dmi.dk/dmi/tr06-17	www.clim-community.eu	http://www.knmi.nl/research/research-regional-climate/models/racmo.html	http://www.smhi.se/en/research/research-departments/climate-research-rossby-centre2-552/rossby-centre-regional-atmospheric-model-rca4-1.16562

2.4 Simulations Analyses

The following climate indices are used to assess the climate change signals: the maximum number of consecutive wet days (c wd), maximum number of consecutive dry days (c dd). A day is considered as wet/dry when the precipitation is greater/less than 1 mm/day. Additionally, days where rainfall is higher than 50 mm is computed as extreme rainfall events in JAS season. The changes represent the difference between the scenario period (2071–2100) and the reference period (1971–2000).

3 Projections of Precipitation Characteristics

The changes in the number of extreme rainfall days are displayed in Fig. 2. Uncorrected REMO and RACMO22T output show a decreasing tendency of extreme rainfall events with RCP8.5 (–35 days) particularly in the eastern basin for RACMO22T and in the majority of basin for REMO. However, for these same models and scenario, the bias correction seems to alter the climate signal for RACMO22T; as for REMO the bias correction decreases the signal. With RCP4.5, the changes are similar to those found with RCP8.5 for RACMO22T. For REMO, the north-western part of the basin depicts the highest changes with RCP4.5. Furthermore, the southern basin exhibits an increase of extreme rainfall in all data and scenarios for REMO and RACMO22T. For HIRHAM5, it is found a clear increasing tendency of heavy precipitation for both datasets and scenarios (up to 35 days in the southern basin). As for RCA4, corrected and uncorrected data show identical spatial patterns with RCP4.5 where the Guinean Highlands depict a decrease of extreme precipitation by contrast above 12°N, slight increase (5 to 10 days) is found in the majority of the basin. In case of RCP8.5, both RCA4 simulations have similar spatial patterns where in some parts show a slight increase and others show a decrease. In addition, the basin is likely to experience generally an increase of heavy rainfall events with RCP4.5 (raw and corrected RCA4 data). As with RCP8.5, both CCLM4 simulations show an increase below 13°N; however, above this latitude, decreased high rainfall events seem to dominate the changes that are more pronounced with uncorrected data.

The changes of the maximum consecutive dry days are displayed in Fig. 3. Considerable increases of dry days (up to 60 days particularly in the northern basin) are found in REMO, HIRHAM5, CCLM4, and RACMO22T in both scenarios. However, this later model shows slight decrease of dry day length in the south-eastern basin. Moreover, RCA4 model shows considerable decrease of c dd to more than 20 days in the majority of the basin, except around the mountainous areas where a slight increase of c dd is found. For all models, both simulations show similar spatial patterns.

Figure 4 shows the changes in the consecutive wet days. All the models except RCA4, project a general decline of wet days (up to more than 20 days). RCA4 model exhibits an increased pattern of wet days length. RACMO22T model also show slight

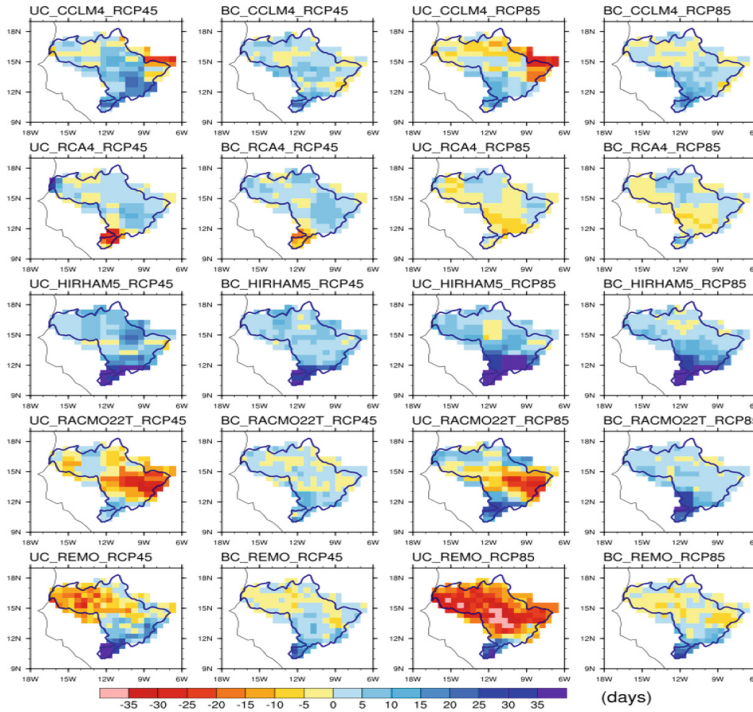


Fig. 2. Changes in the number of daily rainfall greater than 50 mm [UC: uncorrected, BC: bias corrected]

increase in the southern basin. These findings are consistent with above results for the dry days spells. Bias corrected and uncorrected data show similar spatial pattern even though some differences exist in the magnitude of the projected signals.

4 Discussion

The projected changes of extreme precipitation (Fig. 2), RACMO22T shows a decrease particularly in some localized parts of the basin and a possible alteration of the climate signal due to bias correction. This finding shows the limitation of the correction and its impact on the climate change signal as suggested by [15]. As for REMO, the decreased extreme rainfall is also marked up by a reduced bias corrected signal.

However, heavy rainfall is projected in the wettest part of SRB (e.g. Guinean Highlands) that benefit to orographic precipitation; as well as CCLM4 that depicts also similar findings in this part of the basin. Additionally, it is only with HIRHAM5 that the basin is likely to experience a very noticeable increase of extreme rainfall. The differences between models are mainly due their physical parameterizations.

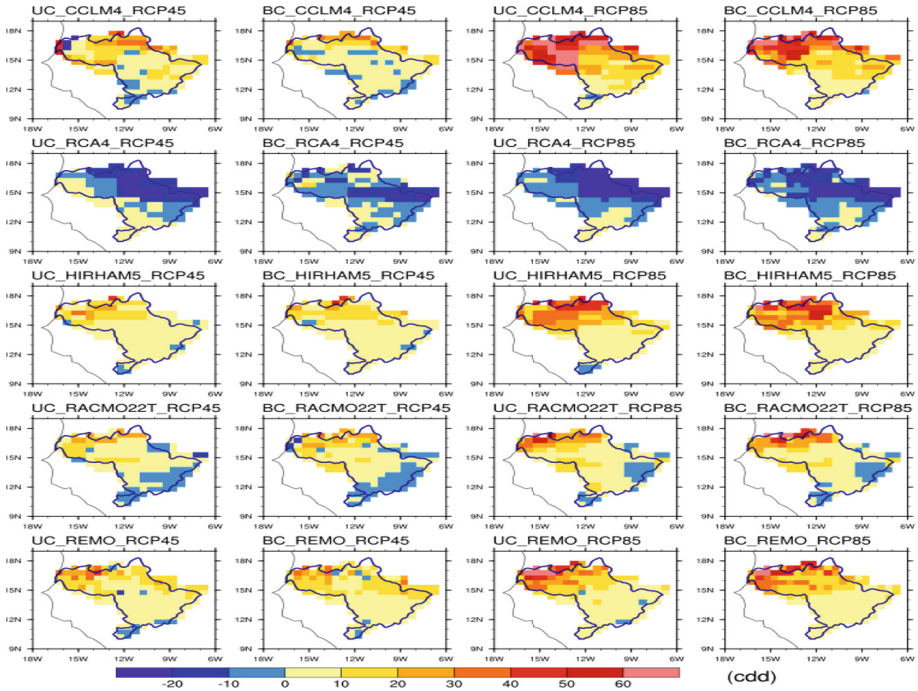


Fig. 3. Changes in the consecutive dry day [UC: uncorrected, BC: bias corrected]

The changes of cdd and cwd as displayed in Figs. 3 and 4, respectively, show considerable increase of dry days and decrease of wet days by all models, except, RCA4 which shows increased wet days. This increase of dry days spells will lead to a drying of the basin that is likely to be more acute in the north. These RCMs divergence is mainly due to the convection scheme used within the models as it was suggested by [4].

Moreover, there is a consistency between these two climate indices in all models; this means that while cdd increases, cwd decreases. The drying of the basin as projected in most of the RCMs can be due to the results of low rainfall intensity and low high rainfall events. This drying is more pronounced with RCP8.5 than with RCP4.5 suggesting the influence of the chosen radiative forcing. In case of the impact of the bias correction, it affects mainly the magnitude of the signals even though some alterations may occur in localized parts of the basin.

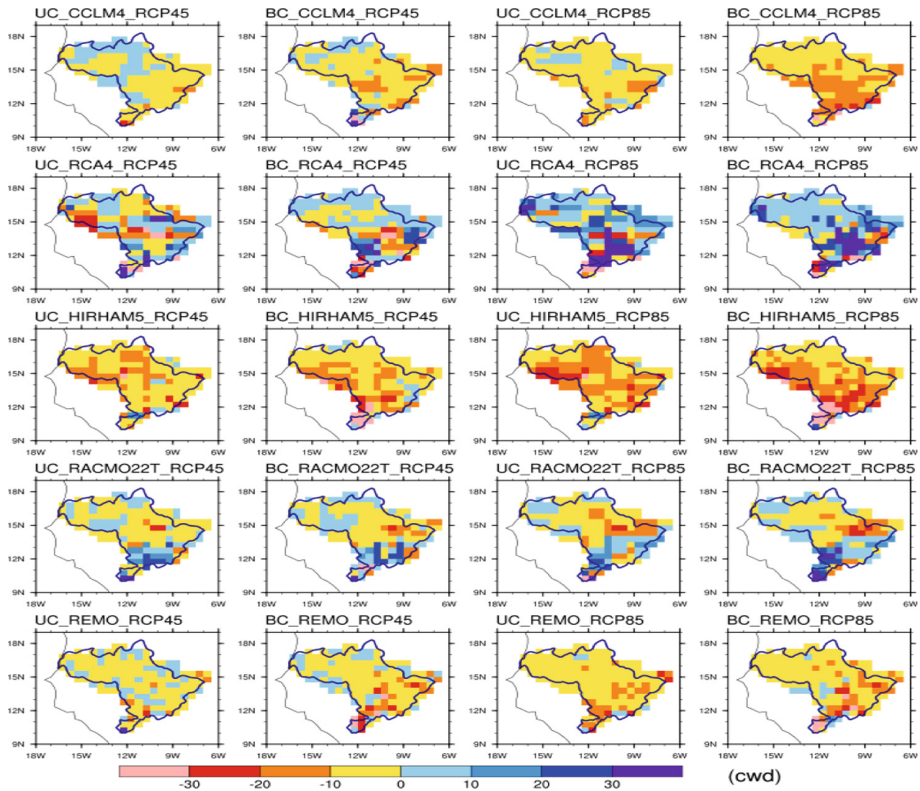


Fig. 4. Changes in the consecutive wet days [UC: uncorrected, BC: bias corrected]

5 Conclusion

In this study, an analysis of the implicit effect of statistical bias correction of five RCMs output on extreme precipitation over the Senegal River Basin was provided. During the present day climate (not shown here), the applied bias correction technique shows generally considerable added value over SRB by successfully removing the RCMs biases during the evaluation part. With respect to the projected climate change signals on extreme precipitation uncorrected and bias corrected data depict substantial changes by the end of 21st century. Divergence was found between models in predicting extreme precipitation over the basin where some models project a decrease and others show an increase. Moreover, the basin may experience substantial drying due to increased number of dry days and decrease of wet spells by four RCMs by contrast to RCA4 which shows more rainfall events. The differences between models output changes were mainly due their different convection scheme. The northern basin is subject to the more pronounced drying. As regard to the radiative forcing, the extreme scenario (RCP8.5) exhibits the higher changes. However, with the most troublesome variable which is precipitation in climate modeling over West Africa, the bias correction may change the signal in specific areas. The main findings of this work is a

drying of the basin due to a likely decrease of wet days and an increase of dry days as it is projected by all RCMs (except for the RCA RCM). Therefore the projected changes over the SRB have to be carefully interpreted due to the various sources of uncertainties that are associated, especially those arising from the choice of the bias correction methodology, the various RCM models and their driving GCMs. To reduce these uncertainties, it would be desirable to conduct analogous analyses with more different RCMs output, different bias correction techniques which would enable uncertainty analyses in future investigations.

Acknowledgments. We thank the Laboratoire Physique de l'Atmosphère et de l'Océan (LPAO/ESP/UCAD/SENEGAL) and the Laboratoire d'Océanographie, des Sciences de l'Environnement et du Climat (LOSEC/UFR ST/Physics Department/UASZ/SENEGAL) where this work has been done. Many thanks to Stefan Hagemann, Andreas Haensler, Tobias Stacke and Christopher Moseley for their supports during our stay in Hamburg (Germany).

References

1. Vizy, E.K., Cook, K.H.: Mid-twenty-first-century changes in extreme events over Northern and Tropical Africa. *J. Clim.* **25**, 5748–5767 (2012)
2. Hagemann, S., et al.: Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth Syst. Dynam.* **4**, 129–144 (2013). <https://doi.org/10.5194/esd-4-129>
3. Giorgi, F., Jones, C., Asrar, G.: Addressing climate information needs at the regional level: the CORDEX framework. *World. Meteorol. Org. Bull.* **58**, 175–183 (2009)
4. Klutse, N.A.B., Sylla, M.B., Diallo, I., et al.: Daily characteristics of West African summer monsoon precipitation in CORDEX simulations. *Theor. Appl. Climatol.* **123**, 369 (2016). <https://doi.org/10.1007/s00704-014-1352-3>
5. Haensler, A., Hagemann, S., Jacob, D.: The role of simulation setup in a long-term high resolution climate change projection for southern African region. *Theor. Appl. Climatol.* **106**, 153–169 (2011). <https://doi.org/10.1007/s00704-011-0420-1>
6. van Roosmalen, L., Sonnenborg, T.O., Jensen, K.H., Christensen, J.H.: Comparison of hydrological simulations of climate change using perturbation of observations and distribution-based scaling. *Vadose Zone J.* **10**(1), 136–150 (2011). <https://doi.org/10.2136/vzj2010.0112>
7. Ly, M., Traore, S.B., Alhassane, A., Sarr, B.: Evolution of some observed climate extremes in the West African Sahel. *Weather Clim. Extremes* **1**, 19–25 (2013). <https://doi.org/10.1016/j.wace.2013.07.005>
8. Faramarzi, M., et al.: Modeling impacts of climate change on freshwater availability in Africa. *J. Hydrol.* **480**, 85–101 (2013)
9. Sarr, M.A., Zorome, M., Seidou, O., Bryant, C.R., Gachon, P.: Recent trends in selected extreme precipitation indices in Senegal – a change-point approach. *J. Hydrol.* **505**, 326–333 (2013)
10. Sarr, M., Seidou, O., Trambly, Y., El Adlouni, S.: Comparison of downscaling methods for mean and extreme precipitation in Senegal. *J. Hydrol. Reg. Stud.* **4**, 369–385 (2015)
11. McSweeney, C., New, M., Lizcano, G.: UNDP Climate Change Country Profiles: Senegal (2010a). <http://country-profiles.geog.ox.ac.uk/>. Accessed 24 Feb 2018

12. McSweeney, C., New, M., Lizcano, G., Lu, X.: The UNDP climate change country profiles improving the accessibility of observed and projected climate information for studies of climate change in developing countries. *Bull. Am. Meteor. Soc.* **91**, 157–166 (2010)
13. Giraldo Osorio, J.D., García Galiano, S.G.: Building hazard maps of extreme daily rainy events from PDF ensemble, via REA method, on Senegal River Basin. *Hydrol. Earth Syst. Sci.* **15**, 3605–3615 (2011). <https://doi.org/10.5194/hess-15-3605-2011>
14. Haerter, J.O., Hagemann, S., Moseley, C., Piani, C.: Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci.* **15**, 1065–1079 (2011). www.hydrol-earth-syst-sci.net/15/1065/2011/, <http://doi.org/10.5194/hess-15-1065>
15. Hagemann, S., Chen, C., Haerter, J.O., Heinke, J., Gerten, D., Piani, C.: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *J. Hydrometeorol.* **12**, 556–578 (2011). <https://doi.org/10.1175/2011JHM1336.1>
16. Schoetter, R., Hoffmann, P., Rechid, D., Schlünzen, K.H.: Evaluation and bias correction of regional climate model results using model evaluation measures. *J. Appl. Meteor. Climatol.* **51**, 1670–1684 (2012). <https://doi.org/10.1175/JAMC-D-11-0161.1>
17. Chen, C., Haerter, J.O., Hagemann, S., Piani, C.: On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. *Geophys. Res. Letters* **38**, L20403 (2011)
18. Dosio, A., Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: evaluation on the present climate. *J. Geophys. Res.* **116**, D16106 (2011). <https://doi.org/10.1029/2011JD015934>
19. Piani, C., et al.: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.* **395**, 199–215 (2010). <https://doi.org/10.1016/j.jhydrol.2010.10.024>
20. OMVS: Sdage du fleuve Sénégal, rapport de phase 1: état des lieux et diagnostic. version finale Décembre, 457 pp. (2009)
21. Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J., Viterbo, P.: The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.* **50**, 7505–7514 (2014). <https://doi.org/10.1002/2014WR015638>
22. Mbaye, M.L., Haensler, A., Hagemann, S., Gaye, A.T., Moseley, C., Afouda, A.: Impact of statistical bias correction on the projected climate change signals of the regional climate model REMO over the Senegal River Basin. *Int. J. Climatol.* **36**, 2035–2049 (2016). <https://doi.org/10.1002/joc.4478>
23. Baldauf, M.: Stability analysis for linear discretisations of the advection equation with Runge-Kutta time integration. *J. Comput. Phys.* **227** (2008). <https://doi.org/10.1016/j.jcp.2008.03.025>
24. Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Weather Rev.* **117**, 1779–1800 (1989)
25. Kain, J.S., Fritsch, J.M.: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.* **47**, 2784–2802 (1990)
26. Kain, J.S., Fritsch, J.M.: Convective parameterization for Mesoscale models: the Kain-Fritsch scheme. In: Emanuel, K.A., Raymond, D.J. (eds.) *The Representation of Cumulus Convection in Numerical Models*. MM, pp. 165–170. American Meteorological Society, Boston, MA (1993). https://doi.org/10.1007/978-1-935704-13-3_16



Climate Change Mitigation Potential in Agricultural and Forestry Sector: The Impact of Expanded Woody Biomass Co-firing on Global Climate Stabilization

Aklesso Y. G. Egbendewe^(✉)

Faculty of Economics and Management Sciences (FASEG), University of Lome,
01 BP 1515, Lome, Togo
emaklesso@gmail.com

Abstract. The impact of expanded woody biomass co-firing in electric power production on global climate stabilization is studied using the Global Biosphere Model (GLOBIOM) and the World Induced Technical Change Hybrid (WITCH) model. The study finds that, even with a ratio of biomass to total feedstock less or equal to 10%, biomass co-firing can help achieve the climate policy goal of 2 °C temperature increase above the pre-industrial level by the end of the century at a lower cost. The policy cost can be further reduced if the ratio of biomass to total feedstock increases via technical progress or biomass supply increases. The study also shows that there is enough biomass potential from agriculture and forestry to progressively replace current nuclear energy supply with bioenergy from co-fired plants. However, replacing current nuclear energy supply with bioenergy produced from co-firing in order to deal with nuclear energy production safety concerns will lead to a high policy cost in terms of total GDP loss. The study finally reveals that future biomass trade from sub-Saharan Africa & Latin America to Europe, North America, and China may be needed for climate policy goals to be reached via biomass co-firing.

Keywords: Global woody biomass supply · Global food price
Nuclear energy phase out

1 Introduction

It is well acknowledged in the literature [1] that mitigation in agricultural and forestry (AF) sector can occur through changes in crop management practices, improvement in livestock management, alteration of crop mix, promotion of good forest management practices, and expansion of bioenergy production as substitutes to traditional fossil energy sources for transportation (via liquid biofuels) & electric power production (via biomass co-firing with coal). While it is desirable that all the above AF mitigation options be deployed to increase mitigation potential, co-firing woody biomass (from short rotation tree plantations and industrial forest logging residues) with coal appears to be an obvious option for short and medium run mitigation for technological and environmental reasons. Technologically, biomass co-firing requires only modest

incremental investment to retrofit existing coal-fired plants or building new biomass co-firing plants [2, 3]. Environmentally, supply of perennial short rotation trees for woody biomass (e.g. Poplar and Eucalyptus) requires little or no tillage before planting and may require scant fertilizer usage to reach maturity [4]. Hence, landscape level environmental impacts such as soil erosion and nutrients loss into surface and groundwater are minimized with woody biomass feedstock production as compared to biomass feedstock production from annual crop residues [4–6].

Alternative uses of woody biomass for energy production which includes converting woody biomass feedstock into second generation liquid biofuels has been a disappointment [7]. Furthermore, woody biomass co-firing is recognized to have high greenhouse gas (GHG) mitigation potential. In fact, Life-cycle analysis (LCA) studies demonstrate that when the quantity of fossil fuels displaced is accounted for, the mitigation potential from woody biomass co-firing could reach up to a carbon sink rate of $400 \text{ g CO}_2\text{e-C.m}^{-2}$ [8, 9]. Climate change mitigation through agricultural and forestry (AF) sector, particularly through biomass co-firing with coal, could be a bridge to the future for short and medium run GHG emissions reduction [10]. For significant but not immediately available investments in low carbon technologies (e.g. wind and solar) are needed to stabilize the global climate by the end of the century. These investments are expensive and may occur late in the century because of the slow capital turnover in the energy sector and the cost ineffectiveness of the current low carbon technologies. Given the current and the future potential threats of climate change [11], there is urgency for short and medium run mitigation actions to reduce GHG emissions.

The objective of this paper is to evaluate the impact of biomass co-firing on global climate stabilization at the CO_2 concentration level of 450 parts per million (ppm) by the end of the century. Specifically, the study does the following: (a) estimate the world economic potential for forest residues production from industrial forest logging and short rotation tree plantations for biomass, (b) evaluate the implications of woody biomass production on land use change, food price increase, and land use CO_2 emissions, (c) measure the impact of the woody biomass co-firing on global CO_2 emissions and GHG abatement. To reach these objectives, three fundamental research questions are to be addressed: (i) what are the drivers of woody biomass production from short rotation tree plantations and forest logging residues?, (ii) what are the implications of such biomass production activities on land use change, food prices and land use CO_2 emissions?, (iii) what is the contribution of such biomass utilization to the climate stabilization goals?

To address these fundamental research questions, this paper proceeds by linking two global models. The Global Biosphere Management Model (GLOBIOM), an agricultural and forestry sector optimization model that includes global crops, livestock and biomass production activities is used to estimate the global potential woody biomass production. The impact of the woody biomass co-firing on emissions reduction and the global climate stabilization is analyzed using the World Induced Technological Change Hybrid (WITCH), an integrated assessment model. This study departs from the previous literature (e.g. [12]) in that: (a) land use has been implicitly modeled to minimize competition between woody biomass production and food production by including forest logging residues in the stock of biomass feedstock, (b) Cellulosic

grasses such Switchgrass and Miscanthus are not included in the feedstock potential to consider only hard wood with high energy content for coal-fired power plants system.

2 Materials and Methods

2.1 Description of GLOBIOM

The Global Biosphere Management Model (GLOBIOM) is a global agricultural and forestry sector optimization model built to tackle several global policy challenges related to bioenergy, food security and environmental management [13]. The structure of the model is similar to the U.S agricultural and forestry sector optimization model [14] and it maximizes the total welfare from agricultural and forestry sector under several resource constraints. Crop yields are simulated from the Environmental Policy Integrated Climate (EPIC), a biophysical crop growth model [15]. The model has a total of 17 crops modeled through EPIC based on fertilization level and irrigation. Other crops not simulated by EPIC are in number of 17 as well. These latter crops are modeled according to two management systems (irrigated or rain fed). In addition, two types of woody biomass sources including short rotation trees and residues from industrial forest logging are simulated.

The structure of land management allows short rotation trees to be grown on land that are marginal to crop production and therefore minimizing direct competition between cropland and short rotation tree plantations. However, minimal land competition may occur when areas cleared through deforestation are reconverted into short rotation tree plantation instead of food crops. This land competition occurs only if the marginal value of biomass production is superior to the marginal value of crop production. To further make sure that land competitions between food crops and biomass crops is minimized, the model requires that food be produced to dynamically satisfy minimum predetermined food calorie demand constraints up until the last simulation year (2100). The model is calibrated so as to replicate the base year (2000) data as reported by the Food and Agriculture Organization (FAO). GLOBIOM is a recursive model which runs up to 2100-time horizon with 10 years' time steps. The model is flexible for outputs to be aggregated into sub-regional dataset of interests to be passed on to energy models such as WITCH.

2.2 Description of WITCH Model

The World Induced Technical Change Hybrid (WITCH) model is an integrated assessment model (IAM) built on the optimal growth modeling framework. Consumption is assumed to be maximized under the constraint that it must equal production net of investments [16]. Production and investment functions are defined to represent a medium complexity energy sector. A build-in climate module allows accounting for GHG emissions throughout the whole economy. Emission caps are defined for various climate stabilization policy targets. One feature that distinguishes WITCH from other IAM is the game theoretical framework that is utilized to find the equilibrium for the 13 regions of the model. The 13 regional aggregations are CAJAZ

(Canada, Japan and New Zealand), CHINA (including Taiwan), EASIA (East Asia including Indonesia), INDIA, KOSAU (South Korea, South Africa, and Australia), LACA (Latin America, Mexico and Caribbean), MENA (Middle East and North Africa), WEST EUROPE (Old EU countries), EAST EUROPE (new EU countries), SASIA (South Asia), SSA (Sub-Saharan Africa), TE (Transition Economies which includes non-EU Eastern European countries and Russia), and the USA.

In WITCH, energy is an input to good and services production and comes as non-electric (NEL) and electric (EL) energy. Non-electric energy is from oil, gas and first generation liquid biofuels as well as traditional biomass. Electric energy comes from hydroelectric, nuclear, wind, solar, and fossil fuels nest (coal, oil, and gas). Woody biomass and coal are modeled as one electric energy nest with infinite elasticity of substitution. Each can operate under two technologies that differ by costs. Coal/biomass can be pulverized (ELP) or gasified through integrated gasification combined cycle (IGCC) with carbon capture and storage (CCS) to produce electricity. The original WITCH model functions with separate combustion of coal and biomass feedstocks. Co-firing coal with biomass can improve the efficiency of the mix (40%–50%). The original WITCH model is then modified to simulate two biomass co-firing technologies [2]. The first consists of directly co-firing coal with woody biomass in a single boiler. As a result, a mixture above 10% of biomass would reduce the efficiency [17]. The second consists of indirectly co-firing biomass by first converting it into a gaseous fuel before burning it with coal. In this case, there is no physical limit to the percentage of biomass in the mixture. Any excess of biomass not co-fired in the model is used in a stand-alone biomass plants even though the efficiency is reduced. A brief description of biomass and coal nest for electric power production in WITCH with the newly added co-firing technologies (in dash) is given in Fig. 1.

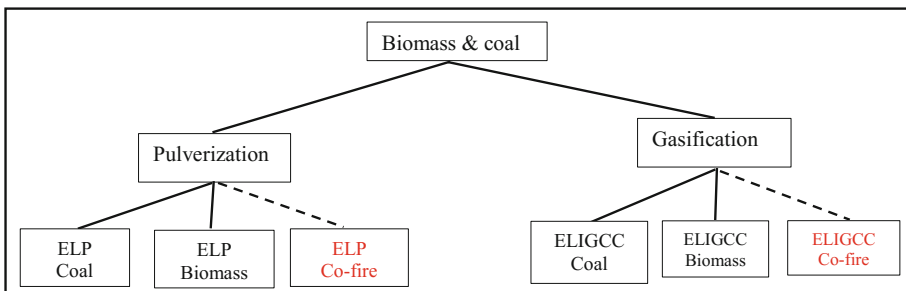


Fig. 1. Coal and biomass electric power production technology (all technologies function with carbon storage and capture technology).

3 Data and GLOBIOM Simulation Results

3.1 Industrial Forest Logging Residues Biomass

Biomass from forest logging residues originates from managed forests around the globe. In GLOBIOM, forest is managed for saw logs, pulp logs, other industrial logs, traditional

fuel woods and biomass for bioenergy. Biomass price is simulated from \$7 US/GJ to \$40 US/GJ. The world potential biomass supply from forest logging residues varies from 0 EJ in 2000 to about 146 EJ by 2100. While this supply function is less sensitive to time trend (because of the fixed managed forest areas) it is sensitive to prices. The highest potentials could be reached when price incentives are at their highest levels. Policies that encourage utilization of biomass feedstock in co-firing (e.g. the renewable fuel standard (RFS) policies in several states in the U.S. and in Europe) could increase market incentives for woody biomass production and trade. The group of regions including LACA, SSA and EASIA are the big producers and regions such as CHINA, USA, INDIA and WESTERN EUROPE may be potential future importers because of their high dependence on coal for electric power supply. This may open future biomass trade businesses across continents for low carbon electric power production.

3.2 Short Rotation Trees Biomass Production, Land Use and World Food Prices

Biomass from short rotation trees is obtained through plantation of poplar trees on land areas previously covered by natural vegetation, grassland and unmanaged forests. Natural vegetation land cover and grassland are marginal to agriculture crop production. However, since biomass from short rotation trees is grown on land areas previously covered by other existing vegetation and grasses, indirect land use emissions may occur from such biomass production activities. Biomass supply potential from short rotation tree plantations varies from 1.2 EJ in 2000 to about 158 EJ in 2100. In contrast with the supply schedule of biomass from forest logging residues, biomass from short rotation trees is less sensitive to price but more sensitive to time. This means that the full potential could not be reached based on price incentives alone. More time is needed for full conversion of available land areas into short rotation tree plantations. Of course, the potential estimated for short rotation trees biomass is linked to yield assumptions. Gradual increases in yield over the simulation period due to technical change could increase the supply potential.

To reach the economic potential of short rotation tree plantations, the land area needed will vary from 0 ha in 2000 to 5.6 billion ha in 2100. This land area is more than three times the current global crop land areas. Therefore, significant dedicated land areas will be required to reach high biomass supply potential from short rotation tree plantations unless yields improve significantly. Given the weak substitutability between land use from crops and short rotation trees in GLOBIOM, the potential impact of biomass production on food price is reduced. It is important to notice that the group of regions such as LACA and SSA which are not big coal users are endowed with more land areas for biomass production than other regions in the model. Therefore, other may need to import biomass processed into chips from LACA and SSA to implement significant co-firing policy. These land use changes have no significant effects on food prices. The world food prices index increases to reach 1.2 in 2020 due significantly to first generation biofuel mandates in various parts of the world including the U.S and Europe. The model assumes that first generation biofuel mandates are not continued after 2020. Prices decrease slowly to reach their pre-mandate levels by the end of the century.

3.3 Land Use CO₂ Emissions and Co-firing Technical Data

WITCH does not have endogenously estimated values of land use CO₂ emissions. They are computed in GLOBIOM based on emissions from deforestation, afforestation and indirect land use change emissions. To capture the relationship between land use CO₂ emissions, carbon price, and woody biomass production; higher order polynomial functions with interactive terms and time trend have been econometrically estimated to predict the level of land use CO₂. This allows land use CO₂ levels to be endogenous in WITCH and reflected in climate change policy scenarios. To implement the co-firing experiments in the WITCH model, several parameters related to the investments in retrofitting existing coal plants as well as the operating and maintenance costs are drawn from the literature [2, 18].

4 WITCH Model Simulation Results

The paper conducts four policy experiments all based on the climate policy target of 2 °C above the pre-industrial level by 2100. In the first policy experiment, coal-fired plants producing electricity using pulverization and gasification technologies are set to progressively phase out by constraining investments in these technologies to zero starting from 2020 with 10% biomass in the feedstock mix (Experiment 1). In the second experiment, the ratio of biomass to total feedstock is progressively increased up to 30% for pulverized plants and up to 100% for gasified plants (Experiment 2). In the third experiment, the supply of woody biomass is allowed to increase by doubling land areas dedicated to short rotation trees (Experiment 3). In the fourth policy experiment, nuclear energy is allowed to progressively phase out by setting investments in that technology to zero starting from 2050 (Experiment 4). Simultaneously with this last experiment, the supply of woody biomass for co-firing is increased by setting land use for all years at the predicted 2050 level (an average of 3.8 billion ha). This fourth policy experiment is justified by safety concerns of nuclear energy production due to past nuclear accidents (Chernobyl in 1986, Fukushima in 2011, etc...). Also, in the last three experiments, progressive pure coal fired plants phase out by 2020 is maintained. The obtained results in the four policy experiments are presented in terms of changes in CO₂ emissions in electric sector, total GHG abatement, and policy costs measured as the total world GDP loss relative to the situation without the co-firing policy in place. Note that the base scenario functions via a cap-and-trade to meet climate policy targets without biomass co-firing.

4.1 Changes in CO₂ Emissions from the Electric Sector

The progressive replacement of pure coal plants by biomass co-firing plants is expected to lower in average CO₂ emissions in the electric power production sector. However, progressive replacement of nuclear energy production plants by biomass co-firing plants is expected to increase on average the total CO₂ emissions from the electric sector. The results are presented in Fig. 2. The graphs in Fig. 2 show that the policy

experiment 3 in which biomass feedstock supply is increased via doubling of land use dedicated to short rotation trees has produced on average the lowest CO₂ emissions from the electric power production sector. The reason behind this result is that many countries in Europe, North America and Asia who rely on pure coal plants for electric power production do not have enough woody biomass feedstock to implement the policy. Therefore, an increase of biomass supply is necessary to reach the objectives of the co-firing policy. Hence, the main limitation of the first two policy experiments is the unbalance between the supply and the demand of biomass in many regions. As a consequence, the trade of biomass between regions can solve the biomass supply-demand unbalances.

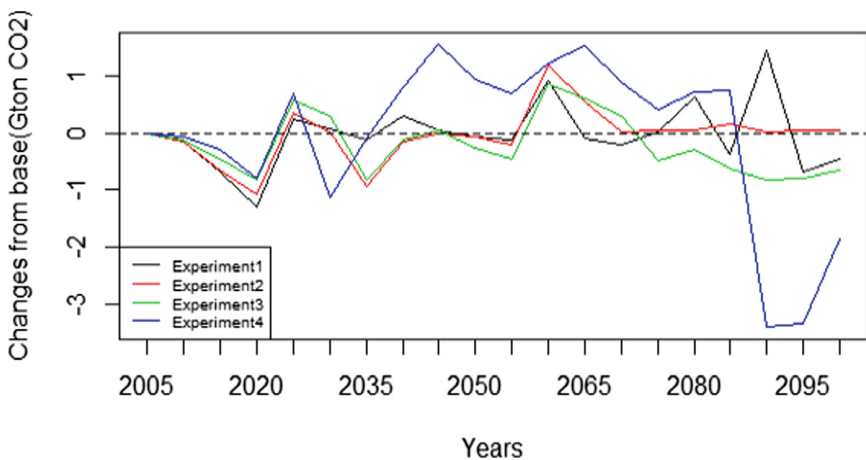


Fig. 2. CO₂ emissions from the electric power production sector measured as the difference of co-firing policy experiments relative to the base emissions with pure coal plants

4.2 GHG Abatement

The total GHG abated is expected to increase as CO₂ and other GHG (e.g. SO₂) emissions drop in average with co-firing policies in place. The GHG abatement trajectories are given in Fig. 3. Significant GHG abatement will only occur starting from 2045 for both the policy experiment 3 where biomass supply is increased through land use and the policy experiment 4. However, the abatement occurring under the policy experiment 4 related to the phase out of the nuclear energy is due to the use of CCS in the last years of the century making the abatement more costly. The first two experiments cause less significant abatement of GHG relatively to the base with pure coal plants.

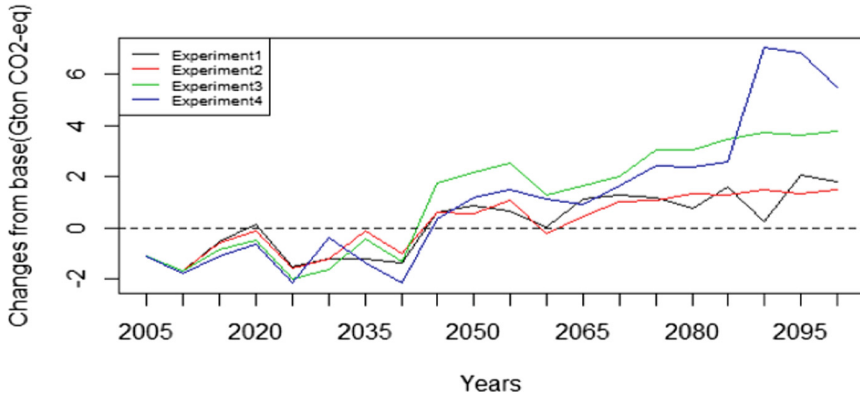


Fig. 3. GHG abatement measured as the difference of each co-firing policy experiment relative to the base emissions with pure coal plants

4.3 Policy Costs

The estimated discounted costs of the four co-firing policy experiments show that progressively phasing out pure coal plants starting from 2020 and replacing them by biomass co-firing (experiment 1) will cost 5.39% of the world GDP versus 5.43% when the co-firing technology is not in place. Increasing biomass to coal ratio up to 30% for pulverized plants and up to 100% for gasified plants (experiment 2) will cost 5.38% of the world GDP or 0.01% less than in the first experiment. Not surprisingly, the policy cost under policy experiment 3 (where biomass supply is increased) is 5.31% of the world GDP which is the lowest of all the experiments. The policy cost is high in the policy experiment 4. In fact, phasing out nuclear energy will increase the policy cost to 6.09% of world GDP given that nuclear energy production is a low carbon technology relatively to any co-firing technology.

5 Conclusion and Discussion

This paper studies the impact of biomass co-firing on global climate stabilization at 2 °C temperature increase above the pre-industrial level by the end of the century. To do this, the global woody biomass supply potential is estimated using the Global Biosphere Management Model (GLOBIOM). The estimated supply of woody biomass is then passed on to the World Induced Technical Change Hybrid (WITCH) model to measure the impact of woody biomass co-firing on CO₂ emissions and GHG abatement. Four policy experiments are conducted under various biomass co-firing settings. In the first policy experiment, coal plants are set to progressively phase out by 2020 in replacement with biomass co-firing plants. The biomass to total feedstock ratio is assumed to increase due to technical progress in the second policy experiment. In the third policy experiment, biomass supply quantity is increased by raising land resources dedicated to woody biomass production. Nuclear energy production is allowed to

progressively phase out in replacement with biomass co-firing plants in the fourth policy experiment.

The results suggest that biomass co-firing can help achieve climate stabilization at 2 °C temperature increase above the pre-industrial level by 2100 with lower policy costs. Under the first policy experiment of woody biomass co-firing (with a maximum of 10% biomass constraint) the policy cost is estimated at 5.39% of the world GDP against 5.43% of the world GDP if the co-firing policy is not in place. An increase in woody biomass co-firing ratio (with more than 10% biomass in the mix) will slightly improve the policy cost to 5.38% of the world GDP. The reduction in policy cost relative to the base is higher when supply of woody biomass is allowed to increase through raising land resources dedicated to woody biomass production. In fact, when woody biomass land use is increased from 1 to 2 billion hectares throughout to the end of the century, the policy cost dropped to 5.31% of the world GDP. However, phasing out nuclear energy production system in replacement with biomass co-firing plants will cost 6.09% of the world GDP due to the replacement of a low carbon technology by a more carbon intensive technology though partially renewable energy production sources. The study also shows that the implementation of biomass co-firing technology will require the participation of all the world regions into a global biomass trade. Land abundant regions such as sub-Saharan Africa (SSA) and Latin America have comparative advantages in the production of woody biomass from short rotation trees. Yet these two regions use less coal in the production of energy. Therefore, production and pre-processing of biomass may occur in these regions and ship to others regions that are more dependent on coal in energy production.

Future research that considers more use of woody biomass not only in electric power production but also in transportation fuels production is warranted. These future studies will help find better ways to use agricultural and forestry resources to find solutions to energy and climate change problems in the short and medium terms.

Acknowledgement. Part of this paper was written when the author was a Senior Researcher at Fondazione Eni Enrico Mattei (FEEM) in Milan, Italy. I wish to thank researchers from FEEM, CMCC, IIASA and Mark Musumba at the University of Florida, Gainesville.

References

1. McCarl, B.A., Schneider, A.: Greenhouse gas mitigation in U.S. agriculture and forestry. *Science* **294**, 2481–2482 (2001)
2. IEA-ETSAP&IRENA: Biomass co-firing: technology brief. International Energy Agency (IEA)-The Energy Technology Systems Analysis Programme (ETSAP) and The International Renewable Energy Agency (IRENA), Paris, France (2013)
3. Roni, M.S., Chowdhury, S., Mamun, S., Marufuzzaman, M., Lein, W., Johnson, S.: Biomass co-firing technology with policies, challenges, and opportunities: a global review. *Renew. Sustain. Energy Rev.* **78**(1), 1089–1101 (2017)
4. Robertson, P.G., Hamilton, S.K., Del Grosso, J.S., Parton, W.J.: The biogeochemistry of bioenergy landscapes: carbon, nitrogen and water considerations. *Ecol. Appl.* **21**(4), 1055–1067 (2011)

5. Egbendewe-Mondzozo, A., Swinton, S.M., Izaurralde, R.C., Manowitz, D.H., Xuesong, Z.: Biomass supply from alternative cellulosic crops and crop residues: a spatially explicit bioeconomic modeling approach. *Biomass Bioenergy* **35**(11), 4636–4647 (2011)
6. Love, B.J., Nejadhashemi, P.A.: Water quality impact assessment of large-scale biofuel crops expansion in agriculture regions of Michigan. *Bioenergy Biomass* **35**, 2200 (2011)
7. Schnoor, J.L.: Cellulosic biofuels disappoint. *Environ. Sci. Technol.* **45**(17), 7099 (2011)
8. Adler, P.R., DelGrosso, J.S., Parton, W.J.: Life-cycle assessment of net greenhouse-gas flux for bioenergy cropping systems. *Ecol. Appl.* **17**(3), 675–691 (2007)
9. Heller, M.C., Keoleian, G.A., Volk, T.A.: Life cycle assessment of a willow bioenergy cropping system. *Biomass Bioenergy* **25**(2), 147–165 (2003)
10. McCarl, B.A., Sands, R.D.: Competitiveness of terrestrial greenhouse gas offsets: are they a bridge to the future? *Clim. Change* **80**(1–2), 109–126 (2007)
11. Field, C.B., et al. (eds.): IPCC: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge/New York (2014)
12. Popp, A., Dietrich, J.P., Lotze-campen, H., et al.: The economic potential of bioenergy for climate change mitigation with special attention given to implications for the land system. *Environ. Res. Lett.* **6**(3), 1–9 (2011)
13. Havlik, P., Schneider, U.A., Schmid, E., et al.: Global land-use implication of first and second generation biofuel targets. *Energy Policy* **39**(10), 5690–5702 (2011)
14. Adams, D., Alig, R., Callaway, J., McCarl, B.: Forestry and agricultural sector optimization model description. Final Report to the EPA, Climate Change Division. Washington, DC (1994)
15. Zhang, X., Izaurralde, R., Manowitz, D., et al.: An integrative modeling framework to evaluate the productivity and sustainability of biofuel crop production systems. *GCB Bioenergy* **2**, 258–277 (2010)
16. Bosetti, V., Carraro, C., Galeotti, E., Massetti, E., Tavoni, M.: WITCH: a world induced technical change hybrid model. *Energy J.* **1**(1), 13–38 (2006)
17. GCEP: An assessment of biomass feedstock and conversion research opportunities. Stanford University, Stanford, California, Global Climate & Energy Project (2005)
18. IEA-ETSAP: Biomass for heat and power. International Energy Agency (IEA), Energy Technology Systems Analysis Programme (ETSAP), Paris (2010)



Climate Change May Result in More Water Availability in Parts of the African Sahel

Ousmane Seidou^{1,2(✉)}

¹ Department of Civil Engineering, University of Ottawa,
161 Louis Pasteur Office A113, Ottawa, ON K1N 6N5, Canada
oseidou@uottawa.ca

² Centre for Water, Environment and Health, United Nations University,
204-175 Longwood Rd S, Hamilton, ON L8P 0A1, Canada

Abstract. The African Sahel is known for its climate variability that often translates into recurrent droughts. Rainfall has drastically decreased substantially across the Sahel from the 1950s until at least the late 1980s. It is unclear from the literature and from the fifth IPCC assessment report whether the trend in annual rainfall in the next decades would be decreasing as observed throughout the 20th century or increasing as suggested by a significant number of climate models. There is however a low to medium confidence that extreme rainfalls would increase. The objective of this paper is to demonstrate that both possibilities (an increase or a decrease in rainfall in the future) may result in more opportunities to mobilize water for populations in the Sahel. To demonstrate that, the ability of 20 regional climate models is evaluated based on their ability to reproduce key parameters of the rainy season in Niger, West Africa. The outputs of the 10 best models are then downscaled at 52 climate stations in the country to generate precipitation projections up to year 2100. Results show that a wetter climate is more likely than a drier climate at horizons 2021–2050, 2051–2075 and 2071–2100 compared to the 1979–2014 period; The paper also examines the so-called ‘Sahelian Paradox’, an observed counter-intuitive phenomena where decrease in rainfall resulted in a higher surface runoff, pointing to opportunities for water harvesting in the eventuality of a drier climate.

Keywords: Climate change · Rainfall · General circulation models
Regional climate models · Runoff · Sahelian Paradox

1 Introduction

The African Sahel is well known for its rainfall variability which results in recurrent droughts and food insecurity and, from time to time, humanitarian crises. The entire region has the scares of persistent and severe drought episodes that occurred in the 70s and the 80s, following a very wet period in the 50s and the

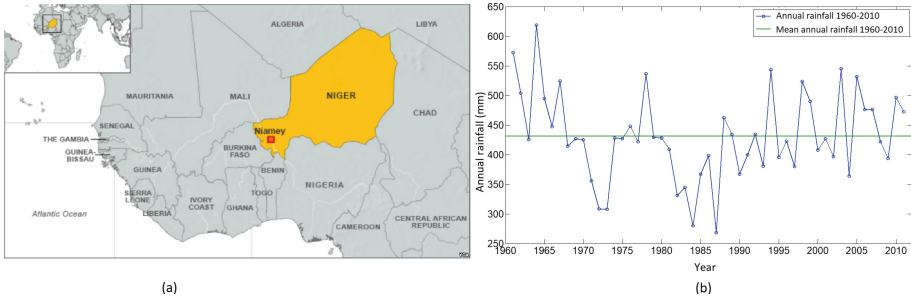


Fig. 1. Localisation of the Niger republic (a) and annual precipitation in Niger from 1960 to 2010 (b)

60s. The Niger republic (Fig. 1, panel a) is a landlocked country of 1,267,000 km² which has been struggling to achieve food self-sufficiency.

The vast majority of the population rely on rainfed agriculture for their subsistence, without any water storage infrastructure, hence their vulnerability to rainfall anomalies. The discharge of the only major river flowing through the country (the Niger river) has decreased drastically as result of lower rainfall in the watershed, but also because of the construction of large dams and irrigation schemes upstream (Fig. 1, panel b). The flow of the Niger river is reported to be 34% lower at Niamey in the 1970–1989 period compared to the 1950–1969 period [9]. Surprisingly, according to several authors [e.g. 3], surface runoff coefficients and stream flows have increased in most Sahelian areas because of land use change and the rise of the water table resulting from an increased infiltration. The increase in runoff despite the decrease in rainfall is called the ‘*Sahelian Paradox*’ is an example of a counter-intuitive impact of a change in climate normal: the drought had at least one positive impact: it actually improved the opportunities to capture runoff in surface water reservoirs. On the other hands, a majority of global circulation models from the CMIP5 experiment suggest a wetter climate for the Sahel, with a few ones pointing to a dryer climate.

The objective of this paper is to examine past and future trends in precipitation and runoff in the Niger republic and assess how these trends may affect water availability within the country in the future. First, trends in rainfall and runoff are examined in a small watershed in central Niger to highlight the Sahelian Paradox. A rainfall runoff model is afterward developed and calibrated on the watershed. The outputs of an ensemble of 10 regional climate models of the CORDEX experiment were bias-corrected and fed into the rainfall runoff model to obtain a streamflow ensemble.

2 Materials and Methods

2.1 Study Area

The Study area is the Maggia river watershed located in the center south of the Niger republic. The watershed (Fig. 2) has an area of 2238 km² and has an

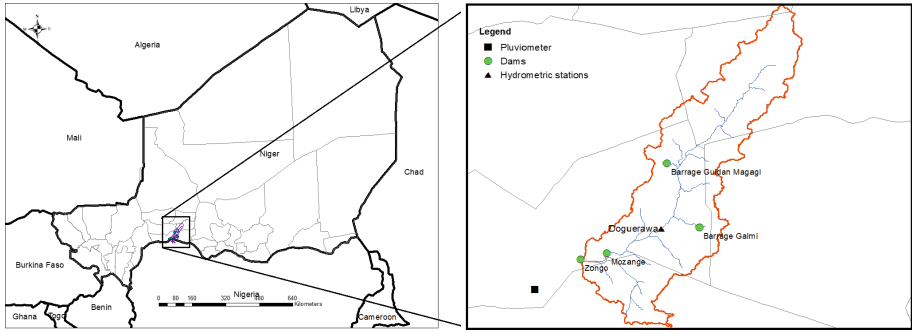


Fig. 2. Localisation of the Maggia river watershed

annual average precipitation which varies from 400 mm in dry years to 850 mm in wet years. The precipitation is concentrated in a 4 month rainy season (June to September) while the rest of the year is dry. The closest rainfall measurement station is that of Birinin’Konni while the only operational hydrometric station is that of Doguerawa, which was in operation between 1954 and 1979, and between 2010 and 2016. Four small dams on the watershed with capacity going from 1.5 millions m^3 to 30 millions m^3 were built in the 80s on the river network to insure irrigation in both the rainy season and part of the dry season. The filling of the dam and the success of irrigation in the dry season depends of the annual precipitation on the watersheds, and it is common for the area to experience crop production losses due to insufficient rainfall and river flows. One interesting feature in the watershed is the relationship between precipitation and streamflow. The time series of annual precipitation at Birnin’Konni between 1954 and 2014 is superimposed on Fig. 3 with the time series of annual streamflow at the Doguerawa station for the same period. While precipitation has been decreasing between 1950 and 1980, recorded streamflow kept increasing in a counterintuitive manner. The increase in runoff despite the decrease in precipitation has been documented in several other locations in the Sahel [3, 10]. This phenomenon, mainly attributed to land use changes, is called the ‘Sahelian Paradox.’ The prolonged drought in the Sahel had at least one positive impact: it actually improved the opportunities to capture runoff in surface water reservoirs.

2.2 Climate Change Projections in the Study Area

According to the International Panel on Climate Change [6], projected rainfall change over sub-Saharan Africa in the mid- and late 21st century is uncertain as the ensemble of climate models used for the assessment disagree in the direction of change over many areas in the region [8]. CMIP5 is an Intercomparison experiment that studies the strength and weaknesses of all major coupled atmosphere-ocean general circulation models. The outputs of CMIP5 were extensively used in the preparation of the fifth IPCC assessment report [6]. The CMIP5 multi-model average suggests either a slight increase in precipitation, or a strong agreement

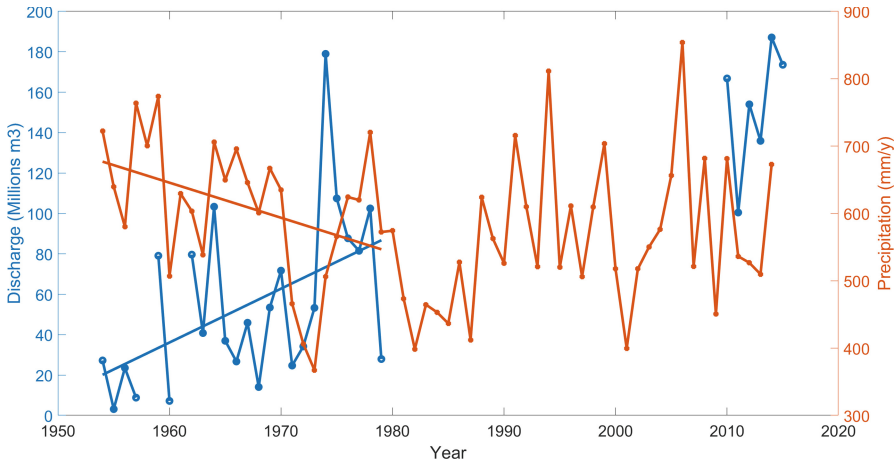


Fig. 3. Time series of annual precipitation at Birnin’Konni and annual runoff at Doguerawa, 1954–2014. The straight lines represent the trend in precipitation and runoff between 1954 and 1979 and illustrate the so-called ‘sahelian paradox’

about a wetter future in most of the country in both periods, with a few spots where model prediction diverge. There is therefore a significant possibility that the country experiences a wetter climate in the coming decades, although the opposite is possible. One criticism often heard about general circulation models is that their spatial resolution is too low to describe the features of regional climates.

Dynamical or statistical downscaling is often used to post-process the outputs of general circulation models to reduce the bias due to the coarse spatial scale and inevitable simplifications in the physical processes description in the model algorithms. The CORDEX [COoRdinated Downscaling Experiment: 5] aims to simultaneously evaluate global climate models and produce climate projections for use in impact and adaptation studies, mainly through dynamical downscaling. To corroborate the projections of general circulation models, a set of 40 regional climate models outputs (20 different climate models listed in Table 1, two RCP scenarios: RCP4.5 and RCP8.5) were obtained from the CORDEX team.

The RCMs were evaluated on their ability to reproduce the following key rainy season characteristics in Niger: (a) start of the rainy season; (b) end of the rainy season; (c) length of the rainy season; (d) average annual rainfall; (e) time distribution of daily rainfall. At the end of the intercomparison, the following 10 models (in order of decreasing performance) were retained:

1. REMO2009-v1 IPSL-IPSL-CM5A-LR
2. RCA4-v1 IPSL-IPSL-CM5A-MR
3. REMO2009-v1 ICHEC-EC-EARTH
4. RCA4-v1 CNRM-CERFACS-CNRM-CM5
5. WRF331-v1 NCC-NorESM1-M

Table 1. List of RCMs used in the study

MODEL	CANRCM4-CanESM2	GCLM-4-8-17-IHEC-EC-EARTH	GCLM-4-8-17-OHC-HadGEM2-ESM	GCLM-4-8-17-MPI-M-MPI-ESM-LR4-v1	HIRHAM5-v1-NCC-NorESM1-M	RACMO22T-IHEC-EC-EARTH	CCMA-CanESM2	CNRM-CERFACS-CNRM-CM5	CSIRO-QCCCE-CSIRO-Mk3-6-0	RCA4-v1-IHEC-EC-EARTH
RCM										
INSTITUTION	CANRCM4 (Canadian Centre for Climate Modelling and Analysis, Victoria, BC, Canada)	GCLM-4-8-17 (Climate Modelling Community)	GCLM-4-8-17 (Climate Modelling Community)	GCLM-4-8-17 (Climate Modelling Community)	HIRHAM5-v1 (Danish Meteorological Institute)	RACMO22T (Royal Netherlands Meteorological Institute)	CCMA-CanESM2 (Canadian Centre for Climate Modelling and Analysis, Victoria, BC, Canada)	CNRM-CERFACS-CNRM-CM5 (Centre National de Recherches Meteorologiques, France)	CSIRO-QCCCE-CSIRO-Mk3-6-0 (Commonwealth Scientific and Industrial Research Organisation, Australia)	RCA4-v1 (Rohini Centre for Atmospheric Research, Sweden)
GCM	CanESM2	IHEC-EC-EARTH	MOHC-HadGEM2-ES	MPI-M-MPI-ESM-LR	NCC-NorESM1-M	IHEC-EC-EARTH	CCMA-CanESM2	CNRM-CERFACS-CNRM-CM5	CSIRO-QCCCE-CSIRO-Mk3-6-0	IHEC-EC-EARTH
MODEL	IPSL-IPSL-CM5A-MR-CanESM2	MIROC5-IHEC-EC-EARTH	MOHC-HadGEM2-ES-MOHC-HadGEM2-ES	MPI-M-MPI-ESM-LR	NCC-NorESM1-M-NCC-NorESM1-M	NOAA-GFDL-ESM2M-IHEC-EC-EARTH	ICHEC-EC-EARTH	IPSL-IPSL-CM5A-LR-CNRM-CERFACS-CNRM-CM5	MOHC-HadGEM2-ES-CSIRO-QCCCE-CSIRO-Mk3-6-0	NCC-NorESM1-M-IHEC-EC-EARTH
RCM										
INSTITUTION	IPSL-IPSL-CM5A-MR (Institut Pierre-Simon Laplace, France)	MIROC5 (Model for Interdisciplinary Research on Climate, Japan)	MOHC-HadGEM2-ES (Met Office Hadley Centre, UK)	MPI-M-MPI-ESM-LR (Max Planck Institute for Meteorology, Germany)	NCC-NorESM1-M (Norwegian Centre for Environmental Modelling and Prediction, Norway)	NOAA-GFDL-ESM2M-IHEC-EC-EARTH (National Oceanic and Atmospheric Administration, USA)	ICHEC-EC-EARTH (Institute for Climate and Earth System Studies, Ireland)	IPSL-IPSL-CM5A-LR-CNRM-CERFACS-CNRM-CM5 (Institut Pierre-Simon Laplace, France)	MOHC-HadGEM2-ES-CSIRO-QCCCE-CSIRO-Mk3-6-0 (Met Office Hadley Centre, UK)	NCC-NorESM1-M-IHEC-EC-EARTH (Norwegian Centre for Environmental Modelling and Prediction, Norway)
GCM	IPSL-IPSL-CM5A-MR	MIROC5	MOHC-HadGEM2-ES	MPI-M-MPI-ESM-LR	NCC-NorESM1-M	NOAA-GFDL-ESM2M-IHEC-EC-EARTH	ICHEC-EC-EARTH	IPSL-IPSL-CM5A-LR	MOHC-HadGEM2-ES	NCC-NorESM1-M

6. RCA4-v1 MPI-M-MPI-ESM-LR
7. RCA4-v1 MIROC-MIROC5
8. CCLM-4-8-17 MPI-M-MPI-ESM-LR
9. RCA4-v1 CCCma-CanESM2
10. RCA4-v1 NOAA-GFDL-GFDL-ESM2M.

The outputs of the best models were further statistically downscaled at 53 cities across the country, using a quantile matching algorithm. The minimum, median and maximum percent change in precipitation were calculated at each of the cities over the 2021–2050, 2052–2075, 2076–2100 periods, then linearly interpolated to create maps of expected changes in precipitation over the country. These changes are presented in Figs. 4 (RCP4.5) and 5 (RCP8.5).

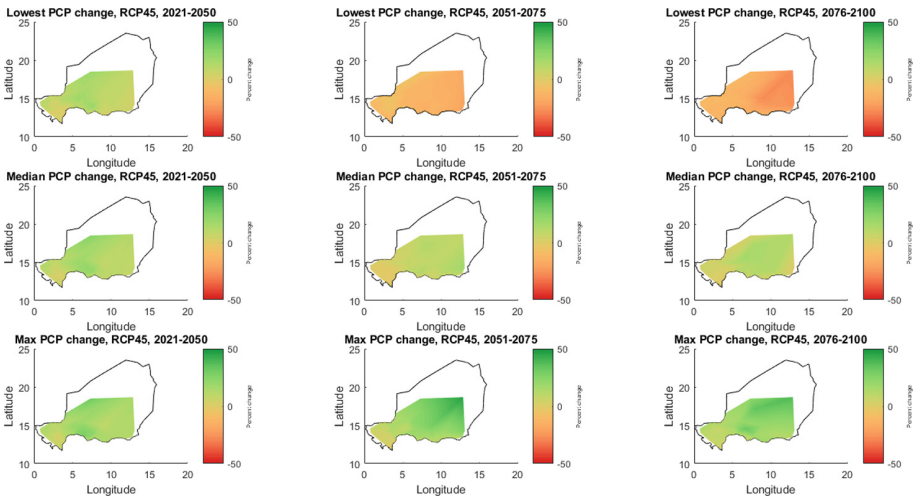


Fig. 4. Projected minimum, median and maximum change in precipitation over Niger under RCP4.5, for the 2021–2050, 2051–2075, 2076–2100 periods

2.3 Hydrological Modelling

In order to further understand the impacts of climate change on the hydrology of the Maggia river, a SWAT [Soil and Water Assessment Tool: 1] model of the watershed was developed. SWAT is a popular semi-distributed rainfall model commonly used for regional scale hydro-ecological modeling. The model was set-up with the following data:

1. Digital elevation model: Shuttle Radar Topography Mission [7].
2. Land use maps: Global Land Cover Facility [2].
3. Soil maps: Food and Agricultural Organization of the United Nations Organization for Education, Science and Culture [4].

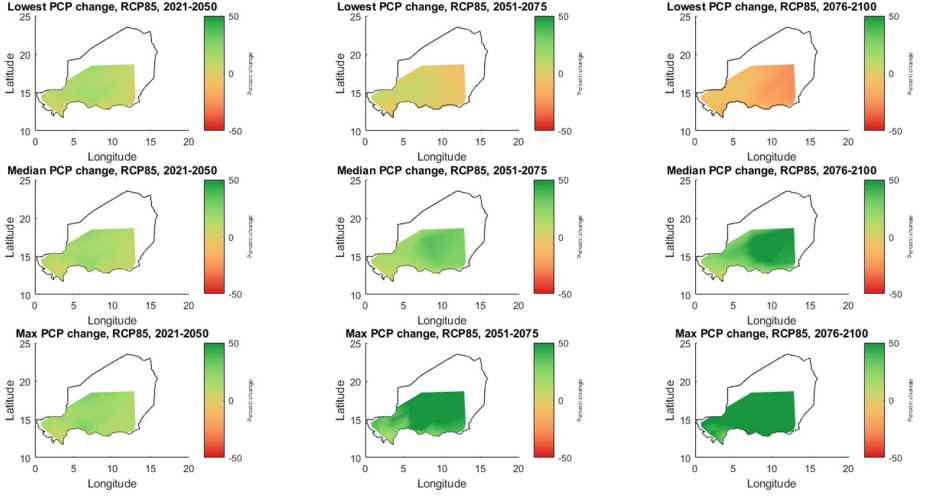


Fig. 5. Projected minimum, median and maximum change in precipitation over Niger under RCP8.5, for the 2021–2050, 2051–2075, 2076–2100 periods

4. Climate data: WATCH-Forcing-Data-ERA-Interim [11].
5. Streamflow data: recorded streamflow data at the Doguerawa hydrometric station.

The SWAT model was calibrated using the most recent streamflow observations (2010–2016) to better represent the current state of the watershed. Given the short length of the observed streamflow time series, the model was not validated. The fit between the observed and simulated time series was evaluated using the Nash Suttcliffe model efficiency coefficient E :

$$E = 1 - \frac{\sum_{t=1}^T (Q_t^m - Q_t^o)^2}{\sum_{t=1}^T (Q_t^o - \bar{Q}_o)^2} \quad (1)$$

where Q_t^m is the modelled streamflow at time t , Q_t^o is the observed streamflow at time t and \bar{Q}_o the average streamflow. E can vary from $-\infty$ to 1, and a good model typically have a Nash coefficient of 0.7 and above. The calibration in this project was found to be satisfying given the quality of the data, with a Nash Suttcliffe coefficient of 0.76. The SWAT model was therefore forced with the downscaled outputs of the 10 selected RCMs under RCP4.5 and RCP8.5, and the projected changes in the discharge at doguerawa and the number of days the dams contain more than the average volume (25 millions m^3) were calculated.

3 Projected Changes in Hydrological Variables

The boxplots in Fig. 6a show that streamflow at doguerawa is expected to be higher in the future under both scenarios, with more and more variability

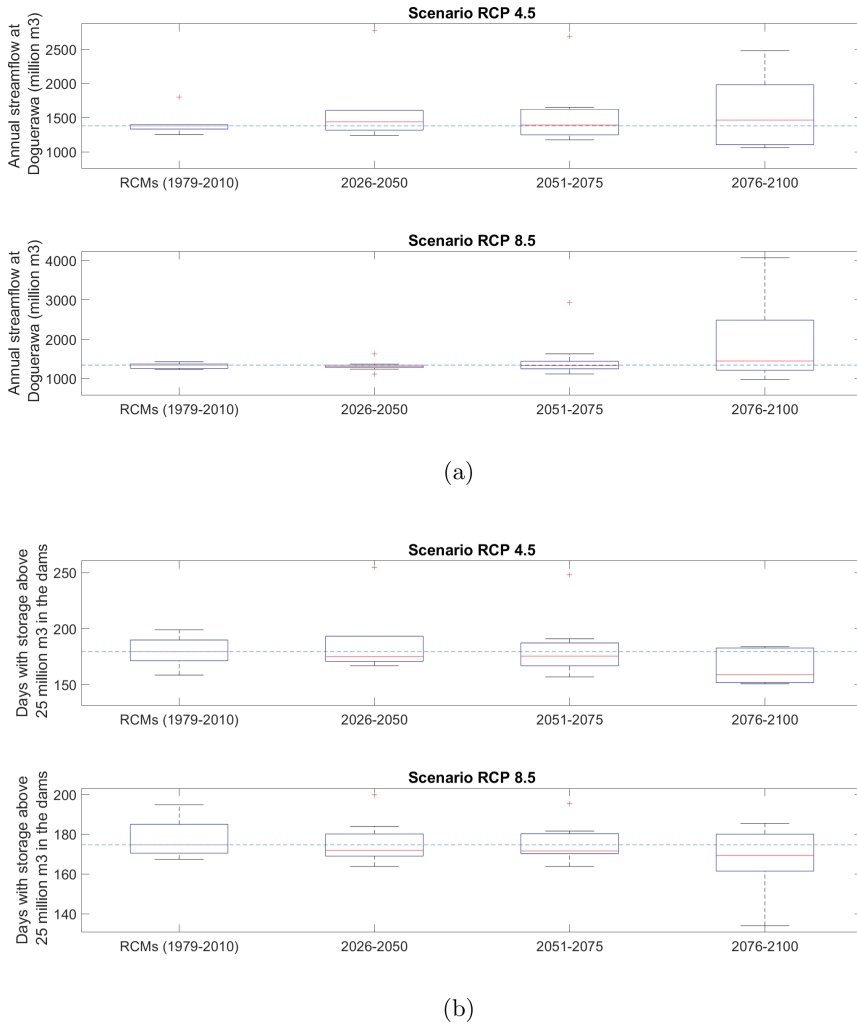


Fig. 6. Projected changes in (a) the discharge at the Doguerawa station and (b) the number of days per year the storage in the dams is above average under RCP4.5 and RCP8.5

(uncertainty) in the future. These results are consistent with the projections in precipitation in the area, and point to the possibility of more agricultural production. Another impact of climate change is the increase in evaporation because of the warmer climate. Evaporation losses is an important component of the water budget in the study area, and even in the current climate around 32 millions m³ of water are lost each year to evaporation. An increase in evaporation can hence offset the increase in rainfall. To check that hypothesis, the average numbers of days where the cumulative volume of water in all dams is above average was calculated under the current and future climates (Fig. 6b). As expected, the number

of days where the storage is above average decrease in the figure, suggesting that storing water in large dams may not be the solution for the future. The reason water losses are hard to master in large reservoir is that their shape is imposed by nature and they generally have a large surface to volume ratio. Evaporation control is much easier in smaller reservoirs because (a) they can be engineered so that their surface to volume ratio become small and (b) there are a number of evaporation control technologies at different phases of development that may be applicable to small reservoirs. That's why Sahelian farmers should consider alternative techniques for harvesting and storing water which are less sensitive to evaporation than large open water reservoirs. These techniques ranges from infiltration promotion to covering the water body with a plastic firm or a layer of chemical products. While these techniques are expensive nowadays, they are likely to become cheaper and more efficient in the coming decades.

4 Conclusions

Past and future trends in precipitation and runoff over the Niger republic were examined in this paper. Much as the rest of Sahel, the country has seen a drastic decrease in precipitation during the 70s and the 80s, followed by an apparent recovery. Surprisingly, the drought resulted in an increased surface runoff in some areas in the Sahel, suggesting that new opportunities for water harvesting may arise in a dry climate, if evaporation is adequately taken care of. The paper also examined the predictions of 10 different regional climate models over the Maggia watershed in the center south of the Niger republic, and found that runoff is likely to increase but that the increase will be offset by evaporation in large dams. In both cases, the use of decentralized storage systems and evaporation suppression techniques should be seriously considered as a way to maintain and increase water supply and agricultural production.

References

1. Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R.: Large area hydrologic modeling and assessment part I: model development. *J. Am. Water Resour. Assoc.* **34**(1), 73–89 (1998). <https://doi.org/10.1111/j.1752-1688.1998.tb05961.x>
2. Bartholome, E., Belward, A.S.: GLC2000: a new approach to global land cover mapping from earth observation data. *Int. J. Remote Sens.* **26**(9), 1959–1977 (2005). <https://doi.org/10.1080/01431160412331291297>
3. Descroix, L., et al.: Impact of drought and land - use changes on surface water quality and quantity: the Sahelian Paradox. In: Bradley, P. (ed.) *Current Perspectives in Contaminant Hydrology and Water Resources Sustainability*. InTech, February (2013). <https://doi.org/10.5772/54536>
4. FAO/UNESCO: Digital soil map of the world and derived soil properties, rev. 1 cd rom (2003). http://www.fao.org/catalog/what_new-e.htm
5. Giorgi, F., John, C., Asrar, G.: Addressing climate information needs at the regional level: the CORDEX framework. *World Meteorol. Org. Bull.* **58**(3), 175–183 (2009)

6. IPCC: Global Sectoral Aspects. Number Impacts, Adaptation, and Vulnerability; Part A in Climate Change. Cambridge University Press, New York (2014)
7. Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E.: Hole-filled SRTM for the globe version 4. Integrated Natural Ressources Planning and Management (2008). Available from the cgiar-csi srtm 90m database. <http://srtm.csi.cgiar.org>
8. Parry, M.L.: Climate Change 2007 - Impacts, Adaptation and Vulnerability: Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge (2007)
9. Servat, E., et al.: Identification, caractérisation et conséquences d'une variabilité hydrologique en Afrique de l'Ouest et Centrale. In: Water Resources Variability in Africa During the XXth Century. IAHS, Abidjan (1998)
10. Séguis, L., Cappelaere, B., Milési, G., Peugeot, C., Massuel, S., Favreau, G.: Simulated impacts of climate change and land-clearing on runoff from a small Sahelian catchment: climatic and anthropogenic effects on Sahelian runoff. *Hydrol. Process.* **18**(17), 3401–3413 (2004). <https://doi.org/10.1002/hyp.1503>
11. Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J., Viterbo, P.: The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-interim reanalysis data. *Water Resour. Res.* **50**(9), 7505–7514 (2014). <https://doi.org/10.1002/2014WR015638>



Evaluation and Update of Two Regional Methods (ORSTOM and CIEH) for Estimations of Flow Used in Structural Design in West Africa

Sehouevi M. D. Agoungbome^{1(✉)}, Ousmane Seidou², and Moussa Thiam²

¹ African Institute for Mathematical Sciences,
Mbour Km 2 Road of Joal, Mbour, Senegal
sehouevi.m.d.agoungbome@aims-senegal.org

² Department of Civil Engineering, University of Ottawa,
161 Louis Pasteur Office A113, Ottawa, Canada

Abstract. The ORSTOM and CIEH methods are the most popular methods used in West Africa to estimate the design flow, needed in the design of hydraulic structures. However, these methods, based on hydrological data collected before 1965 and 1983 for ORSTOM and CIEH respectively, showed some shortcomings and deviations in the estimation of the forecasts. This study, therefore, assesses the design flow on data collected from some watersheds in Benin and Niger using a frequency analysis. These estimations are therefore compared with estimations obtained using the ORSTOM and CIEH methods on the same watersheds. As result, we observe an important under-estimation of the flows with the ORSTOM method, whereas the CIEH method gives higher values. The use of the log linear regionalization method, considering the area of the watershed basin and the global slope index, gives very interesting results and good perspectives of estimation on few or no data watersheds basins.

Keywords: Design flow · Watershed · Log linear regionalization

1 Introduction

As any other area in the world, the prosperity of Africa relies on the quality of its infrastructures. Unfortunately, it is very common to witness dams, bridges and communities being destroyed by floods. While flood damages are expected to happen everywhere, the high frequency of destruction points to inadequate design flow calculations. The design flow or project flow is the expected flow resulting from the worst combination of meteorological and hydrological conditions considered to be reasonably characteristics of the concerned region (aquaportail.com). The ORSTOM and CIEH methods are the most popular design flow estimation methods in West and Central Africa. These are regional

methods (not directly based on observed flow on the site) because of the low data availability in most African regions. The ORSTM method was developed using meteorological and hydrological data recorded up to 1965 while the CIEH was developed using meteorological and hydrological data recorded up to 1983 [1]. But it is clear that the climate in Africa and particularly the western part, has shifted since then: a decrease of 15–35% has been reported for rainfall; a 200 km shifted of isohyets to the south has been observed; the streamflows in most watersheds experienced a decrease of 40–60% [2]. Olivry et al. [3] indicated that surface water resources in rivers have declined since the beginning of 1970 and have halved in the 1980s. Therefore, the application of the CIEH and ORSTM methods in their original proposed form may not reflect the current level of hydrological risk. They should be reassessed and updated in order to prevent damage or financial losses that may result from an over-sizing or under-sizing of hydraulic structures.

The aim of this work is to reassess these two methods and propose an alternative more accurate method that can be used to estimate the design flow.

2 ORSTM Method

Also called Rodier-Auvray method, it is a deterministic method of flow estimation proposed in 1965 by the ORSTM's hydrologists ("Office de la Recherche Scientifique et Technique Outre-Mer"). This institution was later replaced with the "Institut Régional pour le Développement" (IRD). The project flow is defined as a flood caused by a rainfall that is equalized or exceeded in average once a decade for a watershed in the Sahelian and tropical dry lands [4]. This method, based on 65 watersheds with surface areas between 10 and 120 km², is applied in theory to all the West African's basins, between 150 and 1600 mm in annual rainfall and in the range of some hectares to 1500 km² surface area. The study was able to identify the main factors that explain the floods. Those are the height of the generating rainfall, the area of the watershed, the infiltrability of the soil and the relief.

The peak flow of the ten-year flow (m³/s) is given by the relation:

$$Qr_{10} = m.A.P_{10}.Kr_{10}.\alpha_{10}.S/Tb_{10} \quad (1)$$

where A is the abatement coefficient, m: the coefficient of increasing, P_{10} : the ten-year daily rainfall (m), Kr_{10} : the runoff coefficient corresponding to the ten-year flow (%), α_{10} : the peak coefficient corresponding to the ten-year flow, S: the area of basin (m²), Tb_{10} : the base time or time of the flood (s).

The ORSTM method uses several abacus, tables and graphs with fixed values to estimate those parameters. Estimating each term of this expression independently of each other gives less accuracy on the overall quality of the estimate such that the errors made on each parameter appear as a multiplier of the error on Q_{10} [5].

3 CIEH Method

Proposed by Puech and Chabi-Gonni [5], the CIEH (“Comite Interfricain d’Etudes Hydrauliques”) method is based on a statistical multi-regression analysis of the data collected in the collection of Dubreil [6] entitled “Recueil des données de base des bassins representatives et expérimentaux” supplemented by information from the member countries of the CIEH. This study was carried out on a set of 162 small and medium-sized catchments located in 14 West and Central Africa countries [5] and highlighted the main factors which can be quantified without ambiguity.

- Physical factors: area, length, slope, compactness of the watershed;
- Climatic factors: annual and daily precipitation, type of climate, temperatures;
- Soil factors: runoff that depends on pedology, vegetation cover, soil moisture.

The ten-year’s peak flow Q_{10} (m³/s) based on a multiple regression scheme is given by the relation:

$$Q_{10} = a.S^s.Pan^p.Ig^i.Kr_{10}^k.Dd^d$$

(2)

with a, s, p, k, d some coefficients to be determined, Pan: annual average precipitation (m), Kr_{10} : the runoff coefficient (%), Ig: the global slope index (m/km), S: the area of basin (km²), Dd: the drainage density (km⁻¹).

The parameters that have to be included in the model are not limiting, and it may take into account more parameters.

However, recent studies (1994) was carried out using multiple regressions to express more accurately the CIEH formula by defining the most interesting and important factors to consider, by regional climatic groupings and according to the quality (or reliability) criteria. As a result, here is a partial correlation matrix (Table 1) for the following parameters S, L, Ig, Pan, P_{10} , Pm_{10} , Dd, Kr_{10} and Q_{10} .

Table 1. Partial correlation matrix [4]

	Q_{10}	S	L	Icomp	Dd	Ig	Pan	P_{10}	Kr_{10}
Q_{10}	1								
S	0.623	1							
L	0.654	0.963	1						
Icomp	0.151	0.157	0.365	1					
Dd	−0.079	−0.459	−0.383	0.155	1				
Ig	−0.309	−0.693	−0.639	0.076	0.537	1			
Pan	−0.264	−0.103	−0.078	0.098	0.058	0.341	1		
P_{10}	−0.161	0.032	0.047	0.090	−0.112	0.170	0.812	1	
Kr_{10}	0.472	−0.044	−0.041	−0.064	0.064	−0.078	−0.271	−0.168	1

For the purpose of this study, the following parameters with respect to their correlation coefficient r were selected to estimate the project flow:

- The area S ($r = 0.623$);
- The global slope index I_g ($r = -0.309$);
- The precipitations P_{an} and P_{10} The runoff coefficient;
- Kr_{10} ($r = 0.472$) (which has to be taken into account for better estimates when P_{an} is greater than 1000 mm in order to reach significant levels).

Considering these parameters, the following mathematical formulation emerges:

$$Q_{10} = a.S^s.I_g^i. \quad (3)$$

Like the ORSTOM method, the CIEH method requires the estimation of certain parameters in situ. The accuracy of the final result depends, in a large extent, on the precision and accuracy of those parameters. Moreover, these formulations do not take into account the internal heterogeneities of the basin.

4 Frequency Analysis

The statistical treatment of the data in order to estimate the ten-year flow rate was carried out using a frequency analysis which consists in studying past events characteristic of a given process (hydrological or other) in order to define the probability of a future occurrence.

4.1 Return Period

In West Africa, ORSTOM hydrologists based on the relatively small watershed in the tropical and Sahelian regions and available measurements, defined the ten-year return period. It is a statistical estimator which may be observed once in ten years [4].

$$T = \frac{1}{1 - F(x)} \quad (4)$$

with $1 - F(x)$: the probability of overpassing of the event.

4.2 Choice of the Flood Distribution Model

Several models (normal distribution, log-normal distribution, Gumbel distribution, etc.) are used to describe flood phenomena, but the choice is based on some criteria such as the theoretical conditions of the distribution, the asymptotic behavior, some suitability tests, and the use of various diagrams. The most used distribution in hydrology to describe extreme events, such as rainfall and in our case floods, is the Gumbel distribution [7]. It is a double exponential distribution, the limiting form of the distribution of the maximum value of a sample of n values. The Gumbel distribution's cumulative density function is defined as:

$$F(x) = \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right] \quad (5)$$

where $a \in \mathbb{R}$ is the location parameter, b : the scale parameter, x : the describable variable.

The quantile function which is the inverse cumulative distribution x_q is given by:

$$x_q = a - b \ln[-\ln[F(x)]] \quad (6)$$

The Gumbel distribution allows describing the annual maxima of the flow rate on the considered basin.

4.3 Adjustment of the Model

The adjustment of the model to the data set consists of finding the parameters a and b of the above-mentioned distribution. For this purpose, we define a reduced variable u :

$$u = \frac{x - a}{b}. \quad (7)$$

The adjustment or determination of the parameters can be done by the graphical method or the method of the moments.

4.3.1 Graphical Method

In the case of the adjustment according to Gumbel distribution, the graphical method based on the linearity of the expression of a quantile corresponds to a straight line equation ($x = f(u)$). The parameters a and b are then obtained by estimating the coefficients of the regression line (Fig. 1) which fits the most to the points. But, to find the variable u , some expressions are proposed in order to compute an experimental value of $F(x)$. Simulations have shown for the Gumbel distribution that the appropriate expression is Hazen formula which is defined as:

$$F(x_r) = \frac{r - 0.5}{n} \quad (8)$$

where r is the rank of the value in the sample; n is the sample size; x is the describe variable.

The data is decreasingly ordered. Each value is assigned its rank and we use Eq. 8 to compute $F(x)$. Then, we deduce the reduce variable u using the Eq. 6:

$$u = -\ln[-\ln[F(x_r)]] \quad (9)$$

Therefore we plot on the same graph the flow x in terms of u and the regression line.

The advantage of this method is that it provides a visual representation and the adjustment of the data, essential aid for the judgment of the adequacy of the chosen distribution and the data processing. Once we have the parameters a and b , we equate the Gumbel formula with the expression of $F(x)$ deduced from the return period formula. This gives us:

$$F(x) = 1 - \frac{1}{T} \text{ and } u = -\ln[-\ln[\frac{1}{T}]] \quad (10)$$

Hence, we easily compute the ten-year project flow using the Eq. 6.

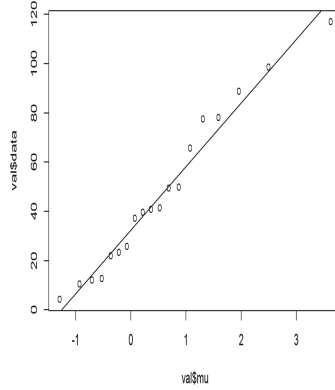


Fig. 1. Regression line. (Aguimo watershed)

4.3.2 Method of Moments

The method of moments consists of equating the sampling moments and the theoretical moments of the chosen distribution.

Let x_1, \dots, x_n be the sample available and let $\hat{\mu}$ the standard estimators of the mean and $\hat{\sigma}^2$ the variance of the sample. The first two theoretical moments of the Gumbel distribution are expressed in terms of the location and scale parameters as follows:

$$\hat{\mu} = a + b\gamma \text{ and } \hat{\sigma}^2 = \frac{\pi^2 \cdot b^2}{6} \quad (11)$$

where

$$\gamma = 0.5772 \text{ (Euler constant).}$$

Therefore, we easily deduce a and b:

$$b = 2.45 \frac{\hat{\sigma}}{\pi} \text{ and } a = \hat{\mu} - b\gamma \quad (12)$$

So with a and b, we compute the ten-year project flow using the Eq. 6.

5 Simulations and Results

5.1 Target Watersheds of the Study

All the data, we used in this research has been collected from the database of AMMA-CATCH. It is a hydro-meteorological observatory service, whose aim is to document the long-term climate, hydrological and ecological changes in West Africa.

Seven watersheds are located in the north of Benin between the positions 9.954 N; 1.819 W; 2.399 E; 9.7106 S. They are parts of the big coastal watershed basin of Oueme-Yewa (Benin) [8].

Three basins belong to the Niger watershed located in Niger between the positions: 13.884 N; 2.63 W; 2.7001 E; 13.6445 S. The areas of the watershed vary from 0.048 to 0.16 km².

5.2 Result Using ORSTOM and CIEH Methods

All the simulations were performed using the software R Studio. For the ORSTOM method, the obtained results are in the range of [0.67–63.24 m³/s]. The smaller design flows are obtained in the region of Niger, with the smallest one in Wankama-amont watershed. The relatively bigger ten-year flow are in Benin region with the biggest ten-year flow obtained in Sani-a-sani basin.

The results obtained using the CIEH method, are in the range of [3.3–101.1524 m³/s]. The smaller design flows are obtained in the region of Niger, with the smallest one in Wankama-amont watershed. The relatively big ten-year flows are in Benin region with the biggest project flow obtained in Sani-a-sani basin.

5.3 Flood Peak Estimation Assuming a Gumbel Distribution

A plot of the histogram and the density of the data are presented in Fig. 2. The choice of Gumbel distribution is based on the samples distribution of our data which is left skewed with skewness 0.56 and platykurtic with kurtosis -0.88. It belongs to the Gamma distribution’s function.

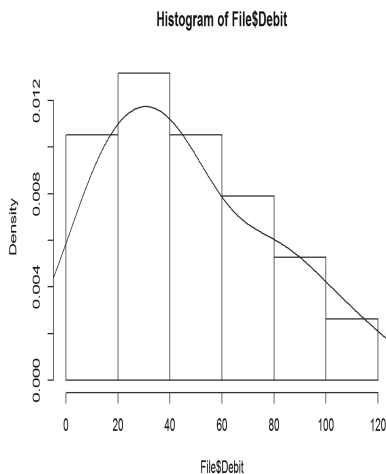


Fig. 2. Sample distribution. (Aguiamo watershed)

The available data varies from a watershed to another. It varies from 4 years to 19 years according to the accessible data we got from AMMA-CATH DB. The

results obtained with this statistical analysis are in the range of $[0.3\text{--}122.5\text{ m}^3/\text{s}]$. The biggest value is obtained on Donga-pont watershed whereas the smallest one is obtained on Tondikiboro-amont.

5.4 Discussion

The following table summarises all the results obtained from the frequency analysis (FA), ORSTOM and CIEH methods (Table 2).

Table 2. Summary of simulations

Location	Station	Area (km ²)	FA (m ³ /s)	ORSTOM (m ³ /s)	CIEH (m ³ /s)
Benin	Aguimo	368.77	90.6	45	84.2
Benin	Sani-a-sani	743.3	43.8	63.24	101.15
Benin	Tebou	529.86	87.86	53.94	95.28
Benin	Wewe	276.55	64.7	39.91	77.45
Benin	Ara-pont	10.93	12.8	12.31	21.75
Benin	Donga-pont	586.95	122.5	57.02	96.85
Benin	Donga-route-kolokonde	101.783	41.6	39.95	56.2
Niger	Tondikiboro-amont	0.16	0.3	1.3	4.2
Niger	Wankama-amont	0.048	1.3	0.67	3.3
Niger	Wankama-ZeAmont	0.062	1.6	0.87	3.7

As shown, we clearly see that for almost all the stations, the ORSTOM method estimations are less than what we could expect using the frequency analysis. We noticed some differences in the range of $[0.14\text{--}20]$. The CIEH method as the ORSTOM, doesn't give better estimation of the project flow. We notice for almost all the stations an increasing of the flow Q_{10} .

The smallest mean square error is observed for the CIEH method (21.26) whereas the mean square error, using ORSTOM is a bit bigger (29.18). Then we clearly see that we make less error with the CIEH method. So, as projected, ORSTOM and CIEH methods based on the data we have, and the analysis we made on it, are both no more accurate like in the time there were built. The ORSTOM method under-estimates the design flow whereas the CIEH method over-estimates the design flow. This result confirms the same observation made on some watersheds basins in Burkina Faso [1].

Therefore, the need of their updating is urgent in order to avoid problems as describe in the introduction caused by an under-estimation or over-estimation of the design flow. At this stage, it is important to find an alternative method.

6 Alternative Method

Adequate estimation of extreme hydrological variables is essential for the rational design and operation of a variety of hydraulic structures, due to the significant

risk associated with these activities. Local frequency analysis is commonly used to estimate extreme hydrological events on sites where adequate amount of data is available. In practice, it frequently happens that little or no stream flow data is available on a site of interest (where a dam is going to be erected for example). In such cases, hydrologists use a regional flood frequency procedure, relying on data available from other basin with a similar hydrological regime or region [9].

According to the previous work done by CIEH and ORSTOM hydrologists, and based on the partial correlation matrix (Table 1), they revealed that the ten-year flow is a parameter which highly depends on the area of the watershed and the global slope index and in some extent on the precipitation and the runoff coefficient. Thus, we will use a regression approach which will take into account those four parameters. The regression method is simple, fast and allows using different distributions for the different sites of the region. It is also non-sensitive to the heterogeneity that may exist in the region [9,10]. Then, a logarithmic transformation of the variables is introduced in order to linearize the relation of the power type. Therefore, the equation below is deduced:

$$Log(Q) = a_1.log(S) + a_2.log(Ig) + a_3.log(Pm_{10}) + a_4.log(Kr_{10})$$

(13)

where S is the area of the watershed (km²), Ig: the global slope index (m/km), P₁₀: the average ten-year daily precipitation (mm), Kr₁₀: the runoff coefficient (%).

The model doesn't give better results. The mean square error obtained with this estimation is 16.87. Then we compute the correlation matrix which shows that the more expressive parameters of the design flow are the area and the global slope index which confirm the previous updating work done by CIEH and ORSTOM hydrologists in 1994.

Hence, we divide the dataset into two, according to the two regions (Benin and Niger). Therefore, we obtain the best results; with a mean square error of 0.96. The result is reported in Table 3.

Table 3. Log linear model results

N	Station	Flow (m ³ /s)	New flow (m ³ /s)
1	Aguimo	90.6	90.09
2	Tebou	87.86	89.69
3	Wewe	64.7	63.65
4	Ara-pont	12.8	12.97
5	Donga-pont	122.5	123.39
6	Donga-route-kolokonde	41.6	41.63
7	Tondikiboro-amont	0.3	1.36
8	Wankama-amont	1.3	0.97
9	Wankama-ZeAmont	1.6	0.37

7 Conclusion

This work aims to re-evaluate the two most popular methods of design flow estimation in West Africa (ORSTOM and CIEH methods). Ten watershed basins have been considered, seven in Benin and three in Niger. The frequency analysis using Gumbel distribution gives the observable ten-year design flow considered as the true values which were compared to the ten-years flow estimated with the ORSTOM and CIEH methods. The error made using these former methods in terms of mean square error, is respectively 21.26 and 29.18 for the CIEH and ORSTOM method.

Then, it appears obviously that these two methods are no more accurate in project flow estimation.

An alternative method for updating was proposed based on the regional log linear model. The model takes as parameters the watershed area and the global slope index. It gives better estimation when we identify and divide the dataset into two different regions, with 0.96 mean square error.

References

1. Bambara, T.: Contribution à la révision des normes hydrologiques au Burkina Faso, 2iE (2011)
2. Mahé, G.: Variabilité pluie-débit en Afrique de l'Ouest et Centrale au 20ème siècle, Changements hydro-climatiques, occupation du sol et modélisation hydrologique. Université des Sciences et Techniques Montpellier 2 (2006)
3. Olivry, J.C., et al.: Vers un appauvrissement durable des ressources en eau de l'Afrique humide. In: *Hydrology of Warm Humid Regions*, vol. 216, pp. 67–78, July 1993
4. FAO: Crues et apports: manuel pour l'estimation des crues décennale et des apports annuels pour les petits bassins versants non jaugés de l'Afrique Sahélienne et tropicale sèche. Bulletin FAO et de Drainage, 54 (1996)
5. Puech, C., Chabi-Gonni, D.: Méthode de calcul des débits de crue décennale pour les petits et moyens bassins versants en Afrique de l'Ouest et Centrale. Bulletin CIEH (1983)
6. Dubreuil, P., Chaperon, P., Guiscafré, J., Herbaud, J.: Recueil des données de base des bassins représentatives et expérimentaux. ORSTOM (1972)
7. Musy, A.: Analyse fréquentielle École polytechnique fédérale de Lausanne (2005)
8. FAO: Profil du bassin versant de l'ouémé et caractérisation des sites pilotes (analyse des données). Rapport FinalFAO (2009)
9. Ouarda, T.B.M., et al.: A review of recent developments in regional frequency analysis of hydrological extremes. *Revue des sciences de l'eau* **21**(2), 219–232 (2008)
10. Martel, B.: Comparaison de méthodes d'analyses fréquentielle régionale appliquées aux crues automnales du Québec. Université du Québec (2009)



Statistical Downscaling of Global Climate Model MIROC_4h Outputs to Precipitation in Rwanda

Nkusi Pearl and Mutabazi Alphonse^(✉)

Climate Change Program, Single Project Implementation Unit,
Rwanda Environment Management Authority, KG 7AVE, Kigali, Rwanda
nkusipearl@gmail.com, mutalpo@hotmail.com

Abstract. Statistically downscaled models for precipitation based on multiple linear regression were developed and validated in the four different climatological zones of Rwanda for period, 1961–1990. The validation of these monthly models was majorly based on the ability of the model to reproduce trends of rainfall for an independent period. The downscaled simulations suggest that there will generally be increases in precipitation for majority of the months and increases of over 15% in the known onset months of rain seasons. The dry months are also expected to experience slight variations for the Eastern zone.

Keywords: Statistical downscaling · MIROC_4h · Precipitation
Rwanda

1 Status of Rwanda's Climate

Rwanda is a country in Tropical Eastern Africa whose economy depends mainly on agriculture. It is famously known as ‘the land of a thousand hills’. Due to landscape terrain with high slopes and high population density, Rwanda is highly exposed to climate change effect and. There is, therefore, need for availing accurate climate-related information to stakeholders throughout the country mainly for long-term planning purpose. Over the recent pasts, Rwanda has experienced extreme events in various parts of countries such as the destructive floods that affected the entire country in May 2012, and the recent Gakenke flood in May 2016.

In Rwanda, there is a high likelihood that rainfall quantity will increase by the end of 21st century. However, model predictions are averages for long periods; daily, monthly and annual variability are very uncertain. While this rainfall increase is predicted to be between 10 and 20% of observed mean rainfall in 1961–1990, there is no indication whether the temporal rainfall distribution will enough to meet future water requirements (Rwanda Country Situational Analysis 2011).

1.1 Problem Statement

Precipitation in Rwanda and East Africa as a whole is a highly variable climate parameter in space and time; as recently studied in Ongoma and Chen (2017). Due to

this reason, precipitation forecasts and projections require to be at high spatial resolution in order to improve their accuracy. Stake holders in government institutions/agencies and various sectors of the economy also require precise climate related information for short and long periods into the future. Therefore, there is still need for information on climate projections across the country to facilitate national and district-level strategic development plans. The readily available Global Climate Model simulations have very coarse spatial resolution and do not account for local climate aspects. To address this gap in available climate information, statistical downscaling is performed in this study.

1.2 General Objective of Study

To generate downscaled climate projections for the various climatological zones of Rwanda and provide future climate scenarios information on precipitation.

2 Materials and Methods

The observational data (predictand) was monthly precipitation for reference climatological stations of Rwanda; obtained from the National Meteorological Service, METEO Rwanda. This is stationed data collected over time periods, 1961–2014. The stations were selected to represent homogeneous climatological rainfall zones of Rwanda (Prioul and Sirven 1980) The model output data (predictors) was obtained from the data portal of Lawrence Livermore National Laboratory, Department of Energy USA.

MIROC4h, was chosen from WCRP's Coupled Model Intercomparison Project Phase 5 (CMIP5). The model output prepared for CMIP5 historical was used for the 1961–1990 model training period while those prepared for CMIP5 RCP 4.5 were used for development of downscaled projections (2015–2035). The validation period, 1991–2014 made use of both sets of data.

MIROC_4h was chosen based on its high spatial resolution; considering Rwanda's size of 26,338 km². High spatial resolution of $(1.0 \times 1.0)^\circ$ or higher allows for more grid points represented for Rwanda. These models present predictor data for 22 vertical levels for the representation of the atmosphere. 4 vertical levels were selected for this study. MIROC_4h simulations for the period 2015–2035 are obtained from esgf.llnl.gov data portal. Representative Concentration Pathways (RCP) scenario that is selected, appropriate for Rwanda is RCP 4.5. The MIROC_4h model has a high spatial resolution of $(0.56^\circ \times 0.56^\circ)$ and the output data sets from which data was extracted for the training period were 1850–2005 (historical) and 2006–2035 (simulations).

In this study, the predictand is monthly rainfall totals from four reference climatological stations for each climatological region of Rwanda. According to Prioul and Sirven (1980), there are four climatologically homogenous zones in Rwanda (Fig. 1).

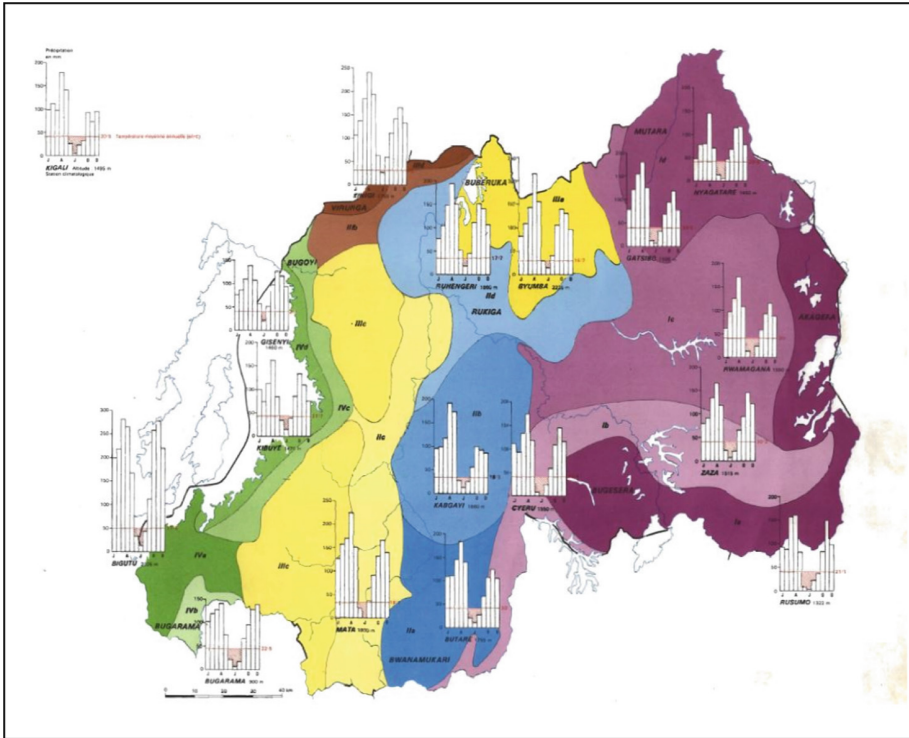


Fig. 1. Map of Rwanda showing climatic regions by colour (Prioul and Sirven 1980)

In this study, we have selected one station for each climatological homogenous zone as indicated below

1. Kanombe station: Warm and dry Eastern Low land
2. Rubona Station: Central Plateau
3. Byumba station: Montane climate
4. Kamembe station: Lake Kivu shores' climate.

There were four main steps taken for this downscaling technique for each reference climatological station:

1. Correlation of predictand (precipitation) and predictors (model output values) to select four best potential precipitation predictors. The empirical relationships are formed based on highest correlation coefficients between potential predictors and predictand
2. stepwise regression to select the best three predictors to yield empirical statistical models

3. Evaluation of statistical models using observations of predictand (precipitation) values for an independent data period
4. Future projections of precipitation.

MIROC_4h historical run, r1i1p1 was used for the period, 1961–1990 and 13 potential rainfall predictors were selected.

They were assigned variable names as follows:

Precipitation P, northward wind at surface Vs, eastward wind at surface Us, temperature T at 850 hPa, temperature T at 700 hPa, eastward wind U at 850 hPa, eastward wind U at 700 hPa, northward wind V at 850 hPa, northward wind at V 700 hPa, relative humidity H at 850 hPa, relative humidity H at 700 hPa, geopotential height Z at 850 hPa and, geopotential height Z at 700 hPa.

Various statistical equations were used to conceive a computer program that finds the best possible correlation coefficients, among the selected predictors that represents the physical relationship between the predictors and predictand. According to Sachindra et al. (2014) and Osman and Mawada (2016), selection/screening of potential predictors is the most important step in statistical downscaling. Four main potential predictors were selected from model outputs of MIROC4h to be run by the program and considered appropriate to characteristics of Rwanda's climate:

Wind Velocity, W at 850, 700 mb levels and surface; Wind direction D at 850, 700 mb levels and surface; Relative Humidity H at 850 and 700 mb levels; Air temperature T at 850 and 700 mb levels.

The generated model equations are then assessed against observational values of precipitation and temperature for an independent time period, 1991 to 2016. The main objective of this process is to determine to what percentage of accuracy, the models developed can reproduce the precipitation values observed in this period. The validation process shows the potential of the statistical model developed to make future projections of precipitation. Both graphical and statistical methods are used to validate the models' performance (Sachindra et al. 2014). The methods that are used in the study include: Model evaluation using graphical representation, Model evaluation and assessment of accuracy of empirical statistical models using contingency tables, development of terciles for both predicted and observed climate parameters as per each climatological region. The developed, evaluated and verified statistical models are used for simulations to generate future projections over Rwanda from 2015 to 2035 for precipitation.

3 Results and Discussion

3.1 Screening of Predictors and Development of Statistical Models

The following predictors were selected using step wise regression and the models developed by multiple linear regression techniques. Example of coding used to name variables: *HJY4* where

H: Humidity;
 J: MIROC_4h (Japan);
 Y: Grid Point Y;
 4: 4th Vertical level (700 hPa).

A FORTRAN program that performs multiple linear regression equivalent to statistical and graphical analysis software (SYSTAT). However, the advantage of the program is its ability to run correlation and multiple linear regressions on hundreds of predictors and is limitless on the number of predictors to include in the model equation. The predictor with the highest frequency of selection included humidity predictors at 700 hPa, and wind direction at the surface. Below is a table of selected predictors for each station (Table 1).

Table 1. Selected Predictors for each month for the period of 1961 to 1990 at each reference station

	Selected precipitation predictor per month											
	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
KIGALI	HJY4	HJY4	HJY4	HJU4	HJV4	HJV4	TJY4	HJU4	HJX4	HJY4	HJA4	HJY4
	DJA4	DJD0	DJG4	DJX3	WJK0	WJU0	TJN3	DJK0	DJW4	TJN3	WJJ0	DJA0
KAMEMBE	HJU4	HJY4	DJO4	HJY4	HJY4	HJO4	HJY4	HJY4	HJY4	HJY4	DJR0	HJX4
	DJB0	DJG0	DJR4	HJQ4	WJN0	HJJ3	WJT3	DJU4	DJO0	HJU4	DJY3	TJK4
RUBONA	HJU4	HJY4	HJL4	HJG4	HJY4	HJY4	TJY4	HJX4	HJY4	HJY4	HJD4	HJO4
	DJD3	TJK4	DJG0	DJY0	HJA4	DJO3	TJN3	DJX4	DJG0	HJS3	DJS0	WJP4
BYUMBA	HJW4	HJV4	HJY4	HJY4	HJY4	HJY4	HJY4	HJE4	HJY4	HJS4	HJX4	HJW4
	HJH4	WJN0	DJS0	TJN3	DJK4	TJA4	TJN3	DJU0	DJX0	DJT3	DJI3	TJN3

3.2 Statistical Model Evaluation and Development of Future Projections

The MIROC4h model projections that are downscaled in this study depict variability in precipitation anomalies for the next two decades (upto 2035). In reference to similar studies on East Africa Shongwe et al. (2010), the trends of precipitation changes tend to agree especially with the general trend across the known MAM and SON rainy seasons. In reference to the year, 2015, the downscaled projections of MIROC_4h suggest that by 2035, there will be an increasing trend in precipitation during majority of the months apart from January, March, July and November for Kanombe station. Validation was based on the model's ability to graphically reproduce trends of rainfall for the validation period, 1991–2014.

KIGALI:

There will generally, be wetter April-May months of the long rain season and a slight increase in precipitation changes over the dry December-January-February. They also suggest that there will be a general increase in precipitation anomalies for months January-February, April-May, July-August and September-October-December. March, June and November show a decreasing trend (Fig. 2).

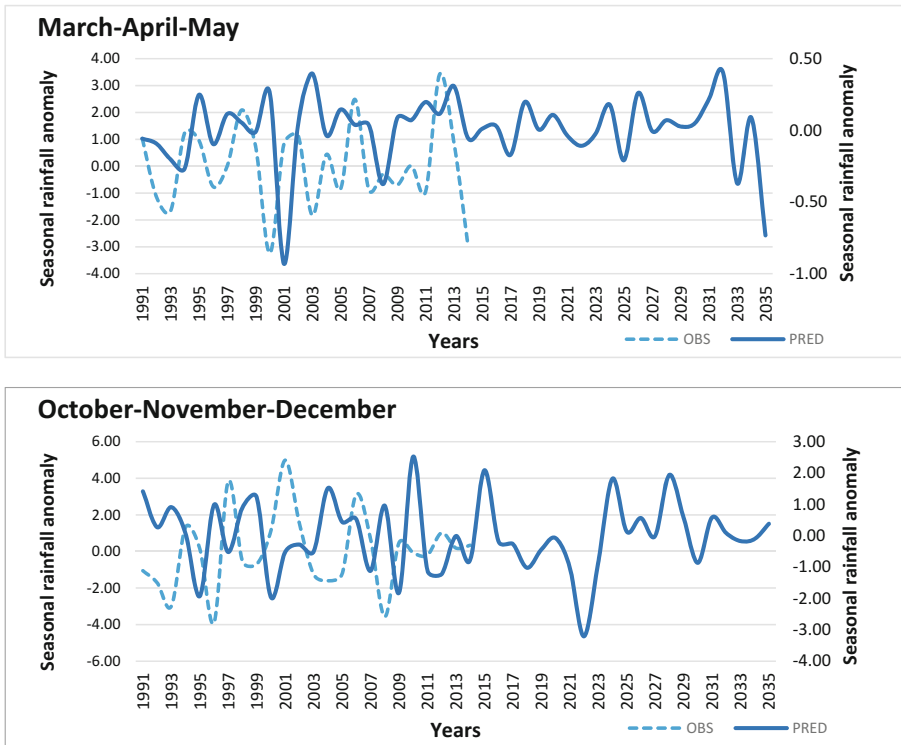
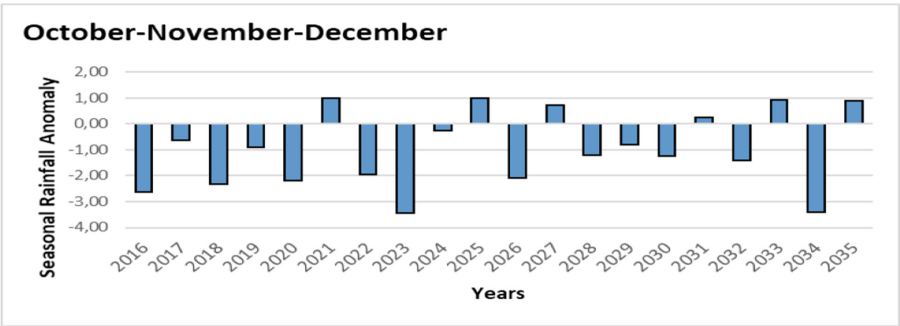
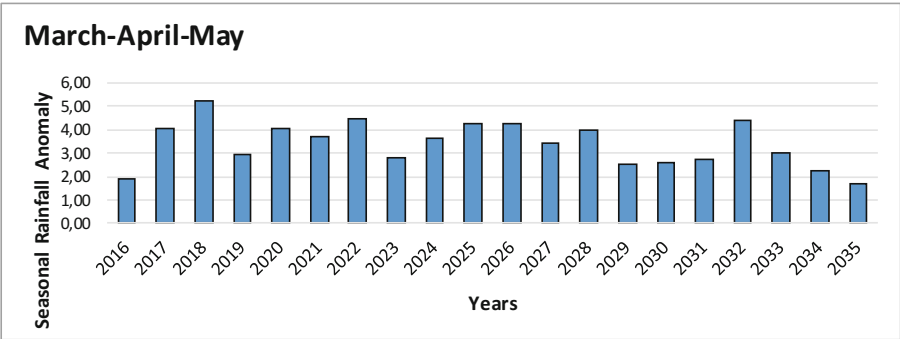


Fig. 2. Seasonal validation (1991-2014) and projections of precipitation (standardized anomalies) from MIROC_4h; Kanombe station (2015 to 2035)

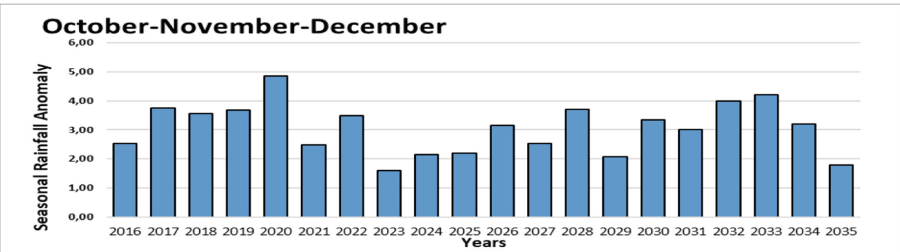
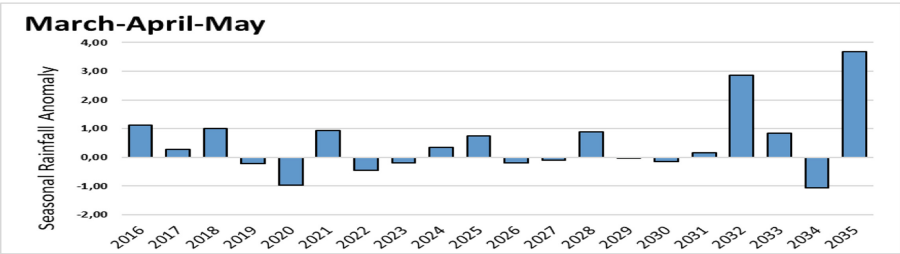
The trends depict a generally wet October-November-December season and a slight increase in precipitation changes over the usually dry January-February and July-August period; showing consistence with projections of mean and extreme precipitation in Africa region under global warming (Shongwe et al. 2010).

Trends in precipitation changes are relatively in agreement with projections of future climate over the Albertine Rift Valley region upto the year 2030 under the SRES A2 scenario. Seimon and Phillipps (2011) reported an approximate increase in precipitation of about 3–9 mm in Rwanda (as part of the Albertine rift) for the months; January, February, May, November and December in the year 2030, relative to 1990. Downscaled MIROC_4h projections from this study suggest an average decrease ranging from 10–13 mm within the Eastern dry climatic region.

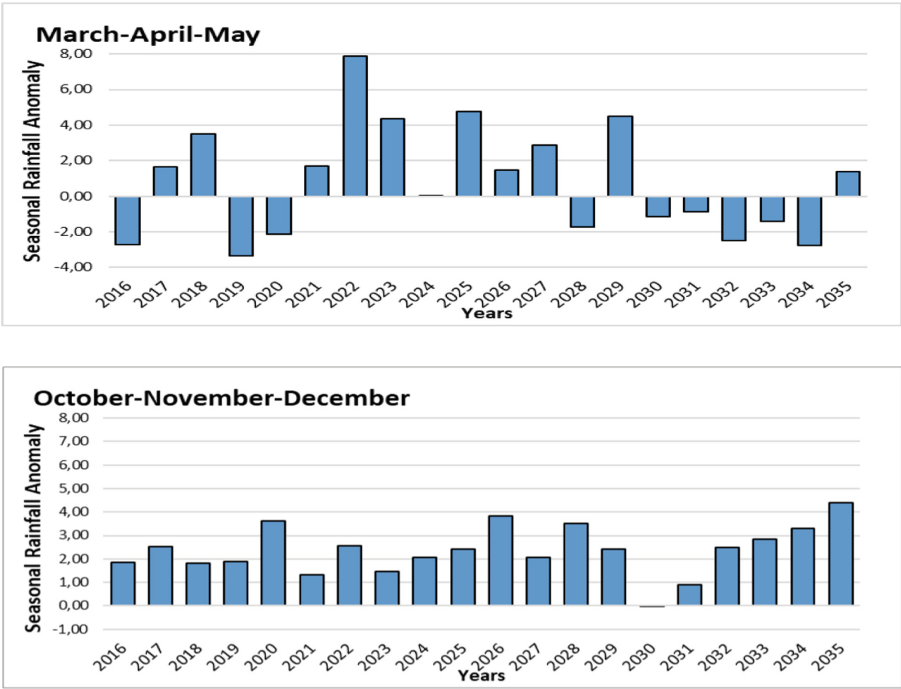
BYUMBA:



KAMEMBE:



RUBONA:



Throughout the country, the reference stations used in this study show relatively varying future changes in precipitation. An average increase of 14%, 12% and 35% increase is expected for stations Kanombe, Rubona and Kamembe. The long dry months of June-August will experience a slight increase in precipitation in Kigali and a decrease for other stations (Table 2).

Table 2. Percentage change of monthly projected precipitation changes by 2035 with reference to 2015

Stations				
Months	KIGALI	RUBONA	BYUMBA	KAMEMBE
January	-17.0	1.0	19.0	0.25
February	18.0	-22.0	-3.0	-1.0
March	-9.0	-3.0	14.0	-49.0
April	3.0	-63.0	-24.0	-8.0
May	0.05	-1.0	-2.0	1.0
June	-13.0	-29.0	-11.2	-1.2
July	-42.0	-18.0	-6.0	-2.0
August	56.0	-6.0	75.0	36.0
September	23.0	2.0	-11.0	4.0
October	2.0	-4.0	13.0	-14.0
November	-1.0	-1.0	8.0	13.0
December	18.0	15.0	96.0	-42.0

4 Conclusion and Recommendations

From the simulations now tailored to local climate characteristics of Rwanda using observational data, we expect that there will be a general increase in monthly totals of precipitation across the country and significant increases in the onset months of rainfall seasons, MAM and SON. Downscaling of GCM simulations before use in impact assessment studies, or their consideration in national strategies and plans is highly recommended. More GCMs need to be used by climate researchers in Rwanda to increase on the base of information on future climate scenarios.

References

- Prioul, Ch., Sirven, P.: Atlas du Rwanda (1980)
- Sachindra, D.A., Huang, F., Barton, A., Perera, B.J.C.: Statistical downscaling of general circulation model outputs to precipitation – Part 2: bias-correction and future. *Int. J. Climatol. Banner* **34**, 3282–3303 (2014). <https://doi.org/10.1002/joc.3915>
- Seimon, A., Philipps, G.: Regional climatology of the Albertine Rift in “Long Term Changes in Africa”s Rift valley: impacts on biodiversity and ecosystems (2011). https://www.researchgate.net/profile/Michael_Wilson3/publication/216460597_Longterm_changes_in_the_ecological_factors_surrounding_the_chimpanzees_of_Gombe_National_Park/links/02e7e51ded180402e1000000/Long-term-changes-in-the-ecological-factors-surrounding-the-chimpanzees-of-Gombe-National-Park.pdf
- Shongwe, M.E., Van Oldenborgh, G.J., Van Den Hurk, B.: Projected changes in mean and extreme precipitation in Africa under global warming, Part II: East Africa, p. 40 (2010)
- Ongoma, V., Chen, H.: Temporal and spatial variability of temperature and precipitation over East Africa from 1951 to 2010. *Meteorol. Atmos. Phys.* **129**, 131 (2017). <https://doi.org/10.1007/s00703-016-0462-0>
- Rwanda Country Situational Analysis: Compiled by Alphonse Mutabazi Climate Change Consultant for Camco, Nairobi, Kenya, May 2011. http://rema.gov.rw/rema_doc/Climate%20change/Situational%20Analysis%20-%20RWANDA%20Final.pdf

CNRIA Workshop



An Internet of Things Infrastructure for Rainfall Monitoring in Dakar

Abdoulaye Kama^(✉), Cheikh Modou Noreyni Fall, Mamadou Simina Drame,
and Moussa Diallo

Université Cheikh Anta DIOP (UCAD), Dakar, Senegal
`abdoulaye.kama@ucad.edu.sn`

Abstract. Rainfall is a very important climatic phenomenon for the Sahelian economies. In Senegal, it has a very tremendous impact on agriculture and human life, which justifies the need for effective monitoring systems. In fact, the country has been experienced number of extreme events such as floods and public health problems. Despite the efforts made, it remains challenging to have real-time observations, which impact directly forecasts quality. In addition, systems used are not efficient and they are often expensive for implementation and deployment throughout the country. In this paper, we present an automatic rainfall station adapted. A prototype has been designed and implemented to ensure the reliability and availability of data in real time. In addition, a validation study is carried out in order to know its performances. For this, the designed prototype is installed on the same site near an approved industrial station, between July 1st and October 31st, 2017. The comparison between obtained results of these two stations is very satisfactory with a correlation of 99%.

Keywords: Internet of Things · Rainfall monitoring
Connected rain gauge · Automatic weather station
Real-time observations

1 Problematics and Motivations

1.1 Impact of Rainfall Disasters

Since the end of drought of the 1980s in the Sahel [1], we observe an intensification of the rainy season reflected by an increase of extreme events. Because of this, largest cities suffer a lot, especially in Senegal. These extreme phenomena are amplified since the beginning of the 2000s [2]. They have very great impacts on the environment, the survival of the population as well as the economy. Dakar, the administrative and economic capital of Senegal, is not left out. Indeed, 40.3% of Senegalese population is centered in urban areas like Dakar, according to the “Agence Nationale de Statistique et de la Démographie” (ANSD) in 2013¹. With

¹ <http://www.ansd.sn/>.

the problems of planning and the increasing urbanization in these agglomerations, floods are more and more recurrent causing sometimes inestimable damage such as insalubrity in public places, inaccessibility, traffic jams, collapse of buildings and even losses of human life [2]. The Fig. 1 shows an example of flood consequences in Dakar city.



Fig. 1. Impact of flood and rain disasters in Dakar during rainy season

In addition, there is the sewers backflow with all the health risks they imply. Environment and health of people are both concerned too. During rainy season, many diseases proliferate such as malaria, plague, dermatological diseases, diarrhea, etc. Wastewater, contents of garbage cans, dirt and other dangerous and sometimes toxic substances are often found in water. People crossing these waters are then exposed to all kinds diseases and infections. In this context, the United Nations (UN) has set up the World Food Program (WFP) with major packages including Agro-Meteorological Watch (AMW) which shows the need to have warning stations and aims to better inform users and to better predict the drastic consequences on agricultural yields, human health and socio-economic activities [3].

In view of all these problems, in-depth rainfall monitoring becomes a necessity in Senegal cities. In this paper, we propose an improvement of the rainfall data collection process. We design an automatic monitoring systems with real time data visualization, simple, adapted and efficient.

1.2 Process Used in Senegalese Institutions

Senegal has a Sahelian climate composed of two seasons: a shorter rainy season (from June to October) with a peak in August-September and a longer dry season. This country has a long history of climate data with more than a century of observations. First rainfall measurements date from 1854 in Saint-Louis [4]. Since then, many structures were born, such as “Agence Nationale de l’Aviation Civile et de la Météorologie” (ANACIM)², “Centre d’étude régional pour l’amélioration de l’adaptation à la sécheresse” (CERAAS) which is a center of “Institut sénégalais de recherches agricoles” (ISRA)³, and “Laboratoire de

² <http://www.anacim.sn/>.

³ <http://www.isra.sn/>.

Physique de l'Atmosphère et de l'Océan - Siméon FONGANG" (LPAO-SF)⁴ of "Université Cheikh Anta DIOP" (UCAD).

Nevertheless, systems used by these structures don't allow real-time measurements and require human intervention for data collection. ANACIM has a fairly diversified network of meteorological observations with 24 synoptic weather stations and more than 300 rainfall stations, of which 150 are operational and distributed throughout Senegal. These stations record data which are transferred with an important delay. Indeed, ANACIM has several agents responsible for collecting and sending data to the agency by email or SMS (Short Message service) at the end of each day. In addition, the number of stations is far from achieving recommended standards of the World Meteorological Organization (WMO) and insufficient to ensure protection's missions of people and goods and also provision of climate services. LPAO-SF has a network of 33 rain gauges acquired as part of AMMA (African Monsoon Multi-disciplinary Analysis) international project⁵. These stations are spread over three zones: the North Zone with 8 stations, the South Zone with 9 stations, and the dense zone in the center-west of the country with 16 stations. This network is installed to cover the space centered on the NPOL radar (-17.09804 West, 14.65654 North) in Kawsara. It aims to document well the spatio-temporal variability of rainfall in the far west of Africa and also the contribution of different types of convective systems in the annual cumulative rainfall. However, the data are not collected in real time.

2 Design and Implementation of Proposed Rainfall Measurement System

2.1 Network Architecture

Our used network consists of several entities. The general architecture is shown in the Fig. 2.

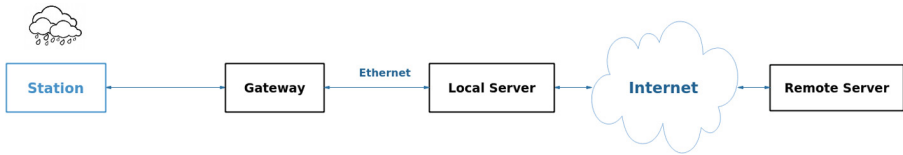


Fig. 2. Network architecture

- **Station:** It allows rainfall data acquisition and transmission to the Gateway. Each station is composed of different entities. The Design and material used are presented in the following part.

⁴ <http://www.esp.sn/?p=1658>.

⁵ <http://www.amma-international.org>.

- **Gateway:** This element is responsible for the mediation between stations and servers. It ensures the protocols conversion, retrieves and sends station data to the server via an Ethernet network. The data transmission is based on HTTP (HyperText Transfer Protocol). The Gateway is placed in Computer Science Department of “Ecole Supérieure Polytechnique” (ESP).
- **Local Server:** It receives data from the Gateway and stores them instantly. It hosts an Apache2 web server, a MySQL database server and a PHP application. Finally, data are sent to the remote Server using HTTP through Internet. They are temporarily stored in a buffer when Internet access isn’t available. After 15 min, the local server retries sending all the data and emptying the buffer if successful.
- **Remote Server:** Its role is to allow data exploitation and visualization in addition to stations monitoring. It hosts an Apache2 web server, a MySQL database server and a PHP application with some JavaScript technology (AJAX, REST, HIGHCHARTS). Data are visualized as real time graphs. A monitoring and administration interfaces are also integrated.

2.2 Designed Station

Acquisition system is the most important element in the network. It carries out the collection of rainfall data and consists of the deployed automatic stations. An automatic station is a smart-sensor, with several units (Fig. 3).

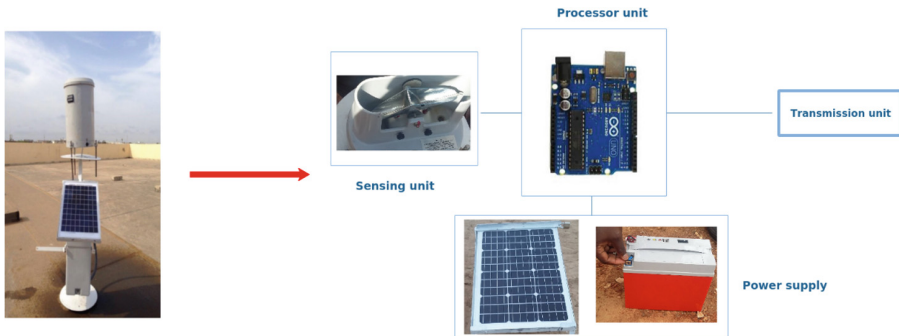


Fig. 3. Station components

Sensing Unit: Sensing unit is generally composed at least one sensor. Our stations are each equipped with a rain gauge which is the oldest sensor [5] and the most used instrument in meteorology to measure the amount of rain [6]. In this work, we use the tipping bucket rain gauge. It has two parts: a funnel-shaped collector and a container for receiving water. The collector directs rainwater towards a kind of small swing in the container, formed of two small metal buckets on each side of a horizontal axis. Water that falls is accumulated and then causes tilts. Each tilt corresponds to 0.2795 mm of rain.

Processor: It is responsible for station's intelligence. It retrieves the data from sensors, performs the processing, and formats messages. After that step, it sends data to the remote server via the communication unit. It consists mainly of an Arduino UNO board which is a microcontroller based on ATmega328P. It counts the tilts coming from the rain gauge.

Power Supply: An automatic station is supposed to be placed in areas which are difficult to access. Then, it must have a minimum of autonomy to work as long without human intervention. The proposed station model has a solar power supply thanks to a solar panel. A battery is added for storing energy and allowing station to operate when there is no sun.

Data Transmission: To communicate with the central server, the station must be equipped with a transmission module. Processing unit fully controls it. For our stations, communication unit can use various means of transmissions:

Wireless Sensor Network: We use essentially two types of wireless network transmission.

- **RF 433:** 433 MHz band is part of ISM (Industrial, Scientific and Medical) band. It is a free band intended for the use industrial, scientific and medical purposes [7] for communications via radio waves. RF band 433 allows low-coast wireless transmissions with low cost [8,9].
- **LoRa:** LoRa (Long Range) is a wireless communication technology belonging to the Low Power Wide Area Network (LPWAN) category. It allows long distance transmissions with low power consumption, low cost and a bit rate of less than 50 kbps [10]. LoRa commonly refers to two different layers: a physical layer (also called LoRa) using Chirp Spread Spectrum modulation (CSS) [11], and a LoRa Wide Area Network (LoRaWAN) protocol which provides a mechanism for controlling access and using LoRa modulation [12]. LoRa is used on ISM band 433, 868 or 915 MHz depending on the geographical zone. In early 2017, we conducted coverage tests with LoRa in Dakar, with a maximum range of 10 km [13].

Cellular Phone Networks

- **SMS:** Cellular phone networks, particularly 2G (Second Generation), are generally a choice to replace wireless sensor networks for very long distance communications. GSM (Global System for Mobile Communications) is the most widely used 2G standard and allows SMS (Short Message Service) short text messages. In Senegal, GSM covers almost the entire national territory. As a consequence, SMS is very convenient for sending data in a near real-time way [14,15].
- **GPRS:** GPRS (General Packet Radio Service), still qualified as 2.5G or 2G +, is an extension of the GSM network. It has the advantage of being faster than SMS and less expensive [14].

- **3G:** 3G has started to be used in order to satisfy users in terms of speed. Offering bits rate up to 42 Mbps, it is often an ideal candidate for a real-time system with a lower cost [16,17].

ADSL: Station can also use the Asymmetric Digital Subscriber Line (ADSL) when placed far from Gateway. In Senegal, ADSL is widely used, well before the advent of 3G, offering speeds of up to 10 Mbps. Thus, the station sends data directly over the Internet to the Gateway via the wired network, which is more stable and more available than wireless networks.

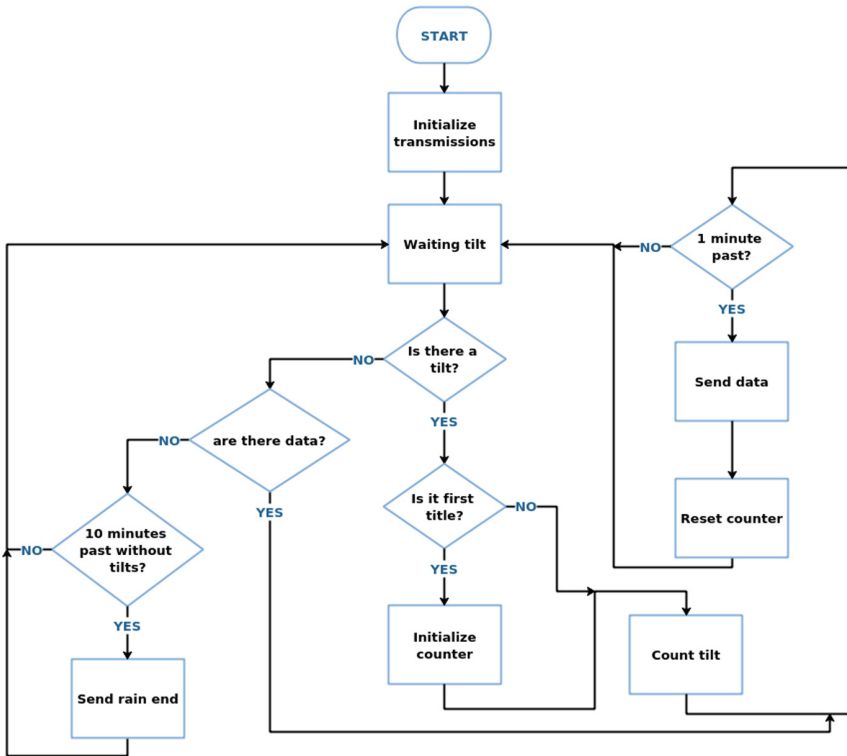


Fig. 4. Rainfall data processing flowchart

2.3 Data Collect

Data collection is done in several phases, from station to servers.

Algorithm in Station: In initial state, system waits for tilts. Then when rain starts, system considers a new rain event and then starts counting number of

tilts. After each **1 min**, a message is formatted and composed of the total number of tilts, date and time, and is sent to Gateway. After this, counter is then reset. When there is no tilts after **10 min**, the system considers it is the end of rain event, and sends then a rain end message to Gateway. This process is described by algorithm in Fig. 4.

Transmission Process: Once the message is formatted by station, it is sent to the gateway which stores it in its internal memory after cleaning up errors. Finally, data are transmitted to central site. Data is then sent to remote server which stores it in its turn. Redundancy in storage allows us to ensure data availability and reliability. It allows our system to be completely independent from Internet.

2.4 Monitoring Application

In order to exploit data, a web application is integrated in the solution. The purpose of this application is mainly to:

- allow real-time viewing of data from automatic stations.
- enable researchers to consult the historical data and to be able to download them.
- allow system administration, accounts creation and deletion, consultation of details of stations and also setting.

User who comes on the application can directly observe the evolution of rainfall in real time, in the form of cumulation on a graph. He can also see total amount per rain event for the current day for all stations in the network.

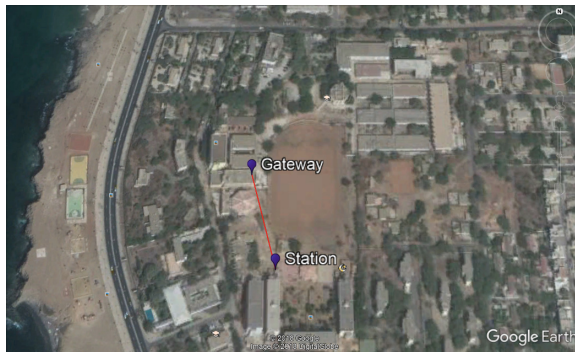


Fig. 5. Gateway and station positions in Google earth map. This image show the exact position of the Gateway which is placed in Computer Science Department and the station placed in Social Campus, in ESP

3 Validation Study

In this part, we present the validation of performances of our station. Concretely, we compare collected data of our station to data of an industrial station, both placed side by side. We collected rain data during the entire rainy season of 2017, from July 1st to October 31st, 2017.

3.1 Prototype of Our Automatic Station

To validate, we placed one of our stations in ESP Social Campus, at a distance of 110 m from the Gateway as shown in the Fig. 5.

That station uses RF 433 module for wireless communications. It consists of a CC1101 module presented in Fig. 6. This module we have already used in another monitoring project [18] has satisfactory performance and is easy to use. Communication between Processing unit and RF module is done via **SPI** (Serial Peripheral Interface) protocol. This protocol is a synchronous serial data bus which operates in full-duplex mode according to the master-slave model and compatible with Arduino cards. A CC1101 module is also used in the Gateway.



Fig. 6. CC1101 module used in our prototype for validation

3.2 Industrial Station

The industrial station used consists of a rain gauge from **LAMBRECHT meteo GmbH**⁶. It is shown in the Fig. 7 on the right. It has a precision of ± 0.1 mm. It requires at least a supply voltage equal to 9.8 V.

Data collection from this rain gauge is done using a wireless transmission too but on a commercial frequency. Station counts and saves total number of tilts every 10 min and sends data daily to a remote server in LPAO-SF. It uses its own transmission protocol.

⁶ http://lambrecht.net/de/niederschlag/meteorologie_hydrologie.



Fig. 7. Our automatic station designed (left) and Industrial rain gauge (right)

3.3 Data, Periods and Method

We have deployed a station placed with an industrial station side by side. These two stations have measured rainfall from July to October 2017. They recorded a total of 31 rainy events. We retrieved all these data and did analysis and comparison. To do this, we used **Python** programming language which is quite easy, open source and very powerful as well as **MATLAB**.

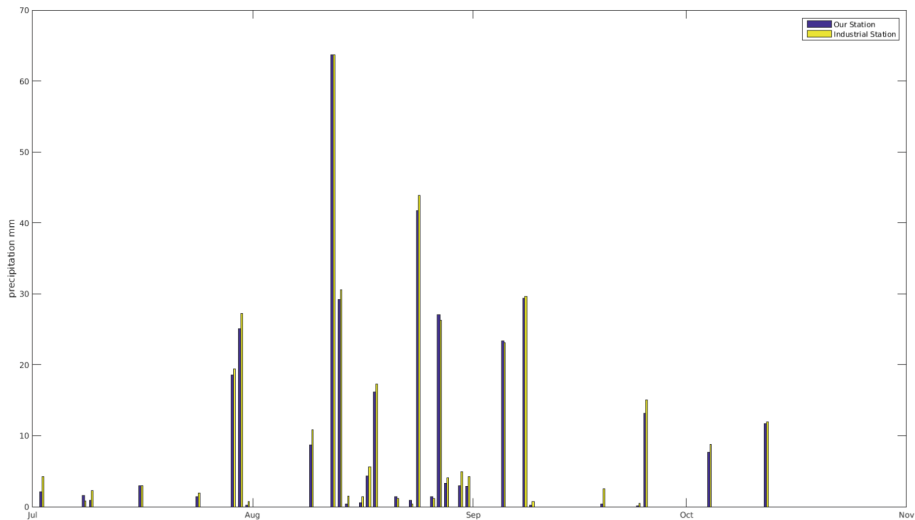


Fig. 8. Rainfall seasonal cycle. In blue, we have our station and in yellow, we have industrial station. (Color figure online)

3.4 Correlations and Results

First, we made a comparison of daily total measurements. In Fig. 8, we represent the daily measurements without any treatment. We have in yellow the data of the industrial station and the collected data of our automatic station in blue. We find that these two stations measure daily quantities substantially equal.

We obtain a correlation coefficient r equals to 0.99 between the collected data of our station and industrial station. The difference in total cumulation is less than 10 mm. So it is very satisfying as result. Cumulation is shown in Fig. 9.

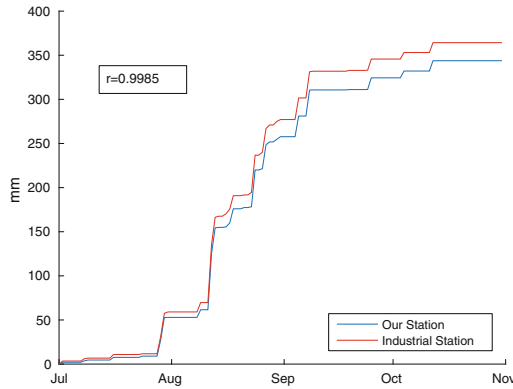


Fig. 9. Rain daily accumulation throughout the season. Our station data is blue color and red color represents industrial station data. (Color figure online)

We also studied the daily differences in the measures (see Fig. 10). In this figure, we represent the difference between the total quantity of rain measured by our station and total quantity of rain measured by industrial station used. During the rainy season, the difference fluctuate between 0 and 2.5 mm in absolute value. Nevertheless, in the majority of cases, this difference is negative (in real terms), in other words, the daily quantity measured by the industrial station is usually greater than the quantity measured by our station.

4 SenPluvio Project Overview

In what precedes, we presented a prototype of an automatic rainfall station, autonomous in energy and low cost implementation. It is placed in ESP next to an approved industrial station during 2017 rainy season under optimal weather conditions for validation. As pointed out earlier, our aim is to overcome rain damage especially floods. It is with this in mind that we initiate the SenPluvio project in LPAO-SF to design and to deploy a wide automatic rainfall stations network, less expensive, reliable and suitable for rainfall monitoring in Dakar city. On the one hand, we will improve existing data collection process used in

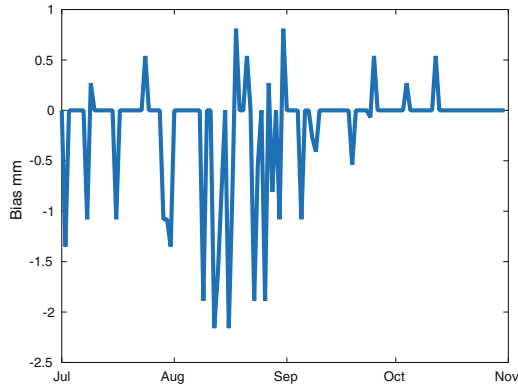


Fig. 10. Biases or difference between total quantity of rain measured by our station and total quantity of rain measured by industrial station used.

LPAO-SF. The station model that we proposed in this paper will be used. On the other hand, we will monitor and exploit the rain in real time. Every Internet users could view data in real time. An alert system will be also used to warn the competent authorities of the risk of floods. The automatic stations will be deployed in various areas of Dakar in order to constitute a very homogeneous network.

5 Conclusion

In this paper, we propose a rainfall collection system adapted, efficient and at a lower cost compared to industrial systems. We designed and implemented an automatic station for periodic measurements with a real-time storage and visualization system. After that, we studied the performance of our station using an industrial station. The result obtained reveals that our station has very coherent performances and accuracies compared to the industrial station.

We also presented a short overview of SenPluvio project whose objective is to ensure the deployment of many stations like the prototype presented here, throughout Dakar city. It's also the next step of our work. On the one hand, we are going to work and add many improvements in communication part, in particular data compression before transmission, intelligence in stations, and also security which is necessary if we want to ensure integrity. On the other hand, we will proceed to the deployment of stations in several sites in the city, chosen strategically in order to represent rain phenomenon. Finally, a warning system for populations and competent authorities will be set up with a real-time decision-making and forecasting system.

References

1. Alhassane, A., Salack, S., Ly, M., Lona, I., Traoré, S.B., Sarr, B.: Évolution des risques agroclimatiques associés aux tendances récentes du régime pluviométrique en Afrique de l'ouest soudano-sahélienne. *Science et changements planétaires/Sécheresse* **24**(4), 282–293 (2013)
2. Mbengue, R., Thiam, M., Diaw, A.: Évolution, contexte climatique et récurrence des inondations sur le littoral de Ngor au Sénégal (2014)
3. <http://fr.wfp.org/>. June 2017
4. Dacosta, H., Kandia, K.Y., Malou, R.: La variabilité spatio-temporelle des précipitations au Sénégal depuis un siècle, pp. 499–506 (2002)
5. Strangeways, I.: A history of rain gauges. *Weather* **65**(5), 133–138 (2010)
6. Messer, H., Sendik, O.: A new approach to precipitation monitoring: a critical survey of existing technologies and challenges. *IEEE Sig. Process. Mag.* **32**(3), 110–122 (2015)
7. Karl, H., Willig, A.: *Protocols and Architectures for Wireless Sensor Networks*. Wiley, Hoboken (2007)
8. Vermesan, O., Friess, P.: *Internet of Things-From Research and Innovation to Market Deployment*, vol. 29. River Publishers, Aalborg (2014)
9. Buyya, R., Dastjerdi, A.V.: *Internet of Things: Principles and Paradigms*. Elsevier, New York (2016)
10. Vangelista, L., Zanella, A., Zorzi, M.: Long-range IoT technologies: the dawn of LoRaTM. In: Atanasovski, V., Leon-Garcia, A. (eds.) *FABULOUS 2015*. LNICST, vol. 159, pp. 51–58. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-27072-2_7
11. Reynnders, B., Pollin, S.: Chirp spread spectrum as a modulation technique for long range communication, pp. 1–5 (2016)
12. Augustin, A., Yi, J., Clausen, T., Townsley, W.M.: A study of LoRa: long range & low power networks for the Internet of Things. *Sensors* **16**(9), 1466 (2016)
13. Seye, M.R., Gueye, B., Diallo, M.: An evaluation of LoRa coverage in Dakar Peninsula. In: 2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (2017)
14. Raghava, T.K.V., Wani, S.P.: Internet enabled tipping bucket rain gauge. In: 2014 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–5. IEEE (2014)
15. Retamar, A.E., Garcia, F.C.C., Yabut, J.J.M., Javier, J.C.: Design and development of a remote station for real-time monitoring of urban flooding. *Proc. Asia-Pac. Adv. Netw.* **38**, 99–114 (2014)
16. Torii, Y., Otsuka, T., Ito, T.: A diversity sensor connection capability WSN for disaster information gathering system. In: 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), pp. 1–6. IEEE (2016)
17. Hsieh, H.-C., Lai, C.-H.: Internet of Things architecture based on integrated PLC and 3G communication networks. In: 2011 IEEE 17th International Conference on Parallel and Distributed Systems (ICPADS). IEEE, pp. 853–856 (2011)
18. Kama, A., Diallo, M., Dramé, M.S., Ndiaye, M.L., Ndiaye, A., Ndiaye, P.A.: Monitoring the performance of solar street lights in Sahelian environment: case study of Senegal. In: The 10th International Conference on the Developments in eSystems Engineering. IEEE (2017)



A Parallelized Spark Based Version of mRMR

Reine Marie Ndéla Marone^{1(✉)}, Fodé Camara², and Samba Ndiaye¹

¹ Department of Mathematics, Cheikh Anta Diop University, Dakar, Senegal
reine.marie.marone@ucad.edu.sn

² Department of Mathematics, Alioune Diop University, Bambey, Senegal
fode.camara@uadb.edu.sn

Abstract. Nowadays, we are surrounded by enormous large-scale high dimensional data called big data and it is crucial to reduce the dimensionality of data for machine learning problems. That's why feature selection plays a vital role in the process of machine learning because it aims to reduce high-dimensionality by removing irrelevant and redundant features from original data. However some characteristics of big data like data velocity, volume and data variety have brought new challenges in the field of feature selection. In fact, most of existing feature selection algorithms were designed for running on a single machine (centralized computing architecture) and do not scale well when dealing with big data. Their efficiency may significantly deteriorate to the point of becoming inapplicable. For this reason, there is an increasing need for scalable yet efficient feature selection methods. That's why we present here a distributed and effective version of the mRMR (Max-Relevance and Min-Redundancy) algorithm to face real-world problems of data mining and evaluate the empirical performance of the proposed algorithms in selecting features in several public datasets. When we compared the efficiency and the scalability of our parallelized method in comparison with the centralized one we have found out that our parallelized method have given better results.

Keywords: Feature selection · Filter method · Parallel computing
Apache Spark · mRMR · SVM

1 Problematic and Related Works

1.1 Introduction

Feature selection is a fundamental preprocessing step that aims to reduce input dimensionality in machine learning and pattern recognition [1]. Many domains use feature selection: for example in bioinformatics it is an important topic because it is critical to define informative features from complex high dimensional biological in disease study, drug development, etc.

Unfortunately, most feature selection algorithms are designed for centralized computing and do not scale well with large-scale datasets [3]. To tackle these problems, distributed computing framework like Apache Hadoop, which implements the MapReduce model, can be a solution [3].

But the Apache Hadoop is not adapted to feature selection because it lacks built-in support for the iterative process [4]. So, an alternative to Hadoop has been presented to overcome these problems. It is Spark, a memory-based iterative computation framework that improves the IO read/write performance issue by processing intermediate data in-memory [4].

In regard to that, in this paper, we propose a parallel version of the centralized mRMR algorithm that we have named SFS-mRMR (for Spark Feature Selection method based on mRMR), on the framework Spark to ameliorate its efficiency. The choice of mRMR is motivated by the fact that minimum-redundancy-maximum-relevance (mRMR) selector is considered one of the most relevant method for dimensionality reduction due to its high accuracy.

The results that we obtained show that our algorithm is scalable and outperforms the classical mRMR feature selection method.

The rest of the paper is structured as follows:

Section 2 reviews previous works. Section 3 deals with the formulation of the problem. Section 4 presents the centralized mRMR. Section 5 gives the metrics we used in our proposal. Section 6 consists of the presentation of our algorithm. Section 7 describes the working environment. Section 8 presents and analyzes the results of the experiments. Section 9 concludes and gives futures researchs.

1.2 Related Works

Feature selection is a fundamental preprocessing step to reduce input dimensionality.

There are 3 general categories of feature selection methods: Filter, Wrapper and embedded [5].

Filters methods use some criterion to score each feature and provide a ranking to evaluate the features which determine their relevance or discriminant powers with the outcome variable [5].

In the wrapper methods the accuracy of classifier is estimated to select the features [5]. Although computationally expensive the wrapper is the best feature selection method for accuracy [5].

In Embedded methods a given model is used to guide the feature selection process, and select the most relevant features when building the model [5].

Filter methods offer better computational complexity but do not take into account the interactions among the variables, which cannot be ignored.

mRMR is one of the most famous filter method. But mRMR is a centralized method and do not scale well with ultrahigh dimensional datasets. So it is fundamental to optimize the mRMR algorithm by using efficient parallelization [7]. That's why, proposals have been made on the parallelization of mRMR algorithm the interest of which is to decrease the training time and ameliorate the accuracy of the machine-learning tasks.

The work in [1] present a parallelization of many methods based on information theory including mRMR in Apache Spark.

The Experimental results show that this methods scale well and efficiently with ultra-high-dimensional datasets.

The work in [8] proposes to extend mRMR by using a number of approaches to better explore the feature space and build more robust predictors. To deal with the computational complexity of those approaches, authors implement and parallelize functions in C using the openMP Application Programming Interface. These methods show significant gains in terms of run-time.

Authors in [9] present a two-stage selection algorithm by combining ReliefF and mRMR. In the first stage, ReliefF is applied to find a candidate gene set; In the second stage, mRMR method is applied to directly and explicitly reduce redundancy for selecting a compact yet effective gene subset from the candidate set. The experimental results show that the mRMR-ReliefF gene selection algorithm is very effective.

In [10], authors present three implementations of an extension of mRMR named fast-mRMR in several platforms, namely, CPU for sequential execution, GPU (graphics processing units) for parallel computing, and Apache Spark for distributed computing using big data technologies.

In [11], authors combined dynamic sample space with mRMR and proposed a new feature selection method. In each iteration, the weighted mRMR values are calculated on dynamic sample space consisting of the current unlabelled samples. The feature with the largest weighted mRMR value among those that can improve the classification performance is selected in preference. Five public datasets were used to demonstrate the superiority of this method.

It is clear that the methods presented in these different works use a greedy approach by iteratively add or remove features into a set of features. Our method selects a set of relevant and non-relevant features on the dataset using only one iteration. That allows a more significant reduction of the learning time and an improvement of classification accuracy.

1.3 Formulation

Our work focuses on classification with 2 classes. Let c be the class label with 2 possible values 0 or 1. Let S refer to the input dataset with a high number n of features $\{i_1, \dots, i_n\}$ and m instances. V is an example defined by a vector (v_1, \dots, v_n) , where v_j is the value of the feature i in V . Let $O(S, D)$ denotes the evaluation function. A subset S' of S is evaluated by O with the data D . Let S_1 and S_2 be 2 subset of features in S . $O(S_1, D) > O(S_2, D)$ means that S_1 is more interesting than S_2 .

Our proposed algorithm, called SFS-mRMR, is a distributed version of the mRMR method that we based on Spark, a parallel programming framework. Our method aims to find a subset S' of features from S that maximize the function O .

2 Improvement of MRMR

2.1 The Classical MRMR

mRMR means Minimum Redundancy and Maximum Relevance. The concept of mRMR is to select the features so that they are mutually maximally dissimilar and maximally relevant with the class label l [12]. Let i and j represent 2 features of S . The mutual information between i and j is denoted by $M(i, j)$. $M(c, i)$ stands for the mutual information between the class label c and i .

The redundancy among the features in S , is obtained by calculating the mutual information between the features in S as follow:

$$Q_I(S) = \frac{1}{|S|^2} \sum_{i,j \in S} M(i, j) \quad (1)$$

The relevance of features in S with the class label c is defined as

$$R_I(S) = \frac{1}{|S|} \sum_{i \in S} M(c, i) \quad (2)$$

By optimizing (1) and (2), we obtain S^* the set constituted of features that are the most relevant and less redundant in S . It is done as follows:

$$S^* = \operatorname{argmax}_{S' \subseteq S} [R_I(S') - Q_I(S')] \quad (3)$$

2.2 Our Proposal

SVM is an extremely powerful machine learning technique and one of the best supervised classification techniques [13]. That's why, SVM is used in combinaison with mRMR in our algorithm SFS-mRMR as proposed by authors in [14]. The features are scored by using this combinaison to obtain more performance. Let $\beta \in [0, 1]$ denotes a ratio between SVM scoring and mRMR scoring. The relevancy $R_{F,i}$ of feature i in S is obtained as follow:

$$R_{S,i} = \frac{1}{|S|} \sum_c M(c, i) \quad (4)$$

$Q_{F,i}$ the redundancy of i is calculated as

$$Q_{S,i} = \frac{1}{|S|^2} \sum_{i,j \in S} M(i, j) \quad (5)$$

Let ω_i be the SVM weight of i .

The final score d_i of i is obtained as follow:

$$d_i = \beta |\omega_i| + (1 - \beta) \frac{R_{S,i}}{Q_{S,i}} \quad (6)$$

3 Our Method

The method that we propose (SFS-mRMR) takes as input:

- a set of data S composed of x attributes and y observations,
- the number T of attributes to select in S
- β , a ratio between SVM scoring and mRMR scoring,
- and z the desired partition number for the dataset.

Let D be the set of features in S .

Our method return S' the subset constituted of T best features namely the features that have the highest d_i values.

SFS-mRMR follows seven steps:

▪ **Stage 1:**

1. Construct $classes = \{c_1, \dots, c_y\}$ the set of the class labels in S .
2. Construct $values = \{\{v_i^1, \dots, v_i^y\}, i=1 \text{ to } x\}$
 v_i^j represents value of the i -th feature in the j -th observation:
3. Construct z subspaces of features SD_t , $t = 1..z$ from D .
4. Construct z subspaces sub_t of $\{\{v_i^1, \dots, v_i^y\}, i \in SD\}$
5. Send each sub_t at a worker.

▪ **Stage 2:**

On each worker t :

6. Create several sets for each feature i in sub_t by mapping i with each other feature j in D as follows:

$$i \Rightarrow \{i, \{v_i^1, \dots, v_i^y\}, \{v_j^1, \dots, v_j^y\}, \{c_1, \dots, c_y\}\}$$

The resulting set constituted of $\{i, \{v_i^1, \dots, v_i^y\}, \{v_j^1, \dots, v_j^y\}, \{c_1, \dots, c_y\}, i=1 \text{ to } x, j=1 \text{ to } y\}$ is called *rdd2*.

▪ **Stage 3:**

In this step, we use each element of *rdd2* to calculate mutual information M_{ij} between each feature *i* and another feature *j* of *D*. We compute also the relevance R_i (mutual information with his class label) of *i*. We proceed as follows:

For each element $e \in rdd2$

7. $rdd [(i, M_{ij}, R_i)] = mapToPair (e \Rightarrow \{ i, M_{ij}, R_i \})$

$M_{ij} = MutualInformation (\{v_i^1, \dots, v_i^y\}, \{v_j^1, \dots, v_j^y\})$

$R_i = MutualInformation (\{v_i^1, \dots, v_i^y\}, \{c_1, \dots, c_y\}) / x$

where $c_{k \in 1..y}$ is the class label of the observation *k*.

End For each

The constituted set of each feature *i*, the mutual information between *i* and a feature *j* of *D* and the mutual information R_i between the class label and *i*, will be called *rdd3*.

▪ **Stage 4 :**

To compute redundancy of each feature *i*, the mutual information between *i* and other features in *D* are aggregated. It gives us a new set and we name it *rdd4*. Each value in *rdd4* consists of $\{i, sumM_{ij}, R_i\}$, where

$sumM_{ij}$ is the sum of mutual information between *i* and other features in *D* and R_i the mutual information between *i* and the class label.

It is done as follow:

For each $(i, M_{ij}, R_i) \in rdd3$

8. $rdd [(i, sumM_{ij}, R_i)] = reduceByKey (_+ _)$

$$sumM_{ij} = \sum_{i=1}^x M_{ij}$$

End For each

▪ **Stage 5:**

In this step compute for each feature *i*, its SVM weight ω_i as follows:

For each *i* in *D*

9. $rdd [(i, \omega_i)] = map (i \Rightarrow \{ i, \omega_i \})$

ω_i in ω where $\omega = SVMWeight(F)$

End For each

▪ **Stage6:**

In this step calculate for each feature i its ranking measure d_i and sent all d_i scores to the master.

It corresponds to the following instructions:

For each element $(i, \text{sum}M_{ij}, R_i) \in \text{rdd4}$

```
10.rdd [(i, di)] = mapToPair ({i, sumMij, Ri} => {i, di})
```

```
Qi = sumMij / (n * n);
```

```
di = β + ωi + ((1 - β) * (Ri / Qi));
```

```
/* ωi is SVM weight of feature i */
```

End For each

11. All workers send d_i values to the master

▪ **Stage 7:**

In this stage the features are collected and ordered by the master. Master then returns the T attributes with the best scores d_i .

This corresponds to the following sentences:

On the master:

12. Collect and ordered

13. Return S' : the set of T attributes that have the best d_i values.

4 Experimental Setup and Results

4.1 Data Description

The classifier used is SVM (support vector machine) and the data used for the experiments are in LibSVM format.

Datasets used here, are from mldata.org [15].

Table 1 gives us details of the datasets.

Table 1. Characteristics of benchmark datasets

Name	Number of features	Number of instances
Colon-cancer	2000	62
Colon-tumor	2000	60

For our experiments we have used a cluster of 4 nodes then a cluster of 6 nodes. Each node has 8 cores and run at 2.60 GHz, with 56 GB memory and a 382 GB disk, Each node run at the Linux-based HDInsight (Spark) cluster.

4.2 Performance Evaluation

In this part, we will first discuss the scalability of our solution then we will compare the execution time of our proposal with the one of centralized mRMR.

Figures 1, 2 and 3 show respectively how the execution time varies according to the number of nodes when we select 25%, 50% or 75% of the dataset.

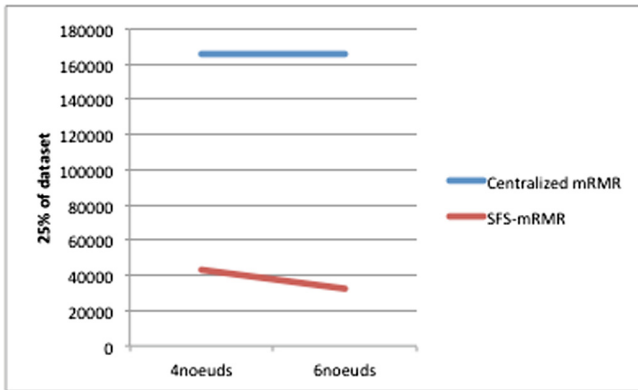


Fig. 1. Scalability of SFS_mRMR and classical mRMR with 25%.

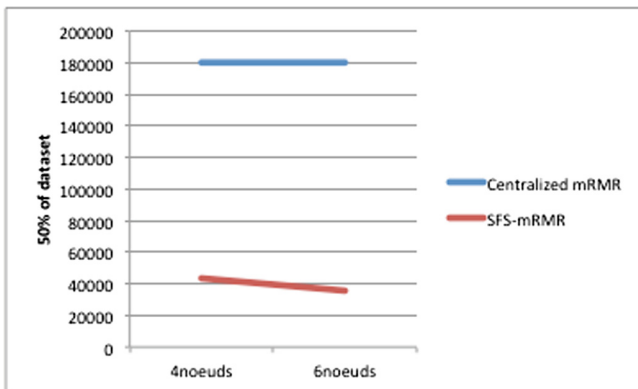


Fig. 2. Scalability of SFS_mRMR and classical mRMR with 50%.

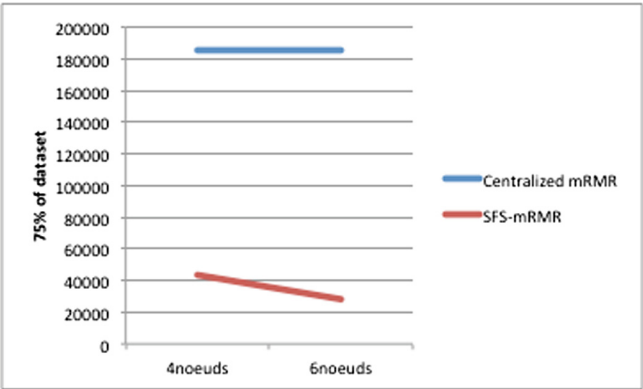


Fig. 3. Scalability of SFS_mRMR and classical mRMR with 75%.

The concerned figures clearly show that the execution time of our proposal considerably decreases when the number of nodes increases whereas the time taken by classical mRMR remains constant.

We have used 4 then 6 nodes for the scalability. And for every case we have run the tests using the same environment.

For every dataset we first select 25% then 50% and after 75% of features.

• **Colon-cancer**

Figures 4 and 5 shows the time taken by our method comparatively to the one of centralized mRMR for respectively 4 and 6 nodes.

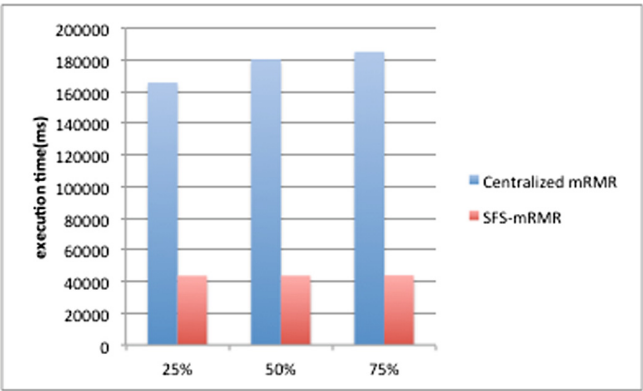


Fig. 4. Time taken for colon-cancer with 4 nodes

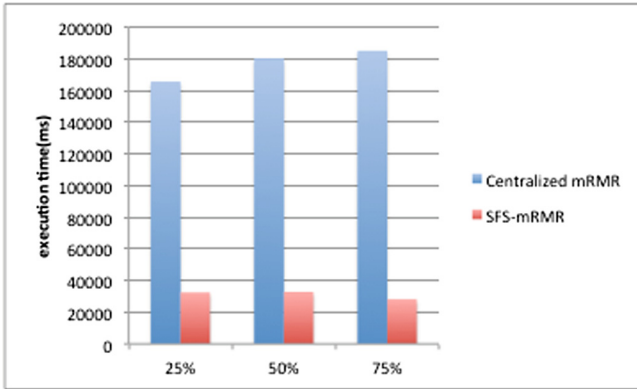


Fig. 5. Time taken for colon-cancer with 6nodes

As we can notice, the execution time of our method is at least 4 times shorter compared to the one of centralized mRMR.

- **Colon-Tumor**

For colon-tumor the results obtained with 4nodes are the following ones given in Fig. 6:

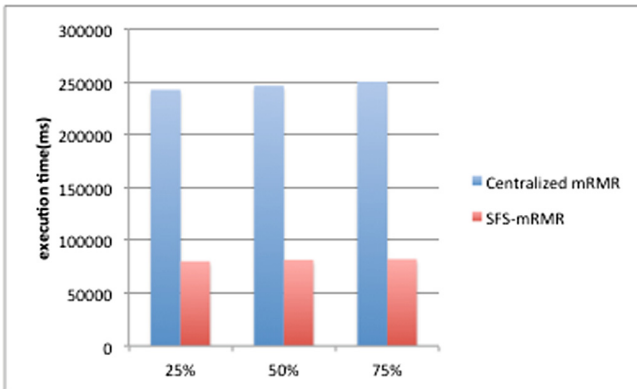


Fig. 6. Time taken for colon-tumor with 4nodes

With a cluster of 6nodes the execution time is stated in Fig. 7.

As for the colon-cancer we can notice that the execution time of our method SFS-mRMR is also 4 times shorter at least.

Therefore, we can conclude from these experiments that our solution outperforms the centralized mRMR method in terms of execution time. Moreover, the more we increase the number of nodes, the shorter the execution time becomes in our method whereas the one of centralized mRMR remains constant.

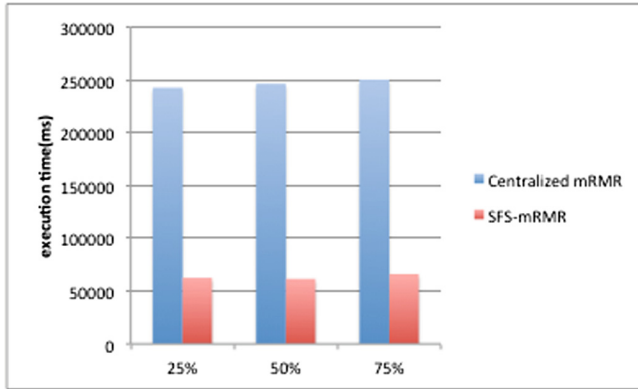


Fig. 7. Time taken for colon-tumor with 6nodes

In our experiments we have used datasets limited to 2000 features, because beyond that number, the centralized mRMR takes too much time to run. For example, for certain datasets above 2000 mRMR can take days to run completely.

5 Conclusion

In this paper, we have proposed a parallel and scalable version of a centralized feature selection method named mRMR that we developed with the Spark framework.

In our method, a score is given to each feature to evaluate its redundancy with the others features of the dataset and its relevance relatively to the class label. Then the features presenting the highest score are returned.

Performance evaluation of our method demonstrates that our parallel algorithm can improve the accuracy of classification and reduce the time taken by instance selection. The performance results also show that our method scale well and efficiently with big data.

In the future, we plan to compare our method with other parallelized methods in the litterature.

Acknowledgment. In this work, Microsoft Azure has sponsored us and we would like to take this occasion to express them our thanks.

Without their help we would not have been able to test our algorithms in a cluster and we would not have reached our goals.

References

1. Ramirez-Gallego, S., et al.: An information theory-based feature selection framework for big data under apache spark. *J. Latex Class Files* **13**(9) (2014)
2. Chahar, V., Chhikara, R., Gigras, Y., Singh, L.: Significance of hybrid feature selection technique for intrusion detection systems. *Indian J. Sci. Technol.* **9**(48) (2016). <https://doi.org/10.17485/ijst/2016/v9i48/105827>

3. Zhao, Z., Cox, J., Duling, D., Sarle, W.: Massively parallel feature selection: an approach based on variance preservation. In: Flach, P.A., De Bie, T., Cristianini, N. (eds.) ECML PKDD 2012. LNCS (LNAI), vol. 7523, pp. 237–252. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33460-3_21
4. Singh, D., Reddy, C.K.: A survey on platforms for big data analytics. *J. Big Data* **2**(1), 8 (2015)
5. Liu, C., Wang, W., Zhao, Q., Konan, M.: A new feature selection method based on a validity index of feature subset. *Pattern Recogn. Lett.* **92**(1), 1–8 (2017)
6. Wenyan, Z., Xuewen, L., Jingjing, W.: Feature selection for cancer classification using microarray gene expression data. *Biostat. Biom. Open Access J.* **1**(2), 555557 (2017)
7. Jaseena, K.U., David, J.M.: Issues, challenges, and solutions: big data mining. In: Sixth International Conference on Networks & Communications. <https://doi.org/10.5121/csit.2014.41311>
8. De Jay, N., Papillon, S., Olsen, C., El-Hachem, N., Bontempi, G., Haibe-Kains, B.: mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics* **29**(18), 2365–2368 (2013). <https://doi.org/10.1093/bioinformatics/btt383>
9. Zhang, Y., Ding, C., Li, T.: Gene selection algorithm by combining reliefF and mRMR. *BMC Genom.* **9**(Suppl 2), S27 (2008). <https://doi.org/10.1186/1471-2164-9-S2-S27>
10. Ramírez-Gallego, S., et al.: Fast-mRMR: fast minimum redundancy maximum relevance algorithm for high-dimensional big data: fast-mRMR algorithm for big data. *Int. J. Intell. Syst.* (2016). <https://doi.org/10.1002/int.21833>
11. Yang, Y., Li, H., Lin, X., Ming, D.: Recursive feature selection based on minimum redundancy maximum relevancy. In: 2010 Third International Symposium on Parallel Architectures, Algorithms and Programming (PAAP) (2010). <https://doi.org/10.1109/paap.2010.52>
12. Mandal, M., Mukhopadhyay, A.: An improved minimum redundancy maximum relevance approach for feature selection in gene expression data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2016)
13. Chang, Y.-W., Lin, C.-J.: Feature ranking using linear SVM. In: Proceedings of the Workshop on the Causation and Prediction Challenge at WCCI 2008, PMLR, vol. 3, pp. 53–64 (2008)
14. Mundra, P.A., Rajapakse, J.C.: SVM-RFE with MRMR filter for gene selection. *IEEE Trans. Nanobiosci.* **9**(1), 31–37 (2010)
15. <http://mldata.org/repository/data/viewslug/ovarian-cancer-nci-pbsii-data/>



Multi-scenario Modelling of Learning

Guy Merlin Mbatchou^{1,2(✉)}, François Bouchet¹,
and Thibault Carron¹

¹ Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, LIP6,
75005 Paris, France

{guy.mbatchou, francois.bouchet, thibault.carron}@lip6.fr

² Université Assane Seck de Ziguinchor, Ziguinchor, Sénégal
guy.mbatchou@univ-zig.sn

Abstract. Designing an educational scenario is a sensitive and challenging activity because it is the vector of learning. However, the designed scenario may not correspond to some learners' characteristics (pace of work, cognitive styles, emotional factors, prerequisite knowledge, ...). To personalize the learning task and adapt it gradually to each learner, several scenarios are needed. Adaptation and personalization are difficult because it is necessary on the one hand to know in advance the profiles and on the other hand to produce the multiple scenarios corresponding to these profiles. Our model allows to design many scenarios without knowing the learner profiles beforehand. Furthermore, it offers each learner opportunities to choose a scenario and to change it during their learning process. The model ensures that all announced objectives have enough resources for acquiring knowledge and activities for evaluation.

Keywords: Adaptation · E-learning · Learning scenario · Instructional design
Learning path

1 Introduction

This work is in the field of personalization and adaptation of technology enhanced learning to make the process of acquiring knowledge more effective. Many researches are carried out in this direction: they are interested in learner models [1], intelligent tutoring systems [2], analysis of learning traces [3] or adapting educational scenarios according to the learners' profiles [4], sometimes according to multiple sources [5].

The pedagogical scenario is the description of a learning sequence, its educational goals and the means to implement it to achieve these goals. The educational scenario is a key element in learning because it is the vector of learning [6]. In a context of lifelong or even initial training, it is extremely difficult to design a scenario for each learner. Some authors rely on learners' profiles to reduce the number of possibilities. To determine learners' profiles, learning data must be available and analysed, which is time-consuming. After determining the profiles, it can happen that we have found several profiles, making the number of scenarios to conceive always enough. Moreover, during learning, the knowledge acquired by a learner and interactions with the learning environment can change their profile. For example, a learner without much computer experience at the beginning of the learning session will have a poor

performance that will improve during their learning as they acquire new computer skills. This evolution of the profile may render the initially proposed scenario inappropriate. Moreover, nothing can ensure that this new profile will match one of the identified profiles. Therefore, the teacher would need to be regularly designing new personalized scenarios as new profiles are identified, which is difficult to do.

Thus, there is a problem of designing several scenarios to adapt to the particularities of learners. We choose to break down an educational scenario into a learning scenario (related to learner) and a coaching scenario (for the teacher) that should be structured, coherent and combined to drive learning [7]. In this paper, we are interested in the learning scenario part, which is the description of the proposed learning activities, their articulation in the learning sequence as well as the expected results of learners [8]. Although this scenario is intended for the learners, its design is to be done by the teachers. Our research focuses on providing teachers with conceptual and technological tools to design a course with several learning scenarios without knowing in advance the profiles of learners.

The remainder of this paper is organized as follows. In Sect. 2, we describe the state of the art of scenario models and more specifically learning scenarios. In Sect. 3, we present our multi-scenario model of a course. In Sect. 4 we describe the implementation of our system into the LMS (Learning Management System) Moodle. In Sect. 5, we report the results of an acceptability questionnaire filled by teachers to assess the quality of their educational productions, their predispositions to personalize learners' follow-up and their resistance to change their teaching method. Finally, in Sect. 6 we present the results of an experiment involving teachers using our system during a multi-scenario course design workshop.

2 State of Art

The design of an educational scenario integrating ICT (Information and Communication Technologies) is a fundamental activity to guarantee the quality of learning by considering the training system. Based on the EML (Educational Modelling Language), most models of learning scenarios are designed as a succession of activities or tasks that the learner needs to perform to reach their learning goal. Some models divide activities based on educational goals [9]. Others propose a division based on teachers' intentions [10] take into account activities to be done by learners, teachers' intentions and interactions [11].

To carry out the division based on teachers' intentions, a set of questions must be asked to make relevant pedagogical choices. Brassard and Daele have identified 17 dimensions of questions organized into 4 categories [12]. To consider the learners' specificities, they suggest a dimension which proposes alternative or variable paths linking the activities in the scenario. The difficulty of this implementation relies on the "a priori knowledge" of learners' characteristics (cognitive styles, emotional factors, prerequisite knowledge...). Moreover, it would be tedious to implement a pedagogical scenario with these 17 dimensions, to produce as many scenarios as possible (learners' categories).

In order to produce new scenarios, Riad et al. [13] propose the reuse and the adaptation of the existing scenario to create new ones. Nevertheless, the weakness of

their approach is the impossibility to modify the scenario structure. Their adaptation consists only in modifying included resources. Using the principle of design patterns, Marne and Labat [14] propose to see activities with several states of input and output. The connections among states depend on the prerequisites between the activities and objectives achieved by the learner. The advantage of this approach relies on its flexibility in the sequence of activities, but it does not take into account a learners' profiles and is not intended to define several scenarios for a same session.

The Competence-based Knowledge Space Theory (CbKST), an extension of the Knowledge Space Theory (KST) [15] proposes a knowledge structure model based on competences for the personalization of learning [16]. The model considers precedence's relationships between competences to establish the notion of knowledge state (set of skills acquired in a field). From the different states, the CbKST allows to establish several learning paths to achieve the same goal. Thus, the CbKST provides a framework for designing multiple learning scenarios in a transparent way for teacher.

3 Multi-scenario Model

3.1 Main Objectives

Our model is inspired by the CbKST given its many successes in various fields such as medicine [17], metacognition [18], education [19] and more specifically Serious Games [20]. However, models based on CbKST have three main weaknesses:

- **Lack of support for activities with multiple competences**
In studied models, the activities allow to work and acquire only one competence at a time. In our context, activities with multiple competences (such as studies case [21]) provide to learners the possibility to use diverse or even low-level competences to solve higher-level problems. It also allows the learner to acquire new competences from those acquired. These are complex activities highly recommended in a training.
- **No temporal constraints**
Although the learning process has for general objective acquiring and evaluating acquired knowledge, this must be done within a given time defined by the conditions of the training. But the models of the CbKST do not take this constraint into account.
- **No levels or thresholds of learning satisfaction**
In models based on the CbKST, the acquisition of the competence is boolean (true or false). However, in a system of initial or continuous training, the acquisition of a competence is subject to a minimum threshold of satisfaction that the learner must reach. Moreover, a competency not acquired can be obtained by compensation as advised by the system (Bachelor - Master - Doctorate) in higher education.

The goal of our model is to provide for the teacher a tool to design several learning scenarios taking into account these different learning constraints. Our model is based on a set of initial concepts that we describe and justify below, and which have been validated by the teaching staff as we will show in Sect. 6.

3.2 The Initial Concepts

Inspired of the teachers' practice, our concepts are based on learning objects. Relationships among concepts are represented by the following class diagram (Fig. 1):

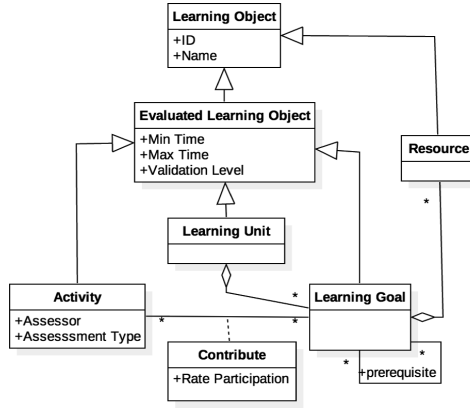


Fig. 1. Class diagram of learning objects

3.2.1 Learning Decomposition in Unit

To be close to teaching practices, a learning or training module is divided into learning units with precedence relationships. These units correspond to the notions of chapter, part, title, etc. Each unit contains a set of learning goals.

3.2.2 Learner-Centered Pedagogy and Structuration by Learning Goals

Most of current pedagogical approaches structure content in parts, chapters, titles, etc. However, our model structures content in learning goals. Each goal has a set of learning resources (R_j^i) for knowledge acquisition and a set of learning activities (A_j^i) for validating acquired knowledge. The acquisition and validation of the knowledge associated with each objective has a duration (T_i) and is conditioned by a satisfaction threshold (S_i). An objective O_i is defined as follows:

$$O_i = \{T_i, S_i, \{R_1^i, R_2^i, \dots, R_{P_i}^i\}, \{A_1^i, A_2^i, \dots, A_{N_i}^i\}\} \text{ with } (P_i, N_i) \in \mathbb{N}^2 - \{(0, 0)\}.$$

P_i is number of learning resources and N_i is number of learning activities.

The goals have prerequisite relationships among them.

3.2.3 Indexing Activities by Learning Goals

The model defines for each activity, the necessary goals for its realization. Likewise, the model ensures that each goal has enough activities to assess and validate acquired

knowledge. For this, the model has a matrix (A_j^i) where activities are in line and goals in column. The matrix contains participation rates of each activity for assessing and validating each goal. Thus, each activity A_j^i participates in the validation of the objective O_i with a rate P_j^i where $\sum_{j=1}^{N_i} P_j^i \geq 100\%$. The model can therefore handle activities with multiple goals.

3.2.4 Acquisition and Validation of Knowledge

Although the acquisition of knowledge is done by using the learning resources, our model does not take into account the fact that learners really use resources. This choice is justified by the fact that, on the one hand, we do not have means to ensure that the resource is actually being used; on the other hand, a learner may have already acquired the knowledge contained in the resource in a previous training. To ensure that knowledge is acquired, the model validates it by learning activities. An activity A_j^i is validated if the obtained score V_j^i is greater than or equal to the threshold S_j^i of validation of the activity.

An objective O_i is validated if there is a time t such as $t < T_i$ $\sum_{j=1}^{N_i} P_j^i V_j^i \geq S_i$.

3.3 Determination of Learning Paths

To determine learning paths, first the knowledge structure containing knowledge states must be generated.

Algorithm of generation of the knowledge structure

Input: **G** Graph of prerequisites among learning goals

Output: **K** Knowledge structure (set of knowledge states)

Variables: A and B are learning goals; Q, E are knowledge state (set of learning goals)

$K = \{\emptyset\} \cup \{Q\}$

For any unmarked state E of K

For each goal A \in E do

If there is no goal B in G such that $A \rightarrow B$ then

$K = K \cup \{E - \{A\}\}$

End For

Mark(E)

End For

From the knowledge structure, the learning paths are determined from the notions of internal and external fringes defined in the KST [15]. The internal (respectively external) fringe of a knowledge state K is the set of goals P such as deleting them (respectively adding them) to K , we obtain another state of knowledge which is immediately lower (respectively higher).

3.4 Impact of Activities with Multiple Goals in Learning Paths

According to KST, a validated knowledge implies its acquisition. Regardless of their current learning state, if the learner decides to do an activity with multiple goals and validates it, then they acquire the goals targeted by this activity. This validation is conditioned by the fact that the score obtained on the activity allows for the validation threshold of each goal to be exceeded.

Example: Considering learning goals a, b, c, d, e, f , and g with their prerequisite relationships, as shown on the graph in Fig. 2. By applying the CbKST approach, we can generate the learning paths (Fig. 3). Suppose that an activity targets goals c and d . The state of knowledge $\{c, d\}$ is not admissible (possible) because:

- the acquisition (validation) of c is conditioned by acquisition of b
- the acquisition (validation) of d is conditioned by acquisition of a .

So, the knowledge state associated with the acquisition of c and d is the state $\{a, b, c, d\}$. It is accessible from any state which is inferior to it. It is possible by the validation of an activity with multiple goals (green lines in Fig. 4).

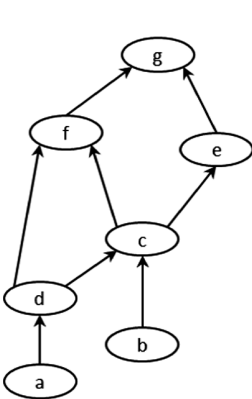


Fig. 2. Graph of prerequisite relationships among goals

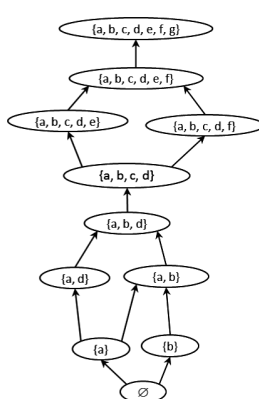


Fig. 3. Graph of learning paths

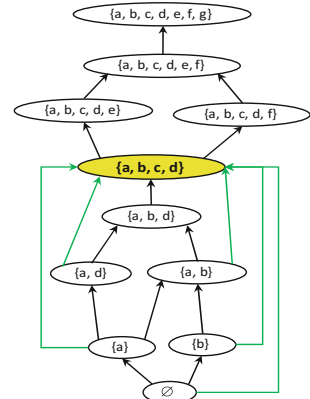


Fig. 4. Graph of augmented learning paths (Color figure online)

The activity with multiple goals increases new learning paths, leading us to the notion of augmented learning path, and the associated notions of augmented knowledge state and augmented link, defined as follows:

- **Augmented knowledge state:** a knowledge state is augmented (yellow state in Fig. 4) if it can be obtained from the validation of an activity with multiple goals. Because it's augmented, from any state that is inferior to it, the learner can access it without going through the intermediate states.
- **Augmented link:** A link from state E1 to state E2 is augmented (green link in Fig. 4) if E2 is an augmented state and E1 is not an immediately lower state of E2.
- **Augmented learning path:** A learning path is augmented if there is at least one augmented link in its list of links.

Augmented learning paths offer challengers or talented learners the opportunity to progress quickly in the acquisition of competences. An acceptable competence level as defined by the teacher can be reached the same way during this quick progress.

4 Implementation: The EGbKST Plugin

Although the previous model is independent of any learning platform, we decided to implement it as plugin for the MOODLE¹ platform (Modular Object-Oriented Dynamic Learning Environment), used in our university. The plugin is named EGbKST (Educational Goals Based Knowledge Space Theory). To show the difference between our new system and the current system, we will first present the system based on the current pedagogical model.

Table 1. Structuring learning in our current pedagogical model

General information	This section situates the course in the training and contains information about the authors
General goals	The different general goals of the course
Work instructions	Work instructions before, during and after learning
Preliminary activities	This section contains the prerequisites of course, an entrance test, keywords, course summary, bibliography, webography and a tool to choose a team for collaborative work
Communication	The different communication tools to use during the learning
Sequence 1	The course is divided in sequence (part, chapter, title, section,...). Each contains a title, a duration, a set of specific goals, a set of resources and a set of exercises
...	
Sequence N	

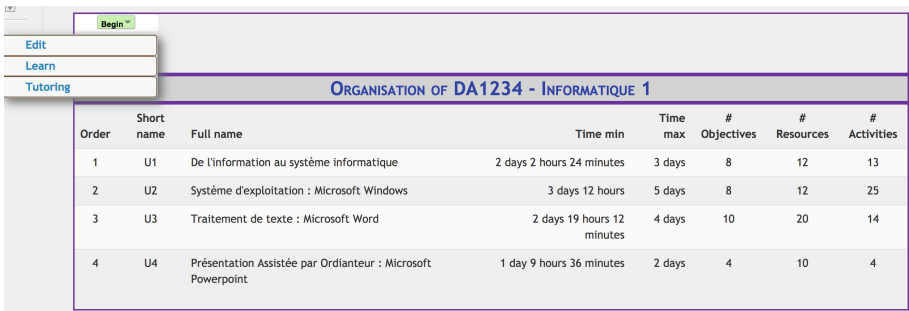
In our current pedagogical model (represented in MOODLE system), learning and evaluation are organized in sequences. The model does not ensure that the defined goals have resources for knowledge acquisition and exercises for evaluation. The evaluation made at the end of the sequence does not allow to anticipate learners' failure because the sequence contains many goals and its duration is of approximatively 2 weeks: any failure is therefore detected only when it is too late.

¹ <https://moodle.org/>.

In our system, we chose to keep the first 3 sections (from Table 1) to allow learners to have access to general information, general objectives and work instructions before starting learning. We added a fourth section including the EGbKST plugin used to design a course according to our model. The plugin integrates 3 roles: teacher (course design), student (learning) and tutor (tutoring). In this paper, since we are interested in course design, we will only present the views associated to the teacher role. To design their course, the teacher has many interfaces and proceeds as follows:

1. Adding metadata such as general goals, prerequisites, bibliography, keywords, ...
2. Adding learning units. For example, “The exercise of political power”
3. Adding learning goals. For example, “Distinguish theories of sovereignty”
4. Adding precedence links among goals. For example, “Identify limits of powers separation” is a prerequisite for “Describe relativity of separation”
5. Adding learning resources by goal. For example, a document, web link, video, etc.
6. Adding learning activities with specifying participation rate of associated goals
7. Generation of knowledge structure and learning paths.

To facilitate course editing, we designed an Excel workbook² that allows teachers to enter all course data. The workbook contains the course information cited above. Its content is exported to CSV (Comma-Separated Values) files and imported into the system. The teacher must generate knowledge structure and learning paths (Fig. 5).



The screenshot shows a web interface for editing a course. On the left, there is a sidebar with buttons: 'Begin' (green), 'Edit' (blue), 'Learn' (blue), and 'Tutoring' (blue). The main area displays a table titled 'ORGANISATION OF DA1234 - INFORMATIQUE 1'. The table has columns: Order, Short name, Full name, Time min, Time max, # Objectives, # Resources, and # Activities. It lists four learning units (U1 to U4) with their respective durations and counts.

Order	Short name	Full name	Time min	Time max	# Objectives	# Resources	# Activities
1	U1	De l'information au système Informatique	2 days 2 hours 24 minutes	3 days	8	12	13
2	U2	Système d'exploitation : Microsoft Windows	3 days 12 hours	5 days	8	12	25
3	U3	Traitement de texte : Microsoft Word	2 days 19 hours 12 minutes	4 days	10	20	14
4	U4	Présentation Assistée par Ordinateur : Microsoft Powerpoint	1 day 9 hours 36 minutes	2 days	4	10	4

Fig. 5. Interface after editing (importing) the educational structure of course

Our system allows the teacher to focus on only one goal at the time during content producing. The organisation of contents is done by system. It is easy to reuse this content in another course. The system ensures that all goals have content and are evaluated. The distribution of learning time by goal allows the teacher to better estimate the workload of learners.

² <https://drive.google.com/file/d/1jVVAIQecZQgiKsaiJ6yUOZymBG9qSQkHo>.

5 Assessing the Acceptability of the Model by Teachers

Before proposing the model to the teachers for designing their courses, we wanted to assess its acceptability and teachers' willingness to use it. For this, a survey³ was submitted to university teachers on the following aspects:

- **Educational productions:** self-assessment of the quality of their courses in terms of (1) structuring, (2) content and evaluation according to the goals of course and (3) organization of the course notions.
- **Interest in customizing the students' learning progress**
- **Resistance to change in teaching method**

The survey has been sent to all teachers and tutors⁴ of our university. Out of 125 persons contacted, we have received $N = 64$ answers⁵. The participants were from 16 departments of university, their age varying between under 25 to over 60 years ($M = 39.25$, $SD = 7.99$) and their teaching seniority varying from less than 2 years to more than 30 years ($M = 10.26$, $SD = 6.64$).

Participants reported that their pedagogical productions are organized mostly in chapters (78.70%) and often in part (23.40%), title (21.30%) and other (4.20%). Nearly 25% of participants believed that certain learning goals have no learning resources clearly identifiable by learners to acquire knowledge. It is also true for exercises used to evaluate acquired knowledge. This confirms the interest to justify association to each goal, resources and exercises to better structure the teaching and facilitate learning.

Regardless of their seniority, 83% of teachers believed that they did not have the best educational scenario. As a result, we believe that the best scenario will depend on the learners since they are the main beneficiary of the teaching.

To follow learners in their chosen scenario, nearly 90% of teachers declared they were ready to cater to learners late in their learning and 55% were willing to follow learners progressing faster in their learning. 63% of them declared being willing to spend some time to help learners outside of the scheduled sessions. Those results confirm the validity of our approach to give opportunity to all learners to finish their learning considering their cognitive characteristics.

Surprisingly, as we expected many teachers to be reluctant to changing their teaching method, 80% of teachers thought it was better to organize the teaching or learning by educational goal, instead of by chapters or parts as usual. More than 90% of them estimated that exercises should be classified by goal to facilitate their resolution. Nearly 80% found that assessment by goal would be better than assessment by period (generally at the end of a chapter, part or even semester).

³ <https://goo.gl/forms/ne1Uua4UeYPW3EeO2>.

⁴ Person responsible of the educational follow-up in the online training platform.

⁵ Consulted at 11-24-2017.

6 Assessing the Usability of the System

To evaluate the usability of EGbKST and its underlying model, we organized a workshop to use our system, attended by 16 teachers from 3 higher education institutions and 8 specialties (Economics, Geography, Management, Computer Science, Applied Foreign Languages, Modern Letters, Management and Law Sciences). We had thus the opportunity to test our model in different domains and therefore to validate the genericity of our approach.

6.1 Methodology

The experiment was organized into 4 parts. In part 1, we presented to the participants the previous survey to get their opinion before the experiment. In the part 2, we asked participants to interpret the result of survey and to criticize the current model according to them. We then exposed the need to improve the pedagogical model. The improvement focused on the possibility of having several scenarios in a course to consider learners' specificities. The concepts of our model were presented to allow them to understand their logic and usefulness. In part 3, participants had to redesign their own courses according to model. This was done through the Excel workbook designed for this purpose. At each stage, we explained to the participants the expected results. Participants' productions were presented to all assistance for verification and improvement. Part 4 of the experiment dealt with another survey⁶ (a posteriori) to collect the appreciation of model and difficulties of implementation.

The experiment was focused more on the pedagogical part (production of course) than the technological part (implementing course on Moodle platform).

6.2 Results

Pedagogically, our model allowed to detect in educational productions some knowledge taught before their prerequisites. These imperfections have been corrected using precedence relationships established between goals.

The graph of prerequisite among goals showed that many courses have several educational goals without or with only few prerequisites. For example, Table 2 shows among the 16 courses currently in production, the teaching unit "Constitutional Law 2" taught in the first year of the Bachelor's degree in Business Law. The teaching is structured in 2 lessons in which the first has 7 goals and the second 11. Figure 6 illustrates the low number of prerequisites between goals found for that course.

In many cases, the teachers realized that they forced a pedagogical scenario although several other scenarios were just as valid. In the example on Fig. 6, we realize that on the 18 goals, 9 have no prerequisite. A learner can begin learning by the end of the course (according the teacher's scenario) without risk.

⁶ <https://goo.gl/forms/eSaZjajB2x744RdQ2>.

Table 2. The learning goals of “Constitutional Law 2” course

Lesson 1	Constitutional organization of democratic power
Goal 101:	Distinguish theories of sovereignty
Goal 102:	Describe institutional consequences of democratic theories of sovereignty
Goal 103:	Describe perverse effects of national sovereignty
Goal 104:	Interpret utopia of popular sovereignty
Goal 105:	Explain the amalgam of democratic theories of sovereignty
Goal 106:	Recognize the main voting methods
Goal 107:	Explain the political implications of the main voting methods
Lesson 2	The exercise of political power
Goal 201:	Distinguish theories of the powers separation
Goal 202:	Explain objectives of the powers separation
Goal 203:	Determine the fundamental principles of powers separation
Goal 204:	Identify limits of powers separation
Goal 205:	Identify different political regimes
Goal 206:	Describe particularity of parliamentary regime
Goal 207:	Summarize origin of the parliamentary system
Goal 208:	Discuss characteristics of the parliamentary system
Goal 209:	Describe relativity of separation
Goal 210:	Name characteristics of presidential regime
Goal 211:	Interpreting complexity of political regimes application in Africa

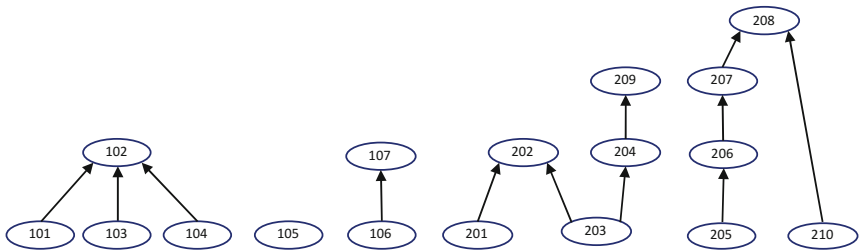


Fig. 6. Graph of prerequisite goals of “Constitutional Law 2” teaching unit

In the final survey, 100% of the teachers declared appreciating our approach to better structure the lessons and organize evaluations, but 42.8% declared using our model was difficult and would have needed more time to experiment. The main difficulty was to index educational resources and activities by educational goals.

7 Conclusion

Designing a learning scenario (by teacher) is a time-consuming activity, making it difficult for teachers to build several scenarios. However, learners with different characteristics may have difficulties to follow the unique scenario defined by teacher.

We have therefore proposed a model to design multi-scenario in courses based on prerequisite relationships between educational goals. Our model allows to design easily several scenarios without knowing in advance the specific characteristics of each learner.

From an experiment made with our system, the teachers have both detected contradictions contained in their productions and realized that several goals of their courses were not related to others. These findings have led some teachers to review their course design and to generate new scenarios. Most teachers realized that learning assessment does not cover all goals. The proposed activities cover even very few goals and generally focus on case studies. But to do this kind of activity, it is necessary to make sure that learners have really acquired the basic competences. This is done through particular activities defined around one skill or learning goal. Our experience was inconsistent in the choice of participants because their competence in e-learning was very different.

In future works, we intend to integrate into the model and the tool, an analysis of the learning scenarios chosen by learners that we will present to the teachers. This analysis will probably allow them to detect hidden dependencies. On the other hand, the quality and the achieving time of learner's results will make possible to better set learning durations for goals and to have a more accurate feedback on the effective implementation of their learning scenarios. To allow teacher to follow the learners in their different scenarios, we will design the coaching scenario.

References

1. Greer, J.E., McCalla, G.I.: Student Modelling: The Key to Individualized Knowledge-Based Instruction. Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-662-03037-0>
2. Psotka, J., Massey, L.D., Mutter, S.A. (eds.): Intelligent Tutoring Systems: Lessons Learned. Lawrence Erlbaum, Hillsdale (1988)
3. Carron, T., Marty, J.-C., Heraud, J.-M., France, L.: Helping the teacher to re-organize tasks in a collaborative learning activity: an agent-based approach. In: Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies, Kerkrade, The Netherlands, pp. 552–554 (2006)
4. Lefevre, M., Jean-Daubias, S., Guin, N.: Personnaliser des séquences de travail à partir de profils d'apprenants. In: Environnements Informatiques pour l'Apprentissage Humain, Le Mans (2009)
5. Monterrat, B., Yessad, A., Bouchet, F., Lavoué, É., Luengo, V.: MAGAM: a multi-aspect generic adaptation model for learning environments. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) EC-TEL 2017. LNCS, vol. 10474, pp. 139–152. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66610-5_11
6. Mbatchou, G.M.: Accompagnement de l'apprentissage/formation : modélisation du scénario pédagogique. In: Iksal, S., Michel, C., Pelissier, C. (eds.) 6èmes Rencontres Jeunes Chercheurs en Environnements Informatiques pour l'Apprentissage Humain, pp. 95–100. Association des Technologies de l'Information pour l'Education et la Formation (ATIEF), Montpellier (2016)

7. Quintin, J.-J., Depover, C., Degache, C.: Le rôle du scénario pédagogique dans l'analyse d'une formation à distance Analyse d'un scénario pédagogique à partir d'éléments de caractérisation définis. Le cas de la formation Galanet. In: EIAH 2005, Montpellier, France (2005)
8. Paquette, G.: L'ingénierie pédagogique : pour construire l'apprentissage en réseau. Presses de l'Université du Québec (2002)
9. Dalziel, J.: Using LAMS version 2 for a game-based learning design. *J. Interact. Media Educ.* **2008**(2), Art. 22 (2008). <https://doi.org/10.5334/2008-24>
10. Emin, V., Pernin, J.-P., Guéraud, V.: Scénarisation pédagogique dirigée par les intentions. *Rev. STICEF* **18** (2011). ISSN 1764-7223. http://sticef.univ-lemans.fr/num/vol2011/01-emin-tice/sticef_2011_tice_emin_01.htm
11. Mariais, C.: Modèles pour la conception de Learning Role-Playing Games en formation professionnelle (2012). <https://tel.archives-ouvertes.fr/tel-00702237>
12. Brassard, C., Daele, A.: Un outil réflexif pour concevoir un scénario pédagogique intégrant les TIC. Presented at the Environnements Informatiques pour l'Apprentissage Humain 2003 April 17 (2003)
13. Riad, B., Mourad, H., Nourredine, G., Hamid, S.: The scenarization: a new task for teachers. *Procedia - Soc. Behav. Sci.* **31**, 732–737 (2012)
14. Marne, B., Labat, J.-M.: Model and authoring tool to help adapt serious games to their educational contexts. *IJLT* **9**, 161–180 (2014)
15. Falmagne, J.-C., Cosyn, E., Doignon, J.-P., Thiéry, N.: The assessment of knowledge, in theory and in practice. In: Missaoui, R., Schmidt, J. (eds.) *ICFCA 2006. LNCS (LNAI)*, vol. 3874, pp. 61–79. Springer, Heidelberg (2006). https://doi.org/10.1007/11671404_4
16. Heller, J., Steiner, C., Hockemeyer, C., Albert, D.: Competence-based knowledge structures for personalised learning. *Int. J. E-Learn.* **5**, 75–88 (2006)
17. Breen, D., Shorten, G., Aboulafia, A., Zhang, D., Hockemeyer, C., Albert, D.: Defining a competency map for a practical skill. *Clin. Teach.* **11**, 531–536 (2014)
18. Steiner, C.M., Albert, D.: Competence-based knowledge space theory as a framework for intelligent metacognitive scaffolding. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) *AIED 2011. LNCS (LNAI)*, vol. 6738, pp. 563–565. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21869-9_102
19. Falmagne, J.-C., Albert, D., Doble, C., Eppstein, D., Hu, X. (eds.): *Knowledge Spaces: Applications in Education*. Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-35329-1>
20. Melero, J., El-Kechaï, N., Labat, J.-M.: comparing two CbKST approaches for adapting learning paths in serious games. In: Conole, G., Klobučar, T., Rensing, C., Konert, J., Lavoué, É. (eds.) *EC-TEL 2015. LNCS*, vol. 9307, pp. 211–224. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24258-3_16
21. Marfisi-Schottman, I., Labat, J.-M., Carron, T.: Building on the case teaching method to generate learning games relevant to numerous educational fields. In: *International Conference on Advanced Learning Technologies, ICALT*, Beijing, China, pp. 156–160 (2013)



RailMon: Distance, Temperature and Location Railway Monitoring Using IoT Technologies

Pa Saffiong Kebbeh^{1(✉)}, Madoune R. Seye², Bassirou Ngom², Bamba Gueye²,
and Moussa Diallo²

¹ University of The Gambia, Banjul, The Gambia
pakebbeh@utg.edu.gm

² Universite Cheikh Anta Diop, Dakar, Senegal

Abstract. Internet of Things (IoT) is an emerging technology and have great potential to be applied in critical environments. In that regard, IoT is a remarkable solution to the challenge of collecting data from physical environments, thanks to their flexibility, low cost and ease of deployment. It has always been a dependent factor for early alert signals for the possibility of unforeseen technical issues that may occur within the mass transport networks. Our proposed infrastructure, called “Rail-Mon” and based on IoT, is able to detect either rail track damages due to unbearable temperature or determine fixed object in a far-away distance in a railway network. Thus, an early warning system is deployed to avoid unforeseen fatal accidents.

Keywords: Internet of Things · Event detection
Structural health monitoring

1 Introduction

Since the evolution of wireless sensor networks can be a dependent technology for safety mass transport monitoring, introducing monitoring techniques in its different forms (damage detection, traffic monitoring, reliability assessment, etc.) can save costs by improving the reliability and efficiency of mass rail transport infrastructures [1, 2]. Indeed, granting the possibility to convey early warnings of problems that may develop within the mass transport networks.

In Senegal, there are two main critical issues of the railway network. Firstly, the lack of modern security monitoring system of the railway infrastructure. For example, the untimely crossing of the train route by the pedestrians, vehicles, and the uncontrolled movement of livestock within the local neighborhood can cause a high risk of train obstructions and accidents during movement. Currently, there exist no automatic detection mechanism of either a fast approaching train towards a pedestrian intersection crossing, or an unattended fixed object on the railway line.

Secondly, the men force who work as railway guards and track engineers have no technology instrument to make real-time and accurate assessment of rail track damages due to unbearable temperature on the tracks, or a determination of a fixed object in a far-away distance within the railway perimeter area. As a result of the consequence of this challenge, this paper presents “RailMon”, a wireless sensor network model for the continuous monitoring and detection of obstructive materials within the rail network using Wireless Sensor Networks (WSN) [3]. In our approach, we aim to deploy, analyze and evaluate the use of wireless sensor networks for a reliable geo-location data collection of object detection, monitor a moving train’s location, and temperature measurement of the rail track infrastructure. This study was conducted in Dakar, using the rail network infrastructure, called “Petit Train Bleu” which serves Dakar to Rufisque. It should be noted that our proposition is also suitable for the Dakar-Bamako railway network. Principally, “RailMon” architecture focuses on practical engineering solutions, where “Dragino” IoT sensor devices [4] are used for their respective strengths to achieve the desired solutions of our study to in order avoid unforeseen fatal accidents.

The layout of the paper is presented as follows. Section 2 discusses the related work. Section 3 describes the “RailMon” deployed architecture. Section 4 highlights the experimental setup and technical specifications. Section 5 analyses the achieved results. Finally, Sect. 6 draws the conclusions and possible future work for a more scalable railway monitoring technique.

2 Related Work

Railway structures are design in various forms based on environment, infrastructure design, model requirement, and core needs. According to the authors of [1, 2], rail tracks are made of steel and can expand when temperature constantly rises to a certain degree, meaning they get longer and deform in shape. An 1800-foot length of rail will expand almost one foot with an 80-degree change in temperature. Heat-related expansion places a lot of stress on the ties and eventually, the tracks will buckle under the force. While the air temperature might be 31 °C, the transport rail corps engineering standards and procedure publications [1] proofs that temperatures on tracks can exceed 50 °C during heat waves. According to the survey of intelligent transport systems [5], different experimental setups of WSN purposely for railway infrastructure monitoring were experimented. In their analysis, the authors proposed a real-time rail condition monitoring by using microcontrollers as sensor nodes, which transmits to base station using either GSM or GPRS as a communication medium. While the authors provide a generic architecture of a typical monitoring infrastructure using WSN where data is produced continuously or periodically, the paper did not address the specifics of which communication medium best provides the most effective transmission cost. In a similar deployment of railway monitoring systems, recent works [6, 7] investigated and proposed an “LED-LDR” based design using an Arduino MCN, GPS and a GSM module to collect data and send it

to a mobile phone for analysis. While the use of this technique was found to produce very accurate results in lab based testing, data transmission from the remote monitored location could be both expensive, and possibly be compromised by intermittent GSM signals or poor network coverage. Additionally, the author's approach of using DC Motors for distance traversing on the rail tracks, is not only uneconomical but could also be obstructive for a moving train. In an evaluation of long range wireless transmission sensor devices, the authors of [8] conducted a survey on the evaluation of the LoRa (Long Range) wireless sensor antenna transmission coverage within the Dakar Peninsula. While the authors evaluated their most optimized results within a combined distance of over 40 km in between 4 base stations, it was concluded that their evaluated results proofs that a LoRaWAN gateway can cover a transmission range of up to 10 km with a packet loss ratio less than 30% through a line of sight.

3 RailMon: Design Overview

The model of our architecture as compared to various related works provides continuous real-time data acquisition and improves data accessibility, with the capability of simultaneously collecting and processing data from combined variation of sensors in an integrated frame. While other authors use wireless sensor networks for real time monitoring of either event detection, temperature monitoring or malicious object detection at different times in a discrete method, the core competence of our deployment model is the capability of providing real time information of both the temperature data and information of a possible malicious object within the railway line, as well as the geo-location data of a moving train towards an endangered location instantaneously. Using sensor network mechanism, both temperature data and the location of a non-moving malicious object within the railway line can be analyzed and interpreted at the command center for further decision making.

3.1 RailMon Architecture

Figure 1 depicts the RailMon alert system that primarily encompasses a sensor network which monitors the railway track for deformation and obstruction before the train reaches the impending danger zone, and a module that monitors the geo-location of the moving train in order to avoid any impending disaster. With the RailMon monitoring mechanism, the following tasks were being measured and evaluated:

- (a) Measure the distance between a detected object within the parallel railway tracks using an ultrasonic sensor, measure the degree of temperature level at given sensor node locations; and subsequently send the combined data to a remote command center as illustrated in Fig. 1.
- (b) Monitor the location of a moving train in the form of geo-location coordinates including date and time stamp, through a GPS and consequently send the data to the command center in a separate frame from (a) above, for any further determination.

- (c) This technology solution is designed with relevance and effective low power consumption with long-range communication capability by means of aggregating both temperature and distance data in a single frame along the way to the sink.

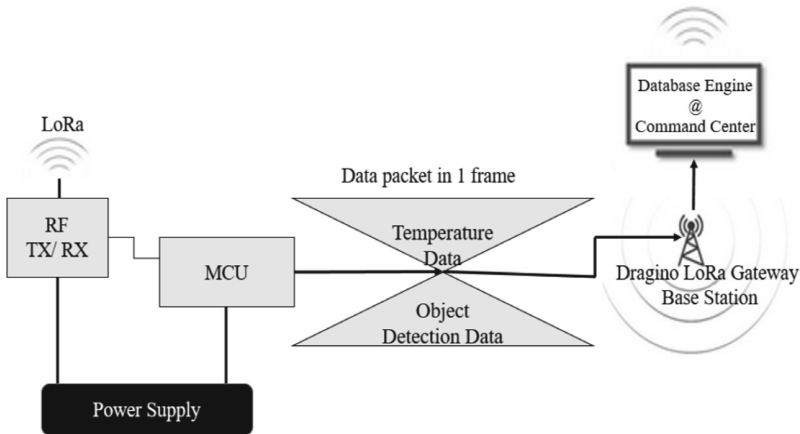


Fig. 1. Block diagram of the DTL WSN

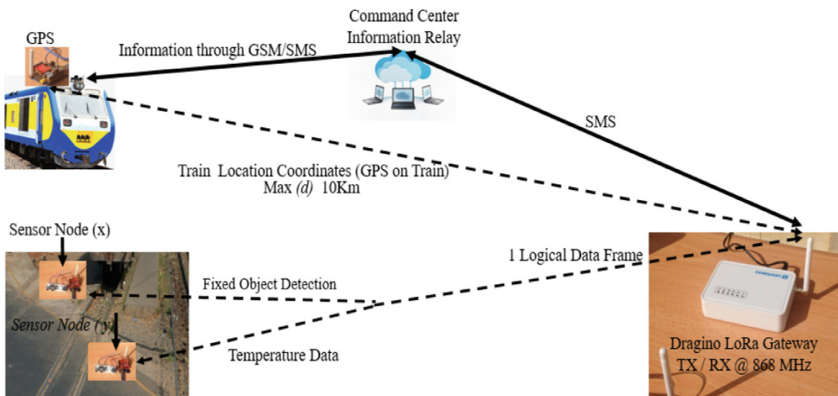


Fig. 2. RailMon architecture

Figure 2 illustrates the deployed RailMon architecture. While sensors “*x*” or “*y*” detects different scenarios (i.e., temperature and object depending on its location), each node can sense both scenarios within its range, and send it to the gateway in 1 logical frame.

3.2 Occurrence Detection in RailMon

Based on Fig. 2, sensor nodes are used to collect data from the environment and send it to a sink node or gateway using a broadcast messaging architecture. Irrespective of the fact that the data transmission device that we used in our deployment transmits data in a broadcast mode by default, (where the transmitted data can be received by similar devices operating on the same frequency), in our architecture, we programmed the sensor nodes and the gateway to transmit and receive on frequency 868 MHz. In our adopted method, data is collected from outlying nodes through a direct spanning tree to the gateway. Typically, some function is applied to the incoming data from each node to identify its location and other related attributes. The goal being that eventually, the gateway will obtain data only from nodes that are associated to the setup.

The approach is often called stepwise refinement:

- i. Initially, all nodes send input to the gateway @ every 1 min upon receiving Data (d) from Nodes (x or y):
- ii. Upon the gateway receiving Data (d) from Nodes (x, or y),
- iii. The sink gateway then forwards the data to the Database
- iv. Append (d,) to the database.

With the above algorithm, if the base station (sink node) is interested in finding the maximum or average temperature in a region, each microcontroller node monitoring the region can easily aggregate the data generated by its sensors and simply send one frame containing the result. The attributes of such frame are “ma” (temperature), “mm” (distance) and “addr” (Location). For the fact that the GPS is in on different microcontroller board on-board a moving train, the GPS data can also be aggregated by collecting coordinates, time stamp and date in a separate frame. Given a set of sensor nodes $S = S_x, S_y, \dots, S_{n-1}$ in the network, with S_{n-1} being the sink node as the gateway, where each node has a data item that it wants to send to the sink node, this implies that the we had to programmed the sensors to transmit at every 1 min but strictly on time slots to avoid bottleneck and interference. As the monitoring network scales up, this algorithm can dramatically reduce the total number and size of packets sent, because each node sends 1 frame and the total number of packets sent is always equal to $N - 1$ for every matured slot.

4 Experimental Setup

There are several sensors types used in railway condition monitoring for analyzing different aspects of the rail infrastructure. According to the proposed topology in [9], the authors setup a single-hop network architecture with application of a converge cast messaging algorithm. For RailMon-based approach, the gateway is positioned in the center of the sensor network where data is transmitted directly to the gateway through a single hop mode. The rail track measurement kit comprises of an Arduino microcontroller board which embeds several sensor kits as specified in Table 1.

4.1 Evaluation of LoRa Transmission Protocol

We compared two contending wireless IoT transmission devices; being ZigBee and LoRaWan. Based on our investigations, it was prudent that after an in-depth comparison between the two technologies. Indeed, LoRaWAN is more suited for long range transmission with very low power consumption, and long-lasting battery operated autonomy.

Table 1. Deployed sensor kits and components for RailMon

Measurement kits		
Kit description	Specifications	
Dragino LoRa shield	Microcontroller (Arduino)	ATmega328
	Operating voltage	5 V
	Digital I/O pins	14 (6 provide PWM output)
	Analog input pins	6
GPS sensor	Based on MT3339	
	Programmable bit rate up to 300 kbps	
	Frequency band: 868 MHZ/433 MHZ/915 MHZ	
Dragino gateway	Open Source OpenWrt system	
	Auto-provisioning built-in web server managed by web GUI	
	Internet connection via LAN, WiFi, 3G or 4G	
DHT11 temperature sensor	Detects surrounding environment of the humidity and temperature	
	Humidity measurement range: 20%–90%	
	Temp Measurement range: 0 °C–60 °C	
HC-SR04 ultrasonic sensor	Power supply: +5 V DC	
	Ranging distance: 2 cm–400 cm/1''–13 ft	
	Resolution: 0.3 cm, measuring angle: 0°	
Batteries	5 V batteries/mobile power bank	

Both LoRa and ZigBee can operate on similar frequency bands of 868 MHz, and the latter being more efficient in short range transmission. The LoRaWAN is especially more efficient for use-cases where we have more uplink (device sending to cloud or gateway) updates than downlink, e.g. long distance condition monitoring. Thus, justifying our need of choosing LoRa as our data transmission mechanism. As illustrated in Fig. 2, the network is typically laid out in a star topology, where the sensor nodes transmit directly to the gateway that in turn, is connected to a the database engine via standard Internet technologies.

Furthermore, Fig. 1 depicts our measurement kit which is formed by a micro-controller and a couple of sensor components. The components are depicted as follows:

- **GPS:** L80 GPS (base on MTK MT3339) is designed for applications that use a GPS connected via the serial ports to the Arduino such as timing applications that require GPS information. The module can calculate and predict orbits automatically using the ephemeris data (up to 3 days) stored in internal flash memory, with automatic antenna switching function.
- **Dragino Gateway:** LG01 LoRa Gateway allows bridging LoRa wireless network to an IP network based on WiFi, Ethernet, 3G or 4G cellular [4]. LG01 runs an open source embedded Linux system with full Ethernet and 802.11 b/g/n WiFi capabilities and can process/send to IoT server in an IP network.
- **DHT11 Temperature Sensor:** DHT11 is a low cost humidity and temperature sensor, which generates calibrated digital output and can be interface with any microcontroller like Arduino, Raspberry Pi, etc.
- **Ultrasonic Sensor:** HC-SR04 ultrasonic sensor module offers excellent non-contact range detection with high accuracy and stable readings in an easy-to-use package. From 2 cm to 400 cm or 1" to 13 ft, its operation is not affected by sunlight or black material like Sharp rangefinders.

4.2 Sensors Network Setup

According to our monitoring setup, wireless sensors are placed along the railway line at a distance of 4 m away from the rail track in order to monitor both the pre-defined temperature and object obstruction of the tracks. The microcontroller board (Fig. 3) enables the sensors to react to both an ultra-object detected within a range of 4 m away from the rail track and a pre-defined temperature range of 0 °C to 60 °C. The data collected is transmitted by LoRa to the gateway @a TX/RX frequency of 868 MHz. When data is sent to the gateway, it includes the sensor node location coordinates and the error type detected.

Unlike basic condition monitoring where information is triggered to the command center based on only abnormal event detection, with our RailMon, [10] data from sensor nodes are treated as time series, i.e. either periodically or continuously. For example, at every sensor node, data is generated and sent to the gateway every 1 min as a condition of no critical error. Otherwise, data will be generated and sent every 5 s due to a constant error of either a malicious fixed object detected within the perimeter of the railway lines, or a measured temperature of more than 60 °C is being realized, which is enough to deform a railway track.

To achieve our desired solutions, we formulate a data frame structure in order to identify and interpret the received data from fixed sensor nodes. The frame packet as illustrated in Table 2 is associated with the data types, (as implicit to a condition mode).

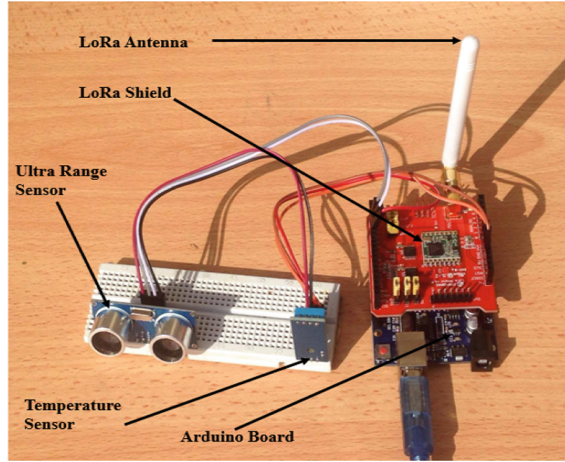


Fig. 3. Mico-controller and sensor components setup

4.3 Analysis of the Error Codes

Depending on a given circumstance of a monitored location, the data frame that is being sent to the command center is segmented into different blocks, as illustrated in (Table 2). The first X is denoted for a segment of a data frame which describes the error type of the detected anomaly within the railway line. The error codes ranges from type 0 to 3. With the interpretation of the frame segments, error type 0 implies that there exist no fault or critical issue for action. In scenario 1, knowing the maximum threshold of the DHT11 temperature sensor coupled with the fact that temperature on tracks can exceed 50°C during heat waves, therefore, if a temperature measurement is equal to or surpass the most extreme temperature of 60°C , then an error code of type 1 is generated continuously for every 5s which is transmitted to the command center for an alert of a possible “sun kinks”, with high risk of causing railway accident. In this scenario, we set the temperature threshold to 35°C in order to proof our use case and generate an error code, since it was impractical to witness a 60°C temperature under normal circumstance.

```
The frame is: 0:1:1362.72:35.00
The frame is: 0:1:33754.35:35.00
The frame is: 0:1:33903.78:35.00
```

Fig. 4. An illustration of a captured data frame sent to gateway with no errors

In scenario 2, at a monitored location, while the temperature measurement may still be within a normal range, there could exist a situation of a constant fixed object within a distance of 0–4 m, thus, causing a possible derailment of a

```

The frame is: 1:1:1440.24:37.00
1
Data Sent
The frame is: 1:1:1004.87:37.00
1
Data Sent
The frame is: 1:1:1427.83:37.00
1
Data Sent

```

Fig. 5. An illustration of a captured data frame sent to the gateway with error type 1

speeding train. In such a situation, an error code type 2 is constantly transmitted to the command center, until the object is moves by default or physically removed. In scenario 3, at a monitored location, if both sensors (temperature and ultra-range) detects an anomaly outside the pre-defined variable parameters of the sensing function, then error code type 3 is generated and transmitted to the command center for an urgent reaction. Which implies that there is an issue with both temperature measurement and the presence of a fixed object on the railway tracks. It worth noticing that the “X” value labeled in Table 1 means a segment of the data frame illustrated in Figs. 4, 5, 6 and 7.

```

The frame is: 2:1:93.67:38.00
2
Data Sent
The frame is: 2:1:77.35:38.00
2
Data Sent
The frame is: 2:1:92.31:38.00
2
Data Sent

```

Fig. 6. An illustration of a captured data frame sent to the gateway with error type 2

```

The frame is: 3:1:969.68:38.00
3
Data Sent
The frame is: 3:1:1128.46:38.00
3
Data Sent
The frame is: 0:1:33867.40:38.00
The frame is: 0:1:33835.78:38.00

```

Fig. 7. An illustration of a captured data frame sent to the gateway with error type 3

Table 2. Data packet frame interpretation

Interpretation of the data packet frame TX/RV at the gateway from sensor nodes			
X:	X:	XXXX.XX:	XX.XX
The first block of the frame is the description of the error type:	The second block of the frame is the identifier (Location Address) of a sensor node	The third block of the frame is the description of the range or distance of an object away from the rail track (up to a max distance of 4 m	The fourth block of the frame is the description of location temperature of the monitored area

4.4 Data Transmission Architecture

Since our architecture of data transmission is through a wireless signal, it is evident that in wireless data transmission, attenuation due to unpredictable climate conditions may also occur at certain times of the year. Our “RailMon” model assumes the most ideal propagation conditions of clear Line-Of-Sight (LOS) between the transmitting sensor nodes and receiving gateway. Thus, not all packet frames transmitted from a given node may all the time be successful to the gateway.

With the gateway located in the center of the network, all the distributed sensor nodes are positioned to send data directly to the gateway in a single hop transmission. Our data transmission mechanism from the sensor nodes to the gateway relates the effectiveness of our architecture to an energy efficient (EE) routing protocol [11]. In the architecture, much delayed packets or slow packets are removed from the queue to the gateway as it is impractical to traverse those packets to the destination, thereby saving the energy of nodes. With this procedure, the routing protocol calculates the expected delay for the current packet to reach the destination and decides whether to remove or not; as illustrated in the below formula for the current packet p at the current node x to reach the destination $d(Txd)$ is given by the formula:

$$T_{sx}(p) = D_x d(p) - D_{sx}(p) \quad (1)$$

As shown in Fig. 1, while the data transmission speed relied on the performance of both the sensor nodes and the gateway, $D_{sx}(P)$ is the distance already covered by the packet p from node x or y , to the gateway. $D_x d(p)$ denotes the remaining terrestrial distance that the current packet p from sensor node x or y should span to the destination d , after the several repeated attempts of trials to reach the gateway. Therefore, $T_{sx}(p)$ gives the delay for the packet to reach to the gateway, and after 5 attempts, the packet will be discarded and be replaced by a newer real time packet for a reliable informed decision at the gateway.

4.5 Temperature Variation

The variation in temperature depends on the amount of heat exerted or absorbed around sensed zone. The Temperature Formula is given by:

$$\Delta T = \frac{Q}{WA} \quad (2)$$

where:

- ΔT is the temperature difference,
- Q is the amount of heat absorbed or exerted,
- W is the weight of the rail track,
- A is the average heat of the track.

To determine an outside temperature of heat on a rail track mass of 5 kg on an average heat of 0.5/kg °C?

$$\text{Given: Heat absorbed } Q = 100 \text{ J}, \quad (3)$$

$$\text{Weight of rail track } W = 2 \text{ kg}, \quad (4)$$

$$\text{Average Heat } A = 0.5/\text{kg } ^\circ\text{C} \quad (5)$$

$$\text{The temperature is given by } \Delta T = QWA \quad (6)$$

$$= (100 \text{ J})(5 \text{ kg})(0.5/\text{kg } ^\circ\text{C}) \quad (7)$$

$$= 100 ^\circ\text{C}. \quad (8)$$

5 Results

The final results achieved is an autonomous monitoring technique, thus, making it possible for a moving train to avoid reaching an unexpected disruption point within the railroad, We deployed multiple sensors on a single MCU for multiple monitoring functions of fixed locations. The RailMon model is very simple and robust. The failure of one node does not influence the operation of the whole network. This survey proofs that at a given time of the day, while the day temperature rises between 09:30 am to 19:00 pm, mass movement of either people or objects can still be detected within a range of 4 m away from the rail tracks, and the maximum temperature recorded after various sampling test was 38 °C.

6 Conclusion

We reviewed wireless sensor network applications for environmental monitoring and implement an effective WSN with several scenarios experimented. From the study, it has been proven that an alternative method of replacing a conventional routine of using men force in environment monitoring can be substitute with a more pragmatic technique to fulfill such functional requirement by using wireless sensor networks. The proposed “RailMon” architecture based on IoT, is able to

detect either rail track damages due to unbearable temperature or determine fixed object in a far-away distance within track zone.

We plan to deploy and analyze multi sink base stations as cluster heads for coordinated processing in larger network environments coupled with more robust efficient and effective routing protocol.

Acknowledgement. The authors gratefully acknowledges the World Bank - Africa Center of Excellence STEM project, in collaboration with the Ministry of Higher Education, Research, Science and Technology of The Government of The Gambia and the University of The Gambia for supporting this research work.

References

1. Temperton, J.: Why does hot weather make railways bend (2015). <http://www.wired.co.uk/article/uk-heatwave-rails-buckle>
2. Kerr, M., Wilson, A.: Engineering Manual Track, TMC 226, Rail Defects Handbook, Version 1.2. Transport Rail Corps Engineering Standards and Procedure Publications, June 2012
3. Othman, M.F., Shazali, K.: Wireless sensor network applications: environment monitoring system. In: International Symposium on Robotics and Intelligent Sensors (2012)
4. Tindie. <https://www.tindie.com/products/edwin/lora-iot-development-kit/>
5. Hodge, V.J., O’Keefe, S., Weeks, M., Moulds, A.: Wireless sensor networks for condition monitoring in the railway industry: a survey. *IEEE Trans. Intell. Transp. Syst.* **16**(3), 1088–1106 (2015)
6. Somalraju, S., Murali, V., Saha, G., Vaidehi, V.: Robust Railway Crack Detection Scheme (RRCDS) using LED-LDR assembly. In: ICRTIT 2012. IEEE (2012). ISBN 978-1-4673-1601-9/12/\$31.00
7. di Scalea, L., et al.: Non-contact ultrasonic inspection of rails and signal processing for automatic defect detection and classification. *NDT Cond. Monit.* **47**(6), 346–353 (2005). Special Issue on NDT of Rails
8. Seye, M.R., Gueye, B., Diallo, M.: An evaluation of LoRa coverage in Dakar Peninsula. In: IEEE IEMCON 2017 (2017)
9. Abhisekh Jain, S., Arvind, S., Balaji, B.S., Ram Viyas, N.P.: Onboard dynamic rail track safety monitoring system. In: International Conference on Advanced Communication Systems, ICACS 2007, Madurai, India, 10–12 January 2007. GCT, Coimbatore (2007)
10. <http://www.rfwireless-world.com/Terminology/LoRa-vs-Zigbee.html>
11. Wixted, A.J., Kinnaird, P., Larijani, H., Tait, A., Ahmadinia, A., Strachan, N.: Evaluation of LoRA and LoRAWAN for wireless sensor networks. In: 2016 IEEE, Glasgow (2016). 978-1-4799-8287-5/16/\$31.00



An Encoding for the Theta Model of Elliptic Curves

Nafissatou Diarra^{1(✉)} and Emmanuel Fouotsa²

¹ Cheikh Anta Diop University, Dakar, Senegal
fifiramatou@gmail.com

² The University of Bamenda, Bamenda, Cameroon
emmanuel Fouotsa@yahoo.fr

Abstract. The use of elliptic curves in cryptography requires to be able to transform an information (generally a bit string) to a point of an elliptic curve. This transformation, called encoding, must be such that the encoded message can be easily and uniquely recovered from the corresponding point. In this paper we propose a new encoding that maps an element of \mathbb{F}_q to a point on the theta model for elliptic curves $E_\lambda : 1 + x^2 + y^2 + x^2y^2 = \lambda^2xy$ recently introduced in [9]. In particular, we show that this new encoding is efficiently computable (deterministic and polynomial-time). We also present a Sage software implementation to ensure the correctness of the encoding on this curve.

Keywords: Theta model · Elliptic curves · Deterministic encoding

1 Introduction

Many elliptic curve-based cryptographic schemes require to hash into the group of points of an elliptic curve, such as password-based authentication protocols (SPEKE (Simple Password Exponential Key Exchange), PAK (Password Authenticated Key exchange)), as well as various signature schemes based on the hardness of the DLP (Discrete Logarithm Problem). The main idea for constructing a hash function into elliptic curves is the following: the image of a message (an arbitrary string) m by the hash function F is $F(m) = f(h(m))$, where h is a classical hash function and f is an encoding function that maps a point of \mathbb{F}_q to an element of the curve. But a problem arises: given any elliptic curve E over any finite field \mathbb{F}_q , how to construct, in a *deterministic* way, a non-zero point of the curve? Some authors have proposed algorithms to answer this question. But before 2006, only probabilistic solutions were known. The paper [3] of Boney and Franklin in 2001 was one of the first that required hashing into (supersingular) elliptic curves; in fact, the public key of their identity-based

This work is supported by the Pole of research in Mathematics with Applications to Information Security (PRMAIS, SubSaharan Africa) sponsored by Simons Foundation and LIRIMA-MACISA Project.

encryption is a point on the curve. In 2006, Shallue and Van de Woestjine proposed the first algorithm [15] that maps in a deterministic way an element of \mathbb{F}_q (with odd characteristic) to a point of any elliptic curve over \mathbb{F}_q . Their algorithm is based on the Skalba's equality theorem and requires to compute a square root in \mathbb{F}_q ; but it can constructs only $(q - 4)/8$ points of the curve.

In 2009, Icart defined a new encoding function [13] for Weierstrass form of elliptic curves, based on a very simple idea: intersect the line $y = ux + v$ with the equation of the curve. He showed that his algorithm works in $O \log^3(q)$ operations in \mathbb{F}_q and conjectured (it was proven later by Tibouchi and Fouque [12]) that the size of the image set is approximately $\frac{5}{8}$ of the size of the curve.

Some authors have also proposed constructions of encoding functions for special families of elliptic curves, such as Hessian curves (by Farashahi in [10]), Edwards curves (elligator functions by Bernstein et al. in [1]), or Huff curves (by Diarra *et al.* in [8]).

Our goal on this paper is to continue this line of research, by proposing an encoding function for the theta model for elliptic curves $E_\lambda : 1 + x^2 + y^2 + x^2y^2 = \lambda^2xy$, recently introduced by Fouotsa and Diao [6]. In particular, we will show that this new encoding is efficiently computable (deterministic and polynomial-time).

The rest of the paper is structured as follows: In Sect. 2, we recall a special mathematical concept needed in the work. We briefly define elliptic curves and present the theta model for elliptic curves. together with an overview of main existing encodings into elliptic curves. The Sect. 3 describes our new encoding on the theta model and describes its properties. A numerical example is given with a code written with the Sage software to ensure the correctness of the encoding. We conclude our work in Sect. 4.

2 Preliminaries

2.1 Quadratic Character

Let $p \neq 2$ be a prime and $\mathbb{F}_{p^n} = \mathbb{F}_q$ the finite field of $q = p^n$ elements (where $n \geq 1$ is an integer). An element $a \in \mathbb{F}_q$ is a quadratic residue if there exists $r \in \mathbb{F}_q$ s.t. $a \equiv r^2 \pmod{q}$. We define the quadratic character as follows: $\chi : \mathbb{F}_q \rightarrow \mathbb{F}_q : a \mapsto \chi(a) = a^{(q-1)/2}$; it verifies: $\chi(a) = 1$ if a is a non-zero quadratic residue, $\chi(a) = 0$ if $a = 0$ and $\chi(a) = -1$ otherwise. The following properties are also verified: $\chi(ab) = \chi(a) \cdot \chi(b)$ for any $a, b \in \mathbb{F}_q$; $\chi(a^2) = 1$ for any $a \in \mathbb{F}_q^*$; and if $q \equiv 3 \pmod{4}$, $\chi(-1) = -1$, $\chi(\chi(a)) = \chi(a)$, for any $a \in \mathbb{F}_q$. If $q \equiv 1 \pmod{4}$, then $\chi(-1) = 1$.

2.2 The Theta Model for Elliptic Curves

Elliptic Curves. An elliptic curve E over a field \mathbb{K} is the set of solution in $\mathbb{A}^2(\overline{\mathbb{K}})$ of the equation

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \text{ with } (a_1, a_2, a_3, a_4, a_6) \in \mathbb{K}^5 \quad (1)$$

together with a rational point \mathcal{O} and the condition $\Delta \neq 0$ where $\Delta = -d_2^2d_8 - 8d_4^3 - 27d_6^2 + 9d_2d_4d_6$ with $d_2 = a_1^2 + 4a_2, d_4 = 2a_4 + a_1, d_6 = a_3^2 + 4a_6, d_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2$.

The quantity Δ is called the discriminant of E and the condition $\Delta_E \neq 0$ ensures that the curve E is smooth. In the set of point of an elliptic curve, it is very easy to set an additive group structure using the chord-and-tangent method (see [16] for complete details). When an elliptic curve is defined over a finite field, the resulting group presents a difficult mathematical problem known as the Discrete Logarithm Problem stated as follows: Given a point Q multiple of another given point P , find the integer n such that $Q = nP$. This problem justifies the use of elliptic curves in cryptography for the construction of several secure cryptosystems. The model of elliptic curve given by Eq. (1) is called the Weierstrass model and is the commonly used in the literature. Several other models exists in the literature such as the Edwards model [7], Hessian curves [14], Huff curves [5] and the Jacobi curves [2, 4] including the theta model recently introduced by Fouotsa and Diao [6]. Although these curves are birationally equivalent to each other, in an algorithmic point of view and for security and efficiency purposes, careful choices need to be made on which model of elliptic to use in cryptography.

The Theta Model. This model of elliptic curves was proposed by Fouotsa and Diao [9]. The model is obtained from theta functions and the equation is given as follows $E_\lambda : 1 + x^2 + y^2 + x^2y^2 = \lambda^2xy$. They showed that their model is birationally equivalent to the Weierstrass model $v^2 = u^3 - (1 + c^4)u^2 - 4c^4u + 4c^4(1 + c^4)$. This model enjoys many other properties such as unified formulas (addition and doubling of points use the same formulas) and presents competitive addition formulas over binary fields. More details can be found in [6, 9].

2.3 Existing Encodings for Elliptic Curves

In this section, we give a short overview of existing methods to encode into elliptic curves.

Trivial Encoding: For an elliptic curve $E_{a,b} : y^2 = x^3 + ax + b$ over the field \mathbb{F}_q , the simplest way to construct a point of $E_{a,b}$ from an element of \mathbb{F}_q is to use the trivial encoding, also known as the *try-and-increment* method. The idea is to pick a x -coordinate and try to deduce the y -coordinate by computing a square root: choose a random element $u \in \mathbb{F}_q^*$ and compute $u^3 + au + b$; and then test whether $u^3 + au + b$ is a square in \mathbb{F}_q . If it is the case, then returns

$(x, y) = (u, \pm\sqrt{u^3 + au + b})$ as a point of the curve. Otherwise, one can choose another u in \mathbb{F}_q and try again. But this method has at least one drawback as it cannot *run in constant time*: the number of operations depends on the input u . In practice the input u is the message m we want to hash; thus running this algorithm can allow the attacker to guess some information about m .

Icart's Encoding: Let $q \equiv 2 \pmod{3}$. The map $x \mapsto x^3$ is a bijection and then computation of a cubic root can be done as an exponentiation. In [13], Icart defined a new encoding function, based on the following idea: intersect the line $y = ux + v$ with the Weierstrass curve $E_{a,b} : y^2 = x^3 + ax + b$, with $a, b \in \mathbb{F}_q$. He defined the encoding function:

$$\begin{aligned} f_{a,b} : \mathbb{F}_q &\rightarrow E_{a,b} \\ u &\mapsto f_{a,b}(u) = (x, ux + v) \end{aligned}$$

where $x = (v^2 - b - \frac{u^6}{27})^{1/3} + \frac{u^2}{3}$ and $v = (3a - u^4)/6u$.

As shown in the paper, this function presents many interesting properties. In fact, it can be implemented in polynomial time with $O(\log^3 q)$ operations. The inverse function $f_{a,b}^{-1}$ is also computable in polynomial time. Icart also showed that $|f_{a,b}^{-1}(P)| \leq 4$, given a point P on the elliptic curve.

Other Existing Encodings: There exist many other encodings for special families of elliptic curves, such as supersingular curves (by Boneh and Franklin in [3]), Hessian curves (by Farashahi in [10]), Edwards curves (Elligator functions by Bernstein *et al.* in [1]), Huff curves (by Diarra *et al.* in [8]), etc.

3 A New Encoding for the Theta Model

3.1 The Algorithm

In this section, we propose a deterministic algorithm that given an element r (with additional conditions) of \mathbb{F}_q , constructs a point on $E_\lambda(\mathbb{F}_q)$. From this algorithm, we define the new encoding function which does not cover all points of \mathbb{F}_q , unless we make some additional hypothesis on the underlying field \mathbb{F}_q . Nevertheless, we can send all elements of \mathbb{F}_q that are not in the set of definition (there are at most 6 such points) of the encoding to the point at infinity.

Algorithm 1. Encode-Theta-Model

Input : q be a prime power, $c \in \mathbb{F}_q$ s.t. $c(1-c^4)(1+c^4) \neq 0$, $u \in \mathbb{F}_q$ s.t. $\chi(u) = -1$,
 $r \in \mathcal{R} = \{r \in \mathbb{F}_q : 4ur^2c^6 \neq (1-c^4)(1+3c^4), 4ur^2c^6 \neq$
 $-(1-c^4)^4, (4ur^2c^6)(1+3c^4) \neq (1-c^4)^3\} \subseteq \mathbb{F}_q$

Output: A point (x_λ, y_λ)

$$v = c \cdot \left(\frac{4ur^2c^6 + (1-c^4)(3+c^4)}{4ur^2c^6 + (1-c^4)(-3c^4-1)} \right);$$

$$\varepsilon = \chi((c^2 - v^2)(1 - c^2v^2));$$

$$X = \frac{1}{2} \left((1 + \varepsilon)v + (1 - \varepsilon) \left(\frac{(-c^4 - 1)(v + c) - c(1 - c^4)}{2c^3(v + c) + 1 - c^4} \right) \right);$$

$$Y = -\varepsilon \sqrt{\frac{c^2 - X^2}{1 - c^2X^2}};$$

$$x_\lambda = \frac{X + 1}{X - 1};$$

$$y_\lambda = \frac{Y - 1}{Y + 1};$$

return (x_λ, y_λ) ;

Theorem 1. *The output (x_λ, y_λ) of Algorithm 1 is a point of the curve E_λ : $1 + x^2 + y^2 + x^2y^2 = \lambda^2xy$, where $\lambda^2 = \frac{4(1+c^2)}{1-c^2}$ and $\lambda(\lambda^2 - 4)(\lambda^2 + 1) \neq 0$.*

Proof. 1. v is well-defined from the definition of \mathcal{R} .

2. Let us show that $\varepsilon \neq 0$. Suppose that $\varepsilon = 0 \Leftrightarrow \mathbf{c}^2 = \mathbf{v}^2$ or $\mathbf{c}^2\mathbf{v}^2 = 1$.

(a) $\mathbf{c}^2 = \mathbf{v}^2 \Rightarrow c = \pm v \Rightarrow 4(1-c^4)(1+c^4) = 0$ (impossible by the choice of c) or $4ur^2c^6 = (1-c^4)^2$ (impossible since u is not a square).

(b) $\mathbf{c}^2\mathbf{v}^2 = 1 \Rightarrow cv = \pm 1 \Rightarrow 4ur^2c^6 = (c^2 + 1)^4$ or $4ur^2c^6 = (c - 1)^4(c + 1)^4$; this is impossible since u is not a square.

So $\varepsilon \neq 0$ and then $\varepsilon = \pm 1$.

3. Let us show that X, Y are well-defined. For this, we consider the two cases

$\varepsilon = 1, \varepsilon = -1$ and show that the quantity $\frac{c^2 - X^2}{1 - c^2X^2}$ is a square.

$\varepsilon = 1 \Rightarrow X = v$ and $\chi\left(\frac{c^2 - X^2}{1 - c^2X^2}\right) = \chi\left(\frac{c^2 - v^2}{1 - c^2v^2}\right) = \varepsilon = 1$; so X and Y are well-defined.

$\varepsilon = -1 \Rightarrow X = -\frac{(-c^4 - 1)(v + c) - c(1 - c^4)}{2c^3(v + c) + 1 - c^4}$. Now let $H(X) = \frac{c^2 - X^2}{1 - c^2X^2} = \frac{(c - X)(c + X)}{1 - c^2X^2}$; we want to express $H(X)$ in term of $H(v)$ and use the value

$\chi(H(v)) = \chi\left(\frac{c^2 - v^2}{1 - c^2v^2}\right) = \varepsilon = -1$. Now to have an expression of $H(X)$ in terms of $H(v)$, we compute separately the value of $c - X, c + X$ and $1 - c^2X^2$ and find that:

$[2c^3(v+c) + (1-c^4)](c-X) = (v+c)(c^4-1)$,
 $[2c^3(v+c) + (1-c^4)](c+X) = (v+c)(3c^4+1) + 2c(1-c^4)$,
 and $[2c^3(v+c) + (1-c^4)]^2(1-c^2X^2) = (1-c^2v^2)(c^4-1)^2$. This leads to

$$\begin{aligned}
 H(X) &= \frac{(v+c)(c^4-1) [(v+c)(3c^4+1) + 2c(1-c^4)]}{(1-c^2v^2)(c^4-1)^2} \\
 &= \frac{H(v)}{c-v} \left(\frac{(v+c)(3c^4+1) + 2c(1-c^4)}{c^4-1} \right)
 \end{aligned}$$

Since $(v+c)(3c^4+1) + 2c(1-c^4) = 2c \left(\frac{8ur^2c^6(1+c^4)}{4ur^2c^6 + (1-c^4)(-3c^4-1)} \right)$ and
 $(c-v)(c^4-1) = 4c \left(\frac{(1-c^4)^2(1+c^4)}{4ur^2c^6 + (1-c^4)(-3c^4-1)} \right)$, then we can rewrite
 $H(X)$ as follows:

$$H(X) = H(v) \left(\frac{4ur^2c^6}{(1-c^4)^2} \right)$$

and thus $\chi(H(X)) = \chi(H(v)) \cdot \chi \left[(4ur^2c^6)(1-c^4)^2 \right] = -\chi(u) = 1$. In other words, $\frac{c^2-X^2}{1-c^2X^2}$ is a square and Y is well-defined.

4. Since X and Y are well-defined (for both cases $\varepsilon = 1$ and $\varepsilon = -1$), we can compute $Y^2 = \frac{c^2-X^2}{1-c^2X^2} \Rightarrow X^2 + Y^2 = c^2(1+X^2Y^2)$. Now we have to show that x_λ and y_λ verify the relation $1 + x_\lambda^2 + y_\lambda^2 + x_\lambda^2 y_\lambda^2 - \lambda^2 x_\lambda y_\lambda = 0$, where $\lambda^2 = \frac{4(1+c^2)}{1-c^2}$. In fact, we have:

$$\begin{aligned}
 1 + x_\lambda^2 + y_\lambda^2 + x_\lambda^2 y_\lambda^2 - \lambda^2 x_\lambda y_\lambda &= 1 + \left(\frac{X+1}{X-1} \right)^2 + \left(\frac{Y-1}{Y+1} \right)^2 + \left(\frac{X+1}{X-1} \right)^2 \left(\frac{Y-1}{Y+1} \right)^2 - \lambda^2 \left(\frac{X+1}{X-1} \right) \left(\frac{Y-1}{Y+1} \right) \\
 &= \frac{1}{(X-1)^2(Y+1)^2} \left[2(X^2+1)((Y+1)^2 + (Y-1)^2) - \lambda^2(X^2-1)(Y^2-1) \right] \\
 &= \frac{1}{(X-1)^2(Y+1)^2} \left[4(1+X^2Y^2 + c^2(1+X^2Y^2)) - \frac{4(1+c^2)}{1-c^2} (1+X^2Y^2 - c^2(1+X^2Y^2)) \right] \\
 &= \frac{4}{(1-c^2)(X-1)^2(Y+1)^2} \left[(1-c^2)(1+X^2Y^2 + c^2(1+X^2Y^2)) - (1+c^2)(1+X^2Y^2 - c^2(1+X^2Y^2)) \right] \\
 &= \frac{4}{(1-c^2)(X-1)^2(Y+1)^2} \left[2c^2(1+X^2Y^2) - 2c^2(1+X^2Y^2) - c^4(1+X^2Y^2) + c^4(1+X^2Y^2) \right] \\
 &= 0. \text{ Moreover } \lambda \text{ verifies } \lambda(\lambda^2-4)(\lambda^2+1) \neq 0 \text{ from the conditions on } c.
 \end{aligned}$$

Definition 1. *The encoding function for the theta model for elliptic curves is the function*

$$\begin{aligned}
 f_\lambda : \mathcal{R} \subseteq \mathbb{F}_q &\rightarrow E_\lambda(\mathbb{F}_q) \\
 r &\mapsto (x_\lambda, y_\lambda),
 \end{aligned}$$

where x_λ and y_λ are defined by Algorithm 1. If $r \in \mathbb{F}_q \setminus \mathcal{R}$, we set $f_\lambda(r) = \mathcal{O}_\infty$.

3.2 Size of the Set \mathcal{R}

- Our encoding covers a subset \mathcal{R} of \mathbb{F}_q ; this means that only elements of \mathcal{R} can be encoded. And from the definition of \mathcal{R} , it is easy to see that at most 6 elements of \mathbb{F}_q can not be encoded, that is $\text{card}(\mathbb{F}_q \setminus \mathcal{R}) \leq 6$. Since in practice q (the size of the field) is much greater than 6, thus we can state that our encoding f_λ covers a great proportion of \mathbb{F}_q . For example, for $q = 503$ (see the Appendix for the complete example), we find that $\text{card}(\mathcal{R}) = 501$ and thus f covers more than 99% of \mathbb{F}_q .
- To cover \mathbb{F}_q (that is $\mathcal{R} = \mathbb{F}_q$), one can choose an element $c \in \mathbb{F}_q$ such that $\chi((1 - c^4)(1 + 3c^4)) = 1$ and -1 is a square (for example when $q \equiv 2 \pmod{3}$), and then $\mathcal{R} = \mathbb{F}_q$.

Remark 1. – The choice of c (and u) does not have any impact on the running-time of the algorithm, since one must choose a suitable c (and u) before starting the algorithm. When $q \equiv 3 \pmod{4}$, one can choose $u = -1$; if $q \equiv 5 \pmod{8}$, one can choose $u = 2$.

- Compared to many existing encodings, we do not put any requirements on q (the authors of [13] proposed for example to choose $q \equiv 2 \pmod{3}$, in order to compute efficiently cubic roots).

3.3 Properties of Our Encoding

Lemma 1 (Polynomial time).

The function f_λ can be implemented in deterministic polynomial time, with approximately $O(\log^3(q))$ operations over \mathbb{F}_q .

Proof. – The function f_λ is deterministic in the sense that, once the parameters q, c and u are fixed, then any input $r \in \mathcal{R}$ will always give the same output $P = (x_\lambda, y_\lambda)$. In fact, the algorithm does not involve any random value.

- To show that the algorithm is also computable in polynomial time, we must evaluate its complexity. Globally, the computation of $P = (x_\lambda, y_\lambda)$ requires some inversions, some multiplications, one computation of the quadratic character χ and one square root computation. The computation of χ can be replaced by an exponentiation (to test if a is square, just compute a to the exponent $(q - 1)/2$), which requires $O(\log^3(q))$ operations. Computing a square root in \mathbb{F}_q requires $O(\log^3(q))$ operations when $q \equiv 3 \pmod{4}$; and more generally, it can be done in probabilistic polynomial time by using the Tonelli-Shanks algorithm. The inversions can be made efficiently by using extended Euclid algorithm or avoided by using projective coordinates (excepted for the last two inversions in x_λ and y_λ). So globally, we can expect our algorithm to run in polynomial time (with approximately $O(\log^3(q))$ operations). \square

From definition (1) and from Algorithm (1), it is easy to see that given any point $P \in \text{Im}(f_\lambda)$ such that $f_\lambda(r) = P$ (for a certain $r \in \mathcal{R}$), we have $f_\lambda^{-1}(P) = \{r, -r\}$. This results from the fact that the definition of the encoding f_λ only involves r^2 . Hence, if $f_\lambda(r) = P$, then $f_\lambda(-r) = P$ also. And one can

show that $r, -r$ are the only points in the set $f_\lambda^{-1}(P)$ (like in [1] or in [8], this property of f_λ is called *almost-injectivity*). Moreover, we can invert f_λ as follows.

Lemma 2 (Inverting f_λ).

Given a point $P = (x_\lambda, y_\lambda) \in \text{Im}(f_\lambda)$, we can compute its preimage $r \in \mathcal{R}$ as follows:

$$\begin{aligned} - \text{ if } \chi(y_\lambda) = -1, \text{ then } r &= \frac{1}{2c^3} \sqrt{\frac{c(x-1)(3+c^4) + (1+x)(1+3c^4)}{u[x(1-c) + 1+c]}}; \\ - \text{ if } \chi(y_\lambda) = 1, \text{ then } r &= \frac{1}{2c^3(1-c^4)} \sqrt{\frac{(1+x)(5c^4-1) + c(1-x)(3+c^4)}{u[(1-c^4)(1+x+c(x-1))]} } . \end{aligned}$$

The proof is similar to those in [1, 8].

Remark 2. Encodings into elliptic curves can be used in several ways. For example, Bernstein *et al.* [1] used an almost-injective encoding, namely Elligator-2, to make uniform strings indifferentiable from random. When the encoding function is well-distributed, it can be used to design an indifferentiable hash function into $E(\mathbb{F}_q)$. We can use these two applications for our encoding, since it is:

- *almost-injective*: injective when restricted to a certain subset S of \mathbb{F}_q . In fact, we can characterize the image set of f_λ and show that given a point P in $\text{Im}(f_\lambda)$, $f_\lambda^{-1}(P) \in \{r, -r\}$ for some $r \in \mathbb{F}_q$. When \mathbb{F}_q is a prime field, one can just set $S = \{0, 1, \dots, \frac{q-1}{2}\}$;
- and *well-distributed*: in [11], Farashahi *et al.* showed that any deterministic encoding into elliptic curves can be transformed into a well-distributed one.

Example 1. We consider an example with the following parameters: $q = 503$, we set $u = -1$ and $c = 3$. A code for the implementation is given in appendix as well as the outputs for this example.

4 Conclusion

In this work, we described the first known encoding for the theta model for elliptic curves $E_\lambda : 1 + x^2 + y^2 + x^2y^2 = \lambda^2xy$. And we showed that this new encoding is efficiently computable (deterministic and polynomial-time). A numerical example is also given to ensure the correctness of our encoding. Like existing encodings for other models of curves, our encoding has some interesting features, like almost-injectivity and inversibility. Such properties can be used to design indifferentiable hash functions in the group of points of the curve, or to design IBE-schemes.

A An Implementation of Theta-Model-Encoding in Sage

```

class ThetaModel():
    def init(self,q,u,c): ##assuming q is prime
        self.F=FiniteField(q,'a')
        self.q=q
        self.u=u
        self.c=c
    def verificationParameters(self):
        F=self.F
        q=self.q
        u=self.u
        c=self.c
        if F.characteristic()==2:
            print 'Error: The characteristic is egal to', F.characteristic()
            return False
        else:
            if (F(c).is_zero() is True) or (F(1-c^4).is_zero() is True) or (F(1+c^4).is_zero() is True):
                print 'Error: bad value for c'
                return False
            else:
                if (F(u).is_zero() is True) or (F(u).is_square() is True):
                    print 'Error: u={} is zero or is a square'.format(u)+' in the finite field of {} elements'.format(F.order())
                    return False
                else:
                    return True
    def setOfDefinition(self,value):
        F=self.F
        c=self.c
        u=self.u
        r=F(value)
        if ((F(4*u*r*r*(c^6)-(1-c^4)*(1+3*c^4)).is_zero() is True) or
            ( F(4*u*r*r*(c^6)+(1-c^4)^4).is_zero() is True ) or ( F(4*u*r*r*(c^6)*(1+3*(c^4)-(1-c^4)^3).is_zero() is True):
            print 'r={} is not in the set of definition'.format(r)
            return False
        else:
            return True
    def quadraticCharacter(self,value):
        F=self.F
        try:
            residu=value.is_square()
        except:
            residu=F(value).is_square()
        if residu is True:
            return 1
        else:
            return -1
    def encodeTheta(self,value):
        if self.verificationParameters() is True:
            if self.setOfDefinition(value) is True:
                F=self.F
                r=F(value)
                c=self.c
                u=self.u
                v=F( c*(4*u*r*r*(c^6)+(1-c^4)*(3+c^4))/(4*u*r*r*(c^6)+(1-c^4)*(-3*(c^4)-1)) );
                e=self.quadraticCharacter(((c*c-v*v)*(1-c*c*v*v)));
                X=F( v*(e+1)/2 + (((-c^4-1)*(v+c)-c*(1-c^4))/(2*(c^3)*(v+c)+1-c^4))*(e-1)/2 );
                try:
                    a=F((c*c-X*X)/(1-c*c*X*X))
                    root=a.square_root()
                    Y=e*root
                    x=(X+1)/(X-1)
                    y=(Y-1)/(Y+1)
                    return (F(x),F(y))
                except:
                    print 'Error when computing the square root of f(x)'
                    return {}

```

B Example with $q = 503, u = -1, c = 3$:

t=ThetaModel()	r = 4 =====>(x,y)= (212, 73)	r = 17 =====>(x,y)= (283, 123)
t.init(501,-1,3)	r = 5 =====>(x,y)= (307, 100)	r = 18 =====>(x,y)= (327, 188)
if t.verificationParameters() is True:	r = 6 =====>(x,y)= (23, 298)	r = 19 =====>(x,y)= (411, 365)
for i in t.F:	r = 7 =====>(x,y)= (42, 171)	r = 20 =====>(x,y)= (480, 265)
if t.setOfDefinition(i) is True:	r = 8 =====>(x,y)= (311, 239)	r = 21 =====>(x,y)= (25, 485)
print 'r=',i,'=====>(x,y)=' ,	r = 9 =====>(x,y)= (491, 332)	r = 22 =====>(x,y)= (386, 153)
t.encodeTheta(i)	r = 10 =====>(x,y)= (415, 385)	r = 23 =====>(x,y)= (176, 404)
	r = 11 =====>(x,y)= (125, 490)	r = 24 =====>(x,y)= (368, 142)
	r = 12 =====>(x,y)= (330, 470)	r = 25 =====>(x,y)= (121, 62)
r = 0 =====>(x,y)= (430, 121)	r = 13 =====>(x,y)= (328, 205)	r = 26 =====>(x,y)= (40, 422)
r = 1 =====>(x,y)= (441, 382)	r = 14 =====>(x,y)= (315, 483)	r = 27 =====>(x,y)= (461, 50)
r = 2 =====>(x,y)= (386, 240)	r = 15 =====>(x,y)= (77, 31)	r = 28 =====>(x,y)= (318, 46)
r = 3 =====>(x,y)= (283, 274)	r = 16 =====>(x,y)= (352, 369)	r = 29 =====>(x,y)= (170, 76)

```

r= 30 =====>(x,y)=(217, 310)
r= 31 =====>(x,y)=(171, 12)
r= 32 =====>(x,y)=(187, 49)
r= 33 =====>(x,y)=(436, 313)
r= 34 =====>(x,y)=(31, 98)
r= 35 =====>(x,y)=(43, 350)
r= 36 =====>(x,y)=(382, 73)
r= 37 =====>(x,y)=(490, 125)
r= 38 =====>(x,y)=(338, 435)
r= 39 =====>(x,y)=(124, 141)
r= 40 =====>(x,y)=(378, 387)
r= 41 =====>(x,y)=(416, 457)
r= 42 =====>(x,y)=(63, 32)
r= 43 =====>(x,y)=(86, 51)
r=44 is not in the set of definition
r= 45 =====>(x,y)=(470, 330)
r= 46 =====>(x,y)=(259, 352)
r= 47 =====>(x,y)=(153, 460)
r= 48 =====>(x,y)=(318, 339)
r= 49 =====>(x,y)=(237, 69)
r= 50 =====>(x,y)=(220, 380)
r= 51 =====>(x,y)=(442, 330)
r= 52 =====>(x,y)=(213, 166)
r= 53 =====>(x,y)=(495, 110)
r= 54 =====>(x,y)=(393, 8)
r= 55 =====>(x,y)=(185, 164)
r= 56 =====>(x,y)=(287, 364)
r= 57 =====>(x,y)=(378, 13)
r= 58 =====>(x,y)=(472, 405)
r= 59 =====>(x,y)=(232, 304)
r= 60 =====>(x,y)=(51, 310)
r= 61 =====>(x,y)=(216, 139)
r= 62 =====>(x,y)=(258, 199)
r= 63 =====>(x,y)=(60, 476)
r= 64 =====>(x,y)=(258, 91)
r= 65 =====>(x,y)=(382, 441)
r= 66 =====>(x,y)=(352, 259)
r= 67 =====>(x,y)=(350, 117)
r= 68 =====>(x,y)=(109, 149)
r= 69 =====>(x,y)=(102, 336)
r= 70 =====>(x,y)=(467, 375)
r= 71 =====>(x,y)=(425, 154)
r= 72 =====>(x,y)=(339, 318)
r= 73 =====>(x,y)=(231, 209)
r= 74 =====>(x,y)=(379, 362)
r= 75 =====>(x,y)=(116, 334)
r= 76 =====>(x,y)=(252, 502)
r= 77 =====>(x,y)=(149, 109)
r= 78 =====>(x,y)=(235, 278)
r= 79 =====>(x,y)=(466, 338)
r= 80 =====>(x,y)=(457, 416)
r= 81 =====>(x,y)=(16, 123)
r= 82 =====>(x,y)=(495, 471)
r= 83 =====>(x,y)=(141, 215)
r= 84 =====>(x,y)=(470, 157)
r= 85 =====>(x,y)=(127, 94)
r= 86 =====>(x,y)=(334, 490)
r= 87 =====>(x,y)=(311, 181)
r= 88 =====>(x,y)=(301, 409)
r= 89 =====>(x,y)=(458, 67)
r= 90 =====>(x,y)=(440, 471)
r= 91 =====>(x,y)=(360, 253)
r= 92 =====>(x,y)=(350, 43)
r= 93 =====>(x,y)=(304, 245)
r= 94 =====>(x,y)=(322, 192)
r= 95 =====>(x,y)=(173, 33)
r= 96 =====>(x,y)=(18, 342)
r= 97 =====>(x,y)=(410, 179)
r= 98 =====>(x,y)=(485, 25)
r= 99 =====>(x,y)=(190, 488)
r= 100 =====>(x,y)=(475, 161)
r= 101 =====>(x,y)=(225, 122)
r= 102 =====>(x,y)=(8, 393)
r= 103 =====>(x,y)=(73, 382)
r= 104 =====>(x,y)=(342, 18)
r= 105 =====>(x,y)=(69, 237)
r= 106 =====>(x,y)=(442, 157)
r= 107 =====>(x,y)=(443, 27)
r= 108 =====>(x,y)=(403, 290)
r= 109 =====>(x,y)=(12, 453)
r= 110 =====>(x,y)=(148, 214)
r= 111 =====>(x,y)=(291, 430)
r= 112 =====>(x,y)=(289, 486)
r= 113 =====>(x,y)=(135, 361)
r= 114 =====>(x,y)=(250, 401)
r= 115 =====>(x,y)=(369, 493)
r= 116 =====>(x,y)=(305, 376)
r= 117 =====>(x,y)=(167, 143)
r= 118 =====>(x,y)=(493, 369)
r= 119 =====>(x,y)=(212, 441)

r= 120 =====>(x,y)=(125, 116)
r= 121 =====>(x,y)=(38, 235)
r= 122 =====>(x,y)=(490, 334)
r= 123 =====>(x,y)=(196, 337)
r= 124 =====>(x,y)=(263, 43)
r= 125 =====>(x,y)=(324, 93)
r= 126 =====>(x,y)=(357, 77)
r= 127 =====>(x,y)=(337, 290)
r= 128 =====>(x,y)=(337, 196)
r= 129 =====>(x,y)=(164, 185)
r= 130 =====>(x,y)=(460, 153)
r= 131 =====>(x,y)=(14, 224)
r= 132 =====>(x,y)=(245, 412)
r= 133 =====>(x,y)=(110, 495)
r= 134 =====>(x,y)=(271, 91)
r= 135 =====>(x,y)=(478, 18)
r= 136 =====>(x,y)=(151, 134)
r= 137 =====>(x,y)=(435, 189)
r= 138 =====>(x,y)=(467, 279)
r= 139 =====>(x,y)=(404, 483)
r= 140 =====>(x,y)=(489, 279)
r= 141 =====>(x,y)=(409, 301)
r= 142 =====>(x,y)=(493, 259)
r= 143 =====>(x,y)=(315, 176)
r= 144 =====>(x,y)=(28, 478)
r= 145 =====>(x,y)=(237, 226)
r= 146 =====>(x,y)=(157, 442)
r= 147 =====>(x,y)=(465, 268)
r= 148 =====>(x,y)=(394, 354)
r= 149 =====>(x,y)=(460, 240)
r= 150 =====>(x,y)=(198, 202)
r= 151 =====>(x,y)=(202, 198)
r= 152 =====>(x,y)=(189, 466)
r= 153 =====>(x,y)=(87, 339)
r= 154 =====>(x,y)=(78, 454)
r= 155 =====>(x,y)=(240, 386)
r= 156 =====>(x,y)=(213, 100)
r= 157 =====>(x,y)=(143, 250)
r= 158 =====>(x,y)=(224, 36)
r= 159 =====>(x,y)=(51, 86)
r= 160 =====>(x,y)=(435, 338)
r= 161 =====>(x,y)=(215, 132)
r= 162 =====>(x,y)=(411, 390)
r= 163 =====>(x,y)=(376, 305)
r= 164 =====>(x,y)=(275, 355)
r= 165 =====>(x,y)=(316, 349)
r= 166 =====>(x,y)=(307, 166)
r= 167 =====>(x,y)=(272, 294)
r= 168 =====>(x,y)=(502, 252)
r= 169 =====>(x,y)=(485, 161)
r= 170 =====>(x,y)=(463, 118)
r= 171 =====>(x,y)=(381, 38)
r= 172 =====>(x,y)=(217, 86)
r= 173 =====>(x,y)=(476, 60)
r= 174 =====>(x,y)=(264, 317)
r= 175 =====>(x,y)=(333, 427)
r= 176 =====>(x,y)=(88, 81)
r= 177 =====>(x,y)=(298, 175)
r= 178 =====>(x,y)=(67, 458)
r= 179 =====>(x,y)=(281, 119)
r= 180 =====>(x,y)=(338, 466)
r= 181 =====>(x,y)=(199, 258)
r= 182 =====>(x,y)=(362, 288)
r= 183 =====>(x,y)=(128, 14)
r= 184 =====>(x,y)=(384, 222)
r= 185 =====>(x,y)=(98, 357)
r= 186 =====>(x,y)=(290, 403)
r= 187 =====>(x,y)=(466, 189)
r= 188 =====>(x,y)=(186, 239)
r= 189 =====>(x,y)=(104, 69)
r= 190 =====>(x,y)=(36, 128)
r= 191 =====>(x,y)=(132, 124)
r= 192 =====>(x,y)=(46, 87)
r= 193 =====>(x,y)=(225, 268)
r= 194 =====>(x,y)=(263, 117)
r= 195 =====>(x,y)=(475, 25)
r= 196 =====>(x,y)=(487, 229)
r= 197 =====>(x,y)=(404, 176)
r= 198 =====>(x,y)=(328, 265)
r= 199 =====>(x,y)=(82, 390)
r= 200 =====>(x,y)=(104, 226)
r= 201 =====>(x,y)=(229, 220)
r= 202 =====>(x,y)=(488, 190)
r= 203 =====>(x,y)=(342, 28)
r= 204 =====>(x,y)=(161, 475)
r= 205 =====>(x,y)=(334, 116)
r= 206 =====>(x,y)=(489, 375)
r= 207 =====>(x,y)=(480, 205)
r= 208 =====>(x,y)=(403, 196)
r= 209 =====>(x,y)=(380, 487)

r= 210 =====>(x,y)=(278, 381)
r= 211 =====>(x,y)=(175, 238)
r= 212 =====>(x,y)=(375, 467)
r= 213 =====>(x,y)=(401, 167)
r= 214 =====>(x,y)=(288, 371)
r= 215 =====>(x,y)=(87, 46)
r= 216 =====>(x,y)=(271, 199)
r= 217 =====>(x,y)=(371, 379)
r= 218 =====>(x,y)=(50, 491)
r= 219 =====>(x,y)=(186, 181)
r= 220 =====>(x,y)=(10, 244)
r= 221 =====>(x,y)=(146, 426)
r= 222 =====>(x,y)=(31, 77)
r= 223 =====>(x,y)=(117, 263)
r= 224 =====>(x,y)=(440, 110)
r= 225 =====>(x,y)=(189, 435)
r= 226 =====>(x,y)=(502, 2)
r= 227 =====>(x,y)=(441, 212)
r= 228 =====>(x,y)=(465, 122)
r= 229 =====>(x,y)=(279, 469)
r= 230 =====>(x,y)=(483, 315)
r= 231 =====>(x,y)=(17, 228)
r= 232 =====>(x,y)=(91, 271)
r= 233 =====>(x,y)=(94, 127)
r= 234 =====>(x,y)=(169, 387)
r= 235 =====>(x,y)=(453, 42)
r= 236 =====>(x,y)=(15, 45)
r= 237 =====>(x,y)=(116, 125)
r= 238 =====>(x,y)=(20, 99)
r= 239 =====>(x,y)=(357, 98)
r= 240 =====>(x,y)=(310, 217)
r= 241 =====>(x,y)=(346, 61)
r= 242 =====>(x,y)=(226, 104)
r= 243 =====>(x,y)=(73, 212)
r= 244 =====>(x,y)=(471, 440)
r= 245 =====>(x,y)=(169, 13)
r= 246 =====>(x,y)=(412, 232)
r= 247 =====>(x,y)=(32, 63)
r= 248 =====>(x,y)=(332, 461)
r= 249 =====>(x,y)=(478, 28)
r= 250 =====>(x,y)=(82, 365)
r= 251 =====>(x,y)=(16, 274)
r= 252 =====>(x,y)=(16, 274)
r= 253 =====>(x,y)=(82, 365)
r= 254 =====>(x,y)=(478, 28)
r= 255 =====>(x,y)=(332, 461)
r= 256 =====>(x,y)=(32, 63)
r= 257 =====>(x,y)=(412, 232)
r= 258 =====>(x,y)=(169, 13)
r= 259 =====>(x,y)=(471, 440)
r= 260 =====>(x,y)=(73, 212)
r= 261 =====>(x,y)=(226, 104)
r= 262 =====>(x,y)=(346, 61)
r= 263 =====>(x,y)=(310, 217)
r= 264 =====>(x,y)=(357, 98)
r= 265 =====>(x,y)=(20, 99)
r= 266 =====>(x,y)=(116, 125)
r= 267 =====>(x,y)=(15, 45)
r= 268 =====>(x,y)=(453, 42)
r= 269 =====>(x,y)=(169, 387)
r= 270 =====>(x,y)=(94, 127)
r= 271 =====>(x,y)=(91, 271)
r= 272 =====>(x,y)=(17, 228)
r= 273 =====>(x,y)=(483, 315)
r= 274 =====>(x,y)=(279, 469)
r= 275 =====>(x,y)=(465, 122)
r= 276 =====>(x,y)=(441, 212)
r= 277 =====>(x,y)=(502, 2)
r= 278 =====>(x,y)=(189, 435)
r= 279 =====>(x,y)=(440, 110)
r= 280 =====>(x,y)=(117, 263)
r= 281 =====>(x,y)=(31, 77)
r= 282 =====>(x,y)=(146, 426)
r= 283 =====>(x,y)=(10, 244)
r= 284 =====>(x,y)=(186, 181)
r= 285 =====>(x,y)=(50, 491)
r= 286 =====>(x,y)=(371, 379)
r= 287 =====>(x,y)=(271, 199)
r= 288 =====>(x,y)=(87, 46)
r= 289 =====>(x,y)=(288, 371)
r= 290 =====>(x,y)=(401, 167)
r= 291 =====>(x,y)=(375, 467)
r= 292 =====>(x,y)=(175, 238)
r= 293 =====>(x,y)=(278, 381)
r= 294 =====>(x,y)=(380, 487)
r= 295 =====>(x,y)=(403, 196)
r= 296 =====>(x,y)=(480, 205)
r= 297 =====>(x,y)=(489, 375)
r= 298 =====>(x,y)=(334, 116)
r= 299 =====>(x,y)=(161, 475)

```

```

r= 300 =====>(x,y)= (342, 28)
r= 301 =====>(x,y)= (488, 190)
r= 302 =====>(x,y)= (229, 220)
r= 303 =====>(x,y)= (104, 226)
r= 304 =====>(x,y)= (82, 390)
r= 305 =====>(x,y)= (328, 265)
r= 306 =====>(x,y)= (404, 176)
r= 307 =====>(x,y)= (487, 229)
r= 308 =====>(x,y)= (475, 25)
r= 309 =====>(x,y)= (263, 117)
r= 310 =====>(x,y)= (225, 268)
r= 311 =====>(x,y)= (46, 87)
r= 312 =====>(x,y)= (132, 124)
r= 313 =====>(x,y)= (36, 128)
r= 314 =====>(x,y)= (104, 69)
r= 315 =====>(x,y)= (186, 239)
r= 316 =====>(x,y)= (466, 189)
r= 317 =====>(x,y)= (290, 403)
r= 318 =====>(x,y)= (98, 357)
r= 319 =====>(x,y)= (384, 222)
r= 320 =====>(x,y)= (128, 14)
r= 321 =====>(x,y)= (362, 288)
r= 322 =====>(x,y)= (199, 258)
r= 323 =====>(x,y)= (338, 466)
r= 324 =====>(x,y)= (281, 119)
r= 325 =====>(x,y)= (67, 458)
r= 326 =====>(x,y)= (298, 175)
r= 327 =====>(x,y)= (88, 81)
r= 328 =====>(x,y)= (333, 427)
r= 329 =====>(x,y)= (264, 317)
r= 330 =====>(x,y)= (476, 60)
r= 331 =====>(x,y)= (217, 86)
r= 332 =====>(x,y)= (381, 38)
r= 333 =====>(x,y)= (463, 118)
r= 334 =====>(x,y)= (485, 161)
r= 335 =====>(x,y)= (502, 252)
r= 336 =====>(x,y)= (272, 294)
r= 337 =====>(x,y)= (307, 166)
r= 338 =====>(x,y)= (316, 349)
r= 339 =====>(x,y)= (275, 355)
r= 340 =====>(x,y)= (376, 305)
r= 341 =====>(x,y)= (411, 390)
r= 342 =====>(x,y)= (215, 132)
r= 343 =====>(x,y)= (435, 338)
r= 344 =====>(x,y)= (51, 86)
r= 345 =====>(x,y)= (224, 36)
r= 346 =====>(x,y)= (143, 250)
r= 347 =====>(x,y)= (213, 100)
r= 348 =====>(x,y)= (240, 386)
r= 349 =====>(x,y)= (78, 454)
r= 350 =====>(x,y)= (87, 339)
r= 351 =====>(x,y)= (189, 466)
r= 352 =====>(x,y)= (202, 198)
r= 353 =====>(x,y)= (198, 202)
r= 354 =====>(x,y)= (460, 240)
r= 355 =====>(x,y)= (394, 354)
r= 356 =====>(x,y)= (465, 268)
r= 357 =====>(x,y)= (157, 442)
r= 358 =====>(x,y)= (237, 226)
r= 359 =====>(x,y)= (28, 478)
r= 360 =====>(x,y)= (315, 176)
r= 361 =====>(x,y)= (493, 259)
r= 362 =====>(x,y)= (409, 301)
r= 363 =====>(x,y)= (489, 279)
r= 364 =====>(x,y)= (404, 483)
r= 365 =====>(x,y)= (467, 279)
r= 366 =====>(x,y)= (435, 189)
r= 367 =====>(x,y)= (151, 134)

r= 368 =====>(x,y)= (478, 18)
r= 369 =====>(x,y)= (271, 91)
r= 370 =====>(x,y)= (110, 495)
r= 371 =====>(x,y)= (245, 412)
r= 372 =====>(x,y)= (14, 224)
r= 373 =====>(x,y)= (460, 153)
r= 374 =====>(x,y)= (164, 185)
r= 375 =====>(x,y)= (337, 196)
r= 376 =====>(x,y)= (337, 290)
r= 377 =====>(x,y)= (357, 77)
r= 378 =====>(x,y)= (324, 93)
r= 379 =====>(x,y)= (263, 43)
r= 380 =====>(x,y)= (196, 337)
r= 381 =====>(x,y)= (490, 334)
r= 382 =====>(x,y)= (38, 235)
r= 383 =====>(x,y)= (125, 116)
r= 384 =====>(x,y)= (212, 441)
r= 385 =====>(x,y)= (493, 369)
r= 386 =====>(x,y)= (167, 143)
r= 387 =====>(x,y)= (305, 376)
r= 388 =====>(x,y)= (369, 493)
r= 389 =====>(x,y)= (250, 401)
r= 390 =====>(x,y)= (135, 361)
r= 391 =====>(x,y)= (289, 486)
r= 392 =====>(x,y)= (291, 430)
r= 393 =====>(x,y)= (148, 214)
r= 394 =====>(x,y)= (12, 453)
r= 395 =====>(x,y)= (403, 290)
r= 396 =====>(x,y)= (443, 27)
r= 397 =====>(x,y)= (442, 157)
r= 398 =====>(x,y)= (69, 237)
r= 399 =====>(x,y)= (342, 18)
r= 400 =====>(x,y)= (73, 382)
r= 401 =====>(x,y)= (8, 393)
r= 402 =====>(x,y)= (225, 122)
r= 403 =====>(x,y)= (475, 161)
r= 404 =====>(x,y)= (190, 488)
r= 405 =====>(x,y)= (485, 25)
r= 406 =====>(x,y)= (410, 179)
r= 407 =====>(x,y)= (18, 342)
r= 408 =====>(x,y)= (173, 303)
r= 409 =====>(x,y)= (322, 192)
r= 410 =====>(x,y)= (304, 245)
r= 411 =====>(x,y)= (350, 43)
r= 412 =====>(x,y)= (360, 253)
r= 413 =====>(x,y)= (440, 471)
r= 414 =====>(x,y)= (458, 67)
r= 415 =====>(x,y)= (301, 409)
r= 416 =====>(x,y)= (311, 181)
r= 417 =====>(x,y)= (334, 490)
r= 418 =====>(x,y)= (127, 94)
r= 419 =====>(x,y)= (470, 157)
r= 420 =====>(x,y)= (141, 215)
r= 421 =====>(x,y)= (495, 471)
r= 422 =====>(x,y)= (16, 123)
r= 423 =====>(x,y)= (457, 416)
r= 424 =====>(x,y)= (466, 338)
r= 425 =====>(x,y)= (235, 278)
r= 426 =====>(x,y)= (149, 109)
r= 427 =====>(x,y)= (252, 502)
r= 428 =====>(x,y)= (116, 334)
r= 429 =====>(x,y)= (379, 362)
r= 430 =====>(x,y)= (231, 209)
r= 431 =====>(x,y)= (339, 318)
r= 432 =====>(x,y)= (425, 154)
r= 433 =====>(x,y)= (467, 375)
r= 434 =====>(x,y)= (102, 336)
r= 435 =====>(x,y)= (109, 149)

r= 436 =====>(x,y)= (350, 117)
r= 437 =====>(x,y)= (352, 259)
r= 438 =====>(x,y)= (382, 441)
r= 439 =====>(x,y)= (258, 91)
r= 440 =====>(x,y)= (60, 476)
r= 441 =====>(x,y)= (258, 199)
r= 442 =====>(x,y)= (216, 139)
r= 443 =====>(x,y)= (51, 310)
r= 444 =====>(x,y)= (232, 304)
r= 445 =====>(x,y)= (472, 405)
r= 446 =====>(x,y)= (378, 13)
r= 447 =====>(x,y)= (287, 364)
r= 448 =====>(x,y)= (185, 164)
r= 449 =====>(x,y)= (393, 8)
r= 450 =====>(x,y)= (495, 110)
r= 451 =====>(x,y)= (213, 166)
r= 452 =====>(x,y)= (442, 330)
r= 453 =====>(x,y)= (220, 380)
r= 454 =====>(x,y)= (237, 69)
r= 455 =====>(x,y)= (318, 339)
r= 456 =====>(x,y)= (153, 460)
r= 457 =====>(x,y)= (259, 352)
r= 458 =====>(x,y)= (470, 330)
r=459 is not in the set of definition
r= 460 =====>(x,y)= (86, 51)
r= 461 =====>(x,y)= (63, 32)
r= 462 =====>(x,y)= (416, 457)
r= 463 =====>(x,y)= (378, 387)
r= 464 =====>(x,y)= (124, 141)
r= 465 =====>(x,y)= (338, 435)
r= 466 =====>(x,y)= (490, 125)
r= 467 =====>(x,y)= (382, 73)
r= 468 =====>(x,y)= (43, 350)
r= 469 =====>(x,y)= (31, 98)
r= 470 =====>(x,y)= (436, 313)
r= 471 =====>(x,y)= (187, 49)
r= 472 =====>(x,y)= (171, 12)
r= 473 =====>(x,y)= (217, 310)
r= 474 =====>(x,y)= (170, 76)
r= 475 =====>(x,y)= (318, 46)
r= 476 =====>(x,y)= (461, 50)
r= 477 =====>(x,y)= (40, 422)
r= 478 =====>(x,y)= (121, 62)
r= 479 =====>(x,y)= (368, 142)
r= 480 =====>(x,y)= (176, 404)
r= 481 =====>(x,y)= (386, 153)
r= 482 =====>(x,y)= (25, 485)
r= 483 =====>(x,y)= (480, 265)
r= 484 =====>(x,y)= (411, 365)
r= 485 =====>(x,y)= (327, 188)
r= 486 =====>(x,y)= (283, 123)
r= 487 =====>(x,y)= (352, 369)
r= 488 =====>(x,y)= (77, 31)
r= 489 =====>(x,y)= (315, 483)
r= 490 =====>(x,y)= (328, 205)
r= 491 =====>(x,y)= (330, 470)
r= 492 =====>(x,y)= (125, 490)
r= 493 =====>(x,y)= (415, 385)
r= 494 =====>(x,y)= (491, 332)
r= 495 =====>(x,y)= (311, 239)
r= 496 =====>(x,y)= (42, 171)
r= 497 =====>(x,y)= (23, 298)
r= 498 =====>(x,y)= (307, 100)
r= 499 =====>(x,y)= (212, 73)
r= 500 =====>(x,y)= (283, 274)
r= 501 =====>(x,y)= (386, 240)
r= 502 =====>(x,y)= (441, 382)

```

References

1. Bernstein, D.J., Hamburg, M., Krasnova, A., Lange, T.: Elligator: elliptic-curve points indistinguishable from uniform random strings. In: Gligor, V., Yung, M. (eds.) ACM CCS (2013)
2. Billet, O., Joye, M.: The Jacobi model of an elliptic curve and side-channel analysis. In: Fossorier, M., Høholdt, T., Poli, A. (eds.) AAEC 2003. LNCS, vol. 2643, pp. 34–42. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-44828-4_5

3. Boneh, D., Franklin, M.: Identity-based encryption from the weil pairing. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 213–229. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44647-8_13
4. Chudnovsky, D.V., Chudnovky, G.V.: Sequences of numbers generated by addition in formal groups and new primality and factorization tests. *Adv. Appl. Math.* **7**(4), 385–434 (1986)
5. Devigne, J., Joye, M.: Binary huff curves. In: Kiayias, A. (ed.) CT-RSA 2011. LNCS, vol. 6558, pp. 340–355. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19074-2_22
6. Diao, O., Fouotsa, E.: Arithmetic of the level four theta model of elliptic curves. *Afrika Mathematica* **26**(3), 283–301 (2015)
7. Edwards, H.M.: A normal form for elliptic curves. *Bull. Am. Math. Soc.* **44**, 393–422 (2007). <http://www.ams.org/bull/2007-44-03/S0273-0979-07-01153-6/home.html>
8. Diarra, N., Sow, D., Ould Cheikh Khlil, A.Y.: On indifferentiable deterministic hashing into elliptic curves. *Eur. J. Pure Appl. Math.* **10**(2), 363–391 (2017). (MathScinet: MR3607082) (ZentralBlattMath: Zbl 06701281)
9. Fouotsa, E., Diao, O.: A theta model for elliptic curves. *Mediterr. J. Math.* **14**, 65 (2017). <https://doi.org/10.1007/s00009-017-0840-y>
10. Farashahi, R.R.: Hashing into Hessian curves. In: Nitaj, A., Pointcheval, D. (eds.) AFRICACRYPT 2011. LNCS, vol. 6737, pp. 278–289. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21969-6_17
11. Farashahi, R.R., Fouque, P.-A., Shparlinski, I.E., Tibouchi, M., Voloch, J.F.: Indifferentiable deterministic hashing to elliptic and hyperelliptic curves. *Math. Comput.* **82**(281), 491–512 (2013)
12. Fouque, P.-A., Tibouchi, M.: Estimating the size of the image of deterministic hash functions to elliptic curves. In: Abdalla, M., Barreto, P.S.L.M. (eds.) LATIN-CRYPT 2010. LNCS, vol. 6212, pp. 81–91. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14712-8_5
13. Icart, T.: How to hash into elliptic curves. In: Halevi, S. (ed.) CRYPTO 2009. LNCS, vol. 5677, pp. 303–316. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03356-8_18
14. Smart, N.P.: The Hessian form of an elliptic curve. In: Koç, Ç.K., Naccache, D., Paar, C. (eds.) CHES 2001. LNCS, vol. 2162, pp. 118–125. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44709-1_11
15. Shallue, A., van de Woestijne, C.E.: Construction of rational points on elliptic curves over finite fields. In: Hess, F., Pauli, S., Pohst, M. (eds.) ANTS 2006. LNCS, vol. 4076, pp. 510–524. Springer, Heidelberg (2006). https://doi.org/10.1007/11792086_36
16. Washington, L.C.: Elliptic Curves. Number Theory and Cryptography. Discrete Mathematics and Applications. Chapman and Hall, Boca Raton (2008)



BEDWE: A Decentralized Workflow Engine for Best-Effort Infrastructures

Palakiyem Wallah¹, Cédric Tedeschi²(✉), and Jean-Louis Pazat²

¹ Université de Kara, Kara, Togo

² Univ Rennes, Inria, CNRS, IRISA, Rennes, France
`cedric.tedeschi@inria.fr`

Abstract. We consider the problem of executing composite computing applications called *workflows* on top of unreliable computing infrastructures. Having in mind the situation of the electric delivery in the sub-saharan area, we propose BEDWE, a decentralized workflow engine able to dynamically assign portions of the workflow to currently live compute nodes. More precisely, in a point-to-point manner, each node can receive a part of the workflow and delegate a subpart of it to another node. This mechanism can be repeated recursively until the whole workflow is executed. BEDWE includes a mechanism to support nodes leaving the network due to power outage. We present a software prototype of BEDWE and its experimentation over the French nation-wide Grid'5000 platform.

Keywords: Workflows · Decentralized orchestration
Fault-tolerance · Best-effort infrastructures

1 Introduction

Computing reliably is a major challenge in countries struggling to deliver a constant electric power delivery. For instance, African countries from the sub-Saharan region are used to face power outages, due to an insufficient level of electric injection to satisfy the needs of people, administration and companies. This problem can be solved by cutting electricity in some area/while another one is supplied. These cuts are planned according to a predefined schedule that are publicly announced so people can organise themselves. Being able to ensure the completion of computations running on computers located in such an environment calls for fault-tolerance mechanisms to be injected in the system supporting these applications. This paper explores a specific problem in this area, where we consider a fully-decentralized computing platform composed of compute nodes running in an electric environment subject to power cuts. More specifically, we consider applications which are compositions of building block services, as followed by *service-oriented computing*.

Service-oriented computing has become one of the dominant paradigms to develop applications in both scientific and industrial contexts. This model advocates the composition of services as a programming model to build complex

applications out of existing building blocks, or simpler services [1]. A sibling concept is the *workflow*. A workflow is a *temporal* composition of services completing a specific task. More precisely, each service in a workflow corresponds to a particular step in the workflow and may have precedence constraint dependencies with other services. These dependencies between services can be represented as a directed acyclic graph where nodes are the services and links are the dependencies/precedence constraints between services.

A workflow needs to be enacted. For this, it traditionally relies on a *workflow engine* *i.e.*, a program that takes a workflow's specification and deploys it on compute nodes, starting each service once its dependencies have been satisfied *i.e.*, when the services to get run before completed. This engine is traditionally a centralized, reliable component. If it fails, the whole workflow coordination is undermined and the workflow may not complete. Having this centralized engine is no longer possible when the workflow is supposed to run over unreliable platforms, where having one reliable node on which to run the engine is no longer possible. The problem becomes even more complicated when considering these emerging computing platforms described above. Recently, decentralizing workflow execution has been the focus of some researches aiming at providing solutions where centralized enactment systems cannot be used any more.

The objectives of our work is to propose: (1) a truly decentralized workflow execution whose management is shared by each participant node and (2) a mechanism to support crashes and delays for common workflow patterns. Our solution, called BEDWE (*Best Effort Decentralized Workflow Execution*) runs on every node taking part in the workflow engine. It uses a specific protocol allowing to send, receive and process portions of the workflow.

The remainder of this paper is organized as follows. In Sect. 2, we define our model and describes, by syntactic means, the type of workflows supported. Section 3 describes BEDWE execution model and illustrates its execution and deployment over the nodes for different workflow patterns. In Sect. 4, we discuss the problems brought about by potential crashes (for instance due to power cuts) and devise a simple solution, based on a heartbeat protocol to deal with failures in BEDWE. Experimental results using the Montage workflow over the Grid'5000 platform are presented in Sect. 5. Finally, Sect. 7 presents some applications perspective and concludes the paper.

2 Workflow Patterns and Grammar

Workflow computing is a major paradigm in both business process management and many scientific fields. We distinguish between (1) a workflow specification module that allows to describe services and the ordering amongst them, and (2) a workflow enactment system that coordinates the execution of the services in the correct order. Our contribution is focused on the workflow enactment system level which should handle most workflow models through an appropriate language.

2.1 Workflow Patterns

We here describe the type of workflows that we want to support in BEDWE. A workflow is a finite set of services that are composed in some specific logical order to accomplish a specific process/application. Most commonly, patterns found in workflows are the three following models, as illustrated in Fig. 1: (a) the sequential pattern, (b) the parallel pattern, and (c) conditional pattern.

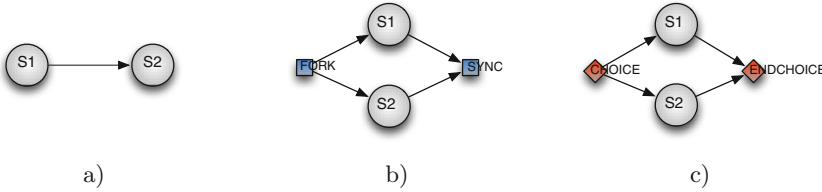


Fig. 1. The three basic workflow patterns.

In order to express more complex patterns, these basic patterns can be combined. Therefore, the loop structure will not be considered in this dissertation. As an example, Fig. 2 shows a concrete workflow which integrates 10 services. The first two, S1 and S2 services are sequentially processed. After service S2, the execution diverges into two parallel branches. The execution of branches will be merged before the execution of service S9. Also, after the execution of service S3, the two outgoing branches are associated with the choice condition and either S3 S4 S6 or S3 S5 S6 path will be selected. Consequently, after S3, a sequential order process is obtained with S6.

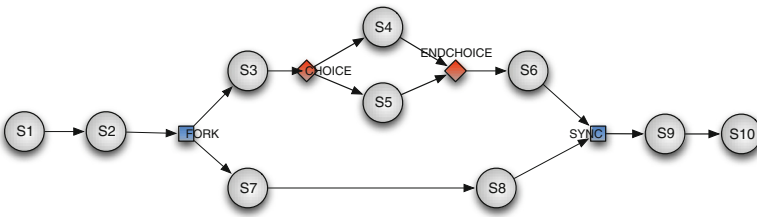


Fig. 2. Example of workflow.

2.2 Workflow Description Language

Defining a workflow consists in describing the patterns that compose it and the relationships among them so, any programmable analyzer can interpret it. As

mentioned in Sect. 2.1, a workflow pattern is defined by control-flow dependencies between services. In order to build our workflow grammar, we summarize the possible workflow definition as follows: (1) a workflow is composed of patterns; (2) a pattern is either a parallel, a conditional or a sequential model, or a composition of two or three different models; (3) a parallel workflow pattern starts with the FORK keyword which splits the workflow, followed by a list of branches and ends with the SYNC keyword which merges incoming result of branches; (4) a conditional workflow pattern starts with the CHOICE keyword which splits the workflow, followed by a list of switch branches and ends with the ENDCHOICE keyword indicating a simple merge step. (5) a sequential workflow pattern is one or a set of services with a basic dependency (a simple transition). Using some compiling languages principles and rules it is easy to convert above summary in a grammar annotation similar to the BNF notation that is depicted in Table 1.

Table 1. Workflow grammar.

$WF \rightarrow t \{WF_s \mid WF_p \mid WF_c\}^*$	with:
$WF_s \rightarrow t^+$	t: service identifier, BRC: branch
$WF_p \rightarrow \text{FORK}(\text{BRC } \{, \text{BRC}\}^+) \text{SYNC}$	WF_s : sequential pattern
$WF_c \rightarrow \text{CHOICE}(\text{BRC } \{, \text{BRC}\}^+) \text{ENDCHOICE}$	WF_p : parallel pattern
$\text{BRC} \rightarrow WF_s \mid WF_p \mid WF_c$	WF_c : conditional pattern

According to this grammar, the workflow in Fig. 2 is easily described as follow, where Si is a service identifier:

S1 S2 FORK(S3 CHOICE(S4, S5)ENDCHOICE S6, S7 S8)SYNC S9 S10

We developed a parser program based on Flex and Bison to validate this grammar [2].

3 BEDWE Workflow Processing

3.1 Pattern Extraction Model

BEDWE decentralized procedure relies on workflow extraction: any workflow w in our grammar can be written as $w = y.z$, y being the first atomic service to be executed on the local node, and z the rest of the workflow definition to be sent in a *message pack* to another node able to execute it. More precisely:

- if y is a sequential pattern, then the first service of the sequence is extracted;
- if y is a parallel pattern, then a list of its branches is extracted;
- if y is a conditional pattern, then from the extracted list of branches, one is chosen according to the choice condition.

3.2 Workflow Execution in BEDWE

In the following, we describe how BEDWE operates, by illustrating its behaviour on a particular workflow composed of the three basic patterns. We assume a set of agents (processes) running on a set of possibly distributed compute nodes, and that are able to communicate through message passing. All these agents run the same software stack; they can invoke services, check the well-formedness of the workflow (or part of a workflow) received and communicate via message-passing. These agents are thus interchangeable and any one of them can be selected to manage the workflow or part of it, as we will now describe. Let us use the example of Fig. 2 that contains the three basic patterns to illustrate different main phases of the process illustrated in Figs. 3, 4, 5, 6 and 7.

Initialization. A first contact node (N1 in Fig. 3), gets a complete description of a workflow from the client, it extracts the first pattern, here S1, it assembles with the rest of definition into a message pack which will be sent to a node N2 that can invoke S1.

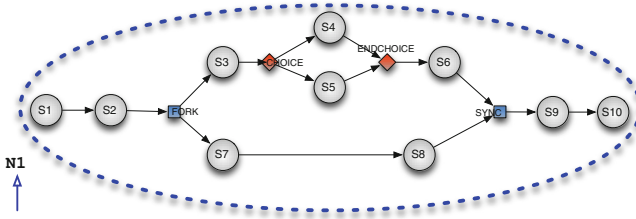


Fig. 3. BEDWE workflow execution (initialisation).

Processing of a Sequential Pattern. As illustrated in Fig. 4, some node N2 gets a message pack from N1, it invokes service S1 and, in the same way extracts from the content of definition the first pattern (here it is S2), it assembles with the rest of definition into a message pack which will be sent to a node N3, that can invoke S2.

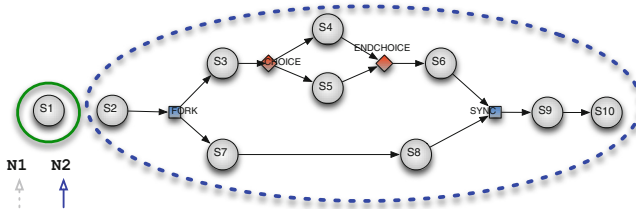


Fig. 4. BEDWE workflow execution (sequential pattern).

Processing of a Parallel Pattern. As illustrated in Fig. 5, some node N3, receives a message pack from N2, it invokes service S2 and try to extract from the content of definition the first pattern, here it is a parallel model. It extracts all branches belonging to that pattern, here there are “S3 CHOICE(S4, S5)ENDCHOICE S6” and “S7 S8”. It then assembles message pack from each branch and sends them to different nodes (here are N4 and N5) which are respectively able to invoke service S3 and service S7. It keeps the following patterns of the parallel model (here: S9 S10) that will be processed after synchronization. Nodes N4 and N5 receive their message packs from the node N3 and they concurrently process to it. Node N5 will process as sequential pattern while node N4 will operate on a conditional pattern after service S3.

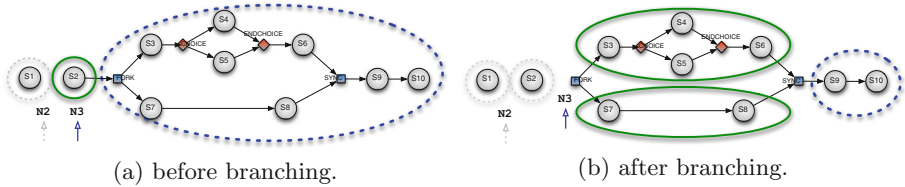


Fig. 5. BEDWE workflow execution (parallel pattern).

Processing of a Conditional Pattern. A conditional pattern execution is illustrated in Fig. 6. Node N4, after invocation of service S3 and tries to extract from the content of definition the first pattern, here it is a conditional model. It extracts the two branches belonging to that pattern, here it is S4 and S5. Then, it evaluates the condition of selection, we suppose S4 is selected. The selected conditional branch forms with rest of the branch a sequential model (S4 S6) which will be processed as sequential phase pattern.

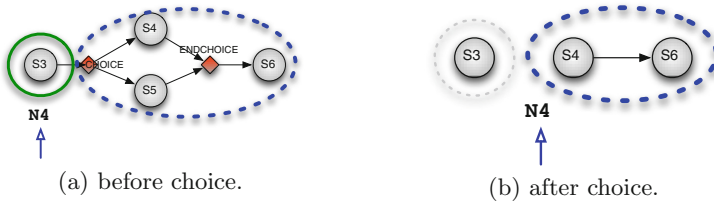


Fig. 6. BEDWE workflow execution (conditional pattern).

Processing of a Synchronization. As illustrated in Fig. 7, the nodes which has invoked services S6 and S8 send their results to the node N3 who had split branches for parallel processing. At the end of synchronization, node N3 reads the following of achieved parallel pattern and processes it as any pattern. For our example case, we have a pattern S9 S10 which is a sequential model.

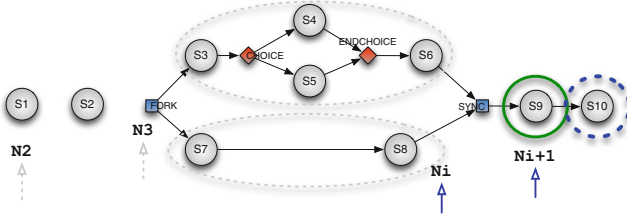


Fig. 7. BEDWE workflow execution (synchronisation pattern).

Ending Output. Still as shown in Fig. 7, after the invocation of service S10, the workflow has completed its execution. The output result of S10 is sent to the client as such a good process of the workflow.

4 Fault Tolerance in BEDWE

A service's execution time can significantly vary between invocations, even for different invocations within the same workflow. This can be due to network's and CPU's load fluctuation. Incidentally, processes may crash. This raises the common question of detecting crashes of computing nodes hosting services: how much time should we wait for an answer before considering a service invocation as *failed*. In BEDWE, we rely on the classical *heartbeat* mechanism and fix a particular duration above which no heartbeat received from a process makes it considered as *failed*.

4.1 BEDWE's Resilience Principle

The implementation of the heartbeat protocol in BEDWE is illustrated in Fig. 8. Assume Node 1 sends a message to Node 2 to request it to process the next workflow pattern. We can consider that Node 1 becomes the client for Node 2 which becomes the server for this particular part of the workflow. During the service's execution on Node 2, Node 2 sends a particular heartbeat message to Node 1 periodically, informing it that it is still running the task. Upon receipt, Node 1 updates its the liveness status of Node 2 as *alive*. Also, periodically, Node 1 checks this liveness status and resets it to *crash*. If, at the next checking time, the status is not back to *alive*, it means Node 2 did not send the heartbeat and thus will be considered as *crashed* by Node 1. When, the task completion's notification is received, Node 1 stops its periodic checking of Node 2's liveness. Note that a node can be both a server and a client for different tasks.

4.2 Fault Tolerance for Each Pattern

Let us now review more precisely how the resilience is done for each possible pattern. Note that the resilience mechanism is recursively done in recursive patterns

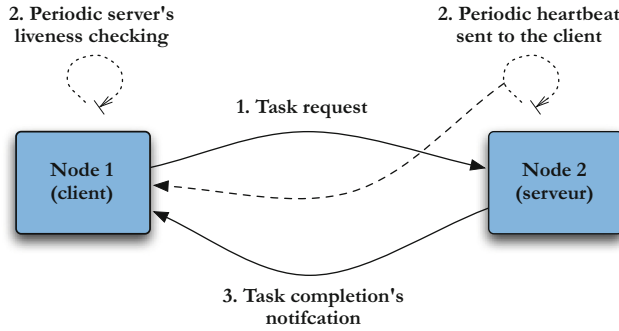


Fig. 8. BEDWE's resilience principle.

such as nested parallel patterns. The key assumptions in the following are that: (1) A node cannot be responsible for the execution of two consecutive services in a sequence. (2) The two nodes responsible for two consecutive services cannot fail at the same time.

Sequential Pattern. In a sequence of services, the heartbeat protocol presented previously is repeated between each pair of nodes managing consecutive services. Given a sequence of three services whose execution is hosted by N1, N2 and N3 respectively, the process is first started when N2 starts executing, N2 being the server and N1 the client. Once N2 completes, it sends the subsequent part of the workflow to N3. At this point, the heartbeat mechanism is also triggered between N2 which is now the client and N3 which is the server. Once the mechanism is started between N2 and N3, N2 sends the notification of completion and the heartbeat protocol between N1 and N2 is stopped.

Parallel Pattern. Assume a node N1 starting a parallel pattern whose first nodes are respectively N2 and N3. In this case, N1 sends one distinct workflow branch to N2 and N3. Two concurrent instances of the heartbeat protocol are started between N1 and N2 on one hand and N1 and N3 on the other hand. When N2 (respectively N3) completes its service and if there are other services in this branch, then N2 (resp. N3) forwards the residual branch to another node say N4, (resp. N5). When N4 and N5 receive their respective residual branch, N2 and N3 stops sending heartbeats to N1 but two new heartbeats protocols are started between N2 and N4 on one hand, and N3 and N5 on the other hand. The process of shifting the heartbeat along the branch is done in parallel inside the two branches until reaching their ends.

Conditional Validation. A conditional pattern is converted to sequential one after the selection test. So, their validation is similar to the sequential case.

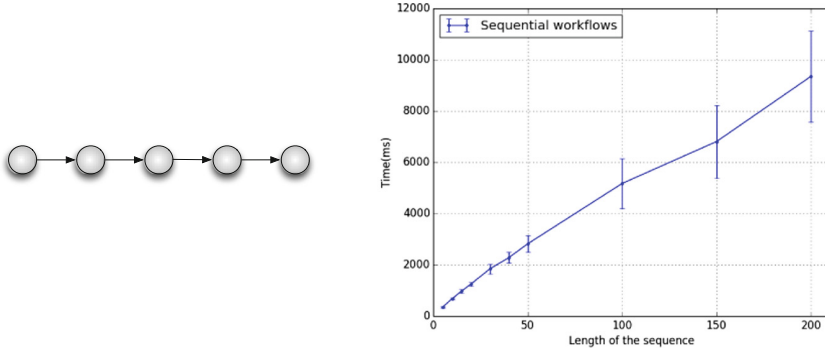


Fig. 9. Experiments with sequential workflows.

5 Experimental Validation

To validate BEDWE, we developed a Java-based software prototype implementing the algorithms described above. It represents more than 1700 lines of code. The prototype has been enhanced with the resilience mechanism described in Sect. 4. Deploying BEDWE means deploying BEDWE agents (as described in Sect. 3) implementing the algorithms described above, and using sockets to communicate portions of workflows, acks and heartbeats. The prototype was deployed over the Grid'5000 platform which gathers more than 8000 compute cores distributed over 8 geographically distributed sites [3]. In the following experiments, a set of BEDWE agents were deployed over computing cores of Grid'5000, and different workflows were submitted to these agents. All the experiments were conducted on the *parapide* cluster located in Rennes, composed of Intel Xeon X5570 CPUs with 8 cores and 24 GBs of memory each. For whole set of experiments, 40 BEDWE agents were deployed over 40 different cores. Agents are initially *idle*: they are listening for an incoming workflow (or sub-workflow) to be submitted.

The first experiments were done on sequential workflows, *workflows* composed of a single sequential pattern whose length varied between 5 and 200, as illustrated in the right part of Fig. 9). The results are given in the right part of Fig. 9. We observe an execution time growing linearly with the number of services. This is to be expected in the sense that such a workflow does not require to have many agents to work at the same time. At most two agents are required to be active at the same time due to the resilience mechanisms: an agent that completed a service waits for the completion of the next service in the workflow, running in another agent, before becoming idle again.

The second set of experiments conducted was performed using parallel workflows. These workflows were composed of one fork giving birth to a number of branches varying between 2 and 20. The number of sequential services within a branch was kept constant, at 10. Such a workflow is illustrated for a number of branches of 5 in the left part of Fig. 10. The objective of this experiment was to

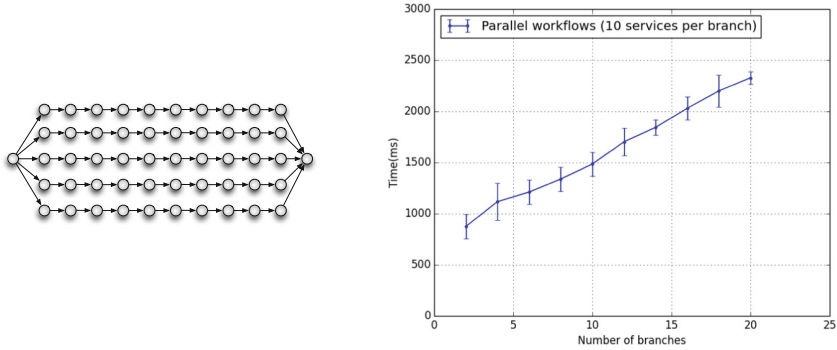


Fig. 10. Experiments with parallel workflows.

test the ability of BEDWE to tackle high parallelism. The result is shown in the right part of Fig. 10. It suggests that the BEDWE prototype is able to leverage the parallelism: recall that in the previous, sequential experiment, 200 sequential services were completed in more than 9 s. Now, when we have 20 branches, we have also 200 services, but the length of the branches are only 10, and the completion time is between 2 and 2.5 s, which is far closer to the case of a sequential case of 10 services (even if still significantly longer).

The last experiment was performed on workflows with nested forks. The workflows used in this part are all composed of 100 services, but the number of nested forks varies from 1 to 5. This supposes to adapt the number of services per branch for each case. The particular case illustrated in the left part of Fig. 11 is for the case of 2 nested forks, resulting in 25 services in each branches. For cases where the total number of services cannot be divided by the number of branches, few isolated services were added between forks. In our experiments, as plotted in the right part of Fig. 11, we observed a slightly increasing completion time, which can be explained, in spite of the higher degree of parallelism, by the increased complexity of forks management. In particular, the more nested forks, the more nodes waiting for the completion of the fork they are responsible for, making less nodes available to run tasks and an increase in sequentiality.

6 Related Work

Workflow execution is a topic which is not settled yet, as highlighted by the recent articles covering it in literature [4, 5]. Traditionally, workflow execution is centralized, would it be in an industrial [1] or more academic/scientific context [6, 7]. While very mature and efficient, these tools cannot be used in a context where the continuous presence of an orchestrating tool is mandatory.

The idea of describing direct interactions between services in the execution of a composition was proposed in the concept of *choreography of web services* [8]. Choreography allows to describe at once all the interactions that will take place at execution time. It allows to be more precise than orchestration in which the

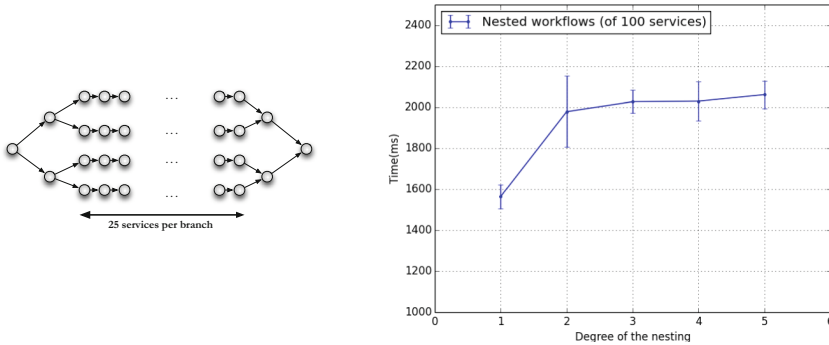


Fig. 11. Experiments with nested parallel workflows.

execution is seen from a single point of view, and to describe interactions between different organizations each providing part of the services to be combined [9].

Decentralising the workflow execution by relying on direct interactions (based on messaging) between services has been proposed in [10–12]. In particular, A continuation-passing style, where information on the remainder of the execution is carried in messages, has been proposed in [11]. Nodes interpret such messages and thus conduct the execution of services without consulting a centralised engine. However, nodes need to know explicitly which nodes to interact with and when, in a synchronous manner. A similar idea, based on the dynamic partitioning of the workflow as its execution moves forward has also been studied in [13]. Bedwe follows a similar principle but focus on parallel splits and choices, while extending such mechanisms with a particular protocol for fault-tolerance.

7 Conclusion

This paper has proposed BEDWE to decentralize workflow execution over unreliable platforms. The platform envisioned in this work is a set of compute nodes in regions subject to power cuts according to a predefined schedule, due to an insufficient level of electric injection to satisfy all the needs. Ensuring the completion of workflow execution on computers located in such an environment calls for decentralization and fault-tolerance. The BEDWE engine is supposed to run on every node taking part in the workflow engine. Engines use a specific protocol allowing to send, receive and process portions of the workflow, to split the workflow into several parallel executions, and synchronize their outcomes. Nodes taking care of contiguous portions of the workflow are watching each others. A BEDWE prototype was implemented in Java and validated over the Grid’5000 platform. Our future work will include devising a solution based on BEDWE for executing workflows in an African context, specifically for the city of Lome, Togo, which is subject to planned power cuts. With such a solution, a workflow

execution will dynamically move from power outages area to some neighborhood supplied with power. While the targeted platform is very different from the platform used for the experiment, the present article was about validating the approach.

References

1. Weerawarana, S., Curbera, F., Leymann, F., Storey, T., Ferguson, D.F.: Web Services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging and More. Prentice Hall PTR, Upper Saddle River (2005)
2. Levine, J., John, L.: Flex & Bison, 1st edn. O'Reilly Media Inc., Sebastopol (2009)
3. Bolze, R., et al.: Grid'5000: a large scale and highly reconfigurable experimental grid testbed. *Int. J. High Perform. Comput. Appl.* **20**(4), 481–494 (2006)
4. Hi-WAY: Execution of Scientific Workflows on Hadoop YARN, 21–24 March 2017
5. Marozzo, F., Duro, F.R., Blas, F.J.G., Carretero, J., Talia, D., Trunfio, P.: A data-aware scheduling strategy for workflow execution in clouds. *Concurr. Comput. Pract. Exp.* **29**(24) (2017)
6. Wolstencroft, K., et al.: The Taverna workflow suite: designing and executing workflows of web services on the desktop, web or in the cloud. *Nucl. Acids Res.* **41**(Webserver–Issue), 557–561 (2013)
7. Ludäscher, B., et al.: Scientific workflow management and the Kepler system. *Concurr. Comput. Pract. Exp.* **18**(10), 1039–1065 (2006)
8. Barros, A., Dumas, M., Oaks, P.: A critical overview of the web services choreography description language. *BPTrends* (2005)
9. Qiao, X., Wei, J.: A decentralized services choreography approach for business collaboration. In: *International Conference on Services Computing (SCC 2006)*, Chicago, USA, pp. 190–197 (2006)
10. Micillo, R.A., Venticinquie, S., Mazzocca, N., Aversa, R.: An agent-based approach for distributed execution of composite web services. In: *Proceedings of the 17th IEEE International Workshops on Enabling Technologies*, Rome, Italy, June 2008
11. Yu, W.: Consistent and decentralized orchestration of BPEL processes. In: *Proceedings of the 24th ACM Symposium on Applied Computing (SAC)*, Honolulu, March 2009
12. Downes, P., Curran, O., Cunniffe, J., Shearer, A.: Distributed radiotherapy simulation with the webcom workflow system. *Int. J. High Perform. Comput. Appl.* **24**, 213–227 (2010)
13. Atluri, V., Chun, S.A., Mukkamala, R., Mazzoleni, P.: A decentralized execution model for inter-organizational workflows. *Distrib. Parallel Databases* **22**(1), 55–83 (2007)



A Robust Process to Identify Pivots Inside Sub-communities in Social Networks

Joseph Ndong¹ and Ibrahima Gueye²(✉)

¹ Department of Mathematics and Computer Science,
University Cheikh Anta Diop, BP 5005 Fann Dakar, Senegal

joseph.ndong@ucad.edu.sn

² Computer Engineering and Telecom, Polytechnic School, Thies, Senegal
igueye@ept.sn

Abstract. In this work, we extend a previous work where we proposed a suitable state model built from a Karhunen-Loeve Transformation to build a new decision process from which, we can extract useful knowledge and information about the identified underlying sub-communities from an initial network. The aim of the method is to build a framework for a multi-level knowledge retrieval. Besides the capacity of the methodology to reduce the high dimensionality of the data, the new detection scheme is able to extract, from the sub-communities, the dense sub-groups with the definition and formulation of new quantities related to the notions of energy and co-energy. The energy of a node is defined as the rate of its participation to the set of activities while the notion of co-energy defines the rate of interaction/link between two nodes. These two important features are used to make each link weighted and bounded, so that we are able to perform a thorough refinement of the sub-community discovery. This study allows to perform a multi-level analysis by extracting information either per-link or per-intra-sub-community. As an improvement of this work, we define the notion of pivot to relate the node(s) with the greatest influence in the network. We propose the use of a thorough tool based on the formulation of the transformation of a suitable probabilistic model into a possibilistic model to extract these pivot(s) which are the nodes that control the evolution of the community.

Keywords: Social network analysis · Community detection · Energy
Pivot · Influencer

1 Introduction

Social networks describe web-based services that allow users (individuals) to connect with other users, communicate, share or publish contents within the network [1, 2]. The rise of web 2.0 has come with the ease in the production and sharing of content that allowed the social networks development. Users themselves become the producers of web content [1, 3]. In a formal and simple description, social network can be seen as a graph consisting of nodes (as individuals) and links (as social links) used to represent social relations on social network sites [4]. These last years, Social networks sites have

become some important sources of online interactions, contents sharing [5] and communication means. While giving these opportunities, they are affordable and universally acclaimed. Social network sites are commonly known for information dissemination, personal activities posting [6], product reviews, online pictures sharing, professional profiling, advertisements, subjectivity [6], assessments [7], approaches [8], influences [9], observations [10], feelings [3], opinions and sentiments expressions [11], news, remarks, reactions, or some many other content type [12]. News alerts, breaking news, political debates and government policy are also posted and analyzed on social network sites.

Observations [13] show that more and more people are becoming interested and involved in social networks. Sometimes, users of social networks exploit them for making decisions. These decisions can be made based on know or familiar users or, in some case, based on unfamiliar users [11]. Such a situation increases the degree of confidence in the credibility of these sites. The social network has transformed the way different entities procure and retrieve valuable information regardless of their location. The social network has also given users the privilege of giving their opinion.

The massive data generated or produced by all the social networks users allows for thorough analysis to efficiently extract useful knowledge such as trends, opinion leaders, influencers, feelings, etc. Such an analysis can be done as a whole, which means the representations of all the social network actors then proceed to the detection of communities and/or opinion leaders. The main aim of community detection methods is to partition the network into dense regions of the graph. Those dense regions typically correspond to entities which are closely related. The determination of such communities is useful in the context of a variety of applications such as customer segmentation, recommendations, link inference and influence analysis [14].

In this work, we focus on community influencers identification. We extend our previous work [15] in which, we derive a suitable state model from Karhunen-Loeve Transformation [16]. From this state model we built a new decision process, which allows us to extract useful knowledge and information about the identified underlying sub-communities from an initial network.

This extension is interested in the extraction of the most “important” nodes in given sub-communities already detected or known. Identifying sub-groups under a social network is a great challenge in the area of social network analysis [17–19], but analyzing the intrinsic behavior of each sub-group might be an important need if one desires to focus on the quality of nodes/actors [20]. If sub-communities are identified, decisions could be taken on a group of nodes regardless the underlying contents of the group itself since it can be viewed as a single homogeneous entity in which all items are of same behavior. One could also see the sub-community as a machine where the dynamic structure and the evolution depend strongly on the nature and quality of its combined pieces. If this second point-of-view is adopted, we need a tool to identify the “most important” pieces whose actions are more influential than those of the rest of the group. We define these pieces as “pivots” and the quality of the network is as important as it contains influential pivots with high degree of influence. In the following, we define pivots and how they could be important for a manager, a advertiser and so on.

Inside a sub-community, nodes share links so, identifying pivots can be helpful in many situations:

- (i) if in a sub-community, the pivot(s) is/are linked to a few other nodes, the manager can conclude that this network is of less quality simply because the “most” important nodes have less interactivity and thus less influence. But in the contrary, he can pay more attention to the underlying evolution of the group;
- (ii) the different sub-communities can be classified in an ascending order based on their quality – the first group is the one with the greater number of pivots, and so on;
- (iii) in a sub-community, we can find other sub-communities in the following manner: If the pivots themselves are not linked to each other, each of them can be the center of a group/entity formed by all the nodes linked to a pivot.

The procedure for sub-communities detection/identification we built in our previous work [15] consisted in extracting a decision variable from a state-space model we defined with a Karuhen-Loeve Transformation (KLT). This have served also to reducing the dimensionality of the dataset in order to maintain the only relevant part of the data. Then, after defining some new quantities as “energy” of an actor/node and the “co-energy between” actors, we apply a decision process to identify the sub-communities and the features inside each of them. The improvement we bring in this present work is the ability to learn more about the impact of node’s energy to see if some nodes could be selected as those with the highest degree of influence. The notion of highest degree must be properly define to achieve this aim. We found that a suitable definition of a probability density function (pdf) related to the defined node’s energy can be done. Then, a second definition of a possibilistic model will be able to discover the relevant node(s) we classify as pivot(s).

The rest of this paper is organized as follow: In Sect. 2 we do a thorough study of the related work; in Sect. 3 we detail the methodology and algorithm for the sub-community detection; in Sect. 4 we present our use of possibility theory to identify the pivots; in Sect. 5 we present the results of our experimental validation of our propositions, and we conclude this paper in Sect. 6.

2 Background

Social sites have undoubtedly bestowed unimaginable privilege on their users to access readily available never-ending uncensored information. Twitter, for example, permits its users to post events in real time way ahead the broadcast of such events on traditional news media. Also, social network allow users to express their views, be it positive or negative [20]. Organizations are now conscious of the significance of consumers’ opinions posted on social network sites to the patronage of their products or services and the overall success of their organisations. On the other hand, important personalities such as celebrities and government officials are being conscious of how they are perceived on social network. These entities follow the activities on social network to keep abreast with how their audience reacts to issues that concerns them [21].

Opinion of influencers on social network is based largely on their personal views and cannot be hold as absolute fact. However, their opinions are capable of affecting the

decisions of other users on diverse subject matters. For example, Rihanna's rant has cost Snapchat \$800 million in one Day [22]. Opinions of influential users on Social network often count, resulting in opinion formation evolvement. Clustering technique of data mining can be used to model opinion formation by assessing the affected nodes and unaffected nodes. Users that depict the same opinion are linked under the same nodes and those with opposing opinion are linked in other nodes. This concept is referred to as homophily in social network [23]. Homophily can also be demonstrated using other criteria such as race and gender [24].

Researchers in social network analysis are facing many other research issues and challenges such as in *Linkage-based and Structural Analysis*, or in *Dynamic Analysis and Static Analysis*.

Linkage-based and Structural Analysis consists of analyzing of the linkage behaviour of the social network so as to ascertain relevant nodes, links, communities and imminent areas of the network - Aggarwal [20].

Static analysis, such as in bibliographic networks, is presumed to be easier to carry out than those in streaming networks. In static analysis, it is presumed that social network changes gradually over time and analysis on the entire network can be done in batch mode. Conversely, dynamic analysis of streaming networks like Facebook and YouTube are very difficult to carry out. Data on these networks are generated at high speed and capacity. Dynamic analysis of these networks are often in the area of interactions between entities such as in Papadopoulos et al. [25] and Sarr et al. [26]. Dynamic analysis is also covered through temporal events on social networks such as in Adedoyin-Olowe et al. [27] and Becker et al. [17], and we also have such analysis through the study of evolving communities - Fortunato [18], Sarr et al. [19].

3 Methodology and Algorithm for the Sub-communities Detection

This methodology tracks and detects sub-communities based on the analysis of a huge number of features corresponding to events/activities for which a group of actors/nodes participate. First, we aim at finding the main features, to incorporate in our model, by means of extended principal component analysis. The second relevant issue of this methodology is related to the specification of a new detection procedure consisting of merging all the relevant features into a single process we will label as a "Decision Variable" (DV). By analyzing this process for the sub-community tracking operation, we can discover subgroups of actors using a multi-level thresholding and the notion of "energy dissipation" of an actor over the events.

We consider a community of R actors $\Omega = (a_1, \dots, a_R)$ which perform activities on a set of K initial correlated events (e_1, \dots, e_K) . For each event e_k , we have a column vector of size R containing the amount of participation of all R actors to the corresponding activity. This operation gives us the $R \times K$ matrix of correlated random variables $X = (X_1, \dots, X_K)$. In other words, one observes these random variables through R independent realization vectors $x^i = (x_1^i, \dots, x_K^i)$ $i = 1, \dots, R$.

After extracting the relevant components from the Karhunen-Loeve transformation [15], we can build our decision variable as a row vector $DV = (y_1, \dots, y_K)$. Then we can set a certain number of concepts for our methodology. We introduce the notion

of “energy dissipation” (Ed) to quantify the degree of importance a given actor puts on a series of events. This notion is simple and intuitive. When considering the set of events/activities, the events for which the actor puts a high degree of importance constitutes his energy. For example, we can consider money as energy. When someone goes to buy some products, we can say that he/she is dissipating a certain amount of his/her energy. In this case, he/she should buy a “product A”, and consequently buy another “product B” necessary to use the product A. Here, we can see the notion of correlation between these products/variables. When an athlete performs several disciplinary exercises in sport, we can view his actions as the dissipation of his energy over the different events, in order to win a medal. The energy of an actor is then quantifiable, its a measure of the strength of his participation to the a series of activities.

If the actor participates actively to all or most of the activities with a high intensity, then his energy increases, otherwise we say that this actor has less energy according to the ensemble of events happening at a given period of time.

Since the DV variable contains the aggregated amount of all actors participation to all events, the energy dissipation Ed of an actor i is the row vector defined as:

$$Ed_i = \{k, /x_k^i \geq DV[k], \forall k = 1, \dots, K\} \quad (1)$$

Ed_i contains all the index of events for which the energy dissipation is greater than the reference DV. Consequently, we can calculate the total energy of the actor i as:

$$E_i = \frac{|Ed_i|}{|DV|} \quad (2)$$

where $|\cdot|$ indicates the size of a vector.

We also refer to the notion of “co-energy” dissipation (CED) as the amount of energy between two actors according to their participation to the same set of activities. This quantity is a measure of the mean energy produced simultaneously by the two actors on the same activities:

$$CED_{ij} = \frac{|(Ed_i \cap Ed_j)|}{|DV|} \quad (3)$$

Finally, our detection procedure boils down to fix a threshold α and put a link between actor i and actor j if the rate of their co-energy exceeds the limit α . This means the following inequality must be held to add the link:

$$CED_{ij} \geq \alpha \quad (4)$$

When Eq. 4 holds, the value of CED_{ij} becomes the weight of the link between actor i and actor j . And then, this link is bounded by the interval $[\min(E_i, E_j), \max(E_i, E_j)]$. By varying the threshold $\alpha \in [0; 1]$, one can build many different sub-communities with the same dataset, each sub-community with a score α which measures its degree of realization. The algorithm to achieve our aim is described as follow:

4 Using Possibility Theory to Identify Pivot(s)

Our pivot identification methodology relies mainly in the feature of energy we defined for the actors/nodes. Pivots refer to nodes with “high” energy. These nodes have the opportunity to control the dynamic evolution of the network. If the energy of a pivot decreases or increases, the structure of the network might evolve towards a new direction allowing to suppress or add links between nodes. A simple question arises from the perspective of pivots identification: how much energy is necessary for a node to be classified as a pivot? We believe that it is very difficult to answer this question by analyzing only the amount of energy of each node. If someone would like to do so, he/she should build a kind of threshold and apply the decision to put the label “pivot” on a node if its energy exceeds this limit. This methodology weakens the objectivity of pivots’ identification. To surround this difficulty, we propose a more robust identification scheme based on a link between probability theory and possibility theory.

Algorithm 1. Sub-Community Discovering

Input : C_t , a community
 $\Omega(C_t)$, the sets of actors within C_t
 $x^i = (x_1^i, \dots, x_K^i)$ the vector of participation of actor i
 $DV = (dv_1, \dots, dv_K)$ the decision variable
 α , the link detection threshold

Output : V , a sub-community

```

1 /* Calculate Co-Energy dissipation between actors and apply
   threshold to add link*/
2 begin
3   foreach  $(k, l) \in \Omega(C_t), k \neq l$  do
4     /* Apply Eq. (1) */
5      $Ed_k = \{p, /x_p^k \geq dv_p, \forall p = 1, \dots, K\}$ 
6      $Ed_l = \{p, /x_p^l \geq dv_p, \forall p = 1, \dots, K\}$ 
7     /* Apply Eq. (3) */
8      $CED_{kl} = \frac{|(Ed_k \cap Ed_l)|}{|DV|}$ 
9   /*Apply threshold to decide to put a link, Eq. (4)*/
10  if  $CED_{kl} \geq \alpha$  then
11     $addLink(V, k, l)$ 
12  return  $V$ 
```

The energy property can be viewed as a continuous random variable $X : \Omega \rightarrow V$, where Ω is the set of actors/nodes and V the measurable function giving the real value of the energy as defined in Eq. 2. X does not return a probability. But we want to know, if it were the case, could this probability help achieving our goal.

Probability theory is a valuable quantitative tool to study randomness/uncertainty in random phenomena. In this area, we can find the probability of occurrence of different possible outcomes in an experiment. If a random variable returns a probability P , then

$V = [0, 1]$ and $\sum_u P(X = u) = 1$. As we defined the energy in Eq. 2, the energy E_i of actor i is always between 0 and 1. So, we can take $V = [0, 1]$, but we do not have $\sum_i^K E_i = 1$ (K is the number of nodes).

4.1 Defining Probability to Characterize Energy

For the purpose to return a probability from X , we have just to normalize the energy by the size of \mathbb{E} , where $\mathbb{E} = \{E_i, i = 1 \dots, |\Omega|\}$. At this point, the random variable X returns a probability since the equation $\sum_u P(X = u) = 1$ holds. We build this probability to characterize each energy of a node with a probability of occurrence in $[0, 1]$. Now, using this probability definition, our scope is to build a test to identify the potential pivots under each sub-community. Generally, a typical test consists of applying a threshold on the outcomes (probabilities of energy) to decide to put the label pivot on a node if the probability distribution takes a value higher than this threshold. The task to build this threshold is not straightforward even if, for given nodes if the probability of their respective energy is 0.95, 0.6 and 0.3 for example, how could we decide to label a node as pivot, based only on the not obvious notion of “high” probability. Is 0.6 a “high” or “low” probability? By pointing out this example, we simply want to show that it is not evident to know the “best” value of the probability threshold in order to conclude if the node is a pivot or not. Nevertheless, one could, for simplicity, build an heuristic decision process where the threshold is set manually. With probability, the only evident decision we can take is, when the probability of a node is $P(X = u) = 1$. If this case happens, it means that there’s only one pivot in the entire sub-community, since the other nodes have probability 0 and so this sub-community is built with only one node. Finally, we believe that pivots might be nodes with any probability, different to zero, “sufficient” to become member of a sub-community and to have the potential to change the dynamic evolution of the network. To go towards the direction of extracting the “best” level of probability measure, we think of an alternative related the area of possibility theory, which gives as another tool to represent uncertainty in a qualitative fashion. This tool can also helps to learn more about the **incompleteness** that reflects the lack of information. We have seen above that, our defined probability doesn’t give us the information about the limit to apply to detect pivots. So, the idea behind the use of this new scheme is to associate to each node, in accordance to its energy, a degree of possibility which quantifies the level of “importance” of that node among the others.

4.2 Detecting Pivot by Possibility Degree

Between probability and possibility, we can state a consistent principle is this terms: “*what is probable should be possible*” [28]. This requirement can then be translated as follow:

$$P(A) \leq \Pi(A) \quad \forall A \subseteq \Omega \quad (5)$$

where P and Π are, respectively, the probability and the possibility measure on the domain Ω . In this case, Π is said to **dominate** P . With Π , given nodes should have

the maximum possibility degree, i.e. the value 1. In possibility theory the equality $\sum_u \Pi(X = u) = 1$ is not guaranteed. So, for any node, when its possibility degree reaches the maximum, we can robustly say that this node is a pivot since it is entirely sure that it might exist in the network. So, we see that the notion of “high” probability can be defined properly, since it corresponds to the maximum degree of possibility.

A possibility measure, [29], Π on V is characterized by a possibility distribution $\pi : V \rightarrow [0, 1]$, and is defined by:

$$\forall A \subseteq V, \Pi(A) = \sup\{\pi(v), v \in A\}. \quad (6)$$

If V is a finite set, thus: $\forall A \subseteq V, \Pi(A) = \max\{\pi(v), v \in A\}$. The key concept of a possibility distribution is the preference ordering it establishes on V . Basically, π designates what one knows about the value of a given variable X , and $\pi(v) > \pi(v')$ states that $X = v$ is more plausible than $X = v'$. When $\pi(v) = 0$, thus, v is an impossible value of the variable X while $\pi(v) = 1$ means that v is one of the most plausible values of X . For us, identifying pivots is just a process to searching at these most plausible values.

Transforming a probability measure into a possibilistic one then amounts to choosing a possibility measure in the set $\mathfrak{P}(P)$ of possibility measures dominating P . This should be done by adding a strong order preservation constraint, which ensures the preservation of the shape of the distribution:

$$p_i < p_j \Leftrightarrow \pi_i < \pi_j \quad \forall i, j \in \{1, \dots, q\}, \quad (7)$$

where $p_i = P(\{E_i\})$ and $\pi_i = \Pi(\{E_i\})$, $\forall i \in \{1, \dots, K\}$. It is possible to search for the most specific possibility distribution verifying (5) and (7). The solution of this problem exists, is unique and can be described as follows. One can define a strict partial order \mathbf{P} on Ω represented by a set of compatible linear extensions $\Lambda(\mathbf{P}) = \{l_u, u = 1, L\}$. To each possible linear order l_u , one can associate a permutation σ_u of the set $\{1, \dots, q\}$ such that:

$$\sigma_u(i) < \sigma_u(j) \Leftrightarrow (\omega_{\sigma_u(i)}, \omega_{\sigma_u(j)}) \in l_u, \quad (8)$$

The most specific possibility distribution, compatible with the probability distribution (p_1, p_2, \dots, p_K) can then be obtained by taking the maximum over all possible permutations:

$$\pi_i = \max_{u=1, L} \sum_{\{j | \sigma_u^{-1}(j) \leq \sigma_u^{-1}(i)\}} p_j \quad (9)$$

Finally, the vector $(\pi_1, \pi_2, \dots, \pi_K)$ gives us all the possibility degrees for the K nodes, corresponding to the (p_1, p_2, \dots, p_K) vector of probability of their energies. And pivots are nodes for which the possibility degree exceeds a given rate δ . For this study we set this threshold $\delta = 1$, it corresponds to the maximum value a possibility degree might be set. A more flexible and non-heuristic method might be to set δ to a value less than the maximum. But, here, we set δ to 1 in order to show that in all situations or scenario, a pivot must be found.

For a thorough view of possibility theory, we recommend the reader to [28–30].

5 Validation

We validate our approach on the real world collection of data coming from **Reddit.com** [31]. We use several samples of different sizes and, build four scenarios A, B, C and D with dimension $(N \times K, N$ the number of actors and K the number of events) 10×15 , 10×150 , 10×500 and 10×1200 respectively. In Table 1, we give an idea on the content of the data, in each column vector, we have the total amount of submissions to an image by the set of actors.

Table 1. Activities and amount of actor participation to submissions on events. Scenario A.

Actors	Events														
	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}	e_{11}	e_{12}	e_{13}	e_{14}	e_{15}
1	11	0	11	4	0	2	0	4	18	2	0	6	16	1	0
2	5	0	0	0	0	0	1	0	0	0	2	0	2	0	0
3	1	0	2	0	3	1	1	0	1	0	0	2	1	0	2
4	4	1	0	0	0	1	2	0	7	0	1	0	9	1	0
5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
7	0	2	0	0	1	0	0	0	0	0	0	4	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

In the previous experiment our previous work [15] (that we can not show again due to a limit page number), we retrieved two levels of information can be retrieved from the results. In the first level, we had the results about the formation of the underlying sub-communities. It corresponds to the natural clustering of the different nodes according to the energy provided by each of them. The second level of information refered to the characteristics of links and nodes inside the given sub-groups. This refinement provides useful information when one wants to emphasize and explore some parts of the network.

Here in below, we discuss the obtained results about the pivots identification.

Discussion About the Pivots Identification Results

We apply the detection process to identify pivots to the data samples for scenarios A, B, C, D, E, F and G. We have arbitrarily chosen the size of the sample and the interval where data come from. We just want to show cases where there's one pivot or more. The Table 2 resumes the data selection scheme, and in we put the results of the detection procedure in Table 3 where the variable NBE refers to the total amount of participation on events by nodes, p is the probability of the energy and π the corresponding degree of possibility. The results give many useful information. For the scenarios A, B, C and D, we extract only one pivot which have the particularity to be the node with the highest amount of activities in the network. The pivot also has the highest probability of energy

Table 2. Information about the selection of the different data samples

Scenario	A	B	C	D	E	F	G
Size	15	150	500	1200	155	255	555
<i>interval</i>	[1:15]	[1:150]	[1:500]	[1:1200]	[1100:1254]	[1000:1254]	[700:1254]

Table 3. Detection of pivots when degree of possibility $\pi = 1$

Nodes i	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	Scenario
<i>NBE</i>	75	10	14	26	1	2	7	0	0	0	A
p	0.30	0.10	0.25	0.25	0	0	0.10	0	0	0	
π	1	0.20	0.70	0.70	0	0	0.20	0	0	0	
pivot	Yes	No	No	No	No	No	No	No	No	No	
<i>NBE</i>	591	28	218	142	44	31	117	53	10	8	B
p	0.42	0.06	0.18	0.23	0	0	0.11	0	0	0	
π	1	0.06	0.35	0.59	0	0	0.17	0	0	0	
pivot	Yes	No	No	No	No	No	No	No	No	No	
<i>NBE</i>	1638	108	668	348	120	63	348	170	27	16	C
p	0.39	0.11	0.17	0.22	0	0	0.11	0	0	0	
π	1	0.22	0.38	0.61	0	0	0.22	0	0	0	
pivot	Yes	No	No	No	No	No	No	No	No	No	
<i>NBE</i>	3659	257	1596	773	206	164	890	425	59	40	D
p	0.39	0.11	0.17	0.22	0	0	0.11	0	0	0	
π	1	0.22	0.38	0.61	0	0	0.22	0	0	0	
pivot	Yes	No	No	No	No	No	No	No	No	No	
<i>NBE</i>	428	34	204	95	26	40	52	40	6	4	E
p	0.38	0	0.38	0	0.06	0.06	0.06	0.06	0	0	
π	1	0	1	0	0.25	0.25	0.25	0.25	0	0	
pivot	Yes	No	Yes	No	No	No	No	No	No	No	
<i>NBE</i>	702	50	321	140	38	54	122	82	13	8	F
p	0.32	0	0.32	0.07	0	0	0.07	0.15	0	0.07	
π	1	0	1	0.23	0	0	0.23	0.38	0	0.23	
pivot	Yes	No	Yes	No	No	No	No	No	No	No	
<i>NBE</i>	1569	127	663	335	72	81	405	170	22	16	G
p	0.32	0	0.38	0	0.06	0	0.18	0.06	0	0	
π	0.62	0	1	0	0.12	0	0.31	0.12	0	0	
pivot	No	No	Yes	No	No	No	No	No	No	No	

but, the value of this probability seems to be “very low” in a pure point-of-view of the probability theory. For example, for scenario A, the pivot has a probability $p = 0.30$ but its possibility degree reaches $\pi = 1$. For scenarios A, B, C and D, we find only one pivot

and it corresponds to the node with the highest probability and which have perform the greater amount of activities in the network. But this findings are not a generality since we see, in the scenario G, the pivot has the highest probability but it does not have the greatest amount of activities. This discover says clearly that, learning only the amount of activities is not a sufficient process to analyze a network in order to detect links or to put a level of importance/quality to nodes. In scenarios E and F, we detect twos pivots which have not the same amount of actions but they have the same probability of energy. We think that this situation is do to the fact that the quality of node depends not only to its level of participation on events but, it depends also to its interaction with other nodes. So one must have two (or more) nodes with different level of actions on events but they have the same behavior regarding to other nodes.

6 Conclusion

In this work, we have extent a new technique related to an extended version of principal component analysis to build a methodology for the purpose of community detection in a social network. The initial work built a technique more elaborated to run within stochastic process than the classical PCA which is designed originally to solve the problem of dimensionality reduction for univariate dataset. The main innovation of the work is manifold: (i) we define the notion of co-energy between two nodes to quantify the intensity of their relation, (ii) we can also extract the proper energy of a given node to know how it influences the overall community, (iii) technically, the KL-PCA technique makes possible to build a decision variable and to form a state model from which we apply a decision process to identify each link. The introduction of the notion of energy make possible to see potential intra sub-communities (i.e. nodes with the same co-energy) inside a sub-community; (iv) each detected link is bounded, so we know how much energy is necessary to maintain a link over time. In this complementary study, we show that a possibility distribution can be properly defined from the energy to solve an interesting feature i.e, the problem of identifying pivots which have the main impact in the dynamic nature of the network. From this work, we plane to learn the impact of the number of pivots in a given sub-community and between sub-communities to face the idea related to their impact on a network distributed in many geographic area.

References

1. Adedoyin-Olowe, M., Gaber, M.M., Stahl, F.T.: A survey of data mining techniques for social media analysis. *JDMDH* **2014** (2014)
2. Chen, Z., Kalashnikov, D.V., Mehrotra, S.: Exploiting context analysis for combining multiple entity resolution systems. In: *ACM SIGMOD International Conference on Management of Data, SIGMOD 2009*, pp. 207–218 (2009)
3. Kaplan, A.M., Haenlein, M.: Users of the world, unite! the challenges and opportunities of social media. *Bus. Horizons* **53**(1), 59–68 (2010)
4. Borgatti, S.P.: Social network analysis, two-mode concepts in. In: Meyers, R.A. (ed.) *Computational Complexity*, pp. 2912–2924. Springer, Heidelberg (2012). https://doi.org/10.1007/978-1-4614-1800-9_179

5. Chelmiss, C., Prasanna, V.K.: Social networking analysis: a state of the art and the effect of semantics. In: IEEE International Conference on Social Computing, pp. 531–536 (2011)
6. Asur, S., Huberman, B.A.: Predicting the future with social media. In: Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology, pp. 492–499. IEEE Computer Society (2010)
7. Kim, Y., Hsu, S.-H., de Zúñiga, H.G.: Influence of social media use on discussion network heterogeneity and civic engagement: the moderating role of personality traits. *J. Commun.* **63**(3), 498–516 (2013)
8. Korda, H., Itani, Z.: Harnessing social media for health promotion and behavior change. *Health Promot. Practice* **14**(1), 15–23 (2013)
9. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Identifying ‘influencers’ on Twitter. In: Fourth ACM International Conference on Web Search and Data Mining (WSDM) (2011)
10. Atkinson, N., Chou, W.-Y.S., Hunt, Y.M., Beckjord, E.B., Moser, R.P., Hesse, B.W.: Social media use in the United States: implications for health communication. *J. Med. Internet Res.* (2009)
11. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.* **2**(1–2), 1–135 (2008)
12. Liu, B.: *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, San Rafael (2012)
13. Statista: Number of social media users worldwide from 2010 to 2021 (in billions) (2017). <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. Accessed 09 Oct 2017
14. Khatoon, M., Banu, A.: A survey on community detection methods in social networks. *I. J. Educ. Manag. Eng.* **1**, 8–18 (2015). <https://doi.org/10.5815/ijeme.2015.01.02>
15. Ndong, J., Gueye, I.: A new decision technique for sub-community and multi-level knowledge extraction in social networks. In: Cherifi, H., Gaito, S., Quattrociocchi, W., Sala, A. (eds.) *Complex Networks & Their Applications V. Complex Networks 2016*. SCI, vol. 693, pp. 263–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50901-3_21
16. Gray, R.M., Davisson, L.D.: *An Introduction to Statistical Signal Processing*. Cambridge University Press, Cambridge (2004)
17. Becker, H., Naaman, M., Gravano, L.: Beyond trending topics: real-world event identification on Twitter. In: *Proceedings of the Fifth International Conference on Weblogs and Social Media* (2011)
18. Fortunato, S.: Community detection in graphs. *Phys. Rep.* **486**(3), 75–174 (2010)
19. Sarr, I., Ndong, J., Missaoui, R.: Overlaying social networks of different perspectives for inter-network community evolution. In: Missaoui, R., Sarr, I. (eds.) *Social Network Analysis - Community Detection and Evolution*. LNSN, pp. 45–70. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12188-8_3
20. Aggarwal, C.C.: An introduction to social network data analytics. In: Aggarwal, C. (ed.) *Social Network Data Analytics*, pp. 1–15 (2011). https://doi.org/10.1007/978-1-4419-8462-3_1
21. Castellanos, M., et al.: LCI: a social channel analysis platform for live customer intelligence. In: *ACM SIGMOD International Conference on Management of Data*, pp. 1049–1058. ACM (2011)
22. FR, Y.N.: Rihanna’s rant has cost Snapchat \$ 800 million (2018). <https://fr.news.yahoo.com/coup-gueule-rihanna-fait-perdre-114958692.html>. Accessed 19 Mar 2018
23. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. *Ann. Rev. Sociol.* **27**(1), 415–444 (2001)
24. Jackson, M.O.: *Social and Economic Networks*. Princeton University Press, Princeton (2008)
25. Papadopoulos, S., Kompatsiaris, Y., Vakali, A., Spyridonos, P.: Community detection in social media. *Data Min. Knowl. Discov.* **24**(3), 515–554 (2012)

26. Sarr, I., Missaoui, R.: Temporal analysis on static and dynamic social networks topologies. In: Alhajj, R., Rokne, J. (eds.) *Encyclopedia of Social Network Analysis and Mining*. Springer, Heidelberg (2014). <https://doi.org/10.1007/978-1-4614-6170-8>
27. Adedoyin-Olowe, M., Gaber, M.M., Stahl, F.: TRCM: a methodology for temporal analysis of evolving concepts in Twitter. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) *ICAISC 2013. LNCS (LNAI)*, vol. 7895, pp. 135–145. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38610-7_13
28. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. In: *Fuzzy Sets and Systems*, pp. 3–28 (1978)
29. Dubois, D., Prade, H., Sandri, S.: On possibility/probability transformations. In: *Proceedings of the Fourth International Fuzzy Systems Association World Congress (IFSA 1991)*, Brussels, Belgium, pp. 50–53 (1991)
30. Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**, 338–353 (1965)
31. Gueye, I., Ndong, J., Sarr, I.: An accurate probabilistic model for community evolution analysis in social network. In: *International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pp. 343–349 (2015)
32. McGloin, J.M., Kirk, D.S.: An overview of social network analysis. *J. Crim. Justice Educ.* **21**(2), 169–181 (2010)
33. Wasserman, S., Faust, K.: *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge University Press, Cambridge (1994)

Author Index

- Abanda, Yannick 83
Agoungbome, Sehouevi M. D. 153
Alphonse, Mutabazi 163
- Behrmann, Marlene 72
Bézy, Michel 90
Bilal, Boudy Ould 23, 100
Bouchet, François 199
- Camara, Fodé 187
Carron, Thibault 199
Charki, Abdérafi 23
- Deen, Linda G. 62
Diallo, Moussa 175, 212
Diarra, Nafissatou 224
Diatta, Samo 123
Drame, Mamadou Simina 175
- Egbedewe, Aklesso Y. G. 133
Etienne, Arnelle 72
- Fall, Cheikh Modou Noreyni 175
Fouotsa, Emmanuel 224
- Garba, Aminata A. 52, 62
Gaye, Amadou Thierno 123
Grover, Pulkit 72
Gueye, Assane 3, 100
Gueye, Bamba 212
Gueye, Ibrahima 248
Gumyusenge, Florian 52
- Kama, Abdoulaye 175
Kebbeh, Pa Saffiong 212
Kébé, Cheikh Mouhamed Fadel 23, 41, 100
Kelly, Shawn K. 72
Kenfack, Gutenbert 83
Krishnan, Ashwati 72
- Kumar, Ritesh 72
Kwizera, Emelyne 62
- Mambo, Abdulhameed Danjuma 41
Marone, Reine Marie Ndéla 187
Mbatchou, Guy Merlin 199
Mbaye, Babacar 3
Mbaye, Mamadou Lamine 123
Mico, Darius 62
Mueller, Ulrike 111
Mugisha, Robert 52
- Nayebare, Michael 62
Ndiaye, Ababacar 23, 100
Ndiaye, Pape Alioune 23, 100
Ndiaye, Samba 187
Ndong, Joseph 248
Ngigi, Marther W. 111
Ngom, Bassirou 212
- Ould Mohamed Moctar, Ahmed 13
- Pabame, Zoutene 100
Pazat, Jean-Louis 236
Pearl, Nkusi 163
- Robinson, Amanda 72
- Saint, Martin 52, 62
Sambou, Vincent 23
Sarr, Idrissa 13
Seidou, Ousmane 143, 153
Seye, Madoune R. 212
Shema, Alain 90
- Tarr, Michael J. 72
Tedeschi, Cédric 236
Thiam, Moussa 153
Tiedeu, Alain 83
- Wallah, Palakiyem 236