Challenges and Prospects for Data-Driven Climate Change Mitigation

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Engineering and Public Policy

Lynn Helena Kaack

B.S., Physics, Freie Universität Berlin
M.S., Physics, Freie Universität Berlin
M.S., Machine Learning, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

May, 2019

Acknowledgments

First of all, I would like to thank the people at Carnegie Mellon University (CMU) for creating such a wonderfully interdisciplinary but profoundly technical environment, where it was a pleasure to be a graduate student.

All chapters were supported by the Center for Climate and Energy Decision Making (CEDM) through a cooperative agreement between the National Science Foundation and Carnegie Mellon University (SES-0949710). The work on Chapter One received additional financial support from the Electric Power Research Institute (EPRI). I am grateful to the U.S. Energy Information Administration for hosting a presentation and discussion about this work, in particular Faouzi Aloulou, David Daniels, and John Staub. Chapter Two benefited from the insights of participants in a workshop organized by CEDM at CMU in February 2017. Chapter Three would have not been possible without satellite image data provided by DigitalGlobe, Inc., and without computing resources that were made available through Google Cloud Platform and Microsoft AI for Earth. I would also like to thank Norichika Kanie and Lisanne Groen for hosting me as a visiting research assistant at the United Nations University Institute for the Advanced Study of Sustainability (UNU-IAS) in Tokyo during the summer of 2017.

I have deepest gratitude and admiration for my advisor and committee chair Granger Morgan. He has, often invisibly, shaped my thinking, and gave me the freedom and infinite support to develop my own research persona. Granger is a role model for me professionally but also in his ability to built long-lasting relationships, and to show trust and respect for everyone around him. I am lucky to have worked with and have been advised by my committee members Jay Apt, Parth Vaishnav, Patrick

McSharry, and Inês Azevedo, who have supported me throughout my PhD. Special thanks also to George Chen, who has assumed an advising role in my last year, and who worked substantially with me on Chapter Three.

I would like to further thank my teachers and colleagues, Lisanne Groen, Alex Davis, Max Henrion, David Daniels, Gabriel Katul, Diego Rybski, Amelia Haviland, Cosma Shalizi, Mitchell Small, and Michael Griffin, for shaping my path to becoming a researcher and inspiring me. My sincere thanks to the Engineering and Public Policy Staff for their timely and competent support.

I am grateful for the education that both the German and American societies have provided for me. The German education system enabled me to go my way, independently of my family's income or academic background, which is much appreciated.

I would also like to thank my friends for their incredible support and invaluable insights. I will never forget the countless days and nights spent over homework with my "fellow nerds" Rahul Ladhania and Cristóbal de la Maza. I am so glad that we did this stats and ML marathon together, and while painful, I wouldn't want to miss any of it. I am grateful to Erin Mayfield for being such a close friend, for all those great conversations, and for being a companion also professionally. I am so lucky to know Evan Sherwin, Jeremy Keen, Nichole Hanus (housemates for basically our entire PhD!), Octavio Mesner, Verena Beckert, Peter Tschofen, Priya Donti, and Brian Sergi.

Ringrazio moltissimo Marco Bertolini per essere stato al mio fianco durante questi anni. Per fare $> 30,000$ miglia con la Mustang per vedermi. Per viaggi matti. Per essere così vicino, anche quando abbiamo vissuto in due continenti diversi con troppo poco tempo di vita normale insieme. [*I am deeply grateful to Marco Bertolini for staying by my side during these years. For driving the Mustang* $> 30,000$ *miles to see me. For crazy travels. For being so close even when we lived on two different continents with way too little normal life together.*]

Grazie a Maurizia Bonini e Gianni Bertolini per avermi ospitato quando ho scritto buona parte della tesi. [*I would also like to thank Maurizia Bonini and Gianni Bertolini for hosting me when I wrote a good part of the dissertation.*]

Ganz viel Dank gilt meinen Eltern, Kirsten und Detlef, und meiner Schwester, Alice, die meine lange Abwesenheit ausgehalten haben. Meine Familie hat mich immer unterstützt in allen

Entscheidungen und mich mit unser aller Begeisterungsfähigkeit, Neugierde, Dickköpfigkeit, Diskutierfreudigkeit und den Erklärlawinen perfekt auf einen PhD vorbereitet. Ich möchte auch meiner Oma Herta für das viele Verständnis danken, besonders nachdem ich in acht Jahren nur einmal einen Geburtstag mitfeiern konnte, und meiner Oma Lotti, mit der ich gerne mehr Zeit in den letzten zwei Jahren ihres Lebens verbracht hätte. [*My deepest gratitude to my parents, Kirsten and Detlef, and my sister, Alice, who have endured my long absence. They have supported any decision I've made, and they have prepared me well for a PhD with our curiosity, determination, eagerness to discuss, and the many passionate explanations. I also would like to thank my grandmother Herta for her acceptance, especially after missing all but one of her birthdays in the past eight years, and my late grandmother Lotti, with whom I would have liked to spend more time in the last two years of her life.*]

Abstract

Successful climate change mitigation will require data-driven decision making, but the field faces a diverse set of challenges. In this dissertation, I provide three examples that illustrate how uncertainty is often not adequately characterized, how missing data can pose a barrier to climate-relevant policy making, and how big data and machine learning could be used to obtain important information. I conclude with a survey and a discussion of how artificial intelligence can be applied to climate change mitigation.

In the first chapter, I show how to construct an empirical estimate of the uncertainty of long-term energy forecasts based on past forecast errors, using projections made by the U.S. Energy Information Agency (EIA). This method gives analysts and decision-makers a means to estimate the uncertainty of those forecasts quantitatively. Energy forecasts provide the basis for financial evaluation of energy investments as well as for energy system models. I lay the groundwork for evaluating the performance of these methods in the data-scarce setting of long-term forecasts. The EIA has used my results in their most recent retrospective review.

The second chapter is based on a topical review of policies to decarbonize heavy freight transportation by shifting freight from road to rail and water. I find that while the freight sector is responsible for a large share of greenhouse gas (GHG) emissions, a systematic analysis of the potential to decarbonize with modal shift is still missing from the literature. This is partly due to a lack of publicly available, standardized, and updated data. For a global comparison of modal split and trends, I expanded existing databases with national freight activity from 2000-2017. I find that only less than half of the countries in the world provide such information on road freight activity.

The third chapter provides an example of how big data and machine learning (ML) could be used to fill in information gaps that inhibit climate policy analysis. I use satellite imagery for truck traffic monitoring in areas where this information is otherwise difficult to obtain. I count the number of freight vehicles visible in the images with deep convolutional neural networks, and estimate the average annual truck traffic on roads from those counts by modeling traffic variation patterns.

In a final chapter, I discuss how methods from artificial intelligence can be used to improve socioeconomic, policy, and engineering research for climate change mitigation. I provide a survey of the literature and identify the main barriers and challenges that arise at the intersection of those disciplines. Research in this area demands both careful design of ML algorithms and consideration of domain knowledge. I conclude with proposing a research agenda.

Per Marco

Für meine Familie

Contents

List of Tables

List of Figures

Introduction

More than twenty five years ago, the United Nations Framework Convention on Climate Change (UNFCCC) was established as a foundation for a global effort to minimize the anthropogenic impact on the earth's climate [1]. While international parties have agreed to keep global warming below 2°C [2, 3], it is likely that current pledged reductions will be insufficient to achieve this target [4]. To restrict anthropogenic greenhouse gas (GHG) emissions, global energy systems will need to be rapidly and deeply decarbonized [5, 6, 7]. My work affects two aspects that are critical to achieving an energy system with near-zero GHG emissions: long-term planning and energy services that are hard to decarbonize.

In the first chapter, I address a persistent problem in long-term forecasting, which is that energy forecasts typically consist of point projections and scenarios without associated probabilities. Long-term energy projections, such as those provided by the U.S. Energy Information Administration (EIA), are of tremendous importance for decarbonizing the energy system. It has been shown that targeted and technology-discriminating policies may help to move to a low-carbon energy system and to avoid lock-in into inferior technologies [8], for which the policy-maker would need to pick "winner" and "loser" technologies for years to come. Also, investment decisions in the wrong energy infrastructure can lead to carbon lock-in, as infrastructure is built to last for decades. Long-term energy forecasts, such as those by the U.S. EIA, critically affect these policy and investment decisions, making a good understanding of the uncertainty essential.

Chapters Two and Three focus on long-haul road freight, which is one of the energy services that are hardest to decarbonize [9]. Road freight is a carbon-intensive mode, and trucks were responsible for 7% of total world energy-related CO_2 emissions in 2015 [10]. In Chapter Two, I review five strategies that can be used to decarbonize road freight transportation, and then focus on an international comparison of policies for mitigating climate change through modal shift from road to rail or water. Because of its simplicity, modal shift is one of the most important means to reduce GHG emissions from the freight sector in the near term. Modal shift can be promoted by policies targeting infrastructure investments and internalizing external costs of road freight [11, 12, 13, 14], but not many countries have such policies in place. I also find that less than half the countries in the world report national road freight activity (in tonne-km). Much of the growth of road freight occurs in developing countries and emerging economies, and identifying ways to monitor truck traffic in these regions is one of the first steps to address the problem. Chapter Three provides a proof of concept of how to monitor truck traffic with remote sensing. In this chapter, I explore a method that might be used to provide information on the traffic density by using machine learning and high-resolution satellite images.

While climate change mitigation is at the heart of the topics addressed in this dissertation, there are also methodological and conceptual dimensions to the work. As the policy challenges change through time, so must the method toolbox of policy analysis[1] expand. For example in 1999, Morgan et al. [16] have argued that methods should account for the new spatial and temporal scale that is required for climate mitigation policies. More computing power, advanced analytical methods, and information and communication technologies that offer a wealth of new data sources, have changed the society [17], the economy, and the way we conduct science [18]. While data and statistics have always been important for policy analysis [15], I believe that this new regime of data analytics will have an impact on how policy analysis is conducted. This dissertation addresses two of the data challenges typically found in policy analysis - uncertain data and missing data - and explores how data-driven approaches leveraging statistics, artificial intelligence, and big data could provide answers. It also includes an analysis of how artificial intelligence can be used for policy analysis, forecasting, and engineering to reduce GHG emissions - and which challenges could arise from that approach.

[1]For an introduction to policy analysis see Morgan (2017) [15].

1
Empirical prediction intervals improve energy forecasting

Hundreds of organizations and analysts use energy projections, such as those contained in the U.S. Energy Information Administration (EIA)'s Annual Energy Outlook (AEO), for investment and policy decisions. Retrospective analyses of past AEO projections have shown that observed values can differ from the projection by several hundred percent, thus a thorough treatment of uncertainty is essential. We evaluate the out-of-sample forecasting performance of several empirical density forecasting methods using the continuous ranked probability score (CRPS). The analysis confirms that a Gaussian density, estimated on past forecasting errors, gives comparatively accurate uncertainty estimates over a variety of energy quantities in the AEO, in particular outperforming scenario projections provided in the AEO. We report probabilistic uncertainties for 18 core quantities of the AEO 2016 projections. Our work frames how to produce, evaluate and rank probabilistic forecasts in this setting. We propose a log-transformation of forecast errors for price projections, and a modified non-parametric empirical density forecasting method. Our findings give guidance on how to evaluate and communicate uncertainty in future energy outlooks.

This chapter has been published as: Kaack, L. H., Apt, J., Morgan, M. G., & McSharry, P. (2017). Empirical prediction intervals improve energy forecasting. Proceedings of the National Academy of Sciences, 114(33), 8752-8757. `https://doi.org/10.1073/pnas.1619938114`

1.1 Introduction

Projections of quantities such as electricity and fuel demands, commodity prices, and specific energy consumption and production rates are widely used to inform private and public investment decisions, long-term strategies and policy analysis [19, 20, 21]. Policy analysts and decision-makers often use modeled projections as forecasts[1] with little or no discussion about the associated uncertainty [20, 23, 24]. Here we are concerned with national scale forecasts in the energy industry that span a range from years to decades. Two of the most influential sets of energy projections are those of the U.S. Energy Information Administration (EIA) and the International Energy Agency (IEA), complemented by those made by private oil and gas companies, such as Shell, ExxonMobil and Statoil. When assessed retrospectively, such energy projections have sometimes shown very large deviations from the realized values [25, 26, 27]. Providing information on the likely uncertainty associated with such projections would help individuals and organizations use them in a more informed manner [28, 29, 30].

All of the energy outlooks mentioned above provide point projections without a probabilistic treatment of uncertainty. Often, point forecasts are labeled as a "reference scenario", and are accompanied by alternative scenarios. While scenarios may be used to bound a range of possible outcomes, they can easily be misinterpreted [31] and are typically not intended to reflect any treatment of probability. The fact that most projections in the energy space do not report probability distributions around predicted values, or an expected variance, is a problem that has been frequently noted in the literature [32, 31, 33, 34, 35]. Shlyakhter et al. criticize the EIA for not treating uncertainty in the Annual Energy Outlook (AEO) [32]. Density forecasting is increasingly becoming the standard [34, 36] in a variety of disciplines ranging from forecasts of inflation rates [37, 38, 39], financial risk management and trading operations [40, 41], to demographics [42], peak electricity demand [43] and wind power generation [44, 45]. There are a number of procedures for probabilistic forecasting [40]. Most of these methods take an integrated approach to forecast the whole distribution

[1] Energy outlooks are often referred to as *projections* because they refrain from incorporating future policy changes into the reference scenario. In contrast, the term *forecast* denotes a best estimate allowing for all changes of the state of the world [22]. While we are aware of this difference, our analysis treats the reference scenario as the best estimate forecast. We use the terms forecast and projection interchangeably.

including the best estimate. The empirical methods we use here instead allow analysts or forecast users to attach an uncertainty distribution to a pre-existing point forecast.

The importance of density forecast evaluation has been discussed by several authors [46, 47, 35, 48]. When methods are chosen to generate probabilistic energy forecasts, such evaluation is often omitted. Our work is a step towards making energy density forecasting more feasible and robust by framing how to evaluate a probabilistic forecast in this setting.

1.1.1 Choosing a density forecasting method

We compare different methods by testing how accurately they estimate the uncertainty of data that were not used to train the methods.

We argue that if a forecaster is choosing between different methods, this should be the central criterion, even though others such as usability and ease of explanation might also be relevant. Adopting a frequentist's approach, we view a future observation as a random event around the given forecast. A density prediction is best if it equals the probability density function (PDF) from which this future observation is drawn.

Density forecasts are evaluated by their calibration and their sharpness subject to calibration [47]. By sharpness we mean that narrower PDFs are preferable. Calibration, as a core concept of forecast evaluation, refers to the predictive density representing correctly the true PDF of the observation. Measuring calibration requires the availability of unknown observations. This can be simulated by using an early portion of the time series to train the density prediction and using later actual values as the test observations. This procedure is referred to as out-of-sample forecast evaluation. Dividing the data into these two sets requires a long enough record of historical data and forecasts to draw statistically significant conclusions. While the AEO sample size is small, we see no viable alternative to this procedure, and find that even small sample results can provide useful insights.

As it is a measure of both calibration and sharpness, we use the continuous ranked probability score (CRPS) [48, 49, 50] to compare density forecasts. For point forecast evaluation we work with the average prediction error, here the mean absolute percentage error (MAPE), and the transformed mean absolute logarithmic error (MALE) for prices (Section 1.4: *Materials and Methods*).

1.1.2 Empirical density prediction methods

We compare four different data-driven parametric and non-parametric estimates of forecast uncertainty in the form of PDFs (see Table 1.1 and Section 1.4: *Materials and Methods*). A simple method of empirical prediction intervals (EPI), first published by Williams and Goodman [51], uses the distribution of past forecast errors to create a probability density forecast around an existing point forecast. It relies on the assumption that past errors are a good estimator of the forecaster's current ability to predict the future. EPIs are an established approach and have been employed in a number of fields such as meteorology [52], including the creation of the classic "cone of uncertainty" now routinely produced for likely hurricane tracks [53], future commodity prices [54], and the values of macroeconomic variables such as inflation [38]. There is a continuing interest in the method from researchers in applied mathematics and statistics [36, 55, 56]. We introduce a second non-parametric EPI, which is a modification of Williams and Goodman's EPI, with a centered error distribution. For a third, parametric, prediction method we use the forecasting errors to estimate a Gaussian density forecast. A parametric PDF has the advantage of greater ease of use. We use the volatility of the time series of historical values to inform a fourth probabilistic forecast, which is valuable in cases where the forecasting record is short.

We apply the four different methods to 18 quantities in EIA's AEO [57], which are chosen based on EIA's Retrospective Review [58] (Section 1.4: *Materials and Methods*). The AEO forecasting record spans more than thirty years. Unfortunately, in the context of forecast evaluation a sample size of ~ 30 data points is very small. In addition, because of modifications that EIA makes to its models, and changes in technology, market conditions, and regulations, errors are not likely to be stationary. Because stationarity of past forecasting errors is an essential requirement for good performance of EPIs [56], we test the extent to which PDFs estimated using this procedure provide robust probabilistic forecasts. Previous work has analyzed the forecast errors of EIA's AEO [25, 20, 59, 19, 21, 60] and the projections by the IEA [26]. Generally, authors have focused on a mean percent error and directional consistency of errors, also termed bias. Shlyakhter et al. [32] constructed a parametric density forecast with the retrospective errors of AEOs, similar to what we test in this paper. However, they did not assess the calibration of their prediction intervals.

Table 1.1: Empirical density forecasting methods compared. Details can be found in Section 1.4: *Materials and Methods.*

Method	Parametric	Based on	Median ctrd.
NP1: non-parametric EPI	no	forecast errors	no
NP2: n.-p. centered EPI	no	forecast errors	yes
G1: Gaussian distr.	yes	forecast errors	yes
G2: Gaussian distr.	yes	hist. deviations	yes

We begin by evaluating the point forecast performance of the AEO reference case over our test range of AEO 2003-2014. Using the same out-of-sample AEOs and historical observations, we then compare the calibration and sharpness of the four different density forecasts. The prediction intervals are also compared to the scenarios published in the AEO. We find that over the test range a normal distribution based on past forecasting errors clearly outperformed uncertainties based on the scenarios in the AEO. This conclusion is for the diverse set of all quantities, but depending upon the quantity, in some cases other methods showed better results. We conclude the paper with a comparative discussion of the methods and their applicability to energy forecasting.

1.2 Results

We evaluate the predictive performance of four uncertainty estimation methods (Table 1.1) over the test range of AEO 2003-2014 and observations of 2002-2015, using 1985-2002 as the training range. The test range excludes AEO 2009, which did not provide scenarios for the updated reference case. We determine the number of quantities for which a method performed best. We find that Gaussian densities informed by retrospective errors (G_1) or based on the variability of the historical values (G_2) performed best for the most quantities. The original non-parametric method as in [51] (NP_1), performed best in very few cases. The centered non-parametric distribution (NP_2), which gives the largest weight to the AEO reference case projection instead of the bias, performed better over the test range than NP_1. The respective best empirical uncertainty estimation methods had significantly better calibration than methods based on the AEO scenarios with 95% confidence. In fact, G_1 significantly outperformed the scenarios for all quantities and provided a valid general approach to estimate the uncertainty in the AEO.

While we have performed analysis for 18 quantities forecasted in the AEO, we use two of the quantities, natural gas wellhead price in nominal dollars per 1000 cubic ft. (hereafter natural gas price) and total electricity sales in billion kWhrs (hereafter electricity sales) for illustration purposes (Fig. 1.1 and 1.2). Results for all 18 quantities can be found in the *Appendix* 5.

1.2.1 Error metric and transformation for price quantities

All forecast evaluation scores are computed on the basis of the deviations of the forecasts $\hat{y}$ from historical values y, referred to as error. We found it useful to work with the percent error, or relative error, $\epsilon_{rel} = \frac{\hat{y}-y}{y} = \frac{\hat{y}}{y} - 1$. Percent errors allow us to compare different quantities and they are independent of changes in the currency value. We can conduct the analysis in a similar way with absolute errors. Since the error distributions of price quantities are asymmetric, as prices are typically log-normally distributed [61], we modify the error for price quantities. Drawing an analogy to logarithmic returns, a concept from financial theory, we modify ϵ_{rel} to yield the logarithmic error $\epsilon_{log} = \ln(1+\epsilon_{rel}) = \ln\left(\frac{\hat{y}}{y}\right) = \ln\hat{y} - \ln y$. For prices we compute the comparative statistics and additional transformations, such as centering of the PDF, in ϵ_{log} (*Appendix* 5).

The structure of the relative errors as a function of forecast year and forecast horizon is shown in Fig. 1.3. The horizon H refers to the number of time steps, or years, into the future that the forecast is made. Uncertainty increases with H. AEO projections reflect uncertainty in past values, e.g. for AEO 2016 we therefore refer to 2015 as $H = 0$, and 2016 as $H = 1$.

1.2.2 Retrospective analysis can inform density forecasts

We illustrate examples of the four probabilistic forecasting methods listed in Table 1.1. Fig. 1.1 and Fig. 1.2 compare the non-parametric methods to the methods that performed better for the two example quantities, that is, the two Gaussian predictions.

A non-parametric distribution of the errors (NP_1) results in the EPI shown in Fig. 1.1 (A). Here the median of the errors is not exactly zero, which is often referred to as bias. We see that this results in a second point forecast, or a best estimate forecast that is not equal to the reference case

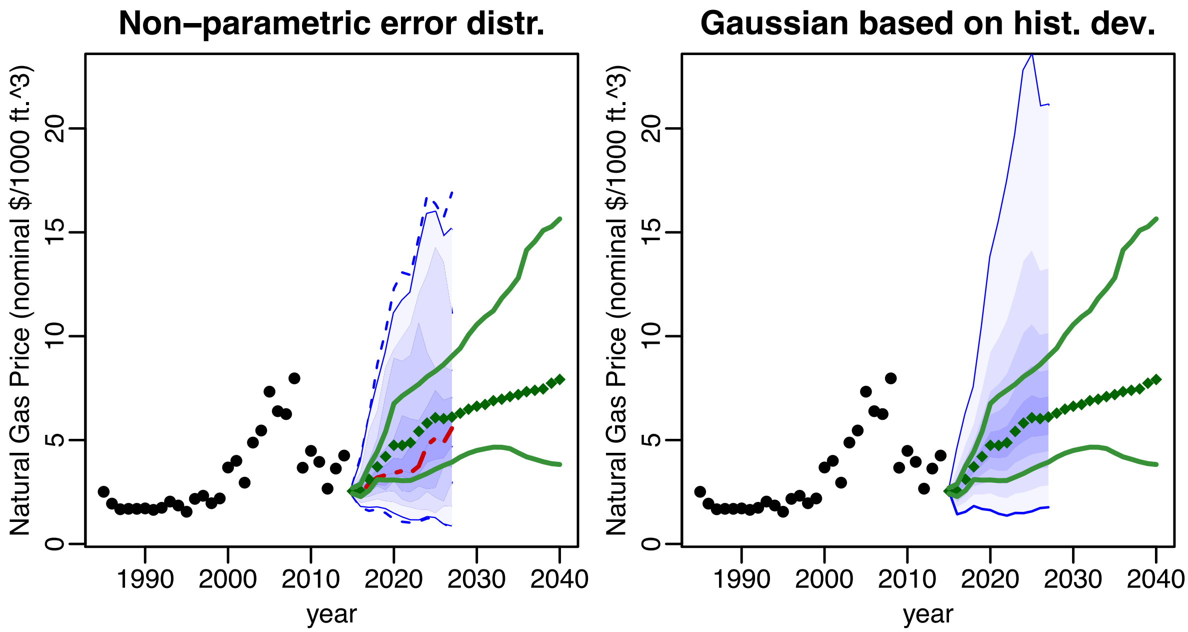


Figure 1.1: Density forecasts for natural gas prices in nominal \$. (A) Non-parametric EPI based on forecast errors (NP_1). (B) Gaussian density forecast based on the variability of historical values (G_2), which tested to be the better estimate. Historical values are indicated by black dots, the AEO 2016 reference case by green diamonds and the density forecast in blue shaded areas. The different shades correspond to the percentiles 2, 10, 20, 30, ..., 80, 90, 98. The outermost dashed lines report the minimum and maximum value of the error samples. AEO 2016 envelope scenarios are in green. Note that in (A) the median of the predictive distribution (dashed red line) does not coincide with the reference case.

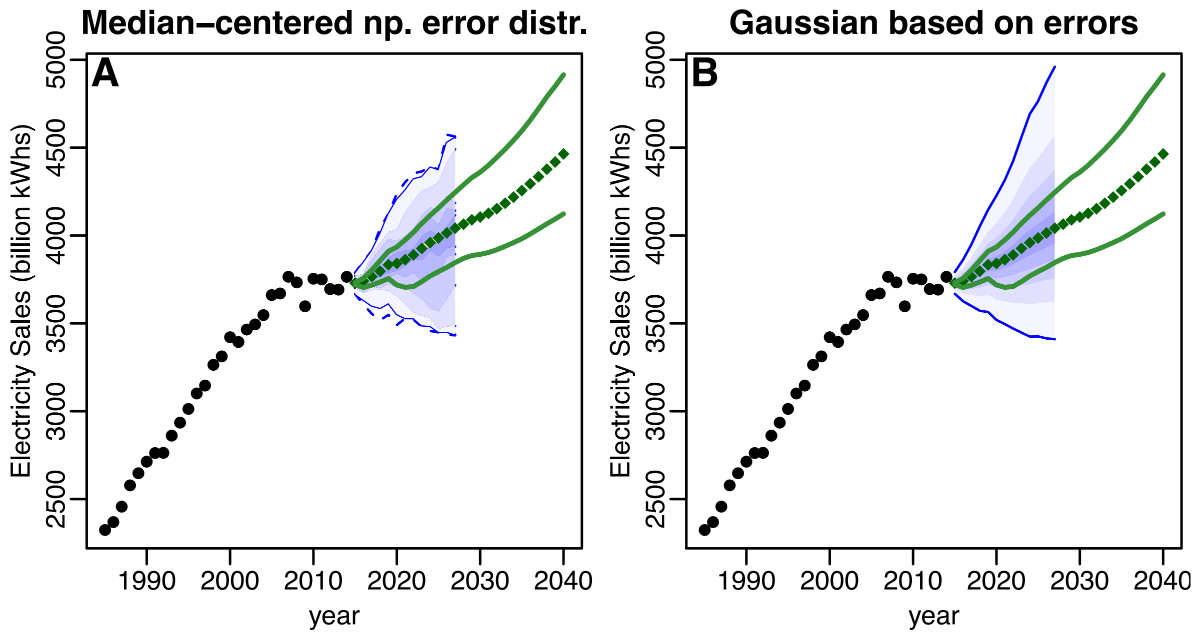


Figure 1.2: Density forecasts for electricity sales based on AEO 2016. (A) For the median-centered non-parametric EPI (NP_2), the median or bias now coincides with the AEO reference case. (B) The Gaussian density forecast based on the SD of the errors (G_1) was the best forecast over the test range. The envelope scenarios are narrower.

scenario. If we can assume that the forecasting errors are stationary, then past and future errors follow the same PDF, and this bias should yield a better point forecast than the reference case. However, we found this is not the case for most quantities.

Modifying the non-parametric distribution in such way that it places the greatest weight on the AEO reference case projection is one approach to combat this problem (NP_2). This centered EPI for electricity sales is shown in Fig. 1.2 (A). In the percent-error space, we center by subtracting the median error m_{rel} from all errors in the distribution $\epsilon_{rel,ctr} = \epsilon_{rel} - m_{rel}$. For the price quantities, we transform the distribution in log-error space. We define the log median $m_{log} = \text{median}(\epsilon_{log}) = \ln(1 + m_{rel})$. The centered log errors are then $\epsilon_{log,ctr} = \epsilon_{log} - m_{log} = \ln\left(\frac{1+\epsilon_{rel}}{1+m_{rel}}\right)$ (*Appendix* 5).

These two non-parametric estimations are compared to two parametric distributions, Gaussians with a mean of zero and the variance of the errors (G_1) (Fig. 1.2 (B)) and with the variance of historical values (G_2) (Fig. 1.1 (B)). When modeling normality, we implicitly make assumptions about the nature of the errors. Extreme errors, which can have large consequences for decision-

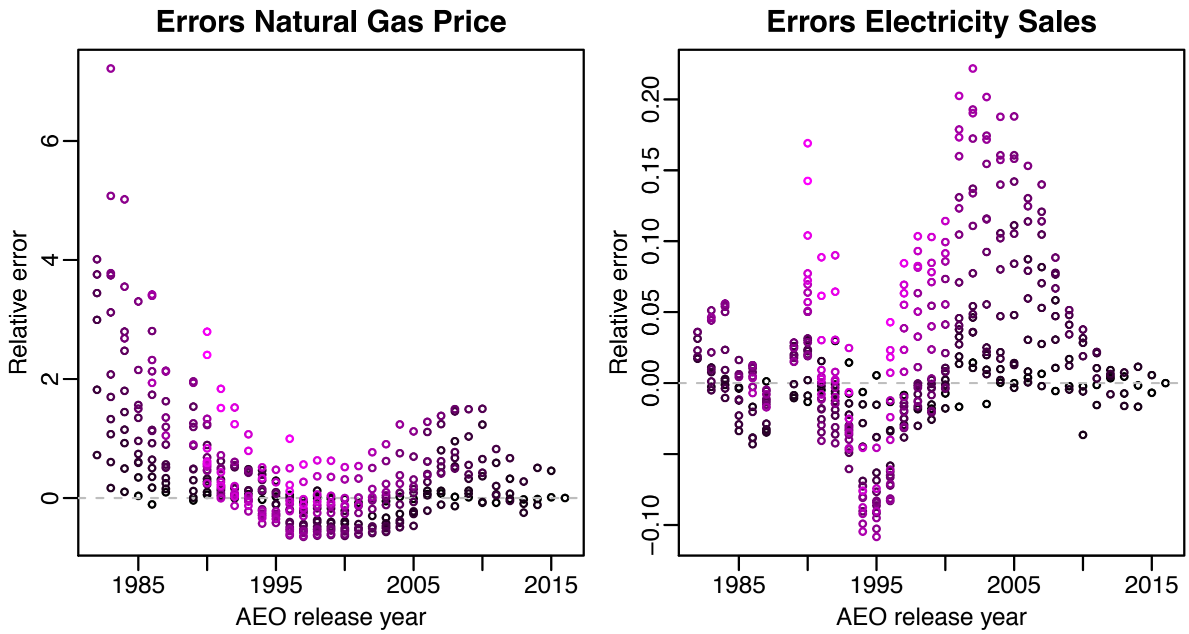


Figure 1.3: Forecast errors by AEO release year. Different colors correspond to forecast horizons ranging from $H = 0$ in black to $H = 21$ in purple. All forecast errors are untransformed. Note the different scale. No AEO was released for 1988.

making, occur frequently in energy forecasting [32]. A Gaussian PDF may not do an adequate job of representing heavier tails and might underestimate the probability of extreme events. However, a parametric distribution will generate longer tails than a non-parametric error PDF. Regarding usability, the simplicity of a two-parameter specification prevails over non-parametric distributions. A discussion of normality and correlation in the errors is provided in the *Appendix* 5.

1.2.3 Past bias in the AEO does not predict future bias

Recently, electricity sales have been flat. Can a forecast be better than a constant prediction using the last observation, i.e. persistence? We can assess the point forecasting skill of the AEO reference case projections by comparing them with benchmark forecasts such as persistence or simple linear regression. To compare different point forecasts, we evaluate the mean absolute percentage error (MAPE) and the mean absolute log error (MALE) for prices. MAPE and MALE are defined as the sum over the absolute value of all observed errors for a given horizon (Section 1.4: *Materials and Methods*). A larger MAPE/MALE indicates that the forecast has performed worse over the test

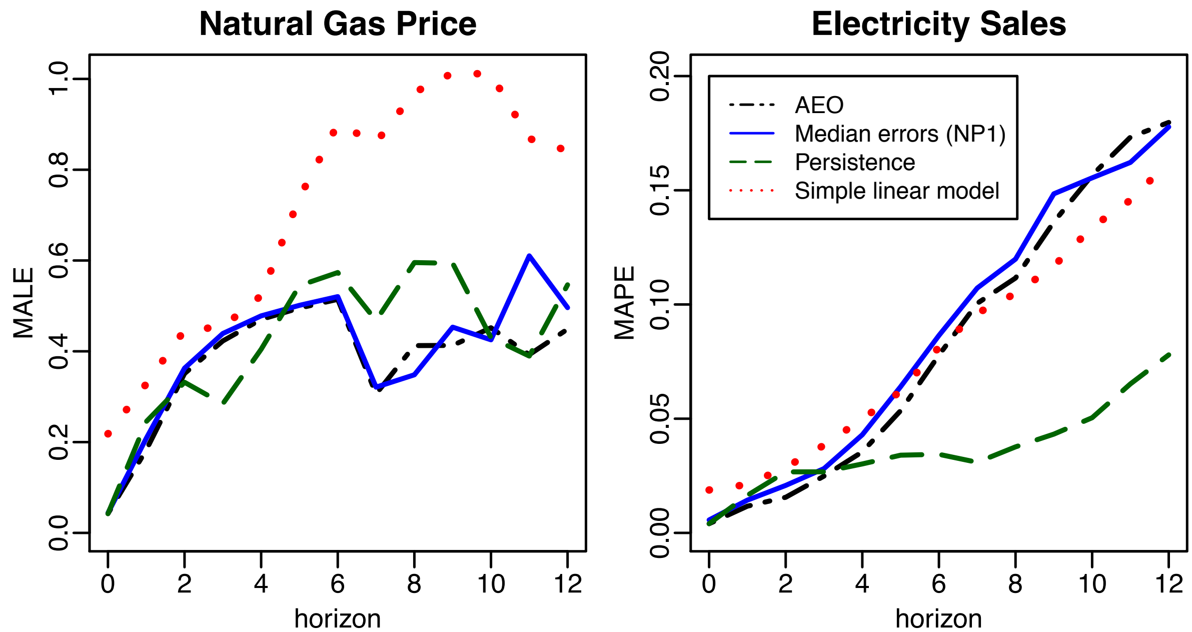


Figure 1.4: The mean absolute percentage or log-error (MAPE/MALE) for the test range 2003-2014. We see that for natural gas prices (in nominal \$), the median of NP_1 performs similarly to the AEO reference case. For electricity sales, the reference case outperforms the median for nearly every horizon. For the test range, a persistence forecasts has clearly been the best forecast for electricity sales, which have recently experienced near zero growth.

range 2003-14 (Fig. 1.4).

We find that persistence performed surprisingly well over the test range of the last decade, outperforming the AEO for 10 of the 18 quantities. This is due to the fact that the recent decade has seen trend changes that are conducive to persistence forecasts. If the length of the fitted window is optimized for the test range, a simple linear regression significantly outperformed the reference case for eight quantities with 95% confidence. Point forecast comparison of the AEO reference case with the median of the errors reveals that correcting for the bias is not a good strategy in most cases. The AEO reference case was a better point forecast than the bias for most of the quantities over the test range, except for coal production and residential energy consumption. We therefore anticipate that centering the non-parametric uncertainty (NP_2) is advised for most quantities except those.

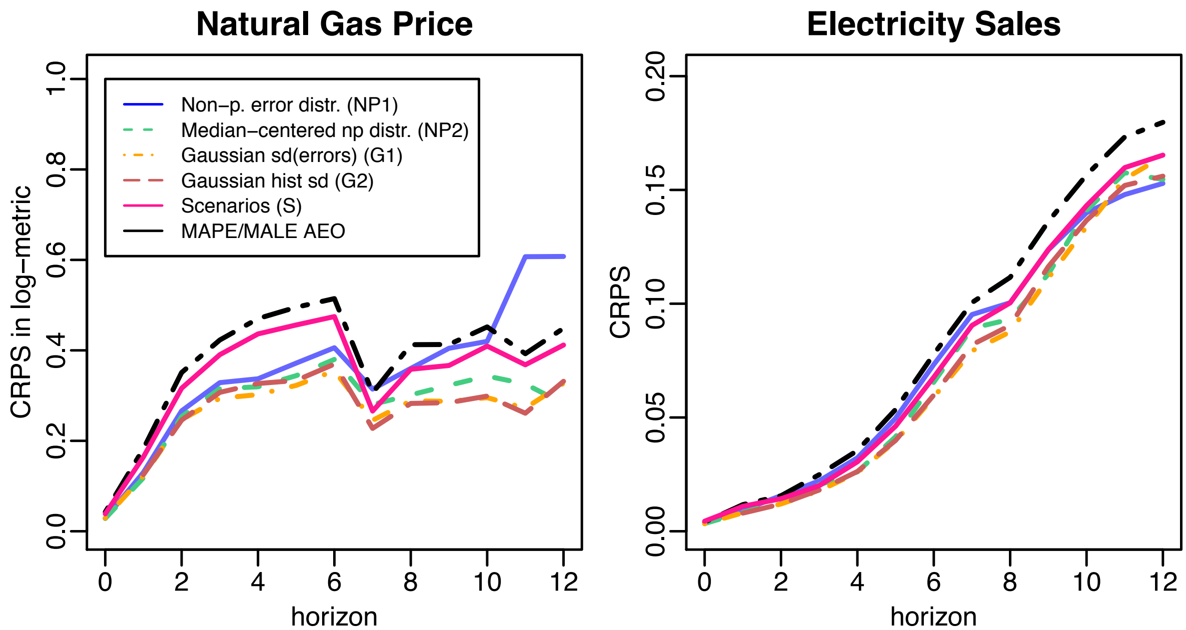


Figure 1.5: The continuous ranked probability score (CRPS) for the test range 2003-2014. A lower CRPS corresponds to a better density or ensemble forecast.

1.2.4 Gaussian density forecasts often perform well

Scoring rules, or scores, provide a means for comparing the performance of different probabilistic forecasts. We use the continuous ranked probability score (CRPS), which is a strictly proper score in this case [49]. It assigns value not only to the predicted probability of an observation but also to the distance of a predicted probability mass from an observation. It is therefore relatively robust to specific functional forms of the density forecasts [48], and allows for comparison with point and ensemble forecasts [49, 50] (Section 1.4: *Materials and Methods*).

The results of the average CRPS over the test range for each horizon in units of relative or log error are illustrated in Fig. 1.5. A standalone value of the CRPS is not meaningful; it serves to provide a comparison between different methods. As the CRPS reduces to the MAPE/MALE for a point forecast, it is informative to compare the results to the MAPE/MALE of the AEO reference case. In Fig. 1.4 and 1.5, we find that the scenarios (S) only marginally improve the prediction with respect to the point forecast. In addition we see that for the natural gas price, NP_1 is larger than the MALE due to poor point forecast performance of the EPI's median.

To find the best density prediction method, we normalize the CRPS of each method by the CRPS of the scenario ensemble (S) for every horizon (Fig. 1.6). For every quantity, we then average over a core range of horizons $H = 2$ to $H = 9$, and rank these aggregated scores. The method with the lowest average rank is considered the best density over the test range for a given quantity. We find that the results barely change if more horizons, modifications to the test range or an alternative ranking method are considered (*Appendix* 5).

The ranking of all quantities shows that the two Gaussian methods perform well for most quantities (Fig. 1.7). G_1 counts as the best method for nine out of the eighteen quantities and G_2 for three quantities. The performance of G_2 is however often similar to G_1 and it is second best for eight quantities. The fact that these parametric methods performed well over the test range is convenient, because there are standard ways to use a normal distribution as a model input. Besides these parametric methods, also NP_2 performed well. As expected, in the two cases of coal production and residential energy consumption, including the bias with NP_1 seemed the best approach over the test range. In the following section, we analyze if the empirical methods performed significantly better than uncertainty estimates based on the scenarios.

1.2.5 AEO scenario ranges are narrower than observed uncertainties

Every AEO includes a number of scenarios, intended as sensitivity studies on the reference case under a small number of varied input assumptions. No value is assigned to the probability that a future outcome will lie within the scenario range. The CRPS allows for comparison of a density forecast with an ensemble forecasts. It assigns every discrete scenario an equal point probability mass (S). Because of the varying number of scenarios in the AEO, we make a simplification and only consider the reference case and the high and low envelope scenarios, which do not correspond to a specific scenario in the AEO (Section 1.4: *Materials and Methods*). In addition, we discuss a Gaussian distribution (SP_1) and a uniform distribution (SP_2) based on the envelope scenarios.

The CRPS scores normalized by the score of S are shown in Fig. 1.6. This figure also includes the scores for the sensitivity cases SP_1 and SP_2. A normalized CRPS of an empirical method that is < 1.0 indicates an improvement over uncertainties based on the scenarios (S). We can find at least

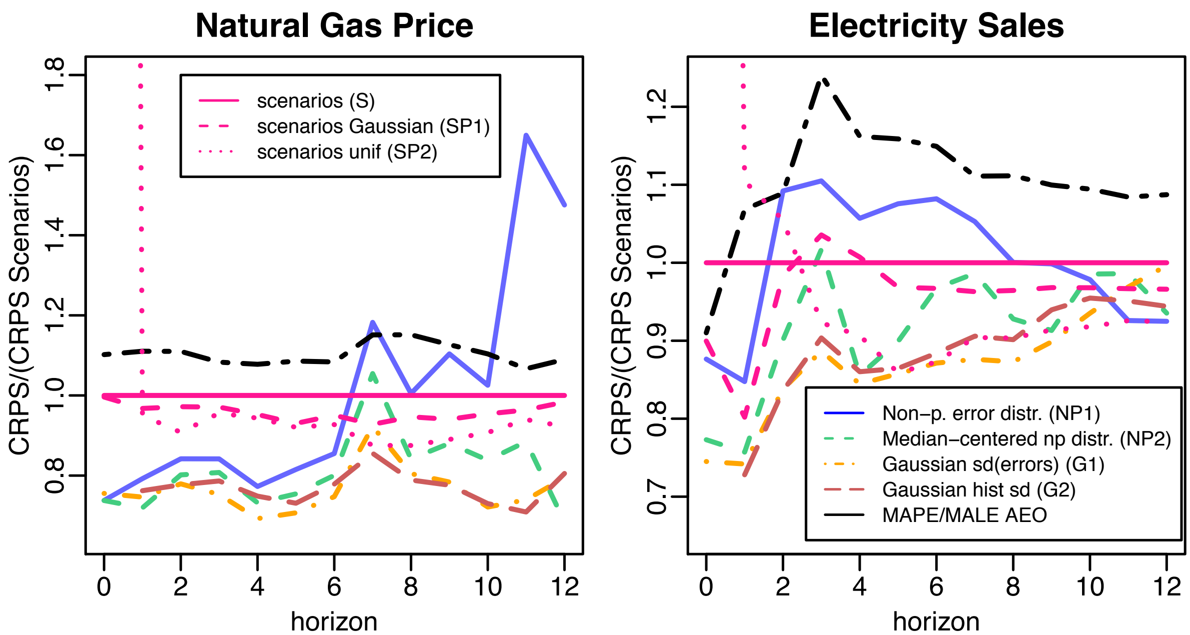


Figure 1.6: Relative improvement of the methods with respect to the envelope scenarios for the test range 2003-2014. Values are plotted as fraction of the CPRS of the scenario ensemble (S). A normalized CRPS lower than 1.0 corresponds to a better density forecast. SP_1 corresponds to a normal distribution with the scenario range as 1 SD, and SP_2 is a uniform PDF between the envelope scenarios.

one density forecasting method for every quantity, which in average over the core horizons performed better than the scenarios. In addition, we conduct a hypothesis test if we can reject that either S or SP_1 were the better probabilistic forecasts over the test range. We find that the best ranked empirical method for a respective quantity was significantly better than both S and SP_1 with 95% confidence. In fact, NP_2, G_1 and G_2 all show significant improvements (Fig. 1.7). These results are likely due to the fact that over the test range on average the scenario range of all AEO quantities covered only 14% of the actual values (*Appendix* 5). The width between highest and lowest scenario, however, changes greatly from one AEO to another and is somewhat correlated to the number of scenarios published.

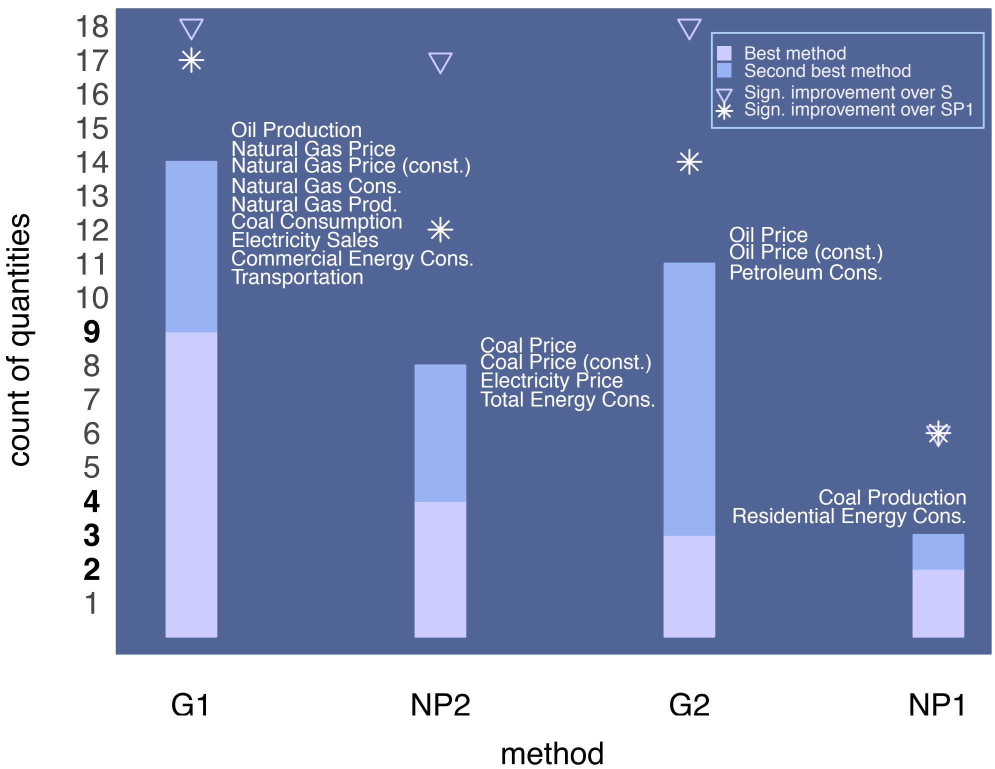


Figure 1.7: Graphical summary of the evaluation results. The methods are ordered by the number of quantities they perform best for (listed in white). The Gaussian based on errors (G_1) performs best or second best for 14 out of 18 and showed significant improvement over the scenarios for almost all quantities. Improvement is more likely over S than SP_1. The non-parametric biased EPI (NP_1), performs worse than the n.-p. centered EPI (NP_2), and the Gaussian based on historical deviations (G_2).

1.3 Discussion and conclusion

This analysis showed that empirical density prediction methods, based on forecasting errors or historical deviations, provide valuable approaches for including an estimate of uncertainty with a forecast. There are empirical methods available for estimating the uncertainty around the AEO reference case, which have proven to be significantly more accurate over the past decade than the scenarios of the AEO. We find that a Gaussian distribution based on past errors (G_1) offers a method with convincing ease of use and good performance over the different quantities (Fig. 1.7). We therefore recommend that the EIA and others producing energy forecasts include the standard deviation of forecast errors in their retrospective reports. We supply the values for the AEO 2016 in the *Appendix* 5. A non-parametric distribution of the observed forecast errors was the better density forecast only in a few cases, confirming that focusing on representing the exact error distribution does not need to provide the better out-of-sample forecast. Point forecast evaluation illuminated that EIA's forecast bias is in most cases not consistent and that using a bias-corrected reference case does typically not lead to the better forecast.

As both the forecasting process and the energy system can be non-stationary, there is no way to be sure that our results will be applicable to future data. However, the way we evaluated and chose a method is a robust procedure. Hence, in the absence of other insights we recommend using one of the Gaussian distributions.

Despite the advantages of probabilistic forecasts, scenarios convey important information about the workings of energy predictions and allow users to better understand and compare the assumptions. We emphasize that the combined use of a density forecast and scenarios would be a fruitful approach to describe the uncertainty of a forecast. Empirical density forecasts are easily reproducible, but other probabilistic methods such as a quantile forecasting could also advance energy projections.

1.4 Materials and methods

See *Appendix* 5 for a detailed description of the materials and methods used.

Data

The data set consists of AEOs 1982-2016 and historical values from 1985 to 2015. Historical data were taken from the EIA Retrospective Review [58] and the AEOs [57], and conversions were applied where necessary. All data are publicly available on the EIA website. Refer to *Appendix 5: Data Description* for more detail. The data analysis was performed in R [62].

1.4.1 List of methods

Point forecasting methods:

- *AEO reference case*: We treat the AEO reference case as a point forecast. The reference case is a projection of the current state of laws and regulations and does not represent a best estimate forecast. Also the EIA chooses the reference case as a best estimate when determining projection errors [58].
- *Median errors (NP_1)*: The median of the EPI with a non-parametric distribution of the errors (NP_1), computed as the reference case adjusted by the median of past forecasting errors.
- *Persistence*: Persistence refers to a constant forecast equal to the last observation. Here, we use the forecasted value at $H = 0$ as the last observation, since on the AEO release date this is the closest approximation to the actual value.
- *Simple linear model:* This benchmark is a simple linear regression with time as the predictor. The quantity is regressed over a moving window of the last 7 historical observations. This size of window is the optimum for the test range.

Density forecasting methods:

- NP_1: EPI with a non-parametric distribution of the forecasting errors and a median different to the reference case. This method was originally published by [51].
- NP_2: EPI with a non-parametric error distribution, which is centered such that the median and $\epsilon = 0$ align. This results in the AEO reference case being the best estimate forecast.

- G_1: A Gaussian distribution with the standard deviation of the past errors and a mean and median of $\epsilon = 0$.
- G_2: Gaussian distribution with a standard deviation based on a sample of all relative deviations between two historical data points which are H steps apart. Mean and median are $\epsilon = 0$.
- S: This ensemble forecast consists of the reference case and the highest and lowest scenario projection in every year. These correspond to the envelope of all scenarios by using only the highest and lowest projected values.
- SP: Two parametric density predictions based on the envelope scenarios in the AEO. We chose a Gaussian distribution with the distance to the farthest scenario as 1 SD (SP_1) and a uniform distribution between the envelope scenarios (SP_2).

1.4.2 MAPE

The mean absolute percentage error (MAPE) is a measure for point forecast performance. This becomes the mean absolute log error (MALE) in the case of price forecasts with log-errors. They are defined as

$$MAPE_H = \frac{1}{n_H} \sum_{t=1}^{n_H} |\xi_{rel,H,t}| = \frac{1}{n_H} \sum_{t=1}^{n_H} \left| \frac{\hat{y}_{H,t} - y_{H,t}}{y_{H,t}} \right|, \tag{1.1}$$

and $MALE_H = \frac{1}{n_H} \sum_{t=1}^{n_H} |\ln \hat{y}_{H,t} - \ln y_{H,t}|$, where there are n_H errors for a particular horizon H. $\hat{y}$ refers to the forecast, while y is the actual observation.

1.4.3 CRPS

The continuous ranked probability score (CPRS) for every horizon, as we use it in this paper, is defined as

$$CRPS_H(F, \epsilon) = \frac{1}{n_H} \sum_{t=1}^{n_H} \int_{-\infty}^{\infty} \left(F_t(\epsilon_t) - I(\epsilon_t \geq \xi_t)\right)^2 d\epsilon_t \tag{1.2}$$

similar to [49]. ϵ_t is a point of the predictive error distribution, while ξ_t is the forecast error of the observation. The CRPS compares the CDF of the density forecast with the CDF of an observation, a step function $I(\epsilon_t \geq \xi_t)$. We compute the score in the respective error metric. The CRPS for a non-parametric CDF is computed like the CRPS for an ensemble forecast of discrete scenarios [50]. For ensemble forecasts, the CRPS can also be written as $CRPS_H(F, \epsilon) = \frac{1}{n_H} \sum_{t=1}^{n_H} \left[E_F \left| \epsilon_t - \xi_t \right| - \frac{1}{2} E_F \left| \epsilon_t - \epsilon_t' \right| \right]$ [49]. In our case, the $CRPS_H$ reduces to the $MAPE_H$ for a point forecast. In this case we have a single $\epsilon_t = 0$, resulting in $E_F \left| \epsilon_t - \xi_t \right| = \left| \xi_t \right|$ and $E_F \left| \epsilon_t - \epsilon_t' \right| = 0$. The CRPS is a strictly proper score here [49], which means that the expected score is maximized if the observation is drawn from the predictive distribution and this maximum is unique. The CRPS has different scales for different quantities or error measures, which is why we normalize the CRPS$_H$ by the CRPS$_{S,H}$ of the scenario ensemble.

1.4.4 Improvement testing

We perform a bootstrap on the single CRPS results in a horizon sample, which then are used to compute the CRPS$_H$, and the aggregated CRPS average for the ranking. For every of the four methods, we determine the portion of resampled results that indicates that S or SP$_1$ is the better forecast. If this portion is smaller than 0.05, we speak of the method as being a significant improvement over the scenarios.

1.4.5 Sensitivity analysis on the ranking results

To test the sensitivity of the ranking, we varied the default assumptions. Instead of first averaging the normalized CRPS and then ranking that result, we alternatively first ranked the CRPS$_H$ and then averaged over the horizons. We also averaged over the full range of horizons $H = 1$ to $H = 12$ instead of the core range that included large H with small sample sizes. In addition, we included AEO 2009 in the test range. The respective best methods did not change with these variations. For some quantities, the performance of the best and second best methods were very similar to each other. This resulted in a sensitivity regarding a change in the test range for three quantities.

2

Decarbonizing intraregional freight systems with a focus on modal shift

Road freight transportation accounts for around 7% of total world energy-related carbon dioxide emissions. With the appropriate incentives, energy savings and emissions reductions can be achieved by shifting freight to rail or water modes, both of which are far more efficient than road. We briefly introduce five general strategies for decarbonizing freight transportation, and then focus on the literature and data relevant to estimating the global decarbonization potential through modal shift. We compare freight activity (in tonne-km) by mode for every country where data are available. We also describe major intraregional freight corridors, their modal structure, and their infrastructure needs. We find that the current world road and rail modal split is around 60 : 40. Most countries are experiencing strong growth in road freight and a shift from rail to road. Rail intermodal transportation holds great potential for replacing carbon-intense and fast-growing road freight, but it is essential to have a targeted design of freight systems, particularly in developing countries. Modal shift can be promoted by policies targeting infrastructure investments and internalizing external costs of road freight, but we find that not many countries have such policies in place. We identify research needs for decarbonizing the freight transportation sector both through improvements in the efficiency of individual modes and through new physical and institutional infrastructure that can support modal shift.

This chapter has been published as: Kaack, L. H., Vaishnav, P., Morgan, M. G., Azevedo, I. L., & Rai, S. (2018). Decarbonizing intraregional freight systems with a focus on modal shift. Environmental Research Letters, 13(8), 083001. `https://doi.org/10.1088/1748-9326/aad56c`

2.1 Introduction

To restrict the increase in global temperatures to less than 2°C above preindustrial levels, rapid decarbonization of the global economy is necessary [5, 6]. The share of global greenhouse gas (GHG) emissions from transportation is rising, with oil demand in the transportation sector accounting for about 22% of global energy-related carbon dioxide (CO_2) emissions in 2015 [10]. The International Energy Agency (IEA) estimates that road freight alone contributed 7% of total world energy-related CO_2 emissions in 2015 [10], more than twice the total emissions from aviation [63]. There have been few systematic studies of deep decarbonization options for global and intraregional freight transport [64] and little political effort to mitigate GHG emissions from freight transportation [65]. For example, only four countries, the United States (U.S.), Canada, China and Japan, regulate the fuel economy of heavy-duty vehicles. The European Union (EU) is currently developing such standards [10]. On the other hand, many countries have standards for particulate matter (PM) emissions, and these standards provide an incentive to improve fuel economy. The Nationally Appropriate Mitigation Actions (NAMAs), submitted to the United Nations Framework Convention on Climate Change (UNFCCC) by developing countries, also lack proposals that target the transportation sector [66]. While some modal shift measures in transportation have been included in Nationally Determined Contributions, proposals are heavily skewed towards passenger transportation [67].

In this paper, we focus specifically on road freight, which has a large and growing share of freight activity in most countries [10, 68], and we explore the potential for modal shift as a decarbonization strategy. We emphasize that the deep decarbonization of the freight sector can only be achieved by combining modal shift with multiple other strategies such as energy efficiency, switching to fuels with low or net zero carbon emissions, and improving operational efficiency. Many such changes are unlikely to occur without top-down or other policies to reduce emissions. We also discuss potential interactions and adverse effects of combining these strategies. Across the developing world, and in parts of the developed world, a key barrier to identifying opportunities for the adoption and deployment of modal shift and intermodal transport is the lack of sufficiently granular high-quality freight data. Governments collect data on country-specific freight activities, and international organizations such as the International Energy Agency (IEA), OECD [69] and the World Bank [70] gather national statistics, but data collection is sparse and inconsistent, with some important efforts from the IEA [10]. For this review, we created a database of freight activity between 2000 and 2016 for every country where data are available.

To our knowledge, this is the first comprehensive review of the modal shares of intraregional freight activities and the potential for modal shift in transport systems globally. We are aware of similar reviews with a regional focus, such as by the U.S. Department of Energy (DOE) [13] and by Woodburn et al. (2007) [11] and Hoen et al. (3013) for Europe [71]. A 2001 report by the OECD took a global perspective on intermodal freight [72] and a United Nations report from 2015 included a chapter on modal shift [14].

We begin by introducing five general strategies to decarbonize global freight transport. After an overview of trends in freight transportation and the global modal structure, we provide an in-depth review of studies of the theoretical aspects of modal shift. We identify which factors influence shippers' mode choices, and we discuss the greatest barriers to increasing the share of rail, water and intermodal freight transportation. We conclude that discussion with an overview of strategies to promote modal shift. We then describe freight transportation systems and modal shift policies in specific regions of the world, including major intraregional freight corridors, their modal share, and their infrastructure needs. We conclude the review by summarizing existing estimates of the potential for modal shift.

We exclude ocean shipping, because fuel use and emissions from ocean shipping have been well studied and analyzed within the limitations of the available data [73]. We also only briefly discuss air transport of freight because it represents a small proportion of total emissions from transport, and offers limited opportunities for modal shift.

2.2 General strategies to decarbonize freight

Here we introduce five broad strategies to decarbonize freight, which we describe below, using the same taxonomy as McKinnon (2016) [74]. These are: 1) reducing the demand for freight; 2) optimizing vehicle use and loading; 3) increasing the efficiency of freight vehicles; 4) reducing the carbon content of fuel used to transport freight; and 5) shifting freight to low carbon-intensity modes. Below we discuss each of these strategies.

2.2.1 Strategy 1: Reducing the demand for freight transport

There is a strong relationship between the growth of freight activity and economic growth [68, 75, 76]. To the extent that this relationship is causal, this makes it difficult to lower the total demand for freight transportation. Inland freight transport is moreover of enormous importance to economic development of low-income countries [14, 77, 78]. While in the past, freight demand grew much faster than the gross domestic product (GDP), the elasticity of trade and GDP is likely to be closer to one in the long term [74]. While there is some evidence of moderate decoupling of total freight activity and GDP in a few developed countries [68, 79], there is no decoupling of road freight activity and GDP [68, 10]. Furthermore, the demand for freight transportation is relatively inelastic to fuel prices [80, 81], suggesting that modest price signals may not induce much reduction in total freight activity. For example, a meta-analysis by de Jong et al (2010) [82] concluded that a 1% rise in fuel price reduces fuel demand by 0.2-0.6%, vehicle kilometers by 0.1-0.3%, and transport activity by 0.05-0.3%.

Another strategy to reduce demand for road freight would be structural changes in supply chain management [74]. Such strategies would entail changes in the production location, in the origin-destination pattern in a supply chain, a reduction of demand for the goods being consumed, and novel production technologies [82]. For example, additive manufacturing (AM), which creates parts by depositing material layer by layer [83], could produce light, hollow products with intricate internal structures to give them strength. As such, it can reduce transport work by making both products [84] and vehicles lighter [85]. Some have argued that it can reduce the distances over which materials and products are transported, by stimulating localized production [86], but it faces issues with economies of scale, fabrication speed, and quality control [87]. Some estimates have suggested that AM could eventually account for less than 5% of total manufacturing output by value [88, 89] and reduce carbon dioxide emissions by less than 1% [90].

Many recent trends in supply chain management, such as just-in-time production, centralization of inventory, spatially fragmented production, and global trade and procurement, have led to an increase, rather than a decrease, in transportation demand [74, 77, 12].

2.2.2 Strategy 2: Optimizing vehicle use and loading

Efficiency gains can be achieved by making better use of modal capacity. This includes efficient routing and supply-chain collaboration, backhauling (less empty runs), more efficient packing, and raising the legal limits for high-capacity vehicles in the trucking sector [74, 10]. Higher efficiency might also be achieved by lowering speeds, as has been the case with ocean shipping [73], and training drivers to reduce idling and drive more efficiently. The IEA has provided a detailed review of these systemic improvements in the road freight sector and their potential for reducing GHG emissions [10].

2.2.3 Strategy 3: Increasing the efficiency of freight vehicles

In Fig. 2.1, we summarize reported estimates of the average carbon intensity (in $gCO_{2,eq}$ per tkm) for different regions and modes. Naturally, these estimates have uncertainty associated with them and depend on the assumptions made. For example, for road vehicles, these estimates will depend on age, fuel type, payload, terrain, driving patterns and other factors. We find that overall rail and inland water freight modes are much less carbon intensive than road for all regions, and that inland water transport is more carbon intensive than rail in most regions. Shipping on inland waterways is typically also more carbon intensive than ocean shipping [64]. For carbon intensities of different types of vehicles without a regional specification we refer the reader to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) (Table 8.3 in 3) [64]. In the next sections we provide more detail regarding the carbon intensity and the technological emissions reduction potentials associated with each mode.

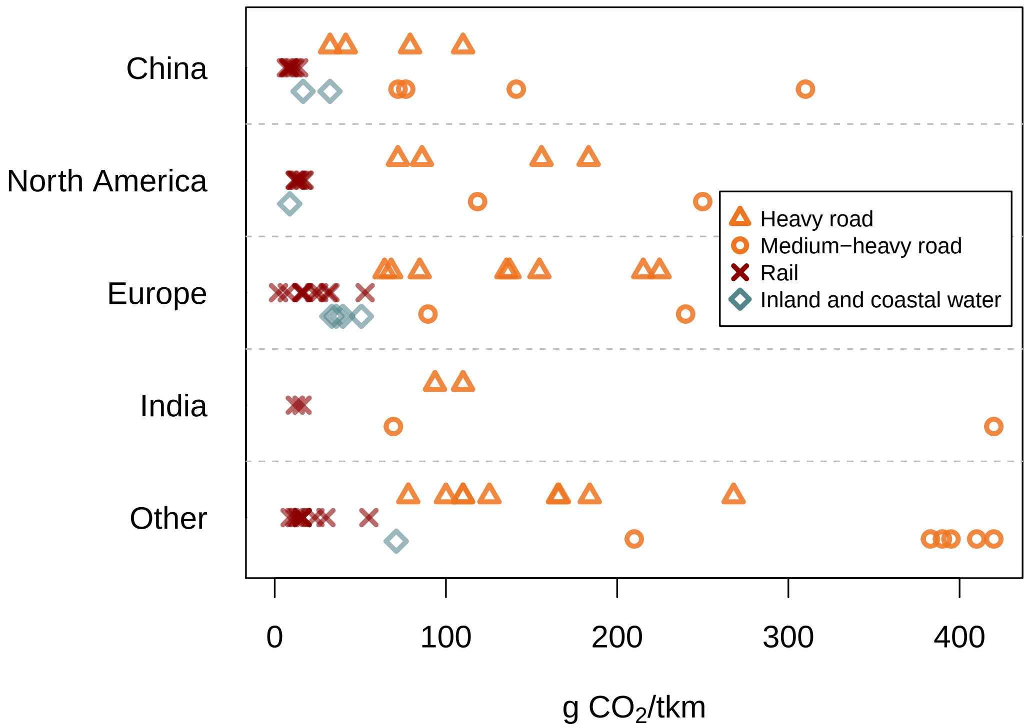


Figure 2.1: Carbon intensity of surface freight modes in different regions. Each marker represents an average estimate for the region or a country in that region from a different data source (refer to supplementary materials for values and sources). Road carbon intensity values distinguish medium- and heavy-duty vehicles. We do not differentiate sources that report life-cycle emissions from those reporting tailpipe emissions as this difference is well within the uncertainty.

Road freight

Despite the uncertainty across the studies in Fig. 2.1, we find that the range of road freight energy intensities varies between the regions. This has to do with economic, geographic and infrastructure constraints that determine which vehicles are in use. For a detailed review on heavy-duty vehicle technology and utilization we refer the reader to [10, 91, 92].

The U.S. National Academies of Science, Engineering and Medicine (NASEM) [91, 93] has performed several assessments regarding the medium and heavy-duty sector in the U.S. In its 2010 study, the Academies suggest that a series of vehicle technologies - ranging from improvements in aerodynamics to the use of low-rolling resistance tires - could reduce fuel consumption for U.S. trucks by up to 15%. Improvements to diesel engines could reduce fuel consumption by up to 20%, and a shift to hybrid drive trains could halve it. The actual savings and the cost of achieving them would depend on the size and function of the vehicle. A recent report by the IEA [10] concludes that the energy intensity of trucks could be reduced by 34% by 2050, relative to a business-as-usual reference scenario.

Another strategy for decarbonizing road freight is by electrification using batteries, hydrogen fuel cells, or electrified roadways. However, due to the relatively low energy density of current battery technologies, electrification with batteries is only feasible for freight vehicles that have short ranges [10]. Indeed, Sripad and Viswanathan [94] find that the cost of a lithium-ion battery pack that gives a large truck (i.e., with a payload of up to 16 tons) a range of 300 miles on a single charge would cost more than US$160,000. By comparison, a diesel-powered semi-truck costs US$120,000. Guttenberg et al. find that, if seven trucks - each one eight feet behind the other - were to travel in a platoon, the range would be extended to 800 miles on a single charge [95].

Electrified roadways include overhead catenary lines, tracks in the road, and inductive transfer of power, which can be combined with other electric powertrain technologies [10]. All systems require additional receiving technology in the vehicle. In partnership with Siemens and several automobile companies, Germany, Sweden and the U.S. are currently testing this technology [96], and a number of other projects are underway in Germany and Sweden [97, 98]. This strategy would only likely be applicable to short, high traffic-density corridors and requires considerable investments. The electrification of a one-mile stretch of a road near the Port of Los Angeles amounted to US$13 million and costs for electrifying a Swedish highway are estimated to around one million euro per kilometer [99].

In the United States and Europe, the transport work done by a truck falls substantially after four to six years [100]. However, old trucks remain in service, and may even enter heavy service in secondary markets if they are sold abroad. As such, in the absence of a policy intervention, new technology may diffuse across the fleet much more slowly than it is introduced.

There are also several important barriers to the adoption of efficient technologies, such as many users wanting a payback of less than two years, lack of access to high upfront capital, lack of information; and doubt regarding the reliability and safety of new technologies [10, 100, 101]. For example, the uncertainty in exploring new technology options have been a market barrier for small fleet shippers (fleet size $<$20) [101], which make up over 97% of all U.S. trucking fleets [102].

Rail freight

Much of the world's rail system is already electrified [64] and thus its potential for decarbonization is closely tied to the decarbonization of the electricity sector. A big exception is the U.S., where diesel-electric locomotives are used on virtually all freight routes. Hoffrichter et al. estimate that the well-to-wheel emissions from the diesel electric locomotives currently in use in the U.S. are 1.02 $kgCO_2$ per kWh of tractive energy delivered, while emissions from electrified locomotives powered by the U.S. grid in 2008 would have been 0.9 kg per kWh [103]. The current U.S. grid is around 20% cleaner than in 2008 [104], and the benefit of a shift to electric traction would be correspondingly greater. However, in countries like the U.S. with very long haul routes, the economics of electrification are not attractive as long as there are no limits on GHG emissions [105].

The Norwegian research organization SINTEF has conducted an assessment of hydrogen and battery-operated alternatives to conventional electrification and concluded that hydrogen or hydrogen-hybrid propulsion would provide a factor of two lower cost solution than installing catenaries on the remaining un-electrified lines in Norway [106]. While somewhat more expensive, in this case, large battery storage units (with reroute replacement/recharge) also dominated electrification in that application.

Inland waterways and coastal shipping

While the carbon intensity on water is lower than on road on average, small vessels carrying less than 250 tons of non-bulk cargo have GHG intensities that are similar to, or higher than, that of trucks [107, 64]. Much of the developing world relies on small vessels for freight transportation [108, 109].

Some zero-cost operational measures may be able to reduce fuel use by up to 33% [110]. Geertsma et al. have reviewed a number of mechanical-electric hybrid propulsion technologies, and conclude that they could reduce fuel use by 10-35% [111]. Vessels that use electrochemical storage in the form of electric batteries have been used for operations with limited ranges [112, 113]. Belgium and the Netherlands plan to bring fully electric and autonomous container barges into operation in 2018 [114].

Air freight

Air freight only handles a small portion the total freight activity in tonne-km, focusing on high-value and time-sensitive goods [13]. Strategies to reduce emissions from air freight overlap with those for passenger air transport, since a large part of air freight is transported by passenger aircraft [115, 116] and most freight aircraft are minor variants of passenger aircraft.

Schäfer et al. [63] find that, for narrow body aircraft, cost-effective measures can reduce emissions by about 20%. Several of these measures such as winglets are already being deployed. Since the dawn of the jet age, aircraft engines have reduced specific fuel consumption by 70%. A recent NASEM study concluded that the overall efficiency of gas turbine engines could be increased by up to 30% relative to the best engines in service today [117]. However, further efficiency gains will likely require a significant change in aircraft configuration [117, 118]. Both U.S. and European agencies aim to produce practical designs that reduce fuel burn by 70% compared to current aircraft by mid-century. Few of the designs developed in response to these goals are likely to achieve them, even after assuming significant breakthroughs in technology, regulatory posture, and customer preferences. Concepts include combinations of laminar flow, open rotors, unswept wings, wing-body hybrids, aeroelastic control surfaces, lower cruise speeds, and hybrid turbo-electric propulsion [119]. It is unlikely that industry will be able to produce motors and generators with sufficient power density (measured in kilowatts per kg) to propel large electric transport aircraft for at least the next two decades, nor is it likely that batteries with sufficient energy density (in kilowatt-hours per kg) will be available [117]. Given the extraordinarily high levels of safety demanded by the flying public, industry and regulators are cautious about introducing new technologies [120, 121].

2.2.4 Strategy 4: Reducing the carbon content of fuel used to transport freight

Low-carbon liquid fuels could help achieve decarbonization of long-haul freight, shipping, and aviation. Drop-in fuels that have characteristics similar to the fuels they replace would avoid the need to rebuild fuel storage and distribution infrastructure, and the need to redesign engines. However, absent a breakthrough in electrofuels or artificial photosynthesis [122] (that is, devising a technology to use sunlight to reduce carbon dioxide to produce sugars and eventually hydrocarbons) [123], pathways to a low-carbon drop-in liquid fuel are constrained by cost [10, 124, **?**] and materiality. In the case of biomass fuels, keeping lifecycle greenhouse gas emissions low requires minimal land use change, which in turn limits the volume of low-emissions fuel that can be produced [124, 125, 126]. Many alternative fuel pathways, including those that use captured CO_2, require the conversion of carbohydrates to hydrocarbons by hydrogenation [117, 127]. The cost of producing hydrogen and the feasibility of doing so without releasing CO_2 constrains the options. It is also technically difficult to build an infrastructure to safely store and transport molecular hydrogen [128].

Farrell et al. [129] suggested that hydrogen could be introduced as a transport fuel to "a small number of relatively large vehicles that are operated by professional crews along a limited number of point-to-point routes or within a small geographic area." Indeed, trucks powered by hydrogen fuel cells are currently in use for drayage at the ports of Los Angeles and Long Beach [130]. Although hydrogen has low volumetric density, a recent study suggested that there would be sufficient space on trucks to store fuel for most medium and heavy-duty vehicles to perform their current daily range of operations [131]. Van Biert et al. conclude that while fuel cells may be attractive for ships undertaking short trips, ships with missions of over 100h are likely to require cryogenic hydrogen storage with a volumetric capacity that is 1.5 to 5 times larger than current vessels [112].

Methane, the primary constituent of natural gas, is a potent GHG. Due to the possibility of leakage, using natural gas as a fuel is unlikely to reduce the greenhouse gas emissions of heavy trucks [132]. Similarly, it would take 30 years - equal to or longer than the useful life of merchant vessels [133] - for the contribution to global warming of a natural gas-fired ship to fall below that of an otherwise identical diesel-fueled ship [134].

Characteristics such as economics, operational conditions, vehicles, and fuel-handling infrastructure are different for each mode, and determine what fuels are feasible. Price sensitivity limits the range of fuels that may serve as economically viable alternatives to fossil fuels in ocean shipping, while high energy density is more important for aviation. For example, even low sulfur (<1%) bunker fuel is much cheaper than jet fuel. While electrification combined with platooning, or the use of hydrogen fuel cells could be used to decarbonize trucks in the next two or three decades, perhaps the only realistic way to substantially decarbonize air freight in that time frame is to directly capture and sequester a volume of CO_2 equivalent to the emissions from the sector [**?**], or to develop a drop-in low carbon fuel [117].

2.2.5 Strategy 5: Shifting freight to low carbon-intensity modes

There is a large difference in the carbon intensity of surface transportation modes, as summarized in Fig. 2.1. Shifting as much freight as possible from road transportation to rail and water is one of the most important means for decarbonizing the freight sector [135]. It is also one of the simplest approaches, as it does not require companies to make large capital investments [71]. For example, the EU has chosen this as a primary strategy to reduce emissions from the freight sector [136, 137].

Often, it is only possible to shift some part of a shipment's journey to a low carbon mode, requiring shipments to be *multimodal* [138]. If a good is transferred in a single unit, e.g. container, during the course of its journey, without unpacking it, this multimodal transport is referred to as *intermodal*[1] [138, 140]. The carbon intensity of an intermodal shipment depends on the types and

[1] Another term is *co-modality*, which is often used in Europe and refers to the efficient use of single or multiple modes in the same transport chain in order to reduce environmental impacts [139, 138]. We will refer to the shipment of goods on multiple modes as intermodal freight transportation in this report.

shares of the modes used, making it difficult to estimate the potential savings in carbon emissions with respect to road [141]. GHG emissions from intermodal shipments hence depend on origin, destination, and the proximity of intermodal terminals [141]. When quantifying the environmental benefits of a shift to intermodal transport, analysts typically explore many different routes to obtain a distribution of carbon intensities. For example, a report commissioned by the International Road Transport Union [142], and cited by the IEA in its 2009 report on transportation [12], found that intermodal shipments in Europe use on average 16% less energy than road shipments, and range from 45% less to over 10% more than road energy use. Craig et al. [141] estimate that rail and truck intermodal shipments have an average carbon intensity of 67 g CO_2/ton-mile compared to 125 g CO_2/ton-mile for trucks, but the distribution has a large variance with a few intensities exceeding the one for trucks due to GHG emissions from terminal operations.

Modal shift focuses on long-haul road freight, as rail and water freight modes have limited use for last-mile delivery [64, 143]. However, decarbonizing the last mile of the shipment by means of technology is comparatively easier, for example through low-carbon vehicles such as rickshaws, tricycles, or electric vehicles [144]. Some authors argue that the much anticipated delivery robots and drones cannot easily reduce freight GHG emissions from urban logistics [145, 10].

2.3 Global overview of freight activity and trends

In this section, we provide a brief overview of global freight activity and important trends using a dataset of road, rail, and domestic water freight activities that we have compiled for 157 countries for the years 2000 to 2017. Freight activity is measured in tonne-km (tkm) and refers to the transport of one metric ton of cargo for one kilometer.[2] The freight activities by mode and country are summarized in Fig. 2.2. While we focus on the three surface transportation modes (road, rail and water), due to the lack of data on water transportation, we often restrict the discussion to the land transportation modes (road and rail) and their modal split. Detailed tables as well as data sources are available as spreadsheets in the accompanying supplementary materials. Any value of freight

[2]There are limitations to using the weight-based metric of freight activity in tonne-km as the primary functional unit, as opposed to a volume- or value-based metric. Often, freight vehicles are full before they reach their weight limit [74]. However, because most public data are available in this unit, using this metric ensures broad comparability.

activity in this review is from this dataset, unless noted otherwise.

2.3.1 Modal structure and trends

Our data show that the global road and rail modal split is 61 : 39 based on data from 75 countries, which are equivalent to 83% of global GDP. Freight activity on inland waterways and coastal shipping is region-dependent and data are scarce. With the exception of a few countries where rail dominates, such as Russia, Australia and Canada, most countries rely heavily on road freight transportation (Fig. 2.3). While countries with large surface areas tend to have high rail shares in freight activity, some South American countries and China are exceptions.

In Fig. 2.4 and 2.5 we show how the share of rail freight activity has changed over time. To make the changes comparable with varying lengths of historical time series, we analyze compound annual growth rates (CAGR) over the available time frames. We find that the share of rail freight has decreased in many countries, particularly in China, India, and Eastern Europe (Fig. 2.4). The shares of rail in Russia, and the U.S. have remained relatively unchanged. The U.S. has shown a recent increase in the share of rail largely due to an estimated decline in road freight activity (see Section 2.5.2). Japan and Australia have increased their rail share. Fig. 2.5, which plots the CAGR of road freight activity against the CAGR of rail freight activity, allows us to identify the mode that drives the change in modal shares. For example, Australia has experienced strong growth in rail, which has largely been driven by growth in iron ore mining activities [146], and some growth in road. This modal shift to rail therefore does not correspond to a decline in road freight. We find that only a few European countries and Japan have experienced an overall decline in land freight activity, while most countries have seen growth with a much higher CAGR for road than for rail. See Section 2.5 for details on regional modal share trends.

2.3.2 Freight infrastructure

While the world's road network grew significantly, with China almost tripling the length of its paved roads [147, 148], the global length of rail track declined between 2000 and 2009. Only China, India and some ASEAN countries added track kilometers. According to the OECD International

Figure 2.2: Road, rail and domestic water freight activity for all countries that provide information on road freight activity, as orange, red and green points, respectively. Countries are ordered by the total land freight activity. The lower figure shows the freight activity of smaller countries at a larger scale.

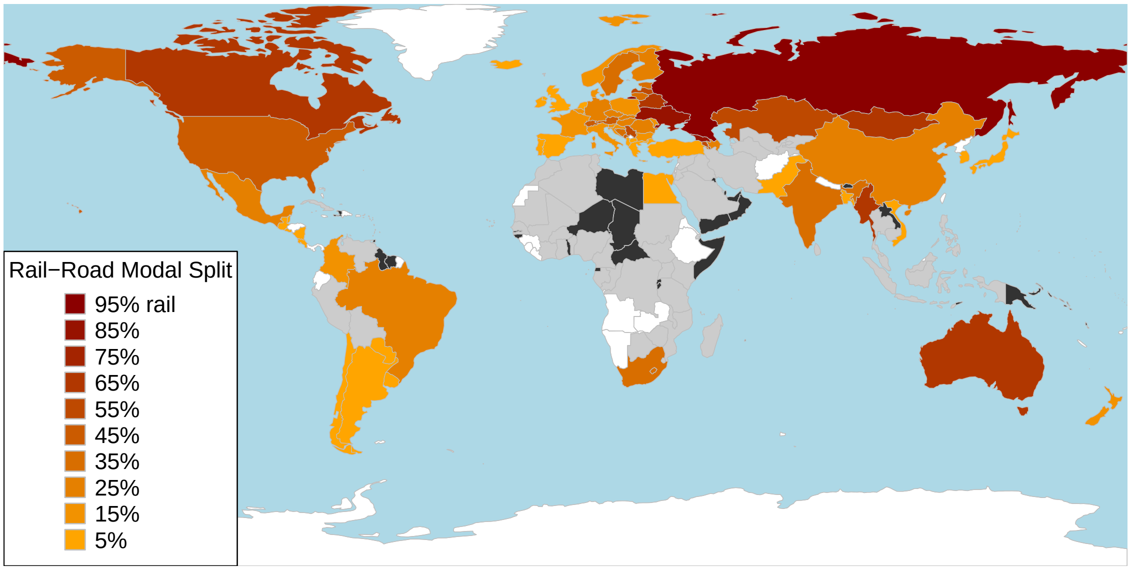


Figure 2.3: Data availability and modal shares of road and rail globally. The color gradient indicates the fraction of road transportation of total land freight activity, with median values of the ranges given in the legend. Orange corresponds to a larger share of road freight. For the grey-colored countries, only the rail freight activity is available. Black indicates countries that do not have a rail system and do not report road freight activity data. Countries with no data are in white.

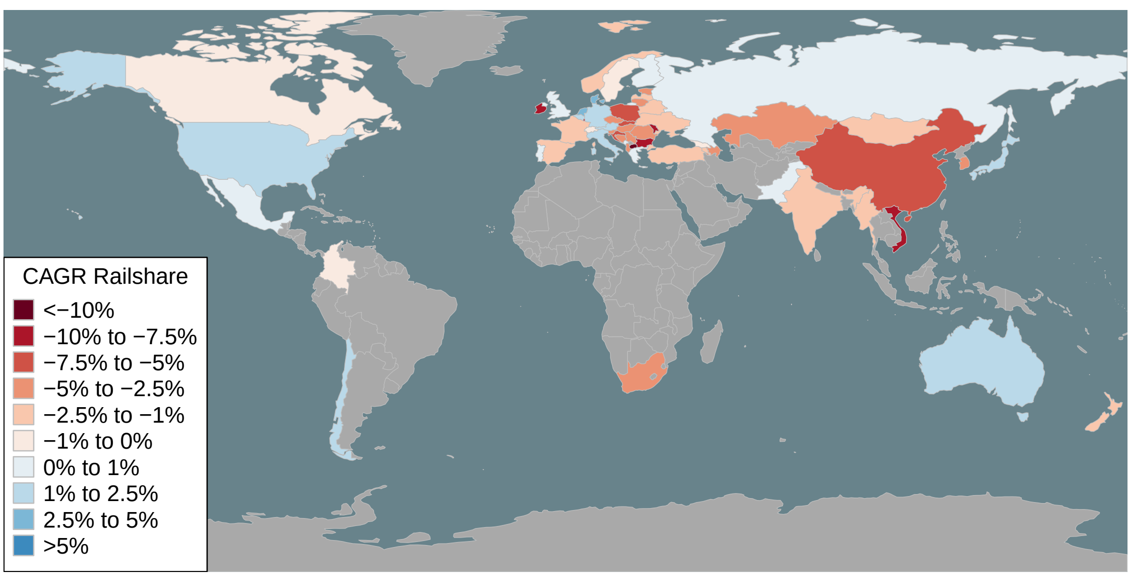


Figure 2.4: Compound annual growth rate (CAGR) in percent for available years between 2000 and 2016. Any country reporting values for road and rail freight activity for more than one year is shown. Blue refers to an increase in the share of rail freight activity with respect to road, and red to a decrease.

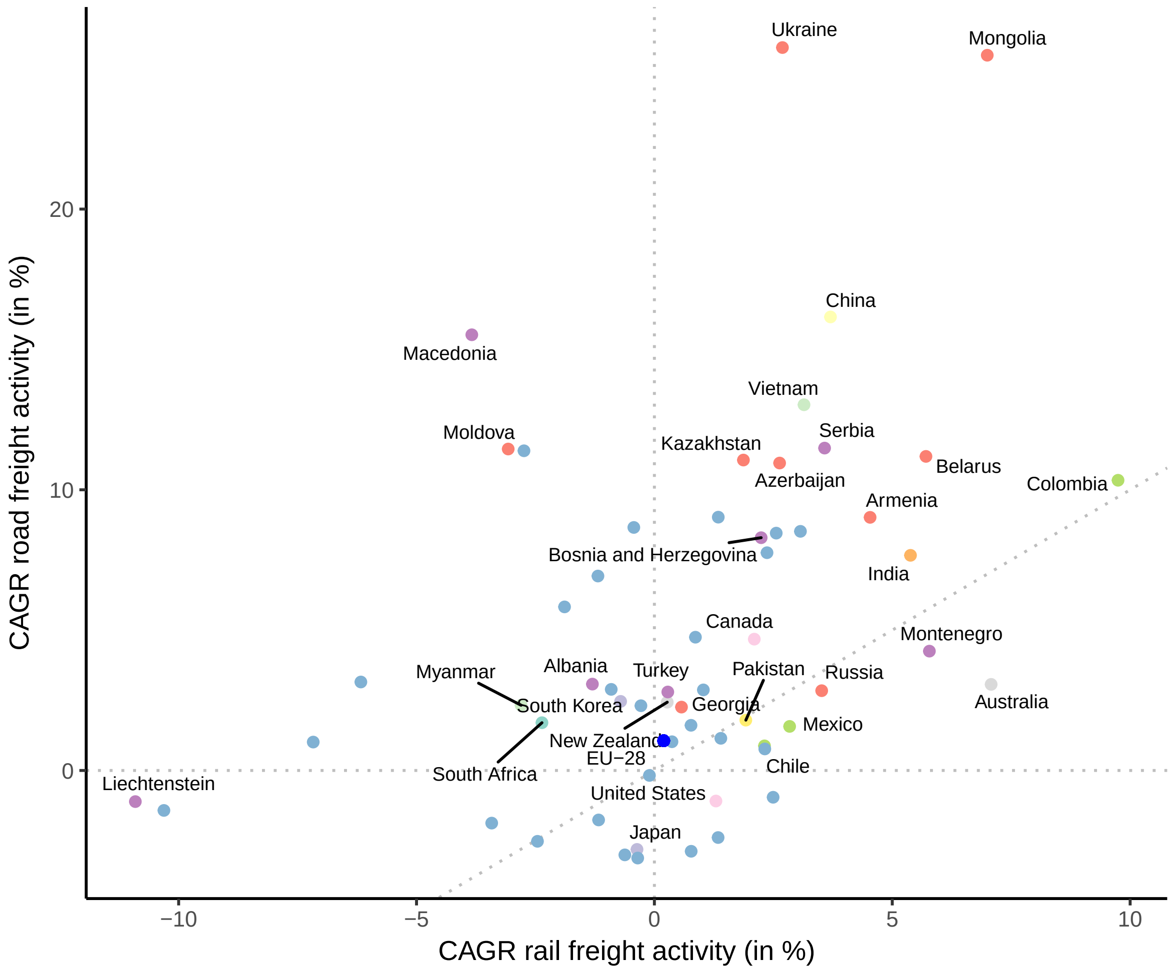


Figure 2.5: Compound annual growth rate (CAGR) of national road and rail freight activity. This figure includes all countries with data for more than one year for both road and rail freight. The colors of the points distinguish the freight regions. EU and EFTA countries are indicated in blue and are not labeled, please refer to Fig. 2.11. The majority of countries experience a large growth in road freight activity, which exceeds the growth in rail freight activity.

Transport Forum (ITF) [149, 150], the largest needs for surface transport capacity expansions are in Asia and Africa followed by Europe [149]. The IEA analyzed world-wide transportation infrastructure needs [148, 147] and estimates that non-OECD countries will account for 85% of land transport infrastructure additions leading up to 2050. The World Bank suggests that to close the freight performance gap between low- and high-income countries, and for exports to be competitive, primarily better - not more - freight infrastructure is necessary [78].

2.3.3 Data availability and estimation methods

Globally there is a lack of data on freight-related key indicators [151, 152], one of which is freight activity. While the World Bank [70] maintains a database of rail freight activity and the OECD [69] reports both road and rail freight activity for many countries, their data are not in agreement and there are large gaps in coverage. We gathered information from governments and international organizations to expand those databases where possible.

The quality of the data is highly mode-dependent. For estimating the potential for decarbonization, data on road freight activity are particularly important, but these data are unavailable for many countries. Freight activity data are especially scarce for low- and middle-income countries. The fragmented nature of the sector hinders the collection of reliable and granular data, which are typically obtained through surveys. The World Bank suggests that a voluntary data-reporting program could be helpful [78]. Freight activity estimation methods are not well-documented. Inconsistencies between surveys might make it difficult to compare data from different countries, but might also be necessary given local specifics of freight systems. For example, the EU requires member states to comply with a standardized framework [153], which excludes trucks with a capacity of less than 3.5 tonnes [153]. In many regions a significant volume of freight is moved by vehicles of this size, such as in Japan [10]. Vietnam's statistics even account for freight on motorcycles [154]. If these countries were to adopt the same framework as the EU, they would underestimate the volume of road freight carried. Rail freight activity is typically better documented as there are usually few rail companies operating in a given country. The most difficult data to obtain are on inland waterways and coastal shipping.[3]

While some analysts adjust raw national or regional statistics, we have not done so. Our database includes only unadjusted figures, and our results reflect unadjusted data but make a mention where others have seen such corrections warranted. Adjustments might for example be necessary for estimating GHG emission data, such as in a recent report by the IEA [10], in order to match top-down and bottom-up approaches. We find them less important for measuring the change in modal shares.

2.4 Strategies to promote modal shift

2.4.1 Mode choice

Shippers choose the mode of transportation based on shipment characteristics, commodity characteristics, costs, modal access, and environmental considerations [13]. For most shippers, service quality is of great importance [155, 156, 13]. Switching carriers is costly if contracts need to be broken, or new information-sharing arrangements need to be put in place [155]. Below, we describe some of the features that influence mode choices.

Shipment and commodity characteristics

Low-weight and high-value goods are frequently shipped by road or air, while heavy, low-value bulk goods, such as coal, grain and gravel, are often transported by rail or water [13, 155]. The decisions on shipment modes are influenced by the limit that a country and operators impose on truck and rail car weights. Shipment volume is constrained by vehicle or container size, and many shipments *cube out* before they *weigh out* [10, 13].

Rail dominates long-distance freight, which is for example defined as more than 500 miles in the U.S. [13]. Shifting freight from short-haul trucking to rail is difficult because local rail lines are not in place and, with the exception of some heavy industries and particular bulk commodities such as

[3] In those cases where coastal shipping cannot be distinguished from general maritime shipping, we have not included it. This might cause a general underestimation of water shipments in our analysis. See the supplementary information for details.

coal, transfer centers or warehouses are mostly not designed for railcar-size deliveries [13].

Some commodity groups, such as live animals or chemicals, require specialized equipment. Shelf life, temperature and humidity requirements, sensitivity to acceleration forces and the risk of accidents can limit the modal choice [157, 13]. Protection of the shipment may lead to shippers only trusting a few carriers to transport high-value goods such as luxury cars [155, 13].

Modal characteristics

Transit time and reliability are critical components of mode choice [13, 156, 158]. Reliability determines whether requirements such as just-in-time, quick response, port deadlines, and hub-and-spoke operations can be met [156]. It is affected by congestion, maintenance, accidents, natural disasters, extreme weather, and mismanagement [158, 13]. Shippers consider trucks the most reliable mode, because of, e.g., the ability to avoid congestion by taking alternative routes [13]. For just-in-time operations, high-value or high-demand goods, also the shipment frequency is important, as it affects inventory costs [13], and shippers place similar importance on service frequency as on reliability [159, 160].

Shipping costs, handling costs and modal elasticity

Some studies argue that shipping costs are less important for modal choice than quality of service attributes [161], while others suggest that shipping costs, especially when including inventory cost or financial risks, are decisive [13, 160]. Clearly the factors discussed above are not independent: for example in-transit carrying costs vary by shipment time, or inventory costs are affected by frequency and reliability.

De Jong et al. (2010) [82] review elasticities and cross-elasticities in the freight sector. They find that transport price changes are more likely to drive changes in modal shares than in absolute freight transport demand. Furthermore, they find considerable variability in the elasticities and cross-elasticities, which are highly depend on shipment characteristics described above and on the market condition of the modes. Similarly, Christidis and Leduc [162] have reviewed a number of studies and found cross elasticities of rail and road that range from 0.3 to 2. This means that a

10% increase in the cost of road freight will increase rail freight by 3-20%. Other authors present cross-elasticity estimates as large as 3 [163]. For intermodal transport, Arencibia et al. find cross-elasticities of greater than 2 with respect to costs, which are smaller than the ones they find with respect to transit time [160]. Elasticities often cannot be transferred from one country to another [82], and those values likely do not hold for low-income countries.

Environmental considerations

Although not a primary concern for most shippers [164], there is increasing pressure on firms to reduce the environmental footprint, including CO_2 emissions, of their logistics supply chains. Programs such as SmartWay by the U.S. Environmental Protection Agency and other international green freight programs [165] aim to make environmental performance more transparent and to ease the integration of environmental impacts into firms' decision-making [101].

2.4.2 Barriers to modal shift and increased use of intermodal transport

Time of delivery and reliability

Rail intermodal shipments have a significant disadvantage compared to trucks when speed and predictable time of delivery are priorities [157, 166, 167]. For example, freight train delays due to extreme weather can induce a shift back to trucks [168]. Winebrake et al. [140] compared costs, transit time and GHG emissions of route options, and the lowest-emissions choice, predominantly rail, was sometimes cost competitive but always had a longer shipping time than other mode options. However, there are also cases where rail intermodal is faster than the conventional lowest cost-option, as for example the Europe-Asia road-rail link that is faster than ship [157]. To speed up intermodal, significant investments in freight capacity, modal connectivity, and efficiency of operation are required [12, 147, 148, 169]. Intermodal would also be more attractive if clients would adjust their strategic and operational decisions to move away from practices like just-in-time manufacturing [170, 12]. In just-in-time production, which originated in Japan, outside suppliers are required to deliver their parts in smaller batches in time for immediate use, thus reducing the need for warehousing [77]. It requires frequent, fast and reliable delivery times, that also resulted in a disadvantage for modes such as rail and inland water [12]. New regulation that incentivizes collaborative logistics, higher inventories and larger shipments could slow down supply chains and enable modal shift [166].

Distorted pricing of transport costs

The external costs of road freight transportation are not adequately represented in transportation costs, and a stronger price signal could promote modal shift. One problem is the lack of granular data, which prevents the accurate assessment of environmental performance and obstructs policy analysis and decision making. A fragmented and informal road freight industry, in particular in the developing world, is hard to regulate [171], and competition makes carriers reluctant to share proprietary information.

2.4.3 Policy, infrastructure investments, and technologies to promote modal shift

We distinguish two main strategies to promote a shift of freight transportation from road to modes with lower carbon intensity: improvements to the infrastructure and operational efficiency of the freight system and incentives. Such policies include infrastructure investments, promotion of new technologies, subsidies for low-carbon freight modes, regulation of the road freight sector and the internalization of external costs, as described below.

Infrastructure investment in rail, waterways and intermodal connectivity

Infrastructure investments in new or dual rail tracks, waterways, intermodal terminals, and inland ports can increase rail and water freight capacity, enhance modal connectivity, consolidate loads, improve quality of service and extend the rail network to new locations. However, such investments are often lacking. The International Transport Forum (ITF) found that in 2014 the average public and private investment in inland transport infrastructure in OECD countries was 0.75% of GDP. The rail share of this investment was less than 30% [150]. As emerging economies and developing countries are experiencing rapid growth in transportation demand, there is a pronounced need for new transportation infrastructure. The IEA states that by investing in a low-carbon infrastructure and inducing modal shift, these countries could meet demand, boost their economies, and reduce costs [147]. In Africa, for example, the little existing railway infrastructure has been poorly maintained, resulting in a drastic reverse modal shift from rail to road [172, 173]. Operational, financial and contractual standards for intermodal terminals, such as those that exist for ports in the World Bank Port Reform Toolkit, are necessary [174].

In industrialized countries, there is a particular need to develop intermodal terminals and dry ports [175, 176, 177]. Dry ports are inland terminals, close to demand, which receive goods directly from the seaport, preferably by rail or inland waterways [176]. They allow containers to be handled close to demand centers (the concept of *extended gateway*) and increase the share of low carbon modes in the goods' journey [178, 179]. Large cargo-handling facilities may set themselves modal split goals that have an impact on port infrastructure designs [149, 180]. For example, the port of Rotterdam aims to ship 65% of incoming sea freight into the European hinterland by rail or inland waterways by 2035 [180]. Besides environmental considerations, those targets may also be in place to increase capacity or expand the port's reach by making longer-distance shipments available [180].

Intermodal operations research and planning

Operations research, which is essential for efficient infrastructure design, supply chain logistics and terminal operations, can reduce the carbon intensity of intermodal transport [181] and foster modal shift [182]. As intermodal transport research has emerged as an independent field, a number of studies have surveyed the state of the literature [183, 184, 185]. Although a number of freight transport models have been developed [186, 187], and tactical and operational issues have been studied, the lack of realistic physical topologies handicaps modeling freight flows. For example, SteadieSeifi et al. [138] note that most analyses model hub-and-spoke freight networks, while many goods are in fact transported along dedicated corridors. In addition, most models optimize for lowest cost: there is a need for models that include other (e.g., environmental) objectives [138, 187].

Integration of services between modes

Besides the *vertical integration* of dry ports, Logistics Service Providers (LSP) can facilitate coordination between different shippers to reduce cost, travel time and GHG emissions *horizontally* [188, 189, 135]. The EU emphasizes efficiency gains through this cooperative approach and the provision and exchange of reliable information on sustainability metrics [190, 191]. LSPs may allow shippers to choose the desired cost, speed, and level of environmental impact of a shipment, and devise a combination of modes and carriers to meet the shipper's requirements [192, 193]. LSPs can also offer more frequent services to all customers.

One of the key notions regarding the integration of services between modes is *synchromodality*, which refers to a situation where modal combinations and operational schedules can be changed after the shipment is on the way, in response to new information [138, 189, 190, 194, 195]. This ensures reliability and prevents data lock-in with a carrier.

Pfoser et al. [193] note that successful synchromodal transportation needs close cooperation between all stakeholders of the logistics chain, a willingness to change existing practices, and harmonized transport regulations, including for data sharing. Contract structures must be developed, which provide legal security for all market participants for the liability for delay, loss or damage. They do not see the development of infrastructure and information technology as significant barriers, although more research and modeling work is needed [193, 195].

Consolidation of smaller shipments that do not fill out a container or truck is offered through the Less-Than-Truckload (LTL) market, which provides door-to-door services where customers can select the transit times and levels of reliability [181].

Enabling efficient intermodal transport through information technology

Information and communication technology (ICT) can be used to improve and automate terminal operations, track shipment locations, improve security and quality control, administer data and aid with routing decisions when optimizing and pricing intermodal shipments [196]. ICT can also be used to automatically assess or collect tolls or constrain heavy vehicles to certain corridors [10, 197]. All of these functions could contribute to decarbonizing the sector [196, 198].

Terminal and port ICTs can address everything from customs information to fully automated freight shuttles between ports and dry ports [196]. Tracking devices include Radio Frequency Identification (RFID) systems, GPS systems, and fieldbus communications networks, the last of which allows the distributed control of widely dispersed sensors [199]. These technologies can track freight trains, railcars [200], or individual intermodal containers [201]. ICTs can also facilitate the use of intermodal for the market of perishable goods by allowing continuous temperature monitoring in refrigerated containers [201]. Limited GPS coverage, signal blockage, dependence on batteries, and maintenance are problems, but technology is evolving fast [201]. Furthermore, the lack of standards for technologies and for data exchange retard the adoption of these technologies [199].

Consolidation in logistics, as for example in the LTL sector, is a complex optimization problem, which benefits greatly from ICT [199, 202]. ICTs could also provide firms involved in logistics - some of whom might be mutual competitors - secure and anonymous platforms to coordinate activities. For example, an Electronic Logistics Marketplace (ELM) can permit different companies to anonymously pool their shipping requirements without sharing confidential information or entrusting it to a third party [196, 203]. Shipping companies and banks are beginning to explore the blockchain technology for transactions with customers and tracking ocean and road shipments [204, 205, 206].

ICT is also essential for improving the cross-border compatibility of rail systems. Since the 1990s, the EU has worked with the private sector to create a standardized rail communication and signaling system with investments of over €770 million [207]. The European Railway Traffic Management System (ERTMS) is one of the most advanced train control systems [200] and it is employed also in many countries outside Europe [208]. Similarly, the EU fostered the development of an ICT system for inland waterways, the River Information System (RIS) [209]. In the future, these Intelligent Transport Systems (ITS) could be combined into an efficient multimodal digital freight system [209, 196]. Many Asian nations have also recently introduced ICT systems to make border crossings faster [210].

While the overall ICT strategies described above would yield some carbon reduction potential, we are not aware of a systematic assessment of what that potential could be. This could be an important contribution to the literature that could support policy decision making for climate mitigation in the freight sector.

Regulation and subsidies of low-carbon freight modes

Third-party rail track access can be publicly regulated. The EU requires infrastructure managers to give access to third-party railway operators, while this is done voluntarily in the U.S. [211]. The lack of track access may make hinder growth in rail traffic by allowing incumbent railroad operators (who in the U.S. also own the tracks) to block new entrants by charging high track access charges [212].

Governments may choose to directly subsidize modal shift because it supports the goal of decarbonization. The Marco Polo Programme of the EU [213] compensated projects with €1-2 per 500tkm shifted (the equivalent of EUR 25-50 per tCO_2 emissions avoided). Although an early assessment has found that the program has underachieved targets for modal shift [214, 215], studies suggest that direct subsidies can be successful [182]. Belgium subsidizes the transfer of containers from sea ports to inland waterways [216].

GHG pricing and internalizing external costs

In transportation, external costs include air pollution, noise, accidents, congestion, infrastructure damage and climate change. While this paper focuses on climate change costs, policies often address multiple, region-specific externalities. If revenues from Pigouvian taxes on road transport are applied to improve other modes, policies can be more effective in promoting modal shift and might gain more public support [217].

Assuming a carbon price of €90 per tonne CO_{2eq} and a discount rate of 3%, the average marginal climate change costs for heavy freight vehicles are estimated at 2.5 to 10.4 euro cents per vkm [218], which is around 0.3 to 0.4 euro cent per tkm but can also be much higher depending on load factors and emission standards. Diesel-powered rail is estimated to cost 0.26 euro cent per tkm for a load of 500t, and electrified rail is likely lower. These values are larger than the estimates of accident costs but substantially smaller than congestion costs. While a Belgian study showed that internalizing all those external costs can induce modal shift to lower-carbon freight modes [219], very high taxes might be needed induce modal shift [13]. This is also demonstrated by the anemic response to the

Marco Polo program discussed above. Combining taxes with other decarbonization policies ("policy packaging") might be more effective [217].

External costs of rail freight and inland waterways are typically lower than for road transportation, but the case of intermodal transportation is less clear. Mostert and Limbourg (2016) [220] find that internalizing external costs can make intermodal shipments more cost-competitive but highlight that this effect diminishes for longer drayage distances. Similarly, Santos et al. [182] show that internalizing external costs can also hurt intermodal transport competitiveness, depending on the length of the road haul.

Van Essen et al. [217] find that fuel and vehicle taxes, vehicle charges and in some cases emission trading are the most common policies to internalize the external costs of transport. They recommend a strong differentiation for charges and taxes and a clearly labeled CO_2 tax or emission trading for climate change costs. Emission trading schemes such as the EU Emissions Trading System still do not include the transportation sector, with the exception of aviation [221]. We focus on fuel taxes and vehicle charges the following sections.

Motor fuel taxes

Harding [222] finds that in almost all OECD countries, diesel fuel taxes are lower than gasoline taxes. She concludes that the externalities associated with the fuels do not justify this tax differential and advises its gradual removal. In particular countries with large distances between cities, or few alternatives to road freight transportation tend to have lower fuel taxes [10]. While some countries still have diesel fuel subsidies in place, a few recently moved to eliminate them [223].

While much scholarship evaluates the effect of fuel prices on the demand for transport and fuel, there is less empirical work on the effect of fuel taxes. An analysis that sought to delineate the consumer response to gasoline taxes from the response to gasoline prices [224] found that the tax-exclusive price elasticity of demand for gasoline was -0.03, only half as large as the tax elasticity of -0.069. This may be because taxes are publicly debated and are therefore more salient or because consumers see tax rises as more permanent than transient price rises. In fact, the tax elasticity is lower in states where the taxes change frequently [224]. We have not encountered studies that separate the effects of prices and taxes for the trucking sector, and an empirically study would be instructive. One could speculate that the contractual arrangements that allow carriers to pass on increases in fuel prices [80] were designed to protect carriers' slim margins from market volatility. That does not apply to predictable taxes. However, some studies suggest that tax increases are rapidly and fully reflected in shipping prices [225]. The majority of the studies focus on the United States, Canada, and Europe, and empirical evidence from large developing countries remains scarce.

Road user charge and corridor regulation

As of 2017, most EU countries have implemented tolls for heavy-duty vehicles [226], and this has proved to be effective in inducing a shift to rail or water in Switzerland, Germany and Austria [227]. The U.S. Department of Energy concluded that direct user charges in the form of tolls for trucks could have the largest impact and it is the policy most likely to be implemented in the U.S. [13]. Apart from aiding modal shift, road user charges may generate revenue for infrastructure maintenance, reduce congestion, increase logistics efficiency, and charge foreign-registered vehicles for the use of national infrastructure and the externalities they impose [197].

Technology can make road pricing more efficient and increase compliance rates. Germany has implemented, and other countries are studying, a satellite-based system to collect heavy-vehicle tolls [197]. This system also includes charges on secondary roads that have previously been used to avoid tolls. Australia's satellite-based Intelligent Access Program (IAP) monitors where, when and how large and heavy vehicles are operated, and could be used to restrict those vehicles to corridors where more sustainable alternatives such as rail or water are unavailable [10].

Labor rules

As labor costs make up a significant portion of trucking costs, reducing the maximum hours of service can induce a shift of 2-3% of the U.S. tonnage from road to rail [13]. The entry barrier to the road freight industry is very low in most developing countries resulting in high competition between numerous single-driver companies with low prices and low labor standards [78, 228]. National minimum wage rules also influence which countries bare the largest share of road freight traffic in a region, as for example in Europe [229]. While labor rules have in the past played a significant role for port and terminal operations, this is no longer as important [77]. However, strikes at container terminals and poor human resource management can impact reliability and productivity [230].

Truck size and weight regulations

Restrictions on truck sizes are in place in most countries, mainly to protect the road infrastructure and to address safety [10]. Road damage increases as the $\sim 2^{nd}$ to $\geqslant 5^{th}$ power of vehicle weight [231]. In some parts of the world, however, regulations are poorly enforced, resulting in high profit margins for truck operators due to overloading, speeding and lax vehicle requirements [232]. Tightening and enforcing regulations, by cracking down on corruption, would reduce the cost-competitiveness of the road freight sector and promote a shift to rail [78].

In many developed countries, the discussion instead focuses on allowing larger and heavier vehicles (LHVs), which are already permitted in Canada, Australia, Brazil and Scandinavia [10]. Increasing the maximum permissible truck size and weight to more than 70 tons can decrease the carbon intensity [163], but could make road freight cheaper and induce a shift away from rail and water [10, 13]. The effect on the relative share of road and rail freight of allowing LHVs varies by country, depending on the associated cost reductions and the fraction of rail freight that is containerized [233, 163]. Predictions of the modal shift potential of LHV policies are difficult, since modal cross-elasticity values are uncertain and many current studies were conducted by groups with an interest in promoting one or the other mode [163, 162].

A study in Sweden found that allowing longer vehicles capable of carrying two containers for intermodal drayage can reduce total intermodal transportation costs by 5-10% [234]. There is a need for more research on role of LHVs for low-carbon intermodal freight transportation [233].

2.5 Freight systems and modal shift by region

China, the United States, India and Europe together account for more than 78% of the reported road freight activity (Fig. 2.6). As data on road freight activity are not available for large parts

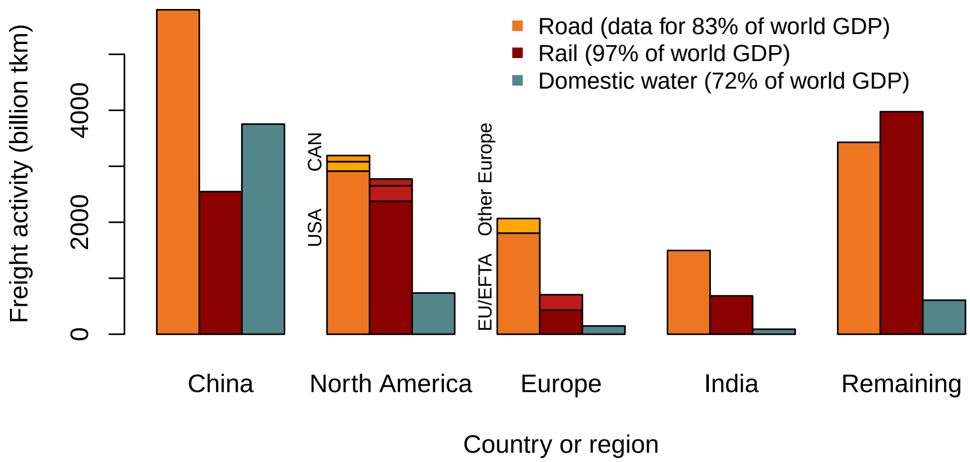


Figure 2.6: The four economies with the largest road freight activity in our dataset. The values for road and water freight activities in the remaining part of the world are most likely larger, since there are many countries for which there are no data available. The road and rail freight activities for Canada are split in domestic (bottom) and international shipments (top).

of the world, the true percentage is likely lower. The road freight sector has grown faster in low- and middle-income countries than in high-income countries [78]. China has recently surpassed the U.S. as the country with the largest road freight activity (Fig. 2.7). Fig. 2.8 shows the modal shares, trends and missing data for the remaining regions. In this section, we summarize the modal shares and political, economic and geographic particularities that have an influence on modal shift in those regions. We devote the greatest attention to those regions with the largest road freight activity.

2.5.1 China

Much of the freight in China is transported long distances, which gives the railway system a competitive advantage. However, while rail used to have the largest share, most freight is transported by road now. Since 2000, China's road freight transport activity has increased by more than eightfold, while domestic water shipping has more than quadrupoled, and rail freight activity has increased by a more modest 70%. There was a revision of the road freight activity reporting in China that resulted in a jump in reported values around 2008 [10]. A large fraction of the total freight activity

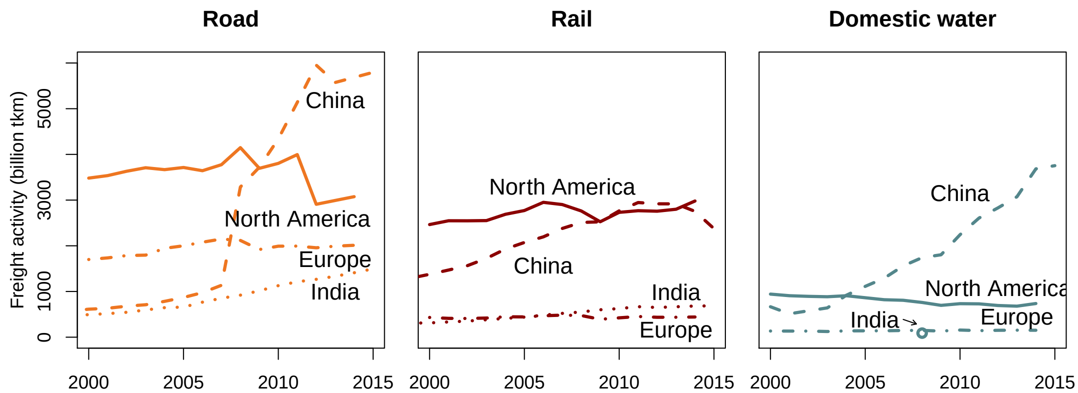


Figure 2.7: Recent development of freight activity of the four economies with the largest freight volume. Indian waterborne freight activity has only been estimated for one year.

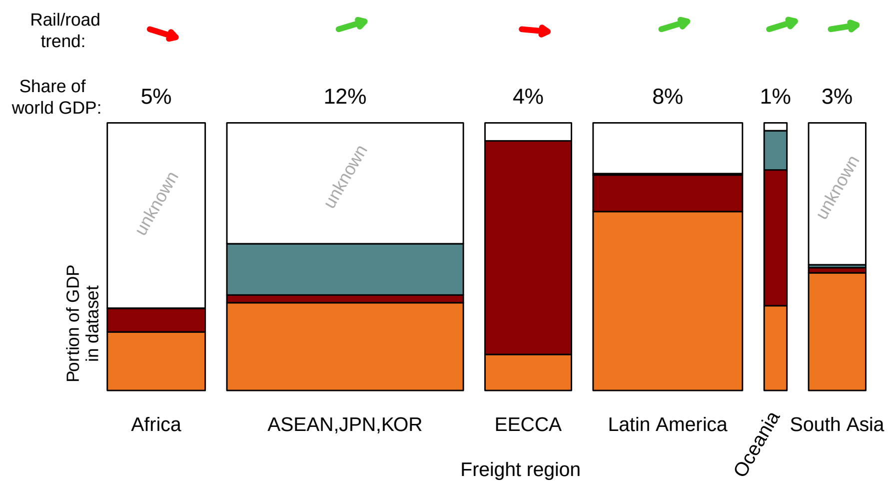


Figure 2.8: Summary of modal share, trends and missing data for each region. There are many countries that do not report road freight activity, and the white part of each bar corresponds to the fraction of regional GDP that such countries make up. The colored portion of the bar is split in the proportion of the total freight activity of road, rail, and water for the countries in the region that at least report road and rail data. We also indicate the proportion of world GDP of each region by the width of the bar. The arrows at the top of the graph show the approximate trend in the rail-road split: a downward arrow suggests that the share of rail is falling. For example, in the ASEAN-Japan-Korea region, both road and rail freight activity is given for four countries, which together make up 55% of regional GDP. In this region, road makes up 60% (orange) of total freight activity, rail 5% (red) and water 35% (blue). The remaining nine countries of the region do not provide road freight data.

is coastal and river shipping [235, 236], which was at 31% in 2015. Duan et al. [237] estimate that in 2012, 88% of transport GHG emissions in China were from the freight sector, mostly from road vehicles. Electrification reduced the rail carbon intensity in the early 2000s [236], although Chinese trains still have relatively high emissions because most electricity is generated from coal [238]. Luo et al. [236] find that, due to differences in economic development, transportation GHG emissions in China differ significantly between regions.

The road transportation sector is fragmented with an average fleet size of three [235] and harsh working conditions for drivers [228]. Trucks are often old and overloaded and there is intense competition, which creates very low prices. GPS tracking and telemetry are largely absent in the road freight sector [228]. Rail intermodal is underdeveloped [239] and it has a lower priority than rail for military, passenger, energy, and food transportation [235].

Select policy instruments in place

Since 1995, China has pursued the development of a high-capacity freight network with intermodal corridors and hubs, connecting economic and industrial centers [210]. The Chinese government is the main investor in infrastructure, particularly in the railway sector, and project developments are fast [235]. China invested 5.4% of its GDP in inland freight infrastructure in 2015, the highest investment rate reported by the ITF. However, more than three quarters of this investment was in road infrastructure [240]. With the *Belt and Road Initiative*, China is also rapidly expanding its transportation infrastructure outside of its own borders [241]. The inefficiency of the inland freight system is one of the main challenges for China's logistics system and is in stark contrast to some of the largest and most modern container ports in the world located in China [228]. One of the main barriers to modal shift is modal connectivity and ports are mainly serviced by trucks [235, 239]. For the line-haul movement of containers, on certain corridors the use of rail is prioritized and other modes are restricted [210].

Conclusion and policy recommendations

Blancas et al. [239] propose to reform the regulatory environment to allow Chinese railways to take a more customer-focused approach, similar to the deregulation in the US. They believe that as new high-speed rail has freed up capacity, such developments become more likely.

Luo et al. [236] suggest a regional approach where wealthier eastern regions could experiment with measures such as fuel taxes while western regions focus on infrastructure development. However, different fuel prices in different regions might encourage smuggling.

Jun and Bensman [228] recommend requiring the trucking industry to improve working conditions, reduce overloading, eliminate falsified paperwork, and retire overaged fleets. Well-designed incentives might induce investment in more efficient modern fleets. We expect that these efforts would both lower the carbon intensity and increase the costs of road transportation. The latter could induce a shift to rail- or waterways.

2.5.2 USA and Canada

The transportation systems of the U.S. and Canada are similar and closely connected. North America's geography is ideal for rail freight, which has a high share of land transportation (63% in Canada and 48% in the U.S. in 2014). Most freight activity in the U.S. is on road, which grew strongly in the 1990s [155]. Both the U.S. and Canada have seen a recent increase in the rail freight activity. In Canada, however, road freight activity has grown considerably faster than rail (Fig. 2.9). Domestic water freight activity has declined in North America[4] and inland water transportation on rivers and the great lakes plays a modest role.

Freight dominates passenger railway operations in North America [244] and the rail freight system is considered safe, efficient and cost-effective [245, 246]. Intermodal transportation in North America grows modestly [155, 242, 247] but it is now the largest single source of revenue for U.S. freight railways, which before was coal transport [248]. The Association of American Railroads attributes this growth to factors including investments, improvements in service quality, and truck driver

[4] Statistics Canada does not provide information on water freight activity but it reports that the domestic tonnage has been fairly stable, while waterborne trade with the U.S. has declined [242, 243].

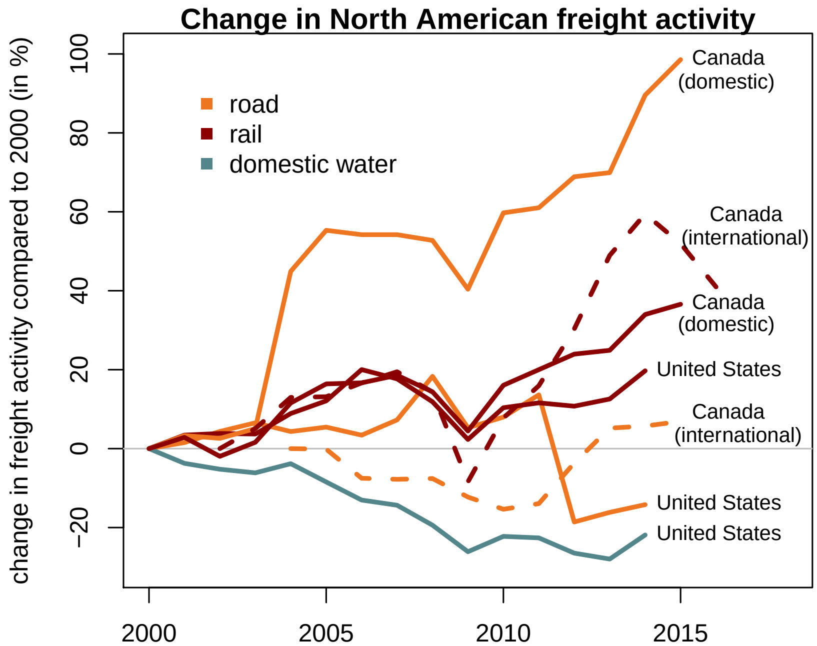


Figure 2.9: The percentage change of road, rail and water freight activity in the U.S. and Canada since 2000. Canadian values are reported separately for shipments that have an international origin or destination ("international") and for those that are only domestic ("domestic"). Values for the U.S. after 2011 are not based on the Commodity Flow Survey. We refer to the SI Section 6.1.1 for details.

shortages among other factors [248]. In Canada, rail transit times are decreasing and intermodal travels faster in average than other commodities [242].

A report by the U.S. Department of Energy (DOE), based on 2007 data [13], suggests that there is an opportunity to move freight carried over distances of 250-750 miles from road to rail. It finds that a substantial share of the freight activity has distances between 250- and 1,500-miles. In a 2017 report, Zhou et al. from Argonne National Laboratory (ANL) [249] evaluate the modal shift potential based on commodity flows and rail level of service information. They estimated that 4.1% of truck freight activity could be shifted to rail, resulting in 4.4% reduction in total freight CO_2 emissions by 2040. Canadian studies have also found that modal shift can contribute to decarbonizing the freight sector [250, 251].

Select policy instruments in place

Intermodal policies have been part of the North American transport strategy to improve the freight system since the 1990s [72]. For example, a detailed analysis of potential policy approaches was undertaken in 1998 [252]. Policies have mostly targeted corridor development and infrastructure financing [213, 242, 253], as well as the development and employment of intelligent transport systems [72, 213]. The U.S. invested about 0.6% of its GDP in transport infrastructure in 2015. Less than 10% of it was in the rail sector [240], one of the lowest shares among OECD countries. For historical reasons, the U.S. Federal Government favored investments in highways over railroads [254]. As a consequence, road freight, which travels over publicly-built and maintained roadways, enjoys an implicit subsidy relative to rail transport, which is privately owned and where freight rates must reflect the private cost of owning, building, paying tax on, and maintaining the infrastructure [255].

For the Canadian government, reducing the carbon emissions from the transportation sector is key to reaching GHG emission targets [242], in particular as some provinces have high shares of renewable energy generation [256], which limits the scope for reducing emissions from the electricity sector.

North American freight rail companies are privately owned and vertically integrated as they own both infrastructure and railway operations [245, 242]. In the U.S., federal deregulation was adopted in the 1980s and 90s. This removed constraints on rail companies such as maximum rate regulations, and allowed them to enter into confidential rates and services contracts and abandon lines [155]. The U.S. and Canada have significantly reduced the extent of their rail track network [147]. There is no track access regulation, and track sharing is voluntary [211, 245].

The U.S. and Canada are among the countries with the lowest motor fuel taxes in the world, translating to a carbon price for diesel of US$2.50 or less per tonne of CO_2 [222]. In the U.S., diesel taxes are lower in states with a larger proportion of employment in the trucking industry [257]. Only four U.S. states (Kentucky, New York, New Mexico and Oregon) have tolls specifically targeting heavy trucks. There are few reports of interest in extending tolls in the U.S. [258] as well as Canada [259].

Canada allows long and heavy combination vehicles with two semi trailers [242]. In the U.S., heavy vehicles are largely restricted to areas near the US-Canada border to facilitate the smooth movement of goods between the two [155]. The U.S. and Canada both have heavy-duty vehicle GHG emission standards [10, 242].

Conclusion and policy recommendations

Although the share of rail freight is high in North America, the competition from the road sector has induced a shift from rail to road. North American investments and policy-making lag Europe, China, and India in efforts to promote a modal shift from road to low-carbon modes. There is a clear opportunity to use motor fuel taxes and road charges to provide such incentives. Heavy-vehicle tolls have been identified by the DOE as an impactful policy for modal shift [13].

Due to the presence of regional rail monopolies in North America, connectivity between different rail segments of long-distance rail shipments is a barrier to modal shift. More terminals that facilitate efficient transfers between different rail lines could accommodate fragmented ownership and relieve congestion in North America [246]. Upgrading intermodal terminals to facilitate the transfer of more commodity groups may also help to reduce GHG emissions [260].

The decline of coal and the surge in oil and natural gas drilling activity in the North America are currently changing the rail business. For example, a single new well pad can require up to 40 rail carloads of equipment and raw materials [261]. In regions, where a pipeline infrastructure is not in place, liquefied natural gas is also transported by rail, as for example in the state of Pennsylvania [261]. At the same time, the volume of coal shipped has declined sharply as demand in the electricity sector has fallen [262]. The transition from coal to intermodal freight as the largest source of revenue [248] is changing geographical patterns of rail freight activity, impacting investment needs, and posing a business challenge for railway companies. This challenge has already been recognized by the Canadian government [242].

2.5.3 Europe

For the European Union (EU), modal shift has long been one of its key strategies to decarbonizing the freight system [72, 263, 137]. Freight demand in Europe is relatively stable, after a period of

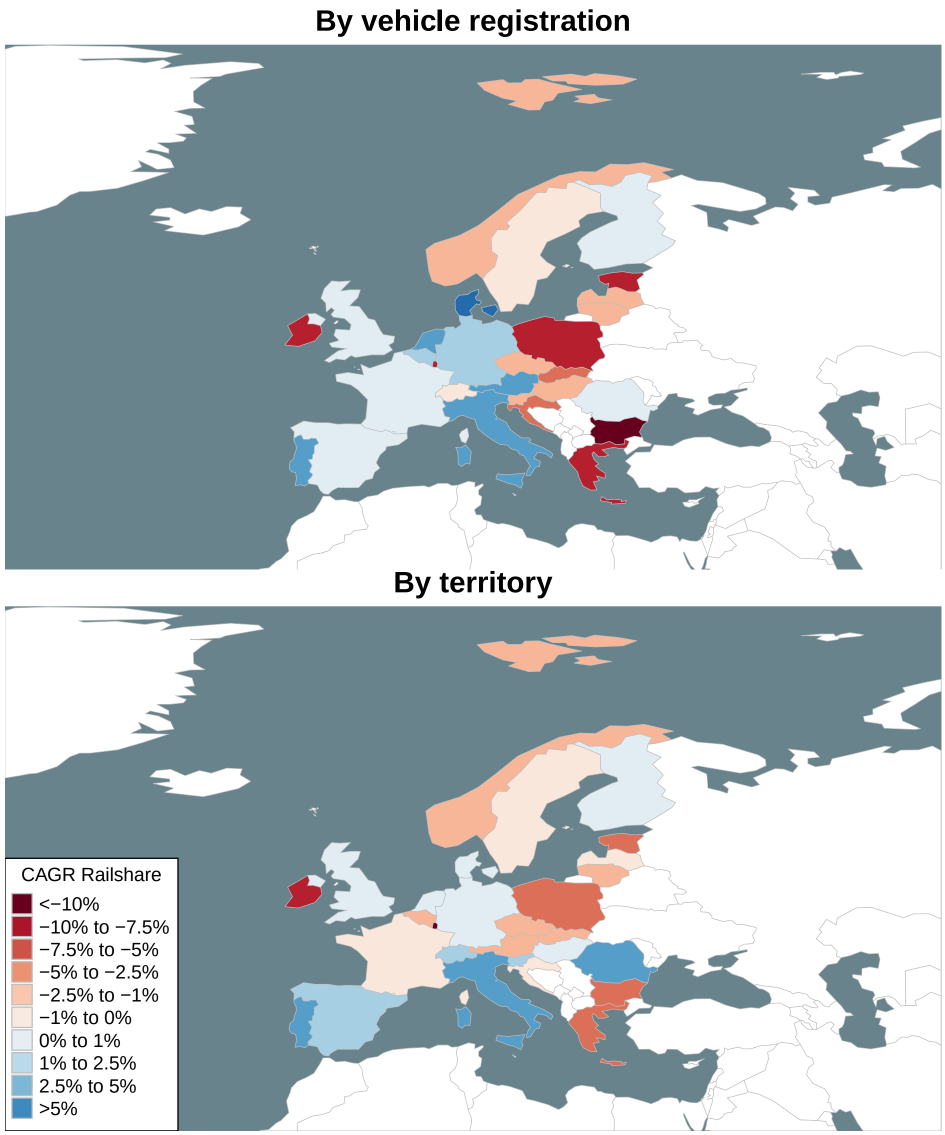


Figure 2.10: The CAGR of the rail share of total surface transportation in Europe for 2005-2015. Countries with a decrease in rail share are colored in red, and with an increase in blue. Above: Accounting for the country where the freight vehicle is registered. These data are used in our global database and in Fig. 2.4 (with a CAGR for 2000- 2015). Below: Accounting for the country where the transport work is done. We see that much of the increases in road transport activity by Eastern European companies is performed in Central European countries, in particular Austria, Germany the Netherlands and Belgium. The data source is Eurostat.

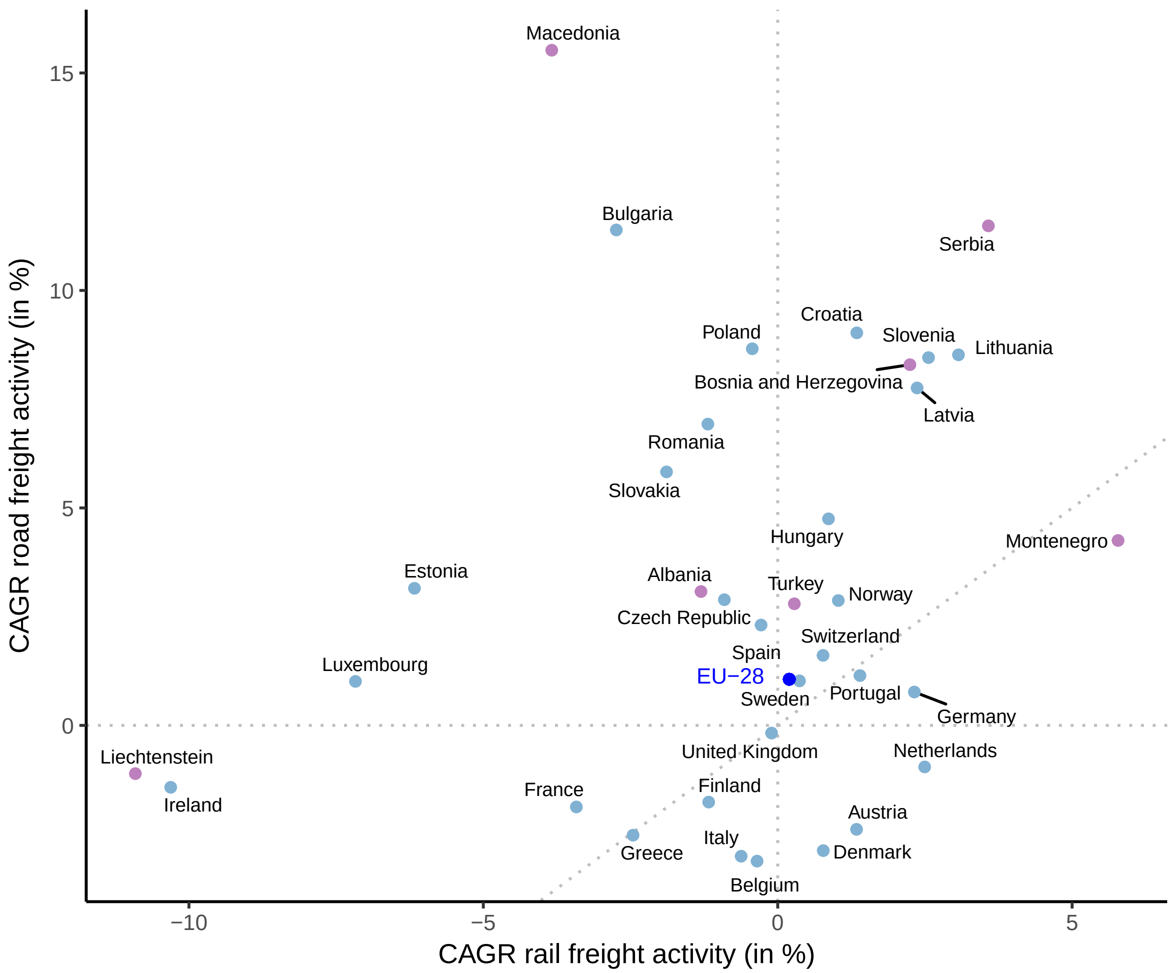


Figure 2.11: The CAGR of land freight activity in Europe since 2000. This graph illustrates the large growth in road freight activity by vehicles registered in Eastern European countries. Some Western European countries have seen a sharp decline in rail freight activity.

growth in the 90s and 2000s due to the expansion of the EU, and business and logistics changes that increased shipping distances [150]. In European countries, the road and rail freight modal split is 82:18 (80:20 in the EU). Taking the average over 36 European countries that provide data rail, road, and water freight through 2000-2014, the share of rail has decreased by around 11% compared to 2000 and the water share has remained approximately constant. A study for the European Parliament [212] notes that modal shares and their trends differ greatly between countries in Europe. Some of the newest member states (EU-13) have seen large growth rates in road freight activity ([212] and Fig. 2.10 and 2.11). In Fig. 2.10, we compare how the modal split of registration-based and territorial freight activity has changed from 2005 to 2015. Companies from Eastern Europe have taken a large share of road freight transportation in the EU, offering truck driver wages that are below the minimum wage in many Western European countries [229]. This gives the appearance of a shift to road in Eastern Europe, and a corresponding shift towards rail in Western Europe.

Studies estimate the current modal shift potential from road to rail in the EU between 1 and 14 percentage points, and a realistic target could be that rail transports 20% of all freight [212]. Other studies believe that European rail transport could grow by 10% to 30% under modal shift policies [166]. Barriers to increasing the share of rail freight include the need for investments to alleviate capacity bottlenecks, relatively poor reliability, and in some countries the higher cost of shipping by rail [212]

Select policy instruments in place

In a 2011 white paper on transportation [136], the EU set the goal of shifting 30% of long-distance road freight (over 300km) to rail or water by 2030, and 50% by 2050. This is to be achieved by improving the rail network, modal connectivity, and the quality of service to make multimodal freight economically attractive. The white paper mentions several EU-funded information systems to improve logistics and traffic management. As discussed in Section 2.4.3, EU-initiated projects are at the forefront of developing intelligent transport systems.

The EU Combined Transport Directive (Council Directive 92/106/EEC as of 1992) regulates intermodal transport of goods between EU member states and ensures for example that vehicle tax reductions for drayage vehicles are granted through member countries. The EU Commission recently proposed amendments to the Directive, reinforcing sustainability goals [137].

The Marco Polo Programme, which offered subsidies to modal shift projects, ran from 2003 to 2013 [215]. As of 2010, 30 billion tkm of freight have been shifted but the performance lagged behind expectations [264]. However, monitoring of the program's results has continued after 2010.

The EU is proactive in internalizing external costs of freight transportation [220, 218, 217]. European countries have higher diesel taxes than most other countries, equivalent to between 120 and 260 euro per tonne of CO_2 [222]. As of 2017, most EU countries have implemented some form of road charges for heavy vehicles [226, 197], which are harmonized by the Directive 1999/62/EC or *Eurovignette Directive* [265]. The tolls are intended both to recover infrastructure costs and to provide a means to induce modal shift [217]. Germany applies a heavy-vehicle toll to any truck with a gross weight of at least 7.5 tonnes using a GPS-based system [266]. Since 2005, this toll has been collected on all major highways (*Bundesautobahnen*) and since 2015 also on a number of secondary roads (*Bundesstraßen*), which were previously used to avoid tolls. The charge depends on the length of travel, the size of the vehicle and its emissions class, and it is between 8.1 and 21.8 euro cents per km, as of 2015. Belgium, Denmark, Luxembourg, the Netherlands and Sweden have a common truck tolling system, the *Eurovignette* system, which is based on time of use instead of distance [226]. Austria and Switzerland have particularly high road charges of up to US$0.50 per km or more [267, 268].

In addition, the EU is integrating modal shift strategies into the development of the Trans-European Transport Network (TEN-T) of core corridors [212, 264, 269]. EU-level and national efforts have led to a large number of new intermodal terminals [175]. The European railway system is primarily geared towards passenger service and rail infrastructure is mostly publicly owned [244]. The EU mandates non-discriminatory access to railway tracks [211]. To create a Single European Railway Area (SERA), the EU has created a standardized European Railway Traffic Management System (ERTMS), replacing existing national automatic train protection systems [270]. The EU's Horizon 2020 research and innovation program funds the Joint Undertaking (JU) *Shift2Rail* to promote the development of rail technology and modal shift strategies [271]. Through the COMCIS project, the EU is identifying and addressing problems of missing collaboration between modes by standardizing data systems and sharing. This includes adopting such concepts as synchromodality and the use of dry ports [272, 273].

Modal shift policies on the national level have been less successful than on the EU level, with the exception of policies targeting intermodal ports [212]. Many European ports only permit sustainable port development and include modal share targets, sometimes even mandated, into their infrastructure planning and terminal concessions [180].

Conclusions and policy recommendations

The EU has extensively promoted modal shift to decarbonize freight, but as the average share of road freight activity has increased, it is not apparent that these policies have been successful [135]. It is unclear whether the modal shift initiatives have been without effect or have potentially prevented a larger shift to road transportation. More research is needed in this area. The EU recently proposed further measures expanding information transparency and new investments in terminals [137].

2.5.4 India

In 2015, the Indian road to rail modal split was 69% to 31%. Road and rail freight activities have grown with a CAGR of 7.7% and 5.4% respectively between 2000 and 2015. This has led to a three-fold increase of road and a doubling of rail freight activity in that period. The Government of India (GoI) expects the total freight activity to grow at 9.7% per year until 2032 [274]. Given the current and forecast GDP growth rates of around 7% [275], GoI's forecast assumes a high elasticity of transport demand, as mentioned in [276], and may overstate the actually likely growth.

India's freight system suffers from congestion [210, 276]. The road freight industry operates at low costs, because entry barriers are low and competition is high, and because overloading is allowed to continue unchecked [274, 78]. It has benefited from recent improvements in the road network [78]. The Indian railway system is publicly owned and operated [277]. The volume of containerized transport on rail has grown considerably [278]. Rail freight tariffs are high as the government uses the revenue from freight movement to subsidize passenger tariffs [279, 280]. The condition of the rail lines and terminals is poor, which together with congestion causes uncertainties in the transit time [276, 279]. In 2012-13, more than 65% of India's rail freight activity was hauled by electric traction [281], but 58% of electricity is generated by coal [282]. Nonetheless, the carbon intensity of rail has fallen significantly since 2000 [283].

Indian inland waterways are underdeveloped [274]. Data on domestic water freight activity are scarce, but it is estimated as significantly less than rail freight activity, e.g. 6% in 2007-08. The vast majority is coastal shipping [210, 276, 283].

Contrary to current growth trends, the GoI aims to achieve equal shares for road and rail by 2032 [274]. As early as 1980, the GoI set an ambitious target of transporting 70% of freight by rail and 30% by road by 2000. It has fallen far short of these goals [280]. Infrastructure capacity shortfalls and poor integration of transport modes are identified as the main barriers [210].

Select policy instruments in place

To achieve its stated target of a rail share of 50% by 2031-32, the GoI aims to modernize the freight system and expand capacity [210, 274]. As much of the infrastructure is yet to be built, long-term planning that accounts for the need to minimize GHG emissions could avoid the construction of an inefficient and carbon-intensive system. The GoI currently invests around 4% of GDP in infrastructure but only a tenth of that in the rail sector [274]. The GoI has initiated large multimodal infrastructure projects [274] and is building Dedicated Freight Corridors (DFC) [274, 276, 284, 285] in an attempt to expand rail capacity. Some of these projects are built with international financial assistance [276, 278, 284, 285] and the use of public-private partnerships (PPPs) [286]. India has more than 50 dry ports for containerized transport [278]. Containerized freight is handled by Container Corporation of India (CONCOR), which is part of Indian Railways and has invested in modernizing its operations with ICT; for example to track containers [278].

In addition to the initiatives for improving rail and road networks, the GoI plans to subsidize water shipments [210] and to develop India's first modern inland waterway on the Ganges river, which flows through the densely populated Indo-Gangetic plain [287]. Approximately 40% of India's traded goods make their way through this region at some point in their journey [287], many of which are trucked to ports on the country's west coast. The World Bank has financed the project with a loan of $375 Million [288].

Fuel taxes are collected by central and state governments with considerable variation between states [289]. Efforts to reduce air pollution and congestion by heavy-vehicle driving restrictions (e.g., a nationwide ban on trucks older than 15 years), fuel emission standards, and green taxes on vehicles that enter urban areas have been adopted in parts of India [289]. These may promote modal shift as well as help to decarbonize the last mile.

Conclusion and policy recommendations

As level of service factors [280] and capacity bottlenecks [210] are the most important barriers to modal shift, the share of infrastructure investments in rail needs to be increased [274]. The United Nations Environment Programme (UNEP) estimates that the planned DFCs can avoid large amounts of GHG emission, mainly by preventing a shift from rail to road in the growing freight sector. These effects could be increased by a carbon pricing strategy [276]. We found little evidence of the use of cost incentives to promote modal shift and adjusted price signals that are important for optimal infrastructure planning [290]. The GoI acknowledges that Indian subsidies, tariffs and taxation policies send distorted price signals [274]. Heavy vehicle tolls have been recommended as a policy to promote modal shift and could be collected electronically [274, 280].

The Goods and Services Tax (GST), a national value-added tax introduced in 2017, eliminates state taxes in some sectors. The new regime is expected to lower the cost of freight, although it is unclear whether it will benefit one mode more than others [291]. Early reports suggest that by eliminating the need to stop at tax checkpoints every time a truck crossed state boundaries, GST has cut the travel time on road by about 25% [292], potentially making road more attractive. A retrospective analysis of the impact of GST on modal choices would be interesting once the tax regime has stabilized and taken root.

Coal is the commodity with the largest freight activity in India, and almost 80% of it is transported by rail [293], which is already contributing to congestion of the rail network. As demand for coal and for the transport of it increases, it may crowd out other goods if infrastructure does not expand rapidly enough [294], and may be increasingly transported by road [295].

2.5.5 EECCA Countries and the Trans-Asian Rail Network

Rail is the dominant mode for Eastern Europe, Caucasus and Central Asia (EECCA) and Mongolia, while inland water transport is negligible (Fig. 2.8). Russia has 90% rail freight activity. Except for in Russia, the road freight sector has grown faster than the rail sector in the region. A number of EECCA countries are landlocked nations, where high transport costs hinder trade and economic development. Upgrading the transport infrastructure and facilitating regional cooperation between transit neighboring countries deserve high priority in this region [296, 297, 298]. To automate and accelerate clearing processes at border crossings, ICT is increasingly used [210]. In addition, the development of new transit corridors is important to prevent bottlenecks [296, 298]. For example in Kazakhstan, there is hope that new corridor developments on the Eurasian land bridge will stimulate the logistics system [299]. In the last five years, Kazakhstan has invested several billion dollars to upgrade its railway infrastructure [300].

While the vast majority of intercontinental freight is transported by ship, the land route between China and Europe is only about half as long as the sea route [300] and can be traveled by rail in much shorter time [157, 300]. First the Silk Road and then the Trans-Siberian Railway were important trade routes in the region, but after the fall of the Soviet Union this corridor has only carried small shares of the freight [300]. However, with the recent increase in trade between China and Europe and the Chinese *Belt and Road Initiative*, the rail connections between Asia and Europe promise to gain importance. In 2016, the rail freight volume between China and Europe was already at 500 thousand tonnes, equivalent to one third of airfreight tonnage traveling to Europe [300]. After raw material and machinery parts, the most important commodities transported along these routes are automotive, high-tech and consumer goods [300].

In the 1960s, the international community expressed interest in a joint railway network for Asia [301] and since 1992 the United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) has worked towards creating the *Trans-Asian Railway Network*. Member states negotiated an Intergovernmental Agreement in 2009 that has resulted in new railway links but development has been slow. Cross-border rail links are fragmented, especially in South and South-West Asia, causing shippers to prefer maritime freight transport [210]. To promote intermodal transportation in the region, UNESCAP adopted the Intergovernmental Agreement on Dry Ports in 2013 [302].

There are many gauge breaks in the Trans-Asian Rail Network that require additional handling [157], further complicating the multiple border crossings and slowing transportation down. Additional challenges for the routes are the extreme temperatures in those regions of continental climate, damage to goods and trains due to poor rail tracks, and theft [303, 157]. Temperature sensitive products are mostly only shippable in spring and fall. A test shipment bringing wine from Europe to China in May 2017 revealed large temperature fluctuations inside the container, which could be mitigated by a special container foil liner [304]. Another difficulty could be the trade imbalance between the two regions. Already in 2016, the amount of freight from China to Europe on this route was 1.5 times the reverse [300].

2.5.6 Africa

African countries have experienced strong growth in freight transportation but from a low base [150]. The share of intraregional trade between African countries is small, but it is expected to grow significantly [232, 305, 306]. Inland surface transport constitutes a large portion of the cost of exported goods, with cartels and regulatory institutions distorting prices [78]. The African Development Bank (AfDB) estimates that more than 80% of freight in Africa is carried by road [172]. Africa has experienced a shift from rail to truck and many rail companies have struggled or failed [307, 173]. Fig. 2.12 shows available data on changes in rail freight transportation in African countries between 2000 and 2016. In sub-Saharan Africa, railways are the least developed transportation mode, are often configured for the export of raw materials (a legacy of their colonial origins), and poorly maintained [172]. Railways in Northern Africa and the Republic of South Africa are generally in a better state [232]. While some countries are improving the operational efficiency of their railways [172], investments in new infrastructure need to be combined with policies and institutions that encourage a shift from road to rail [232]. The AfDB has concluded that there is the

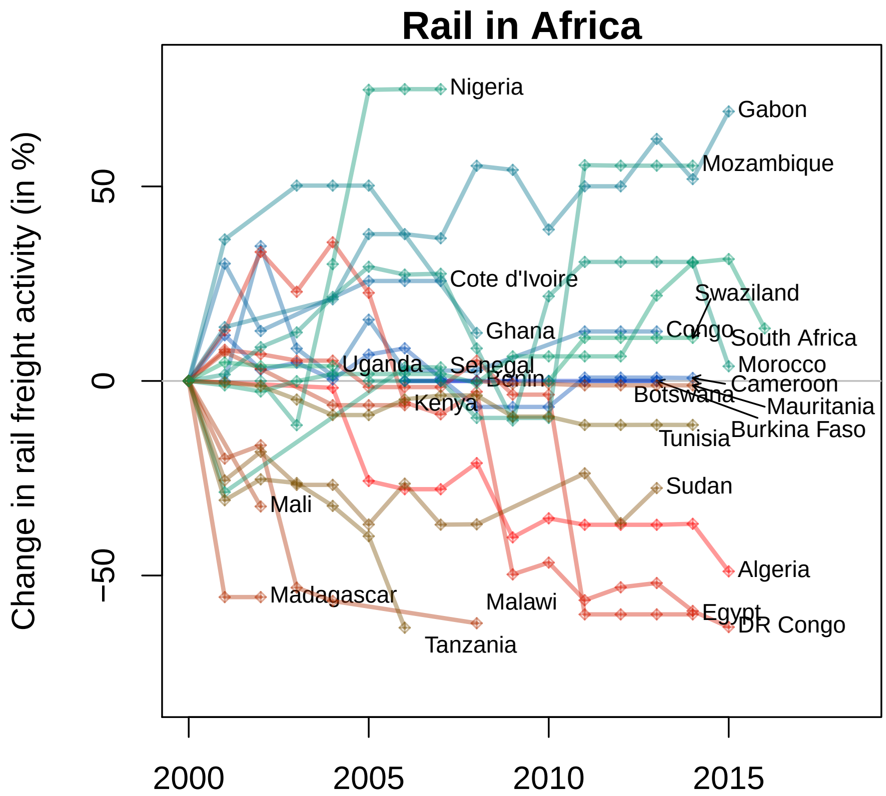


Figure 2.12: Development of rail freight activity in Africa with respect to a base year (2000 or first available data point). Countries with a decrease are in red, and with an increase in blue. Due to the lack of data on road freight transportation, we cannot display the development of modal shares. Only countries with more than one data point are shown.

need for a new approach to strengthen African railways, so as to accommodate recent growth in the economy, population, urbanization, and mining activity [232].

Modal shift in Morocco, the Northern Corridor, and South Africa

Morocco has a national strategy to green the supply chain, in particular by channeling freight transportation in multimodal corridors for more efficient and low-carbon transportation [308]. Since 2000, new rail lines with multimodal nodes, double tracks and dry ports have been constructed [232]. Railways are operated by the national company ONCF and the largest commodity is bulk phosphate [232].

The Northern Corridor in Sub-Saharan Africa links the port of Mombasa with destinations in Burundi, the Democratic Republic of Congo, Kenya, Rwanda, Uganda and South Sudan. The Green Freight Programme of the Northern Corridor Transit and Transport Coordination Authority assists freight operators with collaboration, and overseas infrastructure improvements for better intermodal connectivity, which are planned for 2016 to 2021 [309]. In 2017, Kenya opened a new standard gauge line from Nairobi to Mombasa to replace the century-old 1,067mm gauge railway. It is financed by a $4 billion Chinese loan [310]. Currently, more than 95% of the freight between the port of Mombasa and Nairobi is carried by road [307, 173], but the railway is designed to carry 40% of the projected freight in 2035 [311]. There are concerns that passages of single track and load constraints will make the new rail option as uncompetitive as the old one [307]. Maintenance will be crucial [173], as will policies to boost rail over trucks.

The South African transportation sector is well-developed, and rail has a high share of the freight activity of approximately 40% in 2016. The publicly owned railway company Transnet operates most freight rail [232]. Public institutions do not collect data on road freight activity but some data are available through Stellenbosch University [312, 313]. Havenga and Simpson (2016) [313] estimate a great modal shift potential in the country due to the long distances that bulk commodities are transported. Such a shift could be stimulated by charges that internalize the external costs of road transport.

Conclusion and policy recommendation

Africa largely exports bulk commodities, such as minerals, which are very suitable for rail [172, 232]. There is, however, a need to develop long-distance rail transportation corridors and better cross-border transportation systems [232, 306]. This will serve the many landlocked countries in Africa and foster the envisioned increase in intra-African trade [305]. The benefits of the introduction of a new gauge, as for example in Kenya, are questionable since gauge breaks can pose an obstacle [232]. International institutions are well-suited to play a significant role for providing capital for infrastructure investments, and South-South transfer of experiences with private-public partnerships can be valuable [232].

In addition, an important first step to give rail freight a competitive edge over trucks without undermining economic growth is to introduce and enforce regulations in the road freight sector [232]. Fuel taxes that could be used to fund railway infrastructure have also been recommended [232].

2.5.7 ASEAN and East Asia

Trade in Asia and the Pacific is growing fast, and intraregional trade has already become more important than trade with overseas markets [210]. The developments provide a need for boosting the freight infrastructure [147, 210] and for a regional intermodal policy [213].

Road transport dominates the land freight sector, but coastal and inter-island shipping have a large share in the region (Fig. 8 in the main text and [108, 210]). Japan relies more on light-duty vehicles for freight than other regions [10]. The share of rail freight is small and some ASEAN countries do not have railways at all. While the East Asian countries Japan and South Korea have a highly developed rail system, even here less than 10% of freight is transported on rail. With the Tohoku earthquake and tsunami in 2011 and the subsequent reduction of nuclear power in the electricity mix, the carbon intensity in the mostly electrified rail sector in Japan has strongly increased [283].

Many ASEAN countries do not report freight activity data, with notable exceptions, for example the General Statistics Office of Vietnam. Since the motorcycle is an important mode of transportation in Vietnam, those are also included in the freight surveys [154]. The Asian Development Bank is in the process of publishing a transportation data bank for the region [314].

There is little information on freight policies in the regions but there are indications that many countries have set modal share targets and have emphasized the importance of intermodal [213, 315, 316]. There is a need for stronger policy commitments such as to develop more dry ports in the region [210, 278] and improve rail systems. For example, Thailand plans to connect ports with new rail links [210]. Regional studies have found some potential for modal shift [108, 317, 318]. In an international comparison, fuel taxes in Japan and Korea are in the medium range [222].

As the largest economy in the region, Japan targeted a 50% share of rail and coastal shipping in 2010 [213], which it missed by only 5 percentage points, but nearly accomplished in 2014. More than 44% of freight was transported by coastal shipping that year and only around 5% was on rail. Japan is one of the few countries in our dataset that have experienced a modal shift, mainly because road freight activity has decreased by more than 30% since 2000. The total value of foreign trade has not decreased in this time frame [319]. Close monitoring of the targets, standardization, regulatory reforms, the privatization of the railway sector and improvements in logistics management could be factors that have contributed to Japan's modal shift success [10].

2.5.8 Latin America and the Caribbean

Data is scarce for Latin America, in particular on coastal shipping. By weight, maritime transportation is the dominant mode in the region [320]. In those countries that report freight activity data, most land transportation is by road, but the share of rail transport has increased (Fig. 8, main text). In average, less than 20% of the volume is intraregional trade [320]. The logistics performance of the region is low and transportation costs are high [321]. Freight transport in Latin America is expensive due to issues with technology standards, empty backhauls, infrastructure development and crime [78, 320, 322]. There is a pronounced need to develop a more efficient and multi-modal freight system [320]. Latin America's need for rail infrastructure has attracted foreign investment but ambitious railway projects face obstacles [323] or are abandoned [324].

Mexico has made a number of improvements in the rail sector that created a flourishing intermodal corridor and increased the railway's share of national freight [325, 326]. Mexico has extremely low fuel taxes [325], but has recently removed its gasoline and diesel subsidies [223]. In contrast to many other South-American countries, Argentina, Chile and Brazil have a road freight sector that meets international standards [78]. Brazil is the country with the largest freight activity in the region. While most of its freight is transported on road, coastal shipping and inland waterways have a large share that is growing fast [327]. In Chile, almost all freight is transported by road. Due to long transport distances, the country would benefit from a more competitive freight rail system. The average length of rail haul is short (around $150km$), connecting mines with ports [328]. Chile has the ambitious goal of almost doubling the rail freight volume from 2012 by the end of the decade [329]. Due to its geography, coastal shipping plays an important role in Chile and a *maritime highway* along the coast could shift cargo from the road to coastal shipping [330].

2.5.9 Middle East

Data on surface freight transportation in the Middle East are scarce. While there is information available on rail freight in Iraq, Jordan, Syria and Saudi Arabia, our dataset does not cover any road freight activity in the region. The country with the largest fuel consumption in the transportation sector is Saudi Arabia, followed by Iraq and the United Arab Emirates (UAE) [331]. Many countries in the region do not have railway systems. The UAE are currently building a 1200km rail network [332] that has partially begun operations in 2015. This rail system currently transports almost 500 thousand tonnes of sulfur monthly [333].

2.5.10 Oceania

Oceania's total freight activity is dominated by Australia. The positive development of the rail share in the region (Fig. 8, main text) was largely driven by the growth of bulk transports in Australia, in particular by iron ore in the Pilbara region [146]. In Australia, almost all freight consists of bulk products that are transported over long distances [334], mostly iron ore and coal [146]. Rail carried 64% of land freight in 2015 and road freight accounted for most of the non-bulk goods [146]. Coastal shipping plays a significant role but the total amount shipped on water has not increased in recent years. Australia permits very long and heavy vehicles, e.g. *road trains* and *B-doubles*, which carried nearly 70% of road freight in 2007 [335]. The Australian rail network uses four different gauges and mainly connects cities to the coast. The fragmented rail sector, low reliability, and competitively large road vehicles are regarded as barriers to shifting more non-bulk goods to rail [230].

Freight activity in New Zealand is dominated by the road sector, which is growing. Studies in New Zealand have found a large potential for intermodal [336, 337], even though quality of service is a barrier for shippers [337]. New Zealand does not have a diesel tax, but applies road charges to diesel vehicles as well as all heavy vehicles [222, 338].

Other countries in Oceania are smaller island states that rely heavily on coastal and inter-island shipping [109]. Here, energy efficiency and alternative propulsion options for ships are more applicable than modal shift.

2.5.11 South Asia

Road freight data for South Asian countries (without India) are limited to Pakistan and Bangladesh, and at least as old as 2005. Rail activity is also reported for Iran. Overloading and goods damage in the road freight sector are frequent in Pakistan and Bangladesh [339, 78, 316].

More than 95% of the freight is carried by road in Pakistan. The road freight sector is fragmented, with only 1% of companies having a fleet size of $>$100. Freight rates are extremely low. While rail was the dominant mode in the 1950s, the freight volume has drastically declined due to low infrastructure investment rates, prioritizing of passenger service, and long delivery times [339]. As the average trip distance in Pakistan is long and the country is a natural corridor for transit trade, in particular with China, the potential for modal shift to rail intermodal is high [339].

Road freight activity in Bangladesh has grown faster than the GDP, while inland water and rail could not keep pace, as the government had increasingly favored investments in roads instead of in the two other modes. Inland waterways are much more important for freight than rail and provide access to remote areas in Bangladesh. The government of Bangladesh has begun to put policies in place to promote green freight and intermodal transportation [316].

2.6 Estimating the modal shift potential

There is a clear need for a systematic assessment of the global modal shift potential and its associated GHG emission reductions and costs. While such a comprehensive assessment does not exist, regional estimates provide a starting point on how to think about this issue. Brogan et al. [13] identify three approaches to estimate the modal shift potential: market-segmentation methods, modal cross-elasticities, and mode choice models.

Market-segmentation methods are frequently used given their simplicity. They analyze origin-destination pairs of freight shipments to identify the fraction of shipments that could potentially be transported by each mode. These methods are capable of estimating a maximum feasible modal split under ideal conditions, but do not take into account capacity constraints and modal shift policies. For example, Zhou et al. [249] have used the U.S. Commodity Flow Survey to estimate how many shipments could be shifted based on average freight tonnage, transport distance, and time sensitivity and made assumptions about shipper preferences for shifting to rail. Models that use commodity flows and pricing information to estimate modal elasticities and cross-elasticities are often proprietary [13]. Modal elasticities can be used for policy analysis or as model parameters to predict future modal shares. The range of cross-elasticities found in the literature is wide and values are location-specific [162, 163, 82] (Section 2.4.1), thus making it difficult to use the values for reliable predictions. Mode choice models survey the preferences of individual shippers, which can be used to estimate modal elasticities and also to predict modal shift. Often those studies use discrete choice models, e.g. [160], and data collection might be costly [13].

There are a few estimates of regional modal shift potentials and targets for modal shift in the literature, which we reproduced in Table 2.1. These are expressed for example as target modal shares, or as changes in the share of one mode or the other with respect to an explicit or implicit baseline. We find that some estimates and targets are poorly defined: for example, they do not say if the share is expressed as a percentage of total land transportation or of total surface transportation including water modes. A number of the targets in Table 2.1 are expressed relative to counterfactual, business-as-usual baselines. Unless the current and projected baselines are explicitly stated, it is difficult to interpret such targets, and to quantify their consequences for greenhouse gas emissions reductions. In some cases, even if the current baseline is clearly stated, we find that it disagrees with independent estimates of the current level of activity and modal shares. For example, data by the International Council on Clean Transportation (ICCT) suggest that in 2010 the share of rail in Chinese freight activity was $> 60\%$ [227], whereas OECD data suggest that it was $< 40\%$ [69].

Table 2.1: Estimates of modal shift potential and political targets.

Source	Nature of estimate	Regions	Estimate
ICCT (2012) [227]	Assumption	United States, China, Japan, Canada, South Korea, Australia	20% increase in rail freight activity by 2030 compared to a 2010 baseline
ICCT (2012) [227]	Assumption	EU, India, Brazil, Mexico	40% increase in rail freight activity by 2030 compared to a 2010 baseline
ICCT (2012) [227]	Assumption	North America, Australia, Brazil, Japan, South Korea	Share of rail freight will remain unchanged between 2010 and 2030 given currently planned policies; large potential increase with incentives
ICCT (2012) [227]	Assumption	EU	Share of rail freight will decrease between 2010 and 2030 given currently planned policies; small potential increase with incentives
ICCT (2012) [227]	Assumption	China, India	Share of rail freight will fall by more than 10 percentage points between 2010 and 2030, decrease could be substantially reduced by infrastructure investments
IEA (2009) [12]	Scenario	Global	'BLUE' Scenario: all measures including modal shift result in 15% lower carbon emissions than in the baseline scenario for 2050
IEA ETP (2015) [169]	Scenario	Global	Avoid and Shift policies in the '2DS' scenario reduce transport GHG emissions by 15% or more by 2050
UIC [283]	Target	Global	Rail same share as road by 2030 and 50% greater than road by 2050
Zhou et al. [249]	Estimate	United States	4.1% of road freight activity can be shifted to rail, with 4.4% reduction in GHG emissions compared to base case scenario
European Commission [136]	Target	European Union	Shift 30% of road traffic of $> 300km$ to rail or water by 2030 and 50% by 2050
Studies in Tavasszy et al. [166]	Estimate	European Union	Rail transport could grow 10% to 30%; modal shift potential from road to rail of 1 to 14 percentage points under modal shift policies
Dionori et al. [212]	Estimate	European Union	Rail could realistically transport 20% of all freight activity
Government of India [274]	Target	India	50:50 share of road and rail by 2031-32
Havenga and Simpson [313]	Estimate	South Africa	21% of the transport activity (15% of the tonnage) could be shifted to rail if externalities were considered
Government of Japan [213]	Target	Japan	50% share of rail and water by 2010 (acheived 2014)
Subsecretaría de Transportes Chile [329]	Target	Chile	Double rail freight activity by 2020 based on 2012

2.7 Conclusion

Cost-effective GHG emissions reductions for the transportation sector may be available but in today's markets will likely not lead to the levels of decarbonization that are needed to slow climate change. Thus, additional policies that include either incentives for reductions or penalties for GHG emissions will be needed. We find that modal shift may have the potential to reduce GHG emissions, but that a systematic analysis of the possible emissions reductions and costs is yet to be found in the literature.

To promote modal shift, governments could use two types of policy approaches: infrastructure investments and incentives. However, policy approaches have been largely underused. For example, the share of infrastructure investments in the rail sector is small and should be increased, especially in developing countries. Investments should focus on constructing efficient rail and intermodal terminals and facilitating the use of ICTs for example to track shipments and expedite routing. In low-income countries, a freight system with multimodal and low-carbon infrastructure can be a cost-competitive way to promote economic growth. Incentives should discourage the use of road freight by pricing mechanisms such as tolls, fuel or vehicle taxes, or through tighter regulation. At a minimum, diesel taxes should be as high as gasoline taxes. Some countries might also consider subsidies and R&D programs to promote intermodal freight.

Infrastructure investments are effective approaches to encourage modal shift in less developed freight markets but policies targeting the internalization of external cost should be preferred in well-developed markets in order to maximize welfare [175]. Combining multiple policies (policy packaging) has been emphasized as a valuable approach [217, 340], in particular the combination of cost increases for road transport with decreases in lead time of intermodal transport [263].

The lack of standardized, high-quality data in the freight sector limits informed policy analysis, formulation and validation. Standardized data could be collected using the Common Questionnaire for Inland Transport Statistics by the United Nations Economic Commission for Europe (UNECE), the International Transport Forum (ITF) and Eurostat, which is currently used by around 60 countries [153, 341]. The global potential for reducing GHG emissions through modal shift remains unknown, although there are a wide variety of national and regional estimates, targets and assumptions. A rigorously produced, and economically and politically realistic estimate would be valuable. Also missing from the literature is a marginal GHG abatement cost curve for freight that captures mode shifting and intermodal potentials, and that also accounts for ICT strategies. In addition, there are potentially adverse interactions of other decarbonization strategies with modal shift that need to be studied, such as the effect of more efficient and therefore cheaper trucking. It would also be instructive to examine potential disruptive changes in commodity demands that can result in stranded infrastructure assets, change freight market conditions but also free up rail or water freight capacity. This is particularly relevant as many countries shift away from coal for electricity generation.

3

Truck traffic monitoring with satellite images

The road freight sector is responsible for a large and growing share of greenhouse gas emissions, but reliable data on the amount of freight that is moved on roads are scarce. Many low- and middle-income countries have limited ground-based traffic monitoring and freight surveying activities. We show that we can use an object detection network to count trucks in satellite images and predict average daily truck traffic from those counts. In this proof of concept, we describe a complete model, test the uncertainty of the estimation, and discuss the transfer to developing countries.

3.1 Introduction

As noted by the United Nations, despite an exponential growth in the availability of data in recent decades, many people and critical aspects of their lives and environment remain unmeasured [151]. Especially across the developing world, a key barrier to identifying opportunities for mitigating climate change is the lack of sufficiently granular, high-quality data. Heavy- and medium-duty trucking accounts for 7% of total world energy-related CO_2 emissions [10], with much of the growth occurring in developing countries [342]. In order to successfully implement policies and make targeted investments, reliable data about the volume of freight that is moved on roads is crucial. More than half of all countries do not collect national road freight activity data and where estimates exist, they are typically survey-based and often inadequate [342]. Knowing truck movements is also important for a variety of economic analyses and for road maintenance planning, even if only based on short-duration counts [343], but such ground-based traffic monitoring is costly and not performed in many countries.

In this chapter, we propose a remote sensing approach to obtain vehicle counts from high-resolution satellite images. As satellite images become both cheaper and are taken at a higher resolution over time, we anticipate that our proposed approach will become scalable at an affordable cost within the next few years to much larger geographic regions. We take advantage of recent advances in deep convolutional neural networks for object detection. These methods have already been successfully applied to detecting vehicles in satellite images [344, 345, 346, 347, 348]. Most work has focused on cars, and to a lesser extent on multiple vehicle classes including trucks [349, 344]. Note that a satellite image is only for a single snapshot in time, whereas conventional traffic estimates are taken over a much longer period of time. Thus, our approach separately models how traffic changes with time.

We begin by providing a brief overview of traditional ground-based traffic monitoring and remote sensing alternatives (Section 3.2). We then introduce our framework, which consists of a truck detection model (Section 3.3.1) and a temporal traffic monitoring model (Section 3.3.2). We validate and test our approach using data from the New York Thruway (Section 3.4) and assess how the model transfers to data from Brazil (Section 3.5). We conclude with a qualitative discussion of how well the model translates to developing countries and outline future work (Section 3.6).

3.2 Traffic monitoring and freight surveying

The US Federal Highway Administration (FHWA) highlights the importance of vehicle counting for traffic monitoring, as counting provides statistics such as the Annual Average Daily Truck Traffic (AADTT) [343]. Ground-based automatic vehicle counting devices include pneumatic tubes, inductive loop detectors, magnetic sensors, video detection systems, and several others. Installation and maintenance for some of these systems requires pavement cuts and lane closures. Traffic monitoring is usually based on continuous counts, which also provide the basis for periodic (e.g., hour of the day) factors applied to short duration counts. Typical short duration detection periods are between 24 hours and a week long [343].

Traffic monitoring with remote sensing As ground-based detection devices can be prone to failure and are too costly to install and maintain in some countries, there is a need for alternative monitoring technologies, such as through GPS data from cell phones [350] or with aerial or high-resolution satellite images. Even lower-resolution satellite images can provide sufficient resolution [351, 352]. There is also potential for using drones [353]. With remote sensing, a large number of roads can be covered at the same instance, many of which are not equipped with costly sensors [351, 352] (e.g., rural or remote roads). Also, areas that are difficult to access, for example due to a disaster or conflict, could be monitored [354]. A weakness of the method is that traffic fluctuations on short time scales as well as time-of-day, day-of-week, and seasonal traffic patterns can distort the accuracy of the estimate of the AADTT [343]. Images are only useful during daylight and under cloud-free conditions. In addition, this method requires advanced analytical and computational resources. The uptake of remote sensing methods for transportation applications has been slow but it promises to offer cost-effective and scalable options for a multitude of applications [355, 356].

Freight surveying Data on road freight activity, measured in tonne-km, are typically obtained through national surveys of shipping companies, which need to provide information on origin, destination, weight, and other indicators of all shipments [342]. As the road freight sector is fragmented with most companies operating very few trucks, this approach can be costly and relies on high compliance rates. Less than half of the countries in the world collect this type of information [342].

3.3 Problem setup

Our framework consists of a truck detection model and freight monitoring model. The detection model counts the number of freight vehicles on roadways in a satellite image, and the monitoring model translates these counts into the AADTT.

3.3.1 Detection model

Object detector The object detection model provides the vehicle count from an image. Huang et al. identified three object detection meta-architectures, which are Faster Region-based Convolutional Neural Networks (Faster R-CNN), Single Shot Detectors (SSD) and Region-based Fully Convolutional Networks (R-FCN) [357]. They have tested models based on these meta-architectures for speed and accuracy, and have found that Faster R-CNN often achieves the highest accuracy, while SSD excels in speed. We compare Faster R-CNN [358] with 50- and 101-layer Resnets [359] and SSD Inception V2 [361] for our application. Faster R-CNN first proposes regions with the Region Proposal Network (RPN) and then uses the Fast R-CNN detector [360] for object detection, sharing convolutional layers. While Faster R-CNN first classifies the objectiveness of proposed boxes, and then predicts the class in another network, SSD directly classifies and regresses boxes, which makes it much faster to train and perform inference. We use the default implementations for the COCO image dataset from the Tensorflow Object Detection API [357] and pre-trained convolutional layers.

Road filter We only want to count trucks that are driving on the road that we are interested in, and exclude those that are sitting in parking lots or traveling on other nearby roads. To filter out irrelevant predictions from the detection model, we use geospatial data. Those data are ubiquitous, and also available for main transit highways in developing countries. We count a truck if at least one corner of its bounding box is within a certain distance of the center of the road. If both lanes are indicated, we set this distance to 8 meters, which approximately accommodates a four-lane highway. This filter is applied to both the ground truth validation and test datasets and the predictions from the model.

3.3.2 Monitoring model

To use a snapshot image to approximate ground-based vehicle counts, we assume that all c_I vehicles travel with a constant speed within the interval captured by the image. From that we infer the time t_I that it takes for a vehicle to travel from the start to the end point in the interval. A detector installed in the end point should count c_I vehicles in t_I. The FHWA recommends that traffic density variation factors $f_{h,d,m}$ be applied when using less-than-a-day counts to compute the AADTT [343], so as to account for time-of-day, day-of-the-week and monthly variations. We can approximate the average daily (bidirectional) counts as $AADTT \approx c_I \cdot \frac{24h}{t_I} \cdot f_{h,d,m}^{-1}$. Detailed information about traffic patterns, and access to satellite images taken at different times for the same location, can reduce the error of the estimate. Here, we assume that no information about traffic variation in the test region is given, and we need to approximate the factors and their uncertainty from regions where truck traffic is monitored. For this we use linear regression on data from several regions in distinct states or countries. We use a Monte Carlo method to incorporate the uncertainty of speed and time variation factors. By making assumptions about the distribution of payloads of the freight vehicles [343], one could use this approach to further estimate the freight activity through truck counts.

3.4 Experiments

We validate and test on images and toll data obtained for sections of the NY Thruway. We have two kinds of ground-truth data: the true labels in the satellite images and ground-based counts. Both of the submodels are independent and we validate them separately to choose the best model specifications and parameters. We then test how well the whole model can estimate the AADTT on held-out sections of the Thruway.

3.4.1 Data

We curated our own collection of 31cm-resolution, RGB-color satellite images provided by Digital-Globe, Inc., (Appendix 7.1.2), since a large satellite image database ("xView" [362]) with several thousand labeled truck instances proved too inaccurate and other satellite image datasets contained only small numbers of trucks [348, 363]. For training, we used images of several regions in the Northeastern US, primarily the NY Thruway, with a total of 2050 truck examples. For validation, we worked with images from 3 sections of the Thruway, some partially covered by fog, that contain 216 truck examples (81 on road). For the road filter, we used a shapefile of the Thruway provided by the State of New York [364].

For model selection and to train the factors in the monitoring model, we used hourly ground-based counts for four regions, namely the NY Thruway [366, 365], California [?], Brazil [368], and Germany [?]. While the first are toll data, the latter three are datasets from short-term and continuous counters. Refer to Appendix 7.1.3 for a description of sources and data preparation. We used the toll data from the NY Thruway as the ground truth for testing the results. For the speed, we assumed a mean of 65 mi/hr.

3.4.2 Detection model

Training We trained three different object detection models on our training dataset of ~ 2000 truck images from Northeastern US. The model specifications are:

1. Faster R-CNN Resnet 50, convolutional layers pre-trained on COCO image data, 10,906 iterations.

2. Faster R-CNN Resnet 101, convolutional layers pre-trained on COCO image data, 25,318 iterations.

3. Single Shot Detector with Inception V2, pre-trained on COCO image data, 16,153 iterations.

A model that was not pre-trained on COCO data proved difficult to train. All models were trained over several days, and early stopping was applied based on a small validation portion of the training dataset.

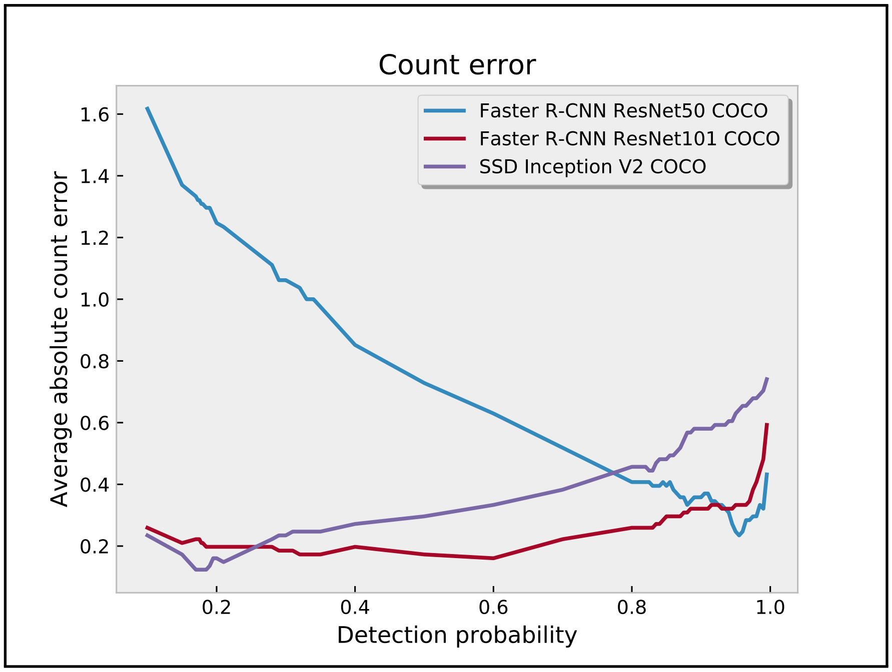


Figure 3.1: Absolute error of total truck counts, which also includes false positives, over detection probability. We see that the SSD achieves lowest count error.

Validation We determined an optimal threshold for the prediction probability in order to minimize the prediction error of total truck counts in an image (which can include true positives and false positives). We compute the mean absolute count error over all validation images as the weighted sum of the relative absolute count error of each of N validation images

$$\epsilon_{Count} = \frac{\sum_{i=1}^{N} |c_{pred}^{(i)} - c_{true}^{(i)}|}{\sum_{i=1}^{N} c_{true}^{(i)}}, \tag{3.1}$$

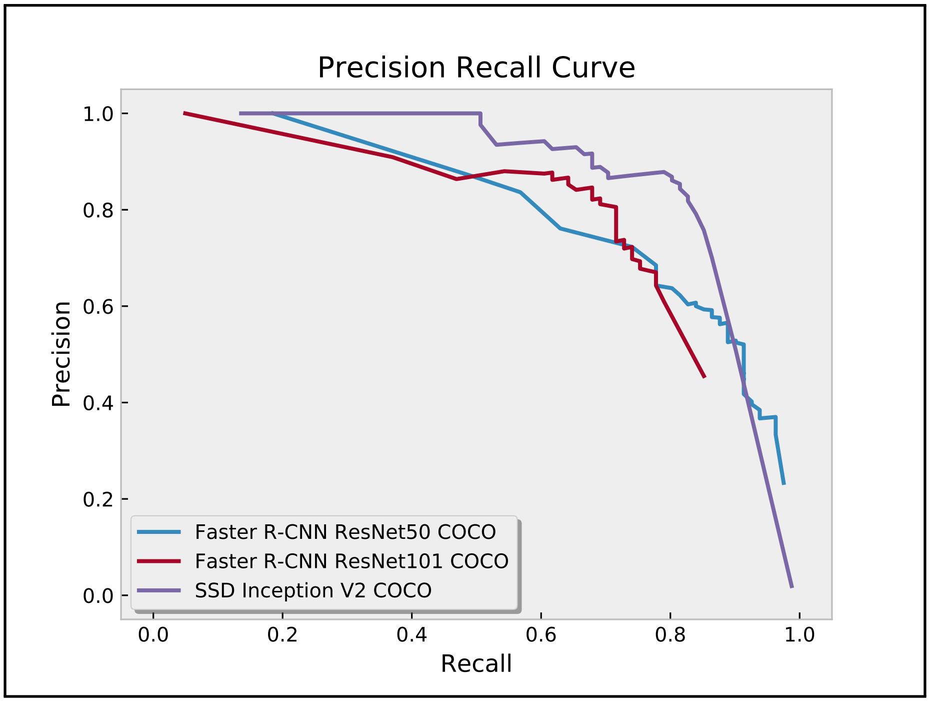
where c is the number of trucks counted in image i. While the mean absolute count error is often used [413], we chose to use a weighted sum to account for the relative importance of images, as images contain a diverse number of truck examples. The measure does not have guarantees to be convex and might not have a unique minimum due to the discrete nature of counting. Fig. 3.1 shows

Table 3.1: Performance for optimal count prediction probability; pre-trained on COCO and fine-tuned on $\sim$ 2000 trucks.

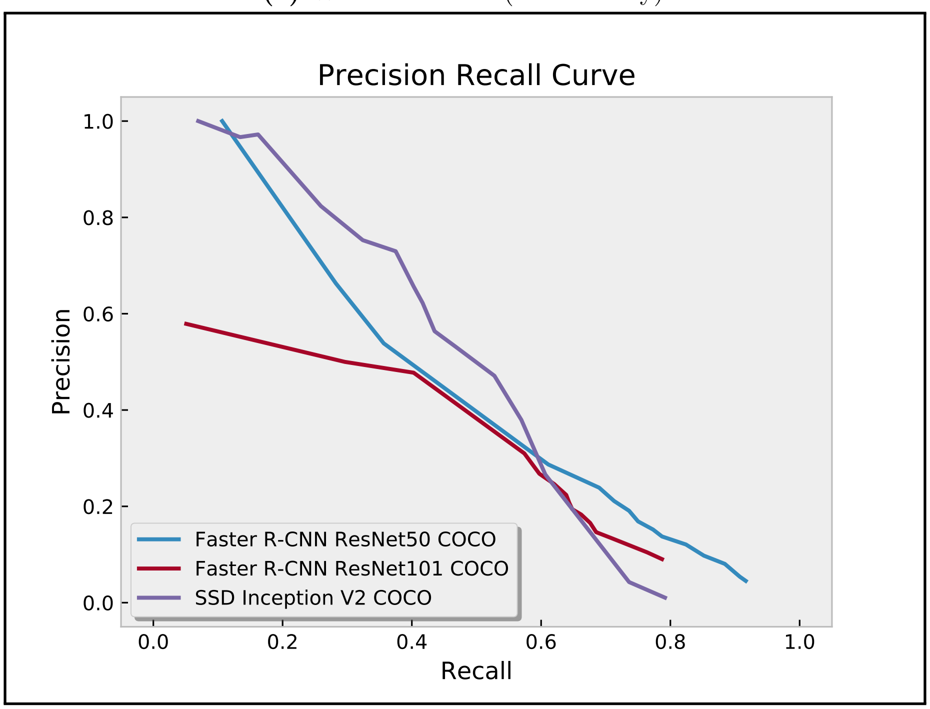
Faster R-CNN	Av. Count Err.	p_{pred}	Av. Prec.	Av. Rec.
With road filter				
ResNet50	0.235	0.955	0.741	0.714
ResNet101	0.160	0.600	0.716	0.806
SSD Inception V2	0.123	0.175	0.802	0.861
On entire image				
ResNet50	0.380	0.985	0.384	0.483
ResNet101	0.463	0.990	0.403	0.478
SSD Inception V2	0.245	0.320	0.458	0.529

the count error for various probability thresholds over the validation dataset (see also Appendix 7.2.1).

In addition, we validated the models using precision and recall on the validation dataset for the whole image, and for the subset of trucks that are on the road (Fig. 3.2). The model with the lowest count error can achieve simultaneously the largest precision and recall. We counted a truck as detected, if its bounding box has an intersection over union (IoU) with the ground truth of at least 0.3. We computed the average precision and recall over all validation images together, and did not average the performance over each image separately. The optimal values are reported in Table 3.1.



(a) With road filter (on Thruway)



(b) On entire image

Figure 3.2: Precision-recall curves for validation images. All of the models performed better when used for on-road predictions, as those often contain less difficult examples.

From Table 3.1 and Fig. 3.2, we can see that the models performed better when the experiment was constrained to the road. The full image can contain more difficult examples, for example clustered trucks on parking lots or less typical trucks in junk yards or construction sites. In our model setting, we only optimized for the on-road counts. SSD Inception V2 achieved the lowest minimal count error on the validation data constrained to the road, and higher precision, recall and speed. We chose to use SSD Inception V2 with prediction probability $p_{pred} = 0.175$ to test the model.

3.4.3 Monitoring model

We trained and validated models to predict the time-varying factors for a test region where we do not have ground truth data. Using the uncertainty of these factors, a distribution of vehicle speeds, and the distance of the section, we used the Monte Carlo method to predict the AADTT.

Factor model selection The count data are seasonal with no trend, which is why we used linear regression models with time fixed effects to estimate the factors of time-of-day, day-of-week, and monthly variation, and the variance of the random component of the traffic counts. Those models were informed by the recommended practices of the FHWA [343, 367]. We created normalized count values by dividing the hourly count data by the mean of all hourly counts in the year.

In the process of selecting the best linear model, we took into account that when the model will be applied, counts data for the test region will not be available, and hence we needed to find a model that predicts well when trained on ground-based counts from other regions. We used a cross-validation procedure to validate the prediction on a held-out region, comparing eight different linear models:

1. normalized count $\sim$ weekend (boolean) + daytime (boolean)
2. normalized count $\sim$ weekend + hour (factor)
3. normalized count $\sim$ weekday (factor) + daytime
4. normalized count $\sim$ weekday + hour

5. normalized count $\sim$ weekday + hour + weekend * factor

6. normalized count $\sim$ weekday + hour + hour * weekday

7. normalized count $\sim$ month (factor) + weekday + hour

8. normalized count $\sim$ month + weekday + hour + hour * weekday.

For the cross validation, we selected the equivalent of 10 continuous ground-based counting stations from each region, where we prioritized those stations that have more data and a higher AADTT. Since some datasets contain short-term counts, to maintain approximate balance, we sampled more counting stations until we had as much data as 10 continuous counters or the dataset was exhausted (Appendix 7.1.3). Since the toll data for the NY Thruway are constrained to one single highway, we used only 6 toll booths here. We ignored inter-year variation.

We trained the model on three of the regions and recorded the mean squared prediction error (MSE) on the held-out fourth region. The MSE averaged over all regions is reported in Fig. 3.3. Some of the models, in particular those that do not have interactions that allow hourly patterns to differ between weekdays and weekend days, produced negative predictions. Since traffic counts are strictly positive, such predictions are infeasible, and we excluded these models. We found that the most complex linear model (Model 8), which includes one interaction term, yielded the lowest cross-validated MSE. We therefore use this linear model to predict time-varying factors.

Fig. 3.4 is a visualization of how Model 8 predicts on each held-out region. Overall, the pattern of truck traffic variability seems to transfer well to other regions. This example, however, also shines light on which types of shortcomings a fit on traffic data from other regions might have. For example, Germany does not allow truck drivers to work on Sundays. We see that the particularly low values on Sundays in Germany, and the corresponding larger variability, are not well predicted by the fit on the other regions that do not have such strict labor rules. In return, the predictions that include German data underpredict Sundays in other regions. This indicates that to improve the performance of the monitoring model, information about local labor rules may need to be incorporated. This figure also illustrates how such traffic models fail to predict deviations from regular patterns, such as holidays (see for example the deviation for German public holidays) [343].

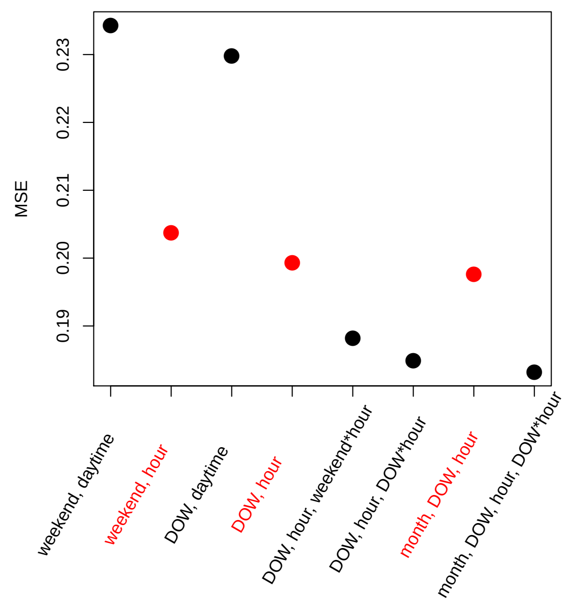


Figure 3.3: Cross-validated MSE of different factor regression models to estimate the normalized hourly count, indicated by the independent variables. Those models that predict infeasible (negative) values are in red.

AADTT prediction and uncertainty analysis We used the predicted vehicle count from the detection model, time-varying factors, and speed to make a prediction of the AADTT with Equation 3.1. With the Monte Carlo method, we predicted a probabilistic uncertainty, where the median value corresponds to the best estimate. We assumed a heuristic for the speed distribution (Gaussian with a standard deviation of 5% of the value). The uncertainty of time-variation factors is composite of the variation of training examples and the confidence of the fitted model and follows a t-distribution. The distribution is the same for all factors.

3.4.4 Test results and discussion

Fig. 3.5 and Table 3.2 show the AADTT estimates from the model compared to the ground-based counts. We find that the model predicts some test cases well. In 3 of the 4 test cases, the value based on traditional ground-based estimation methods is within the interquartile range of the prediction. We see detection model counts that are lower than the true counts in the images, reflected in somewhat lower AADTT predictions (Table 3.2). This suggest that improvements could be made

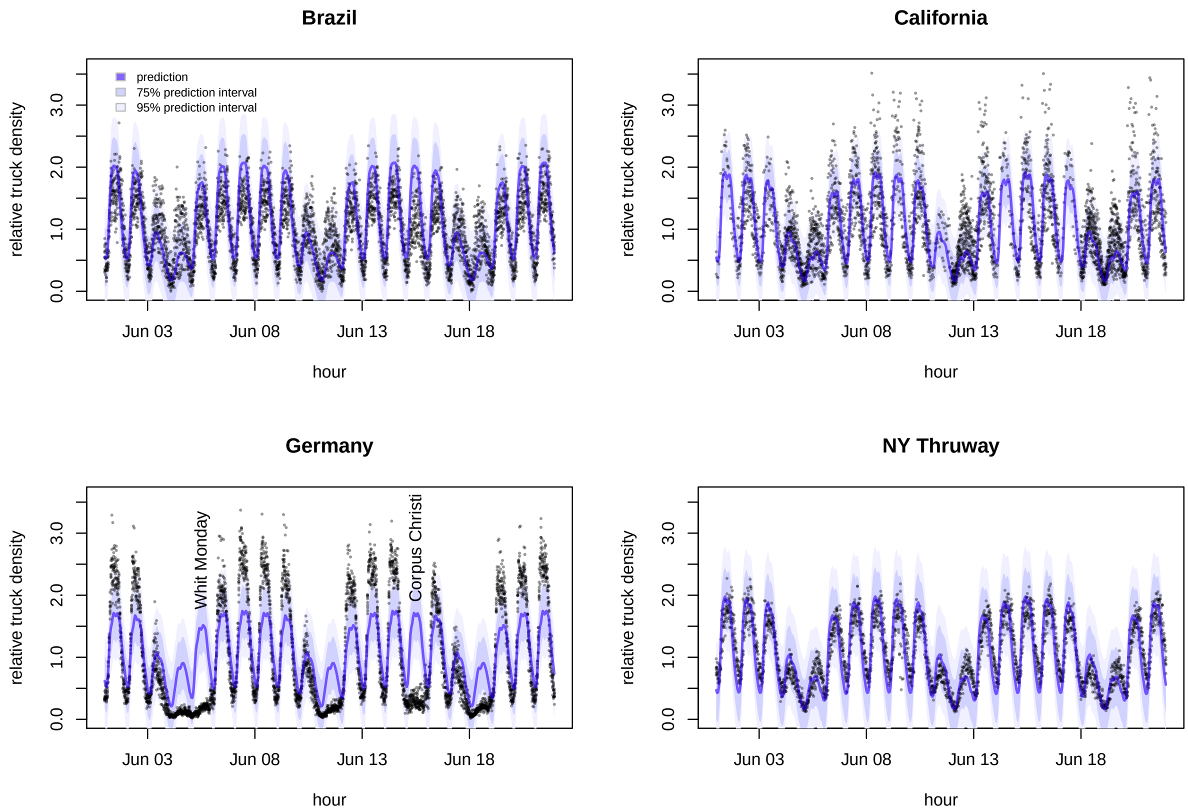


Figure 3.4: Out-of-sample traffic variability prediction for three example weeks in the four different regions. Each plot shows the true normalized hourly vehicle count for all of the randomly selected count stations as scattered points, and the prediction as a blue line. Each model was trained on hourly counts in all three other regions. The prediction interval based on the residual standard deviation is shaded in light blue. German public holidays with strict labor rules are indicated by vertical text. The noise for Corpus Christi stems from the fact that this holiday is not observed in all German states.

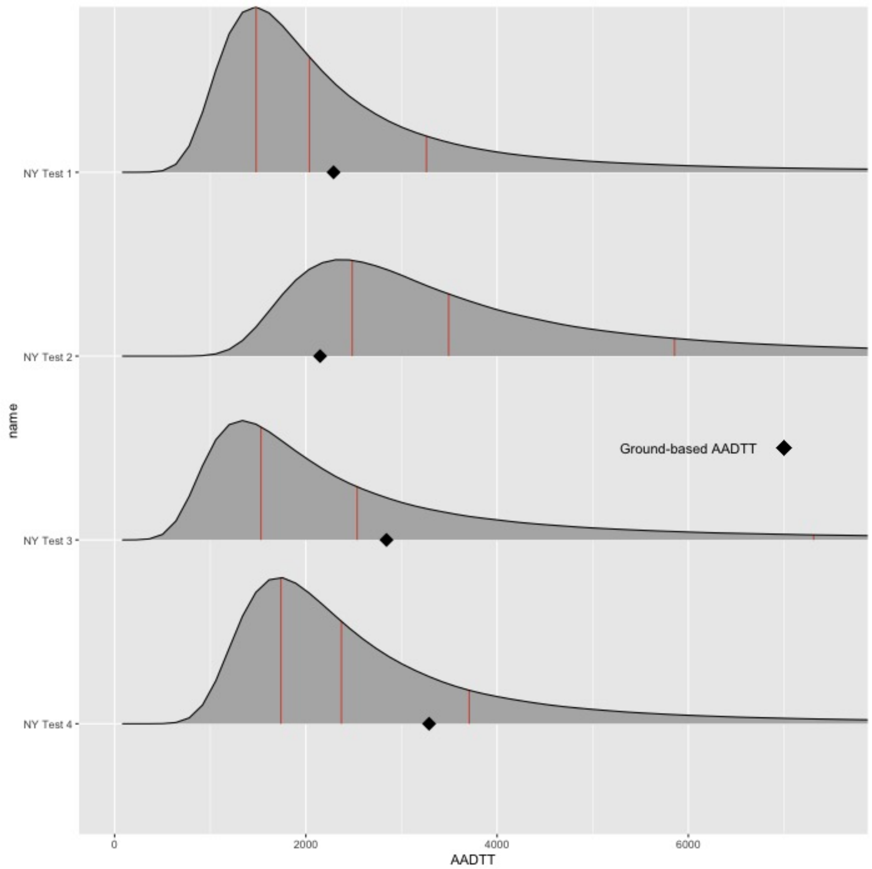


Figure 3.5: Predicted AADTT from satellite images (distributions) and ground-based AADTT (diamond) for different test regions on the NY Thruway. Long upper tails of the distributions are not shown. See Table 3.2 for values.

by increasing the count accuracy of the detection model (see also Appendix 7.2.1 for a discussion). Discrepancy in model performance is expected, given that a snapshot image corresponds to a single, very short counting time, and is sensitive to traffic fluctuations. For example, a very high number of annotated trucks in Test Case 4 can be attributed to trucks parked on the shoulder of the highway that belonged to building grounds. Those were also difficult to detect for the model (Table 3.2).

Our method would provide data for regions where to date no such data exist. As the predictions are uncertain, there is also value in the probabilistic ranges that we provide. Large estimation errors are not uncommon for predictions in energy and transportation, including for well-established conventional methods, and often the uncertainty of the estimate is not communicated. For example, when the U.S. Bureau of Transportation Statistics updated its method to calculate total national

Table 3.2: Test results for the NY Thruway using SSD Inception V2, $\epsilon_{test} = (AADTT_{pred} - AADTT_{true})/AADTT_{true}$.

Exits Section	Time	c_{true}	c_{pred}	$AADTT_{true}$	$AADTT_{pred}$	ϵ_{test}
Exit 25A to 26	08/04/17, Fri, 11:02am	10	8	2290	2039	-11%
Exit 35 to 36	09/18/17, Mon, 11:23am	23	15	2149	3490	62%
Exit 38 to 39	10/06/18, Sat, 11:36am	7	6	2843	2535	-10%
Exit 44 to 45	9/12/17, Tue, 11:27am	35	11	3289	2373	-28 %

road freight activity, the values increased by more than one fourth with the new method [?]. Yet, as this is the only such estimate on road freight activity in the US, it is likely that policy makers routinely make decisions based on those point estimates.

3.5 Generalizing the model to another country

As this model is intended to make predictions in developing countries and emerging economies, we tested how well the model would perform if applied to Brazil. Brazil is suitable as it is an emerging economy but there are sufficient data available to analyze how well the model and each of its components generalize. We were particularly interested to test if additional fine-tuning of the detection model with local images would be necessary.

3.5.1 Data

For the detection model, we used images from DigitalGlobe, Inc., for three different time stamps at the same location. For the monitoring model, we worked with traffic data from continuous and short term counters available through the Brazilian agency Departamento Nacional de Infraestrutura de Transportes (DNIT) [368]. Our test case is a section of the highway BR-116 between two exits, where a counter was located at km 109. For the ground truth we used an AADTT that we computed from all available data as the average of a count of 21 days in 2017 ($AADTT_{true,simple,2017} = 2081$) and 182 days in 2015 ($AADTT_{true,simple,2015} = 3427$) as reported by DNIT [368]. We also retrieved geospatial data of roads in Brazil from DNIT [369]. These data are centered in one of the lanes, which is why we needed to expand the range used in the road filter to 30m to ensure that both lanes pass the filter. This could result in some errors if trucks are parked close to the road. We assumed a

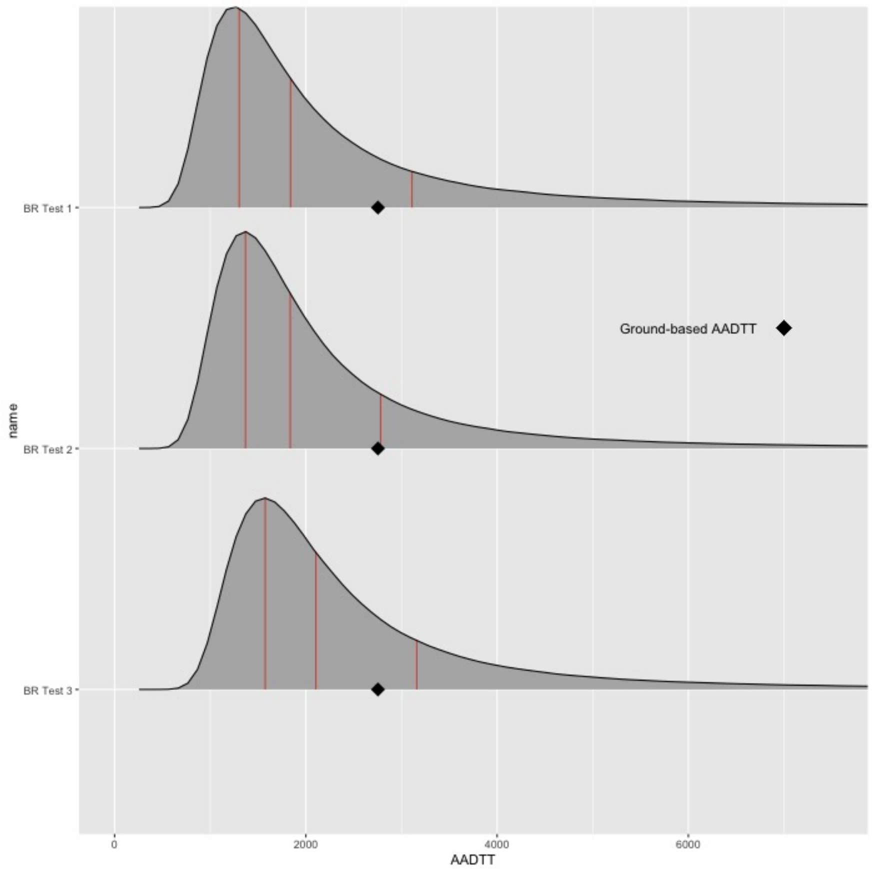


Figure 3.6: Predicted AADTT from satellite images (distributions) and ground-based AADTT (diamond) for a test section on BR-116 km 109 in Brazil for three different times. The detection model is trained on images from Northeastern US. Long upper tails of the distributions are not shown. The vertical lines indicate the 2^{nd}, 3^{rd} (median), and 4^{th} quartile. See Table 3.3 for values of the median.

mean speed of 90 km/hr.

3.5.2 Test results and discussion

We analyzed the performance of the whole model on the Brazil test case, using the trained detection model and parameter settings from the NY Thruway as well as the traffic variation factors trained on count data from the NY Thruway, California, and Germany.

Since all images were taken on weekdays in the same hour for the same section, we expected counts and the predicted AADTT to be comparable. From the results in Table 3.3, we see that the model is able to generate similar AADTT values. We find, however, that the detection model largely

Table 3.3: Test results for Brazil count station BR-116 km 109 using SSD Inception V2 trained on images from Northeastern US. The prediction error is defined as
$\epsilon_{test} = (AADTT_{pred} - AADTT_{true})/AADTT_{true}$.

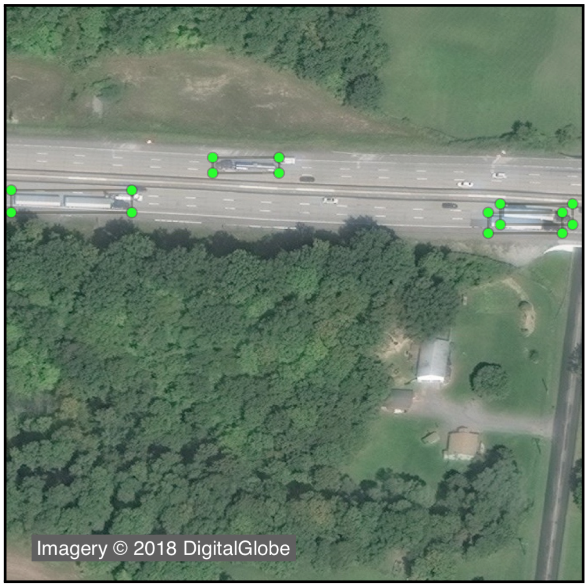
Section	Time	c_{true}	c_{pred}	$AADTT_{true}$	$AADTT_{pred}$	ϵ_{test}
BR-116 km 109	03/12/18, Mon, 10:26am	32	10	2754	1838	-33%
BR-116 km 109	08/11/16, Thur, 10:28am	24	12	2754	1837	-33%
BR-116 km 109	06/16/16, Thur, 10:21am	31	14	2754	2106	-24%

under-predicted the number of trucks in the new images. This translated into an underpredicted AADTT in all test cases (Fig. 3.6 and Table 3.3). If the true number of trucks visible in the images was used, the estimated AADTT was larger than the true value.

Discussion These results clearly indicate that additional fine-tuning of the detection model on images of the new location is necessary. The poor performance of the detection model on data in Brazil is most likely due to the occurrence of new truck types that are specific to Brazil and were not contained in the training dataset (Fig. 3.7). This is also reflected in the generally lower precision recall curve for the Brazil test images in Fig. 3.8.

3.6 Conclusion

We find that we can use machine learning to count trucks in satellite images with reasonable accuracy. Using models of highway traffic patterns that were trained on data in other regions, a snapshot image can yield predictions of average daily traffic volumes that are acceptable, given the data limitations. While these initial results are promising, both the detection model and the monitoring model are likely to entail higher uncertainty when transferred to developing countries. From a test on a highway in Brazil, we found that in particular the detection model does not generalize well to the new images. Distinct truck types and road surfaces likely impact the prediction accuracy of the detection model, and additional training seems necessary. We can attempt to reduce the estimation error by using counts from similar countries or regions, and information from fleet studies or interviews with truck drivers about their driving patterns. Results could also be improved by using multiple satellite images taken of the same section at different times, where available.



(a) NY Thruway

(b) BR-116

Figure 3.7: Green boxes indicate annotated examples. While there is considerable variability in the training data from Northeastern US, which also include winter scenes, the detection model does not generalize well. These images show that trucks seem to look different in Brazil compared to the US. The tractor is more box-shaped in Brazil.

The method currently still requires access to images, knowledge, and computing resources that might be difficult for some countries, but this could change in the near future.

We plan further analysis by using more test data, making improvements to the performance of the detection model, refining the monitoring model by using a non-parametric factor distribution, and performing a sensitivity analysis on the IoU threshold. We will also fine-tune the detection model with training images from Brazil to test if local images lead to performance improvements.

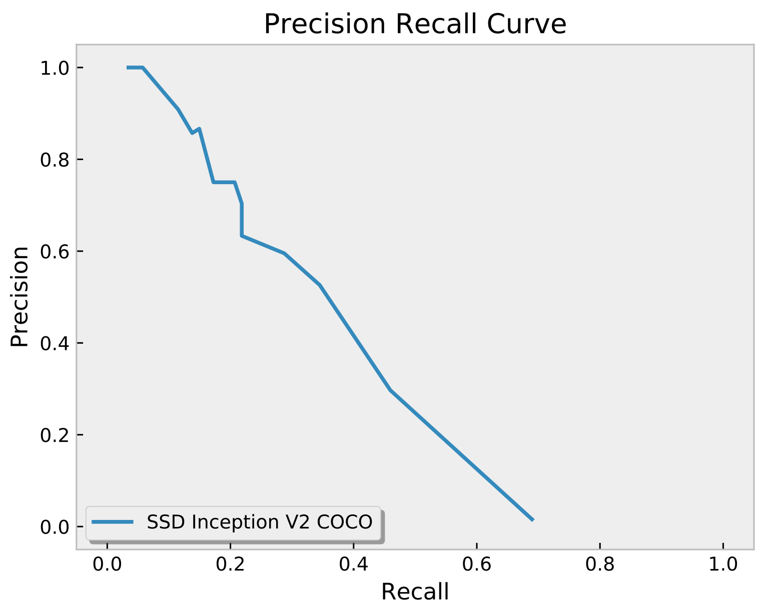


Figure 3.8: Precision recall curve for predictions on the Brazil test images. The curve reflects the performance degradation of the model trained on Northeastern US images when it is applied to images from Brazil. Compare to Fig. 3.2.

4
AI for climate change mitigation

Artificial intelligence (AI), and machine learning in particular, can improve social and engineering research for climate change mitigation - but considerable challenges lie ahead. We argue that the successful application demands careful design of algorithms and consideration of domain knowledge, calling for engineers that are both trained in statistical machine learning and have expertise in the respective climate change mitigation field. We present a survey of studies that apply AI to a wide range of climate change mitigation problems, which demonstrate that AI can improve policy analysis, data availability, forecasting, and energy efficiency in engineering applications. We provide a discussion of challenges that arise at the intersection of the disciplines, and propose a research agenda.

4.1 Introduction

Artificial intelligence (AI) promises to advance a number of research fields. It has contributed an array of powerful new tools to data analytics, leading to some enormous improvements in tasks such as forecasting, computer vision, pattern recognition, or text processing. While researchers are turning to AI to address global challenges in the public sector, the field of environmental sustainability has seen less of this type of work [370]. In this article, we discuss the potential of AI to improve climate change mitigation. In particular, we address three central questions: How can AI be used for climate change mitigation? Which barriers does the application of AI to climate change mitigation face and how could they be addressed? Which research challenges are particular to the intersection of these fields? While there have been conceptual articles addressing AI for the common good [371], for health [372], for developing countries [373], for climate science and modeling [374, 375, 376] and for climate change adaptation [377], the field of climate change mitigation is missing such an analysis.

Researchers in the field of *artificial intelligence* aspire to develop systems that can imitate characteristics of human intelligence, such as the ability to see, understand speech, learn from experience, or generalize. While we will use AI as an umbrella term, many of the methods and challenges discussed here are particular to the subfield of *machine learning* (ML). ML is concerned with algorithms that learn from previous experience to improve a learning task [378]. AI however also includes fields such as social choice, game theory, planning, and natural language processing, which can have impacts on decision-making and analytics in the climate space. As ML heavily draws on and overlaps with statistics, many of the applications and challenges we discuss can also be attributed to statistics. We are interested in those methods that make use of great amounts of data, novel analytical approaches, or computer science, and can therefore advance the field of climate change mitigation in new ways. More traditional statistics approaches like linear regression are part of this method space, but we will not include them in the review and occasionally use them to explicitly contrast with the aforementioned AI methods. Other closely related terms are data analytics and data science, which could also be used to describe much of this method space.

ML methods often require larger numbers of observations to provide the desired accuracy, which is why many studies discussed here work with "big" data. Such data are, for example, provided through vast numbers of new sensors in intelligent infrastructure, they consist of information voluntarily revealed through social media and online platforms, or they can be satellite imagery and other remote sensing data [379]. Characterized by the "V"s (Volume, Velocity, Variety; and sometimes Veracity and Value) [380, 379], big data in combination with AI could enable new research avenues to understand socio-technical factors affecting climate change. Big data has introduced a new paradigm in other related fields, for example geospatial analysis [379], health [381], or economic development [382, 383].

We begin by introducing a number of studies that have used AI to advance research on climate change mitigation (Section 4.2). We group those studies by the methodological role that the AI algorithm takes. In Section 4.3, we discuss potential challenges that can arise with generalizability, capacity, data, and explainability. We conclude by proposing a research agenda in Section 4.4.

4.2 Application domains for climate change mitigation

Here, we present a survey of studies that illustrate where AI has made an important contribution to the field of climate change mitigation.[1] This review is intentionally exploratory, as the boundaries of both the fields of climate change mitigation and AI are fuzzy. We instead focus on structuring what we believe are the most important methodological contributions of AI to research on climate change mitigation. We exclude those studies that make intriguing use of big data, but draw their conclusion by using more traditional data analysis methods. Sections 4.2.1-4.2.4 are concerned with methods that could improve policy analysis and decision making on climate change mitigation. Section 4.2.5 is somewhat distinct, as it describes intelligent - or smart - infrastructure, which is a group of engineering solutions using AI that can contribute to lowering GHG emissions.

[1] We have reviewed recent articles on Scopus.com with the keywords 'climate change' AND 'artificial intelligence' ($>$ 500 articles) and climate change' AND 'machine learning' ($>$ 400 articles). Of the search results, we have only included studies that related to climate change mitigation and actually made use of AI. We excluded climate change science and adaptation studies. We have discovered other articles through our academic network.

4.2.1 Pattern recognition for policy analysis

In fortunate problem settings, a wealth of informative data are available, and the task is to extract knowledge from these data. The field of pattern recognition uses heuristics or, more successfully, machine learning (ML), to identify regularities in data, and for example classifies the data in categories [384]. In ML, an algorithm learns to discover these patterns in a supervised, a semi-supervised, or an unsupervised fashion. In supervised learning, the algorithm is faced with training examples that are labeled by an expert or a ground truth, and it has the task to learn rules that classify these examples correctly. In unsupervised learning, patterns are identified based only on the features that are contained in the data. Here, the algorithm learns to cluster the points, often based on a particular representation of the data, which is hoped to be more informative than the data themselves. Semi-supervised learning lies between these two approaches and uses few labeled examples and many unlabeled data to efficiently discover patterns.

Electricity consumption and efficiency: ML is very powerful in analyzing electricity consumption data that have become more widely available through new meters or home energy monitors,[2] which can help to research and induce energy efficient behavior. For example, analysts can use electricity consumption data and ML to classify households into categories of interest (like the number of occupants) [385] or to cluster them based on similarity [386]. It also provides new methods for energy disaggregation [387], which is the process of understanding individual appliance loads from the total electricity consumption signal. Kolter et al. use additive factorial hidden Markov models [388] and sparse coding [389] for electrical energy disaggregation, and others have used neural networks [390]. Statistical ML also opens up new avenues for causal inference. For example, Burlig et al. [391] use Lasso regression on hourly electricity consumption data from schools in California to find that energy efficiency interventions fall short of the expected savings.

Building energy and heating, ventilating, and air-conditioning (HVAC): ML can provide alternative methods for predicting energy consumption in buildings [392, 393] or for developing

[2] For example `https://sense.com`.

new building energy benchmarks using city-specific data [394]. Clustering is used by Tureczeka et al. to understand consumption patterns in a Danish district heat network [395]. In a global analysis predicting future exposure to extreme heat, Mora et al. [396] use support vector machines (SVMs) to distinguish lethal and non-lethal heat events.

Transportation: AI can improve understanding about passenger travel choices. Some recent studies have shown that supervised ML based on survey data can improve passenger mode choice models [397, 398, 399]. Seo et al. propose to conduct long-term travel surveys with online learning, which reduces the demand on respondents, while obtaining high data quality [400]. Sun et al. [401] use SVMs and neural networks for analyzing preferences of customers traveling with high speed rail in China. Ghaemi et al. [402] cluster public transit users from smart card data. Also, ML makes it possible to estimate origin-destination demand from traffic counts [403]. We found less literature on freight transportation. Spiliopoulos et al. [404] use ML to identify patterns in global ocean shipping.

Climate change perception: Farrell [405] uses a topic model with a latent dirichlet allocation to discover links between corporate funding and climate change polarization. Random forests are used to predict surveyed attitudes towards climate change based on psychological and demographic factors [406]. Other studies turn to twitter data for sentiment analysis of climate change [407, 408] or to debate websites for extracting arguments [409].

Globally comparative studies: Creutzig et al. [410] use decision tree learning to determine urban greenhouse gas emission mitigation potential from a global sample of cities. Ganzenmüller et al. [411] use a clustering technique to analyze national GHG emission pathways globally.

4.2.2 (Remote) sensing

Policy-relevant information remains scarce globally [151]. Here, we introduce AI approaches that provide decision-relevant knowledge from passively collected data, which are often big and geospatial. Li et al. [379] provide a detailed review of the challenges that come with geospatial big data, and emphasize the potential of ML for knowledge discovery. A large part of remote sensing is based on satellite images, where ML, and especially computer vision with deep learning, have made important improvements in object detection, classification and counting [412, 413] and extracting meaningful features that can be mapped to indicators of public interest [414].

Energy systems: Malof et al. [415] and Yu et al. [416] use deep learning to detect rooftop solar photovoltaic (PV) modules in satellite images. This provides information on how much distributed solar generation is installed in an area, and allows to discover correlations with socioeconomic factors at a high spatial resolution. Bogomolov et al. [417] predict energy consumption in Italy by regressing it on cellular communication data with an ensemble ML algorithm. AI methods also help in new wireless methane leak detection technologies [418].

Buildings and transportation: The built environment can be studied well with ML-assisted segmentation or object detection in satellite images or LiDAR data. This allows for example to generate country-wide building footprints data [419] or 3D building models [420]. It has been demonstrated that different ML methods are able to classify building types from LiDAR data [421] or from 3D city models [422], which can be used for estimating building energy demand. Geiß et al. [423] for example use clustering to assess the potential of district heat in a German town based on remote sensing data of the built environment. Vehicles can be counted with high accuracy using deep convolutional neural networks (CNNs) [344, 345, 348, 347], and counts can serve to estimate transportation volume and emissions. In a working paper of our own (Chapter Three), we estimate road freight traffic based on truck counts in satellite images.

Land use, land-use change and forestry (LULUCF): Rudiyanto et al. [424] compare 14 ML models to estimate the carbon stock in tropical peatlands using a number of datasets including satellite images and radar. Similarly, Jardine et al. [425] estimate the carbon in mangrove soils, and find that ML improves their prediction over conventional methods. Hengl et al. [426] use a large number of datasets, many of which are remotely sensed, and apply different ML tools to provide global soil information, which they released as open data. Mascaro et al. [427] also find improvements with ML when remote sensing carbon stocks in tropical forests. Exbrayat and Williams [428] quantify the GHG emissions from deforestation in the Amazonas. Another study uses remote sensing with ML to improve fire prediction in Namibia [429].

4.2.3 Forecasting

We find that especially the electricity sector benefits considerably from probabilistic generation, load, and price forecasting with ML, but there is a large body of AI-based forecasting also in the other energy sectors. Most studies in this section focus on short- or medium-term forecasts. Long-term forecasting is notoriously difficult because of effects such as non-stationarity in the underlying processes [430], and it would be naive to assume that AI can provide better long-term forecasts. Depending on the problem setting, some AI techniques might provide improvements over traditional methods, such as random forests for long-term solar PV growth forecasting [431].

Electricity sector: Especially as solar and wind energy generation is dependent on meteorology, decision-makers in the power sector need reliable short- and medium-term forecasts that also capture the uncertainty of variable renewable generation. AI methods are important for predicting wind power generation [432], and such probabilistic wind power forecasts have been made with sparse Gaussian conditional random fields [433] and neural networks [434, 435]. Similarly, solar radiation and solar PV generation forecasting benefit from ML [436, 437, 438]. Methods such as SVMs [439], hidden Markov models [440], or deep recurrent neural networks [441] are used, just to name a few. Another promising and active area for AI is electric load forecasting, where AI methods can offer higher prediction accuracy and probabilistic forecasts. See [442, 443, 444] for reviews. There are also other types of energy system forecasts that use AI, such as fault prediction with deep learning in power grids that have high solar generation [445]. Cenek et al. [446] combine long short-term memory (LSTM) and neural networks to predict load and renewable energy generation for microgrid management.

Building energy: ML provides an alternative to physical models of energy consumption in buildings, which is sometimes referred to as empirical modeling [447]. The ML approach improves the prediction accuracy of energy use in individual residential buildings as reviewed in [448, 449], for example by employing neural networks [450] or SVMs [447, 451]. Zhang et al. [452] use matching of several datasets with various ML models to estimate residential energy consumption at the neighborhood level. ML methods can also be used for forecasting the demand response potential in buildings [453].

Transportation: There are also a lot of forecasting applications in the transportation sector. For example, AI can improve short-term road traffic flow forecasting [454, 455]. Even though it has received less attention in the literature, there are also fruitful approaches to short-term forecasting of public transit ridership, see for example [456]. Mazloumi et al. predict bus arrival times and uncertainty with neural networks [457]. ML is also used to solve the bikesharing rebalancing problem by improving forecasts of bike demand and inventory [458].

4.2.4 Games, scheduling, and social choice

Fang uses game-theoretic approaches and other methods such as deep reinforcement learning in Green Security Games that can for example combat illegal logging by improving ranger patrolling [459, 460]. Memarzadeh et al. [461] propose a novel method for planning maintenance of wind farms based on reinforcement learning. Tulabandhula and Rudin develop a framework to integrate predictive ML with decision making under uncertainty for power-grid maintenance [462]. AI algorithms also offer ways to understand the social dilemma that is climate change. Hilbe et al. [463] take an evolutionary perspective to stochastic games and show that in face of a depleting resource, cooperation is likely if immediate feedback is provided - which is not the case for climate change where consequences are felt long after the action. Procaccia [464] turns to cake cutting algorithms to explore social welfare and invites to think about climate change negotiations in that way.

4.2.5 Intelligent infrastructure

"Intelligent" or "smart" infrastructure does not only produce vast amounts of data on energy and transportation [379], but it carries the promise of reducing GHG emissions [465]. In contrast to the previous sections that describe how AI can improve policy analysis and decision making on climate change mitigation, intelligent infrastructure refers to a physical system that can be engineered to make energy systems more efficient. Intelligent infrastructure can include sensors, information and communication technology (ICT), analytics that use AI to extract information from the data streams, and control algorithms. In principle, many parts of the energy system can be made "intelligent", giving rise to popular concepts such as smart grid, smart home, smart city, internet-of-things, or intelligent transport systems. Many intelligent engineering concepts have the goal to make operations more efficient, among other objectives such as safety or reliability. We refrain from attempting a survey of this large engineering field, but we provide select examples below that have the ability to reduce GHG emissions.

Many intelligent infrastructure systems are still in the phase of initial development, but analysts often see a large mitigation potential. For example, Pratt et al. estimated possible CO_2 emission savings from smart grids for the U.S. Department of Energy and found that up to 18% of electricity-sector CO_2 emissions in the U.S. could be saved [466]. They found that most of these savings are through the support of additional variable renewable generation sources and electric vehicles, but also through usage diagnostics and feedback for consumers. While smart grids could encourage residential energy savings in this way, intelligent infrastructure in buildings could incur very large additional savings, for example by adjusting HVAC to follow the needs of its occupants more closely [465]. Cities can leverage intelligent infrastructure to monitor transportation, energy, water and waste, and produce detailed estimates of carbon emissions. However, Giest [467] finds that to date urban policy-making based on intelligent infrastructure faces major challenges with data management and analytics. She sees the need to move from small-scale applications towards integrating multiple data sources and domain knowledge for developing GHG-emission-reducing policies. Also, the freight transportation sector is moving towards integrating AI-based engineering applications. Intelligent transport systems of different modes could be combined and enable more efficient multimodal freight transportation [342]. Platooning with autonomous driving and communication technologies could reduce some of the challenges that come with electrifying long-distance road freight [468].

4.3 Challenges and opportunities of AI for climate change mitigation

4.3.1 Generalizability and scale

To move AI from research to application in order for it to make an impact, it would be desirable for models to generalize well. Most models reviewed in the previous section, however, are tailored to specific assumptions and particularities of the local problem setting, which is often desired [373] and necessary to provide highly accurate results. Even studies of a global nature are still particular to the time frame of the data available. ML is designed for the model to generalize to new data of the same distribution (same statistical properties). That does not imply that a model generalizes well to problem settings where the underlying distribution of the data has changed. This has two main implications: 1) Models will need to be re-trained, and often also adapted, when applied to another local problem, which requires analysts who are skilled in statistics and ML. 2) In contrast to the expectation that AI and data availability allow to explore larger hypotheses, e.g., of global scope, scaling the analysis up is not straight forward.

In principle, models can be re-trained (or fine-tuned) and validated when faced with new data where statistical properties might have changed. However, this requires new labeled training data, which are based on expert knowledge or ground truth data, and which can be expensive. In addition, expert knowledge is often case-specific, and training data might only be available for one local problem.

The field of transfer learning therefore aims to understand how models trained on a source domain $\mathcal{D}_S$ can improve the performance of a learning task on a target domain $\mathcal{D}_T$, which has different statistical properties. It assumes that there is some relationship between source and target data, based on which it is possible to derive theoretical error bounds [469]. Pan and Yang [470] provide a good introduction to transfer learning and domain adaptation. A prominent application of transfer learning to poverty mapping is provided by Jean et al. [414].

Another way to deal with the fact that expert knowledge is expensive, is to use methods that rely on as few labeled data points as possible. For example, querying limited but informative labels and making use of unlabeled data in some cases can provide good classification accuracy (active learning [471]). The field of semi-supervised learning explores classification and clustering in the setting, where few fixed labeled data and a lot of unlabeled data are available [472]. Online algorithms can provide solutions, where decisions need to be made repeatedly based on some expert knowledge, and where new labels appear one after the other [473].

In our view, generalizability should be a central concern for most studies of AI applications. For example, it has been documented that smart city projects experience difficulty moving from small-scale insights based on big data to generating large-scale policies that can incur long-term emission reductions [467].

4.3.2 Explainability, Accountability, and Fairness

AI methods can be very powerful for making predictions or allocating resources, but often they are used as an input to answer much more complex questions on climate change mitigation policy and engineering. These usually require human intervention to find the solution or decision, and they are crucially based on trust in the algorithm. Decision makers often base their trust on some understanding of why the algorithm decided one way or another. Here lies another problem of many AI methods: They are very opaque. Neural networks, for example, do not communicate which features the prediction was based on. Even if the data and code are fully disclosed, the model is still a black box [474]. Most refer to this issue as explainability, or interpretability, of AI [475]. Especially for decision support systems in the public sector, there are deep ethical concerns if algorithms cannot be held accountable [476]. This includes creating "fair" algorithms that for example predict equally well across two demographics that are divided by a protected attribute such as gender [477]. In circumstances where full transparency is required, the best option might be to use interpretable models such as decision trees or linear regression, where prediction accuracy might be sacrificed at the expense of interpretability.

New research is exploring ways to increase explainability, without compromising prediction accuracy. There exist many different strategies to increase transparency at the algorithm level, for example by highlighting predictive features in "saliency maps", but they are still under development and often unreliable [475]. Visualizing how neurons activate, e.g., as in [478], provides a lot of information about deep neural networks, and is particularly instructive for computer vision. For policy decisions, post-hoc techniques by examples or counterfactuals might be more relevant [475]. For example, Lai et al. use perturbation of signals to understand how residential electricity consumption responds to changes in building variables [447]. In the future, different approaches might be combined to an explanation object [475].

Also, causal inference based on ML is still difficult [479] but techniques do exist [480]. As policy analysts are often interested in measuring the effect of a certain policy intervention, causal inference can be important for climate change mitigation. For example Zhao et al. [481] use these techniques to estimate the effect of World Bank funding on forest cover.

ML can deliver very high prediction accuracy and new probabilistic approaches to forecasting, but physical modeling of engineering systems is better suited to analyze the workings of energy systems. A promising approach is to use ML methods, such as neural networks, for metamodeling of a physical model. Those are trained on a number of realizations of the simulation model (for example a building energy model). As they present an approximation of the model, they reduce the computational effort and allow to run a much higher number of experiments [482].

4.3.3 Capacity requirements

To date, even inference (using trained ML models to make predictions) requires computing knowledge and resources. As discussed above, to draw relevant information from data with AI algorithms, it is often necessary to verify that the underlying assumptions are still fulfilled if time passes or as data have changed otherwise. In many cases the model will need to be adapted. This requires the organization or project owner to have access to personnel who have the statistics and ML skills to do so. Depending on the application, also computing resources can be a bottleneck, especially if big data are analyzed.

Considering that even companies in cities like New York have difficulties finding skilled personnel [483], attracting labor might be nearly impossible for many public organizations and institution that work on climate change mitigation. Hilbert [382] finds that many countries face a general shortage of labor with analytical skills, such as statistics, data analytics, and computer science. As GHG emissions from emerging economies are outpacing industrialized nations [484], mitigation solutions need to be applicable also in developing countries. According to Hilbert, relative spending in computing power and computer service employees differs widely between countries, and might force some of them to outsource the work. Some organizations, in particular in developing countries, might not be able to afford computing resources, e.g., GPUs for training deep nets. These challenges are often described as the digital divide and might generally compromise the quality of AI applications.

Similar to De-Arteaga et al. [373], we view these challenges as an opportunity to make ML applications more dynamically adaptable and tailor requirements for training and inference to the local circumstances of use cases. Open source applications and collaborative environments could help alleviating some of the challenges with labor [382]. For example, Stoyanovich et al. [474] propose a collaborative analytics platform, "Fides", that combines data management with analytics that adhere to rules of fair, accountable, and transparent analysis. A way to overcome computing bottlenecks is to develop models that rely on lean algorithms that require less memory. Also intelligent infrastructure benefits from lean algorithms. Large sensor networks tend towards distributed computing, in order to only transmit relevant information instead of transmitting the large stream of raw data that are being collected [379].

4.3.4 Data requirements

In addition to the digital divide, there is what Hilbert called the digital big data divide [382]. Many organizations in the world still do not have the capacities to collect and release data that are timely, harmonized, well-documented, and machine-readable - or they might refrain from doing so because they are unaware of the benefits [151, 467]. Big data, or at least datasets that are large enough for learning, are often collected in regions or under circumstances that are not representative of the problem at hand, and therefore exhibit sampling bias. For example, in some countries mobile phone use is constrained to more privileged parts of the population [382].

Often datasets in the real world, in particular when working across countries or institutional layers, are messy with missing data points and implicit biases, and can be quite small. These issues require new approaches and new ML algorithms [373]. Bias in data may also be masked or introduced in pre-processing or at later stages of the analysis process [474]. In addition, there are concerns with privacy [485] as data obtained through sensors often contain sensitive information, such as motion patterns or energy use of people. As AI provides tools to make sense of these data for surveillance, abuse of models with benign intention can also not be excluded.

When moving from a proof of concept to using the models for decision-relevant applications on the ground, researchers should also reflect about data access. While many data are produced and provided through the public sector as open data [382, 383], in other circumstances data can be proprietary or too expensive to acquire for some organizations. For example, high-resolution satellite image data are mostly available through private companies at high cost. Academia should also grapple with the fact that while they might gain access to these types of resources, they are not, or not yet, widely available to all users. In the meantime, alternatives should be fully explored (for example drone imagery instead of satellite images).

Successful applications of AI and big data rely on government regulation to ensure data availability, reliability, and minimize privacy concerns [373, 382]. Examples of local governments, such as cities that implement intelligent infrastructure projects, show that problems can arise due to the lack the skills within the organization and subsequent outsourcing [467]. Many of the challenges with data can be improved by addressing data literacy and data processing in a globally harmonized way, for example through the United Nations [151] or through well-designed collaborative analytics platforms [474]. The climate change mitigation field could perhaps gain insights from fields that moved faster into big data analytical approaches such as genomics [486] or astronomy [18].

4.4 A research agenda for AI for climate change mitigation

In our view, there is a need to embrace the difficulties specific to interdisciplinary research and consider new approaches to building models. We believe that two quite different approaches to developing real-work AI models are promising for the field of climate change mitigation.

In many problem settings in the field of climate change mitigation, there exist only limited ground truth data and labeling is costly. Learning with small data is still a challenge [373]. However, there is often a lot of domain knowledge that can be leveraged, and heuristics about the uncertainty of outcomes can help validating the performance of the model. Data-driven models are therefore often integrated with theoretical models to arrive at a relevant prediction. These are sometimes referred to as "grey box" models, for example for forecasting building electricity use, which couple data-driven models with physical models [448, 449]. This approach is very flexible, and models are often relatively easy to design and train. In fact, many studies reviewed here use AI for a part of the model, and integrate the results with more traditional analysis. We see a need to formalize these approaches.

Another promising avenue is to coordinate between the realms of prediction and decision-making. This can be achieved by training predictive models end-to-end in a manner that accounts for the decision maker's ultimate objective. For instance, a decision maker may seek to reduce power system emissions, for which they may require an ML model producing solar forecasts. No predictive model is perfect, and in such settings over-forecasting often results in very different losses for the decision maker than under-forecasting. An end-to-end ML model makes error-tradeoffs that are tuned towards the decision maker's ultimate task. One approach to end-to-end model training involves embedding the decision-maker's optimization problem into the ML model training. Donti et al. [487] propose such a model of task-based learning and run experiments on two different decision-problems from the electricity sector. While leading to considerable improvements in accuracy, these model architectures are more difficult to train than standard neural networks [487, 488, 489]. End-to-end training might be most suitable where the algorithm is needed to directly propose a decision, such as in engineering applications and some policy applications.

As the challenges described above arise at the intersection of domain knowledge and algorithm design, it becomes obvious that analysts, who have substantial training in both AI and climate change mitigation, are essential to further this type of research. This is needed in addition to fostering collaborative environments between the disciplines [377]. First, research that is firmly anchored in both disciplines is necessary for real-world challenges to drive cutting-edge ML research [373]. Second, without in-depth understanding of AI methods and model assumptions, the outcomes of those approaches can have limited validity for adequately addressing mitigation problems. At the same time, purely data-driven approaches without understanding of the theory can get the solution very wrong [490]. To foster this kind of collaborative environment, computing challenges such as the United Nations Global Pulse Big Data Climate Challenge[3] could encourage interdisciplinary teams. Collaborative platforms such as that proposed by Stoyanovich et al. could help to form teams and address some of the capacity concerns [474].

4.5 Conclusion

Our review and analysis of the literature revealed that there is a wide range of climate change mitigation applications where AI can deliver new insights. Although neural networks and support vector machines seem to dominate, studies take advantage of an array of different methods. Most studies surveyed here are fairly recent, and new data sources seem to drive the research. We find that forecasting across all sectors, remote sensing of LULUCF, and analysis of electricity consumption patterns have most benefitted from AI to date. Although still underrepresented, those studies that use AI to generate and publish national or global datasets relevant to climate change mitigation seem promising. While there is enormous potential for using AI and big data sourced from intelligent infrastructure, few studies use this approach to generate strategies and policies for long-term emission reductions. Here, analyses that integrate different datasets and domain knowledge are urgently needed. Most studies reported that AI improved relevant metrics, such as prediction accuracy, over traditional methods, or that the new approach allowed to circumvent other known difficulties of traditional methods. Other studies could not have been conducted at all without ML, or not at the spatial or temporal scale that was used.

At the same time, significant challenges lie ahead. AI constitutes a set of new methods, and the application of these methods requires engineers who have received formal training and who are able to apply scrutiny to the models they design and work with. To create meaningful research in the climate mitigation space, collaborative environments as well as scholars trained in both domain expertise and data analytics are needed.

[3] `https://www.unglobalpulse.org/data-for-climate-action`

Conclusion

To achieve a reduction in greenhouse gas (GHG) emissions, it ultimately matters that actionable insights are generated and disseminated to policy makers. Here, I discuss how the work in this dissertation is relevant to decision makers in the public sector.

In the first chapter [430], I recommend that the U.S. Energy Information Administration (EIA) made tables available containing the standard deviations of past projection errors. These allow forecast users to construct probabilistic estimates and to understand empirically how wide the forecast uncertainty may be. Used jointly with projections and scenarios that convey information on the workings of the energy system, these can improve decision making in the energy sector. After completing this work, I presented my results to the EIA and my recommendations were adopted in EIA's most recent Retrospective Review (2018) of the Annual Energy Outlook [491].

Chapter Two [342] provides the most comprehensive review to date on national freight transportation activity from countries where data are available. These results should be able to aid international organizations such as the United Nations, who are advancing a globally coordinated effort to standardize reporting in the freight sector. Especially as we move towards deeper decarbonization, globally harmonized and complete data on freight transportation are extremely important. Research of this kind can help in pushing for such data standards.

Chapter Three is one of the few examples of studies in the field of climate change mitigation that employ a deep learning model and develop the full pipeline to generate policy-relevant information.[4] However, this proof of concept is not ready for wide implementation due to a number of technical

[4] As reviewed in Chapter Four.

barriers. First, there is limited availability of high-resolution satellite images, and most organizations would need to pay a steep price for access. Another bottleneck is the computing infrastructure, as even inference on small images covering a few kilometers of road takes around half an hour on a laptop. When the model is applied to a new region, it will need to be re-trained, requiring access to a graphics processing unit (GPU). A minimum of computer science literacy is needed to train models, and to conduct inference with them. Models are currently mainly available as open source Python code. I conclude that public organizations that are looking to leverage intelligent infrastructure and data science currently have little alternative but to employ their own personnel who have training in data management, artificial intelligence (AI), and computer science, as outsourcing to the private sector may be problematic for privacy reasons. Further considerations regarding the use of AI methods for climate change mitigation are discussed in detail in Chapter Four.

It is my hope that this thesis, and especially the proposed agenda in Chapter Four, will spur research at the intersection of data science and climate change mitigation. Naturally, data science alone will not reduce GHG emissions, yet it can provide a tool for addressing many questions that urgently need to be answered to slow dangerous climate change.

Bibliography

[1] United Nations, and Canada, *United Nations Framework Convention on Climate Change*. New York, USA: United Nations, General Assembly, 1992.

[2] "Kyoto Protocol to the United Nations Framework Convention on Climate Change." U.N. Doc FCCC/CP/1997/7/Add.1, 37 I.L.M. 22, 1998. Dec. 10, 1997.

[3] UNFCCC, "Adoption of the Paris agreement - Proposal by the President - Draft decision -/CP.21." Available at `https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf`, December 2015.

[4] Y. Robiou du Pont and M. Meinshausen, "Warming assessment of the bottom-up paris agreement emissions pledges," *Nature Communications*, vol. 9, Nov 2018.

[5] J. Rockström, O. Gaffney, J. Rogelj, M. Meinshausen, N. Nakicenovic, and H. J. Schellnhuber, "A roadmap for rapid decarbonization," *Science*, vol. 355, no. 6331, pp. 1269–1271, 2017.

[6] R. K. Pachauri, M. R. Allen, V. R. Barros, J. Broome, W. Cramer, R. Christ, J. A. Church, L. Clarke, Q. Dahe, P. Dasgupta, *et al.*, *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC, 2014.

[7] H. J. Schellnhuber, S. Rahmstorf, and R. Winkelmann, "Why the right climate target was agreed in paris," *Nature Climate Change*, vol. 6, pp. 649–653, Jul 2016.

[8] M. Kalkuhl, O. Edenhofer, and K. Lessmann, "Learning or lock-in: Optimal technology policies to support mitigation," *Resource and Energy Economics*, vol. 34, no. 1, pp. 1 – 23, 2012.

[9] S. J. Davis, N. S. Lewis, M. Shaner, S. Aggarwal, D. Arent, I. L. Azevedo, S. M. Benson, T. Bradley, J. Brouwer, Y.-M. Chiang, C. T. M. Clack, A. Cohen, S. Doig, J. Edmonds, P. Fennell, C. B. Field, B. Hannegan, B.-M. Hodge, M. I. Hoffert, E. Ingersoll, P. Jaramillo, K. S. Lackner, K. J. Mach, M. Mastrandrea, J. Ogden, P. F. Peterson, D. L. Sanchez, D. Sperling, J. Stagner, J. E. Trancik, C.-J. Yang, and K. Caldeira, "Net-zero emissions energy systems," *Science*, vol. 360, no. 6396, 2018.

[10] J. Teter, P. Cazzola, and T. Gül, *The Future of Trucks*. International Energy Agency, 2017.

[11] A. Woodburn, M. Browne, M. Piotrowska, and J. Allen, "Literature review wm7: Scope for modal shift through fiscal, regulatory and organisational change," *University of Westminster and University of Leeds*, 2007.

[12] International Energy Agency, "Transport, Energy and CO2," 2009.

[13] J. J. Brogan, A. E. Aeppli, D. F. Bagan, A. Brown, M. J. Fischer, L. R. Grenzeback, E. McKenzie, L. Vimmerstedt, A. D. Vyas, and E. Witzke, "Freight transportation modal shares: scenarios for a low-carbon future," tech. rep., 2013.

[14] United Nations Economic Commission for Europe (UNECE), "Transport for Sustainable Development - The case of Inland Transport," 2015.

[15] G. Morgan, "Theory and practice in policy analysis," 2017.

[16] M. G. Morgan, M. Kandlikar, J. Risbey, and H. Dowlatabadi, "Why conventional tools for policy analysis are often inadequate for problems of global change," *Climatic Change*, vol. 41, pp. 271–281, Mar 1999.

[17] E. Horvitz and D. Mulligan, "Data, privacy, and the greater good," *Science*, vol. 349, no. 6245, pp. 253–255, 2015.

[18] G. Bell, T. Hey, and A. Szalay, "Beyond the data deluge," *Science*, vol. 323, no. 5919, pp. 1297–1298, 2009.

[19] J. J. Winebrake and D. Sakva, "An evaluation of errors in US energy forecasts: 1982–2003," *Energy Policy*, vol. 34, no. 18, pp. 3475–3483, 2006.

[20] M. Wara, D. Cullenward, and R. Teitelbaum, "Peak Electricity and the Clean Power Plan," *The Electricity Journal*, vol. 28, no. 4, pp. 18–27, 2015.

[21] A. Q. Gilbert and B. K. Sovacool, "Looking the wrong way: Bias, renewable electricity, and energy modelling in the United States," *Energy*, vol. 94, pp. 533–541, 2016.

[22] Intergovernmental Panel on Climate Change, "Definition of terms used within the DDC pages." `http://www.ipcc-data.org/guidelines/pages/definitions.html`, 2015.

[23] A. Neuhauser, "Wasted energy," *U.S. News and World Report*, May 2015.

[24] C. Harvey, "How we get energy is changing fast — and it's sparking a huge fight over forecasting the future," *Washington Post*, May 2016.

[25] C. Fischer, E. Herrnstadt, and R. Morgenstern, "Understanding errors in EIA projections of energy demand," *Resource and Energy Economics*, vol. 31, no. 3, pp. 198–209, 2009.

[26] H. Linderoth, "Forecast errors in IEA-countries' energy consumption," *Energy Policy*, vol. 30, no. 1, pp. 53–61, 2002.

[27] V. Smil, "Perils of long-range energy forecasting: reflections on looking far ahead," *Technological Forecasting and Social Change*, vol. 65, no. 3, pp. 251–264, 2000.

[28] R. Schlaifer and H. Raiffa, *Applied statistical decision theory.* 1961.

[29] M. G. Morgan and M. Henrion, *Uncertainty : a guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge New York: Cambridge University Press, 1990.

[30] B. Fischhoff and A. L. Davis, "Communicating scientific uncertainty," *Proceedings of the National Academy of Sciences*, vol. 111, no. Supplement 4, pp. 13664–13671, 2014.

[31] M. G. Morgan and D. W. Keith, "Improving the way we think about projecting future energy use and emissions of carbon dioxide," *Climatic Change*, vol. 90, no. 3, pp. 189–215, 2008.

[32] A. I. Shlyakhter, D. M. Kammen, C. L. Broido, and R. Wilson, "Quantifying the credibility of energy projections from trends in past data: The US energy sector," *Energy Policy*, vol. 22, no. 2, pp. 119–130, 1994.

[33] P. P. Craig, A. Gadgil, and J. G. Koomey, "What can history teach us? A retrospective examination of long-term energy forecasts for the United States," *Annual Review of Energy and the Environment*, vol. 27, no. 1, pp. 83–118, 2002.

[34] T. Gneiting, "Editorial: probabilistic forecasting," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pp. 319–321, 2008.

[35] S. P. Vahey and L. Wakerly, "Moving towards probability forecasting," *BIS Paper*, no. 70b, 2013.

[36] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, vol. 1, pp. 125–151, 2014.

[37] F. X. Diebold, A. S. Tay, and K. F. Wallis, "Evaluating density forecasts of inflation: The survey of professional forecasters," in *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger* (R. F. Engle and H. White, eds.), Oxford University Press, 1997.

[38] E. Britton, P. Fisher, and J. Whitley, "The Inflation Report projections: understanding the fan chart," *Bank of England Quarterly Bulletin*, vol. 38, no. 1, pp. 30–37, 1998.

[39] M. Blix and P. Sellin, "Uncertainty bands for inflation forecasts," *Sveriges Riksbank Working Paper Series*, vol. 65, 1998.

[40] A. S. Tay and K. F. Wallis, "Density forecasting: a survey," *Journal of forecasting*, vol. 19, no. 4, pp. 235–254, 2000.

[41] T. J. Linsmeier and N. D. Pearson, "Value at risk," *Financial Analysts Journal*, vol. 56, no. 2, pp. 47–67, 2000.

[42] A. E. Raftery, N. Li, H. Ševčíková, P. Gerland, and G. K. Heilig, "Bayesian probabilistic population projections for all countries," *Proceedings of the National Academy of Sciences*, vol. 109, no. 35, pp. 13915–13921, 2012.

[43] P. E. McSharry, S. Bouwman, and G. Bloemhof, "Probabilistic forecasts of the magnitude and timing of peak electricity demand," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1166–1172, 2005.

[44] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *IEEE Transactions on Energy Conversion*, vol. 24, no. 3, pp. 775–782, 2009.

[45] P. Pinson, "Wind energy: Forecasting challenges for its operational management," *Statistical Science*, vol. 28, no. 4, pp. 564–585, 2013.

[46] F. X. Diebold, T. A. Gunther, and A. S. Tay, "Evaluating density forecasts, with applications to financial risk management," *International Economic Review*, vol. 39, pp. 863–883, 1998.

[47] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.

[48] L. A. Smith, E. B. Suckling, E. L. Thompson, T. Maynard, and H. Du, "Towards improving the framework for probabilistic forecast evaluation," *Climatic Change*, vol. 132, no. 1, pp. 31–45, 2015.

[49] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

[50] H. Hersbach, "Decomposition of the continuous ranked probability score for ensemble prediction systems," *Weather and Forecasting*, vol. 15, no. 5, pp. 559–570, 2000.

[51] W. H. Williams and M. L. Goodman, "A simple method for the construction of empirical confidence limits for economic forecasts," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 752–754, 1971.

[52] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *Power Systems, IEEE Transactions on*, vol. 25, no. 4, pp. 1845–1856, 2010.

[53] NOAA National Hurricane Center, "National hurricane center forecast verification," March 2016. `http://www.nhc.noaa.gov/verification/verify6.shtml`.

[54] O. Isengildina-Massa, S. Irwin, D. L. Good, and L. Massa, "Empirical confidence intervals for usda commodity price forecasts," *Applied Economics*, vol. 43, no. 26, pp. 3789–3803, 2011.

[55] M. Knüppel, "Efficient estimation of forecast uncertainty based on recent forecast errors," *International Journal of Forecasting*, vol. 30, no. 2, pp. 257–267, 2014.

[56] Y. S. Lee and S. Scholtes, "Empirical prediction intervals revisited," *International Journal of Forecasting*, vol. 30, no. 2, pp. 217–234, 2014.

[57] U.S. Energy Information Administration, "Annual Energy Outlook," 2016. `http://www.eia.gov/forecasts/aeo/`.

[58] U.S. Energy Information Administration, "Annual Energy Outlook Retrospective Review," 2015. `https://www.eia.gov/forecasts/aeo/retrospective/`.

[59] B. C. O'Neill and M. Desai, "Accuracy of past projections of us energy consumption," *Energy Policy*, vol. 33, no. 8, pp. 979–993, 2005.

[60] M. Auffhammer, "The rationality of eia forecasts under symmetric and asymmetric loss," *Resource and Energy Economics*, vol. 29, no. 2, pp. 102–121, 2007.

[61] C. M. Sprenkle, "Warrant prices as indicators of expectations and preferences," *Yale Economics Essays*, vol. 1, pp. 178–231, 1961.

[62] R Core Team, *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2015.

[63] A. W. Schäfer, A. D. Evans, T. G. Reynolds, and L. Dray, "Costs of mitigating co2 emissions from passenger aircraft," *Nature Climate Change*, vol. 6, no. 4, pp. 412–417, 2016.

[64] R. Schaeffer, R. Sims, J. Corfee-Morlot, F. Creutzig, X. Cruz-Nunez, D. Dimitriu, and M. e. a. D'Agosto, *Transport, in IPCC (ed), Working Group III contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Climate Change 2014: Mitigation of Climate Change, chapter 8. Geneva.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2014.

[65] F. Creutzig, P. Jochem, O. Y. Edelenbosch, L. Mattauch, D. P. van Vuuren, D. McCollum, and J. Minx, "Transport: A roadblock to climate change mitigation?," *Science*, vol. 350, no. 6263, pp. 911–912, 2015.

[66] T. S. Schmidt and J. Huenteler, "Anticipating industry localization effects of clean technology deployment policies in developing countries," *Global Environmental Change*, vol. 38, pp. 8–20, 2016.

[67] C. Huizenga, S. General, and K. Peet, "Transport and climate change: How nationally determined contributions can accelerate transport decarbonization,"

[68] J. Eom, L. Schipper, and L. Thompson, "We keep on truckin': Trends in freight energy use and carbon emissions in 11 IEA countries," *Energy Policy*, vol. 45, pp. 327–341, 2012.

[69] OECD, "Freight transport (indicator)." doi: 10.1787/708eda32-en (Accessed on 12 February 2018), 2018.

[70] World Bank, "World development indicators," 2017. Available at http://data.worldbank.org/data-catalog/world-development-indicators.

[71] K. M. Hoen, T. Tan, J. C. Fransoo, and G.-J. van Houtum, "Switching transport modes to meet voluntary carbon emission targets," *Transportation Science*, vol. 48, no. 4, pp. 592–608, 2013.

[72] OECD, "Intermodal Freight Transport: Institutional Aspects," 2001.

[73] IMO, "Third IMO Greenhouse Gas Study," 2014.

[74] A. McKinnon, "Freight transport in a low-carbon world: Assessing opportunities for cutting emissions," *TR News*, no. 306, 2016.

[75] E. Bennathan, J. Fraser, and L. S. Thompson, *What determines demand for freight transport?*, vol. 998. World Bank Publications, 1992.

[76] P. Tapio, "Towards a theory of decoupling: degrees of decoupling in the eu and the case of road traffic in finland between 1970 and 2001," *Transport policy*, vol. 12, no. 2, pp. 137–151, 2005.

[77] M. Levinson, *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger, with a new chapter by the author.* Princeton University Press, 2016.

[78] P. Londoño-Kent, "Freight transport for development toolkit: Road freight," *World Bank, Washington DC*, 2009.

[79] A. C. McKinnon, "Decoupling of road freight transport and economic growth trends in the uk: An exploratory analysis," *Transport Reviews*, vol. 27, no. 1, pp. 37–64, 2007.

[80] J. J. Winebrake, E. H. Green, B. Comer, C. Li, S. Froman, and M. Shelby, "Fuel price elasticities in the us combination trucking sector," *Transportation Research Part D: Transport and Environment*, vol. 38, pp. 166–177, 2015.

[81] C. A. Dahl, "Measuring global gasoline and diesel price and income elasticities," *Energy Policy*, vol. 41, pp. 2–13, 2012.

[82] G. de Jong, A. Schroten, H. van Essen, M. Otten, and P. Bucci, "The price sensitivity of road freight transport: A review of elasticities," *Applied transport economics, a management and policy perspective, De Boeck, Antwerp*, 2010.

[83] W. E. Frazier, "Metal additive manufacturing: a review," *Journal of Materials Engineering and Performance*, vol. 23, no. 6, pp. 1917–1928, 2014.

[84] R. Huang, M. Riddle, D. Graziano, J. Warren, S. Das, S. Nimbalkar, J. Cresko, and E. Masanet, "Energy and emissions saving potential of additive manufacturing: the case of lightweight aircraft components," *Journal of Cleaner Production*, vol. 135, pp. 1559–1570, 2016.

[85] H. Helms and U. Lambrecht, "The potential contribution of light-weighting to reduce transport energy consumption," *The International Journal of Life Cycle Assessment*, vol. 1, pp. 58–64, 2006.

[86] J. Holmström and T. Gutowski, "Additive manufacturing in operations and supply chain management: No sustainability benefit or virtuous knock-on opportunities?," *Journal of Industrial Ecology*, vol. 21, no. S1, 2017.

[87] J. B. Roca, P. Vaishnav, J. Mendonca, and M. G. Morgan, "Getting past the hype about 3-D printing," *MIT Sloan Management Review*, 2017.

[88] T. Wohlers, "3d printing and additive manufacturing state of the industry. annual worldwide progress report," *Wohlers Report*, 2014.

[89] J. Manyika, M. Chui, J. Bughin, R. Dobbs, P. Bisson, and A. Marrs, *Disruptive technologies: Advances that will transform life, business, and the global economy*, vol. 180. McKinsey Global Institute San Francisco, CA, 2013.

[90] M. Gebler, A. J. S. Uiterkamp, and C. Visser, "A global sustainability perspective on 3d printing technologies," *Energy Policy*, vol. 74, pp. 158–167, 2014.

[91] National Research Council, *Technologies and approaches to reducing the fuel consumption of medium-and heavy-duty vehicles.* National Academies Press, 2010.

[92] M. N. Taptich, A. Horvath, and M. V. Chester, "Worldwide greenhouse gas reduction potentials in transportation by 2050," *Journal of Industrial Ecology*, 2015.

[93] N. R. Council *et al.*, *Reducing the fuel consumption and greenhouse gas emissions of medium-and heavy-duty vehicles, phase two: First report.* National Academies Press, 2014.

[94] S. Sripad and V. Viswanathan, "Performance metrics required of next-generation batteries to make a practical electric semi truck," *ACS Energy Letters*, vol. 2, pp. 1669–1673, 2017.

[95] M. Guttenberg, S. Sripad, and V. Viswanathan, "Evaluating the potential of platooning in lowering the required performance metrics of li-ion batteries to enable practical electric semi-trucks," *ACS Energy Letters*, vol. 2, no. 11, pp. 2642–2646, 2017.

[96] Siemens, "eHighway - Electrification of road freight transport." Available at `https://www.siemens.com/global/en/home/products/mobility/road-solutions/electromobility/ehighway.html`.

[97] Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit, "Erneuerbar mobil." Available in German at `www.erneuerbar-mobil.de/projekte`.

[98] D. Boffey, "World's first electrified road for charging vehicles opens in Sweden," *The Guardian*, April 2018. Available at `https://www.theguardian.com/environment/2018/apr/12/worlds-first-electrified-road-for-charging-vehicles-opens-in-sweden`.

[99] O. Olsson, "Slide-in electric road system—conductive project report," *Viktoria Swedish ICT, Gothenburg, Sweden*, 2013.

[100] M. Roeth, D. Kircher, J. Smith, and R. Swim, "Barriers to the Increased Adoption of Fuel Efficiency Technologies in the North American On-Road Freight Sector," *Report for the International Council for Clean Transportation. NACFE*, 2013.

[101] C. Bynum, C. Sze, D. Kearns, B. Polovick, and K. Simon, "An examination of a voluntary policy model to effect behavioral change and influence interactions and decision making in the freight sector," *Transportation Research Part D: Transport and Environment*, 2016.

[102] American Trucking Associations, "Reports, Trends and Statistics." Available at `http://www.trucking.org/News_and_Information_Reports_Industry_Data.aspx`.

[103] A. Hoffrichter, A. R. Miller, S. Hillmansen, and C. Roberts, "Well-to-wheel analysis for electric, diesel and hydrogen traction for railways," *Transportation Research Part D: Transport and Environment*, vol. 17, no. 1, pp. 28–34, 2012.

[104] Scott Institute for Energy Innovation, "Power Sector Carbon Index," 2017. Retrieved from `https://emissionsindex.org/`.

[105] A. J. Friedemann, "Can freight trains be electrified?," in *When Trucks Stop Running*, pp. 67–73, Springer, 2016.

[106] F. Zenith, S. Møller-Holst, and M. Thomassen, "Hydrogen and Batteries for Propulsion of Freight Trains in Norway," 7 2016. Presentation slides.

[107] H. Van Essen, O. Bello, J. Dings, and R. van den Brink, "To shift or not to shift, that's the question," *Delft, CE*, 2003.

[108] L. C. Blancas and M. B. El-Hifnawi, *Facilitating Trade Through Competitive, Low-carbon Transport: The Case for Vietnam's Inland and Coastal Waterways*. World Bank Publications, 2013.

[109] P. Nuttall, A. Newell, B. Prasad, J. Veitayaki, and E. Holland, "A review of sustainable sea-transport for oceania: Providing context for renewable energy shipping for the pacific," *Marine Policy*, vol. 43, pp. 283–287, 2014.

[110] M. S. Eide, T. Longva, P. Hoffmann, Ø. Endresen, and S. B. Dalsøren, "Future cost scenarios for reduction of ship CO2 emissions," *Maritime Policy and Management*, vol. 38, no. 1, pp. 11–37, 2011.

[111] R. Geertsma, R. Negenborn, K. Visser, and J. Hopman, "Design and control of hybrid power and propulsion systems for smart ships: A review of developments," *Applied Energy*, vol. 194, pp. 30–54, 2017.

[112] L. Van Biert, M. Godjevac, K. Visser, and P. Aravind, "A review of fuel cell systems for maritime applications," *Journal of Power Sources*, vol. 327, pp. 345–364, 2016.

[113] S. Hanley, "China Launches World's First All-Electric Cargo Ship, Will Use It To Haul Coal," Available at: `https://cleantechnica.com/2017/12/02/china-launches-worlds-first-electric-cargo-ship-will-use-haul-coal/`.

[114] D. Boffey, "World's first electric container barges to sail from European ports this summer," *The Guardian*, January 2018. Available at `https://www.theguardian.com/environment/2018/jan/24/worlds-first-electric-container-barges-to-sail-from-european-ports-this-summer`.

[115] V. Bryan, "Air cargo carriers battle competition from seas, passenger planes," *Reuters*, 2014.

[116] C. Bryant, "Air Freight Has Tummy Trouble," *Bloomberg Gadfly*, 2016.

[117] National Academies of Sciences, Engineering, and Medicine, *Commercial aircraft propulsion and energy systems research: reducing global carbon emissions*. Washington, DC: National Academies Press, 2016.

[118] W. Graham, C. Hall, and M. V. Morales, "The potential of future aircraft technology for noise and pollutant emissions reduction," *Transport Policy*, vol. 34, pp. 36–51, 2014.

[119] E. M. Greitzer, P. Bonnefoy, E. De la Rosa Blanco, C. Dorbian, M. Drela, D. Hall, R. Hansman, J. Hileman, R. Liebeck, J. Lovegren, *et al.*, "N+ 3 aircraft concept designs and trade studies, final report," *Nasa cr-2010-216794/vol2, NASA Glenn Research Center, Cleveland, Ohio*, vol. 44135, 2010.

[120] J. Downer, "The aviation paradox: Why we can 'know'jetliners but not reactors," *Minerva*, pp. 1–20, 2017.

[121] J. Bonnin Roca, E. R. Fuchs, P. Vaishnav, M. G. Morgan, and J. Mendonça, "When risks cannot be seen: regulating uncertainty in emerging technologies," 2016.

[122] N. S. Lewis, "Artificial photosynthesis," *American Scientist*, vol. 83, no. 6, pp. 534–541, 1995.

[123] N. S. Lewis, "Toward cost-effective solar energy use," *Science*, vol. 315, no. 5813, pp. 798–801, 2007.

[124] S. J. Bann, R. Malina, M. D. Staples, P. Suresh, M. Pearlson, W. E. Tyner, J. I. Hileman, and S. Barrett, "The costs of production of alternative jet fuel: A harmonized stochastic assessment," *Bioresource technology*, vol. 227, pp. 179–187, 2017.

[125] R. W. Stratton, H. M. Wong, and J. I. Hileman, "Quantifying variability in life cycle greenhouse gas inventories of alternative middle distillate transportation fuels," *Environmental science & technology*, vol. 45, no. 10, pp. 4637–4644, 2011.

[126] R. W. Stratton, *Life cycle assessment of greenhouse gas emissions and non-*CO_2 *combustion effects from alternative jet fuels.* PhD thesis, Massachusetts Institute of Technology, 2010.

[127] A. Hellemans, "The Netherlands' Carbon Dilemma: Sequester or Recycle?," *IEEE Spectrum*, December 2017.

[128] D. Haeseldonckx and W. D'haeseleer, "The use of the natural-gas pipeline infrastructure for hydrogen transport in a changing market structure," *International Journal of Hydrogen Energy*, vol. 32, no. 10, pp. 1381–1386, 2007.

[129] A. E. Farrell, D. W. Keith, and J. J. Corbett, "A strategy for introducing hydrogen into transportation," *Energy Policy*, vol. 31, no. 13, pp. 1357–1367, 2003.

[130] A. J. Hawkins, "Toyota's hydrogen fuel cell trucks are now moving goods around the Port of LA." Available at: `https://www.theverge.com/2017/10/12/16461412/toyota-hydrogen-fuel-cell-truck-port-la`.

[131] J. Kast, G. Morrison, J. J. Gangloff Jr, R. Vijayagopal, and J. Marcinkoski, "Designing hydrogen fuel cell electric trucks in a diverse medium and heavy duty market," *Research in Transportation Economics*, 2017.

[132] F. Tong, P. Jaramillo, and I. M. Azevedo, "Comparison of life cycle greenhouse gases from natural gas pathways for medium and heavy-duty vehicles," *Environmental science & technology*, vol. 49, no. 12, pp. 7123–7133, 2015.

[133] O. Dinu and A. Ilie, "Maritime vessel obsolescence, life cycle cost and design service life," in *IOP Conference Series: Materials Science and Engineering*, vol. 95, p. 012067, IOP Publishing, 2015.

[134] H. Thomson, J. J. Corbett, and J. J. Winebrake, "Natural gas as a marine fuel," *Energy Policy*, vol. 87, pp. 153–167, 2015.

[135] C. Dong, R. Boute, A. McKinnon, and M. Verelst, "Investigating synchromodality from a supply chain perspective," *Transportation Research Part D: Transport and Environment*, 2017.

[136] European Comission. Directorate-General for Mobility and Transport, *White Paper on Transport: Roadmap to a Single European Transport Area: Towards a Competitive and Resource-efficient Transport System.* Publications Office of the European Union, 2011.

[137] European Commission, Directorate-General for Mobility and Transport, "Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending Directive 92/106/EEC on the establishment of common rules for certain types of combined transport of goods between Member States." COM/2017/0648 final - 2017/0290 (COD), 2017.

[138] M. SteadieSeifi, N. P. Dellaert, W. Nuijten, T. Van Woensel, and R. Raoufi, "Multimodal freight transportation planning: A literature review," *European journal of operational research*, vol. 233, no. 1, pp. 1–15, 2014.

[139] E. C. D.-G. for Energy, *Keep Europe moving: sustainable mobility for our continent: mid-term review of the European Commission's 2001 Transport White Paper.* Office for Official Publications of the European Communities, 2006.

[140] J. J. Winebrake, J. J. Corbett, A. Falzarano, J. S. Hawker, K. Korfmacher, S. Ketha, and S. Zilora, "Assessing energy, environmental, and economic tradeoffs in intermodal freight transportation," *Journal of the Air & Waste Management Association*, vol. 58, no. 8, pp. 1004–1013, 2008.

[141] A. J. Craig, E. E. Blanco, and Y. Sheffi, "Estimating the CO2 intensity of intermodal freight transportation," *Transportation Research Part D: Transport and Environment*, vol. 22, pp. 49–53, 2013.

[142] Institut für Energie- und Umweltforschung Heidelberg and Studiengesellschaft für den kombinierten Verkehr e.V., "Comparative analysis of energy consumption and co2 emission of road transport and combined transport road/rail," 2009.

[143] A. Alessandrini, P. Delle Site, F. Filippi, and M. V. Salucci, "Using rail to make urban freight distribution more sustainable," 2012.

[144] C. M. d. Oliveira, R. Albergaria De Mello Bandeira, G. Vasconcelos Goes, D. N. Schmitz Gonçalves, and M. D. A. D'Agosto, "Sustainable vehicles-based alternatives in last mile distribution of urban freight transport: A systematic literature review," *Sustainability*, vol. 9, no. 8, p. 1324, 2017.

[145] J. K. Stolaroff, C. Samaras, E. R. O'Neill, A. Lubers, A. S. Mitchell, and D. Ceperley, "Energy use and life cycle greenhouse gas emissions of drones for commercial package delivery," *Nature communications*, vol. 9, no. 1, p. 409, 2018.

[146] BITRE, "Australian infrastructure statistics," *Australian Government Department of Infrastructure and Regional Development*, 2017.

[147] J. Dulac, "Global land transport infrastructure requirements," *International Energy Agency, Paris*, 2013.

[148] International Energy Agency, *Energy Technology Perspectives.* International Energy Agency, 2012.

[149] *Capacity to grow - Transport Infrastructure Needs for Future Trade Growth.* OECD International Transport Forum, 2016.

[150] OECD International Transpot Forum, "ITF Transport Outlook 2017," 2017.

[151] Independent Expert Advisory Group on a Data Revolution for Sustainable Development, "A world that counts: mobilizing the data revolution for sustainable development," *United Nations Publication*, 2014.

[152] T. S. Rosenstock, C. Lamanna, S. Chesterman, J. Hammond, S. Kadiyala, E. Luedeling, K. Shepherd, B. DeRenzi, and M. T. van Wijk, "When less is more: innovations for tracking progress toward global targets," *Current Opinion in Environmental Sustainability*, vol. 26, pp. 54–61, 2017.

[153] *Road freight transport methodology 2016 edition.* Eurostat, 2016.

[154] *Decree No. 97/2016 / ND-CP dated 1 July 2016 regulating the content of statistical indicators in the NSIS.* General Statistics Office of Vietnam, 2016. (In Vietnamese).

[155] U. DoT, "Comprehensive truck size and weight study," *Federeal Highway Administration, Washington, DC*, 2000.

[156] A. S. Fowkes, P. E. Firmin, G. Tweddle, and A. E. Whiteing, "How highly does the freight transport industry value journey time reliability—and for what reasons?," *International Journal of Logistics: Research and Applications*, vol. 7, no. 1, pp. 33–43, 2004.

[157] H. Rodemann and S. Templar, "The enablers and inhibitors of intermodal rail freight between Asia and Europe," *Journal of Rail Transport Planning & Management*, vol. 4, no. 3, pp. 70–86, 2014.

[158] F. Robuste and F. Soriguera, *Improving reliability on surface transport networks.* OECD Publishing, 2010.

[159] N. Shinghal and T. Fowkes, "Freight mode choice and adaptive stated preferences," *Transportation Research Part E: Logistics and Transportation Review*, vol. 38, no. 5, pp. 367–378, 2002.

[160] A. I. Arencibia, M. Feo-Valero, L. García-Menéndez, and C. Román, "Modelling mode choice for freight transport using advanced choice experiments," *Transportation Research Part A: Policy and Practice*, vol. 75, pp. 252–267, 2015.

[161] R. Danielis, E. Marcucci, and L. Rotaris, "Logistics managers' stated preferences for freight service attributes," *Transportation Research Part E: Logistics and Transportation Review*, vol. 41, no. 3, pp. 201–215, 2005.

[162] P. Christidis, G. Leduc, *et al.*, "Longer and heavier vehicles for freight transport," *JRC European Commission*, pp. 1–40, 2009.

[163] A. McKinnon, "Improving the sustainability of road freight transport by relaxing truck size and weight restrictions," in *Sustainable Practices: Concepts, Methodologies, Tools, and Applications*, pp. 1265–1278, IGI Global, 2014.

[164] M. J. Meixell and M. Norbis, "A review of the transportation mode choice and carrier selection literature," *The International Journal of Logistics Management*, vol. 19, no. 2, pp. 183–211, 2008.

[165] "Green freight programs." Available at `http://www.globalgreenfreight.org/resources/green-freight`.

[166] L. Tavasszy and J. Van Meijeren, "Modal shift target for freight transport above 300 km: an assessment," *Brussels, Belgium*, 2011.

[167] "Signs point to US shippers' slowing intermodal conversion," June 2015.

[168] J. Ludvigsen and R. Klæboe, "Extreme weather impacts on freight railways in Europe," *Natural hazards*, vol. 70, no. 1, pp. 767–787, 2014.

[169] International Energy Agency, *Energy Technology Perspectives.* International Energy Agency, 2015.

[170] M. Figueroa, O. Lah, L. M. Fulton, A. McKinnon, and G. Tiwari, "Energy for transport," *Annual Review of Environment and Resources*, vol. 39, pp. 295–325, 2014.

[171] International Road Transport Union (IRU) and World Bank, *Road Freight Transport Services Reform.* IRU, World Bank, 2017.

[172] *Tracking Africa's progress in figures.* African Development Bank, 2014.

[173] D. Omondi and M. Kamau, "Despite fanfare, SGR faces same age-old threat that 'killed' Kenya Railway," *Standard Digital*, June 2017.

[174] J. Monios and R. Bergqvist, "Intermodal terminal concessions: lessons from the port sector," *Research in Transportation Business & Management*, vol. 14, pp. 90–96, 2015.

[175] G. Liedtke and D. G. C. Murillo, "Assessment of policy strategies to develop intermodal services: The case of inland terminals in Germany," *Transport Policy*, vol. 24, pp. 168–178, 2012.

[176] V. Roso, J. Woxenius, and K. Lumsden, "The dry port concept: connecting container seaports with the hinterland," *Journal of Transport Geography*, vol. 17, no. 5, pp. 338–345, 2009.

[177] J.-P. Rodrigue and T. Notteboom, "Dry ports in European and North American intermodal rail systems: Two of a kind?," *Research in Transportation Business & Management*, vol. 5, pp. 4–15, 2012.

[178] A. Veenstra, R. Zuidwijk, and E. Van Asperen, "The extended gate concept for container terminals: Expanding the notion of dry ports," *Maritime Economics & Logistics*, vol. 14, no. 1, pp. 14–32, 2012.

[179] V. Roso, "Evaluation of the dry port concept from an environmental perspective: A note," *Transportation Research Part D: Transport and Environment*, vol. 12, no. 7, pp. 523–527, 2007.

[180] R. Van den Berg and P. W. De Langen, "An exploratory analysis of the effects of modal split obligations in terminal concession contracts," *International Journal of Shipping and Transport Logistics*, vol. 6, no. 6, pp. 571–592, 2014.

[181] R. Dekker, J. Bloemhof, and I. Mallidis, "Operations research for green logistics–an overview of aspects, issues, contributions and challenges," *European Journal of Operational Research*, vol. 219, no. 3, pp. 671–679, 2012.

[182] B. F. Santos, S. Limbourg, and J. S. Carreira, "The impact of transport policies on railroad intermodal freight competitiveness–the case of belgium," *Transportation Research Part D: Transport and Environment*, vol. 34, pp. 230–244, 2015.

[183] Y. M. Bontekoning, C. Macharis, and J. J. Trip, "Is a new applied transportation research field emerging?–a review of intermodal rail-truck freight transport literature," *Transportation Research Part A: Policy and Practice*, vol. 38, no. 1, pp. 1–34, 2004.

[184] R. Stahlbock and S. Voß, "Operations research at container terminals: a literature update," *OR spectrum*, vol. 30, no. 1, pp. 1–52, 2008.

[185] T. A. Mathisen and T.-E. S. Hanssen, "The academic literature on intermodal freight transport," *Transportation Research Procedia*, vol. 3, pp. 611–620, 2014.

[186] G. De Jong, I. Vierth, L. Tavasszy, and M. Ben-Akiva, "Recent developments in national and international freight transport models within europe," *Transportation*, vol. 40, no. 2, pp. 347–371, 2013.

[187] J. S. L. Lam and Y. Gu, "Port hinterland intermodal container flow optimisation with green concerns: a literature review and research agenda," *International Journal of Shipping and Transport Logistics*, vol. 5, no. 3, pp. 257–281, 2013.

[188] J. S. L. Lam and Y. Gu, "A market-oriented approach for intermodal network optimisation meeting cost, time and environmental requirements," *International Journal of Production Economics*, vol. 171, pp. 266–274, 2016.

[189] L. A. Tavasszy, B. Behdani, and R. Konings, "Intermodality and synchromodality," 2015.

[190] S. Dalmolen, E. Cornelisse, W. Hofman, H. Basiaansen, M. Punter, and F. Knoors, "Improving sustainability through intelligent cargo and adaptive decision making," e-Freight Conference 2012, Delft, The Netherlands, 2012.

[191] *A European Strategy for Low-Emission Mobility*, vol. COM/2016/0501 final. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS, 2016. Available at `http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52016DC0501`.

[192] B. Behdani, Y. Fan, B. Wiegmans, and R. Zuidwijk, "Multimodal schedule design for synchromodal freight transport systems," 2014.

[193] S. Pfoser, H. Treiblmaier, and O. Schauer, "Critical success factors of synchromodality: Results from a case study and literature review," *Transportation Research Procedia*, vol. 14, pp. 1463–1471, 2016.

[194] K. Verweij, "Synchromodal transport: Thinking in hybrid cooperative networks," *Van Der Sterre, P.J. (Ed.), Logistics Yearbook 2011*, pp. 75 – 88, 2011.

[195] B. Van Riessen, R. R. Negenborn, and R. Dekker, "Synchromodal container transportation: an overview of current topics and research opportunities," in *International Conference on Computational Logistics*, pp. 386–397, Springer, 2015.

[196] I. Harris, Y. Wang, and H. Wang, "Ict in multimodal transport and technological trends: Unleashing potential for the future," *International Journal of Production Economics*, vol. 159, pp. 88–103, 2015.

[197] A. Broaddus and C. Gertz, "Tolling heavy goods vehicles: Overview of European practice and lessons from German experience," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2066, pp. 106–113, 2008.

[198] U. Clausen, C. Geiger, and C. Behmer, "Green corridors by means of ICT applications," *Procedia-Social and Behavioral Sciences*, vol. 48, pp. 1877–1886, 2012.

[199] S. Hernández, S. Peeta, and G. Kalafatas, "A less-than-truckload carrier collaboration planning problem under dynamic capacities," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 6, pp. 933–946, 2011.

[200] G. Shafiullah, A. Gyasi-Agyei, and P. Wolfs, "Survey of wireless communications applications in the railway industry," in *Wireless Broadband and Ultra Wideband Communications, 2007. AusWireless 2007. The 2nd International Conference on*, pp. 65–65, IEEE, 2007.

[201] L. Ruiz-Garcia, P. Barreiro, J. Rodríguez-Bermejo, and J. Robla, "Monitoring the intermodal, refrigerated transport of fruit using sensor networks," *Spanish Journal of Agricultural Research*, vol. 5, no. 2, pp. 142–156, 2007.

[202] T. Sandholm, *Very-large-scale generalized combinatorial multi-attribute auctions: Lessons from conducting $60 billion of sourcing. The Handbook of Market Design*, ch. 16. Oxford University Press, 2013.

[203] Y. Wang, A. Potter, and M. Naim, "An exploratory study of electronic logistics marketplaces and its impact on customised logistics," in *POMS 18th Annual Conference Dallas, Texas, USA, May*, vol. 4, 2007.

[204] M. Arnold, "Banks team up with IBM in trade finance blockchain," *Financial Times*, October 2017. Available at `https://www.ft.com/content/7dc8738c-a922-11e7-93c5-648314d2c72c`.

[205] "Blockchain in Trucking Alliance." Available at `https://bita.studio`.

[206] J. Berman, "UPS becomes a member of the Blockchain in Trucking Alliance," *Logistics Management*, November 2017. Available at `http://www.logisticsmgmt.com/article/ups_becomes_a_member_of_the_blockchain_in_trucking_alliance`.

[207] European Comission Mobility and Transport, "FAQ on ERTMS." Available at https://ec.europa.eu/transport/modes/rail/ertms/general-information/faq_ertms_en.

[208] "The European Rail Traffic Management System." Available at http://www.ertms.net.

[209] G. Schilk and L. Seemann, "Use of its technologies for multimodal transport operations–river information services (ris) transport logistics services," *Procedia-Social and Behavioral Sciences*, vol. 48, pp. 622–631, 2012.

[210] United Nations ESCAP, "Review of Developments in Transport in Asia and the Pacific," *UN ESCAP Transport for Sustainable Development and Regional Connectivity*, 2015.

[211] World Bank Group: Public-Private-Partnership in Infrastructure Resource Center, "Track Access Laws and Regulations." Available at https://ppp.worldbank.org/public-private-partnership/ppp-sector/transportation/railways/railway-laws-and-regulations/track-access-laws-and-regulations/tr.

[212] F. Dionori, L. Casullo, S. Ellis, D. Ranghetti, K. Bablinski, C. Vollath, and C. Soutra, *Freight on road: why EU shippers prefer truck to train.* European Parliament, 2015.

[213] B. E. Horn and T. Nemoto, "Intermodal Logistics Policies in the EU, the US and Japan," *Transport policy studies' review*, vol. 7, no. 4, pp. 2–14, 2005.

[214] European Commission, "Marco Polo." Available at https://ec.europa.eu/transport/marcopolo/index_en.htm.

[215] "Marco Polo II Programme - Frequently Asked Questions (FAQ)," 2013. Available at https://ec.europa.eu/transport/marcopolo/files/calls/docs/2011/faq_call2011.pdf.

[216] European Commission, "Commission authorises Belgium to support intermodal transport by Flemish inland waterways," 2007. Available at http://europa.eu/rapid/press-release_IP-07-653_en.htm?locale=en.

[217] H. Van Essen, B. Boon, A. Schroten, M. Otten, M. Maibach, C. Schreyer, C. Doll, P. Jochem, M. Bak, and B. Pawlowska, "Internalisation measures and policy for the external cost of transport," *paper produced within the study Internalisation Measures and Policies for all external cost of Transport (IMPACT), Deliverable*, vol. 3, 2008.

[218] A. Korzhenevych, N. Dehnen, J. Bröcker, M. Holtkamp, H. Meier, G. Gibson, A. Varma, and V. Cox, "Update of the handbook on external costs of transport," *European Commission DG MOVE*, 2014.

[219] M. Beuthe, F. Degrandsart, J.-F. Geerts, and B. Jourquin, "External costs of the Belgian interurban freight traffic: a network analysis of their internalisation," *Transportation Research Part D: Transport and Environment*, vol. 7, no. 4, pp. 285–301, 2002.

[220] M. Mostert and S. Limbourg, "External costs as competitiveness factors for freight transport—a state of the art," *Transport Reviews*, vol. 36, no. 6, pp. 692–712, 2016.

[221] European Commission, "EU ETS Handbook," 2015.

[222] M. Harding, "The diesel differential: Differences in the tax treatment of gasoline and diesel for road use," *OECD Taxation Working Papers*, no. 21, p. 0_1, 2014.

[223] *Companion to the Inventory of Support Measures for Fossil Fuels 2015*. OECD Publishing, Paris, 2015. http://dx.doi.org/10.1787/9789264239616-en.

[224] S. Li, J. Linn, and E. Muehlegger, "Gasoline taxes and consumer behavior," *American Economic Journal: Economic Policy*, vol. 6, no. 4, pp. 302–342, 2014.

[225] J. Marion and E. Muehlegger, "Fuel tax incidence and supply conditions," *Journal of Public Economics*, vol. 95, no. 9, pp. 1202–1212, 2011.

[226] European Comission Mobility and Transport, "Road Infrastructure Charging - Heavy Goods Vehicles." Available at https://ec.europa.eu/transport/modes/road/road_charging/charging_hgv_en.

[227] C. Façanha, K. Blumberg, and J. Miller, "Global transportation energy and climate roadmap: The impact of transportation policies and their potential to reduce oil consumption and greenhouse gas emissions," *Report, International Council on Clean Transportation, Washington http://www. theicct. org/sites/default/files/publications/ICCT% 20Roadmap% 20Energy% 20Report. pdf*, 2012.

[228] X. Jun, "Trucking in China's Logistics Sector," *Members-only Library*, 2015.

[229] J. Fioretti, "EU challenges Austrian minimum wage law for truck drivers," *Reuters*, April 2017.

[230] H. Ghaderi, S. Cahoon, and H.-O. Nguyen, "The role of intermodal terminals in the development of non-bulk rail freight market in Australia," *Case studies on transport policy*, vol. 4, no. 4, pp. 294–305, 2016.

[231] R. Addis, "Cost 334: effects of wide single tyres and dual tyres," 2002. Available at https://comt.ca/english/programs/trucking/Pavements/European%20Commission%20COST%20334%20Study%20-%20Effects%20of%20Wide%20Single%20Tyres%20and%20Dual%20Tires%202001.pdf.

[232] African Development Bank Group, "Rail Infrastructure in Africa," 2015.

[233] V. S. Rodrigues, M. Piecyk, R. Mason, and T. Boenders, "The longer and heavier vehicle debate: A review of empirical evidence from germany," *Transportation Research Part D: Transport and Environment*, vol. 40, pp. 114–131, 2015.

[234] R. Bergqvist and S. Behrends, "Assessing the effects of longer vehicles: the case of pre-and post-haulage in intermodal transport chains," *Transport Reviews*, vol. 31, no. 5, pp. 591–602, 2011.

[235] D. Cole, T. Furst, S. Daboin, W. Hoemann, M. D. Meyer, R. Nordahl, M. Parker, L. Penne, N. Stoner, and T. Tang, "Freight mobility and intermodal connectivity in China," tech. rep., 2008.

[236] X. Luo, L. Dong, Y. Dou, H. Liang, J. Ren, and K. Fang, “Regional disparity analysis of Chinese freight transport CO2 emissions from 1990 to 2007: Driving forces and policy challenges,” *Journal of Transport Geography*, vol. 56, pp. 1–14, 2016.

[237] H. Duan, M. Hu, Y. Zhang, J. Wang, W. Jiang, Q. Huang, and J. Li, “Quantification of carbon emissions of the transport service sector in china by using streamlined life cycle assessment,” *Journal of Cleaner Production*, vol. 95, pp. 109–116, 2015.

[238] W. Knörr and F. Dünnebeil, “Transport in China: Energy Consumption and Emissions of Different Transport Modes,” *Institute for Energy and Environmental Research Heidelberg, Heidelberg, Germany*, 2008.

[239] L. C. Blancas, G. Ollivier, and R. Bullock, “Customer-driven rail intermodal logistics: Unlocking a new source of value for china,” *China Transport Topics*, no. 12, 2015.

[240] OECD, “Total inland transport infrastructure investment per GDP,” 2018. `http://stats.oecd.org/Index.aspx?DataSetCode=ITF_INDICATORS`.

[241] J. Perlez and Y. Huang, “Behind China’s $1 Trillion Plan to Shake Up the Economic Order,” *The New York Times*, May 2017.

[242] Transport Canada, “Transportation in Canada,” 2016.

[243] Statistics Canada, “Domestic and international shipping - Tonnage loaded and unloaded,” November 2015. Available at `http://www.statcan.gc.ca/pub/54-205-x/2011000/t015-eng.htm`.

[244] H. Pouryousef, P. Lautala, and T. White, “Railroad capacity tools and methodologies in the US and Europe,” *Journal of Modern Transportation*, vol. 23, no. 1, pp. 30–42, 2015.

[245] Association of American Railroads, “America’s Freight Railroads: Global Leaders Under a Balanced Regulatory System.” `https://www.aar.org/BackgroundPapers/America%27s%20Railroads%20Global%20Leaders.pdf`, 2015.

[246] J.-P. Rodrigue, “The thruport concept and transmodal rail freight distribution in north america,” *Journal of Transport Geography*, vol. 16, no. 4, pp. 233–246, 2008.

[247] Oak Ridge National Lab, “Transportation energy data book,” vol. 35, 2016.

[248] Association of American Railroads, “Rail Intermodal Keeps America Moving,” April 2017.

[249] Y. Zhou, A. D. Vyas, and Z. Guo, “An evaluation of the potential for shifting of freight from truck to rail and its impacts on energy use and ghg emissions,” tech. rep., Argonne National Lab.(ANL), Argonne, IL (United States), 2017.

[250] Z. Patterson, G. O. Ewing, and M. Haider, “The potential for premium-intermodal services to reduce freight CO2 emissions in the Quebec City–Windsor Corridor,” *Transportation Research Part D: Transport and Environment*, vol. 13, no. 1, pp. 1–9, 2008.

[251] J. Lof, B. Straatman, and D. B. Layzell, “Meeting Canada’s freight transportation GHG challenge.” Canadian Energy Systems Analysis Research, 2017.

[252] T. R. Board, *Policy Options for Intermodal Freight Transportation: Special Report 252*. Washington, DC: The National Academies Press, 1998.

[253] J. Sugawara, "Port and hinterland network: a case study of the crescent corridor intermodal freight program in the us," *Transportation Research Procedia*, vol. 25, pp. 916–927, 2017.

[254] S. B. Goddard, "The road to now," *The ANNALS of the American Academy of Political and Social Science*, vol. 553, no. 1, pp. 30–41, 1997.

[255] M. F. Gorman, "Evaluating the public investment mix in us freight transportation infrastructure," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 1, pp. 1–14, 2008.

[256] National Energy Board, "Canada's Energy Future 2016: Province and Territory Outlooks." Government of Canada, 2017. Available at `https://www.neb-one.gc.ca/nrg/ntgrtd/ftr/2016pt/prvnc-trrtrl-cmprsn-eng.html`.

[257] C. S. Decker and M. E. Wohar, "Determinants of state diesel fuel excise tax rates: the political economy of fuel taxation in the United States," *The Annals of Regional Science*, vol. 41, no. 1, pp. 171–188, 2007.

[258] J. Stein, "Per mile fee on heavy trucks emerges as tolling alternative for Wisconsin road funding." Available at `https://www.jsonline.com/story/news/politics/2017/06/13/per-mile-fee-heavy-trucks-emerges-tolling-alternative-wisconsin-road-funding/390875001/`, 2017.

[259] Canadian Shipper, "Truck tolls may come to QEW," 2003. Available at `https://www.canadianshipper.com/transportation-and-logistics/truck-tolls-may-come-to-qew/1000026968/`.

[260] M. N. Taptich and A. Horvath, "Freight on a low-carbon diet: Accessibility, freightsheds, and commodities," *Environmental Science & Technology*, vol. 49, no. 19, pp. 11321–11328, 2015.

[261] CDM Smith for Pennsylvania Department of Transportation, "Pennsylvania comprehensive freight movement plan." PA On Track, 2015.

[262] Association of American Railroads, "Railroads and coal," July 2016.

[263] G. Blauwens, N. Vandaele, E. Van de Voorde, B. Vernimmen, and F. Witlox, "Towards a modal shift in freight transport? a business logistics analysis of some policy measures," *Transport reviews*, vol. 26, no. 2, pp. 239–251, 2006.

[264] *Evaluation of the Marco Polo Programme 2003-2010: Final Report*. Europe Economics, 2011.

[265] "Directive 1999/62/EC of the European Parliament and of the Council of 17 June 1999 on the charging of heavy goods vehicles for the use of certain infrastructures.," 1999.

[266] Toll Collect, "Vehicles subject to toll." Available at `https://www.toll-collect.de/en/toll_collect/rund_um_die_maut/mautpflichtige_fahrzeuge/mautpflichtige_fahrzeuge.html`.

[267] GO Maut, "Tarife 2018." Available in German at `https://www.go-maut.at/portal/portal`.

[268] Schweizer Eidgenossenschaft, "LSVA - Allgemeines / Tarife." Available in German at `https://www.ezv.admin.ch/ezv/de/home/information-firmen/transport--reisedokument--strassenabgaben/schwerverkehrsabgaben--lsva-und-psva-/lsva---allgemeines---tarife.html`.

[269] European Comission Mobility and Transport, "TENtec Interactive Map Viewer." Available at `http://ec.europa.eu/transport/infrastructure/tentec/tentec-portal/map/maps.html`.

[270] ERTMS, "The European Rail Traffic Management System." Available at `http://www.ertms.net`.

[271] Shift2Rail. Available at `https://shift2rail.org`.

[272] COMCIS, "Mobilising global freight data." Available at `http://www.comcis.eu`.

[273] *COMCIS: Taking the extended gateway to the next level.* SEVENTH FRAMEWORK PROGRAMME, European Union.

[274] National Transport Development Policy Committee, *India Transport Report: Moving India to 2032*. Routledge, 2014.

[275] OECD, "Real GDP forecast (indicator)." 10.1787/1f84150b-en (Accessed on 13 February 2018), 2018.

[276] P. Pangotra and P. Shukla, "Infrastructure for low carbon transport in India: a case study of the Delhi-Mumbai dedicated freight corridor," *UNEP*, 2012.

[277] A. Budhkar and S. Das, "Finding trend of advanced ticket booking in Indian railways," *Transportation Research Procedia*, vol. 25, pp. 4826–4835, 2017.

[278] S. Hanaoka and M. B. Regmi, "Promoting intermodal freight transport through the development of dry ports in Asia: An environmental perspective," *IATSS Research*, vol. 35, no. 1, pp. 16–23, 2011.

[279] S. Ramachandran, M. S. Nakhava, and K. Pratik, "Logistics in India: Challenges and Scope," *International Review of Research in Emerging Markets and the Global Economy (IRREM)*, vol. 1, no. 2, pp. 364 – 377, 2015.

[280] P. D. Chaudhury, "Modal split between rail and road modes of transport in India," *Vikalpa*, vol. 30, no. 1, pp. 17–34, 2005.

[281] Government of India, Ministry of Railways, Directorate of Statistics and Economics, "Key Statistics (1970-71 to 2012-13)." Available at `http://www.indianrailways.gov.in/railwayboard/uploads/directorate/stat_econ/downloads/Data_Bank.pdf`.

[282] "Power Sector at a Glance ALL INDIA | Government of India | Ministry of Power."

[283] IEA, UIC, "Railway handbook 2016," 2016.

[284] World Bank, "Green Signal for Faster Development: India's New Freight Corridor," February 2017.

[285] "Welcome to DFCCIL." Available at http://dfccil.gov.in.

[286] S. Sinha and V. Sarma, "Indian Railways: On the fast track to growth." McKinsey&Company, September 2016. Available at https://www.mckinsey.com/~/media/McKinsey/Global%20Themes/India/Indian%20Railways%20on%20the%20fast%20track%20to%20growth/Indian%20Railways%20On%20the%20Fast%20Track%20to%20Growth.ashx.

[287] World Bank, "Developing India's First Modern Inland Waterway." Available at http://www.worldbank.org/en/country/india/brief/developing-india-first-modern-inland-waterway.

[288] World Bank, "World Bank Approves $375 Million to Help India Develop its First Modern Waterway." Available at http://www.worldbank.org/en/news/press-release/2017/04/12/world-bank-approves-usd375-million-help-india-develop-its-first-modern-waterway.

[289] L. Malik, G. Tiwari, and D. Mohan, "Promoting low carbon transport in India: Assessment of Heavy Duty Vehicle Characteristics in Delhi," *UNEP DTU Partnership, Centre on Energy, Climate and Sustainable Development*, 2015.

[290] R. Dandotiya, R. Nath Banerjee, B. Ghodrati, and A. Parida, "Optimal pricing and terminal location for a rail–truck intermodal service–a case study," *International Journal of Logistics Research and Applications*, vol. 14, no. 5, pp. 335–349, 2011.

[291] "GST regime: Freight charges expected to drop 4-10%, reckons IFTRT," *The Hindu Business Line*, May 2017.

[292] "GST cuts truck travel time by 25%: Look who all are smiling - The Economic Times."

[293] GIZ, "Urban Freight and Logistics : The State of Practices in India," 2016.

[294] S. Saraf, "India coal: transport bottlenecks as demand is expected to rise," May 2015. Available at https://www.platts.com/news-feature/2015/coal/india-coal-transport/index.

[295] Department of Industry and Science, "Coal in India 2015,"

[296] United Nations ESCAP, "Transit Transport Issues in Landlocked and Transit Developing Countries," *Landlocked Developing Countries Series, No. 1*, 2003.

[297] M. L. Faye, J. W. McArthur, J. D. Sachs, and T. Snow, "The challenges facing landlocked developing countries," *Journal of Human Development*, vol. 5, no. 1, pp. 31–68, 2004.

[298] C. Grigoriou, *Landlockedness, infrastructure and trade: new estimates for Central Asian countries*, vol. 4335. World Bank Publications, 2007.

[299] Z. Raimbekov, B. Syzdykbayeva, D. Zhenskhan, P. Bayneeva, and Y. Amirbekuly, "Study of the state of logistics in Kazakhstan: prospects for development and deployment of transport and logistics centres," *Transport Problems*, vol. 11, no. 4, pp. 57–71, 2016.

[300] Economist, "New rail routes between China and Europe will change trade patterns," *Economist*, September 2017.

[301] United Nations ESCAP, "Trans-asian railway." Available at http://www.unescap.org/our-work/transport/trans-asian-railway/about.

[302] United Nations, "3. Intergovernmental Agreement on Dry Ports," 2013. Available at https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XI-E-3&chapter=11&clang=_en.

[303] D. M. Z. Islam, T. H. Zunder, R. Jackson, N. Nesterova, and A. Burgess, "The potential of alternative rail freight transport corridors between Central Europe and China," *Transport Problems*, vol. 8, no. 4, pp. 45–57, 2013.

[304] W. Shepard, "How European Wine Is Now Going To China Aboard Silk Road Trains," *Forbes*, July 2017. Available at https://www.forbes.com/sites/wadeshepard/2017/07/23/europe-on-the-new-silk-road-european-wine-can-now-be-shipped-to-china-by-rail/.

[305] African Union, "Programme for infrastructure development in Africa," *Interconnecting, Intergrating & transforming a continent*, vol. 2040, 2010.

[306] Y. Motomura, *The Research on the Cross-Border Transport Infrastructure: Phase 3*. Japan International Cooperation Agency, PADECO Co., Ltd., Mitsubishi UFJ Research and Consulting Co., Ltd, March 2009.

[307] The Economist, "Puffed out," *The Economist*, 2016.

[308] *Moroccan Green Logistics: Sustainable development at the heart of logistics dynamics in Morocco*. Agence Marocaine de Développement de la Logístique, 2016.

[309] Northern Corridor Transit and Transport Coordination Authority (NCTTCA), "Northern Corridor Green Freight Programme: For a Competitive and Sustainable Economic Corridor." http://www.ccacoalition.org/en/resources/northern-corridor-green-freight-strategy-competitive-and-sustainable-economic-corridor, 2017.

[310] L. Kuo, "Kenya's $3.2 billion Nairobi-Mombasa rail line opens with help from China," *Quartz Africa*, June 2017. Available at https://qz.com/996255/kenyas-3-2-billion-nairobi-mombasa-rail-line-opens-with-help-from-china/.

[311] Railway Technology, "Mombasa-Nairobi Standard Gauge Railway Project." Available at http://www.railway-technology.com/projects/mombasa-nairobi-standard-gauge-railway-project/.

[312] J. H. Havenga, Z. P. Simpson, D. King, A. de Bod, and M. Braun, "Logistics Barometer South Africa," *Stellenbosch University*, 2016. Available at https://www.sun.ac.za/english/faculty/economy/logistics/Pages/logisticsbarometer.aspx.

[313] J. H. Havenga and Z. P. Simpson, "Freight logistics' contribution to sustainability: Systemic measurement facilitates behavioural change," *Transportation Research Part D: Transport and Environment*, 2016.

[314] "Asian Development Bank to Launch a Transport DataBank for the Asian and Pacific Region," October 2017. Available at https://trimis.ec.europa.eu/news/asian-development-bank-launch-transport-databank-asian-and-pacific-region.

[315] S. Bakker, K. Dematera Contreras, M. Kappiantari, N. A. Tuan, M. D. Guillen, G. Gunthawong, M. Zuidgeest, D. Liefferink, and M. van Maarseveen, "Low-Carbon Transport Policy in Four ASEAN Countries: Developments in Indonesia, the Philippines, Thailand and Vietnam," *Sustainability*, vol. 9, no. 7, p. 1217, 2017.

[316] S. Gota and S. K. Anthapur, "Advancing green freight in Bangladesh: A background paper," October 2015. Available at http://cleanairasia.org/wp-content/uploads/2015/12/Advancing-Green-Freight-in-Bangladesh-A-Background-Paper.pdf.

[317] M. B. Regmi and S. Hanaoka, "Assessment of modal shift and emissions along a freight transport corridor between Laos and Thailand," *International Journal of Sustainable Transportation*, vol. 9, no. 3, pp. 192–202, 2015.

[318] D. Park, N. S. Kim, H. Park, and K. Kim, "Estimating trade-off among logistics cost, CO2 and time: A case study of container transportation systems in Korea," *International Journal of Urban Sciences*, vol. 16, no. 1, pp. 85–98, 2012.

[319] Ministry of Finance Japan, "Trade Statistics of Japan." Available at http://www.customs.go.jp/toukei/.

[320] G. Wilmsmeier and T. Spengler, "The Evolution of Modal Split in Freight Transport in South America, 2000-2013," 2015.

[321] J. A. C. Fernandes, "Why Latin America needs a new approach to trade." World Economic Forum, May 2016. Available at https://www.weforum.org/agenda/2016/05/why-latin-america-needs-a-new-approach-to-trade/.

[322] M. C. Pachon, G. Araya, D. Saslavsky, and J. S. Schwartz, "Road freight in Central America : five explanations to high costs of service provision - background paper (English)," *Background Paper, Sustainable Development Department, Latin America and the Caribbean Region, Banco Mundial, Washington, DC, junio*, 2012.

[323] S. Romero, "China's Ambitious Rail Projects Crash Into Harsh Realities in Latin America," *The New York Times*, October 2015. Available at https://www.nytimes.com/2015/10/04/world/americas/chinas-ambitious-rail-projects-crash-into-harsh-realities-in-latin-america.html.

[324] L. Goy, "Railway to nowhere shows Brazil's infrastructure woes." Available at https://www.reuters.com/article/us-brazil-infrastructure-railway/railway-to-nowhere-shows-brazils-infrastructure-woes-idUSKBN1441XO., 2016.

[325] International Transport Forum, "Freight Railway Development in Mexico," *OECD*, 2014.

[326] G. Stargardter, "Trains, planes and automobiles: Mexico rail freight comes of age," *Reuters*, 2013.

[327] Agencia Nacional de Transportes Aquaviários Brazil, "Movimentação de Carga Geral - Mercado Interno," 2016. Available at http://www.transportes.gov.br/images/BIT_TESTE/Carga/Carga_-_1.pdf.

[328] *Logistics Observatory for Chile: Strengthening Policies for Competitiveness.* International Transport Forum, 2016.

[329] *Plan de Impulso a la carga ferroviaria.* Subsecretaría de Transportes, 2013.

[330] M. R. Brooks and G. Wilmsmeier, "A Chilean Maritime Highway: Is It a Possible Domestic Transport Option?," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2611, pp. 32–40, 2017.

[331] UNESCWA, *Transport for Sustainable Development in the Arab Region: Measures, progress achieved, challenges and policy framework.* October 2009.

[332] National Media Council, Department of External Information, "United Arab Emirates Yearbook 2013," 2013. Available at https://www.mofa.gov.ae/EN/Documents/3556.pdf.

[333] Etihad Rail, "Project Update." Available at http://www.etihadrail.ae/en/project/projectupdate.

[334] *Freightline 1 - Australian freight transport overview.* Department of Infrastructure and Regional Development, Canberra, Commonwealth of Australia, 2014.

[335] Bureau of Infrastructure, Transport and Regional Economics, "Truck productivity: sources, trends and future prospects," *Report 123, Canberra, ACT*, 2011.

[336] J. Asuncion, S. Rendall, R. Murray, and S. Krumdieck, "New Zealand intermodal freight network and the potential for mode shifting," 2012.

[337] H.-C. Kim, A. Nicholson, and D. Kusumastuti, "Analysing freight shippers' mode choice preference heterogeneity using latent class modelling," *Transportation research procedia*, vol. 25, pp. 1109–1125, 2017.

[338] New Zealand Government, NZ Transport Agency, "Road user charges." Available at https://www.nzta.govt.nz/assets/resources/road-user-charges/docs/road-user-charges-handbook.pdf., 2017.

[339] *Road freight transport sector and emerging competitive dynamics.* Trade Related Technical Assistance Programme, 2016.

[340] M. Zhang, M. Janic, and L. Tavasszy, "A freight transport optimization model for integrated network, service, and policy design," *Transportation Research Part E: Logistics and Transportation Review*, vol. 77, pp. 61–76, 2015.

[341] Eurostat, " Common Questionnaire for Inland Transport Statistics." Available at http://ec.europa.eu/eurostat/cache/metadata/en/rail_if_esms.htm.

[342] L. H. Kaack, P. Vaishnav, M. G. Morgan, I. L. Azevedo, and S. Rai, "Decarbonizing intraregional freight systems with a focus on modal shift," *Environmental Research Letters*, vol. 13, no. 8, p. 083001, 2018.

[343] US Federal Highway Administration, "Traffic Monitoring Guide," 2016.

[344] L. W. Sommer, T. Schuchert, and J. Beyerer, "Fast deep vehicle detection in aerial images," in *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, pp. 311–319, IEEE, 2017.

[345] Q. Jiang, L. Cao, M. Cheng, C. Wang, and J. Li, "Deep neural networks-based vehicle detection in satellite images," in *Bioelectronics and Bioinformatics (ISBB), 2015 International Symposium on*, pp. 184–187, IEEE, 2015.

[346] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geoscience and remote sensing letters*, vol. 11, no. 10, pp. 1797–1801, 2014.

[347] Z. Deng, H. Sun, S. Zhou, J. Zhao, and H. Zou, "Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017.

[348] T. N. Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, "A large contextual dataset for classification, detection and counting of cars with deep learning," in *European Conference on Computer Vision*, pp. 785–800, Springer, 2016.

[349] K. Liu and G. Mattyus, "Fast multiclass vehicle detection on aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1938–1942, 2015.

[350] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, 2010.

[351] S. Ø. Larsen, H. Koren, and R. Solberg, "Traffic monitoring using very high resolution satellite imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 75, no. 7, pp. 859–869, 2009.

[352] L. Eikvil, L. Aurdal, and H. Koren, "Classification-based vehicle detection in high-resolution satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 65–72, 2009.

[353] K. Kanistras, G. Martins, M. J. Rutherford, and K. P. Valavanis, "Survey of unmanned aerial vehicles (uavs) for traffic monitoring," in *Handbook of unmanned aerial vehicles*, pp. 2643–2666, Springer, 2015.

[354] A. Gerhardinger, D. Ehrlich, and M. Pesaresi, "Vehicles detection from very high resolution satellite imagery," *International Archives of Photogrammetry and Remote Sensing*, vol. 36, no. Part 3, p. W24, 2005.

[355] R. Bridgelall, J. B. R. PI, and D. D. Tolliver, "Remote sensing of multimodal transportation systems," tech. rep., Mountain Plains Consortium, 2016.

[356] B. Bowen, K. Vlasek, C. Webb, *et al.*, "An assessment of remote sensing applications in transportation," in *45th Annual Transportation Research Forum, Evanston, Illinois, March 21-23, 2004*, no. 208247, Transportation Research Forum, 2004.

[357] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," in *IEEE CVPR*, 2017.

[358] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.

[359] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[360] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[361] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*, pp. 21–37, Springer, 2016.

[362] D. Lam, R. Kuzma, K. McGee, S. Dooley, M. Laielli, M. Klaric, Y. Bulatov, and B. McCord, "xview: Objects in context in overhead imagery," *arXiv preprint arXiv:1802.07856*, 2018.

[363] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 187–203, 2016.

[364] New York State Thruway Authority, *NYSTA Route System*. Albany, New York: New York State Thruway Authority, GIS, 2018. Available at `https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=440`.

[365] New York State Thruway Authority, "NYS Thruway Origin and Destination Points for All Vehicles - 1 Hour Intervals: 2016," February 2018. Available at `https://catalog.data.gov/dataset/nys-thruway-origin-and-destination-points-for-all-vehicles-1-hour-intervals-2016`.

[366] New York State Thruway Authority, "NYS Thruway Origin and Destination Points for All Vehicles - 1 Hour Intervals: 2017," February 2018. Available at `https://catalog.data.gov/dataset/nys-thruway-origin-and-destination-points-for-all-vehicles-1-hour-intervals-2017`.

[367] R. Krile, F. Todt, and J. Schroeder, "Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic Estimation," Tech. Rep. FHWA-PL-16-012, Federal Highway Administration, Washington, D.C., United States, 2016.

[368] Departamento Nacional de Infraestrutura de Transportes , “Plano Nacional de Contagem de Tráfego - Dados de Tráfego.” Available at `http://servicos.dnit.gov.br/dadospnct/DadosTrafego`.

[369] Departamento Nacional de Infraestrutura de Transportes, “Shapefile or roads in Brazil.” Available at `http://www.dnit.gov.br/mapas-multimodais/shapefiles`.

[370] L. N. Joppa, “The case for technology investments in the environment.,” *Nature*, vol. 552, no. 7685, pp. 325–328, 2017.

[371] B. Berendt, “Ai for the common good?! pitfalls, challenges, and ethics pen-testing,” 10 2018.

[372] Z. Obermeyer and E. J. Emanuel, “Predicting the future — big data, machine learning, and clinical medicine,” *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016. PMID: 27682033.

[373] M. De-Arteaga, W. Herlands, D. B. Neill, and A. Dubrawski, “Machine learning for the developing world,” *ACM Transactions on Management Information Systems (TMIS)*, vol. 9, no. 2, p. 9, 2018.

[374] A. Karpatne and V. Kumar, “Big data in climate: Opportunities and challenges for machine learning,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’17, (New York, NY, USA), pp. 21–22, ACM, 2017.

[375] C. Monteleoni, G. A. Schmidt, and S. McQuade, “Climate informatics: Accelerating discovering in climate science with machine learning,” *Computing in Science Engineering*, vol. 15, pp. 32–40, Sep. 2013.

[376] S. Rasp, M. S. Pritchard, and P. Gentine, “Deep learning to represent subgrid processes in climate models,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 39, pp. 9684–9689, 2018.

[377] J. D. Ford, S. E. Tilleard, L. Berrang-Ford, M. Araos, R. Biesbroek, A. C. Lesnikowski, G. K. MacDonald, A. Hsu, C. Chen, and L. Bizikova, “Opinion: Big data has big potential for applications to climate change adaptation,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 39, pp. 10729–10732, 2016.

[378] T. M. Mitchell *et al.*, “Machine learning. 1997,” *Burr Ridge, IL: McGraw Hill*, vol. 45, no. 37, pp. 870–877, 1997.

[379] S. Li, S. Dragicevic, F. A. Castro, M. Sester, S. Winter, A. Coltekin, C. Pettit, B. Jiang, J. Haworth, A. Stein, *et al.*, “Geospatial big data handling theory and methods: A review and research challenges,” *ISPRS journal of Photogrammetry and Remote Sensing*, vol. 115, pp. 119–133, 2016.

[380] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” *Mobile Networks and Applications*, vol. 19, pp. 171–209, Apr 2014.

[381] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, “Big data in health care: Using analytics to identify and manage high-risk and high-cost patients,” *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014. PMID: 25006137.

[382] M. Hilbert, “Big data for development: A review of promises and challenges,” *Development Policy Review*, vol. 34, no. 1, pp. 135–174, 2016.

[383] A. Ali, J. Qadir, R. u. Rasool, A. Sathiaseelan, A. Zwitter, and J. Crowcroft, “Big data for development: applications and techniques,” *Big Data Analytics*, vol. 1, p. 2, Jul 2016.

[384] C. M. Bishop, “Pattern recognition and machine learning,” tech. rep., 2006.

[385] C. Beckel, L. Sadamori, and S. Santini, “Automatic socio-economic classification of households using electricity consumption data,” in *Proceedings of the Fourth International Conference on Future Energy Systems*, e-Energy ’13, (New York, NY, USA), pp. 75–86, ACM, 2013.

[386] A. Kell, A. S. McGough, and M. Forshaw, “Segmenting residential smart meter data for short-term load forecasting,” *Proceedings of the Ninth International Conference on Future Energy Systems - e-Energy ’18*, 2018.

[387] M. Zeifman and K. Roth, “Nonintrusive appliance load monitoring: Review and outlook,” *IEEE Transactions on Consumer Electronics*, vol. 57, pp. 76–84, February 2011.

[388] J. Z. Kolter and T. Jaakkola, “Approximate inference in additive factorial hmms with application to energy disaggregation,” in *Artificial Intelligence and Statistics*, pp. 1472–1482, 2012.

[389] J. Z. Kolter, S. Batra, and A. Y. Ng, “Energy disaggregation via discriminative sparse coding,” in *Advances in Neural Information Processing Systems*, pp. 1153–1161, 2010.

[390] D. Srinivasan, W. S. Ng, and A. C. Liew, “Neural-network-based signature recognition for harmonic source identification,” *IEEE Transactions on Power Delivery*, vol. 21, pp. 398–405, Jan 2006.

[391] F. Burlig, C. Knittel, D. Rapson, M. Reguant, and C. Wolfram, “Machine learning from schools about energy efficiency,” tech. rep., National Bureau of Economic Research, 2017.

[392] B. D’Amico, R. Myers, J. Sykes, E. Voss, B. Cousins-Jenvey, W. Fawcett, S. Richardson, A. Kermani, and F. Pomponi, “Machine learning for sustainable structures: A call for data,” *Structures*, vol. 19, pp. 1 – 4, 2019.

[393] S. Seyedzadeh, F. Rahimian, I. Glesk, and M. Roper, “Machine learning for estimation of building energy consumption and performance: a review,” *Visualization in Engineering*, vol. 6, no. 1, 2018. cited By 0.

[394] S. Papadopoulos and C. E. Kontokosta, “Grading buildings on energy performance using city benchmarking data,” *Applied Energy*, vol. 233-234, pp. 244 – 253, 2019.

[395] A. M. Tureczek, P. S. Nielsen, H. Madsen, and A. Brun, “Clustering district heat exchange stations using smart meter consumption data,” *Energy and Buildings*, vol. 182, pp. 144 – 158, 2019.

[396] C. Mora, B. Dousset, I. R. Caldwell, F. E. Powell, R. C. Geronimo, C. R. Bielecki, C. W. Counsell, B. S. Dietrich, E. T. Johnston, L. V. Louis, *et al.*, “Global risk of deadly heat,” *Nature Climate Change*, vol. 7, no. 7, p. 501, 2017.

[397] H. Omrani, "Predicting travel mode of individuals by machine learning," *Transportation Research Procedia*, vol. 10, pp. 840–849, 2015.

[398] D. Nam, H. Kim, J. Cho, and R. Jayakrishnan, "A model based on deep learning for predicting travel mode choice," in *Proceedings of the Transportation Research Board 96th Annual Meeting Transportation Research Board, Washington, DC, USA*, pp. 8–12, 2017.

[399] J. Hagenauer and M. Helbich, "A comparative study of machine learning classifiers for modeling travel mode choice," *Expert Systems with Applications*, vol. 78, pp. 273 – 282, 2017.

[400] T. Seo, T. Kusakabe, H. Gotoh, and Y. Asakura, "Interactive online machine learning approach for activity-travel survey," *Transportation Research Part B: Methodological*, 2017.

[401] Y. Sun, Z. Jiang, J. Gu, M. Zhou, Y. Li, and L. Zhang, "Analyzing high speed rail passengers' train choices based on new online booking data in china," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 96 – 113, 2018.

[402] M. S. Ghaemi, B. Agard, M. Trépanier, and V. Partovi Nia, "A visual segmentation method for temporal smart card data," *Transportmetrica A: Transport Science*, vol. 13, pp. 381–404, Jan 2017.

[403] W. Ma and Z. S. Qian, "Estimating multi-year 24/7 origin-destination demand using high-granular multi-source traffic data," *Transportation Research Part C: Emerging Technologies*, vol. 96, pp. 96 – 121, 2018.

[404] G. Spiliopoulos, D. Zissis, and K. Chatzikokolakis, "A big data driven approach to extracting global trade patterns," in *Mobility Analytics for Spatio-Temporal and Social Data* (C. Doulkeridis, G. A. Vouros, Q. Qu, and S. Wang, eds.), (Cham), pp. 109–121, Springer International Publishing, 2018.

[405] J. Farrell, "Corporate funding and ideological polarization about climate change," *Proceedings of the National Academy of Sciences*, 2015.

[406] L. F. Beiser-McGrath and R. A. Huber, "Assessing the relative importance of psychological and demographic factors for predicting climate and environmental attitudes," *Climatic Change*, vol. 149, pp. 335–347, Aug 2018.

[407] M. B. Alvi, N. A. Mahoto, M. Alvi, M. A. Unar, and M. A. Shaikh, "Hybrid classification model for twitter data - a recursive preprocessing approach," in *2018 5th International Multi-Topic ICT Conference (IMTIC)*, pp. 1–6, April 2018.

[408] A. Reyes-Menendez, J. R. Saura, and C. Alvarez-Alonso, "Understanding #WorldEnvironmentDay User Opinions in Twitter: A Topic-Based Sentiment Analysis Approach," *International Journal of Environmental Research and Public Health*, vol. 15, no. 11, 2018.

[409] R. Szabo and A. Groza, "Analysing debates on climate change with textual entailment and ontologies," in *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 39–46, Sep. 2017.

[410] F. Creutzig, G. Baiocchi, R. Bierkandt, P.-P. Pichler, and K. C. Seto, "Global typology of urban energy use and potentials for an urbanization mitigation wedge," *Proceedings of the National Academy of Sciences*, vol. 112, no. 20, pp. 6283–6288, 2015.

[411] R. Ganzenmüller, P. Pradhan, and J. P. Kropp, "Sectoral performance analysis of national greenhouse gas emission inventories by means of neural networks," *Science of The Total Environment*, vol. 656, pp. 80 – 89, 2019.

[412] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11–28, 2016.

[413] M. Marsden, K. McGuinness, S. Little, C. E. Keogh, and N. E. O'Connor, "People, penguins and petri dishes: Adapting object counting models to new visual domains and object types without forgetting," 11 2017.

[414] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016.

[415] J. M. Malof, K. Bradbury, L. M. Collins, and R. G. Newell, "Automatic detection of solar photovoltaic arrays in high resolution aerial imagery," *Applied energy*, vol. 183, pp. 229–240, 2016.

[416] J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal, "Deepsolar: A machine learning framework to efficiently construct a solar deployment database in the united states," *Joule*, vol. 2, no. 12, pp. 2605–2617, 2018.

[417] A. Bogomolov, B. Lepri, R. Larcher, F. Antonelli, F. Pianesi, and A. Pentland, "Energy consumption prediction using people dynamics derived from cellular network data," *EPJ Data Science*, vol. 5, no. 1, 2016. cited By 8.

[418] K. Boman, "Machine Learning Could Enhance Methane Leak Detection Technology," *Rigzone*, November 2016. Available at `https://www.rigzone.com/news/oil_gas/a/147295/machine_learning_could_enhance_methane_leak_detection_technology/?all=hg2`.

[419] Microsoft, "Computer generated building footprints for the United States." Available at `https://github.com/Microsoft/USBuildingFootprints`.

[420] G. Gröger, T. H. Kolbe, A. Czerwinski, and C. Nagel, *OpenGIS City Geography Markup Language (CityGML) Encoding Standard, Version 1.0.0.* Open Geospatial Consortium, 1.0.0 ed., 2008.

[421] Z. Lu, J. Im, J. Rhee, and M. Hodgson, "Building type classification using spatial and landscape attributes derived from lidar remote sensing data," *Landscape and Urban Planning*, vol. 130, pp. 134 – 148, 2014.

[422] A. Henn, C. Römer, G. Gröger, and L. Plümer, "Automatic classification of building types in 3d city models," *GeoInformatica*, vol. 16, pp. 281–306, Apr 2012.

[423] C. Geiß, H. Taubenböck, M. Wurm, T. Esch, M. Nast, C. Schillings, and T. Blaschke, "Remote sensing-based characterization of settlement structures for assessing local potential of district heat," *Remote Sensing*, vol. 3, pp. 1447–1471, Jul 2011.

[424] Rudiyanto, B. Minasny, B. I. Setiawan, S. K. Saptomo, and A. B. McBratney, "Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands," *Geoderma*, vol. 313, pp. 25 – 40, 2018.

[425] S. L. Jardine and J. V. Siikamäki, "A global predictive model of carbon in mangrove soils," *Environmental Research Letters*, vol. 9, no. 10, p. 104013, 2014.

[426] T. Hengl, J. Mendes de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, and et al., "Soilgrids250m: Global gridded soil information based on machine learning," *PLOS ONE*, vol. 12, p. e0169748, Feb 2017.

[427] J. Mascaro, G. P. Asner, D. E. Knapp, T. Kennedy-Bowdoin, R. E. Martin, C. Anderson, M. Higgins, and K. D. Chadwick, "A tale of two "forests": Random forest machine learning aids tropical forest carbon mapping," *PLoS ONE*, vol. 9, p. e85993, Jan 2014.

[428] J.-F. Exbrayat and M. Williams, "Quantifying the net contribution of the historical amazonian deforestation to climate change," *Geophysical Research Letters*, vol. 42, no. 8, pp. 2968–2976, 2015. cited By 9.

[429] M. Mayr, K. Vanselow, and C. Samimi, "Fire regimes at the arid fringe: A 16-year remote sensing perspective (2000–2016) on the controls of fire activity in namibia from spatial predictive models," *Ecological Indicators*, vol. 91, pp. 324 – 337, 2018.

[430] L. H. Kaack, J. Apt, M. G. Morgan, and P. McSharry, "Empirical prediction intervals improve energy forecasting," *Proceedings of the National Academy of Sciences*, vol. 114, no. 33, pp. 8752–8757, 2017.

[431] J. Chakraborty and S. Banerji, "Evaluating predictive models for solar energy growth in the us states and identifying the key drivers," *IOP Conference Series: Earth and Environmental Science*, vol. 127, p. 012002, Mar 2018.

[432] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1 – 8, 2012.

[433] M. Wytock and Z. Kolter, "Sparse gaussian conditional random fields: Algorithms, theory, and application to energy forecasting," in *Proceedings of the 30th International Conference on Machine Learning* (S. Dasgupta and D. McAllester, eds.), vol. 28 of *Proceedings of Machine Learning Research*, (Atlanta, Georgia, USA), pp. 1265–1273, PMLR, 17–19 Jun 2013.

[434] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power Systems*, vol. 29, pp. 1033–1044, May 2014.

[435] P. Pinson and G. Kariniotakis, "Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment," in *2003 IEEE Bologna Powzer Tech conference*, vol. 2, pp. 8–pages, 2003.

[436] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. V. Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912 – 928, 2018.

[437] F. H. Gandoman, F. Raeisi, and A. Ahmadi, "A literature review on estimating of pv-array hourly power under cloudy weather conditions," *Renewable and Sustainable Energy Reviews*, vol. 63, pp. 579 – 592, 2016.

[438] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569 – 582, 2017.

[439] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 528–533, Oct 2011.

[440] J. Li, J. K. Ward, J. Tong, L. Collins, and G. Platt, "Machine learning for solar irradiance forecasting of photovoltaic system," *Renewable Energy*, vol. 90, pp. 542 – 553, 2016.

[441] A. Alzahrani, P. Shamsi, C. Dagli, and M. Ferdowsi, "Solar irradiance forecasting using deep neural networks," *Procedia Computer Science*, vol. 114, pp. 304 – 313, 2017. Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, CAS October 30 – November 1, 2017, Chicago, Illinois, USA.

[442] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *Power Systems, IEEE Transactions on*, vol. 16, no. 1, pp. 44–55, 2001.

[443] H. K. Alfares and M. Nazeeruddin, "Electric load forecasting: Literature survey and classification of methods," *International Journal of Systems Science*, vol. 33, pp. 23–34, Jan 2002.

[444] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914 – 938, 2016.

[445] B. Bhattacharya and A. Sinha, "Deep fault analysis and subset selection in solar power grids," 11 2017.

[446] M. Cenek, R. Haro, B. Sayers, and J. Peng, "Climate change and power security: Power load prediction for rural electrical microgrids using long short term memory and artificial neural networks," *Applied Sciences*, vol. 8, p. 749, May 2018.

[447] F. Lai, F. MagoulÃšs, and F. Lherminier, "Vapnik's learning theory applied to energy consumption forecasts in residential buildings," *International Journal of Computer Mathematics*, vol. 85, no. 10, pp. 1563–1588, 2008.

[448] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1192 – 1205, 2018.

[449] H. xiang Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586 – 3592, 2012.

[450] M. Aydinalp, V. I. Ugursal, and A. S. Fung, "Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks," *Applied Energy*, vol. 71, no. 2, pp. 87 – 110, 2002.

[451] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Applied Energy*, vol. 123, pp. 168 – 178, 2014.

[452] W. Zhang, C. Robinson, S. Guhathakurta, V. M. Garikapati, B. Dilkina, M. A. Brown, and R. M. Pendyala, "Estimating residential energy consumption in metropolitan areas: A microsimulation approach," *Energy*, vol. 155, pp. 162 – 173, 2018.

[453] D.-S. Kapetanakis, D. Christantoni, E. Mangina, and D. P. Finn, "Evaluation of machine learning algorithms for demand response potential forecasting," in *15th International IBPSA Building Simulation Conference - IBPSA 2017*, 08 2017.

[454] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3 – 19, 2014. Special Issue on Short-term Traffic Flow Forecasting.

[455] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, pp. 871–882, June 2013.

[456] P. Noursalehi, H. N. Koutsopoulos, and J. Zhao, "Real time transit demand prediction capturing station interactions and impact of special events," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 277 – 300, 2018.

[457] E. Mazloumi, G. Rose, G. Currie, and S. Moridpour, "Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 3, pp. 534 – 542, 2011.

[458] R. Regue and W. Recker, "Proactive vehicle routing with inferred demand to solve the bikesharing rebalancing problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 72, pp. 192 – 209, 2014.

[459] Y. Wang, Z. R. Shi, L. Yu, Y. Wu, R. Singh, L. Joppa, and F. Fang, "Deep reinforcement learning for green security games with real-time information," *arXiv preprint arXiv:1811.02483*, 2018.

[460] F. Fang, P. Stone, and M. Tambe, "When security games go green: Designing defender strategies to prevent poaching and illegal fishing.," in *IJCAI*, pp. 2589–2595, 2015.

[461] M. Memarzadeh, M. Pozzi, and J. Zico Kolter, "Optimal planning and learning in uncertain environments for the management of wind farms," *Journal of Computing in Civil Engineering*, vol. 29, no. 5, p. 04014076, 2014.

[462] T. Tulabandhula and C. Rudin, "On combining machine learning with decision making," 04 2011.

[463] C. Hilbe, Š. Šimsa, K. Chatterjee, and M. A. Nowak, "Evolution of cooperation in stochastic games," *Nature*, vol. 559, pp. 246–249, Jul 2018.

[464] A. D. Procaccia, "Cake cutting," *Communications of the ACM*, vol. 56, p. 78, Jul 2013.

[465] N. Gershenfeld, S. Samouhos, and B. Nordman, "Intelligent infrastructure for energy efficiency," *Science*, vol. 327, no. 5969, pp. 1086–1088, 2010.

[466] R. G. Pratt, P. J. Balducci, C. Gerkensmeyer, S. Katipamula, M. C. Kintner-Meyer, T. F. Sanquist, K. P. Schneider, and T. J. Secrest, "The smart grid: an estimation of the energy and co2 benefits," *Pacific Northwest National Laboratory, Washington*, 2010.

[467] S. Giest, "Big data analytics for mitigating carbon emissions in smart cities: opportunities and challenges," *European Planning Studies*, vol. 25, pp. 941–957, Feb 2017.

[468] M. Guttenberg, S. Sripad, and V. Viswanathan, "Evaluating the potential of platooning in lowering the required performance metrics of li-ion batteries to enable practical electric semi-trucks," *ACS Energy Letters*, vol. 2, pp. 2642–2646, Oct 2017.

[469] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine learning*, vol. 79, no. 1-2, pp. 151–175, 2010.

[470] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[471] M.-F. Balcan and R. Urner, "Active learning - modern learning theory," in *Encyclopedia of Algorithms*, 2016.

[472] X. Zhu, "Semi-supervised learning," in *Encyclopedia of machine learning*, pp. 892–897, Springer, 2011.

[473] A. Blum, "On-line algorithms in machine learning," in *Developments from a June 1996 Seminar on Online Algorithms: The State of the Art*, (Berlin, Heidelberg), pp. 306–325, Springer-Verlag, 1998.

[474] J. Stoyanovich, B. Howe, S. Abiteboul, G. Miklau, A. Sahuguet, and G. Weikum, "Fides," *Proceedings of the 29th International Conference on Scientific and Statistical Database Management - SSDBM '17*, 2017.

[475] A. Preece, D. Harborne, D. Braines, R. Tomsett, and S. Chakraborty, "Stakeholders in explainable ai," 10 2018.

[476] N. Diakopoulos, "Accountability in algorithmic decision making," *Commun. ACM*, vol. 59, pp. 56–62, Jan. 2016.

[477] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, (USA), pp. 3323–3331, Curran Associates Inc., 2016.

[478] C. Olah, A. Satyanarayan, I. Johnson, S. Carter, L. Schubert, K. Ye, and A. Mordvintsev, "The building blocks of interpretability," *Distill*, 2018. https://distill.pub/2018/building-blocks.

[479] S. Athey, "Beyond prediction: Using big data for policy problems," *Science*, vol. 355, no. 6324, pp. 483–485, 2017.

[480] S. Athey and G. W. Imbens, "Machine learning methods for estimating heterogeneous causal effects," *Stat*, vol. 1050, no. 5, 2015.

[481] J. Zhao, D. M. Runfola, and P. Kemper, "Quantifying heterogeneous causal treatment effects in world bank development finance projects," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 204–215, Springer, 2017.

[482] L. V. Gelder, P. Das, H. Janssen, and S. Roels, "Comparative study of metamodelling techniques in building energy simulation: Guidelines for practitioners," *Simulation Modelling Practice and Theory*, vol. 49, pp. 245 – 257, 2014.

[483] S. Cao, "A.I. Buzz Is Causing a Tech Labor Shortage in Both Silicon Valley and New York," *Observer*, May 2018. Available at `https://observer.com/2018/05/artificial-intelligence-buzz-tech-labor-shortage/`.

[484] D. G. Victor, D. Zhou, E. H. M. Ahmed, P. K. Dadhich, J. G. J. Olivier, H.-H. Rogner, K. Sheikho, and M. Yamaguchi, *Intrductory Chapter. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2014.

[485] "A matter of trust," *Nature*, vol. 449, pp. 637 EP –, 10 2007.

[486] M. Khoury and J. Ioannidis, "Big data meets public health," *Science*, vol. 346, no. 6213, pp. 1054–1055, 2014.

[487] P. L. Donti, B. Amos, and J. Z. Kolter, "Task-based end-to-end model learning in stochastic optimization," *In Advances in Neural Information Processing Systems*, pp. 5484–5494, 03 2017.

[488] B. Wilder, B. N. Dilkina, and M. Tambe, "Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization," *CoRR*, vol. abs/1809.05504, 2018.

[489] A. N. Elmachtoub and P. Grigas, "Smart "predict, then optimize"," 10 2017.

[490] D. Lazer, R. Kennedy, G. King, and A. Vespignani, "The parable of google flu: traps in big data analysis," *Science*, vol. 343, no. 6176, pp. 1203–1205, 2014.

[491] U.S. Energy Information Administration, "Evaluation of AEO2018 and Previous Reference Case Projections," 2018.

[492] A. T. Anderson, "Differences between Energy Information Administration Energy Forecasts: Reasons and Resolution," *Energy Information Administration: Iusses in Midterm Analysis and Forecasting*, 1996.

[493] N. M. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[494] N. R. Council, *The National Energy Modeling System*. Washington, DC: The National Academies Press, 1992.

[495] V. Arora, "An evaluation of macroeconomic models for use at eia," 2013.

[496] T. Cogley, S. Morozov, and T. J. Sargent, "Bayesian fan charts for uk inflation: Forecasting and sources of uncertainty in an evolving monetary system," *Journal of Economic Dynamics and Control*, vol. 29, no. 11, pp. 1893–1925, 2005.

[497] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications," *Environmental Modelling & Software*, vol. 15, no. 1, pp. 101–124, 2000.

[498] United States Census Bureau, "Commodity Flow Survey," 2018. Available at `https://www.census.gov/econ/cfs/`.

[499] Inter-American Development Bank (IDB), "Freight Transport and Logistics Statistics Yearbook," 2015. Available at `https://publications.iadb.org/handle/11319/6885`.

[500] U.S. Environmental Protection Agency, "Emission Facts: Greenhouse gas emissions from a typical passenger vehicle," *Environmental Protection Agency Office of Transportation and Air Quality*, 2005.

[501] "World Economic Outlook Database," April 2017. Available at `http://www.imf.org/external/pubs/ft/weo/2017/01/weodata/index.aspx`.

[502] Tzutalin, "Labelimg," 2015. Available at `https://github.com/tzutalin/labelImg`.

[503] Thruway Authority, "Interchange/Exit Listings." Available at `https://www.thruway.ny.gov/travelers/interchanges/index.html`.

5

Appendix: Empirical prediction intervals improve energy forecasting

5.1 Data description

All data come from the Annual Energy Outlook [57] and the Retrospective Review [58] of the U.S. Energy Information Administration (EIA). The data set consists of AEO 1982-2016, with historical, or actual, values for 1985-2015. Historical values are taken from the EIA's Retrospective Reviews with the exception of 2014 and 2015 which are taken from AEO 2016 [57]. Historical values for 2015 are the $H = 0$ projections from AEO 2016, which might be updated in the following AEO. Any missing values are linearly interpolated.

Before 1988, the AEO was released in the end of each fiscal year and after 1988 in April of the following year. This renaming decision led to the fact that there is no AEO 1988. For reasons of simplicity, we will use naming conventions based on the AEOs released after 1988. Horizons in our data set range from $H = 0$ to $H = 21$. As the collection of historical data is not complete when the forecasts are issued, AEOs include estimates of the year before the release and a forecast of the year of the release. AEO 2000 for example has estimates for 1999 and 2000. We refer to these estimates as forecast horizons $H = 0$ and $H = 1$ respectively. The number of forecasting errors for each horizon varies from $n_{H=0\cdots3} = 31$ to $n_{H=21} = 1$. As sample sizes are decreasing with larger horizons and the variance of errors is dependent on the sample size, we chose a maximum horizon for the analysis of $H_{max} = 12$, where $n_{H=12} = 19$ and $n_{H=13} = 16$.

The AEO projections are based on the National Energy Modeling System (NEMS). The EIA ensures that projections match across its products. For shorter time horizons (up to two years ahead), the EIA arranges that the NEMS outputs are consistent with the forecasts in the Short-Term Energy Outlook (STEO) [492]. The STEO is based on a different forecasting system and contains forecasts as opposed to projections. This does, however, not impact our analysis.

As the AEO projects a large number of quantities, we restrict ourselves to eighteen select quantities of EIA's Retrospective Review [58]. The quantity names used throughout the paper correspond to the following AEO naming conventions:

1. *Oil Price (nominal dollars):* Imported refiner acquisition cost of crude oil in nominal dollars per barrel; also crude oil spot prices, crude oil prices, world oil price

2. *Oil Price (constant dollars):* Imported refiner acquisition cost of crude oil in constant 2013 dollars per barrel; also crude oil spot prices, crude oil prices, world oil price
3. *Petroleum Cons.:* Total petroleum consumption in million barrels per day; also liquid fuels: primary supply, product supplied: total product supplied, liquid fuel consumption: total, refined petroleum products supplied: total, petroleum product supplied
4. *Oil Production:* Domestic crude oil production in million barrels per day; also liquid fuels: crude oil: domestic production, domestic crude production, production: crude oil, petroleum production: crude oil
5. *Natural Gas Price (nom.):* Natural gas wellhead prices in nominal dollars per thousand cubic feet; also Henry Hub spot price, average lower 48 wellhead price
6. *Natural Gas Price (const.):* Natural gas wellhead prices in constant 2013 dollars per thousand cubic feet; also Henry Hub spot price, average lower 48 wellhead price
7. *Natural Gas Consumption:* Total natural gas consumption in trillion cubic feet; also natural gas: use by sector: total, consumption by sector: total
8. *Natural Gas Production:* Natural gas production in trillion cubic feet; also dry gas production
9. *Coal Price (nom.):* Coal prices to electric generating plants in nominal dollars per million Btu; also delivered prices: electric power
10. *Coal Price (const.):* Coal prices to electric generating plants in constant 2013 dollars per million Btu; also delivered prices: electric power
11. *Coal Consumption:* Total coal consumption in million short tons; also coal supply: use by sector: total, consumption by sector: total, total consumption
12. *Coal Production:* Coal production in million short tons, this includes waste coal supplied; also production: total and waste coal supplied, production: total, coal production

13. *Electricity Price:* Average electricity prices in nominal cents per kilowatt-hour; also end-use prices: all sectors average

14. *Electricity Sales:* Total electricity sales in billion kilowatt-hours; also electricity sales by sector: total, generation by fuel type: total electricity sales

15. *Total Energy Cons.:* Total energy consumption in quadrillion Btu; also energy use: delivered: all sectors: total, delivered energy consumption: all sectors: total, primary energy consumption: total

16. *Residential Energy Cons.:* Total delivered residential energy consumption in quadrillion Btu; also energy use: residential: delivered energy, residential: total

17. *Commercial Energy Cons.:* Total delivered commercial energy consumption in quadrillion Btu; also energy use: commercial: delivered energy, commercial: total

18. *Transportation:* Total delivered transportation energy consumption in quadrillion Btu; also energy use: transport: delivered energy, transportation: total

We excluded total delivered industrial energy consumption, which is a quantity in the Retrospective Review, based on a change in definition by the EIA which we could not correct for. We are able to generate probabilistic forecasts for total energy related carbon dioxide emissions, but we excluded it from the final analysis due to the shorter forecasting record. The EIA only began to publish carbon dioxide emissions in the AEO 1993.

We analyze each quantity to find the most general approach to creating and evaluating the probabilistic forecasts. We use two of the quantities for illustration purposes in the main article: As prices exhibit a larger degree of volatility than other quantities, we chose to include one price forecast and one other quantity. The **natural gas wellhead price** in nominal dollars per 1000 cubic ft. (hereafter natural gas price) is an important factor for investment decisions. The EIA Retrospective Reviews [58] note the large differences of natural gas price projections and historical values. The Retrospective Review published in 2014 describes that natural gas price predictions influence gas consumption and electricity price forecasts, and recently also coal consumption projections [58]. An example with less volatile historical values are the **total electricity sales** in billion kWhs. The EIA points out the large underestimation of electricity sales in the nineties and the effect on the coal consumption forecasts in its 2008 Retrospective Review [58].

In Fig. 5.3 we see the historical actual data and the past AEO reference case projections for the two quantities selected. This figure also shows the historical values and forecasts for coal prices in nominal dollars, which is an outlier quantity regarding many aspects of the analysis.

5.1.1 Additional data adjustments

Some data required unit or definition adjustments to be consistent over the entire analyzed time frame. Typically, these adjustments needed to be made on reference case and scenario projections alike.

Constant dollar prices were converted to 2013 dollars for the analysis. In some instances, nominal dollar price projections needed to be converted using constant dollar price projections or vice versa by EIA's inflation rates, which were given in the AEO reports or inferred from prices that were reported both in constant and nominal dollars.

Since we analyze **oil production** and **petroleum consumption** in million barrels per day, some of the projected values had to be inferred from values provided in million barrels per year.

Natural gas prices were initially reported as the average lower 48 wellhead price in dollars per thousand cubic feet (AEO 1982-2012). Later AEOs replaced this with Henry Hub spot prices in dollars per million Btu. We converted million Btu into thousand cubic feet with the heat content for dry natural gas reported in the respective AEO. We did not take data from the most recent Retrospective Review (released in 2015) for natural gas prices, since it lists natural gas prices to electric generating plants instead of wellhead prices.

We work with **coal prices** in constant or nominal dollars per million Btu. While most of the AEOs report coal prices in these units, some of them in addition include projections in a mass-based unit of dollars per short ton. AEOs 1983-1993 report coal prices in dollars per short ton. We use approximate heat contents from the outlooks for conversion. Since heat contents vary marginally, and we are not provided a factor for every single forecasted year, we assign the heat content from the nearest forecasted value, or interpolate if the year is between two years with given heat content. We added the waste coal supplied to **coal production** for the projections of AEO 2007-2016. Waste coal was listed separately for these outlooks, while it was included in coal production before. This is consistent with the Retrospective Reviews, except for a discrepancy for the AEO 2013. We chose to use the values directly from in the AEO in this case.

5.1.2 AEO scenario data

Scenario values were taken from the AEO reports. To compute the envelope scenarios, we found the maximum and minimum of all scenarios in every forecasted year and assigned those to what we called high and low scenarios. These resulting envelope scenarios do not correspond to a single projection of the AEO. The scenarios do not include the early release reference cases, but for AEO 2016 we included the "reference case without Clean Power Plan".

The AEO 2009 has been updated after it was published. We work with this updated reference case to find the forecasting errors. The scenarios however have not been updated. This results in a general mismatch between the scenarios and the reference case for AEO 2009, which is why we left it out of the test set.

5.2 Error metrics

It is common to refer to the deviation of the forecast from the actual value as *error*. The EIA for example uses this term in its Retrospective Reviews [58]. We work with the relative error for most quantities and transform the relative error for the price quantities into a log-error, which results in a distribution of price forecast errors that is closer to a normal distribution. The analysis is conducted entirely in the relative and log-error metric, but absolute errors could also be used.

5.2.1 Relative errors

We focus on the relative error or percent error in this analysis, because it enables a comparison between forecasts of different quantities. This choice of error however comes with the typical scaling issues of the percentage metric. It is defined as $\epsilon_{rel} = \frac{\hat{y}-y}{y} = \frac{\hat{y}}{y} - 1$, where $\hat{y}$ refers to the forecast and y to the actual value, or observation. The relative errors for all quantities considered in this analysis are displayed in Fig. 5.4. This is the full set of error samples, also containing the horizons we chose to exclude from the analysis because of their lower sample size. The evolution of the errors over the AEO release years, shown in this figure, makes it easy to identify similarities between the quantities. We can for example see, that electricity price forecast errors look very similar to those of coal price forecasts. In this figure, a large vertical spread indicates that those particular AEO years have resulted in large errors across different horizons. Errors of a similar magnitude over several AEO release years that give the impression to be lined up are in most cases from the same observed value, see for example coal consumption.

We view the forecast densities as a distribution of actual values y around the AEO reference case forecast $\hat{y}$. Also scenarios are treated in this metric. A scenario in our analysis is expressed as the percent error of how much the reference case deviates from that scenario y_S, which is in the resulting relative error metric $\epsilon_{S,rel} = \frac{\hat{y}-y_S}{y_S}$. This means that the errors of high scenarios correspond to $\epsilon_S \leq 0$ and low scenarios to $\epsilon_S \geq 0$. The value of an observation in the relative error metric is computed as $\xi = \frac{\hat{y}-y_{obs}}{y_{obs}}$.

We chose to work with the L_1 loss and mean absolute (percentage) errors instead of the L_2 loss. This means we do not use the root mean square error (RMSE), which is the risk function (or the expected value) of the L_2 loss. This risk is minimized by the mean. By squaring the errors, L_2 loss inflates the weight of errors that are larger, which is desired if attention needs to be paid to outliers in the data. On the contrary, here we wish to find an estimate of the central point of the distribution that is robust to outliers, which the mean is not. The risk function of the L_1 loss is instead minimized by the median. Especially when faced with a skewed distribution, as it is the case for many of the error distributions in our analysis, the median is a better estimator of central tendency because it is less affected by outliers. In addition, the CRPS reduces to the absolute error (the relative or log error in our case) for a point forecasts, which makes both these metrics compatible.

5.2.2 Log-errors

Prices are typically described as log-normally distributed [61]. In Q-Q-plots of historical AEO price quantities, we found that the logarithm of the prices follows a normal distribution closer than the untransformed prices. This supports the assumption that the prices, even though they are given as an annual average, are approximately log-normally distributed. We make the additional assumption that also the price forecasts follow a log-normal distribution, and introduce an error transformation.

For the transformation, we draw an analogy to logarithmic returns, a concept from financial theory. The return is defined as $r = \frac{\text{future value}-\text{present value}}{\text{present value}}$. If values are log-normally distributed, the log return $\ln(1+r)$ follows a normal distribution[1]. To transform the relative errors for prices, we use very similar arguments where instead of the return we work with the relative error $\epsilon_{rel} = \frac{\hat{y}}{y} - 1$. This results in the log error $\epsilon_{log} = \ln(1+\epsilon_{rel}) = \ln\left(\frac{\hat{y}}{y}\right) = \ln\hat{y} - \ln y$. We compute all of the comparative statistics in ϵ_{log}. We termed the mean absolute log error MALE.

How the loss function changes if the absolute percentage error APE is transformed into the absolute log error, ALE, can be seen in Fig. 5.1. Here, we define the loss as APE or ALE.

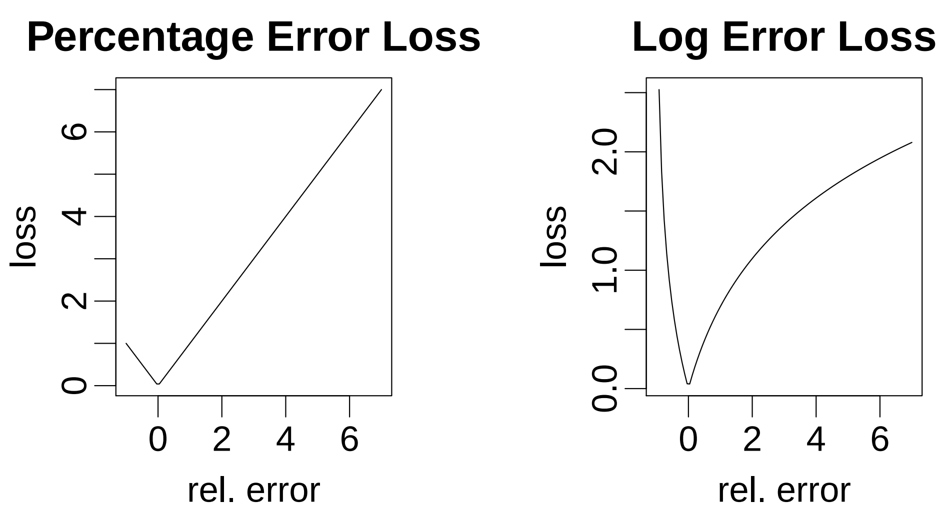


Figure 5.1: Comparison of the two types of errors we work with, with APE on the left and ALE on the right. Loss is defined as the absolute error in the respective metric.

[1]To see this, we start with the definition that if $Z = log(X)$ is normally distributed, X is log normally distributed. So, if FV $\sim$ N and PV $\sim$ N, and noted that $\log(1+r) = \log(1+\frac{\text{FV}-\text{PV}}{\text{PV}}) = \log(\frac{\text{FV}}{\text{PV}}) = \log(\text{FV}) - \log(\text{PV})$ and we know that the sum of normally distributed variables is again a normally distributed variable, we find that $\ln(1+r) \sim N$.

5.3 Summary statistics of the error samples

5.3.1 Normality of the error samples

Here we assess if the errors are normally distributed. Since we use a Gaussian as a parametric density forecast, it is of interest how closely a normal distribution matches the error samples. In addition, we test if the log-errors for the prices are normally distributed, which is the goal of the transformation we apply to price quantities.

We test the assumption that the error samples are normally distributed. We use the Shapiro-Wilk normality test, implemented in the R-package *stats* [62]. The Shapiro-Wilk test is based on the null hypothesis that the sample is normally distributed. The test has the highest power for small sample sizes compared to other tests, even if though the power is fairly low when the true distribution is a symmetric distribution [493]. In Fig. 5.5, we show the test results for the error samples for two different significance levels, 95% and 99%. We see that for most quantities we cannot reject the null that the errors come from a normal distribution. However, there are some quantities, which with 95% confidence do not have normal errors. In particular, petroleum consumption, coal consumption and total energy consumption exhibit deviations from the normal distribution.

To test the assumption that we should transform the price quantities, we also perform the Shapiro-Wilk normality test on price quantities with transformed errors (Fig. 5.5). We see that for almost all price quantities, the log-errors are more likely to be normally distributed than untransformed errors. In further analysis not shown here we found that the log transformation has marginal effect on the production and consumption quantities or makes them less Gaussian. Coal price errors are an exception, which for many horizons are bimodal and therefore clearly not Gaussian, even when transformed to log errors. Electricity price errors behave similarly, as electricity prices are correlated to coal prices. How the distribution is adjusted by the log transformation is shown with histograms in Fig. 5.5. We see that for the example of oil prices, the distribution becomes more Gaussian, whereas the bimodal distribution of coal price errors is largely unaffected by the transformation. Coal prices have been overforecasted for a long period, followed by a period of underforecasting (Fig. 5.3). This resulted in the bimodal error distribution. We also find that changing the confidence level for rejection of the null hypothesis to 99% allows the error samples of many quantities to appear Gaussian for almost all horizons, with the exception of coal prices.

5.3.2 Autocorrelation

We find that autocorrelation of errors is different from quantity to quantity (Fig. 5.6). It is typically lower for smaller horizons, larger horizons all show high correlation that only disappears for long lags. Coal prices and electricity prices have a large autocorrelation even for forecasts with small horizons. This matches the pattern that can be observed for coal prices, where we saw long alternating periods of over and underforecasting, and therefore the errors are more correlated (Fig. 5.3).

In Fig. 5.4, we can see the autocorrelation reflected in the pattern of errors. This figure, as described above, shows the magnitude of the errors over the release year of the AEO that generated the projection. Where we observe a wave pattern, as for example for coal prices, we find that errors of larger horizons are highly correlated from one AEO to another. This pattern is repeated in electricity prices and transportation energy consumption. Quantities with less autocorrelated horizon samples such as residential energy consumption do not exhibit this pattern. In the case of oil production, we find a relatively large autocorrelation for small horizons, which can perhaps be attributed to the recent oil and natural gas boom. The observed values changed systematically and rapidly, which was not picked up by many of the recent AEO projections. This is reflected in the waterfall shape of errors for oil production in Fig. 5.4. Since natural gas production errors were historically larger and more volatile, we do not observe this pattern as clearly here. The pattern of errors that appear lined up, as mentioned in the previous section, does not generally indicate autocorrelation, as this is a result of single outlier observations.

As much as the presence of autocorrelation is a problem for viewing the error series as a random sample, it does not impact the validity of comparing the mean CRPS among the methods. However, for the significance test of improvement of an empirical method over the scenarios, we use the sample of single observation CRPS as a random sample. Here some correlation is to be expected and large correlation could pose a problem. This depends on the autocorrelation of observed values and the AEO forecasts, as well as the similarity of forecast densities from one observation to the other. It is expected to have a similar or lower autocorrelation than the error time series shown in Fig. 5.6. For our purpose, we assume we can view this autocorrelation as negligible.

5.3.3 Grouping the Quantities

In Fig. 5.7, we plot the standard deviation of the error samples against the autocorrelation at a lag of $3yrs$ for every horizon separately. This allows us to potentially identify groups of quantities with similar characteristics. The characteristic form seen in the figure does not change much for an autocorrelation coefficient of a different lag. Most apparent is the large variance of errors of the price quantities. We can identify prices with higher autocorrelation (coal and electricity) and with lower (oil and natural gas). This picture emphasizes that the prices form a distinct group among the quantities. In addition, the standard deviation of price errors has a large spread for the different horizons. The electricity price is the most similar to the other quantities outside this group.

The rest of the quantities has a much lower standard deviation, where zooming in on a section of the plot helps to visualize potential differences. We see that the rest of the quantities are fairly similar in these characteristics. Oil production is somewhat different, in that it has a larger standard deviation at a lower autocorrelation coefficient.

From this and the previous analysis we can conclude that treating the price quantities and the other quantities as two distinct groups, and applying the log transformation only to price errors, seems a valuable approach.

5.4 Details on density forecasting methods

We excluded any historically intractable approaches, i.e. methods, where it is impossible to trace back in retrospect how an analyst would have estimated the uncertainty at the point of decision. A common approach that would fall into that category would be stakeholder elicitation, where the uncertainty range is agreed upon by a number of stakeholders' beliefs about the future. As there is no means of determining how a generic group of stakeholders would have decided at a particular moment in the past, validation and generalization of these types of uncertainty estimates is virtually impossible.

Secondly, we considered but excluded very arbitrary estimates. This could for example be the heuristic of choosing the 10^{th} and 90^{th} percentile as a $\pm 20\%$ error for the forecast five years out. While to our anecdotal knowledge this approach is not uncommon, we chose to exclude it due to the entirely arbitrary nature and the vast number of heuristics that could be employed (e.g., why use 20% and not 15%).

5.4.1 NP$_1$: Non-parametric density forecasts by retrospective errors

This is a detailed description of the empirical density prediction method NP$_1$ as introduced in [51]. Methods NP$_1$, NP$_2$, and G$_1$ are based on the assumption that the past forecast errors are a good estimator for the future forecast errors. Under this assumption, the distribution of past errors provides a probabilistic estimate of a future actual value given a point forecast by the same forecaster [56]. For this EPI (NP$_1$), we use a non-parametric distribution of the errors.

To respect the fact that forecasting gets more and more difficult the further we look into the future, we group the forecast errors by their horizon. For constructing the EPI, we assume that a future forecast error is sampled from the same distribution as past errors. In particular, it is the distribution of all forecast errors with a particular horizon H that determines the uncertainty of the new forecast H years into the future. With the error distributions for a number of consecutive horizons, we obtain a measure for the uncertainty for a time frame $H = 1 \cdots H_{max}$ years into the future. Anchoring the error distribution with $\epsilon = 0$ on the most recent forecast, we obtain a density forecast.

When we create the density forecasts, we need to find the appropriate reconstruction of the predictive density over future real values. In the relative error metric, the statistics of the distribution such as quantiles are reconstructed as actual values y relative to the most recent forecast $\hat{y}$. This is $\epsilon_{rel} = \frac{\hat{y}}{y} - 1 \Leftrightarrow y = \frac{\hat{y}}{\epsilon_{rel}+1}$. When constructing the density forecast from log-errors, we need to use a different expression than if we work with ϵ_{rel}. For log-errors, the density forecast is constructed as

$$
\begin{aligned}
\epsilon_{log} &= \ln(\hat{y}) - \ln(y) \\
\Leftrightarrow \ \ln(y) &= \ln(\hat{y}) - \epsilon_{log} \\
y &= \exp[\ln(\hat{y}) - \epsilon_{log}] \\
&= \hat{y}e^{-\epsilon_{log}}.
\end{aligned}
\tag{5.1}
$$

5.4.2 NP$_2$: Transforming the errors for the median-centered EPI

For method NP$_2$, we center the distribution of errors such that the median of the distribution coincides with $\epsilon = 0$. This prevents the density prediction from creating a second point forecast when bias is present in historical forecasts, as it is the case with method NP$_1$. Here, the goal is to give the largest probability weight to the AEO reference case forecast.

The median-centering is done in percentage points of the errors. This procedure is not based on physical rationale, but it turns out to be a reasonable transformation for small median errors. The centered relative errors are transformed as $\epsilon'_{rel} = \epsilon_{rel} - m_{rel}$. We write ϵ_{ctr} as ϵ' for simplicity. The price forecasts are median-centered in log-errors. Some price quantities have large median errors. If they would be centered in a relative error metric, $\epsilon'_{rel} < -1$ could occur, which is not defined. The log-error metric prevents that situation from occurring.

Centering the error distribution in the log-error metric to ϵ'_{log} changes the relative error as follows below. We center here with the median of the log errors m_{log},

$$
\begin{aligned}
\epsilon'_{log} &= \epsilon_{log} - m_{log} \\
\ln(1 + \epsilon_{rel'}) &= \ln(1 + \epsilon_{rel}) - m_{log} \\
1 + \epsilon_{rel'} &= \exp(\ln(1 + \epsilon_{rel}) - m_{log}) \\
1 + \epsilon_{rel'} &= (1 + \epsilon_{rel})e^{-m_{log}} \\
\epsilon'_{rel} &= (1 + \epsilon_{rel})e^{-m_{log}} - 1.
\end{aligned}
\tag{5.2}
$$

Centering in log-errors retains a crucial property of relative errors, by ensuring that they are defined on the range $-1 < \epsilon_{rel}$. This can be seen by

$$\begin{aligned}\epsilon_{rel'} &= (1+\epsilon_{rel})e^{-m_{log}} - 1 \\ &> (1-1)e^{-m_{log}} - 1 \\ &= -1.\end{aligned} \tag{5.3}$$

How this change in centering changes the resulting width of the uncertainty interval for a range of errors $-1 < \epsilon_{rel} < 7$ is shown in Fig. 5.2. We see here as well that centering in the log error space prevents singularities, which can occur when transforming back to the forecast uncertainty when centering in the relative error space.

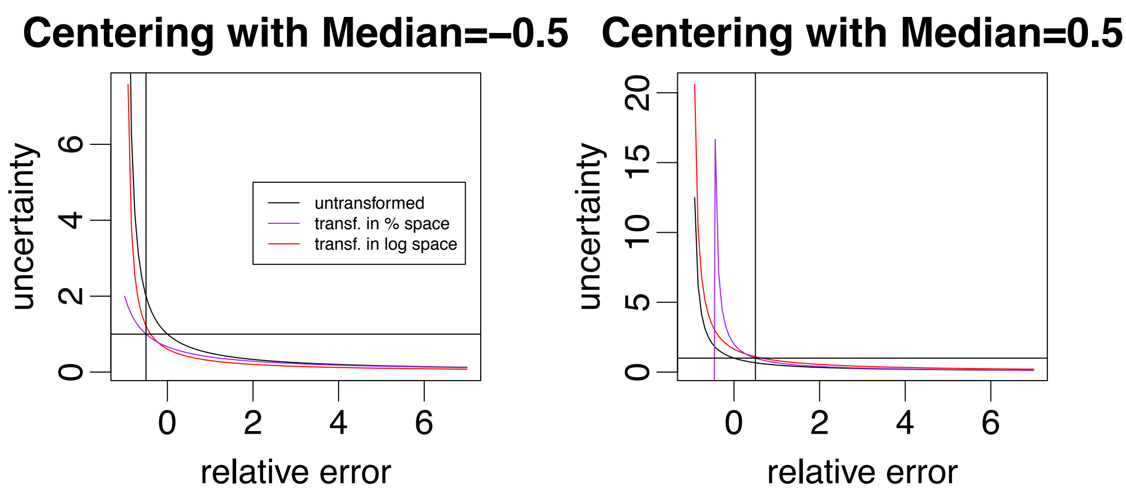


Figure 5.2: Comparison of centering in the two error metrics and the impact on calculating the final uncertainty. To the left with a large negative median error and to the right with a large positive median error. We see that the singularity, that occurs when centering in the relative error space, does not occur for centering in the log error space. Median errors are in units of relative and log error respectively.

5.4.3 G_1: List of standard deviations for all quantities

In Table 5.1, we give the standard deviations of the error samples, which are necessary to implement method G_1 for AEO 2016. We list the standard deviations of the relative errors (or log-errors for prices) for every horizon $H = 0$ to $H = 12$ computed with AEOs 1982-2016.

5.4.4 G_2: Finding the standard deviation of historical values

Method G_2 is a Gaussian uncertainty based on the deviations in the time series of historical values. We find the standard deviation by taking a pair of two historical observations, a horizon H apart, and calculate the relative change of the later value with respect to the earlier value. This is in analogy to the relative error. We find all possible pairs over the time series of historical values for a certain H. The standard deviation of this sample then is the standard deviation that is used to construct the density forecast. For price quantities, we determine the deviation as a log error and then find the standard deviation of those log errors. There is no value for $H = 0$, since $H = 0$ not a real forecast horizon. It corresponds to the error that occurs when in a new AEO the past data has been updated.

5.4.5 Alternative density forecasting methods

The most straightforward integrated approach to obtain a probabilistic forecast is to propagate the uncertainty of both initial conditions and model parameters, most commonly achieved using Monte Carlo simulation. Sensitivity to initial conditions, a feature of many nonlinear systems, is a particular challenge for example for numerical weather prediction. One solution is ensemble weather forecasting, whereby a separate scenario is simulated for each initial condition [44]. These simulation approaches do not consider model misspecification, where the model structure is erroneous, and results depend on the modeler's assumptions about the (future) distribution of input parameters. In the particular case of the AEO projections and the NEMS model, a report by the National Research Council (1992) [494] has recommended the use of multiple probabilistic techniques including Monte Carlo methods and closed-form statistical approaches. They emphasized the need of having reduced-form modules available for shorter run times. The implementation of those methods, however, is considered difficult and might not be feasible. The EIA more recently published a working paper about the use of dynamic stochastic general equilibrium (DSGE) models [495], where the author writes "DSGE models do explicitly incorporate uncertainty and are predominantly forward looking. These models use rational expectations, which imply that consumers are correct on average in forming their expectations about the future values of variables. DSGE models cannot be made very large due to the incorporation of uncertainty, and this limits their usefulness in detailed policy analysis. Their primary uses to date have been in the research work at universities and central banks. Some recent progress has been made in using DSGE models to forecast different macroeconomic variables, but this is an emerging research area."

Other probabilistic forecasting methods are generally very different from the EIA's current forecasting approach, but could in principle give guidance to the AEO scenario selection. Modeling time series data as a stochastic process and methods related to vector autoregressive (VAR) models are common in finance and economics [40]. VAR models might be more suitable for short-term forecasts in the EIA context [495]. There are Bayesian methods that allow for probabilistic forecasting such as Bayesian vector autoregression [496] or Bayesian hierarchical models [42]. In general, many statistical and machine learning methods, such as neural networks [497, 442], can generate density forecasts [36]. When subjective prediction is assessed by expert elicitation, typically the entire predictive distribution is elicited [37]. In principal, an expert elicitation protocol could be modified to quantify the uncertainty around a given point forecast.

5.5 Sensitivity of the method ranking

5.5.1 Normalizing the CRPS

We normalize the average CRPS for every horizon by the average CRPS for every horizon of the scenario ensemble. This is preferable over normalizing every single observation first, since this would unnecessarily bias the result. To illustrate this, we consider a sample of two instances producing the scores for the alternative density forecast $CRPS_{Alt} = \{1, 2\}$ and the scenarios $CRPS_S = \{2, 1\}$. We would in this case like to have an average normalized score of $CRPS_{mean,norm} = 1$. By normalizing for every observation, we would obtain $CRPS_{mean,norm} = mean(\{\frac{1}{2}, 2\}) = 1.25$. However, if we normalize the means, we get $CRPS_{mean,norm} = mean(\{1, 2\})/mean(\{2, 1\}) = 1$.

5.5.2 Main ranking method

To find the best density prediction method for each quantity, we rank the average CRPS after normalizing it by the average CRPS_S of the scenario ensemble. We refer to the scenario methods with the subscript S. For every quantity we then average over a core range of horizons $H = 2$ to $H = 9$, and rank these aggregated scores. The method with the lowest average rank is considered the best density over the test range for a given quantity.

We chose to exclude $H = 0$ and $H = 1$ from the core range of horizons because for most forecast users only future values are relevant. The number of observations per horizon in the test range without AEO 2009 ranges from 11 ($H = 0...2$) over 5 ($H = 9$) to 2 ($H = 12$). We exclude the horizons with a sample size smaller than 5 from the core range, which then is $H = 2$ to $H = 9$.

Table 5.2 summarizes the ranking results for every quantity. It compares the best and second best method of the main ranking procedure, as well has the best method if we average over the larger range $H = 1$ to $H = 12$, employ an alternative ranking method detailed in the next section, or change the test range. We find that the respective best methods do not change much with this sensitivity analysis. Some quantities are however very sensitive to changes in the range of observations since for those quantities two or more methods have very similar scores. For example for natural gas prices and natural gas consumption, the best and second best methods switched after we added the 2015 observation with publication of the AEO 2016. The update of the 2014 observation in AEO 2016 did not have an effect. Those three quantities are an example where the difference is very small. We also see sensitivity for natural gas prices in constant dollars when we remove the first test AEO 2003. The table also lists how much the average normalized CRPS of the best method is lower, and therefore better, than the second method.

5.5.3 Alternative ranking method

To explore the sensitivity of our results for the best density prediction methods for each quantity, we introduce an alternative ranking method. We rank the average CRPS results for each forecasting horizon separately. For every quantity we then average the rank of a method over $H = 0$ to $H = 9$, which results in the final ranking score. The method with the lowest average rank is considered the best density forecasting method for a given quantity. This approach is agnostic about how much the CRPS is improved by a given method over the other. This is the reason why we decided not to use this ranking procedure as the default.

We find that that the method rankings do not change much with the choice of ranking method. The results are listed in Table 5.2. The alternative ranking method ranks the second best method differently to the main ranking method for only one quantity.

5.6 Improvement over scenarios

In Fig. 5.9 we show that we can find a density forecasting method that has a lower mean CRPS than the scenarios for all of the quantities. The only partial exception is petroleum consumption, where that is only true for lower horizons.

5.6.1 Hypothesis test with bootstrap

It is insufficient to know that the aggregated mean CRPS, which we used to rank the methods, is smaller than the aggregated mean CRPS for the ensemble scenarios. Even though a mean might indicate an improvement, the improvement might come for a small fraction of the analyzed observations.

We use a bootstrap method to test how robust, or significant, the indicated improvement is. For each horizon, every observation generates a single CRPS. We resample these scores from the CRPS sample, which depending on the horizon can contain up to 11 elements. We assume complete independence, which means that we do not resample by observation year or make other assumptions about correlation. Under the null hypothesis we assume that the scenarios are the better forecast, i.e. they have the lower aggregated mean CRPS. We test this for every one of the four empirical methods and for every quantity. We resample simultaneously the scores of both the empirical method and the scenario method, which belong to the same observation. We normalize the new mean CRPS by the new mean CRPS_S for every horizon. Averaged over the core horizon range, we obtain a new aggregated normalized mean CRPS. We repeat this a thousand times to find the number of cases where the empirical method could be qualified as worse than then ensemble scenarios, meaning the normalized CRPS is larger than 1. We want this proportion to be smaller than our confidence level of 0.05 to speak of a significant improvement of the empirical method over the ensemble scenarios for the test range.

For all of the quantities, the respective best method is always significantly better than S. Besides performing the hypothesis test for the best methods, we also compared each of the single methods to S. We found that most performed significantly better for all quantities with the exception of NP_2 for constant oil prices which was better with a 92% confidence, and NP_1, which only performed significantly better for six of the eighteen quantities.

We also compared the best methods with the SP_1 method (Gaussian based on scenarios), and found that we can be 95% confident for almost all of the quantities that we found a significantly better uncertainty estimation method for the test range. The only exception is petroleum consumption, where the best method is only better at 74% confidence.

5.6.2 Further analysis of the scenarios

To understand if the scenario range is too narrow, we measure the coverage probability of the range between the envelope scenarios. This corresponds to the percentage of observations that were lower than the highest and higher than the lowest scenario for test AEO 2003-2014, without AEO 2009 (Fig. 5.10). We find that the coverage varies for different quantities and for different horizons, but it is generally very low with an average of 13.7%. This average is for the core horizon range and all quantities. Typical prediction intervals are intended to cover for example one or two standard deviations of a Gaussian distribution, which correspond to about 68% and 95% respectively.

We note that EIA's AEO scenarios are not intended to have a certain coverage probability. They are sensitivity cases on certain input assumptions. Since only one or very few assumptions, such as the impact of a particular policy, are changed at a time, the side cases typically do not differ as much from the reference case as they would if several assumptions were changed simultaneously. If the scenario range would be used for communicating the uncertainty, several assumptions would need to be changed simultaneously and probabilities would need to be attributed. Nevertheless, the

EIA writes for example in its most recent AEO 2017 [57] "EIA addresses the uncertainty inherent in energy projections by developing side cases with different assumptions of macroeconomic growth, world oil prices, technological progress, and energy policies." In our analysis, we use the SP_1 method to account for a wider uncertainty based on the scenarios. The method uses the range to the widest envelope scenario (of both low and high) as one standard deviation to fit a Gaussian distribution with the reference case as the mean. In this case, the observation is expected to be within that range only 68% of the times, which is a lenient interpretation of the scenarios, particularly considering that the scenario range is often asymmetric.

5.7 Point forecast comparison

We compare the mean absolute percentage/log error (MAPE/MALE) of three alternative point forecasts with the AEO reference case, similar to the CRPS significance test. Point forecast comparison allows us to understand that even though in some cases it is better to correct the best estimate forecast with the bias of the EPI, in most cases the AEO and therefore a centered error distribution performs better over the test range AEO 2003-2014 without AEO 2009. We exclude AEO 2009 to make the results consistent with the density forecast. In Fig. 5.8, we show the results for all quantities. From the fact that the reference case seemed to be the better forecast than the median of NP_1 for all quantities but two, we could anticipate that NP_1 would not create a good empirical density forecasts. This was a reason to introduce the centering technique of method NP_2.

5.7.1 Point forecast results

The point forecast comparison is designed to compare the median of the error distribution (bias) to the AEO reference case. In addition, we compare the reference case to two benchmark forecasts. Persistence is the last observation, or here the $H = 0$ forecast. Over the test range it was better than the reference case for 10 of the 18 quantities. This surprisingly good result is probably particular to the recent historical evolution of many quantities. It remains to be analyzed over other test ranges. Another point forecasting method is an interpolation of a simple linear regression over a fixed window. The length of the window has been optimized for the test range, excluding AEO 2009. We tested a window of 5 to 10 years and found that a window of 7 years shows a better forecast for the largest number of quantities, which is 8. This is based on an optimization both on the data pre AEO 2016 and the data updated with AEO 2016. The optimal window range does not change if AEO 2009 is included, but the simple linear regression generally performs worse.

5.7.2 Significance of Point Forecast Performance

We use a similar hypothesis test with bootstrap for the point forecast performance, as described in the previous section for the density forecast performance. Instead of the normalized CRPS, here we normalize the MAPE/MALE as $MAPE_{norm} = \frac{MAPE_{method}}{MAPE_{AEO}}$. We test if this quantity is significantly below 1, which would mean that the alternative method performed better over the test range than the AEO reference case. As before, we resample the absolute percentage error or log error samples for every horizon, and then average to get a MAPE/MALE for every horizon. We then normalize this average and determine the mean over the core horizon range $H = 2$ to $H = 9$. If less than 5% of the values are > 1, we speak of a significant improvement of the method over the AEO reference case for that particular quantity.

5.8 Analysis omitted in the final paper

To evaluate the calibration of the density forecasts, we also produced probability integral transform (PIT) values [46]. The PIT is defined as the value of the predictive CDF that an observation would have. A fundamental property of this variable is that it has a standard uniform distribution, if the historical value is sampled from a distribution that is equal to the density forecast. To assess if the density forecast is well-calibrated over all forecasts and all horizons we can determine if the PIT are sampled from a standard uniform distribution and if they are independent and identically distributed (iid) [46]. We used the Kolmogorov-Smirnov test to compare the distribution of PIT with the standard uniform distribution, and assessed the autocorrelation of the PIT time series. While this procedure was a good visual tool to understand the calibration of the density prediction, it was however not an adequate option to compare different methods quantitatively. We therefore discarded this method in favor of the CRPS.

We have also tried uncertainty estimation methods that weigh the errors depending on their expected relevance for future errors, considering the non-stationary nature of the error time series. We considered a nearest neighbor weighting method and a method that identifies intervals between non-stationarities and assigns weights accordingly. Those methods, however, have only in some cases improved method NP_1 and did not perform as expected. We believe that this approach could be more promising for forecasting problems with more data.

Table 5.1: Standard deviations of the forecast errors from AEO 1982-2016. SD are given as ϵ_{rel} except for the price quantities, which are given as ϵ_{log}. These can be used to construct a Gaussian density with quantile y around a forecast $\hat{y}$, which is defined as $y = \frac{\hat{y}}{\epsilon_{rel}+1}$ or $y = \hat{y}e^{-\epsilon_{log}}$ for relative errors and log errors respectively. Values are subject to change as historical values are updated or additional AEOs are released.

Quantity	H=0	H=1	H=2	H=3	H=4	H=5	H=6	H=7	H=8	H=9	H=10	H=11	H=12
Oil Price	0.029	0.227	0.357	0.423	0.530	0.577	0.668	0.754	0.823	0.893	0.965	1.003	0.988
Oil Price (const.)	0.060	0.229	0.352	0.408	0.505	0.545	0.628	0.703	0.754	0.808	0.865	0.893	0.874
Petroleum Cons.	0.009	0.022	0.038	0.053	0.060	0.070	0.079	0.092	0.105	0.120	0.131	0.139	0.144
Oil Production	0.016	0.050	0.086	0.113	0.131	0.136	0.135	0.133	0.125	0.132	0.144	0.160	0.167
Natural Gas Price	0.051	0.219	0.353	0.428	0.534	0.608	0.673	0.717	0.762	0.761	0.794	0.804	0.770
Natural Gas Price (const.)	0.065	0.218	0.348	0.418	0.518	0.586	0.645	0.679	0.712	0.705	0.725	0.730	0.701
Natural Gas Cons.	0.021	0.042	0.062	0.075	0.083	0.096	0.107	0.114	0.116	0.124	0.129	0.127	0.129
Natural Gas Prod.	0.019	0.040	0.061	0.075	0.090	0.104	0.116	0.124	0.127	0.132	0.129	0.126	0.116
Coal Price	0.060	0.076	0.133	0.187	0.246	0.303	0.362	0.421	0.481	0.535	0.585	0.624	0.641
Coal Price (const.)	0.037	0.076	0.125	0.169	0.220	0.268	0.317	0.365	0.410	0.451	0.486	0.514	0.525
Coal Consumption	0.020	0.045	0.062	0.078	0.097	0.123	0.146	0.162	0.174	0.188	0.190	0.197	0.207
Coal Production	0.019	0.039	0.054	0.059	0.068	0.081	0.092	0.102	0.107	0.117	0.116	0.121	0.130
Electricity Price	0.026	0.049	0.085	0.112	0.142	0.167	0.190	0.214	0.240	0.262	0.285	0.304	0.315
Electricity Sales	0.008	0.015	0.023	0.031	0.037	0.044	0.051	0.059	0.068	0.076	0.080	0.086	0.090
Total Energy Cons.	0.008	0.019	0.028	0.034	0.041	0.051	0.060	0.069	0.080	0.091	0.098	0.103	0.108
Residential Energy Cons.	0.025	0.042	0.039	0.038	0.040	0.048	0.056	0.057	0.064	0.073	0.074	0.076	0.078
Commercial Energy Cons.	0.021	0.033	0.042	0.052	0.056	0.059	0.069	0.078	0.087	0.100	0.103	0.109	0.108
Transportation	0.017	0.026	0.038	0.050	0.065	0.080	0.095	0.111	0.127	0.134	0.150	0.162	0.169

Table 5.2: Ranking results and sensitivity analysis for every quantity. The improvement of the best forecasting method with respect to the second best is measured in percentage difference of the normalized average CRPS. The best methods from various sensitivity analyses are listed to the right. We vary one assumption at a time. Deviations from the default ranking results are indicated in blue. The default ranking is performed on AEOs 2003-2014 without AEO 2009, observations 2002-2015, and over horizons $H = 2$ to $H = 9$.

Quantity	best	second best	$\frac{2^{nd}b.-best}{best}$	with AEO 2009	$H = 1$ to 12	test AEO 2004-2014	no obs. 2015	alt. ranking
Oil Price (nominal $)	G2	G1	0.8%	G2	G2	G2	G2	G2
Oil Price (constant $)	G2	G1	2.3 %	G2	G2	G2	G2	G2
Petroleum Cons.	G2	G1	1.8 %	G2	G2	G2	G2	G2
Oil Production	G1	NP2	4.1 %	G1	G1	G1	G1	G1
Natural Gas Price (nom. $)	G1	G2	0.8 %	G1	G1	G1	G2	G1
Natural Gas Price (const. $)	G1	G2	1.0%	G1	G1	NP1	G2	G1
Natural Gas Consumption	G1	G2	0.2 %	G1	G1	G1	G2	G1
Natural Gas Production	G1	NP1	2.9 %	G1	G1	G1	G1	G1
Coal Price (nom. $)	NP2	G1	6.5%	NP2	NP2	NP2	NP2	NP2
Coal Price (const. $)	NP2	G2	9.0%	NP2	NP2	NP2	NP2	NP2
Coal Consumption	G1	NP2	0.9%	G1	G1	G1	G1	G1
Coal Production	NP1	G2	12.7%	NP1	NP1	NP1	NP1	NP1
Electricity Price	NP2	G1	6.5%	NP2	NP2	NP2	NP2	NP2
Electricity Sales	G1	G2	2.1%	G1	G1	G1	G1	G1
Total Energy Cons.	NP2	G2	1.9%	NP2	NP2	NP2	NP2	NP2
Residential Energy Cons.	NP1	G2	7.2 %	NP1	NP1	NP1	NP1	NP1
Commercial Energy Cons.	G1	NP2	1.3%	G1	G1	G1	G1	G1
Transportation	G1	NP2	4.7 %	G1	G1	G1	G1	G1

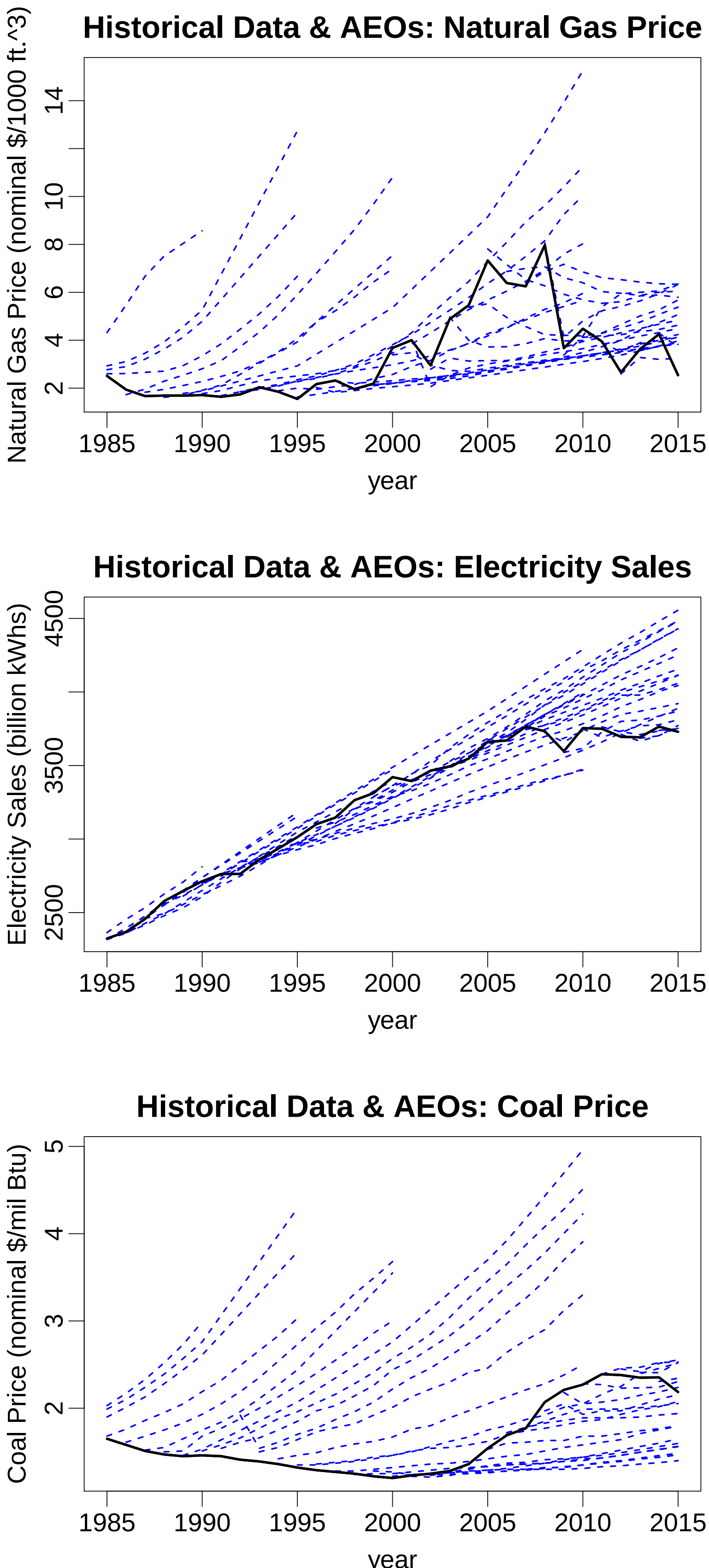


Figure 5.3: The historical values and AEO projections for the example quantities natural gas wellhead prices and total electricity sales, and the outlier case coal prices to electric generating plants. The black solid line indicates the historical yearly averages as listed in the EIA Retrospective Reviews. The annual projections from the AEOs 1982-2016 are shown in blue dashed lines. The unusual coal price projection for 1992 in AEO 1993 is not an error in the data.

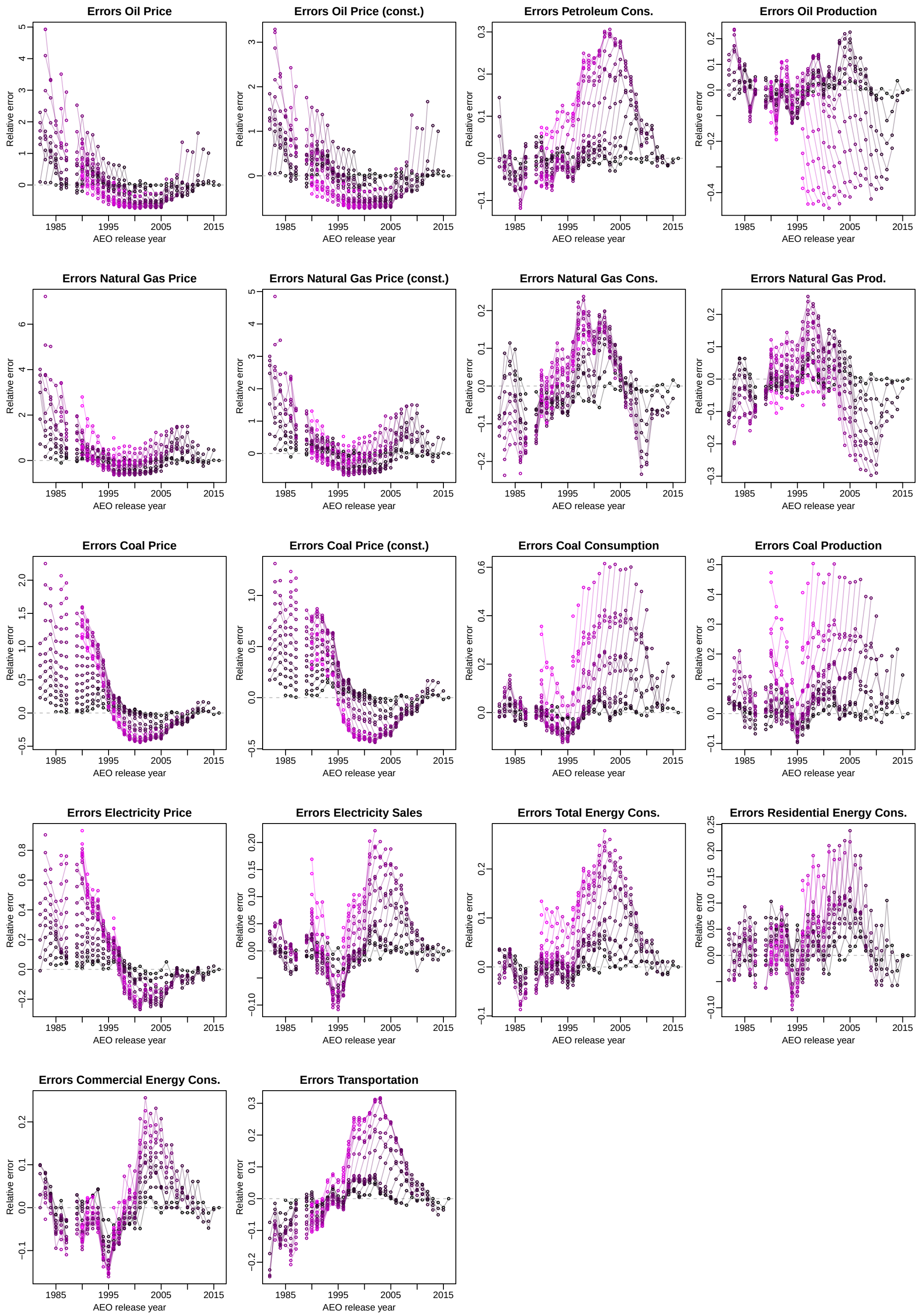


Figure 5.4: The relative errors in this data set for all quantities. Each color connected with a line corresponds to a horizon, ranging from $H = 0$ in black to $H = 21$ in purple. The price forecast errors are untransformed.

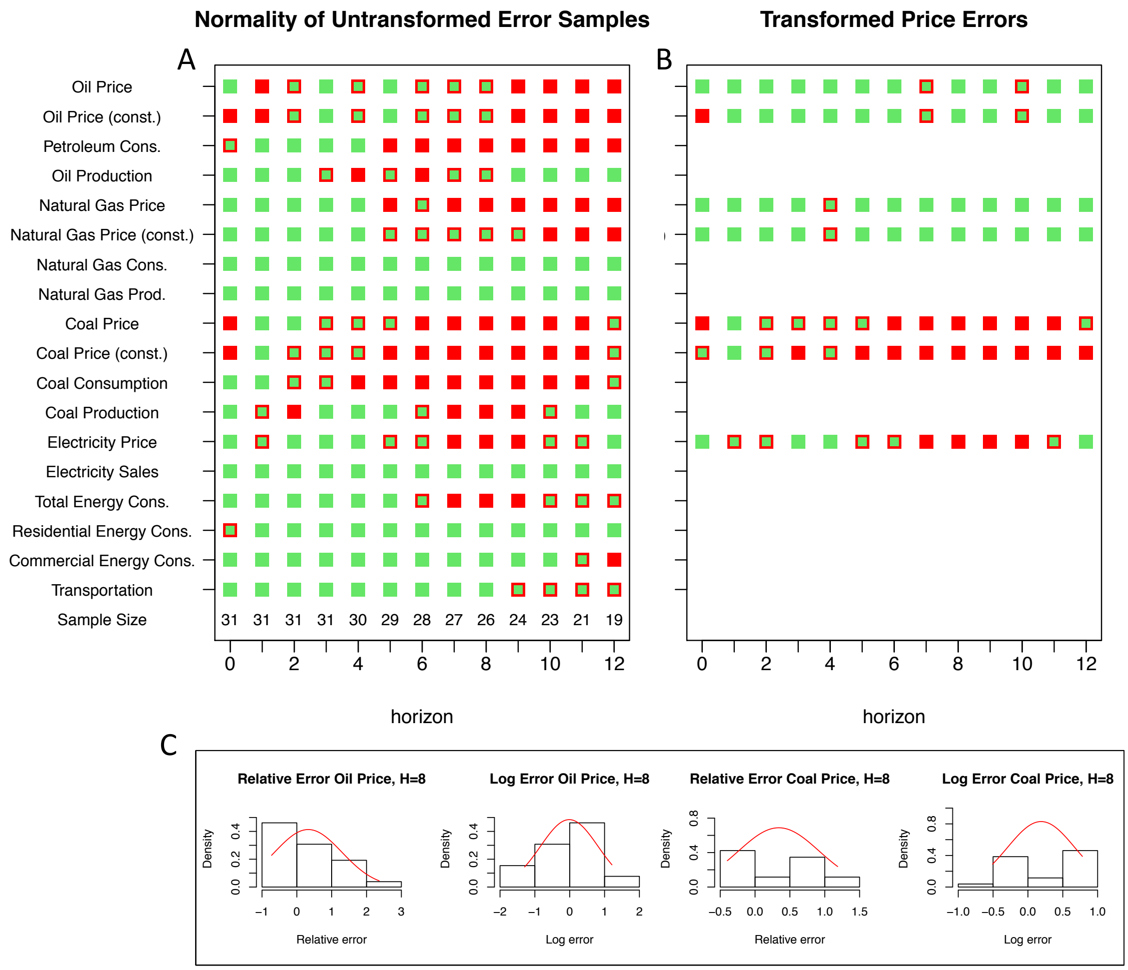


Figure 5.5: The results of the Shapiro-Wilk Normality Test with the original relative errors (A) and the transformed errors for the price quantities (B). Red indicates that the sample is not normally distributed with a certain confidence, while green corresponds to those samples where the null hypothesis of a normal distribution cannot be rejected. The underlying larger square corresponds to rejection with confidence $\alpha = 0.05$, and the smaller to $\alpha = 0.01$. (C) Two example histograms of untransformed and transformed errors with Gaussian fit, illustrating how the log error is much more normally distributed than the relative error for oil prices. The transformation has instead little effect on the bimodal coal prices.

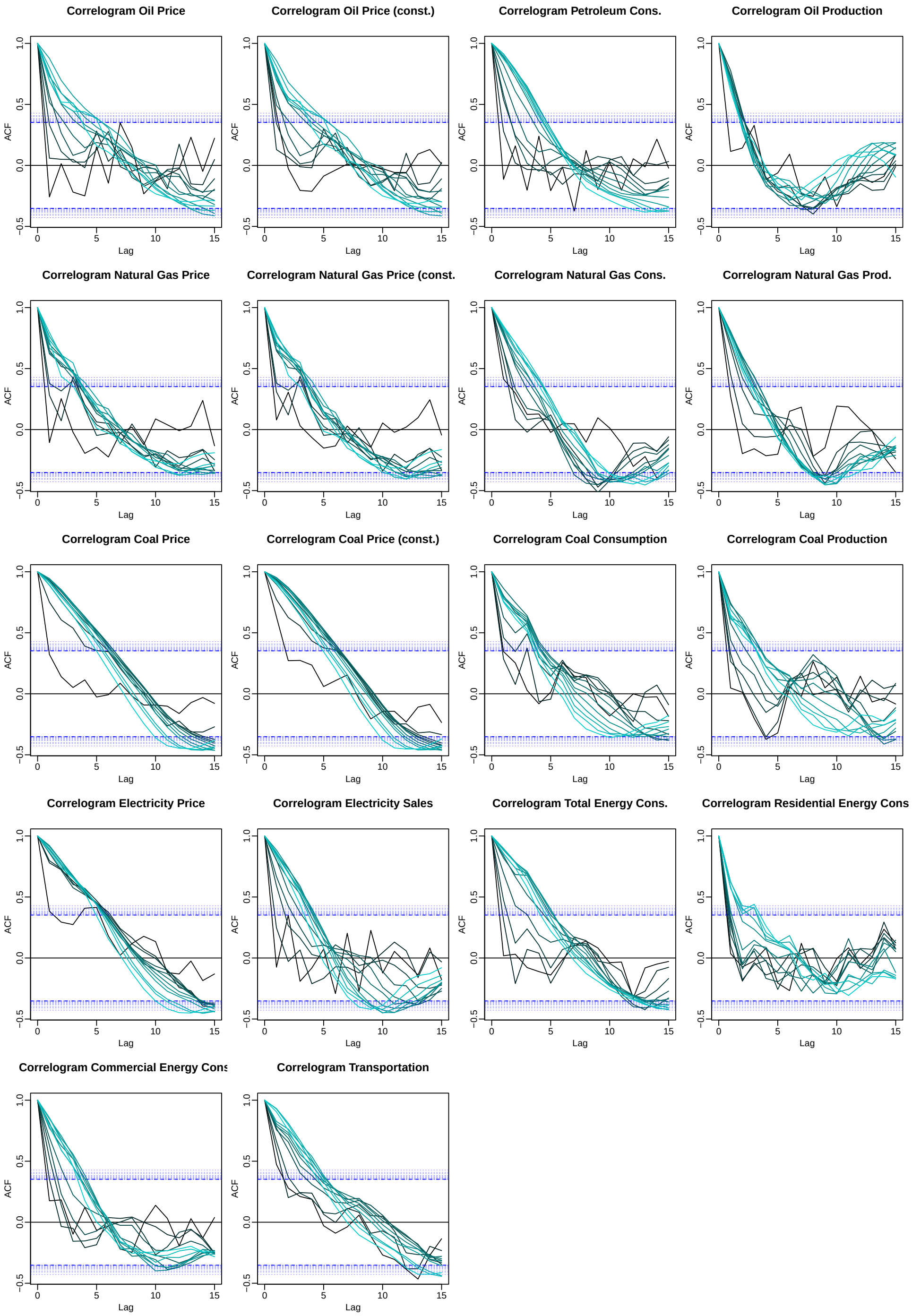


Figure 5.6: The correlograms indicating the autocorrelation in the time series of error samples. Every line shows how the error for a given horizon H is correlated to the error for the same H from a previous AEO. Results for different horizons are summarized in the same plot for every quantity. The colors range from $H = 0$ in black to $H = 12$ in light turquoise. The $\alpha = 0.05$ confidence bands for autocorrelation are indicated in dashed blue lines, they vary for different samples sizes. The confidence region is larger for larger H.

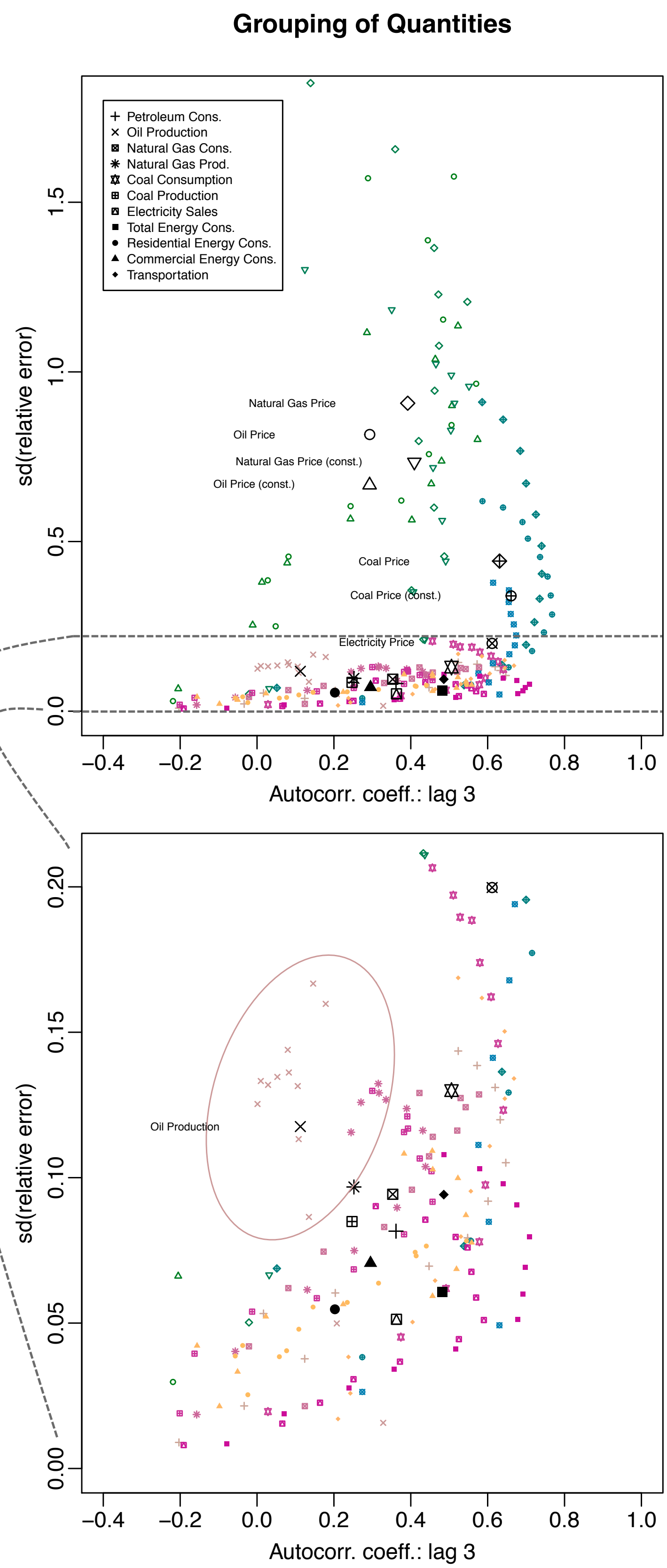


Figure 5.7: The standard deviation of relative error samples for distinct quantities and horizons $H = 0 \ldots 12$ plotted against the autocorrelation coefficient at lag 3 yrs. The top image is at full scale while the bottom image is cropped. The colors correspond to the three variable classes of prices (blue/green), production and consumption (magenta), and energy consumption by sector and total (orange). Every color and shape is assigned to one quantity. The black points indicate a mean forecasting error over all horizons for each quantity. Prices form a distinct group in this graph, with a much larger standard deviation than the other quantities. Coal and electricity price errors have a higher autocorrelation and a lower standard deviation than the other price quantities. The ellipsoid in the lower image highlights that oil production distinct from the other quantities, which is due to its low autocorrelation at lag 3 yrs.

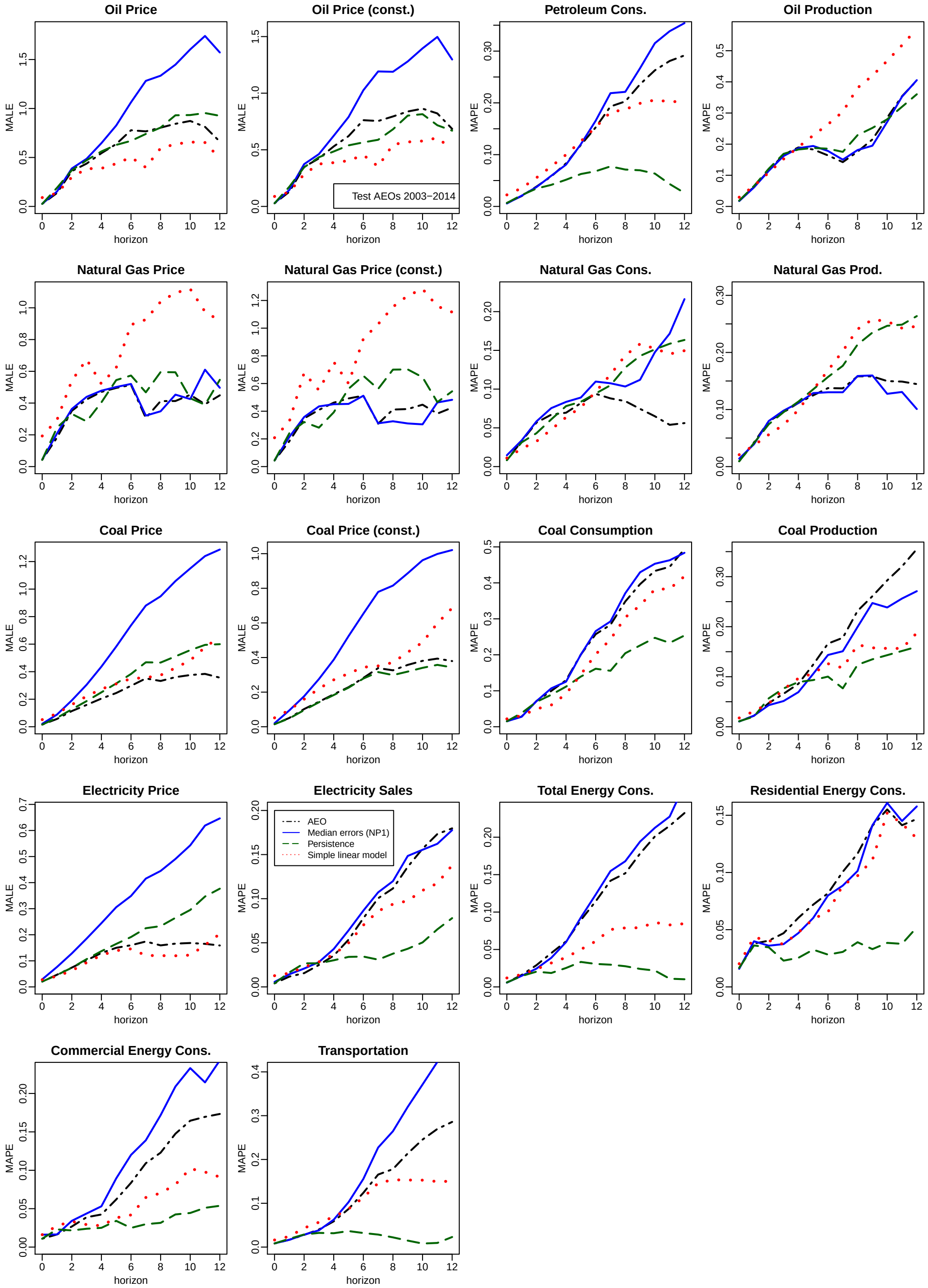


Figure 5.8: The results for the MAPE and MALE for all quantities. This is with the test range AEO 2003-2014, and excluding AEO 2009.

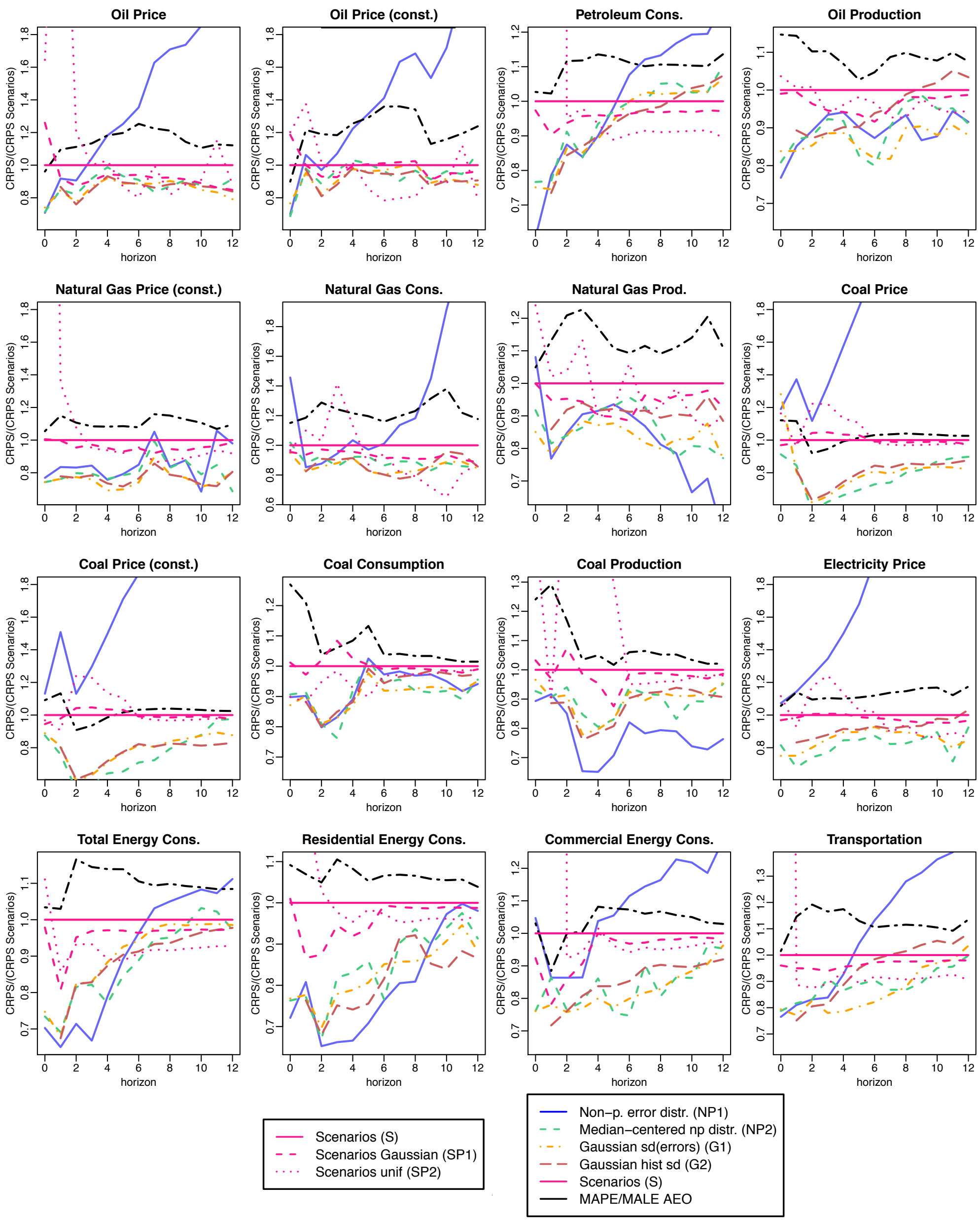


Figure 5.9: Relative improvement of the methods with respect to the highest and lowest scenarios for the test range AEO 2003-2014. Values are plotted as fraction of the CPRS of the scenario ensemble (S). A value lower than 1.0 corresponds to a better density forecast. SP_1 corresponds to a normal distribution with the scenario range as 1 SD, and SP_2 is a normalized CRPS of a uniform PDF between the envelope scenarios.

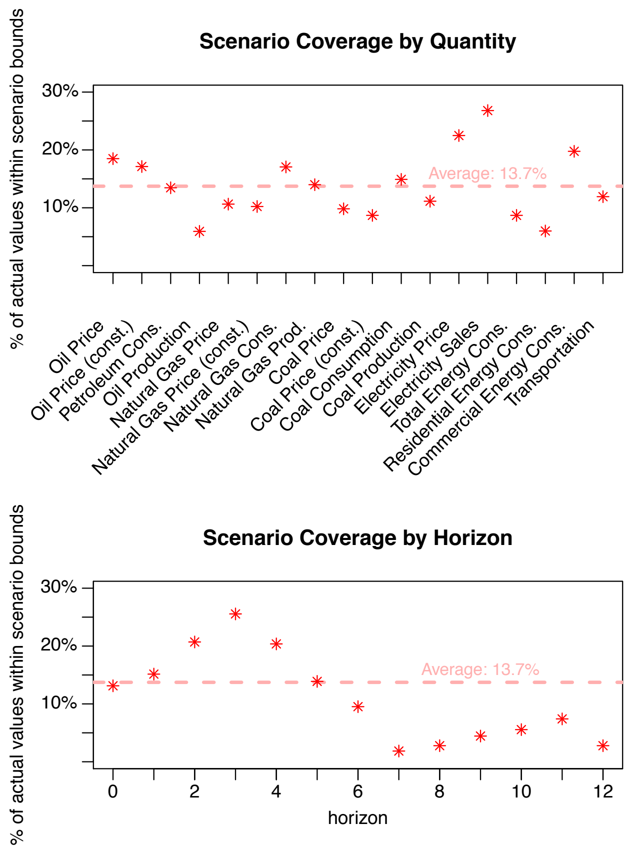


Figure 5.10: The coverage probability of the scenario range over the test range AEO 2003-2014 without AEO 2009. The coverage probability refers to the percentage of observed values within the range between the envelope scenarios. The average is computed as the average over $H = 2$ to $H = 9$ for every quantity (shown in A) and then averaged over the 18 quantities. The coverage for every horizon averaged over all 18 quantities is shown in (B).

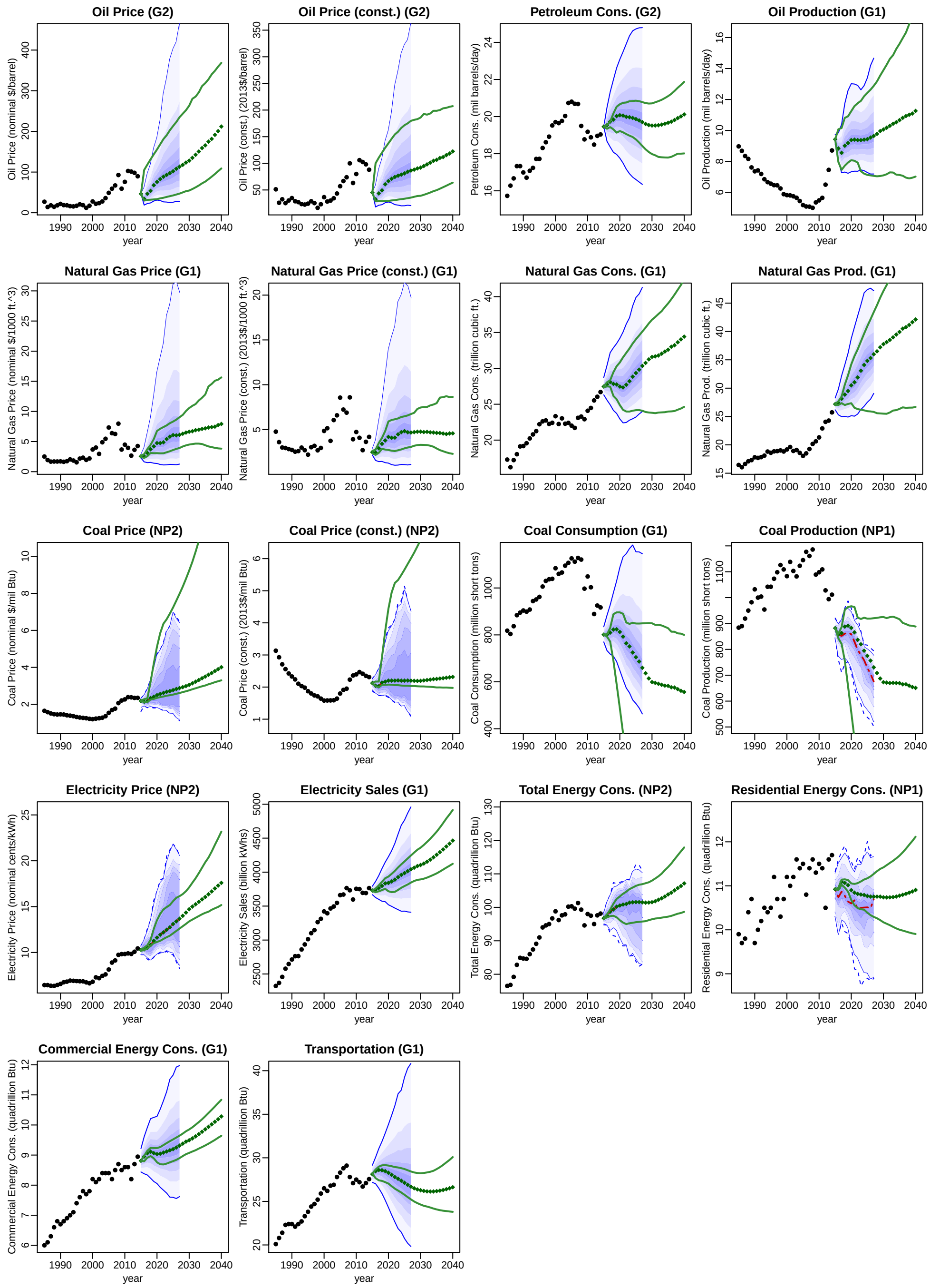


Figure 5.11: Density forecasts with the best method for every quantity based on AEO 2016. The different shades correspond to the percentiles 2, 10, 20, 30, ..., 80, 90, 98. The prediction interval can be very large, since it estimates that only 4% for a future value will fall outside of this interval. The red dashed line indicates the median if different from the reference case. The scenario range (in green) changes greatly from one AEO to another and is somewhat correlated to the number of scenarios published, which is why some AEO scenario ranges might be as wide as the empirical uncertainties. AEO 2016 has a large number of scenarios compared to other AEOs.

6

Appendix: Decarbonizing intraregional freight systems with a focus on modal shift

6.1 Data collection

6.1.1 Freight activity data

The database of freight activities, which is made available with this article in the supplementary materials, contains the country name, ISO3166-1-Alpha-3 code, freight mode, year, the source code, and the respective freight activity (in thousand tonne-km). The file contains a separate sheet with references for the 35 source codes. In this file, we reproduced official national statistics supplied through governments and also included results from peer-reviewed research studies or those that were funded by governments or international institutions. We have not reproduced the data from the OECD, Eurostat and the World Bank, but we have included those in the analysis presented in this Topical Review. We have added references to these publicly accessible datasets in the file (as source codes "1", "2", "8" and "10"). The full dataset covers values for 157 countries (75 for road, 152 for rail, and 51 for water). We also included a note for countries that do not have a railway system.

For our provided dataset and analysis, we have made the following adjustments and assumptions:

- Some data is presented for fiscal years instead of calendar years and given as a range in the in the source document (e.g. 2005/2006). This is for example the case for Australia and Myanmar. In these instances, we assigned the latter of the two years (e.g. 2006).

- American ton-miles, where "ton" refers to short ton, are converted as $1\textit{ton-mile} = 1.460tkm$. Myanmar uses the imperial system, where $1mi = 1.609km$ and $1ton = 1016.047$, resulting in the conversion factor $1\textit{ton-mi} = 1.635tkm$.

- The U.S. Census Bureau and the Bureau of Transportation Statistics conduct the Commodity Flow Survey (CFS) every five years [498]. The next update will feature 2017 data but, because it is yet to be released, we report results from the 2012 survey. The Bureau of Transportation Statistics also reports data for 2012-2014, which are estimates from the Freight Analysis Framework by the U.S. Department of Transportation, Federal Highway Administration.

- The definitions of whether a shipment is domestic or international differ between the US and Canada. In the US, the domestic leg of an international shipment is included in the total domestic freight activity[1]. Canada reports road freight activity for domestic and international shipments separately [242]. "International" freight activity includes everything with an international origin or destination, which is transported on Canadian grounds or by Canadian road carriers. The OECD only includes domestic freight activity of Canada in their database, which is why we adopt a similar definition and conduct all of our analysis with the domestic numbers. However, we show cross-border freight activity from Canadian carriers separately in Fig. 6 in the main text and in the section on North America in Fig. 9 in the main text, as this cross-border activity is quite large. Similarly for rail, Canada also reports international freight activity within Canadian borders with an international origin or destination separately as international[2]. We have also included rail activities in those figures. Cross-border shipments from North American carriers therefore remain not fully accounted for. For the water modes, we have only considered domestic freight activity.

- We include all water freight activity data that are indicated as domestic waterborne trade or coastal shipping. We do not include statistics on maritime shipping, unless they are labeled as domestic shipping. For countries that report both freight activity on inland waterways and coastal shipping, we have added those together and denoted it as domestic waterborne freight activity.

- Some data for Latin America are taken from the Freight Logistics Statistics Yearbook by the Inter-American Development Bank [499] for those countries, where we could not find national statistics. This dataset lists freight activity in tonne-km where available but only assigns a year to some of the data points. Where no information about the year is given, we have assigned it the year of publication, which is 2014.

6.1.2 Carbon intensity data

For Fig. 2.1 in the main text we collected average carbon intensity values from various sources that are listed in the provided data table (supplementary materials). By carbon intensity we refer to

[1]Personal communication with the U.S. Bureau of Transportation Statistics.
[2]Personal communication with Transport Canada.

GHG emissions over freight activity measured in CO_2/tkm. Most sources provide values in this unit, but some needed to be inferred from fuel consumption or energy consumption values. Some sources report average regional values, while others report average values for a vehicle class. We include also reports for single countries (e.g. Germany) in the regional values (e.g. Europe) in Fig. 2.1 in the main text. For sources that report values for more than one year, we use the most recent value. Below, we provide details on data collection and conversion.

Road carbon intensity

- We use the following conversion factors: Emission intensity for diesel is $10180 gCO_2/gal$ and gasoline is $8887 gCO_2/gal$ [500].
- Several sources [10, 92, 238] give regional values for average payload and fuel consumption, which we converted into carbon intensity by above conversion factors.
- One source [68] provides carbon emissions in gC/gal, which we converted to gCO_2/gal.
- Where no information on the road vehicle class was given, for example because the value is an aggregate over the whole road freight sector, we assign it to the heavy road freight vehicle class.

Rail carbon intensity

- We converted diesel rail energy consumption in btu/tkm, which was reported by one source [247], into carbon intensity. For this we used the conversion factor $138700 btu/gal$ for diesel given in the same reference.

6.1.3 GDP data

The GDP data used for this review is from the International Monetary Fund (IMF) given as "Gross domestic product based on purchasing-power-parity (PPP) valuation of country GDP" in current international dollars [501]. The use of these PPP-adjusted GDP values was for example recommended in a World Bank document by Bennathan et al. [75].

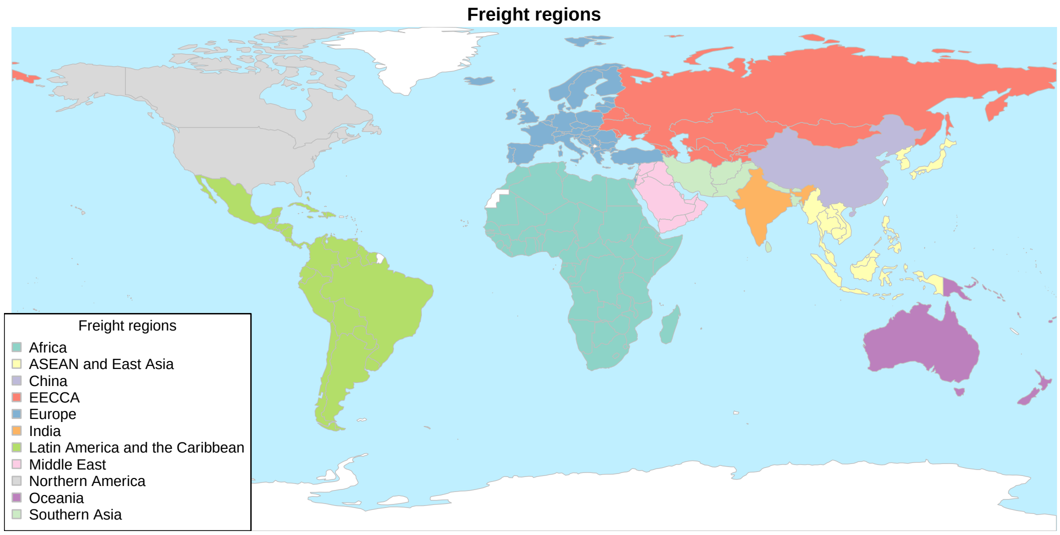


Figure 6.1: The freight regions used for our analysis. We based those region definitions on a combination of regional trade agreements, political and geographic barriers, and trade patterns.

6.2 Defining surface transport regions

To illustrate freight trends in this review, we grouped the countries into eleven freight regions, which are shown in Fig. 6.1 in this appendix. The results in this paper do not depend on this particular categorization. The groups are based on a cluster analysis of intra-continental trade data and a qualitative consideration of political and geographic barriers as well as trade agreements listed by the World Trade Organization.[3] While intra-continental trade in Africa is negligible, we still consider Africa as one region for geographical reasons and for its potential to increase intraregional freight.

6.3 Additional plots

Fig. 6.2 shows the national freight activity divided by the GDP by country.

[3] Available at `https://www.wto.org/english/tratop_e/region_e/rta_participation_map_e.htm`.

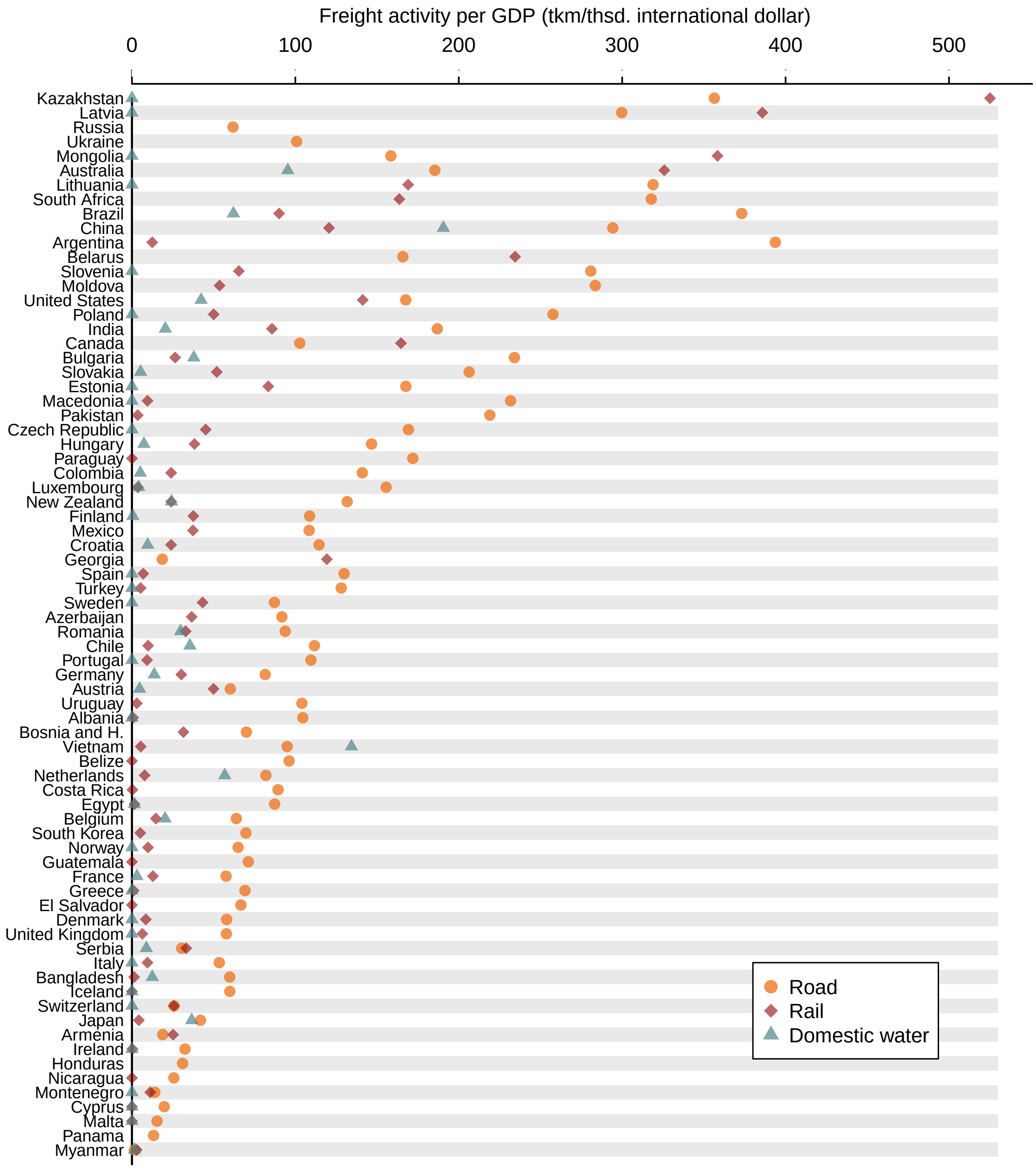


Figure 6.2: Road, rail and domestic water freight activity per GPD for all countries that provide information on road freight activity. GDP values are power purchase parity adjusted.

7

Appendix: Truck traffic monitoring with satellite images

7.1 Data preparation

7.1.1 Satellite images

We use satellite images from DigitalGlobe, Inc., which are taken frequently for many locations and by a number of different satellites. We only work with 3-channel RGB images of the satellite "World View 3" (VW03), as it has the highest resolution of 31 cm. We found that there are difficulties identifying for example black cars in images of the other satellites. The images are not cloud-free but we attempted to select images with nearly no cloud-cover in relevant areas.

For training, we used images of several regions in the Northeastern United States. These include images of the Thruway but also other highways. We also annotated the regions around the highways such as parking lots and logistics centers.

7.1.2 Annotations

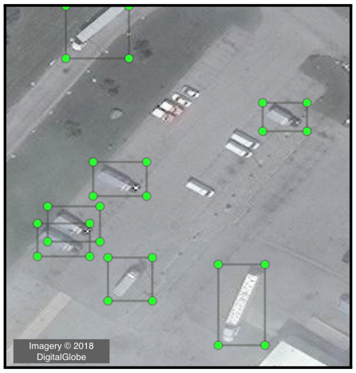
We used the Python-based annotation software "LabelImg" [502] to label the more than 2000 truck examples. We marked each truck with a bounding box and a class label "Truck." Below we describe in detail, which types of vehicles we included as trucks. For cloudy images, we also tried to label those trucks that are hardly visible through the cloud or in the shade of the cloud.

For the training data, we labeled large 3000×3000 pixel images, from which we created 300×300 pixel chips, where we only retained chips with truck examples. Note that this procedure reduces the number of truck examples somewhat with respect to the large images, as bounding boxes are cut and those examples are lost. We chose truck examples conservatively and prioritized accuracy of labeled examples over labeling as many as possible. This means, when in doubt, we chose not to label truck, unless there are very obvious or interesting truck examples in the immediate vicinity. We have sometimes omitted parking areas and junk yards, where vehicles are very close together such that there is no pavement visible. These examples might be less useful for learning on highways but might aid for situations with dense traffic.

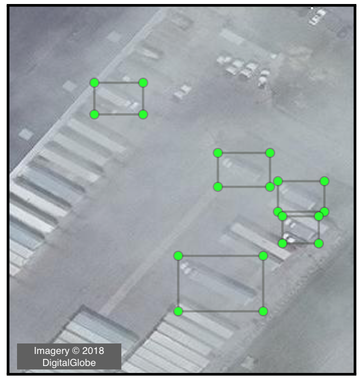
In contrast, for test images it was important to label all likely trucks because the whole image is evaluated with all examples it contains. This includes trucks that are partially obstructed through trees, bridges etc.

We only labeled semi trucks with a trailer, and medium-duty trucks. This also includes car carrier trailers, flatbed trucks or oversized transports such as windturbine blades. We did not annotate pickup trucks, even if they pulled a trailer, and omitted vans, buses, caravans, and RVs. We also did not label tractors or trailers separately, only the combination. Fig. 7.1a shows an example of easily identifiable trucks of different sizes. The examples in Fig. 7.1b and 7.1c are partially obstructed by clouds or trees but they can still be identified as trucks. Buses and RVs can easily be confused with a truck. The example in Fig. 7.1d shows a bus (or a long RV) and something that is likely an RV that is pulling a car, both of which could be taken for a truck. Fig. 7.1e and 7.1f have more of those examples, including a number of yellow school buses. We also include trucks with special trailers such as flatbeds (Fig. 7.1g), or oversized load (Fig. 7.1h). We also labeled trucks with empty trailers, as sometimes load cannot be distinguished from the empty trailer. Smaller trucks and their similarity with vans are particularly difficult (Fig. 7.1i). We considered everything that has a box that is elevated from the driver's cabin as a truck, but errors cannot be excluded. Also, there were many examples of small parked trailers that have a white attachment, which could also be a small drivers cabin (Fig. 7.1j). For trucks that are docked to a building, we excluded those where only the trailer is visible but included those, where the tractor is still attached (Fig. 7.1b).

In the images from Brazil, we found what seemed to be yellow and white busses, which appear in multiple locations Fig. 7.1k). After we have confirmed with Google Street View that such busses frequently travel the highway, we have not labeled these as trucks.

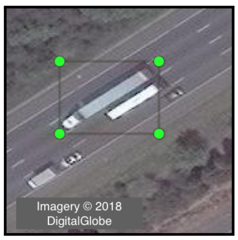


(a) Various trucks



(b) Cloud cover reducing visibility

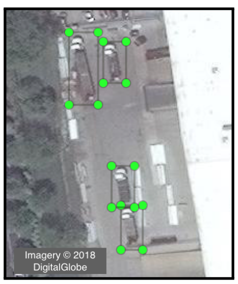
(c) Truck & tree



(d) RV and bus

(e) School busses

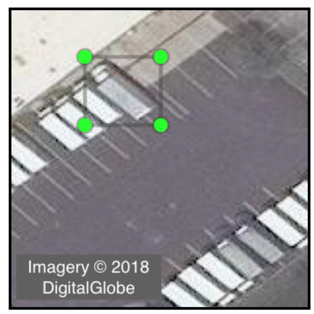
(f) RVs



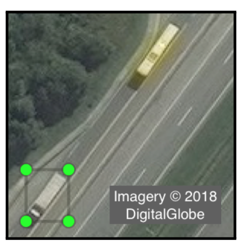
(g) Flatbeds

(h) Special load

(i) Vans



(j) Smaller trailers



(k) Yellow bus in Brazil

Figure 7.1: Image chips that illustrate what is labeled as a "Truck" with a bounding box. Imagery © 2018 DigitalGlobe, Inc.

7.1.3 Vehicle counts

We use count data for four different regions that comprise all counting stations with vehicle class distinction in California, Brazil, and Germany, and toll data from the NY Thruway. We are only interested in highways (or freeways), not smaller roads. Here, we describe each dataset and the respective data preparation, and conclude with a summary. To balance the training data between the regions, we used a sampling method that is explained in Section 7.1.3.

NY Thruway

The toll data for the New York Thruway [366, 365] contain the entrance and exit location for every vehicle and the time it has entered the Thruway as recorded in the toll collect system. We regarded all high vehicles with 3 or more axles as trucks. To determine when a vehicle has passed a location between entrance and exit, we needed to make assumptions about the speed it has traveled and the distance between highway exits. We assumed that every vehicle travels 65 mi/hr. The information about Thruway mileposts are available through the State of New York Thruway Authority [503]. We determined the counts by summing up all the vehicles that have entered a section between two highway exits. For example, if we were interested in analyzing the stretch of road between Exit 30 and 31, we determined the hourly counts by summing up the number of vehicles that pass Exit 30 in one direction and pass Exit 31 in the other direction within that hour. We did not only count the vehicles that pass the particular exit but also the vehicles that enter the Thruway at that exit. The ones that leave at the exit before they enter the section were not counted. See Fig. 7.2 for an illustration.

California

We obtained hourly count data through the California Department of Transportation (Caltrans) Performance Measurement System (PeMS), where we used hourly truck counts from the Caltrans Traffic Census Program [?]. We furthermore used location information of the census stations (weight in motion stations) from PeMS and Caltrans. We also obtained AADTT information from Caltrans for a number of road segments in California [?].

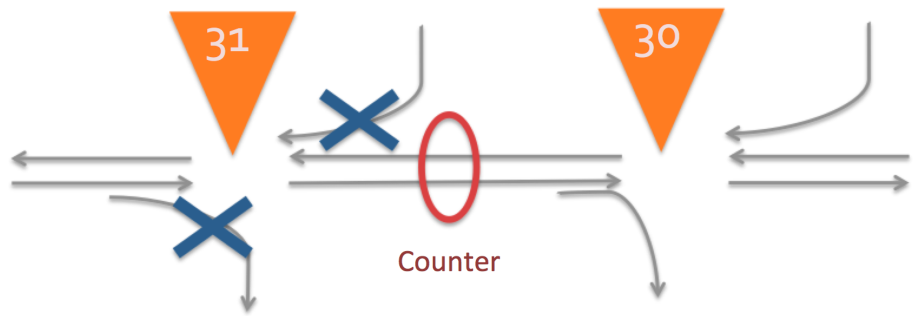


Figure 7.2: Schematic illustration of how we computed the traffic flow for the monitoring model from toll data. The orange cones indicate highway exits, and the grey lines are road sections. We did not consider those that are crossed out.

Hourly data preparation The datasets contain many short-term and some nearly continuous counters. We only used those counters that provide information on both directions, and add those directions up as well as the counts for all lanes. We dropped rows that have vehicle classes $(0, 2, 3, 4, 15)$, as those include vehicles that are too small, or indicate malfunctions of the system. We used the data for the year 2016.

AADTT check As a check, we compared the AADTT computed from the hourly data to those aggregate values provided through [?]. Comparing counting locations in California proved difficult, as precise geographical information is not given. For example, we compared the value for census station 62010, which is weight in motion (WIM) location 73 for Caltrans or *rte. 5, district 6, leg A, JCT.RTE.43* in the 2016 table in [?]. Here, we computed an AADTT of 8578 with the simple AADTT method, and Caltrans gives 8819. We ensured approximate compatibility also for other count stations.

Brazil

The data for Brazil were made available through the Departamento Nacional de Infraestrutura de Transportes (DNIT) [368]. The Brazilian dataset contains short-term as well as continuous counters on Brazilian national highways. The dataset contains counts for several years. We worked with counts for the year 2017 for training the monitoring model because this is the year that has most data available.

Germany

The German agency Bundesanstalt für Straßenwesen publishes hourly count data for count stations on highways (Autobahnen) [?]. The data consist only of time series from continuous counters. We removed those stations that are faulty, indicated by an AADTT of 0. For training the monitoring model, we used the most recent data from 2017.

Count data summary and sampling

To balance between the regions, and ensure sufficiently high AADTT in the training data, we developed our own sampling procedure. We sampled from the stations with the longest count series (most "complete" stations) first, and then sampled from a selection of those with the highest AADTT, to arrive at a dataset of 10 continuous counting stations. As some of the datasets contain short-term counters, we iteratively increased the number of stations sampled until we arrive at a dataset with a number of points equivalent to 10 continuous counting stations.

The two plots in Fig. 7.3 show how complete the count time series are by station and the size of the AADTT, respectively. We see that the number of count stations as well as the length of count series by station differ considerably between the four regions. The figures also show those stations that were randomly selected to be part of the training dataset for the monitoring model. Count stations for test images were excluded.

7.2 Experimental details and further analysis

7.2.1 Detection model

Fig. 7.4 shows the count errors for validation data based on the entire image. We see that the count errors are much larger than for predictions on the road only. This is due to the difficulty of the model to detect trucks that are parked next to each other on parking lots, and more unusual truck types and image environments outside the road.

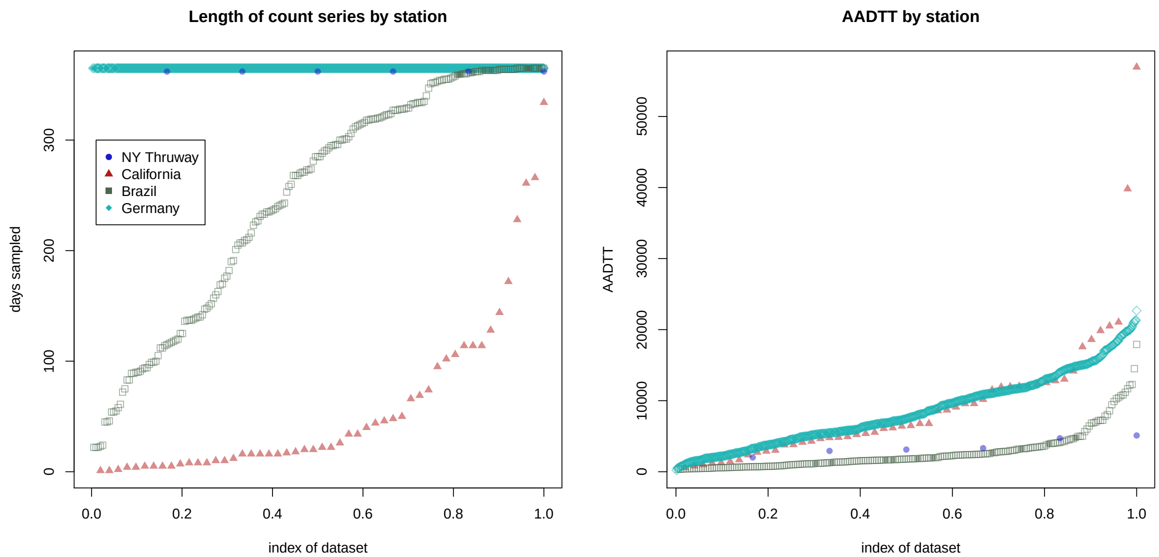


Figure 7.3: The count data vary between the regions. For example, German data contain many continuous counters, while for California mainly short-term counters are included. Filled points indicate those that where selected for the training dataset by the random procedure that prioritizes longer count series and then higher AADTT. We see that for the NY Thruway and California all count stations were used for training.

One issue of the detection model is that it underpredicts the number of trucks. We were interested in understanding if the count error is systematically biased. Fig. 7.5 shows the count error for each validation image with a negative value meaning underprediction. It also shows the mean over all images. As expected, the mean count error is around zero for optimal values. The model that is most balanced is Faster R-CNN with ResNet 101. The SSD, which we used in our analysis, tends to underpredict for the widest range of probability thresholds. This indicates, that it might be useful to further explore if lower validation errors could be achieved by training Faster R-CNN longer.

7.2.2 Monitoring model

Computing the AADTT We can use two methods to compute the AADTT [343]. A simple method computes the average hourly count over all hours in the dataset (ideally a year) and multiplies by 24 to obtain the average daily value. A second method, the AASHTO method, computes average values for every weekday in a month, and then averages over these daily values. This method

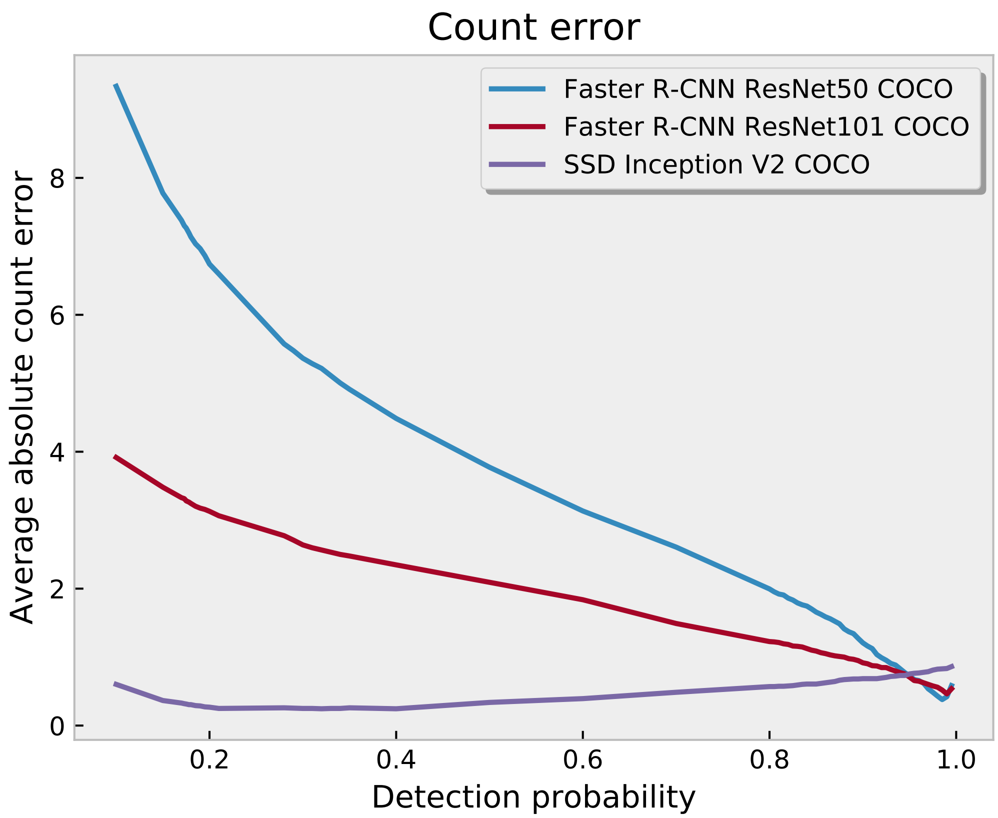


Figure 7.4: Count errors for all three detection models, based on all annotated trucks in the validation datasets, not only on the road.

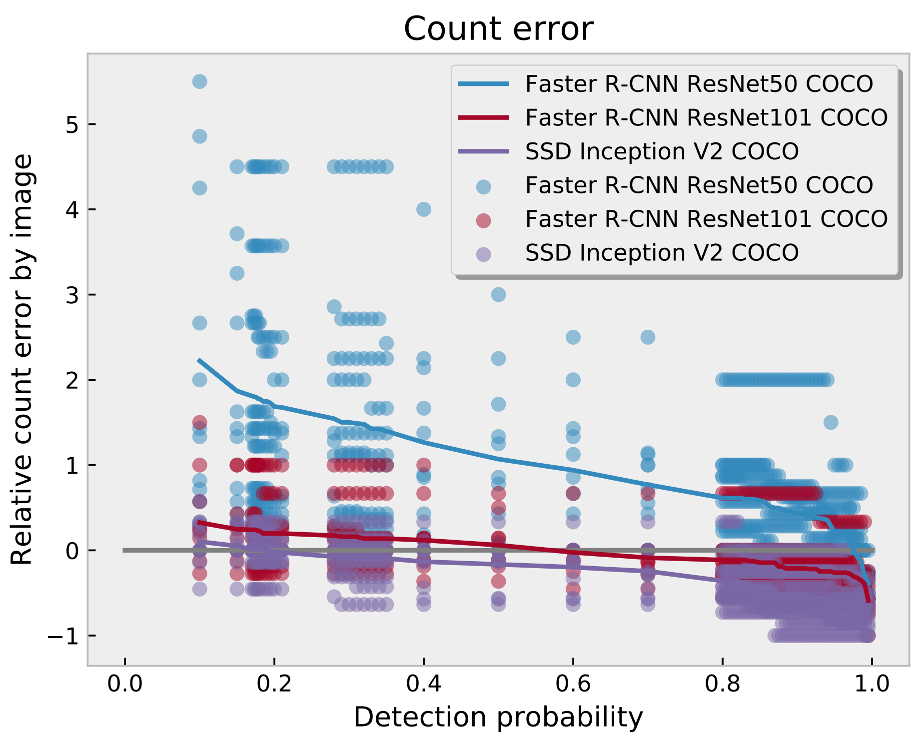


Figure 7.5: Count errors as a fraction of the true count of annotated trucks by image (points). Those counts are only for trucks on the road. Negative values indicate a lower predicted number of trucks than the number of annotated trucks. The lines indicate the mean over all images per model.

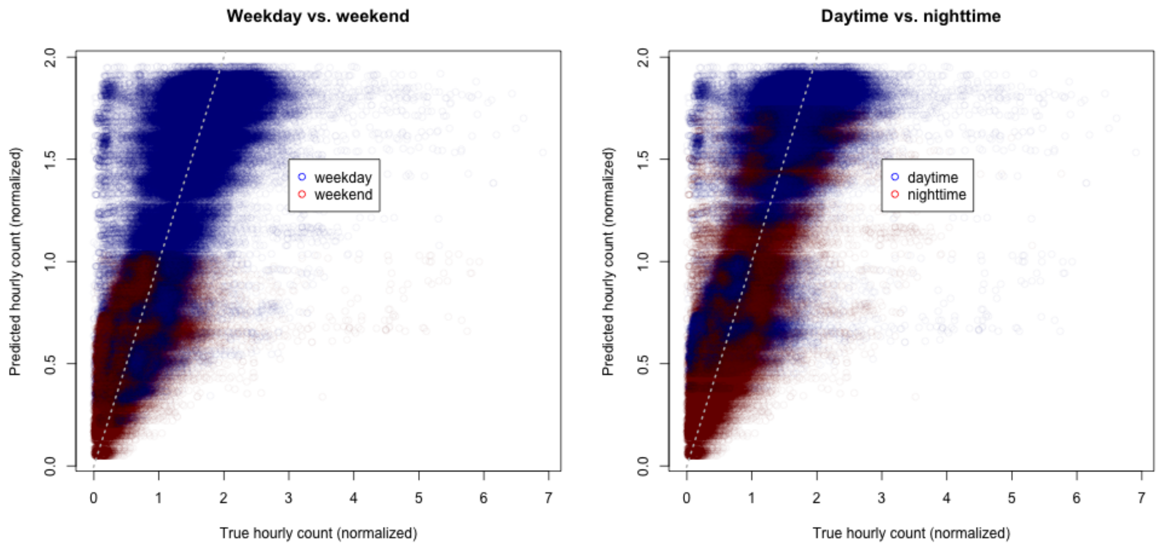


Figure 7.6: This plot of predicted vs. true normalized hourly counts values uses colors to visually investigate if the prediction is biased for certain time frames. A prediction is better if it is closer to the diagonal (white dashed line). We see that there is no indication of bias for weekend vs. weekday counts (left), and the model seems to predict approximately similarly well for nighttime and daytime hours (right).

improves inter-year comparison, as it ensures that every annual value is computed with equal weights between the weekdays. As we had a lot of incomplete and short-term counts in our datasets, we used the first (simple) method to compute the AADTT.

Further analysis and discussion As we are using only factors from daylight time, there was concern that a bias for these particular hours could introduce a bias in the monitoring model. We show the predicted hourly normalized count against the real values in Fig. 7.6. We distinguished daytime and nighttime hourly counts by colors to understand if there could be a bias in the model. From the figure, we see that there is no such concern. Similarly for weekends and weekdays, we do not find such bias.