Statistical Challenges for the Next Generation of NASA's Earth Observing Satellites

Amy Braverman

Jet Propulsion Laboratory, California Institute of Technology

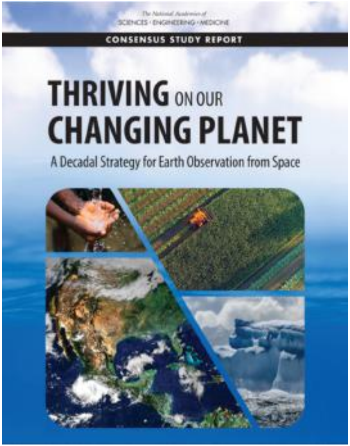
September 20, 2024

- Introduction
- Background
- Surface Biology and Geology (SBG) mission and science
- Conceptual model
- Challenges in science analysis, data processing, and uncertainty quantification
- Summary

Introduction

- Remote sensing data are the most modern, comprehensive data source for global climate and environmental science.
- Remote sensing data products are the result of complex processing chains that involve imperfect inferences and transformations.
- Statisticians have an important role to play in all levels of this enterprise: design, data collection, processing, and analysis.
- Uncertainty is universal, and must be quantified throughout the entire process to provide scientific insight and actionable information.

Introduction



See https://www.nap.edu/read/24938

- A Decadal Strategy for Earth Observation from Space for NASA, NOAA, USGS.
- Statement of task includes: “recommend NASA research activities to advance Earth system science and applications by means of a set of prioritized strategic *science targets*".
- Top-tier designated targets include **Surface Biology and Geology (SBG)**.

Introduction

The Earth Observing System (EOS) is coming to an end. Spatial scales of hundreds of meters; active and passive observation with both multispectral and hyperpectral instruments.

A new era of remote sensing observation is beginning: next generation of NASA Earth observing missions: the Earth System Observatory (ESO).

Spatial scales of tens of meters; multispectral and hyperspectral.

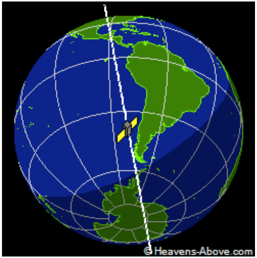
Four new missions: AOS (Atmosphere Observing System), NISAR (NASA-ISRO Synthetic Aperture Radar), **SBG** (Surface Biology and Geology), and Mass Change (no name yet).

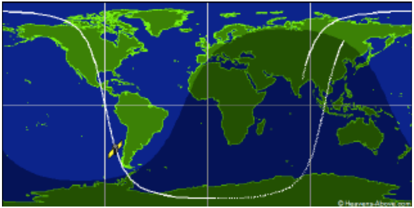
Play https://www.youtube.com/watch?v=FUq5d7dqlVY.

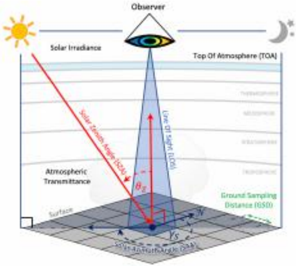
Background

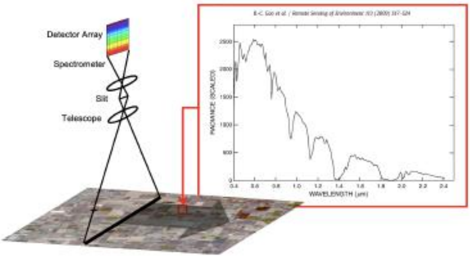
- Passive remote sensing instruments measure photon counts in bins of a discretized electromagnetic spectrum.
- The sun provides incoming photons, which are scattered and absorbed in ways that depend on the media (atmosphere or surface) with which they interact.
- Some media also emit additional photons.
- The instrument discretizes the spatial field into “footprints" and aggregates photons over both spatial footprint and spectral bin.

Background


© Heavens-Above.com


© Heavens-Above.com


Observer
Solar Irradiance
Top Of Atmosphere (TOA)
Solar Zenith Angle (SZA)
Line Of Sight (LOS)
Atmospheric Transmittance
Ground Sampling Distance (GSD)
Surface
Solar Azimuth Angle (SAA)


Detector Array
Spectrometer
Slit
Telescope
RADIANCE (SCALED)
WAVELENGTH (μm)

Background

Remote sensing levels of data processing:

- Level 0: raw photon counts direct from satellite
- Level 1: georectified and calibrated radiances
- **Level 2: estimates of geophysical state**
- Level 3: “statistical summaries" of Level 2 on uniform space-time grid
- Level 4: output of models or data assimilation

Level 2 “data" aren’t “data"; they are inferences!

When drawing scientific conclusions or making policy decisions, it is crucial to take account of uncertainties in these inferences.

Remote sensing observing system:

True state	Forward function	Noiseless radiance	Instru- ment	Obser- vation	Retrieval	State estimate

$$\mathbf{X} \longrightarrow \boxed{\mathbf{F}_0(\cdot, \mathbf{B}_0)} \longrightarrow \mathbf{Y}_0 \longrightarrow \boxed{\mathbf{Y}_0 + \epsilon} \longrightarrow \mathbf{Y} \longrightarrow \boxed{\mathbf{R}(\cdot, \mathbf{F}_1, \mathbf{B}_1 \ldots)} \longrightarrow \hat{\mathbf{X}}$$

Uncertainty quantification is estimation of $P(\mathbf{X}|\hat{\mathbf{X}})$: estimate $\mathbf{X}$ after seeing $\hat{\mathbf{X}}$.

$\mathbf{F}_0$ = nature's true forward function; $\mathbf{B}_0$ = other true quantities.

$\mathbf{F}_1$ = forward model used in retrieval, R; $\mathbf{B}_1$ = other retrieval inputs.

ϵ = instrument measurement error.

$\ldots$ = other retrieval algorithm inputs.

SBG mission and science

Surface Biology and Geology (SBG), launch date $\sim$ 2028:

- Two platforms and two (mostly) separate processing streams for two different spectral regions: visible shortwave infrared (VSWIR) and thermal infrared (TIR). Concurrent (near simultaneous) observation.
- VSWIR: for each 30 meter ground footprint, SBG-VSWIR will observe a 285-dimensional vector of radiances (Thompson et al., 2021).
- TIR: for each 60 meter ground footprint, SBG-TIR will observe an 8-dimensional vector of radiances.
- Multi-stage retrieval will estimate surface reflectances, temperature, and emissivities, from which a large number of physical variables will be derived (Cawse-Nicholson et al., 2021).
- Synergies with legacy instruments (Landsat, Sentinel-2) and ESA's CHIME mission.

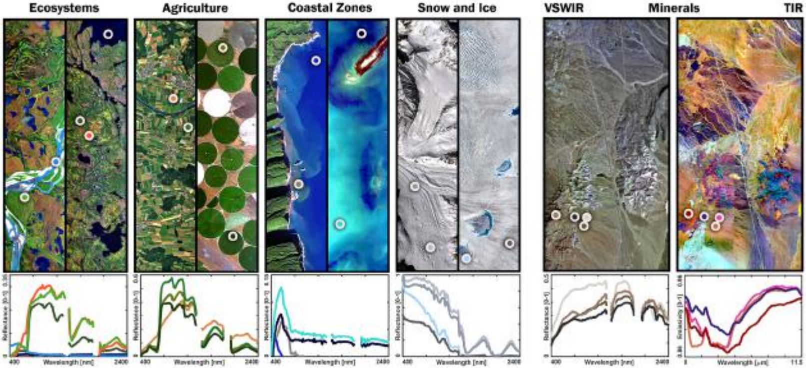
SBG mission and science

Science and applications key focii:

- active surface changes (eruptions, landslides, and evolving landscapes)
- snow and ice accumulation, melting, and albedo
- hazard risks in rugged topography
- effects of changing land use on surface energy, water, momentum, and carbon fluxes
- physiology of primary producers
- functional traits and health of terrestrial vegetation and inland and near-coastal aquatic ecosystems.
- managing agriculture and natural habitats
- water use and water quality
- urban development
- understanding and predicting geological natural hazards
- land-surface interactions with weather and climate.

SBG mission and science

Science and application areas:



Source: https://sbg.jpl.nasa.gov/doc_links/2022-05-17-sbg-community-webinar-8/2022-05-17-8th-sbg-community-webinar/

SBG mission and science

Source: https://sbg.jpl.nasa.gov/doc_links/2022-05-17-sbg-community-webinar-8/2022-05-17-8th-sbg-community-webinar/

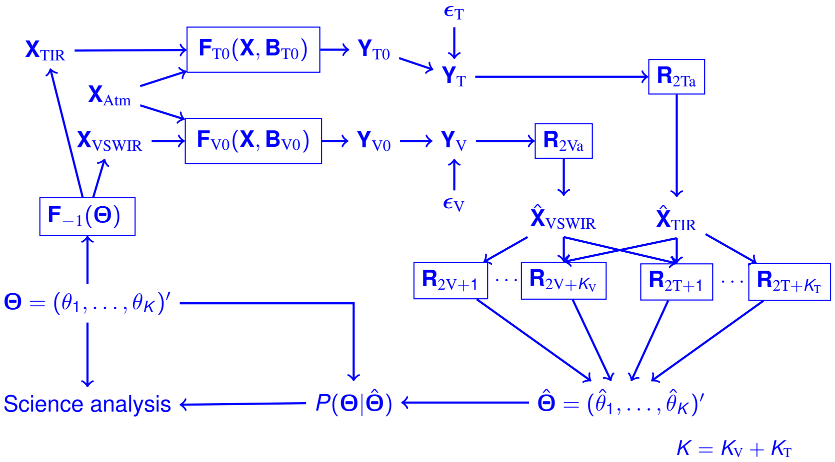
SBG mission and science

	GSD	Revisit	Spectral Range and Resolution	Sensitivity
VSWIR	30 m	16 days Global repeat coverage, ~10 days with partner	0.38 or 0.4 to 2.5μm 10nm or better, Continuous coverage	SNR: VNIR >400 SWIR >250
TIR	60 m	3 days Global repeat coverage, 1 day with partner	8 to 12μm 3 to 5μm >5 Bands desired, 8 recommended	NEdT <0.2 K

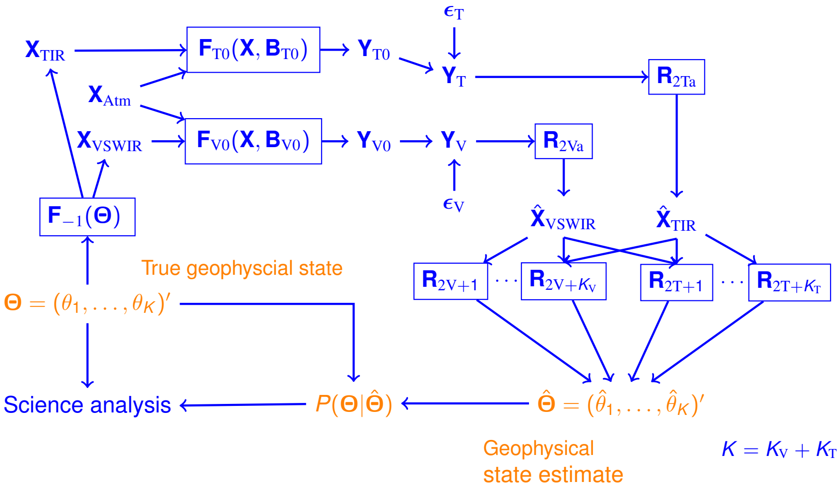
Source: https://sbg.jpl.nasa.gov/doc_links/2022-05-17-sbg-community-webinar-8/2022-05-17-8th-sbg-community-webinar/

- Daily radiance data volume will be between 2.5 and 5 TB depending on final configuration.

Conceptual model

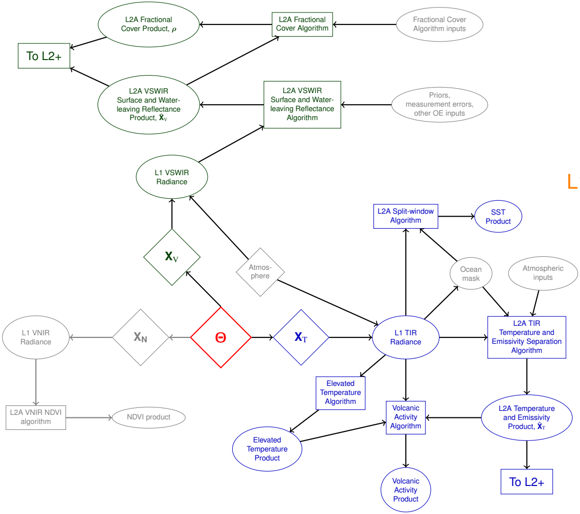


Conceptual model



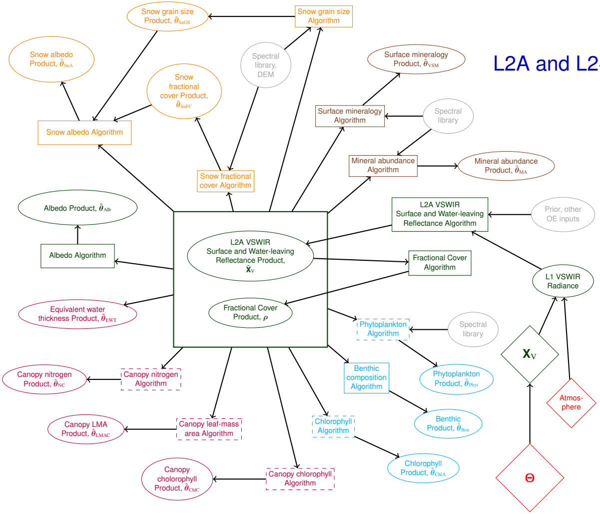
Conceptual model

L1 and L2A dependencies

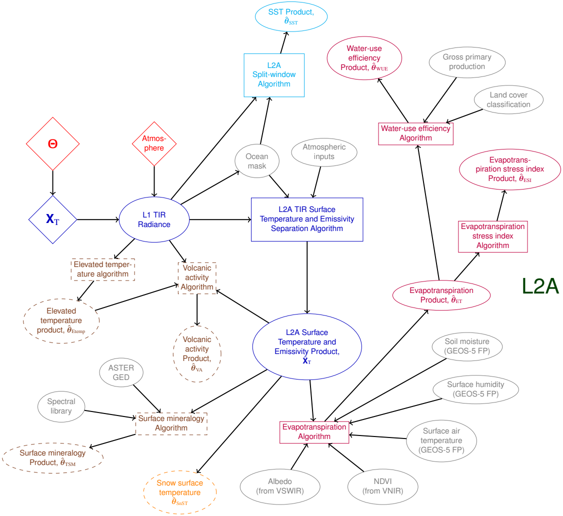


Conceptual model

L2A and L2+ dependencies (VSWIR)



Conceptual model



L2A and L2+ dependencies (TIR)

Challenges

Three kinds of problems here:

- Science analysis: analysis of $\hat{\boldsymbol{\Theta}}$.
- Data processing: statistical problems in getting from $\mathbf{Y}_{\mathrm{V}}$ and $\mathbf{Y}_{\mathrm{T}}$ to $\hat{\boldsymbol{\Theta}}$.
- Uncertainty quantification (ubiquitous).

Challenges: science analysis

A sampling of Decadal Survey science questions for SBG:

- H-1: How is the water cycle changing? Are changes in evapotranspiration and precipitation accelerating... and **how are these changes expressed in the space-time distribution** of rainfall, snowfall, evapotranspiration, and the **frequency and magnitude of extremes** such as droughts and floods?
- H-2: How do anthropogenic changes in climate, land use, water use, and water storage interact and modify the water and energy cycles **locally, regionally, and globally** and what are the **short- and long-term** consequences?
- W-3: How do **spatial variations** in surface characteristics modify transfer between domains (air, ocean, land, and cryosphere) and thereby influence weather and air quality?

Challenges: science analysis

A sampling of Decadal Survey science questions for SBG:

- E-1: What are the structure, function, and biodiversity of Earth's ecosystems, and how and why are they **changing in time and space**?
- E-2: What are the fluxes (of carbon, water, nutrients, and energy) between ecosystems and the atmosphere, the ocean, and the solid Earth, and **how and why are they changing**?
- E-3: What are the fluxes (of carbon, water, nutrients, and energy) within ecosystems, and **how and why are they changing**?
- C-3: How large are the **variations in the global carbon cycle and what are the associated climate and ecosystem impacts** in the context of past and projected anthropogenic carbon emissions?

Challenges: science analysis

Core SBG products:

Level	Domain	Processing stream: VSWIR	State variable	TIR	State variable	NIR
L2A		Surface reflectance Water-leaving reflectance Fractional Cover	$\mathbf{x}_V$ ρ	Temperature Emissivity	$\mathbf{x}_T$	Surface reflectance
L2+	Aquatics	Benthic composition Chlorophyll Phytoplankton	θ_{Ben} θ_{Chlor} θ_{Phyt}	Water temperature*	θ_{SST}	NDVI
L2+	Geology	Mineral abundance Surface mineralogy	θ_{MA} θ_{VSM}	Elevated temperature* Surface mineralogy Volcanic activity	θ_{Etemp} θ_{TSM} θ_{VA}	
L2+	Snow and ice	Snow albedo Snow fraction Snow grain size	θ_{SnA} θ_{SnFC} θ_{SnGS}	Surface temperature	θ_{SnST}	
L2+	Terrestrial ecology	Albedo Canopy chlorophyll Canopy leaf-mass area Canopy nitrogen Equiv. water thickness Fractional cover	θ_{Alb} θ_{CanCh} θ_{CanLMA} θ_{CanN} θ_{EWT} θ_{FC}	Evapotranspiration Evap. stress index Water-use efficiency	θ_{ET} θ_{ESI} θ_{WUE}	

* Derived directly from L1 radiances.

Challenges: science analysis

- Intermediate state estimates (reflectances, temperatures/emissivities) expected to be produced globally. Uncertainties “on the fly" with some form of Simulation-based Uncertainty Quantification (Braverman et al., 2021).

- Geophysical state estimates not expected to be produced globally in all cases. Algorithms in development. Uncertainty quantification approach TBD.

- Best guidance: precursor mission EMIT (Earth Surface Mineral Dust Investigation Experiment) for VSWIR. See https://earth.jpl.nasa.gov/emit/.

- Best guidance: precursor mission ECOSTRESS for TIR. See https://ecostress.jpl.nasa.gov/.

- Both EMIT and ECOSTRESS are on the International Space Station, so not in polar orbit.

Challenges: science analysis

1. Spatial and especially spatio-temporal modeling of noisy, non-smooth, non-stationary fields with missing data.

2. Data fusion: critical to use multiple variables that are produced separately, and with data from other missions (PACE, GLIMR, CHIME, etc).

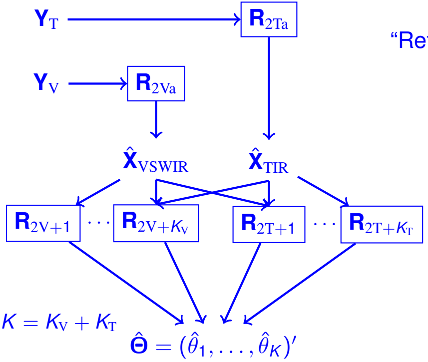
3. Time series analysis of spatially-correlated data (e.g., vegetation, biodiversity). SBG to provide unprecedented temporal frequency.

4. Causal analysis/prediction of extreme (spatial/spatio-temporal) events including harmful algal blooms, volcanic activity, wildfires.

5. Quantifying predictive relationships (in time and space) between mutiple geophysical variables and their high-order interactions.

6. All the above at <u>multiple spatial and temporal scales</u>, and with <u>massive</u> data.

Challenges: data processing

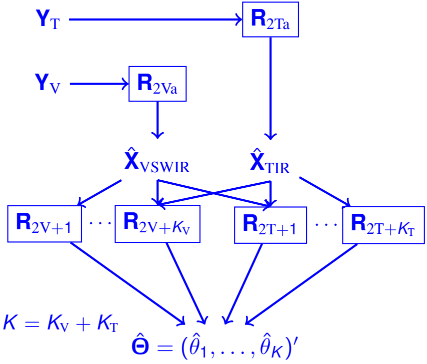


"Retrievals" are classic inverse problems, but:

- Many other inputs for which true values and uncertainties are unknown.
- Bayesian approach(es) can be used when adequate prior information exists.
- Physical forward models do not exist in many R_{2+} algorithms.
- Often: "empirical" (statistical or ML) models trained on field data and lab experiments used instead.

More modern statistical approaches would be beneficial, but speed and scientific relevance/insight are crucial.

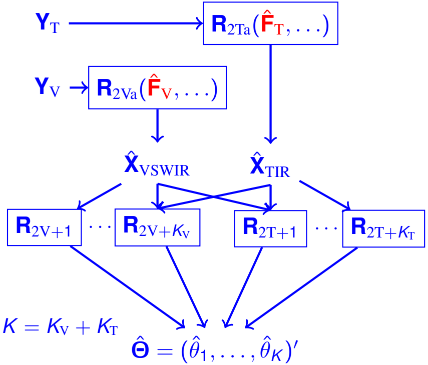
Challenges: data processing



- Starting to experiment with ML (e.g., for aquatics), but training data remains a problem.

- TES Algorithm ($\mathrm{R}_{2\mathrm{Ta}}$; Hully and Hook (2018)) is purely physcial- two-step recursive solution to under-determined set of simultaneous equations.
- Optimal Estimation ($\mathrm{R}_{2\mathrm{Va}}$; Thompson et al., (2018)) requires Gaussian assumptions and is very slow.
- Where radiative transfer is well-understood, there is often very little field data (i.e., snow and ice scenes), and visa versa (i.e., terrestrial vegetation).
- **Linear regression and related methods (partial least squares) are the workhorses for most L2+ algorithms; spatial and temporal dependence not exploited.**

Challenges: data processing

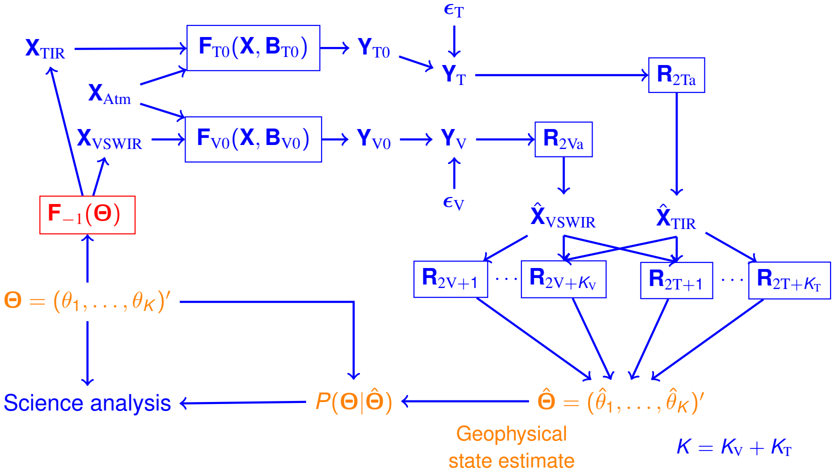


The elephant in the room:

model discrepancy

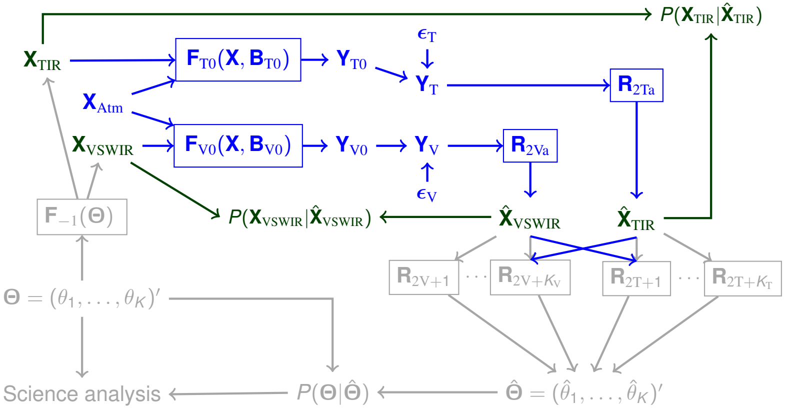
Challenges: uncertainty quantification

Full $\boldsymbol{\Theta}$-to-$\hat{\boldsymbol{\Theta}}$ simulation experiment:



Challenges: uncertainty quantification

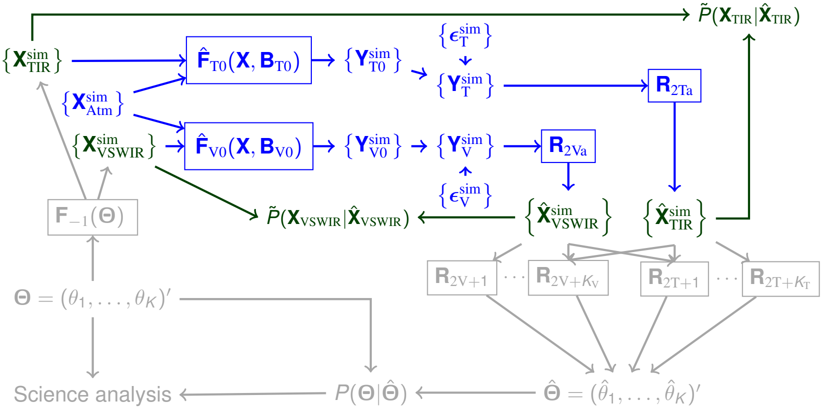
"$\mathbf{X}$-to-$\hat{\mathbf{X}}$":



Hobbs et al. (2017), Braverman et al. (2021).

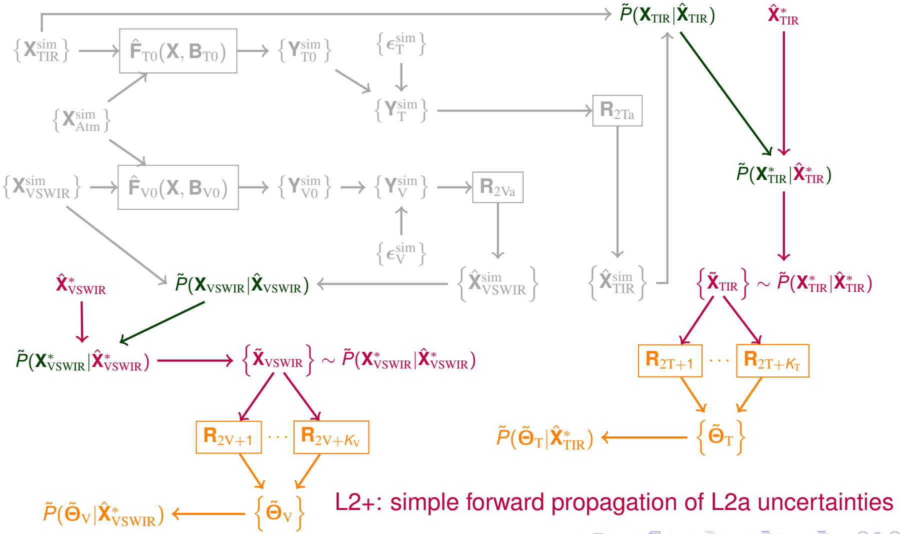
Challenges: uncertainty quantification

L2a: $\mathbf{X}$-to-$\hat{\mathbf{X}}$:



Learn conditional distributions' parameters from a simulation experiment; apply to actual retrievals.

Challenges: uncertainty quantification



L2+: simple forward propagation of L2a uncertainties

Summary

- There are an enormous number of research, implementation, and execution challenges in SBG and ESO. The cause would benefit from greater involvement by the statistics and data science community.
- The statistics and data science community will in turn benefit from the wealth of interesting, high-impact, societally-relevant problems.
- The ESO missions represent an unprecedented opportunity to have an impact on mission design and implementation. We can and should be involved from the beginning.
- SBG is in Phase A right now and open to involvement by statisticians in a way that previous missions have not been.

Summary: top statistical problems

- Emulators for complex, physical (forward) models. Possibly also emulators for retrieval (inverse) algorithms.
- Scalable uncertainty quantification methods for L2a (reflectances and temperatures/emissivities) estimates.
- Faster retrieval (inverse problem) methods, potentially exploiting spatial and spatio-temporal dependence.
- Methods beyond regression and partial least squares for L2+ retrievals.
- Distributed inference (computation).
- Data fusion (L2a: VSWIR, TIR; L2+: products on different grids; SBG + CHIME (ESA); SBG + other NASA missions).
- Dimension reduction: ask me about the "small wiggles" problem.
- Likelihood-free methods for any/all above.

References

Braverman, A., Hobbs, J., Teixeira, J., and Gunson, M. (2021). Post hoc Uncertainty Quantification for Remote Sensing Observing Systems, *Journal on Uncertainty Quantification*, Volume 9, Number 3, pages 1064?1093. doi: 10.1137/19M1304283.

Cawse-Nicholson, K. et al. (2021). NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms, *Remote Sensing of the Environment*, Volume 257, 112349. doi: 10.1016/j.rse.2021.112349.

Hulley, G.C. and Hook, S.J. (2018). ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) Level 2 Land Surface Temperature and Emissivity Algorithm Theoretical Basis Document (ATBD). Jet Propulsion Laboratory, https://lpdaac.usgs.gov/products/eco2lstev001.

Hobbs, J., Braverman, A., Cressie, N., Granat, R., and Gunson, M. (2017). Simulation Based Uncertainty Quantification for Estimating Atmospheric CO2 from Satellite Data, *Journal on Uncertainty Quantification*, Volume 5, Number 1, pages 956?985. doi: 10.1137/16M1060765.

Thompson, D. R. et al. (2018). Optimal estimation for imaging spectrometer atmospheric correction, *Remote Sensing of the Environment*, Volume 216, pages 355-373. doi: 10.1016/j.rse.2018.07.003.

Thompson, D. R. et al. (2021). NASA's Surface Biology and Geology Concept Study: Status and Next Steps, *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS, 2021)*, pages 112-114. doi: 10.1109/IGARSS47720.2021.9554480.

Acknowledgements

This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Contact information: `Amy.Braverman@jpl.nasa.gov`.