National Aeronautics and Space Administration

FY24 Topic Areas Research and Technology Development (TRTD)

New Bayesian Retrieval Methods at Scale

Principal Investigator: Jouni Susiluoto (398); **Co-Investigators:** Regina Eckert (382), David Thompson (382), Youssef Marzouk (Massacusetts Institute of Technology), Kelvin Leung (Massachusetts Insitute of Technology)

Strategic Focus Area: Uncertainty Quantification

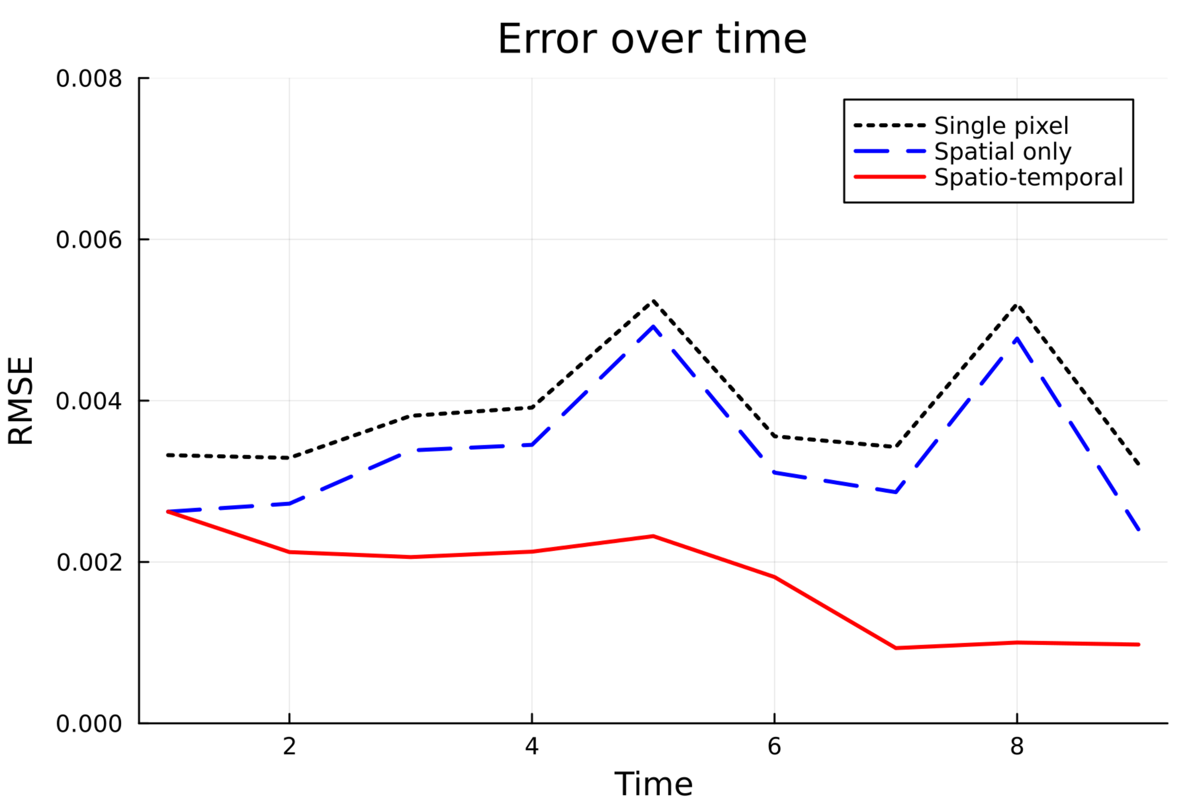
Objectives: The objective of this work is to develop and implement methods for rigorous large-scale spatio-temporal Bayesian inference to reduce errors and bias in retrievals of hyperspectral data from instruments such as AVIRIS-NG, EMIT, and SBG. This was to be carried out in a way that is mindful of maximizing data informativeness for retrieving and inferring downstream biogeophysical quantities. In practie this means using sophisticated algorithms for probabilistic analysis and goal-oriented dimension reduction.

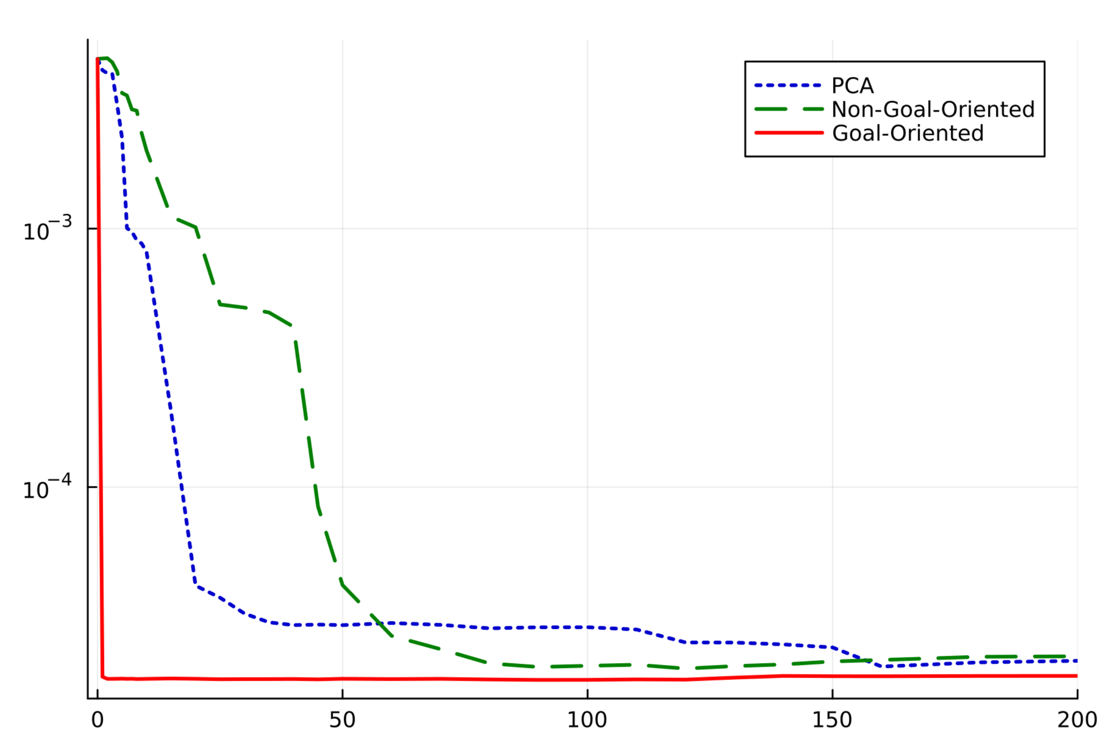
Background:Visible-shortwave infrared (VSWIR) imaging spectrometers have experienced rapid growth in use for Earth science, with instruments including EMIT and the upcoming Surface Biology and Geology (SBG) investigation, and instruments by other space agencies. These sensors aim to measure reflectance properties of the surface, but only directly observe the light incident at the sensor which is influenced by local atmospheric conditions. Consequently, in the analysis (1) at-sensor radiances are analyzed to simultaneously estimate the atmospheric state and reflectance of the surface, after which (2) the retrieved reflectance spectra are used to infer biogeophysical properties across a wide range of science disciplines. The end products, including ecosystem functional traits, surface composition, and snow state, will be a cornerstone of SBG's contribution to NASA's Earth System Observatory in the coming decade. However, current operational retrievals ignore the spatially-dense nature of the acquisition. Moreover, they treat each acquisition independently, ignoring the repeated biweekly observations that will accrue from sun-synchronous orbiters like SBG. In short, they treat all data independently, under-utilizing spatiotemporal covariance information. Accounting for this covariance information, together with goal-oriented dimension reduction that optimally constrains the downstream quantities of interest, dramatically reduces error in the products that these missions generate

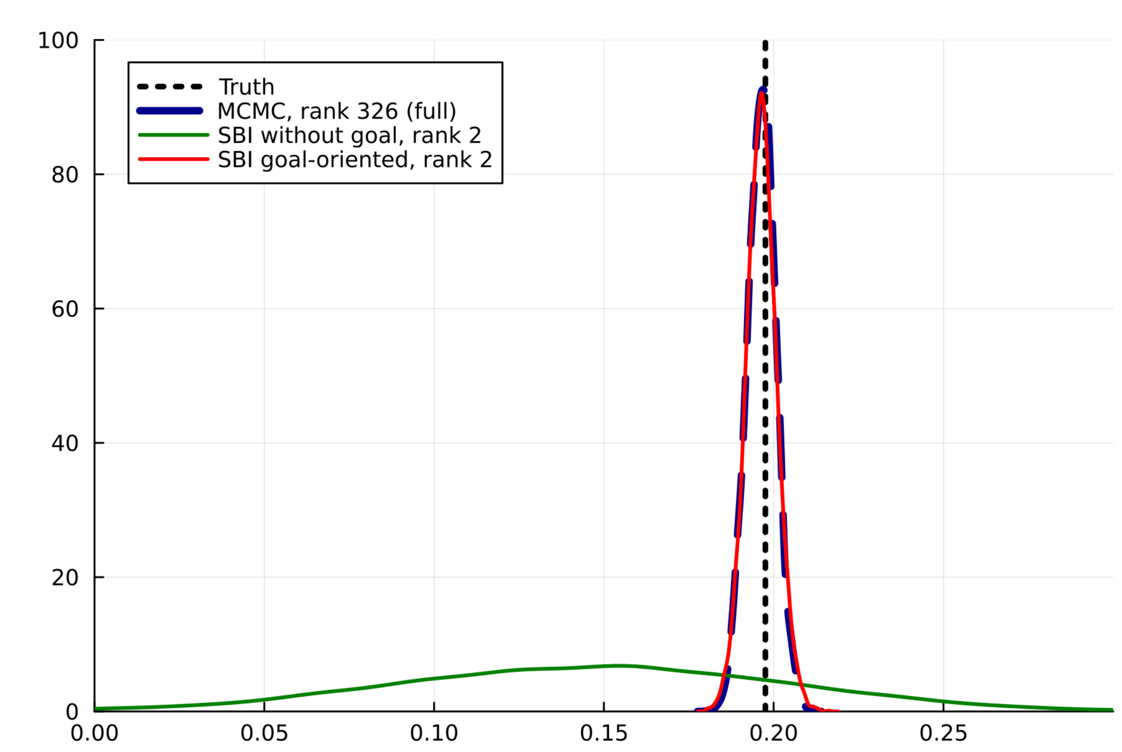
Approach and Results: This work has two parts: large-scale retrievals, and goal-oriented dimension reduction for those retrievals. We formulate the hyperspectral spatio-temporal L2 surface reflectance retrieval as *probabilistic graphical model* for looking at joint probabilities and the conditional independence structure of a collection of random variables. Assuming a spatio-temporal field of instrument-measured high-dimensional radiances, we model the atmosphere as a spatially smooth field, and surface as a temporally slowly-changing field. We use a lightweight Julia-based retrieval framework for the inference, with test data based on the recent SHIFT measurement campaign in the Santa Barbara region. We use the SHIFT (AVIRIS-NG instrument) radiances to generate ground truth states. Comparing spatio-temporal and independent retrievals, errors decrease over time in the former, while this does not happen in the latter. In Fig 1 the error is averaged over all 1200 pixels in the domain of interest, showing average error per spectral band at each time step for pixel-by-pixel retrievals (left) and spatio-temporal retrievals (right). Note the different y-axis scales. The error goes down for later time steps in the spatio-temporal case, whereas such trend is not visible in the beginning. Moving on from the spatio-temporal retrievals, in the goal-oriented setting, reflectance (x) can be treated as a nuisance parameter, since oftentimes we are just interested in a downstream application such as canopy water, minerals, or snow. Framing the problem as a Bayesian Inverse Problem, we take advantage of the dimension of the QoI (z) being low while that of radiance (y) is high. The naive approach (that we use for control) would characterize p(z|x)p(x|y) and marginalize over x with MCMC, which is very slow. For goal-oriented dimension reduction we use the technique due to Spantini et al. (2017), and utilize simulation-based inference for marginalization. More formally, we generate joint sample pairs (y,z), which we then use with Conditional Optimal Transport. While dimension reduction can be done either for parameters or data, here we are interested in data-space dimension reduction. For the high-dimensional vector, already two dimensions capture 99% of the energy of the posterior distribution of canopy water content in our simulations. This is fifteen times less than the standard non-goal-oriented Principal Component-based dimension reduction, while giving much more accurate results. The computational benefits are also substantial: the current code gives MCMC results in an hour, whereas the transport-based inversion gives the same results in three seconds.

Significance/Benefits to JPL and NASA: We have for the first time demonstrated that carrying out spatio-temporal hyperspectral surface reflectance retrievals is computationally feasible. The uncertainty reduction that we obtain is substantial and hence highly significant. The results will help guide future uncertainty quantification efforts of hyperspectral missions. We have further demonstrated how goal-oriented dimension reduction will allow for fast non-Gaussian inference of downstream quantities of interest, verifying the accuracy of the uncertainties produced by comparing to MCMC-based integration over nuisance variables (reflectance). This work will help reinforce the leadership of JPL in the arena of large-scale hyperspectral retrievals. The results and algorithms may be adapted to other missions and retrieval frameworks. The team is working towards introducing / enhancing the spatial and spatio-temporal modeling in upstream operational retrievals based on the lessons learned from developing the algorithms and software.

Figures: **Left**: Spatio-temporal error shows error reductions as time progresses compared to single-pixel or spatial-only modeling. **Center**: goal-oriented dimension reduction is demonstrated to be much more advanced than for example PCA. **Right**: Full-scale MCMC can verify that measure transport-based inference can yield accurate non-Gaussian posteriors, while being 1200x faster (MCMC: 1h, transport: 3s)







National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

RPD-000 Clearance Number: CL#00-0000

Publications:

K. M. Leung, D. R. Thompson, J. Susiluoto, J. Jagalur, A. Braverman and Y. Marzouk. *Evaluating the Accuracy of Gaussian Approximations in VSWIR Imaging Spectroscopy Retrievals*, in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-12, 2024, Art no. 4507512.

PI/Task Mgr. Contact Information:
Dr. Jouni Susiluoto (jouni.i.susiluoto@jpl.nasa.gov)