Jet Propulsion Laboratory
California Institute of Technology

Optimal Estimation for Earth Science Imaging Spectroscopy with Multivariate Uncertainty Analysis

David R. Thompson, Robert O. Green, Joseph Boardman[1], Philip Dennison[2], Michael Eastwood, Mark Helmlinger
Jet Propulsion Laboratory, California Institute of Technology
[1] Analytical Imaging and Geophysics, LLC
[2] University of Utah

 This research has been performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. NASA programmatic support through ESTO and Terrestrial Ecology programs is gratefully acknowledged.

Agenda

1. Review HyspIRI prototype data products
2. Motivation and foundations for uncertainty quantification
3. A simple example using L2 and L3 analyses
4. Initial conclusions and ideas for future investigation

A Growing Operational Data Catalog

Calibrated Radiances

2013 - Surface Reflectances

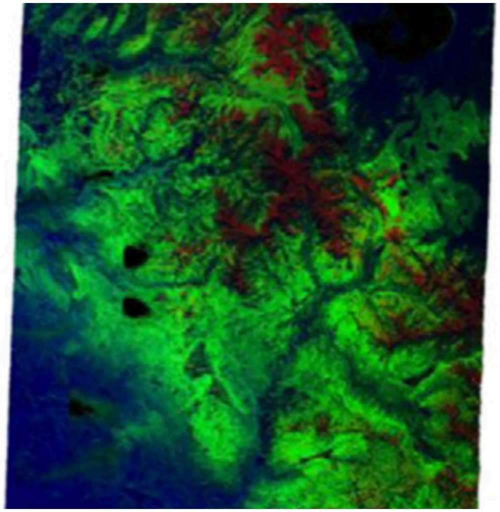
2014 - Liquid, Ice, Vapor H_2O

2015 - CH_4 Point Sources

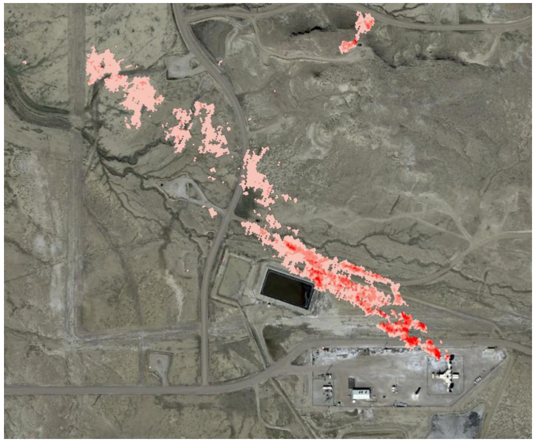
2016 - Benthic Reflectances

2016 - Terrestrial Coverage Fractions (NPV/PV/S)

2016 - Benthic Classification Maps

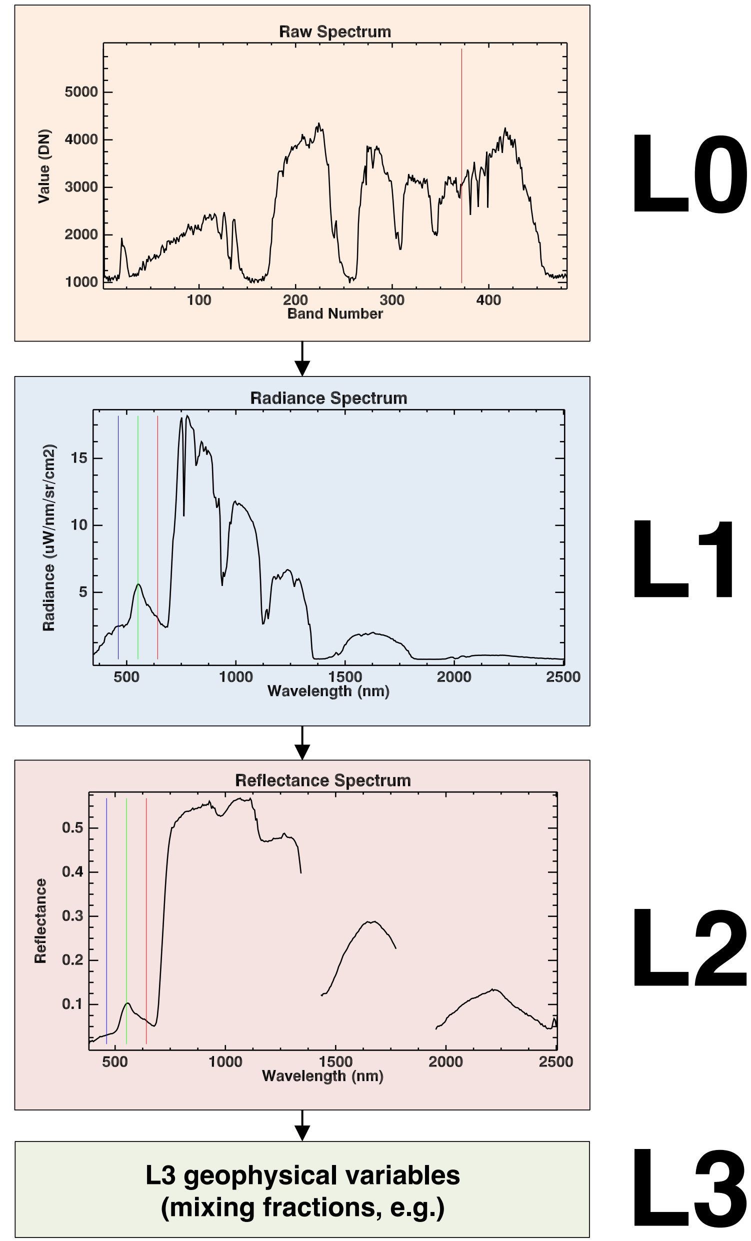


H_2O Ice, Vapor, Liquid
[Thompson et al., RSE 2015]

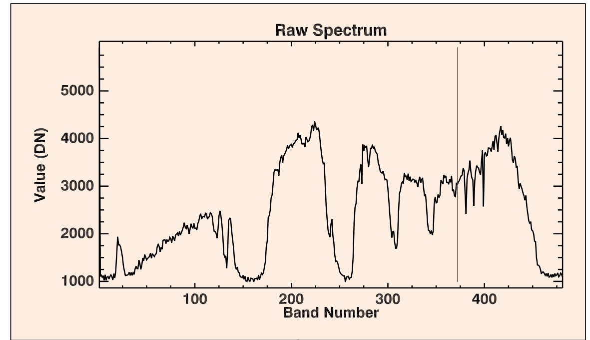


CH_4 enhancement
[Frankenberg et al., PNAS 2016]

HyspIRI prototype analysis

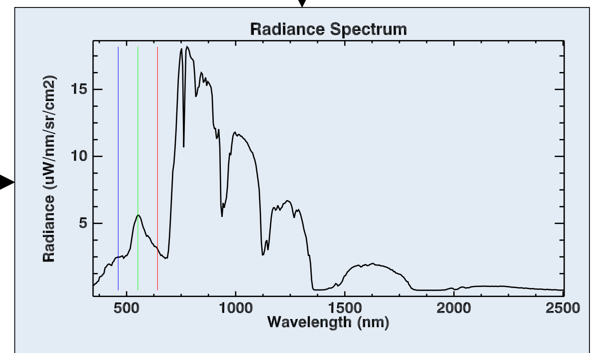


HyspIRI prototype analysis



L0

Laboratory Calibration Data

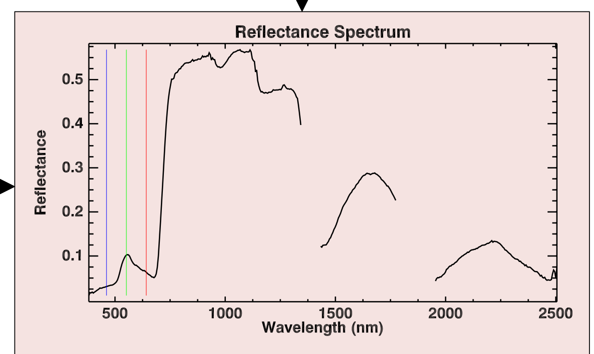


L1

Numerical approximations, decoupling

Atmospheric models

Retrieved atmospheric state



L2

Surface BRDF models

Reflectance libraries

Numerical approximations (linear mixing)

L3 geophysical variables (mixing fractions, e.g.)

L3

Notable previous studies in uncertainty quantification

- Optimal Estimation [Rogers 2002]
- Uncertainty quantification and propagation for OCO-2 and other instruments [Hobbs, Braverman, et al., 2014]
- Comprehensive instrument, measurement & retrieval modeling of CRISM [Parente et al., 2010]
- Data-driven noise estimation [e.g. Meola et al., 2011]

Our Questions

1. **Statistical modeling:** What is the true uncertainty in product inputs and outputs?
2. **Data system design:** How should we summarize uncertainty and communicate it across product levels?
3. **Retrieval algorithms:** How can uncertainty propagation improve downstream analyses?
4. **Performance:** What are the potential accuracy benefits?

Measurements are statistical objects

Red: reported

Point estimate of geophysical variable $\boldsymbol{\theta}$ given measurement x

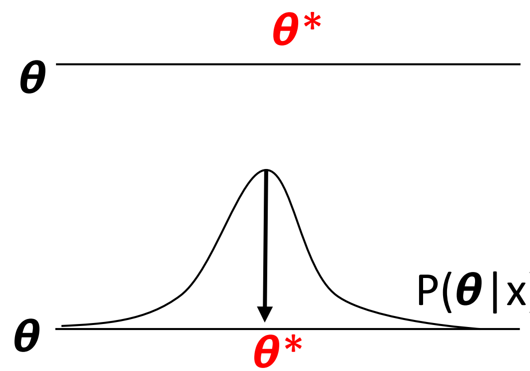
$\boldsymbol{\theta}$ ———— $\boldsymbol{\theta^*}$

Measurements are statistical objects

Red: reported

Point estimate of geophysical variable $\boldsymbol{\theta}$ given measurement x

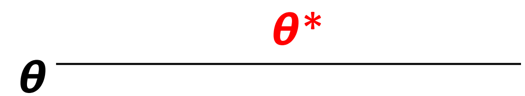
Maximum A Posteriori estimate



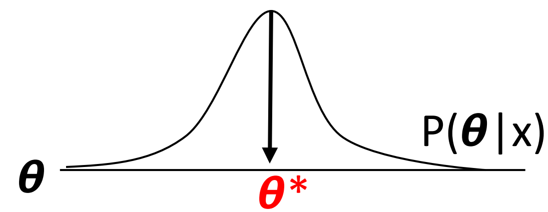
Measurements are statistical objects

Red: reported

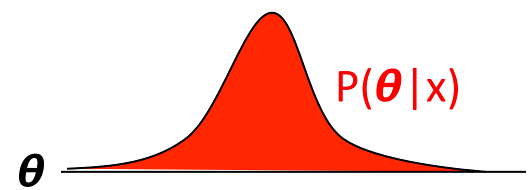
Point estimate of geophysical variable $\boldsymbol{\theta}$ given measurement x



Maximum A Posteriori estimate



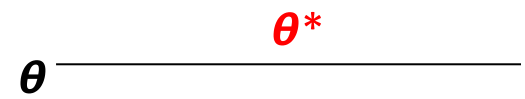
Full or Constrained Gaussian



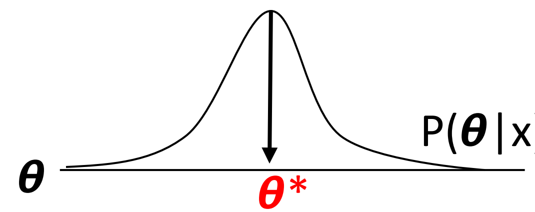
Measurements are statistical objects

Red: reported

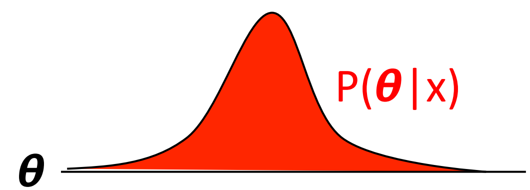
Point estimate of geophysical variable ***θ*** given measurement x



Maximum A Posteriori estimate

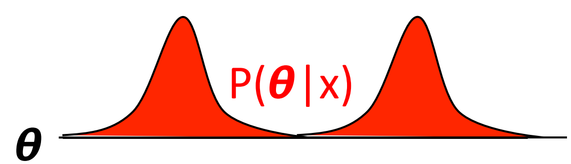


Full or Constrained Gaussian



Full Posterior Distribution

(not necessarily Gaussian)



Conditional independence permits efficient factorization

Probability distributions encapsulated in products allow *tractable* and *exact* L3 posterior distributions.

$$P(L \mid D)$$

Radiometry

$$P(r_s \mid D) = \int_L P(r_s \mid L) \underbrace{P(L \mid D)}_{\text{L1 file}} \, dL$$

Atmospheric correction

$$P(m \mid D) = \int_{r_s} P(m \mid r_s) \underbrace{P(r_s \mid D)}_{\text{L2 file}} \, dr_s$$

Retrieval of geophysical quantities

A simple example: Reporting L2 reflectance uncertainty to L3 unmixing

Simulated retrievals

1. L1 with added noise simulating HyspIRI [Dennison et al.] and Lorentzian PSF tails
2. L2 using RTM model of variable atmospheric conditions
3. L3 based on nonnegative least squares of P, NPV, GV endmembers

Compare three alternative unmixing methods using:

1. No error estimate (point mass probabilities)
2. Multivariate Gaussians, diagonal covariance
3. Multivariate Gaussian distributions, full covariance

Adapting least-squares estimation for correlated input uncertainties

Standard least squares fitting error

$$err(x,\theta)= 1/n \sum \lambda \uparrow (x - x) \uparrow 2$$

Mahalanobis distance – weight by inverse of error distribution

$$err(x,\theta)=1/n \sum \lambda \uparrow ((x - x)-\mu) \uparrow T \; C \uparrow -1 \; ((x - x) - \mu) @ \quad 1/n \sum \lambda \uparrow [((x - \mu) C \uparrow -1/2 - (x - \mu) C \uparrow -1/2)] \uparrow 2$$

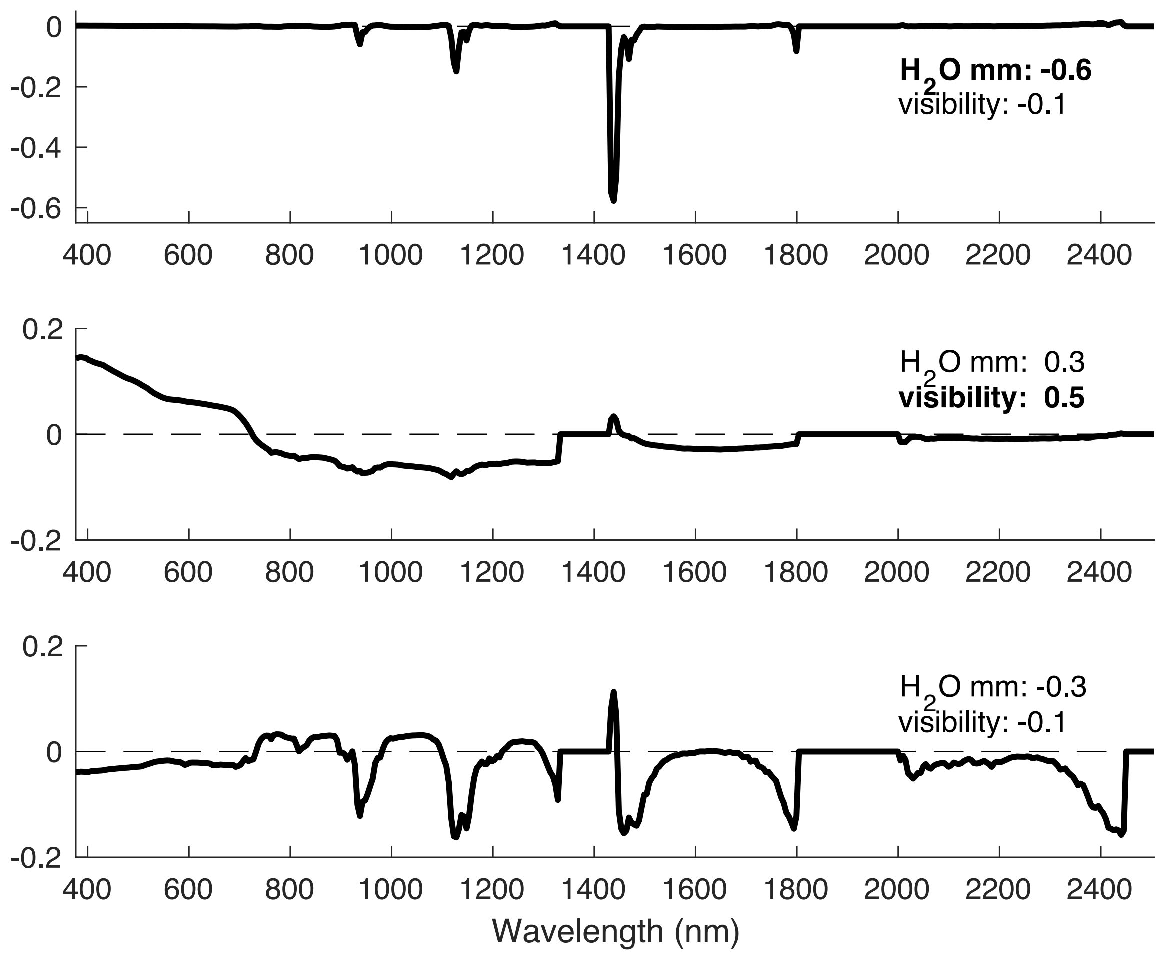
Amounts to a whitening pre-transformation of both libraries and measured reflectance.

Errors in reflectance (eigenvectors) and correlations with errors in atmospheric state

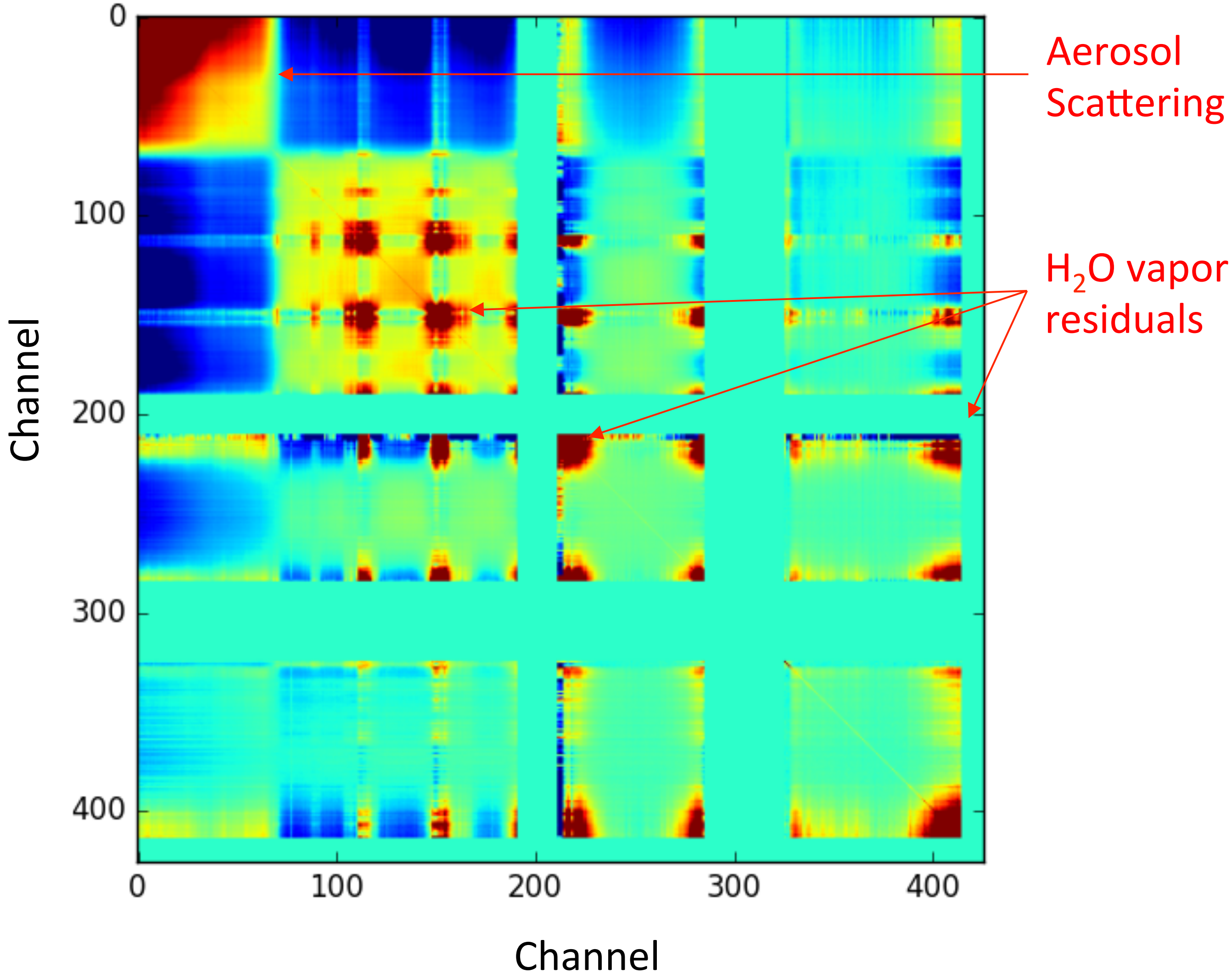
EOF #1

EOF #2

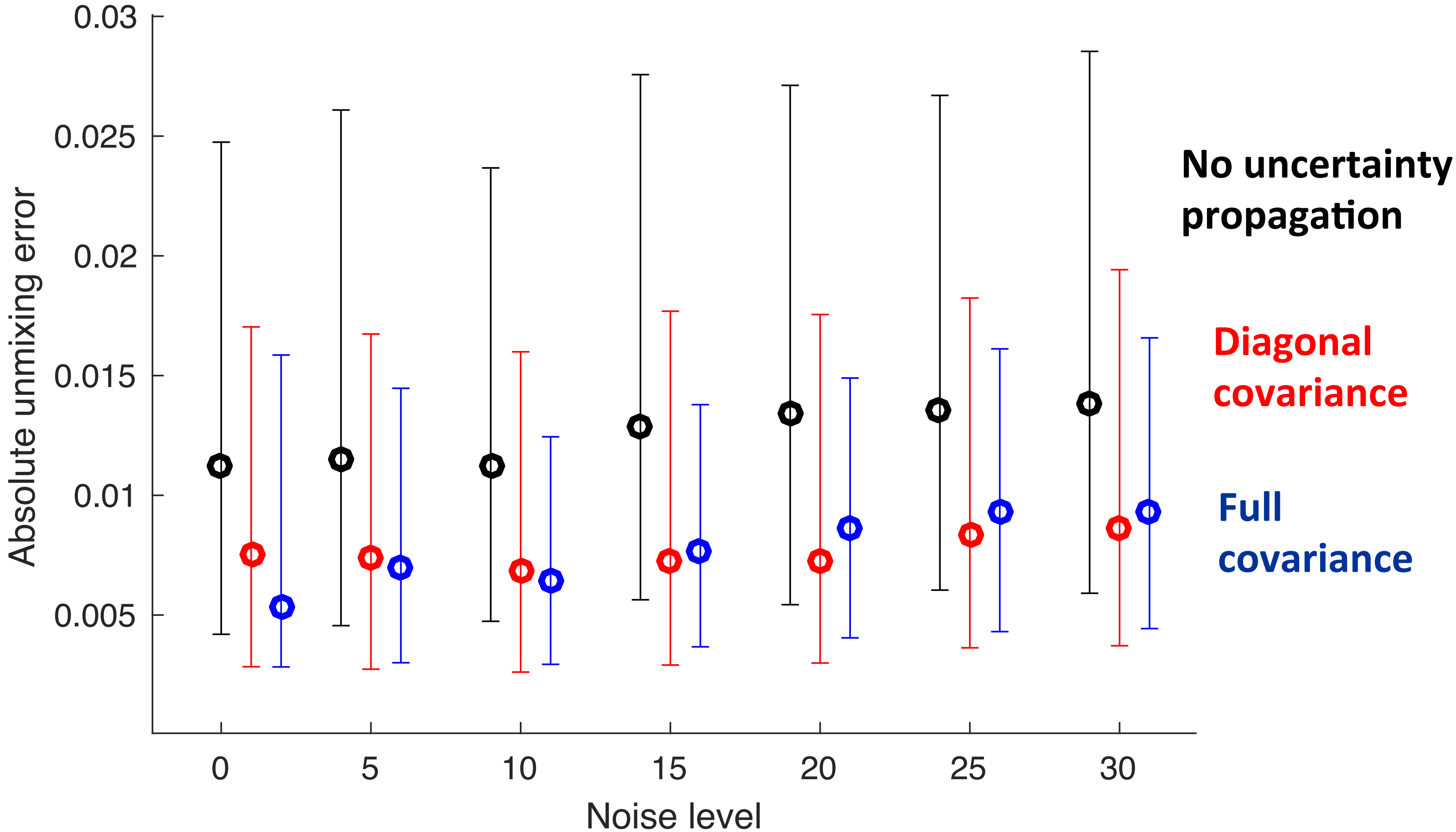
EOF #3



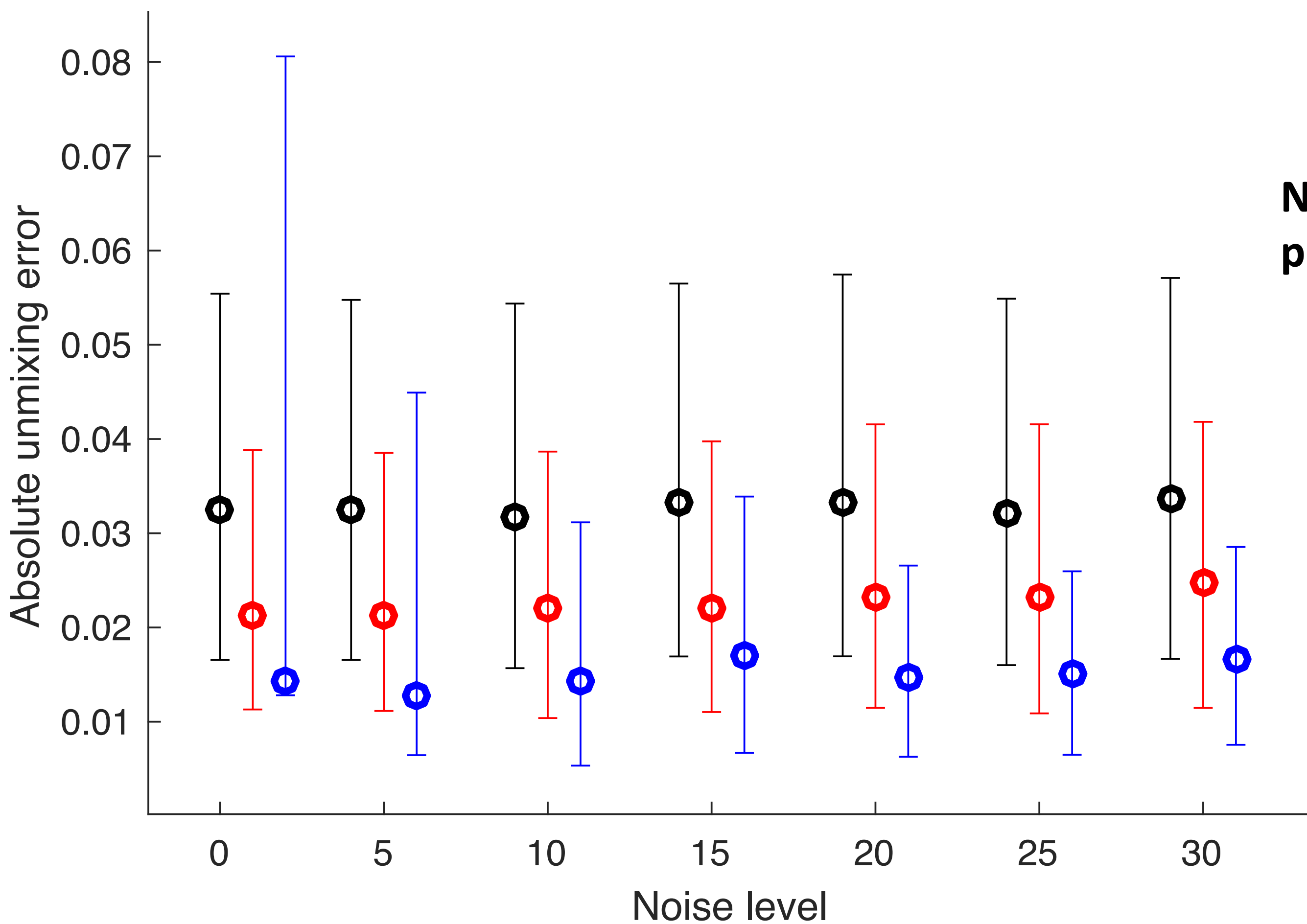
Error covariance matrix: r_s



Impact of uncertainty propagation, Case 1: Perfect radiances



Impact of uncertainty propagation, Case 2: Some radiance errors



No uncertainty propagation

Diagonal covariance

Full covariance

Answers to initial questions

1. **Statistical modeling:** What is the true uncertainty in product inputs and outputs? **Significant correlations** can exist across channels

2. **Data system design:** How should we summarize uncertainty and communicate it across product levels? **Full covariances** may outperform channelwise error reporting

3. **Retrieval algorithms:** How can uncertainty propagation improve downstream analyses? **Simple data pre-transformations** for many existing least-squares algorithms

4. **Performance:** What are the potential accuracy benefits? **Potentially significant**, even for simple idealized cases

Thanks!

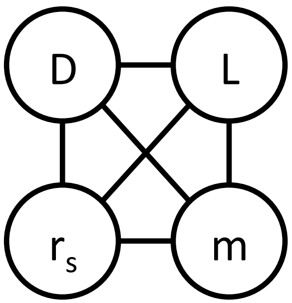
- **NASA Earth Science Division** and HyspIRI preparatory campaign
- **The AVIRIS-C and AVIRIS-NG flight teams**, including Sarah Lundeen, Ian McCubbin, and Charles Sarture. .

AVIRIS-C data is available from
http://aviris.jpl.nasa.gov

AVIRIS-NG data is available from
http://avirisng.jpl.nasa.gov

Backup slides

Is estimating posterior uncertainty even tractable?

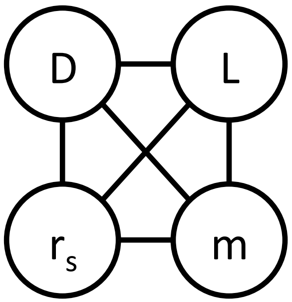


mixtures radiance

$$P(m \mid D) = \int_{r_s} \int_{L} P(m,\ r_s,\ L \mid D)\, dr_s\, dL$$

raw reflectance

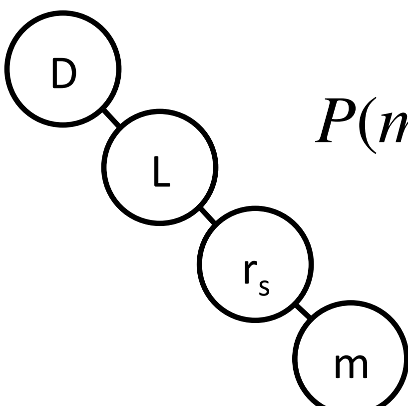
Is estimating posterior uncertainty even tractable?



mixtures radiance

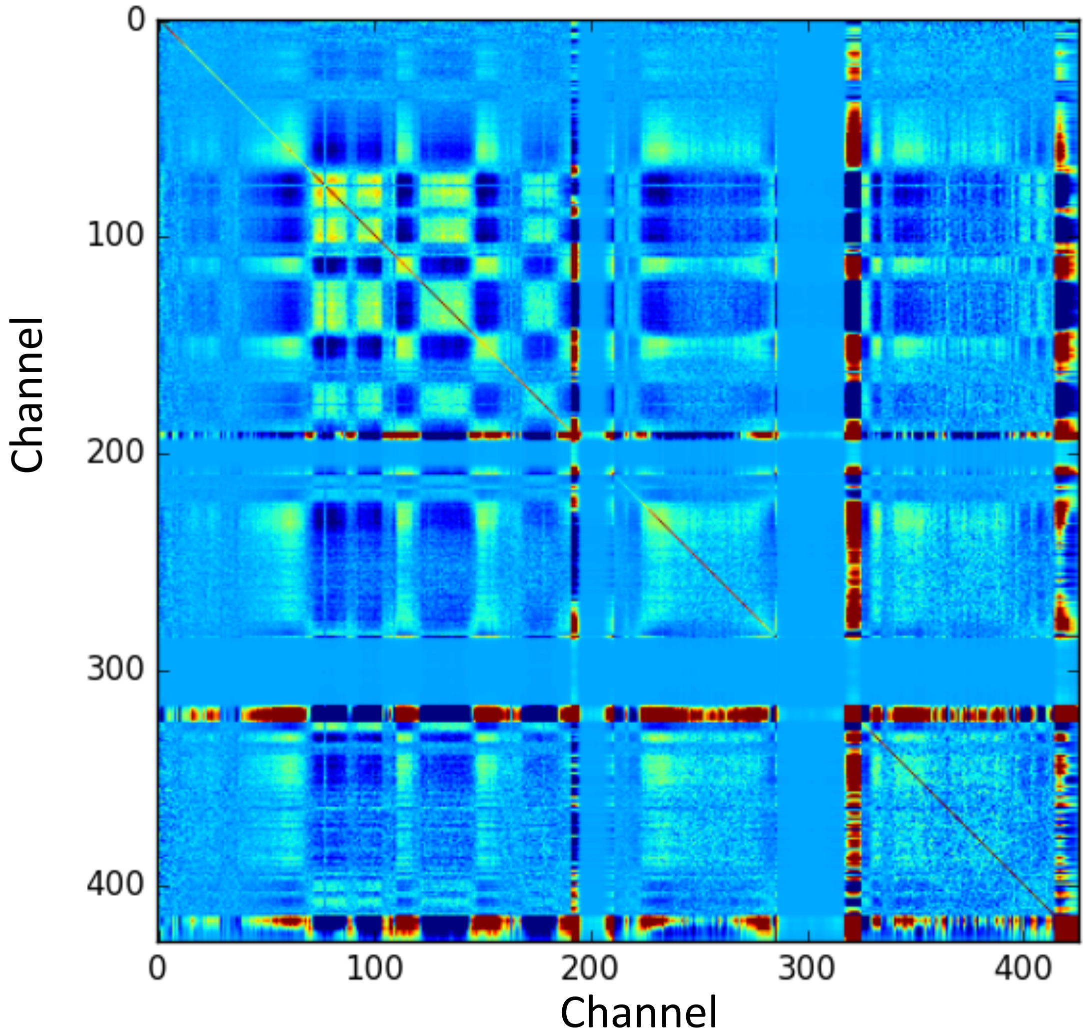
$$P(m \mid D) = \int_{r_s} \int_L P(m,\ r_s,\ L \mid D)\, dr_s\, dL$$

reflectance raw



$$P(m \mid D) = \int_{r_s} \underbrace{P(m \mid r_s)}_{\text{Retrieval of geophysical quantities}} \int_L \underbrace{P(r_s \mid L)}_{\text{Atmospheric correction}} \underbrace{P(L \mid D)}_{\text{Radiometry}}\, dr_s\, dL$$

Error covariance matrix: ρ_{toa}



Example state vector

Parameter	Degrees of Freedom	True Distribution
Radiometric gain	1	Gaussian
Radiometric offset	1	Gaussian
Spectral FWHM	1	Gaussian
Lorentzian PSF fraction	1	Gaussian
Aerosol Optical Depth	1	Exponential
Water vapor	1	Exponential
Constituents	~3	Multinomial discrete
Mixing Fractions	~2	Dirichlet
Total for $P(L_3 \mid L_0, L_2, L_3)$	**15**	**Fully connected graph?**