Optimizing Satellite Datasets for Cloud and Rainfall Assimilation Using Information Theory

Tristan L'Ecuyer, Colorado State University
Tom Greenwald, CIMSS/U. Wisconsin

Acknowledgement: R. Bennartz, C. O'Dell, and A. Heidinger

The Value of Information Content Analyses

- **Data Thinning**: Allows measurements to be used in the assimilation to be optimized to maximize the information they provide while minimizing redundancy (analogous to targeting observations);
- **Error Propagation**: Requires both model and measurement errors to be assessed defining the weights that should be assigned to each measurement in order to accurately represent their information relative to other sources in the assimilation system;
- **Evaluation**: Provides a framework for directly comparing the value of different observational datasets.

Entropy-Based Information Content

- Information content of an observing system is the difference in entropy between an a priori set of possible solutions, S(P1), and the subset of these solutions that also satisfy the measurements, S(P2) (Rodgers, 2000):

$$H = S(P_1) - S(P_2)$$

- If Gaussian distributions are assumed for the prior and posterior state spaces, this can be written:

$$\mathbf{H} = \frac{1}{2}\log_2\left|\mathbf{S}_1\mathbf{S}_2^{-1}\right| = \frac{1}{2}\log_2\left|\mathbf{S}_a\left(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_a^{-1}\right)\right|$$

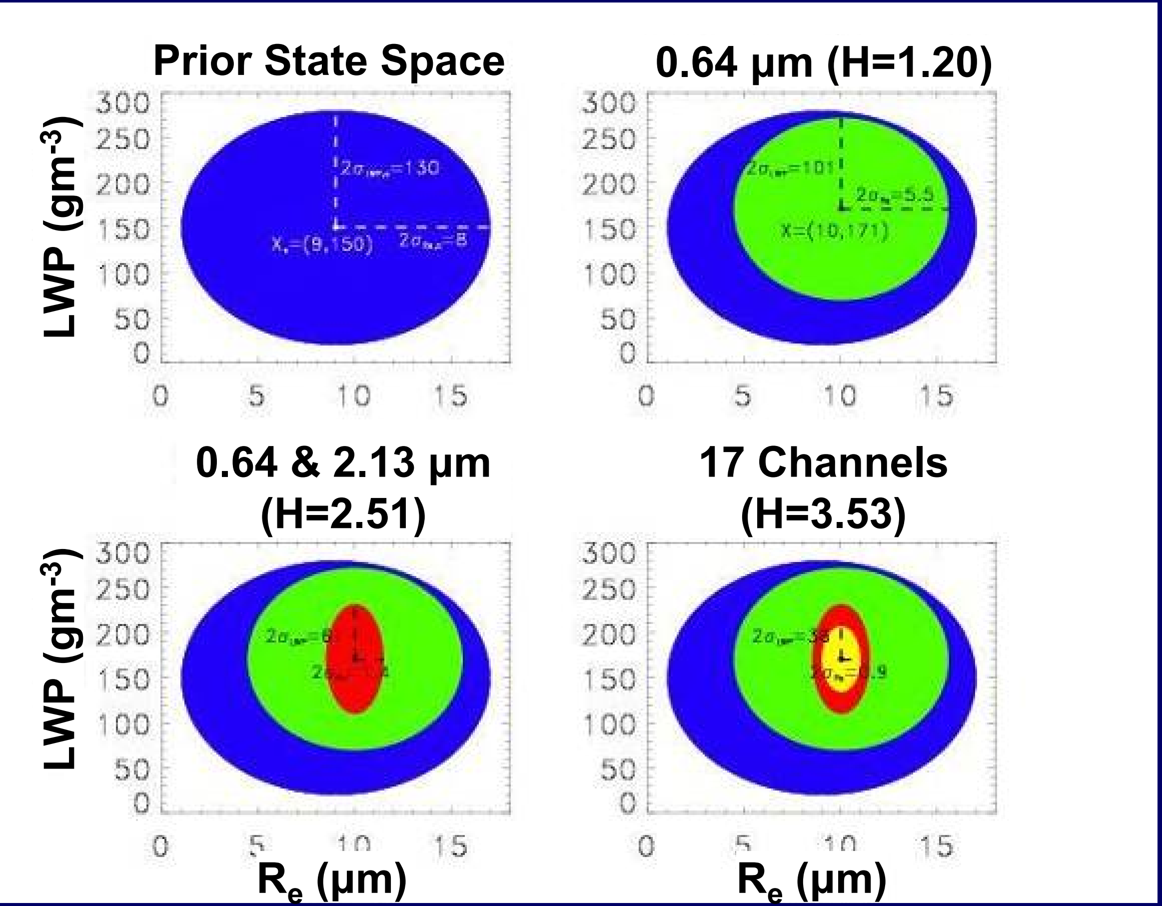
where $\mathbf{S}_a$ is the covariance matrix describing the prior state space and it has been assumed that the posterior space covariance is defined by:

$$\mathbf{S}_x = \left(\mathbf{S}_a^{-1} + \mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K}\right)^{-1}$$

K represents the Jacobian of the measurements with respect to the parameters of interest.

Liquid Cloud Microphysical Properties

- Information content measures the extent to which the solution space is reduced as measurements are added.



Channel Selection

The information content of individual channels can also be assessed via:

$$\mathbf{H}_j = \frac{1}{2}\log_2\left(1 + \mathbf{k}_j\mathbf{S}_a\mathbf{k}_j^T\right)^{-1}$$

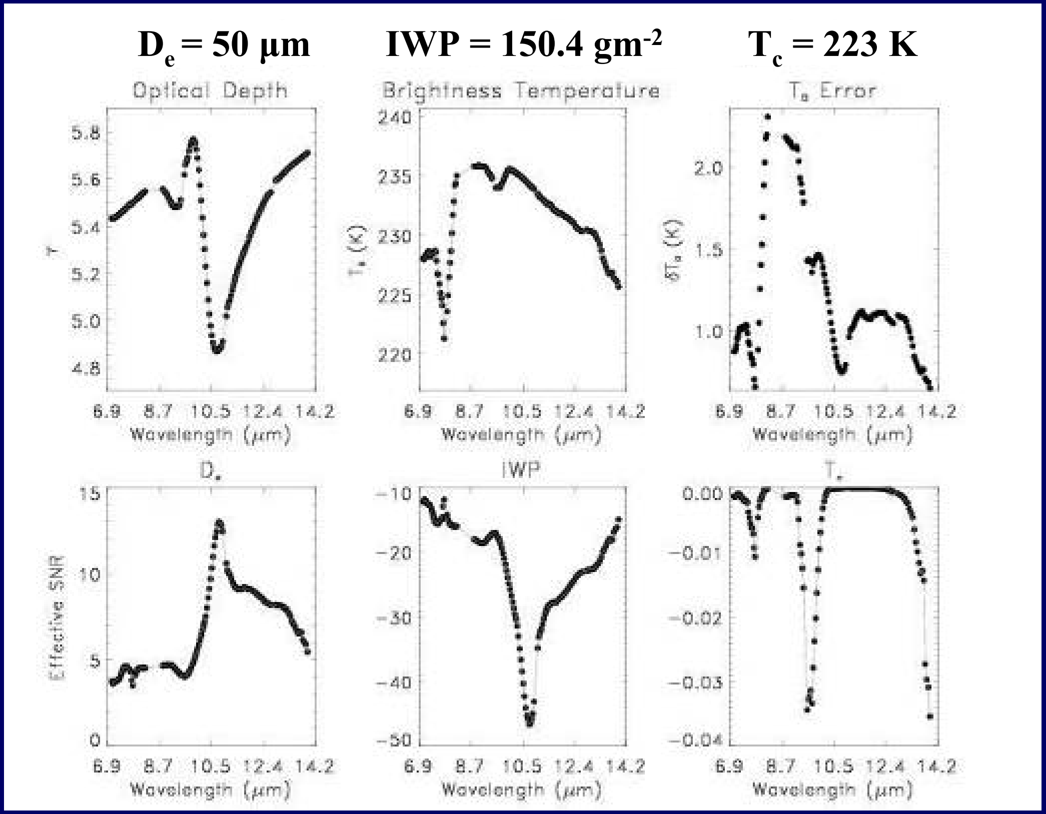
where $\mathbf{k}_j$ is the row of **K** corresponding to channel j. The channels providing the greatest amount of information can then be sequentially selected by adjusting the covariance matrix via:

$$\mathbf{S}_l^{-1} = \mathbf{S}_a^{-1} + \mathbf{k}\mathbf{k}_l^T$$

Infrared Observations of Ice Clouds

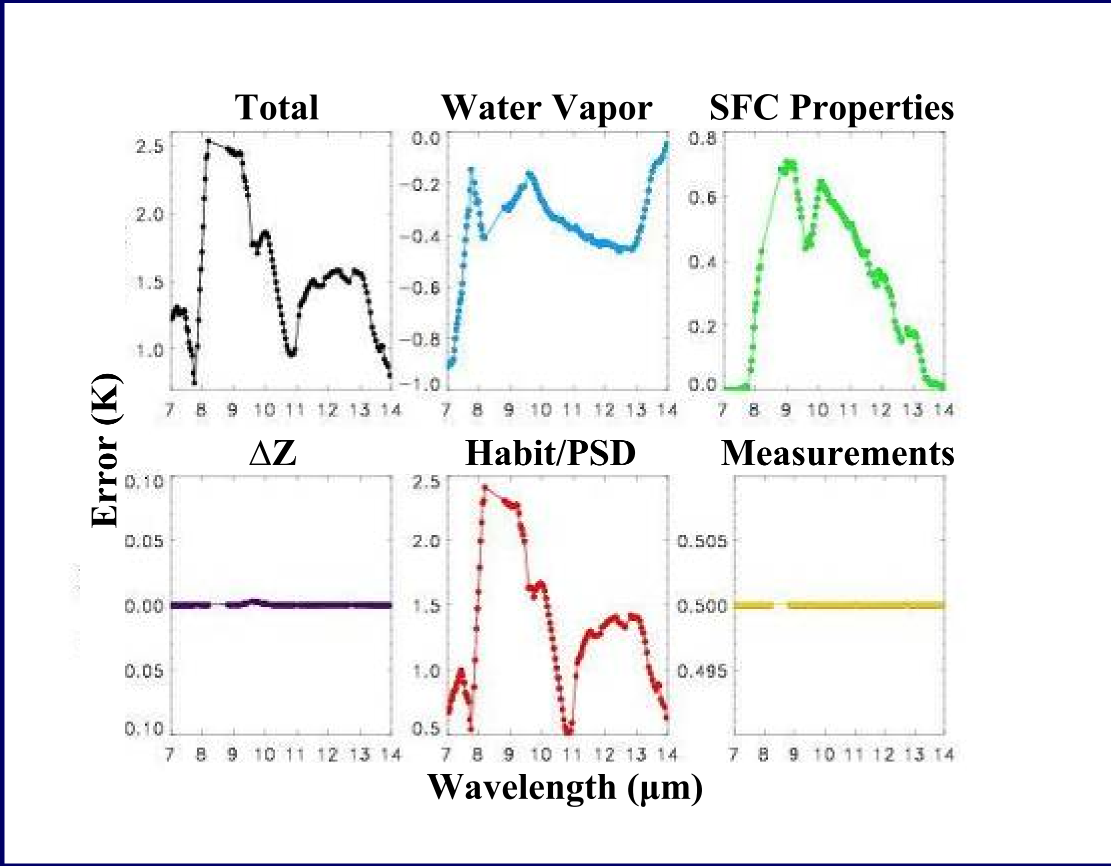
Example: The design of a new IR radiometer

- Retrieval of IWP and effective diameter
- Using 2 or 3 IR bands from 7-14 μm

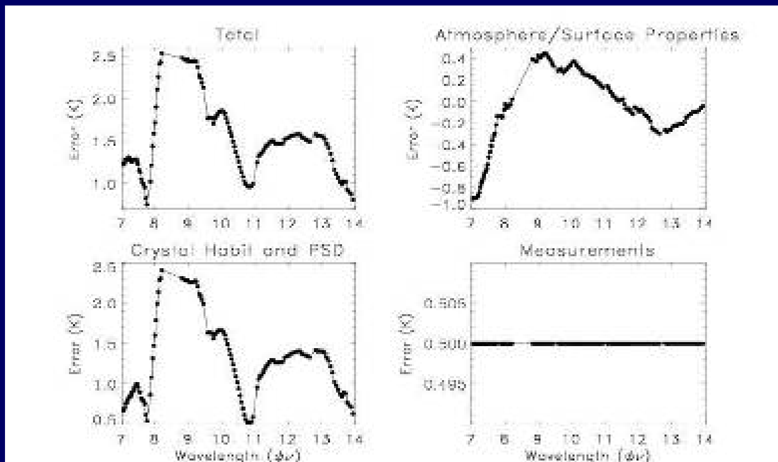


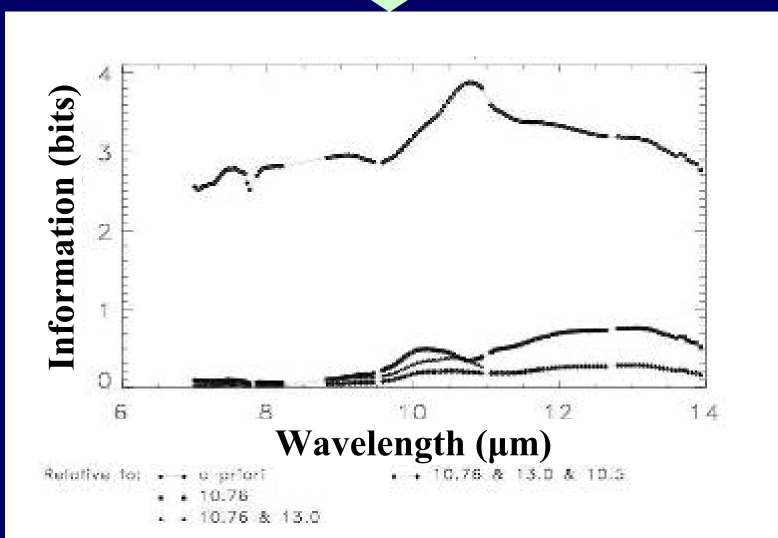
Error Characterization

- Perturb assumed parameters over their expected dynamic range:
  - PSD shape
  - Crystal habit
  - Surface T
  - Surface ε
  - Cloud geometric thickness
  - q profile
- Add measurement error

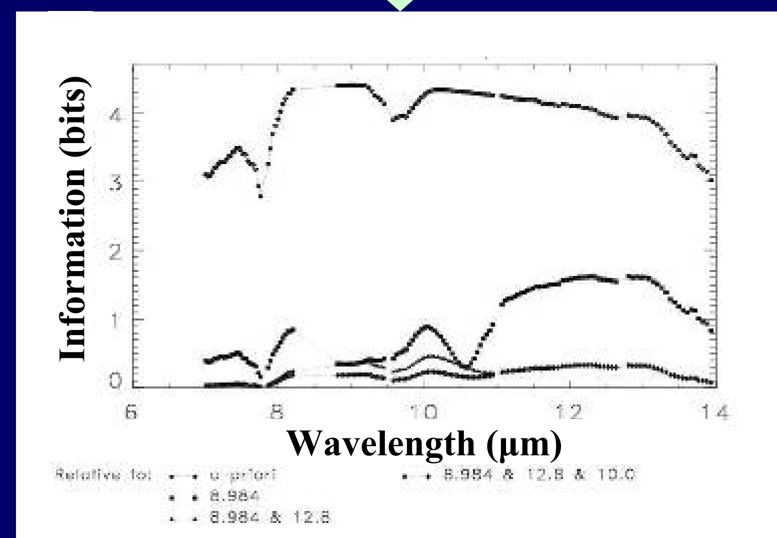


The Importance of Modeling Errors





Measurement Error Only (σ = 0.5 K)



Synthetic Retrievals

- Modeled radiances for all 141 channels are employed to test a number of possible wavelength combinations in an inversion framework that is analogous to the 1D-Var assimilation problem.

$$\frac{\partial \Phi}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}}\left[(\mathbf{x}-\mathbf{x}_a)^T \mathbf{S}_a^{-1}(\mathbf{x}-\mathbf{x}_a) + (\mathbf{y}-\mathbf{F}(\mathbf{x}))^T \mathbf{S}_y^{-1}(\mathbf{y}-\mathbf{F}(\mathbf{x}))\right] = 0$$

  - Retrieval vector $\mathbf{x}$ = (LWP, D_e);
  - $\mathbf{x}_a$ = (120 gm^{-2}, 75 μm), σ_a = (2000 gm^{-2}, 200 μm);
  - $\mathbf{x}_{truth}$ = (75 gm^{-2}, 50 μm);
  - $\mathbf{y}$ = various combinations of channels between 7 and 14 μm.

Synthetic 1D-Var Retrievals

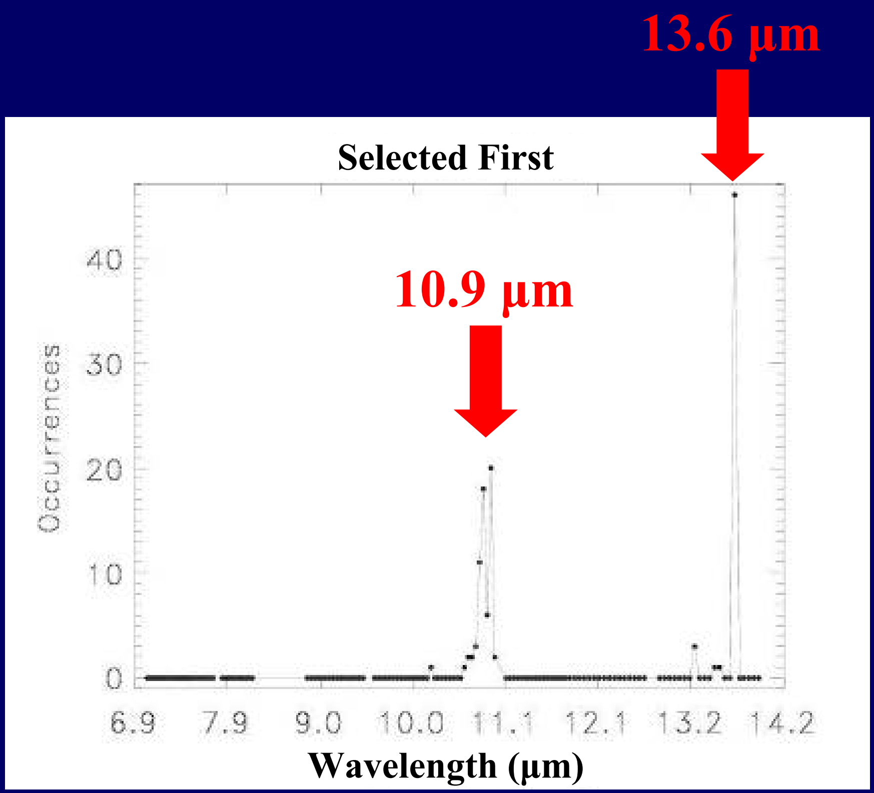
$\mathbf{x}_a$ = (120 gm^{-2}, 75 μm) : $\mathbf{x}_{truth}$ = (75 gm^{-2}, 50 μm)
σ_a = (2000 gm^{-2}, 200 μm)

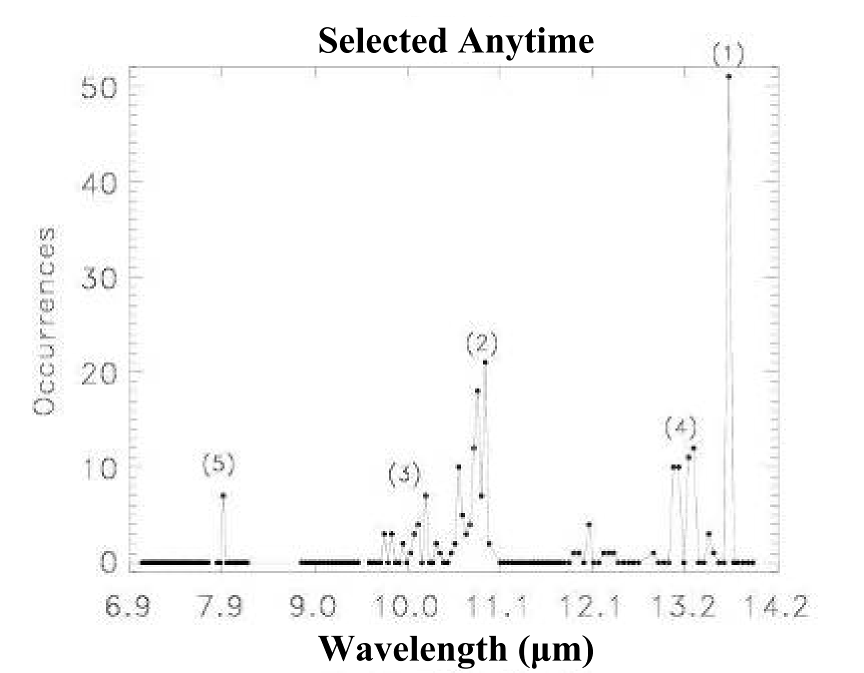
Information Used	IWP	IWP Bias	σ_{IWP}	D_e	D_e Bias	σ_{De}	Computation Time*
All 141 Channels	76.56	1.56	5.7	53.14	3.14	4.8	10.6 days
Ten Channels	76.97	1.97	20.2	53.46	3.46	17.1	18.6 hours
3 Optimal Channels	76.61	1.61	28.7	53.08	3.08	25.2	7 hours
Split-Window	77.73	2.73	42.6	54.01	4.01	37.01	4.6 hours
Single Channel	96.84	21.84	166.5	75.19	25.19	199.3	3.6 hours

*** Time to process a 100x100 array or 10,000 pixels on a 1.4 GHz Linux PC.**

Optimal Channel Spectra

Through analysis of a set of 230 cases representative of the global distribution of cloud properties one can assess the optimal channels for the general retrieval or assimilation problem.





Summary

- **Uncertainties:** The information provided by any measurement is tied to how well it can be simulated from the physical quantities that are predicted by the model.
- **Redundancy Reduction:** Given realistic error estimates, there is often a high degree of redundancy in high-spectral resolution measurements.
- **Lessons from Simpler Problems:** Since they are governed by similar physics, lower dimensional problems like the variational retrieval problem examined here can provide valuable insights for optimizing the more complex 4D-Var assimilation system.
- **Application to Variational Assimilation:** Variational assimilation systems already provide all of the necessary information to perform an objective information content analysis. Robust estimates of the background and measurement error covariances and the sensitivities of the observations to the control variables (i.e. the Jacobian) should allow the best measurements and their optimal spatial and temporal distribution to be determined.