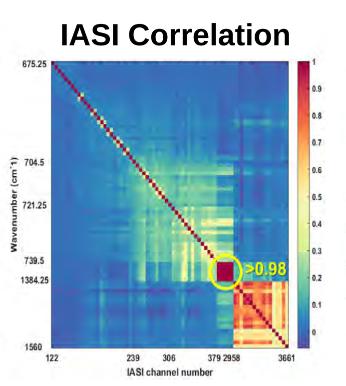
A New Channel Selection Method for Satellite Instruments with Correlated Observation Error

William F. Campbell
U.S. Naval Research Laboratory, Monterey, CA
bill.campbell@nrlmry.navy.mil

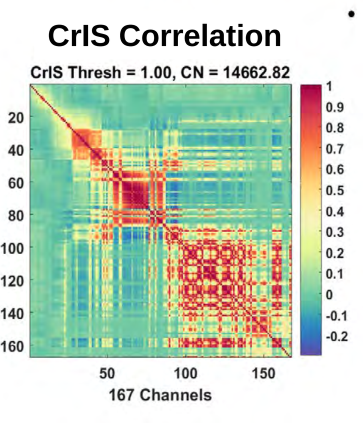
Motivation

Posterior Channel Selection

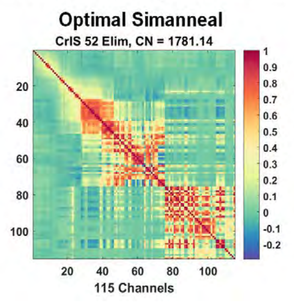
- Water vapor channels 2889, 2944, 2948, 2951, and 2958 have very high error correlation (>0.98)
- The eigenvectors corresponding to the 5 smallest eigenvalues project only on to these 5 channels
- It makes sense to use the Desroziers diagnostic to do *a posteriori* channel selection, to improve the condition number while discarding redundant information
- NAVGEM tests show the retaining only channel 2889 helps solver convergence with no discernable negative impact



Application to CrIS FSR



- **We may benefit from a similar procedure for CrIS channel selection with correlation, e.g. corr < 0.95**



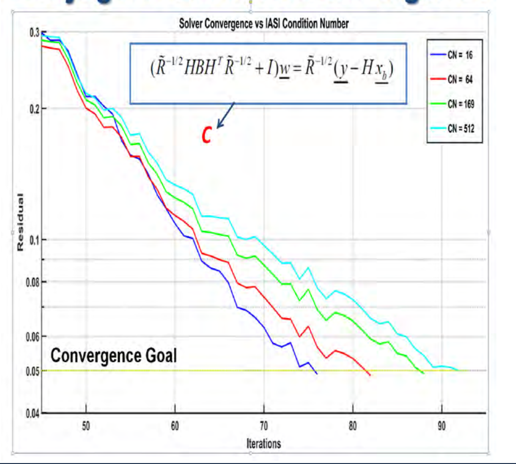
- The **Desroziers** error covariance estimation method can quantify interchannel correlated observation error for satellite radiances from instruments such as IASI, CrIS, and ATMS
- The resulting matrices can be quite **ill-conditioned**, depending on the number of channels (recall that the condition number for a symmetric, positive definite matrix is the ratio of the largest and smallest eigenvalues)
- Ill-conditioned observation error covariance matrices adversely affect the **convergence** of the flexible **conjugate gradient** descent at the heart of NAVGEM, our hybrid 4DVar data assimilation system
- Because of **operational time constraints** at the Fleet Numerical Meteorology and Oceanography Center (FNMOC), there is a **limit** on the **number of iterations** of the conjugate gradient descent

- A variety of **known reconditioning methods** can improve the condition number to the point where operational time constraints are met
- As a consequence of the **Cauchy Interleaving Theorem**, removing channels (i.e. rows/columns) from a symmetric, positive definite matrix can never increase the condition number, and almost always reduces it for the matrices we are concerned with
- **Channels with very high error correlation are by definition not providing much independent information to the analysis system**
- **Our hypothesis is that removing subsets of channels whose errors are highly correlated can improve the condition number of the resulting matrix without adversely affecting the analysis and forecasts**

Condition Number and Solver Convergence

- The **condition number** of a matrix X is defined by $\sigma_{max}(X)/\sigma_{min}(X)$, which is the ratio of the maximum singular value of X to the minimum one. (Singular value == eigenvalue for symmetric X)
- Adding correlated error increases the condition number, slowing down convergence of the solver
- We can control how long the solver takes by constructing an approximate matrix with ***any condition number we choose***

Conjugate Gradient Convergence



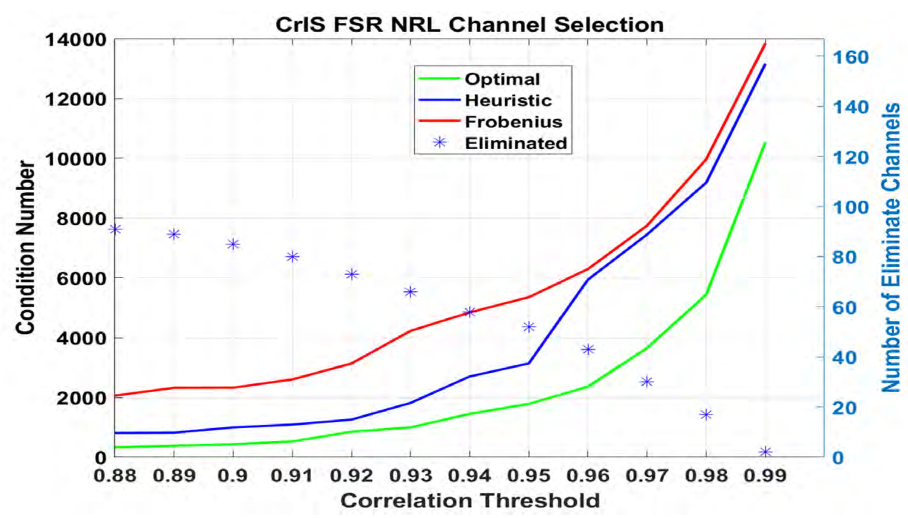
Reconditioning CrIS FSR by Posterior Channel Selection

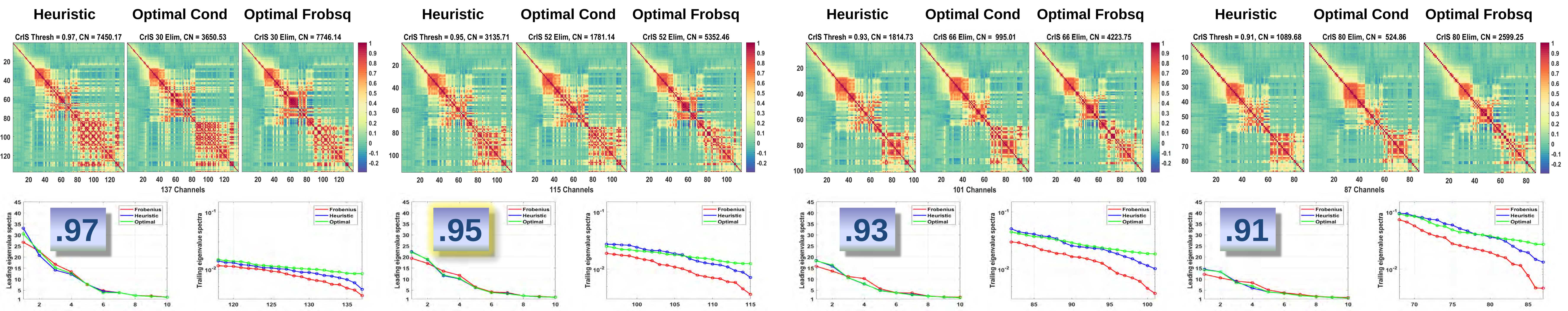
Description of Channel Selection Heuristic

- Compute the sum of the squares (**score**) of the entries in each row of the Desroziers estimate of the CrIS observation error correlation matrix
- Find all rows with at least one value exceeding a **threshold** such as 0.95 (**candidates**)
- Choose the **candidate** with the highest **score**, and **delete** the corresponding row and column
- Update the **scores** of the new, smaller matrix
- Continue finding candidate rows, choosing the highest scoring candidate, and deleting the corresponding row and column **until there are no values exceeding the threshold**
- The **condition number** of the new matrix will be **reduced** considerably, depending on the threshold chosen

Compare with Optimal and Frobenius

- Purely from the viewpoint of condition number, we can find the optimal channel selection, but this cannot be done by brute force for more than a handful of channels
- The condition numbers for **optimal channel selection** were estimated with **simulated annealing**, a global optimization procedure
- A second set of simulated annealing runs was performed with the **Frobenius norm** as the objective function, with no conception of correlation thresholding
- The channel selection **heuristic with correlation thresholding** yielded significantly **better conditioned matrices than pure Frobenius** optimization, lending support to the heuristic





Further Reconditioning for Convergence

Reconditioning Methods

1. Multiplicative preconditioning by diagonal scaling matrices
2. Increase the diagonal values (additively) of the matrix (e.g. Weston et al. (2014))
3. Find the optimal linear combination of the sample covariance matrix and the scaled identity matrix (optimal Steinian linear shrinkage – similar to 2.)
4. Find a positive definite approximation to the matrix by altering the eigenvalue spectrum (constrained minimization of the Ky-Fan p-k norm)
5. Channel selection to eliminate channels with highly correlated error

Steinian linear shrinkage

- Consider the following affine transformation of the eigenvalue spectrum of an NxN sample covariance matrix X with eigenvalues λ_i:

$$\gamma_i=(1-\delta)\lambda_i+\delta c,\ 0\le\delta\le 1$$

where c is equal to $tr(X_{true})/N$. For correlation matrices, c≡1.

- For any desired condition number $(\gamma_{max}/\gamma_{min})$, we can choose a δ that preserves the shape of the eigenspectrum
- This method has more theoretical justification (e.g. Ledoit and Wolf, 2004) than the additive method (as far as I have been able to find), and yields a similar spectrum and reconstructed correlation matrix

Ky-Fan p-k Norm

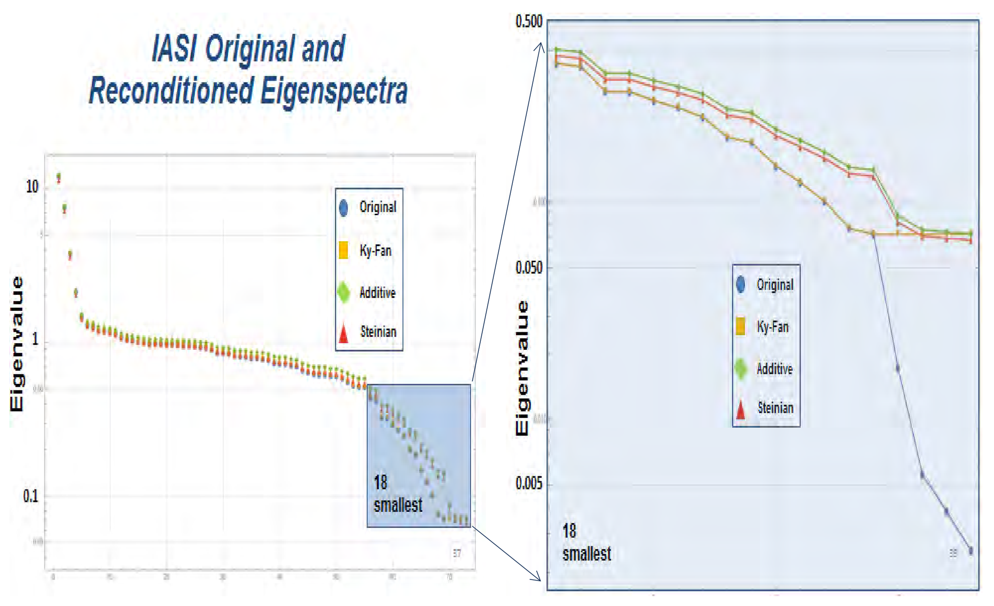
The Ky Fan p-k norm of $X\in\mathbb{C}^{m\times n}\equiv\|X\|_{p,k}=\left(\sum_{i=1}^{k}\sigma_i(X)^p\right)^{1/p}$ where

$\sigma_i(X)$ denotes the i^{th} largest singular value of X
When p=2 and k=N, it is called the Frobenius norm; when p=1 and k=N, it is called the trace norm.

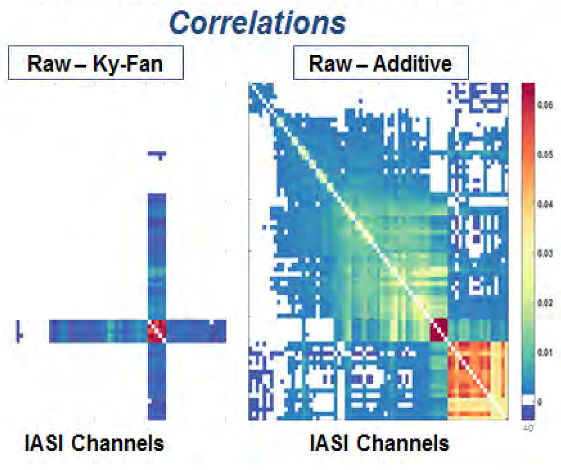
We want to find a positive definite approximation to the matrix $\hat{X}\ni\|\hat{X}-X\|_{p,k}$ is minimized.

In the trace norm, this is potentially as simple as setting all of the smallest singular values equal to the value that gives the desired condition number, and then reconstructing the matrix with the singular vectors to obtain the approximate matrix.

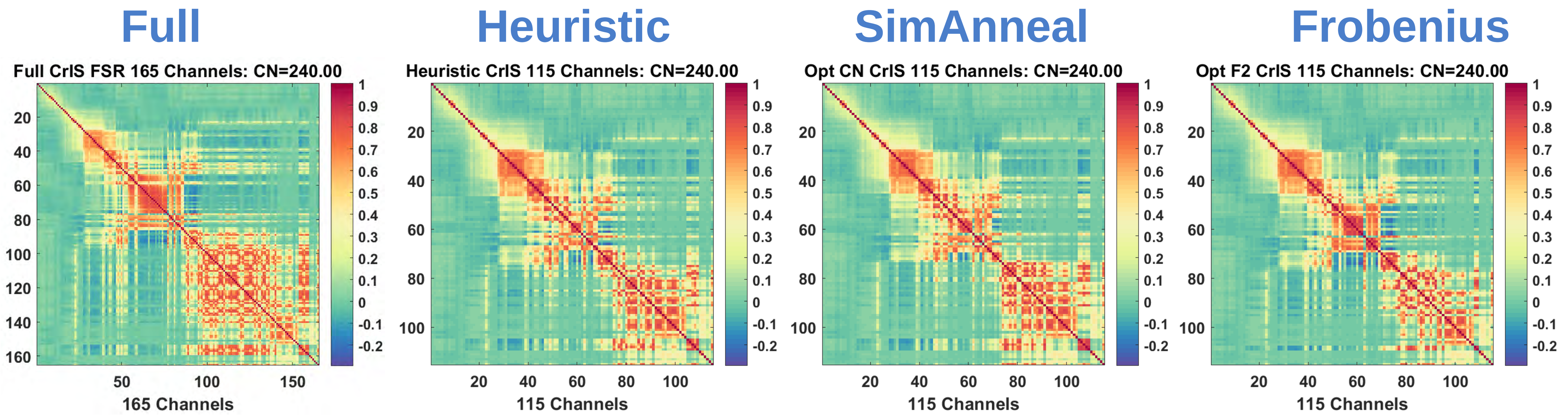
IASI Original and Reconditioned Eigenspectra



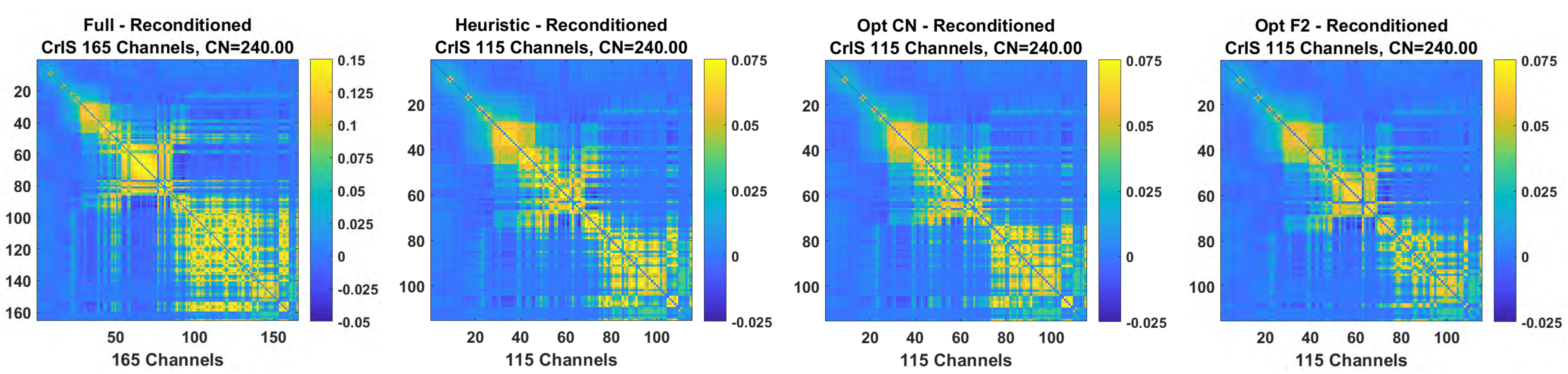
IASI Raw Minus Reconditioned Correlations



Posterior-Selected Steinian Reconditioned CrIS Correlations



Differences



Summary, Conclusions, and Future Work

- **The new channel selection procedure removes subsets of correlated channels until no correlation exceeds a fixed threshold**
- Alternatively, we can remove the same number of channels by **optimizing for minimum condition number**, but we may remove valuable information at the same time, which the heuristic tries to avoid
- Optimal channel selection cannot be done by brute force for more than a handful of channels
- The condition numbers for **optimal channel selection** were estimated with **simulated annealing**, a global optimization procedure

- A second alternative is to **optimize the Frobenius norm**, which has the same potential pitfall as optimizing for condition number
- For maximum correlation thresholds of **0.93** or lower, the **heuristic obtains most of the condition number benefit available from pure channel selection optimized for condition number**
- The resulting matrices have similar patterns of correlation in the retained channels
- For maximum correlation thresholds of **0.95** or lower, the **heuristic has a much better condition number than optimizing the Frobenius norm**

- This makes intuitive sense, as from the results above, most of the improvements in condition number came from increasing the smallest eigenvalue rather than decreasing the largest, which is what one expects Frobenius norm optimization to do
- Comparison of the characteristics of the eliminated channels should be done to determine if there is any underlying pattern that can be exploited
- For moderate channel reduction, the resulting condition number for all three methods is still too high for operational constraints
- Therefore, Ky-Fan optimization or Steinian shrinkage should be applied to the reduced size matrices to further reduce the condition number

- **Cycling data assimilations for channel-subselected CrIS, comparing the control, heuristic subselection, and optimal condition number subselection must still be performed**
- **Each test matrix uses 115 channels, and was reconditioned with Steinian linear shrinkage to a fixed condition number or 240.0 for fair comparison**
- See our paper Campbell et al. "*Accounting for Correlated Observation Error in a Dual Formulation 4D-Variational Data Assimilation System*", *Mon. Wea. Rev.*, **145**, 1019-1032

Poster presented on May 30, 2018 at the 16th JCSDA Technical Review Meeting and Science Workshop on Satellite Data Assimilation in Boulder, CO.