Australian Government
Bureau of Meteorology

Error correlations and hyperspectral sounders

Fiona Smith, Bureau of Meteorology

fiona.smith@bom.gov.au

This talk contains results contributed by my collaborators:

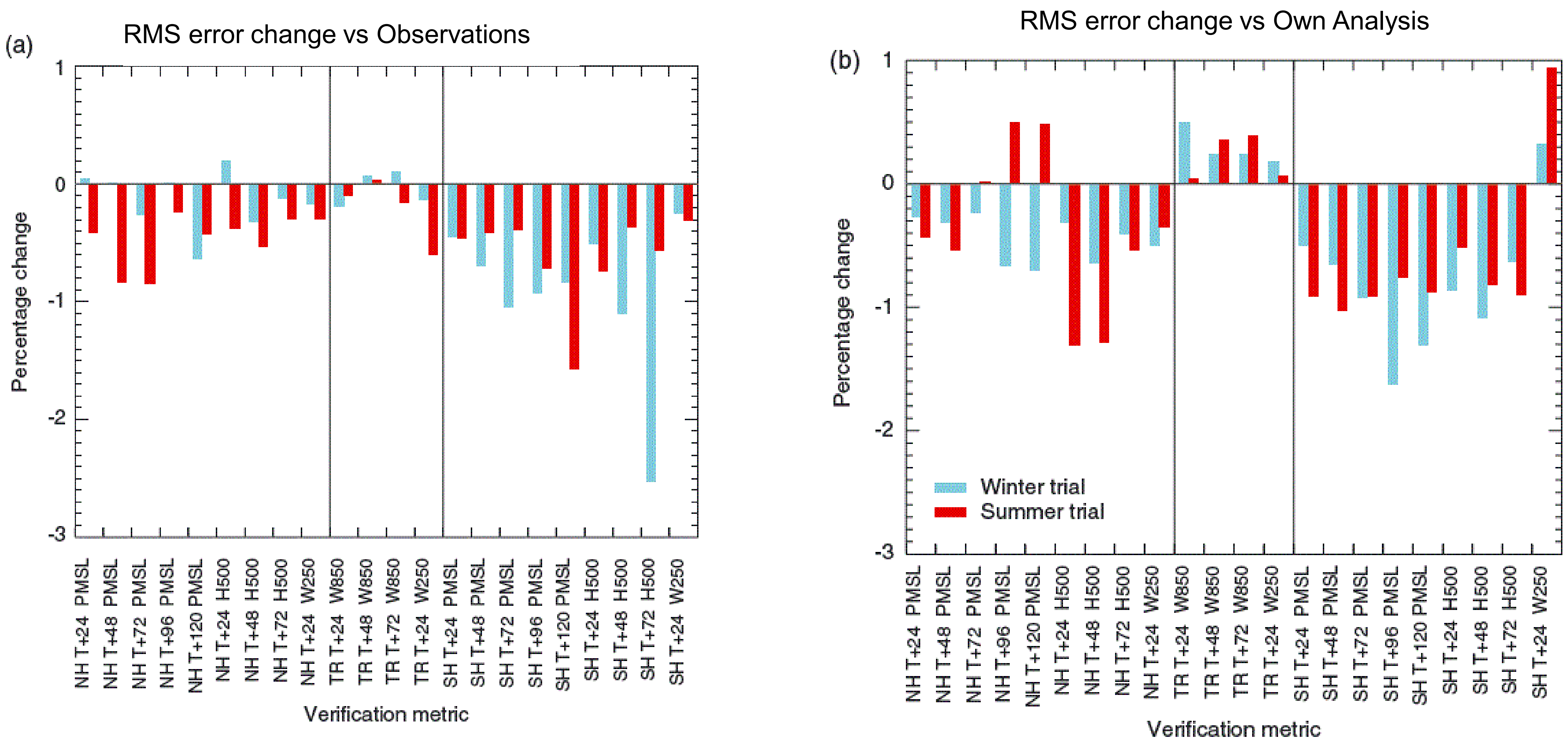
- JMA – Toshiyuki Ishibashi
- NRL – Bill Campbell
- Met Office –Fabien Carminati
- ECCC – Sylvain Heilliette
- Meteo-France – Nadia Fourrié
- NCEP – Kristin Bathmann

- And some material gathered from papers and previous discussions
  - ECMWF – Alan Geer, Niels Bormann, Reima Eresmaa (->FMI)
  - Meteo-France – Vincent Guidard and Olivier Coopmann
  - Met Office – Pete Weston (->ECMWF), Jemima Tabeart (->Edinburgh Uni)

Overview

- Methods used in estimation of error correlations
- Justification for shrinkage and variance inflation
- Comparison of matrices between centres

Implementation of correlated error universally reported as delivering positive impact. [Results from Met Office shown]



https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/qj.2306

Australian Government
Bureau of Meteorology

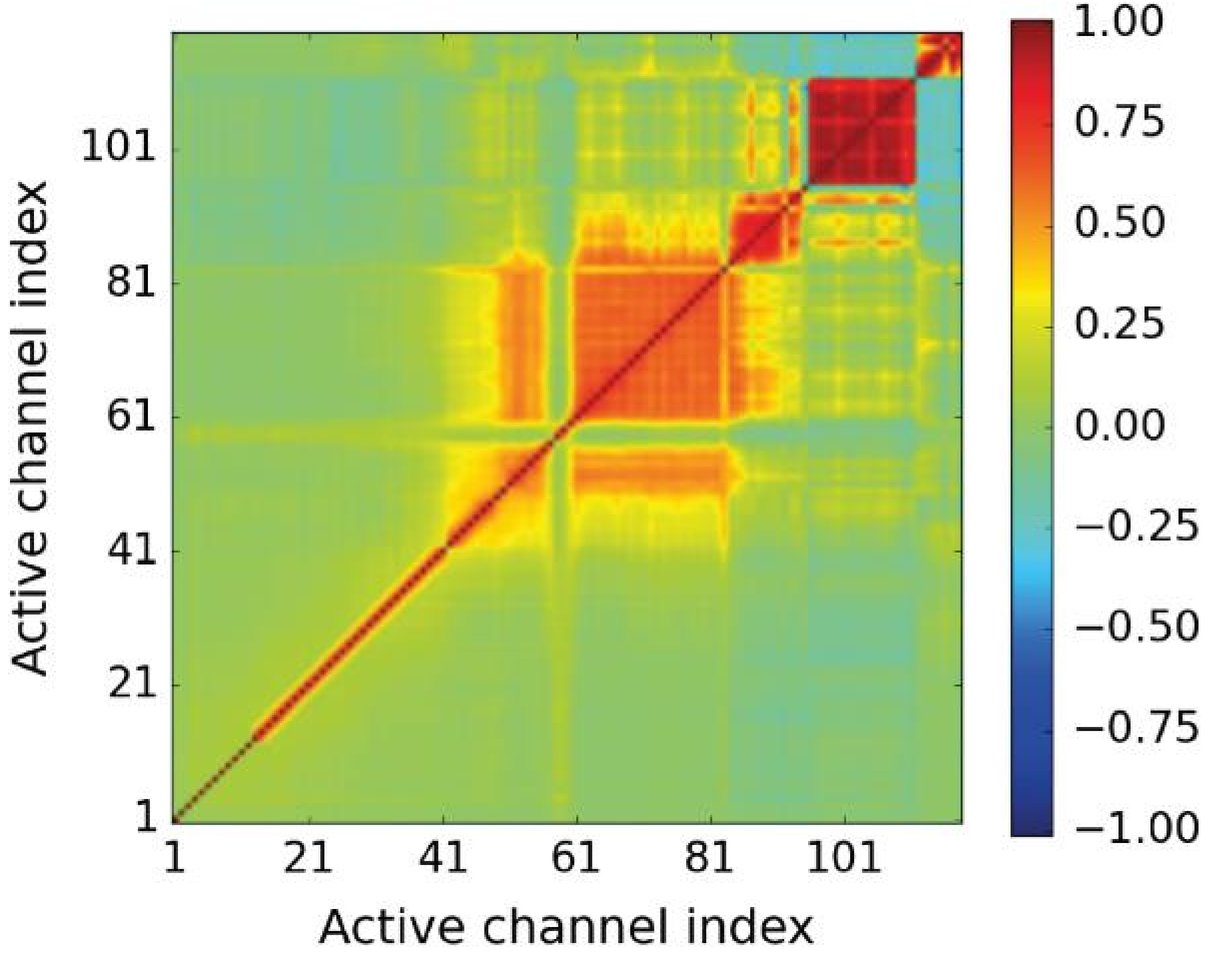
Covariance estimation

Motivation

- What do the error covariances and correlations look like?
    - Consistent between centres?
    - Consistent between instruments?

- What does the consistency (or lack of) tell us about the sources of correlation

- Can we be sure the software we use to estimate covariances is doing a good job?

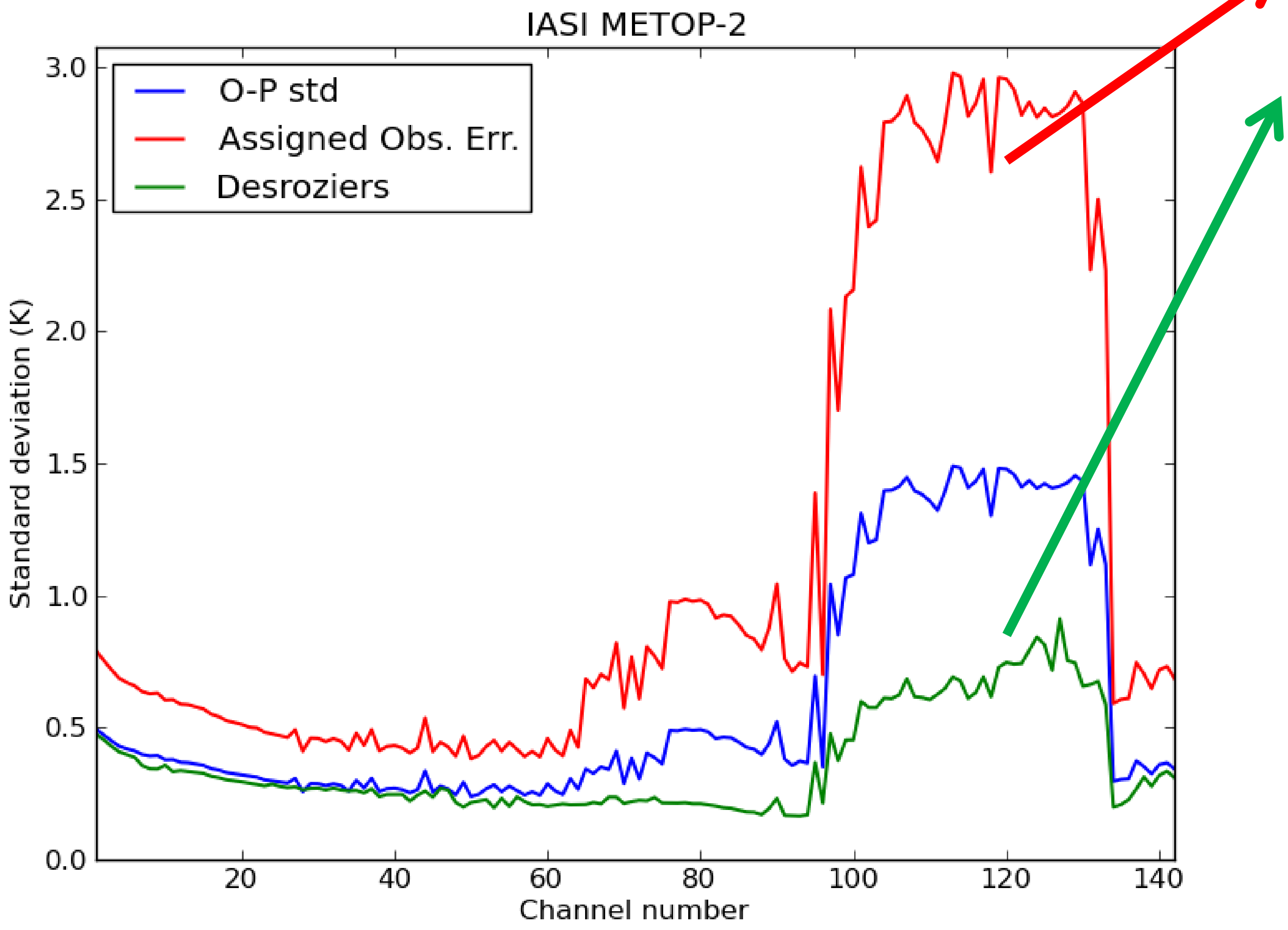


Methods differ in the details but generally similar

- Start with an initial estimate of errors
    - Diagonal
    - Hollingsworth-Lönnberg
    - Desroziers from 1D-Var
    - Possibly multiply by a scaling factor
- Output diagnostics to allow estimation of covariances using Desroziers* method
- Symmetrise Desroziers matrix
    - Covariance or correlation
- Inflate error variance
    - Spectrally variant or invariant multiplicative factor
    - Additive factor (see next step)
- Manipulate covariance matrix to improve conditioning
    - Adjust smallest eigenvalues to reduce spread ("shrink" matrix)

* Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. R.Meteorol. Soc. 131: 3385–3396.

Environment Canada - diagnosed errors for IASI



Chosen errors:
Significant inflation

Desroziers estimate
well below std(O-B)

- Initial estimated errors from Desroziers are usually much smaller than previously used uncorrelated error variances
- Often lie somewhere between the observed SD(O-B) and instrument noise (right)
- Can occasionally be "too small" – below NEDT
- Iterating Desroziers technique can have varying success

- Plot from Environment Canada

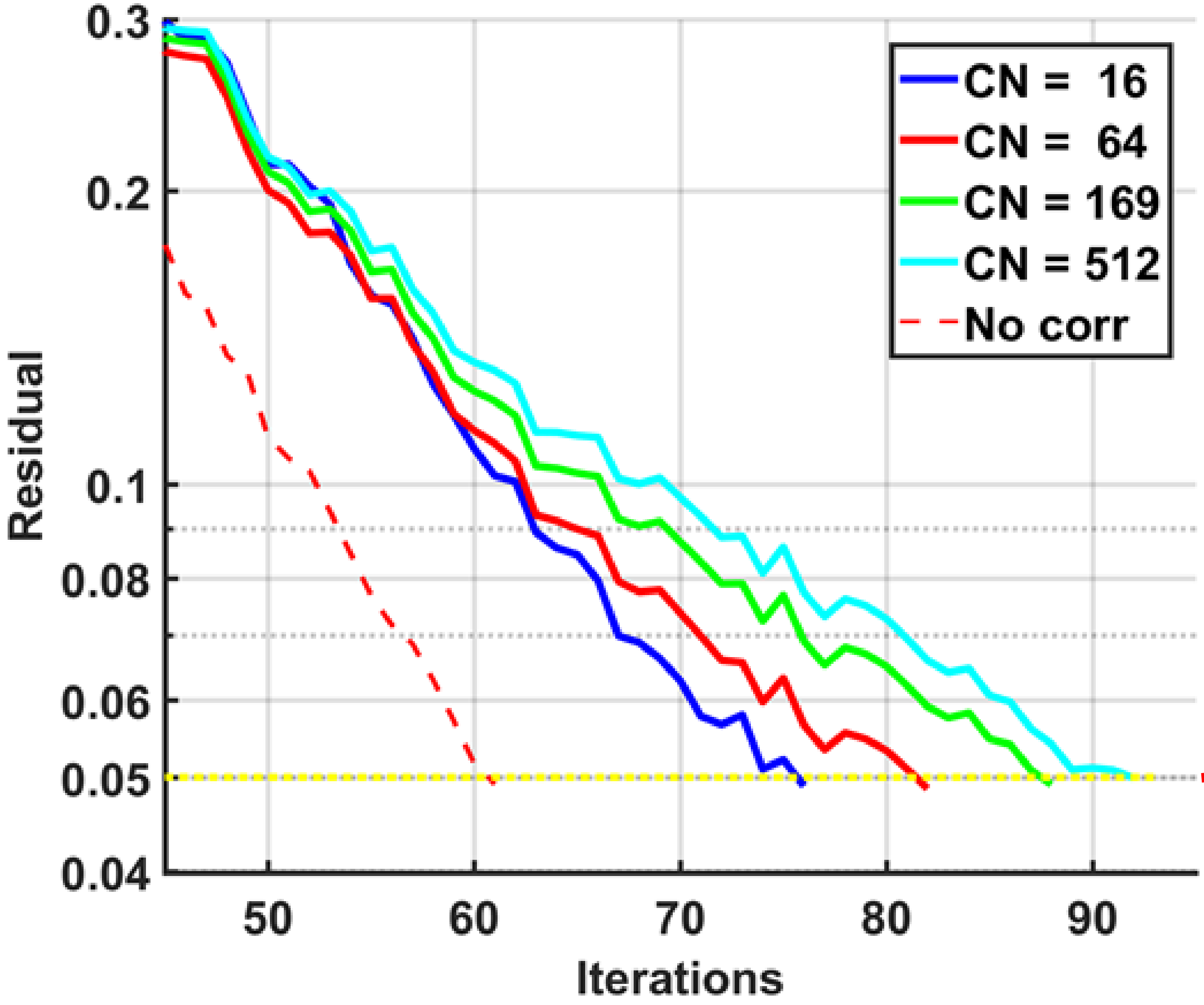
Australian Government
Bureau of Meteorology

Justification for shrinkage and error inflation

What did contributors say?

- Varied statements regarding justification for shrinkage and error inflation
- Most view the process pragmatically
  - A process that must be done to make 4D-Var work effectively
    - Reduce iterations
    - Improve forecast benefit
- Some feel the justification is physical
  - Accounting for errors that are not diagnosed properly by the Desroziers method
    - E.g. quality control problems
    - Mapping of small scale errors onto gravity waves
- Some mathematical justification
  - Large body of work on covariance estimation especially in biostatistics and finance

Impact of conditioning on convergence – "Make the DA work properly"



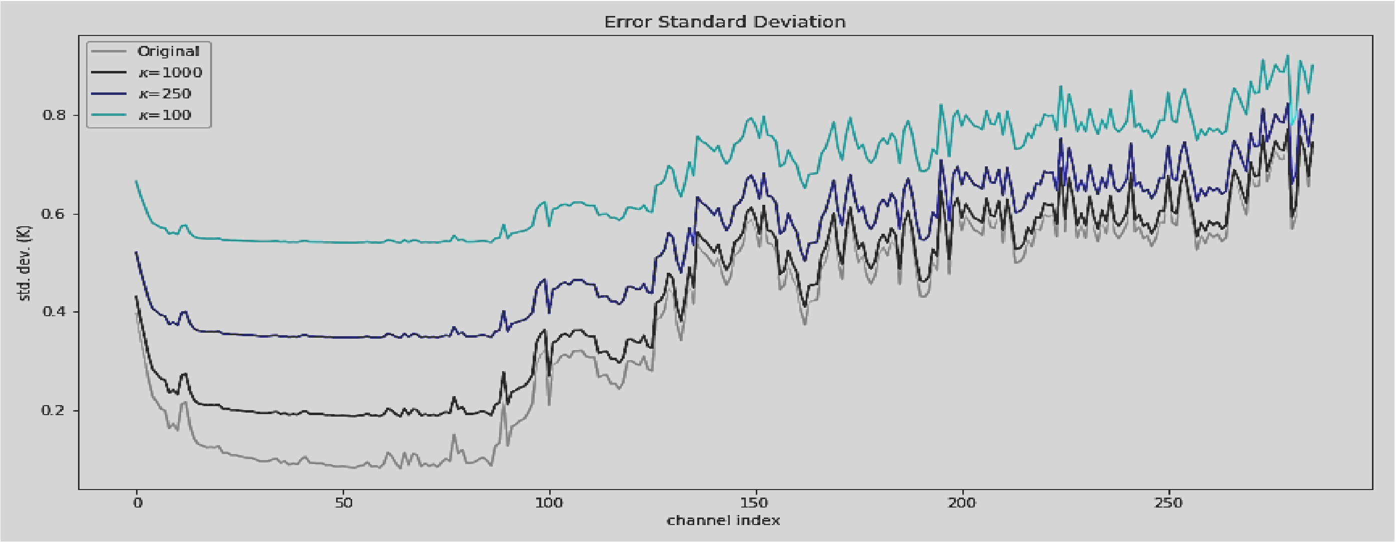
NRL report:
The computational benefits of additive reconditioning outweigh the slightly better forecast performance of increasing small eigenvalues

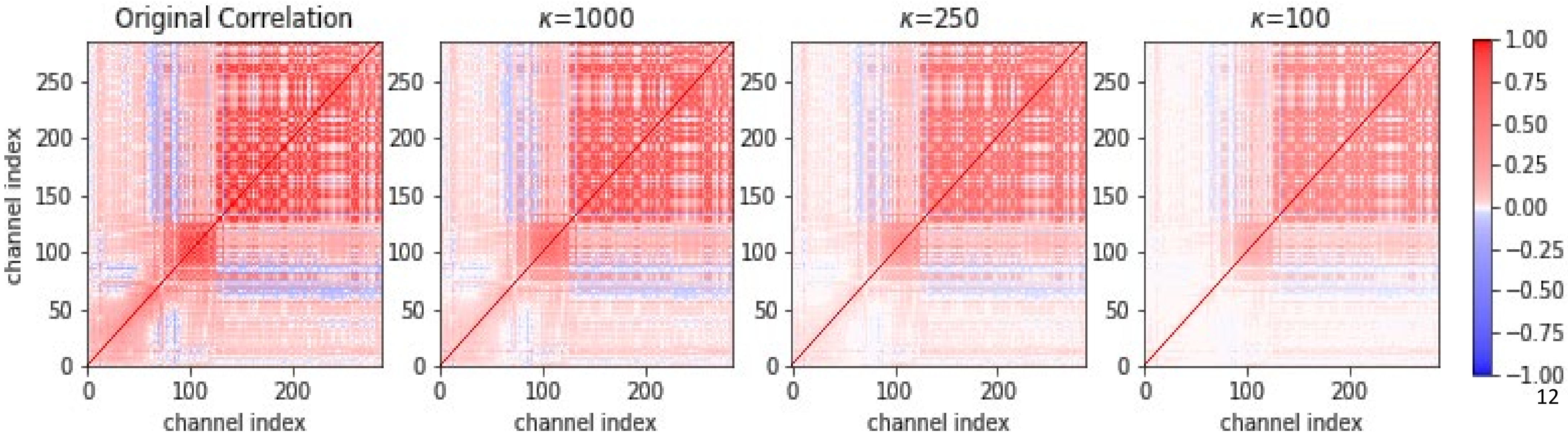
Convergence at Residual = 0.05

Campbell et al., 2017
https://doi.org/10.1175/MWR-D-16-0240.1

Forecast benefit of conditioning – "Improve impact of observations" (1)

CrIS FSR - results from Fabien Carminati

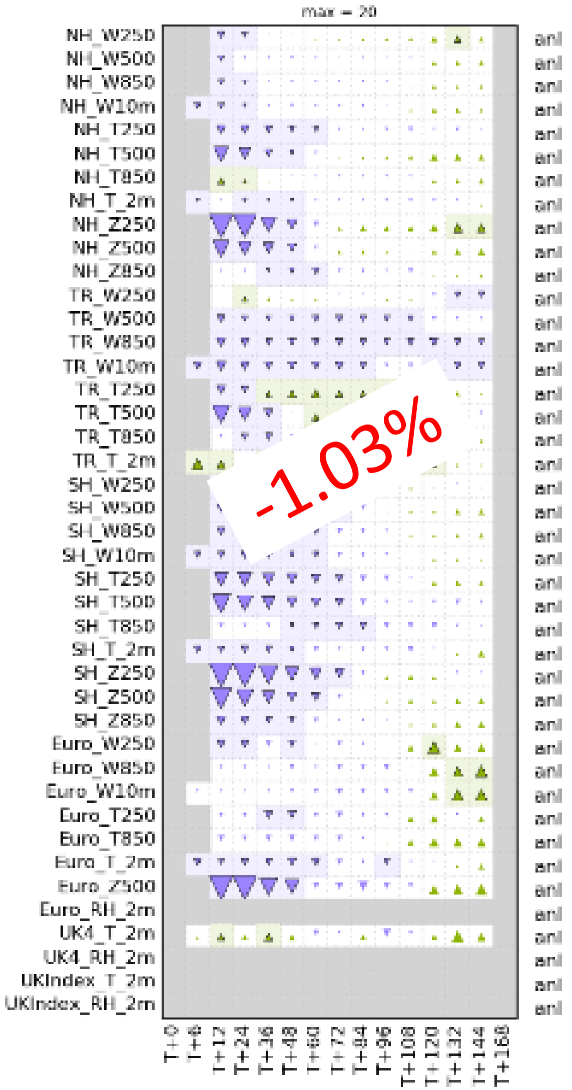




Forecast benefit of conditioning – "Improve impact of observations" (2)

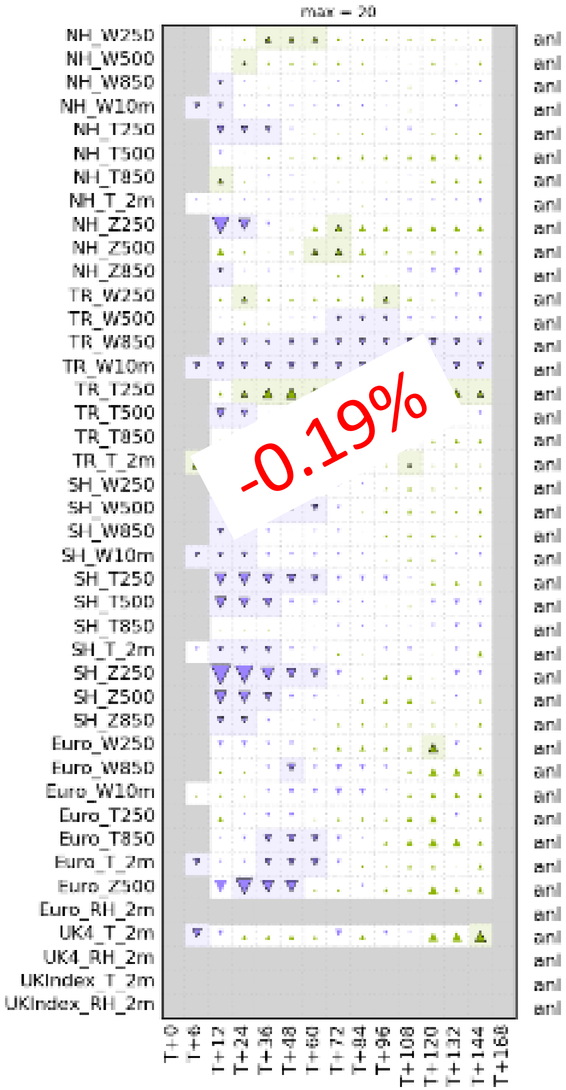
Condition Number 1000

inflation = diagonal + ~0.028



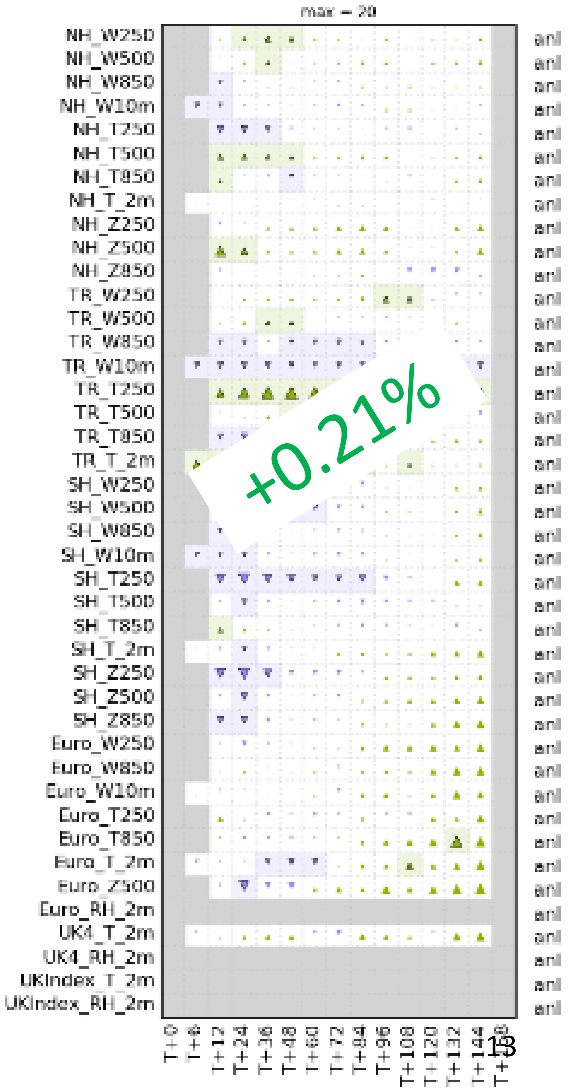
Condition number 250

inflation = diagonal + ~0.113



Condition number 100

inflation = diagonal + ~0.28



Mathematical Justification for Shrinkage

- The estimation of error covariances is inherently "overdispersed"
  - The largest eigenvalues are over-estimated, and the smallest ones are underestimated
  - Covariance matrices perform better if they are "shrunk" – i.e. all eigenvalues are brought towards the mean

- Effron and Morris (1977): https://statweb.stanford.edu/~ckirby/brad/other/Article1977.pdf
- Daniels and Kass (2001): https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0006-341X.2001.01173.x

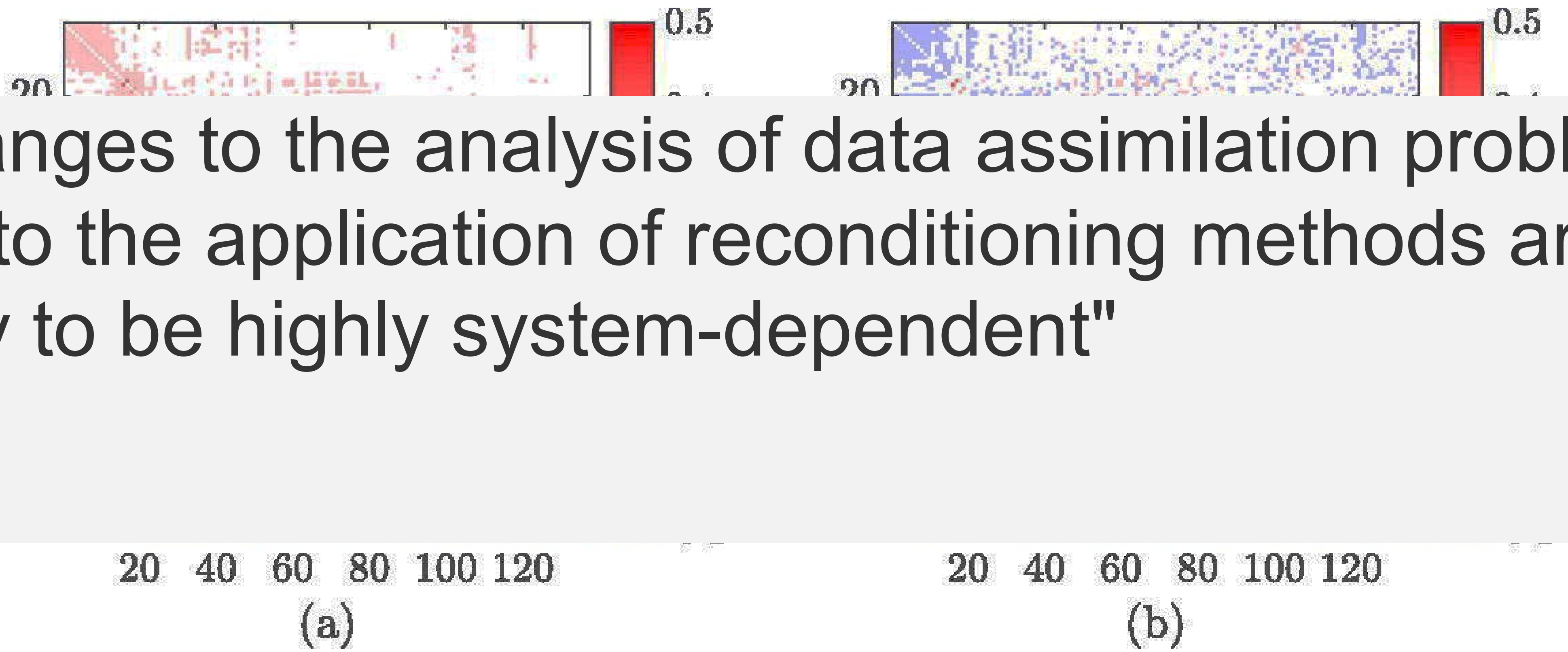
Which method of improving conditioning is more justifiable?

- Some like the idea that increasing smallest eigenvalues is essentially Ky Fan p-k norm covariance adjustment
  - Tanaka and Nakata (2013) https://link.springer.com/article/10.1007/s11590-013-0632-7
  - "Positive definite matrix approximation with a condition number constraint is an optimization problem to find the nearest positive definite matrix whose condition number is smaller than a given constant."
- Adding a constant to the eigenvalues is effectively Steinian shrinkage
  - Ledoit and Wolf (2004) https://www.sciencedirect.com/science/article/pii/S0047259X03000964?via%3Dihub
  - "This paper introduces an estimator that is both well-conditioned and more accurate than the sample covariance matrix asymptotically. This estimator is distribution-free and has a simple explicit formula that is easy to compute and interpret. It is the asymptotically optimal convex linear combination of the sample covariance matrix with the identity matrix."

Which method is better?

Tabeart et al., 2020: https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.3741

- Both methods strictly increase standard deviations, inflation results in a bigger increase than changing only the smallest eigenvalues
  - Inflating all eigenvalues (ridge regression) strictly decreases the absolute value of off-diagonal correlations
  - Increasing only the smallest eigenvalues can increase the absolute value of off-diagonal correlations
  - For the IASI experiment increasing only the smallest eigenvalues leads to smaller changes to correlations.



"Changes to the analysis of data assimilation problems due to the application of reconditioning methods are likely to be highly system-dependent"

Plots show original minus reconditioned

Physical Justification for error inflation (ECMWF)

- Eresmaa et al. 2017 https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.3171
  - It is not fully understood why a scaling factor is needed, nor why it should be higher for CrIS than for IASI. It is our guess that the scaling compensates for sub-optimalities associated with various simplifications needed for practical reasons. These might include ignoring horizontal and temporal error correlation altogether, lack of situation dependency, mis-specification in background-error covariance, and correlation between observation and background errors. Furthermore, our interpretation is that such sub-optimalities are amplified in the case of CrIS, because the uncorrelated observation-error contribution (i.e. instrument noise) is relatively small in the overall error budget.

- Alan Geer – in context of all-sky assimilation:
  - Trailing eigenvectors amplify small inter-channel differences
    - If resulting from biases, these will be incorrectly amplified and generate increments that oscillate in the vertical
    - Even without bias, can amplify signals that map onto vertical temperature oscillations (gravity waves) that DA cannot properly handle

Operational obs error manipulations

Centre	Shrinkage Method	Inflation over Desroziers	Condition number
Met Office + UM Partners	Add constant to all eigenvalues	Effectively: IASI T ~1.5, W.V. ~1.1	IASI 67
NRL	Add constant to all eigenvalues	IASI: T 1.65, WV 1.9	IASI 169
ECMWF	Increase small eigenvalues	IASI: 1.75 CrIS: 2.75	IASI 131 CrIS 4075
Meteo-France		IASI: 2.0	
NCEP	Increase small eigenvalues to condition number IASI: 200 CrIS: 125	T 1.6, WV 1.3, Window 1.8*	IASI 93 CrIS 53
DWD	Increase small eigenvalues	IASI: 1.75	
JMA		1.7**	
ECCC	Ensure positive definite	1.6	

- NCEP find that stricter cloud detection is necessary to get good results with correlated error covariances
- ** JMA justify their inflation with a corresponding deflation of background error by the equivalent factor (1/1.7)

Summary from ECMWF Workshop on Random and Systematic Errors, November 2020

- Everyone uses Desroziers
- Everyone does some manipulation to the output
- Justifications for this manipulation vary

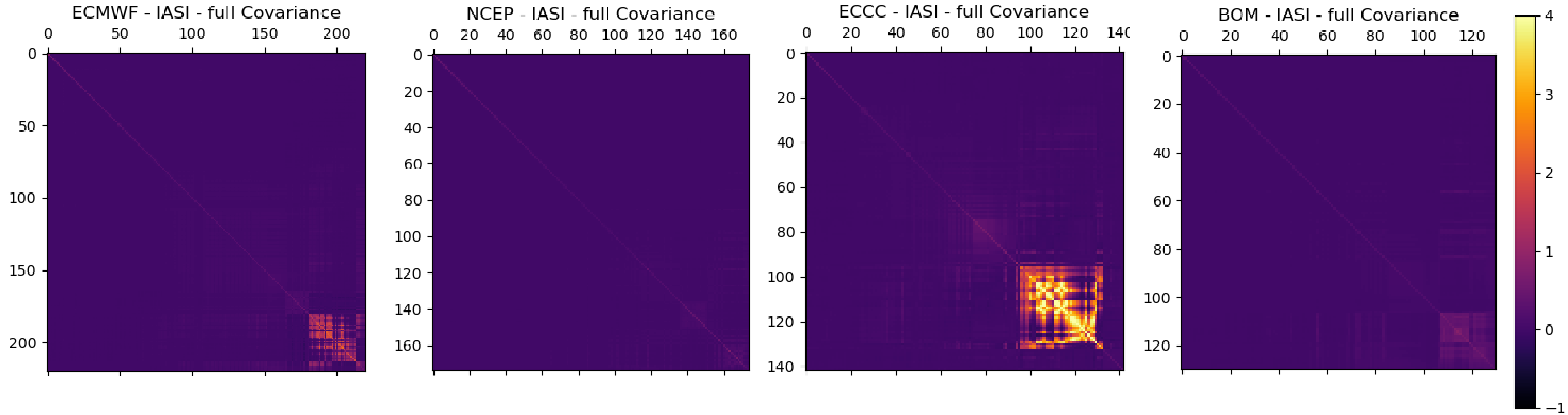
- Results are surprisingly consistent between centres
- All instruments show strong correlations for water vapour channels
- Different sounders have quite different diagnosed correlations
  - Points to different sources of error dominating for each instrument

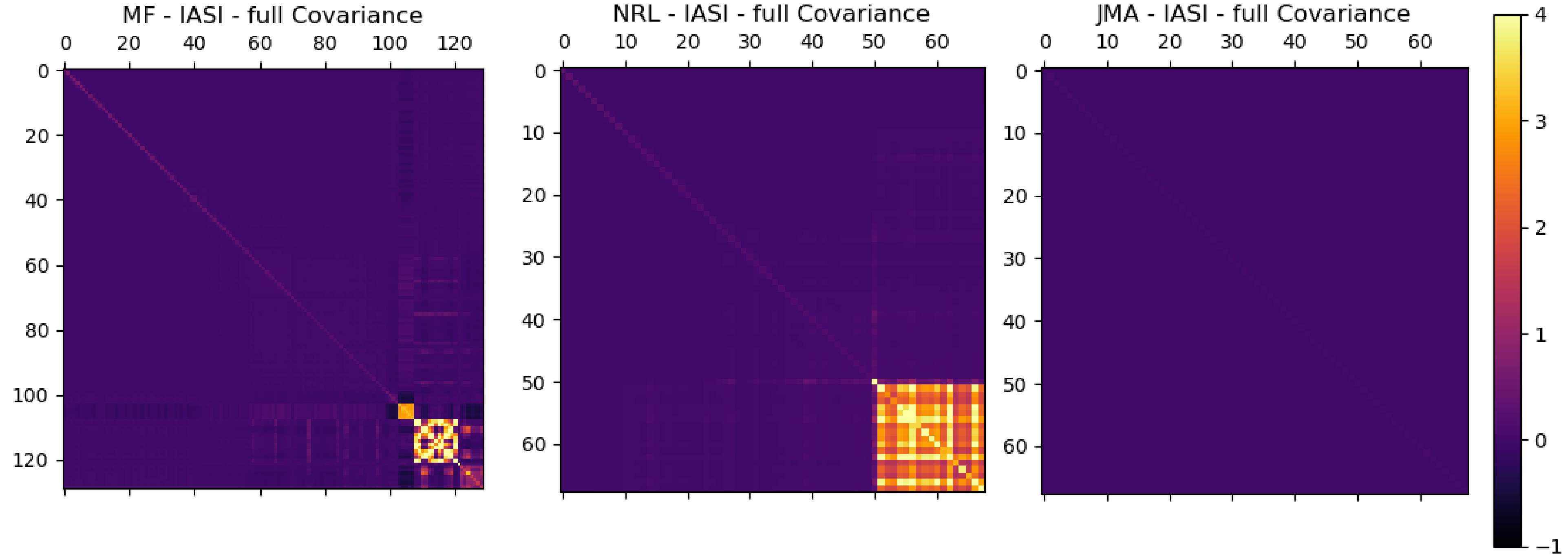
Australian Government
Bureau of Meteorology

New Comparison of Matrices with consistent plotting

Australian Government
Bureau of Meteorology

IASI

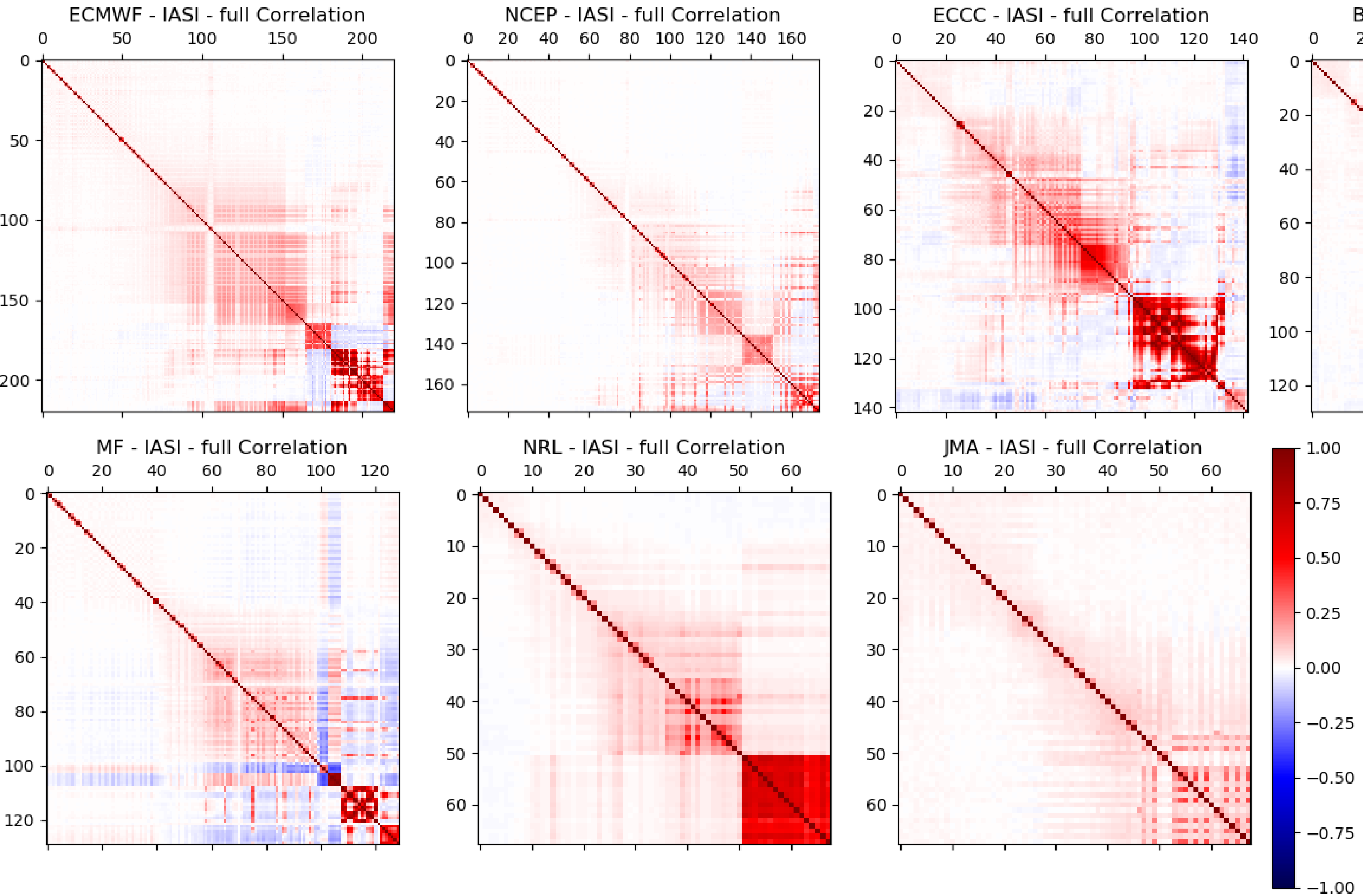




IASI covariances – full covariance matrix for assimilated channels

Differences in water vapour channels apparent.

Temperature channels little correlation for all centres



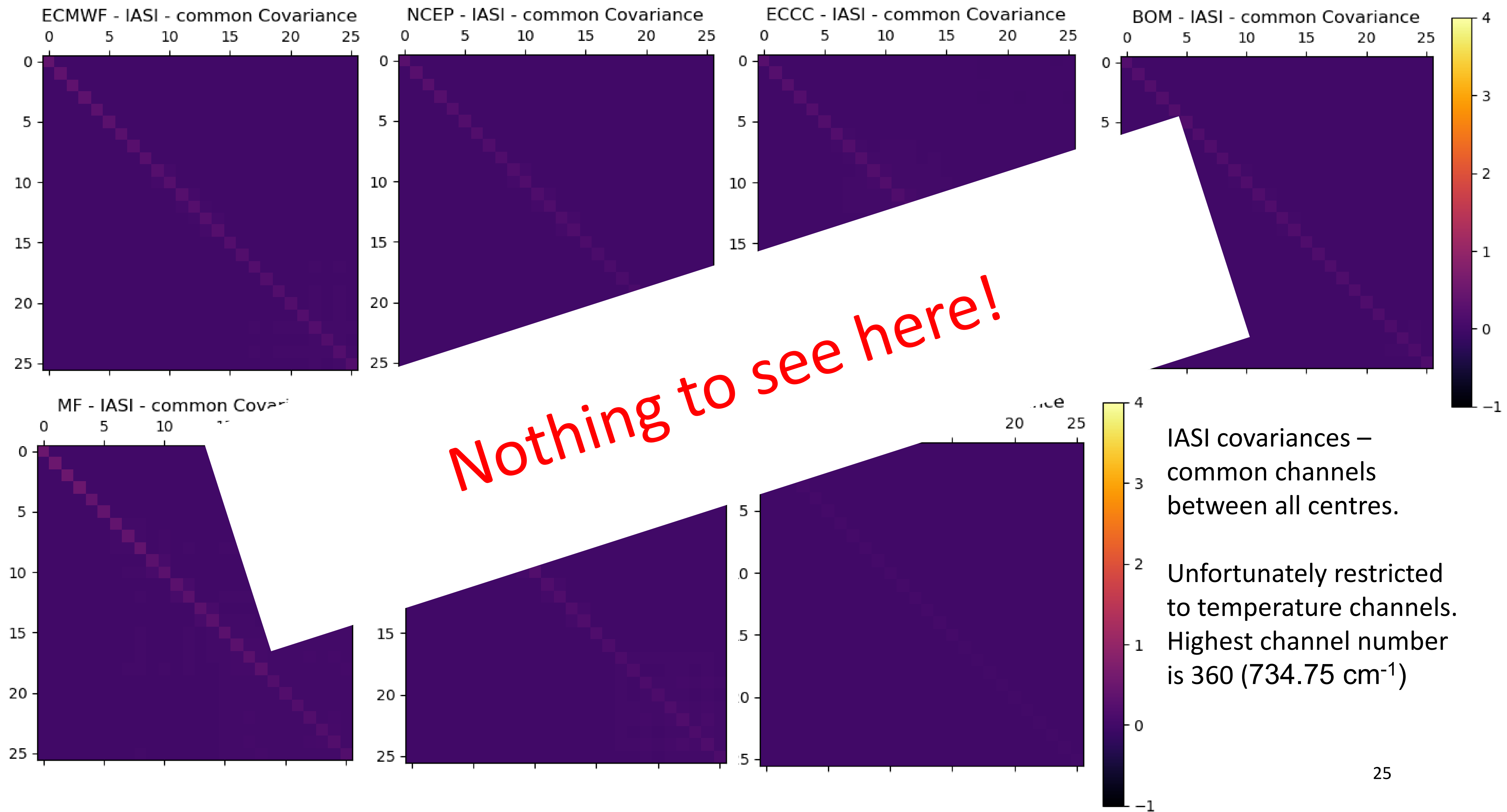
IASI correlations – full covariance matrix for assimilated channels

Differences between centres are clearer though plenty of similarities apparent also.

Need to compare common channels!

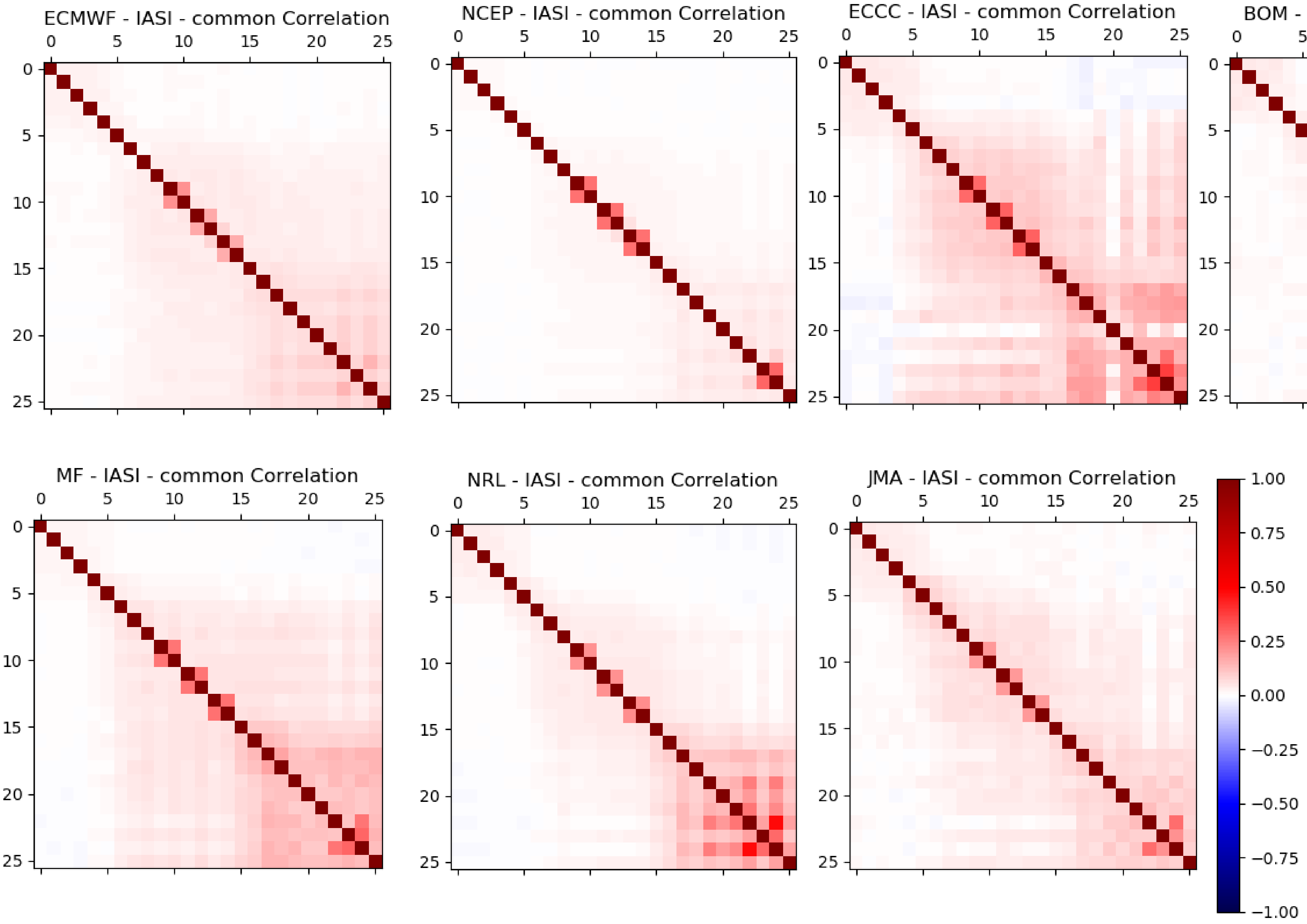
- Some patterns are clear:
  - higher correlations for surface and water vapour channels
  - Higher error variances for water vapour channels

- But it's still hard to see what's going on....
  - So let's look at the channels common to all centres



IASI covariances – common channels between all centres.

Unfortunately restricted to temperature channels. Highest channel number is 360 (734.75 cm^{-1})



IASI correlations – common channels between all centres.

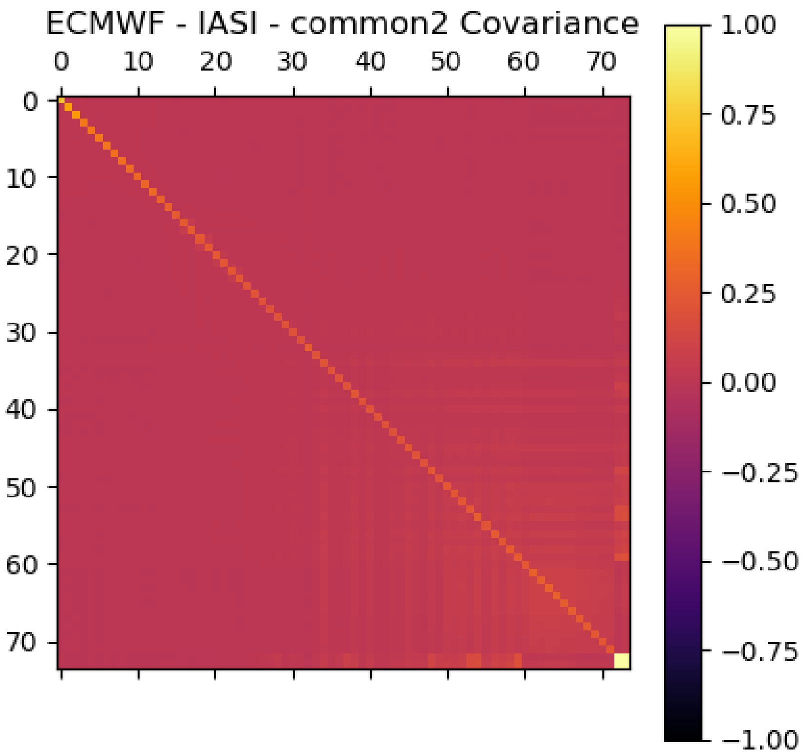
Dominated by instrument noise, with a few interesting differences in low level correlations.

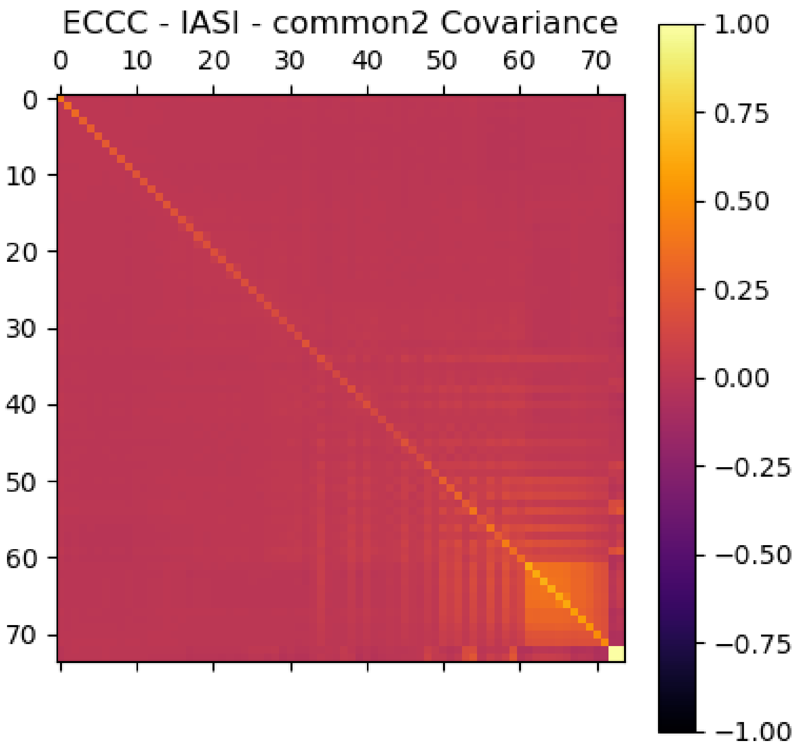
Would be nice to look more at Band 2....?

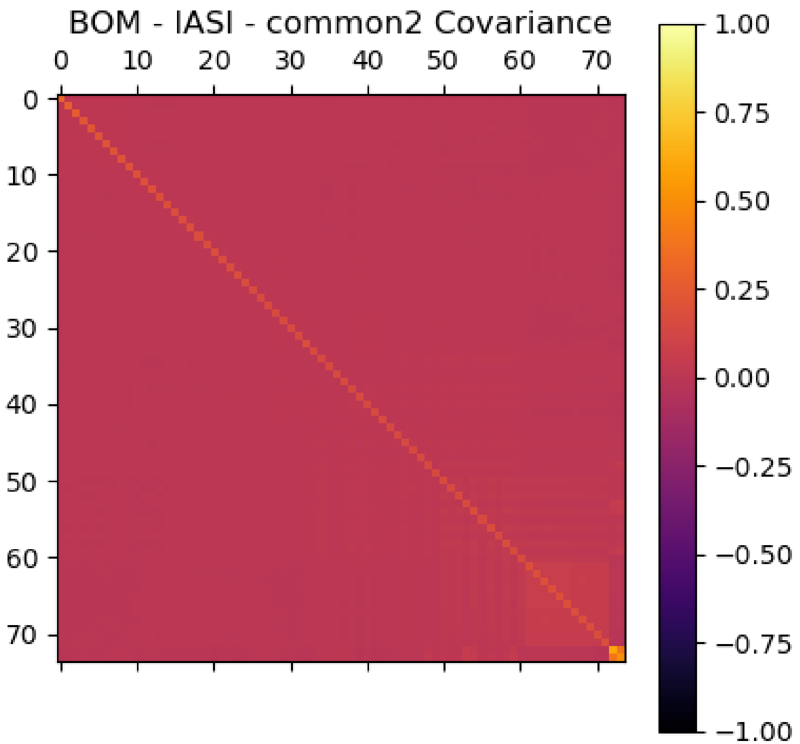
- Subset of centres with more channels in common

IASI covariances – common channels between three centres (note change in colour scale).

Differences in variance and correlation for surface and water vapour channels is evident

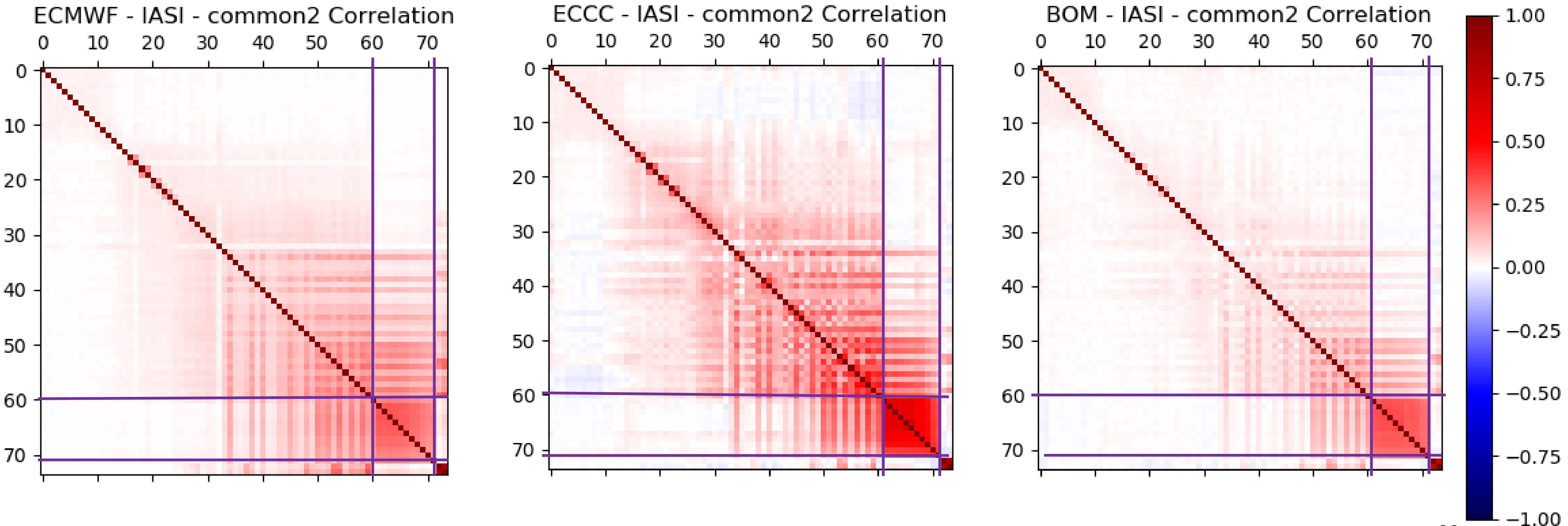






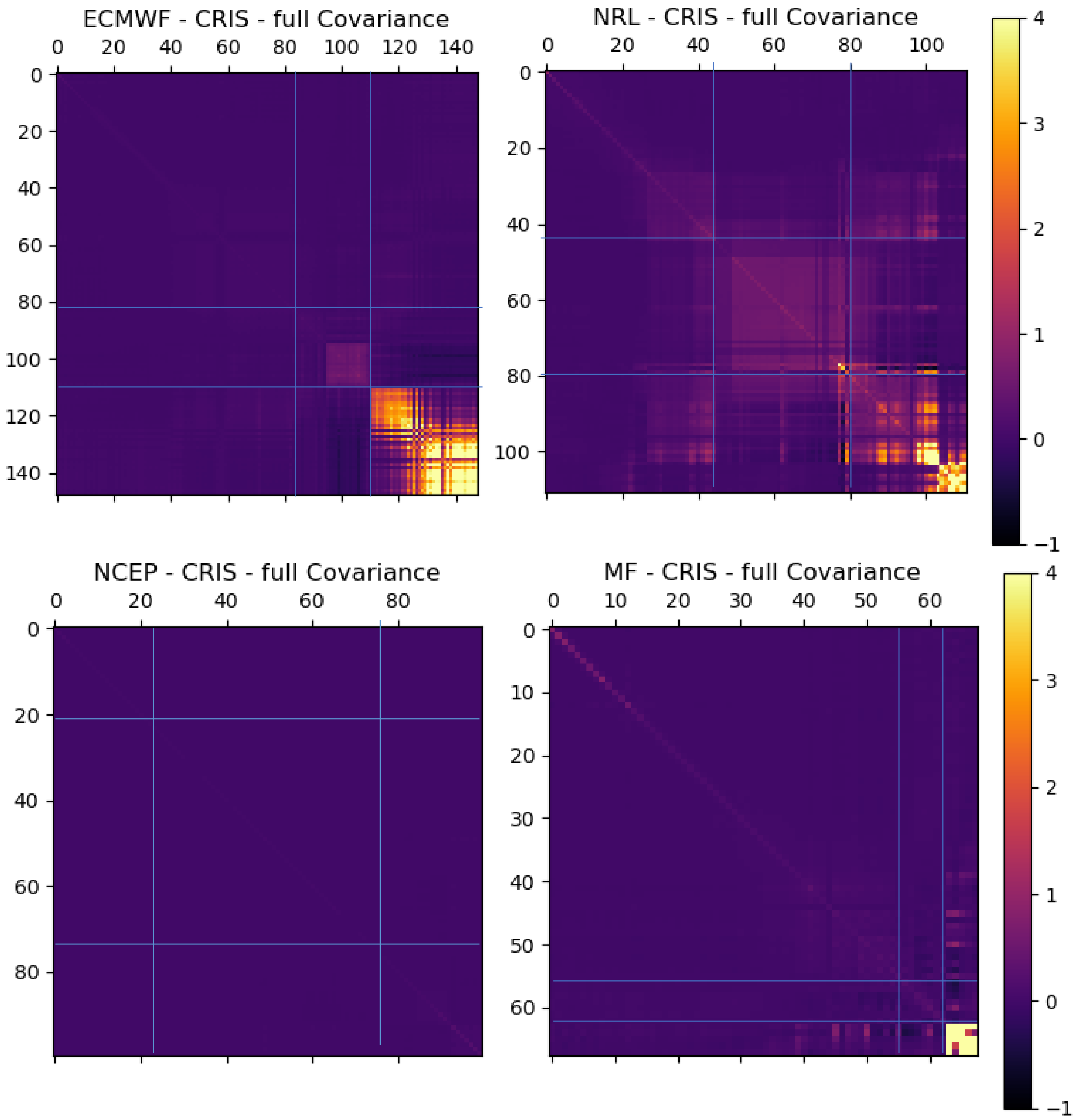
IASI correlations – common channels between three centres.

ECCC matrix shows higher correlations in tropospheric temperature sounding channels and window channels.



Australian Government
Bureau of Meteorology

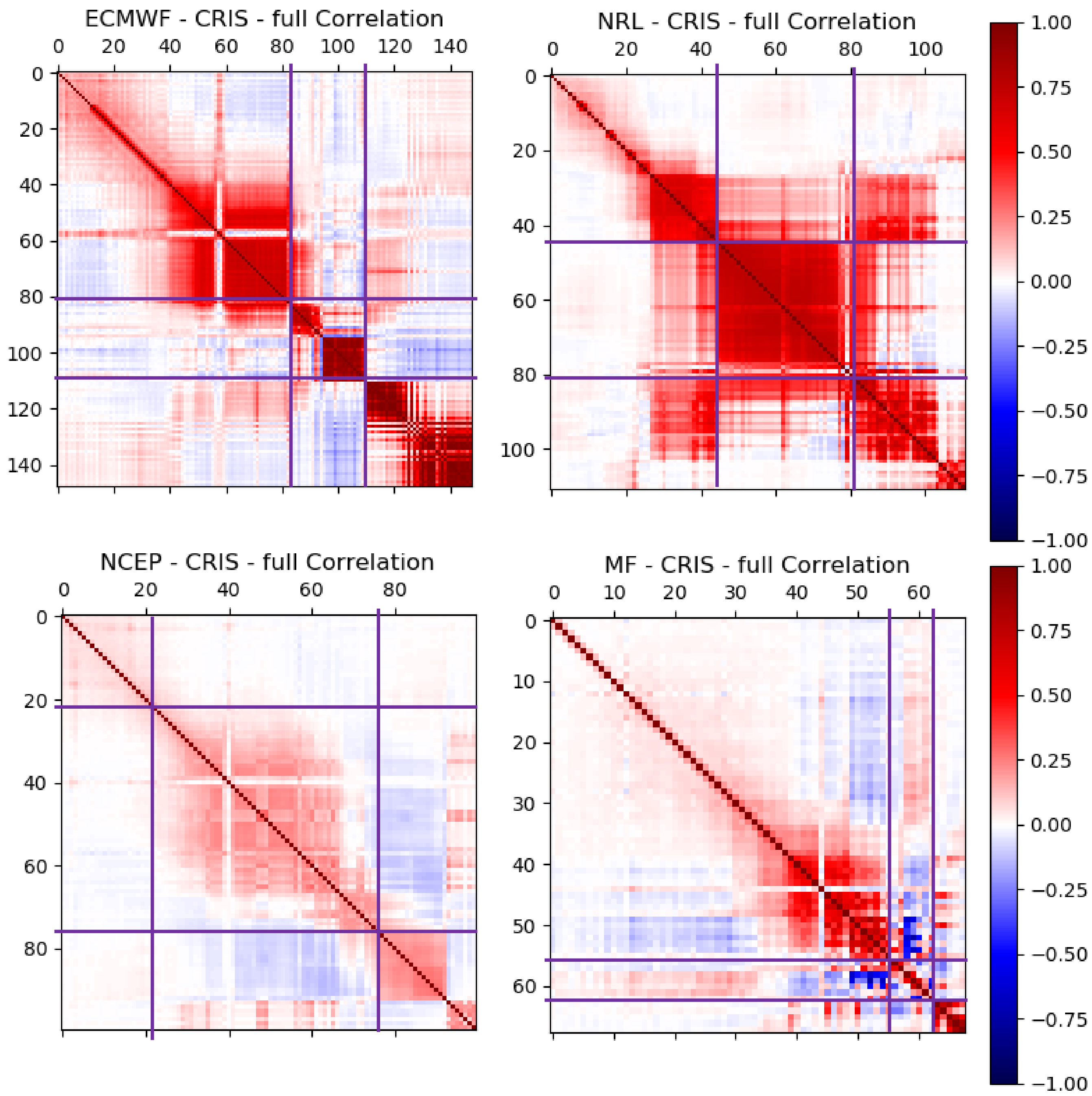
CrIS



Comparison of CrIS FSR Covariance matrices. These appear quite different.

For some centres off-diagonal elements are much more prominent than for IASI.

Not all centres use FSR, and channel selections differ greatly between centres. This makes direct comparison between centres difficult.



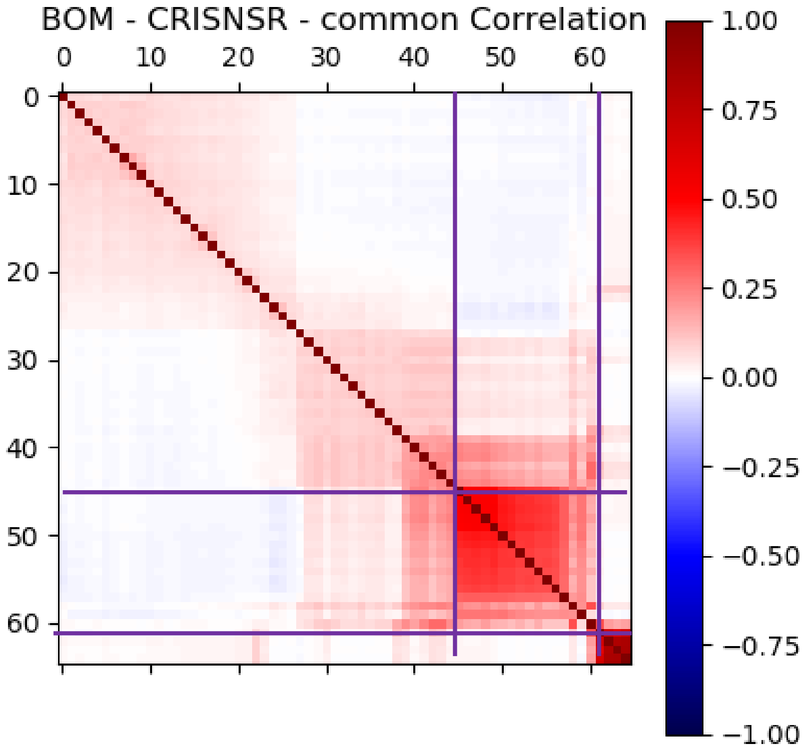
Comparison of CrIS FSR Correlation matrices. These appear quite different.

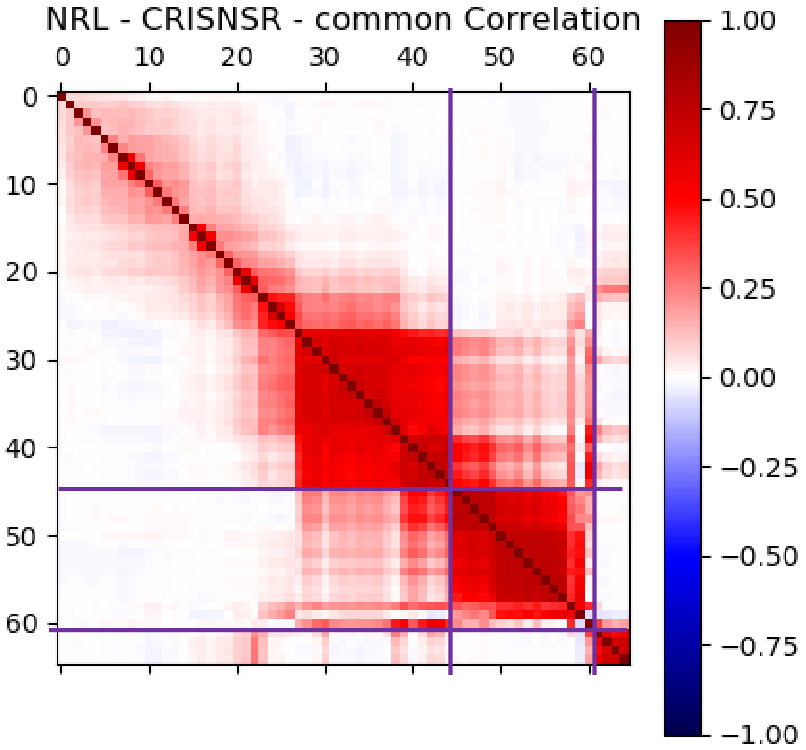
For some centres off-diagonal elements are much more prominent than for IASI.

Difficult to draw any conclusions…

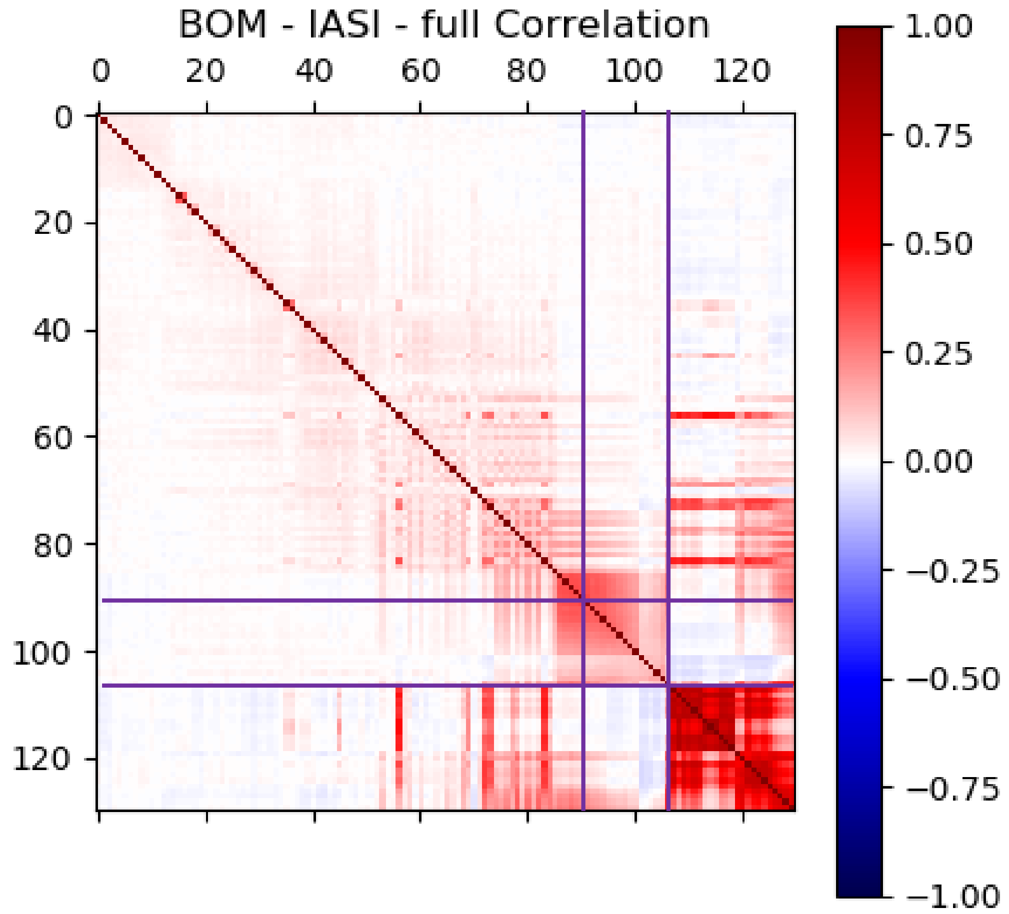
Comparison of CrIS NSR Correlation matrices. Common channels between centres.

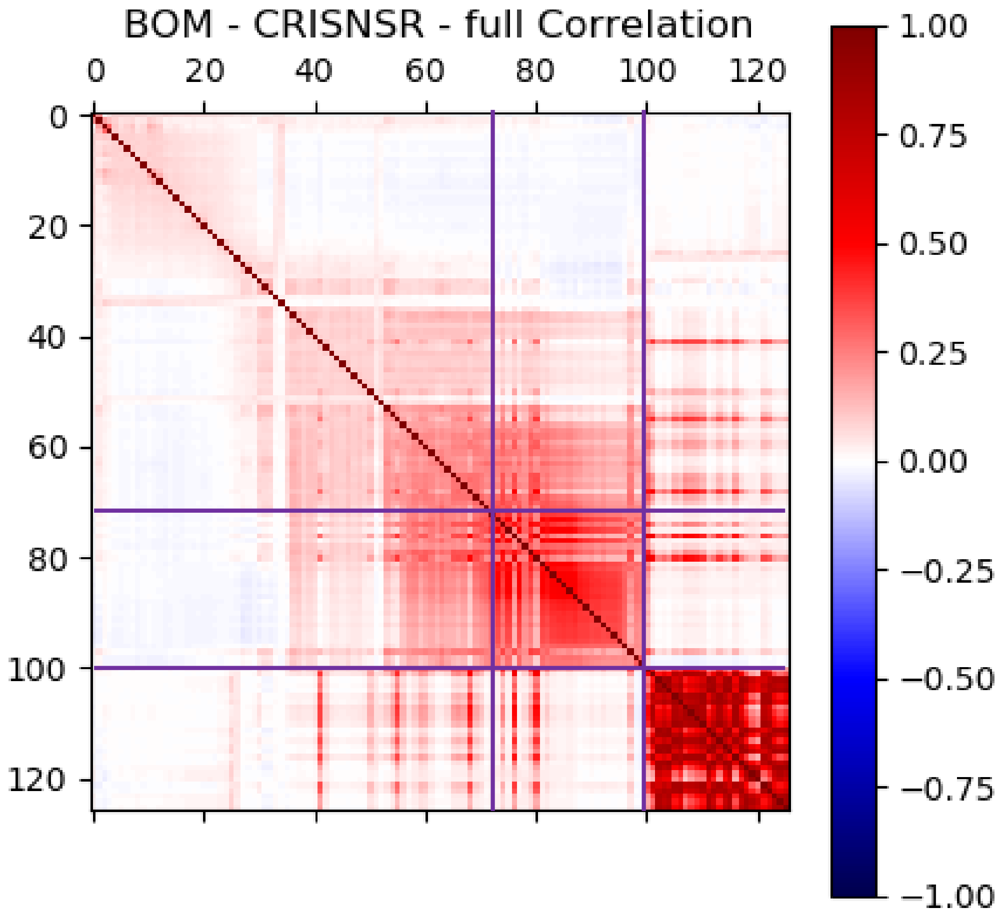
Still rather different – what does this mean for our diagnostic processes?





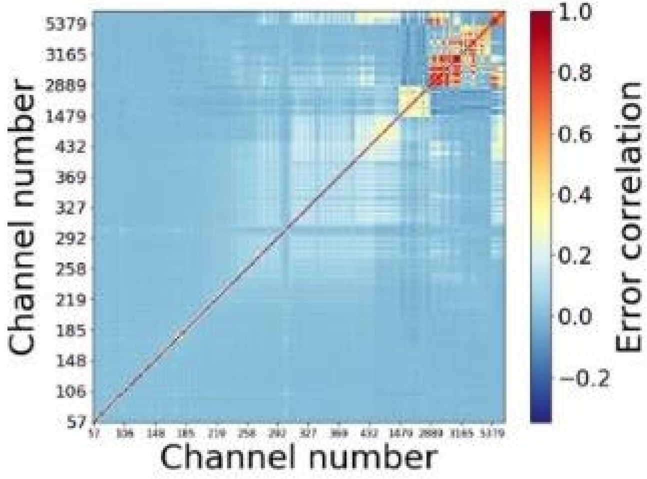
IASI vs CrIS – same centre



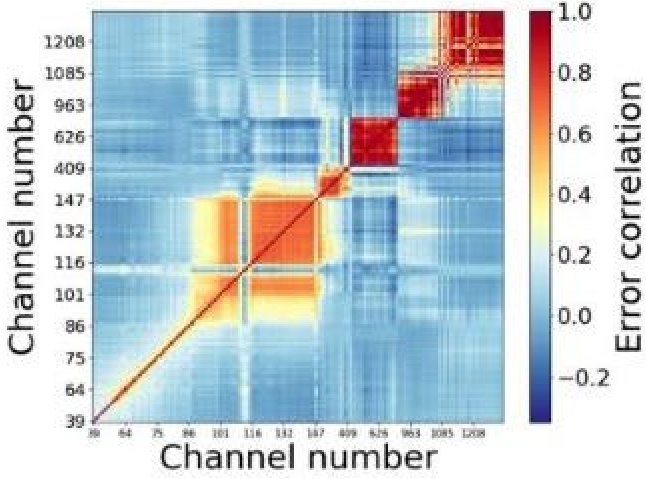


Error correlation matrices for hyperspectral IR instruments

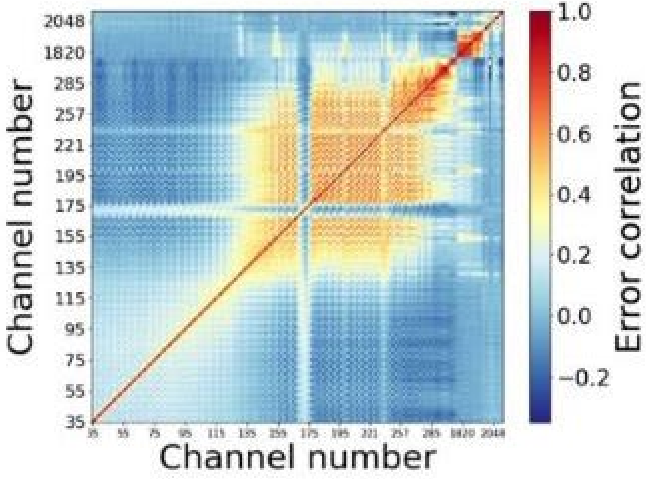
IASI



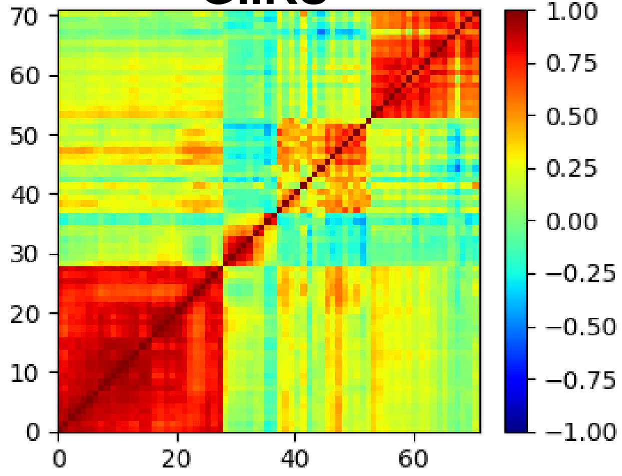
CrIS FSR



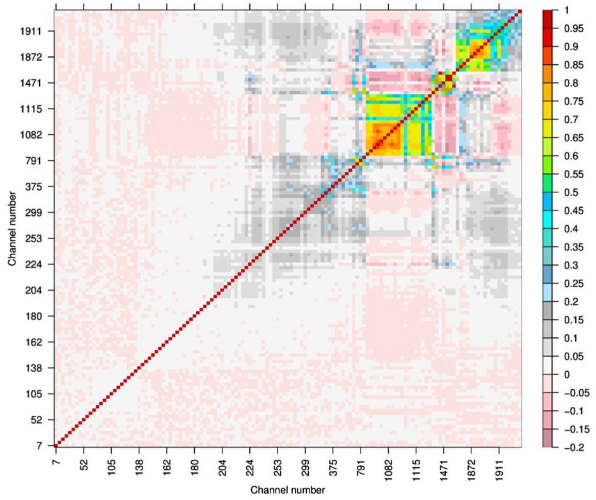
IKFS-2



GIIRS



AIRS



Australian Government
Bureau of Meteorology

New Conclusions

Summary from ECMWF Workshop on Random and Systematic Errors, November 2020

- Everyone uses Desroziers
- Everyone does some manipulation to the output
- Justifications for this manipulation vary

- ~~Results are surprisingly consistent between centres~~
  - Inflation factors are surprisingly consistent between centres
  - Condition numbers are also surprisingly consistent
  - Correlations are not very consistent

- All instruments show strong correlations for water vapour channels
- Different sounders have quite different diagnosed correlations
  - Points to different sources of error dominating for each instrument

Where do we go with this?

- Centres report some dissatisfaction with the lack of confidence that their system for error estimation meets the underlying requirements for Desroziers
- Hardly anyone uses Hollingsworth-Loennberg, and instead contort Desroziers further to estimate errors for channels that are not assimilated
- Everyone wants to know what the best method is….

- Using Desroziers is heavily dependent on the DA system and B-matrix
- Need to progress work to map the error covariances to vertical structures in B-matrix via the Jacobians

- Physical methods for error covariance estimation – work begun many years ago but not progressed.

- Write a paper!

Australian Government
Bureau of Meteorology

Thank You!

Any Questions?

Stein's Paradox in Statistics

The best guess about the future is usually obtained by computing the average of past events. Stein's paradox defines circumstances in which there are estimators better than the arithmetic average

by Bradley Efron and Carl Morris

Sometimes a mathematical result is strikingly contrary to generally held belief even though an obviously valid proof is given. Charles Stein of Stanford University discovered such a paradox in statistics in 1955. His result undermined a century and a half of work on estimation theory, going back to Karl Friedrich Gauss and Adrien Marie Legendre. After a long period of resistance to Stein's ideas, punctuated by frequent and sometimes angry debate, the sense of paradox has diminished and Stein's ideas are being incorporated into applied and theoretical statistics.

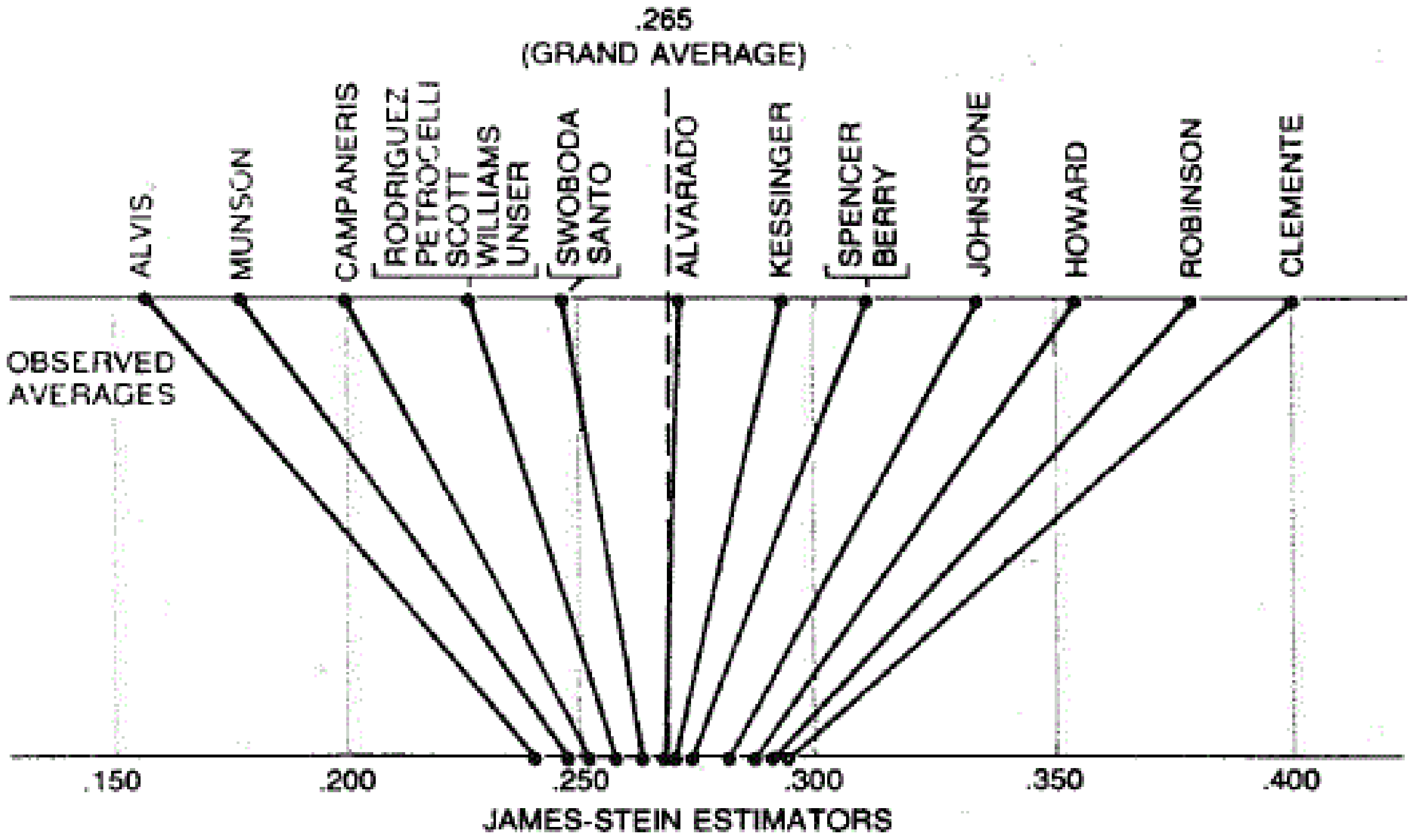
jor-league players as they were recorded after their first 45 times at bat in the 1970 season. These were all the players who happened to have batted exactly 45 times the day the data were tabulated. A batting average is defined, of course, simply as the number of hits divided by the number of times at bat; it is always a number between 0 and 1. We shall denote each such average by the letter y.

The first step in applying Stein's method is to determine the average of the averages. Obviously this grand average, which we give the symbol $\bar{y}$, must also lie between 0 and 1. The essential proc-

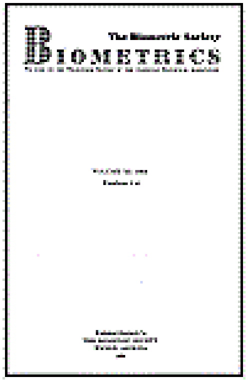
factor c is .212. Substituting these values in the equation, we find that for each player z equals $.265 + .212(\bar{y} - .265)$. Because c is about .2, each average will shrink about 80 percent of the distance to the grand average, and the total spread of the averages will be reduced about 80 percent.

As an example consider the late Roberto Clemente, who was the leading batter in the major leagues when our statistics were compiled. For Clemente y is equal to .400, and z can be determined by evaluating the expression $z = .265 + .212(.400 - .265)$. The re-

Steinian Shrinkage



JAMES-STEIN ESTIMATORS for the 18 baseball players were calculated by "shrinking" the individual batting averages toward the overall "average of the averages." In this case the grand average is .265 and each of the averages is shrunk about 80 percent of the distance to this value. Thus the theorem on which Stein's method is based asserts that the true batting abilities are more tightly clustered than the preliminary batting averages would seem to suggest they are.



JOURNAL ARTICLE

Shrinkage Estimators for Covariance Matrices

Michael J. Daniels and Robert E. Kass

Biometrics
Vol. 57, No. 4 (Dec., 2001), pp. 1173-1184 (12 pages)
Published By: International Biometric Society

https://www.jstor.org/stable/3068250

Cite this Item

- We consider here two general shrinkage approaches to estimating the covariance matrix and regression coefficients. The first involves shrinking the eigenvalues of the unstructured ML or REML estimator. The second involves shrinking an unstructured estimator toward a structured estimator. For both cases, the data determine the amount of shrinkage. These estimators are consistent and give consistent and asymptotically efficient estimates for regression coefficients. Simulations show the improved operating characteristics of the shrinkage estimators of the covariance matrix and the regression coefficients in finite samples. The final estimator chosen includes a combination of both shrinkage approaches, i.e., shrinking the eigenvalues and then shrinking toward structure.

Covariance? Correlation? Inverse Covariance?

- Does it matter whether the shrinkage operation is done on the covariance or the correlation matrix?
  - Plenty of centres shrink the covariance and then inflate the diagonal as well
- Small eigenvalues matter because error covariances are used in their inverted form ($\mathbf{R}^{-1}$ appears in the cost function, not $\mathbf{R}$)
  - 1/ very small number = very big number
  - Think of the small eigenvalues as a mode with a very small error – it's "well measured"… except that as it goes towards zero, you would say there is no information about it at all.
    - This is very confusing!
  - Is it better to shrink the inverse matrix?