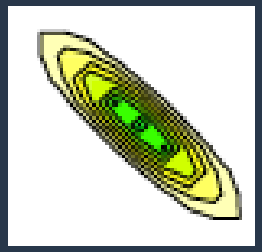
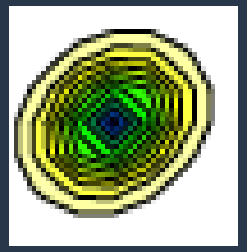
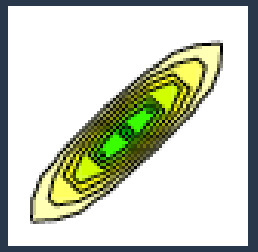







On the interaction of observation and *a-priori* error correlations in Variational data assimilation



Alison Fowler, Sarah Dance, Jo Waller

Data Assimilation Research Centre, University of Reading, UK.

Data Assimilation Research Centre

NERC SCIENCE OF THE ENVIRONMENT

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Motivation

- Background (a-priori) error correlations (BECs) known to be important in DA.
- Until recently observation error correlations (OECs) have been neglected in NWP. To account for this the data has either been thinned, 'super-obbed' or the error variance have been inflated.
  - Therefore, accounting for OECs correctly could allow for denser observations to be assimilated which could be important for high-resolution/high-impact weather forecasting.
- Accounting for inter-channel error correlations in IASI (which previously relied on variance inflation) have led to an improvement in the skill of the analysis and forecast (Weston et al., 2014, QJRMS, Bormann et al., 2016, QJRMS).
- This has motivated the OECs to be estimated for a range of other observations, for example using innovation 'Desroziers' diagnostics.
- Can we expect to see benefit to including OECs in all observation types? Are their cases when thinning the data is still desirable?

Quick literature review

Effect of suboptimal R i.e not accounting for OECs

- Rainwater et al. 2015, QJRMS
- Miyoshi et al. 2013, Inverse Problems in Science and Engineering
- Stewart et al. 2008, International J. for Numerical Methods in Fluids
- Stewart et al. 2013, Tellus A
- Liu and Rabier 2002, QJRMS – also suboptimal H

In this talk I will be assuming that the correct B, R and H are known and used in the assimilation.

Impact of OECs on different scales

- Seaman 1977, MWR
- Rainwater et al. 2015, QJRMS

Quick literature review
Impact of OECs on different scales

Positive error correlations

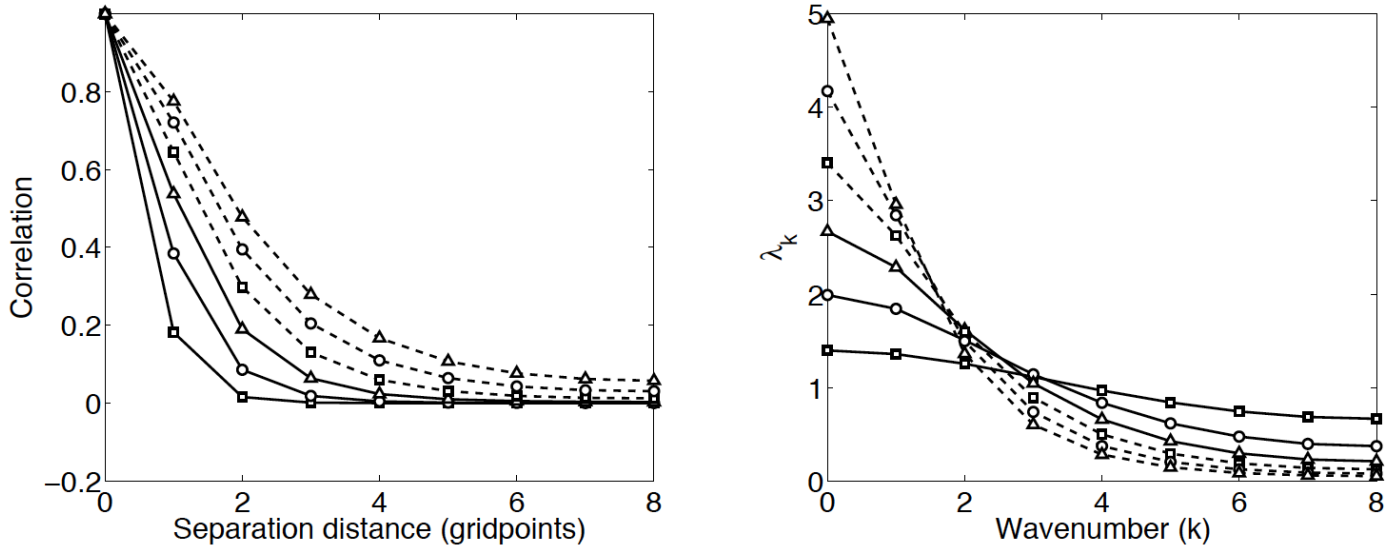
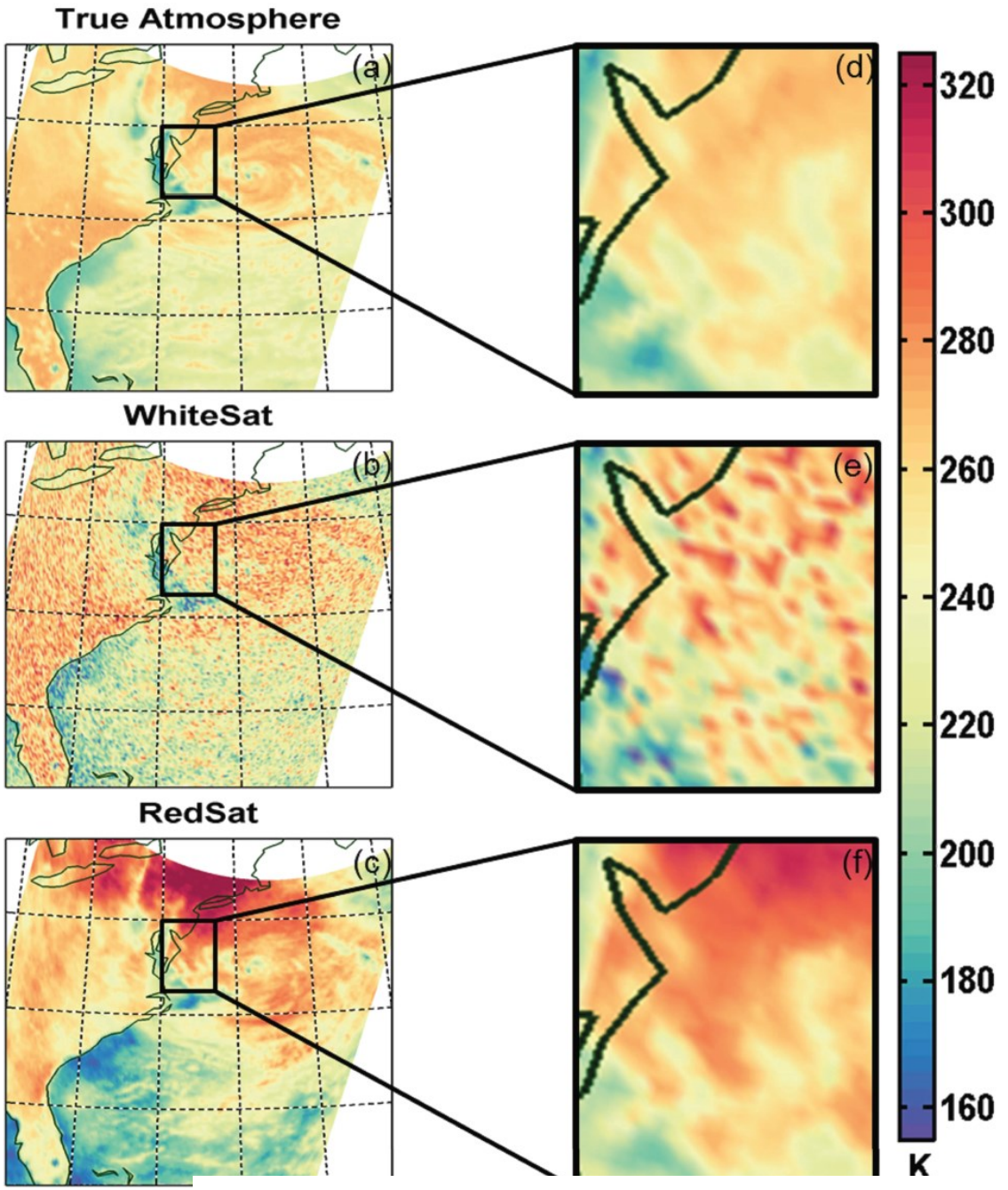


Figure 1: The SOAR function and its eigenspectrum. $L = 2$ (solid line squares), $L = 3$ (solid line circles), $L = 4$ (solid line triangles), $L = 5$ (dashed line squares), $L = 6$ (dashed line circles), $L = 7$ (dashed line triangles).

Taken from Waller et al. 2016, QJRMS



(a) A microwave satellite image of Hurricane Sandy on 24 October 2012, which is treated as truth. (b) Panel (a) plus white (uncorrelated) noise; (c) panel (a) plus red (spatially correlated) noise.

S. Rainwater, C. H. Bishop, and W. F. Campbell. The benefits of correlated observation errors for small scales. *Q. J. R. Meteorol. Soc.*, 141:3439–3445, 2015.

Quick literature review

Impact of OECs in relation to the observation operator

- Miyoshi et al. 2013, Inverse Problems in Science and Engineering
- Terasaki and Miyoshi 2014, SOLA
- Liu and Rabier 2002, QJRMS

Optimal thinning when OECs present

- Liu and Rabier 2002, QJRMS
- Bergman and Bonner 1976, MWR

Few of these papers mentioned, look at how these aspects of the impact of OECs depend upon the background error statistics, beyond noting that there is a sensitivity.

Many have only studied the impact of OECs in terms of one metric, such as analysis rmse.

Aim of this work

- To show how the impact of observations with correlated errors depends on the specification of the background error statistics (B) and the observation operator (H).

- Three different metrics studied
  - The analysis error covariance
  - The sensitivity of the analysis to the observations
  - Mutual information

Measure I: The analysis error covariance matrix

Recall the analysis, x^a, in the case of Gaussian near-linear system is given by (using standard notation)

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \mathbf{K}\left(\mathbf{y} - h(\mathbf{x}^{\mathrm{b}})\right)$$

Where $\mathbf{x}^{\mathrm{b}} \in \mathbb{R}^n$ is the background, $\mathbf{y} \in \mathbb{R}^p$ is the observation vector, $\mathbf{K}$ is the Kalman gain and h is the (near-linear) observation operator

$$\mathbf{K} \in \mathbb{R}^{n \times p} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

The analysis error covariance matrix is

$$\mathbf{P}^{\mathrm{a}} \in \mathbb{R}^{n \times n} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} = (\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}$$

Where $\mathbf{B} \in \mathbb{R}^{n \times n}$ and $\mathbf{R} \in \mathbb{R}^{p \times p}$ are the background and observation error covariance matrices

$\mathbf{H} \in \mathbb{R}^{p \times n}$ is the linearised observation operator

Measure II: The sensitivity matrix

- This quantifies the sensitivity of the analysis (in observation space) to the observations

$$\mathbf{S} = \frac{\partial h(\mathbf{x}^{\mathrm{a}})}{\partial \mathbf{y}} \approx \mathbf{HK}$$

- The diagonal elements are known as the self-sensitivities
- The off-diagonal elements are known as the cross-sensitivities, measure the spread in information.
- dfs=trace(S) i.e. a sum of the self-sensitivities.
- The sensitivity matrix can be related to the analysis error covariance matrix:

$$\mathbf{S} = \mathbf{HP}^{\mathrm{a}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$$

Measure III: Mutual information

- Mutual information measures the reduction in entropy due to the assimilation of observations
- Entropy (uncertainty) is defined as

$$E(P) = -\int P(x) \ln P(x) dx.$$

- The entropy of a Gaussian distribution with covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is therefore

$$E(P) = n \ln(2\pi e)^{1/2} + \frac{1}{2} \ln |\boldsymbol{\Sigma}|$$

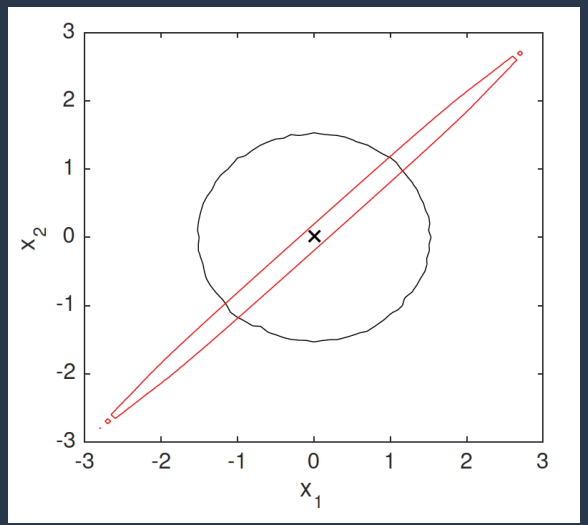


Illustration of the 95th percentile of the region of uncertainty given
by Gaussian distribution when the errors in x_1 and x_2 are uncorrelated (black)and correlated with a coefficient of 0.999 (red).

In the first case the entropy is 2.8379 and in the second case the entropy is 0.8794.

Measure III: Mutual information

- Mutual information is the prior entropy minus the posterior entropy

$$MI = 0.5 \ln |\mathbf{B}\mathbf{P}_a^{-1}|,$$

- This can also be written in terms of the sensitivity matrix as

$$MI = -0.5 \ln |\mathbf{I}_p - \mathbf{S}| = -0.5 \sum_{k=0}^{p-1} \ln(1 - \lambda_k^s),$$

- Unlike dfs, MI is a function of both the self and cross sensitivities.

2 variable experimental design

- In the following experiments, the background and observation error variances are defined as:

$$\mathbf{B} = \beta \begin{bmatrix} 1 & \gamma \\ \gamma & 1 \end{bmatrix} \quad \text{where } -1 < \gamma < 1,$$

$$\mathbf{R} = \rho \begin{bmatrix} 1 & \psi \\ \psi & 1 \end{bmatrix} \quad \text{where } -1 < \psi < 1.$$

- The observation operator is given by:

$$\mathbf{H} = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \quad \text{where } -1 < a < 1.$$

e.g. a=0.5 =>
$y_1=x_1+0.5x_2$, $y_2=x_2+0.5x_1$
(obs themselves are +vely corr in state space.)
a=-0.5 =>
$y_1=x_1-0.5x_2$, $y_2=x_2-0.5x_1$
(obs themselves are -vely corr in state space.)

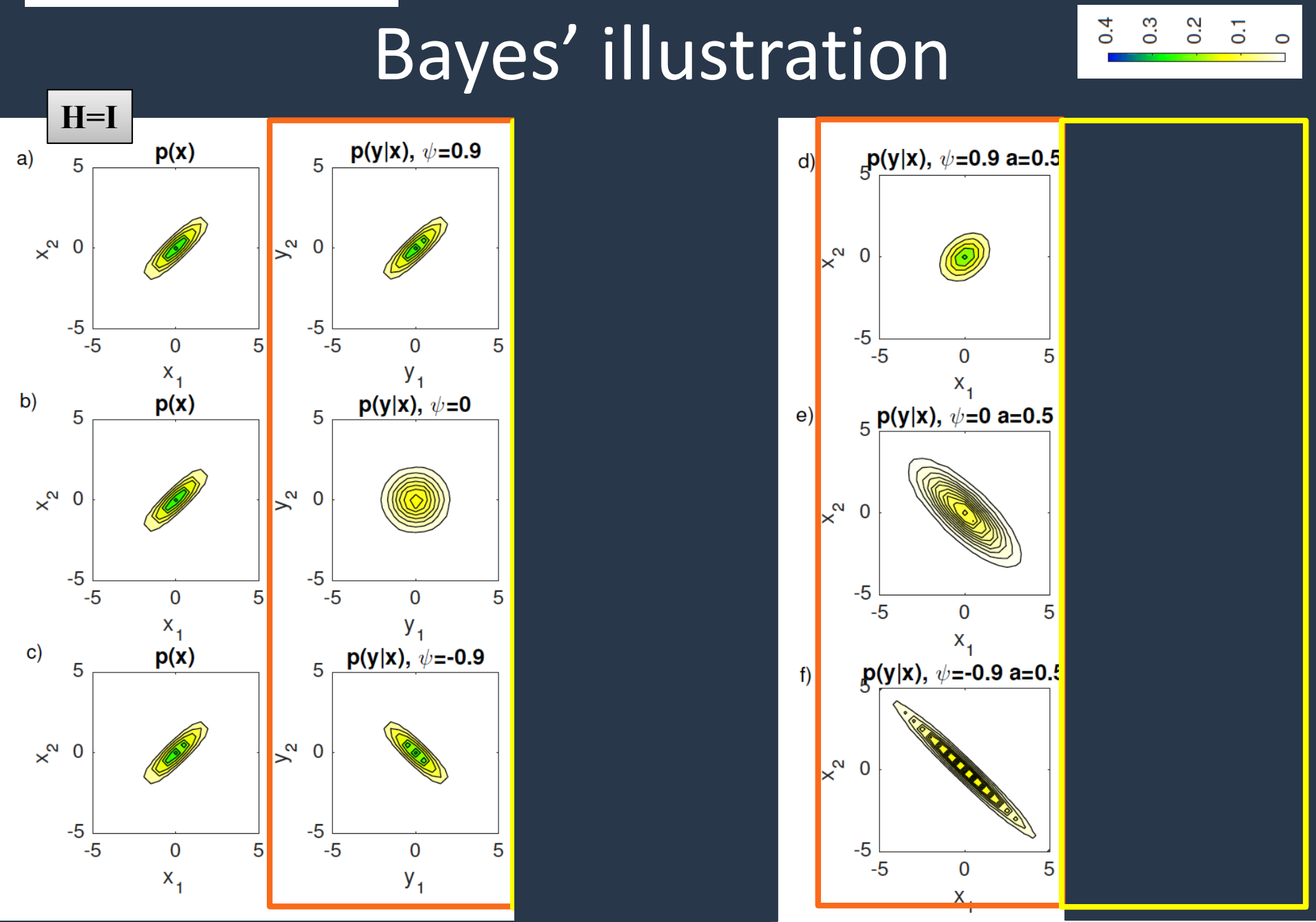
- This setup results in circulant analysis error covariance and sensitivity matrices.

$$\mathbf{P}^{\mathrm{a}} = \begin{bmatrix} P^a_{ii} & P^a_{ij} \\ P^a_{ij} & P^a_{ii} \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} S_{ii} & S_{ij} \\ S_{ij} & S_{ii} \end{bmatrix}$$

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$$

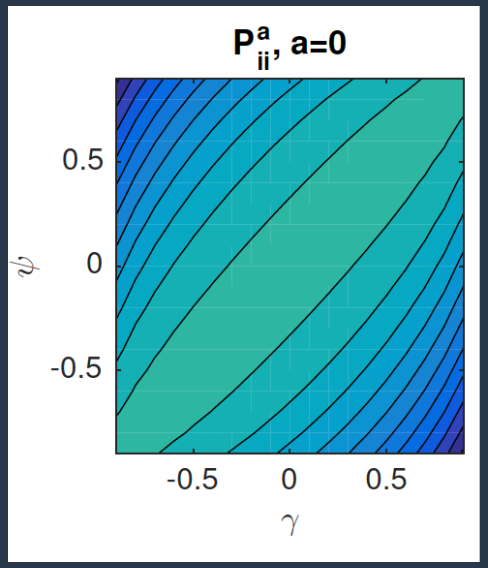
β=ρ=1, γ=0.9

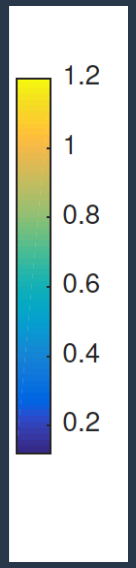
Bayes' illustration

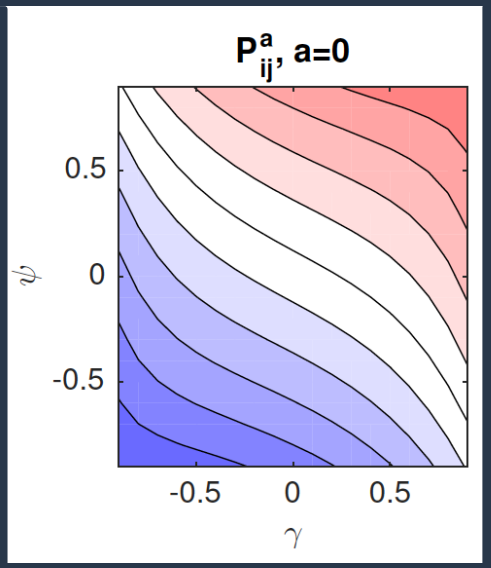


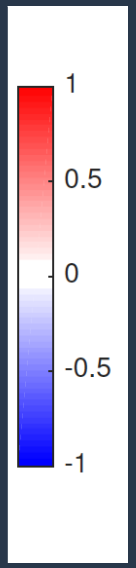
γ Background error correlations
Ψ Observation error correlations

Analysis error covariance, $\beta=2, \rho=1$







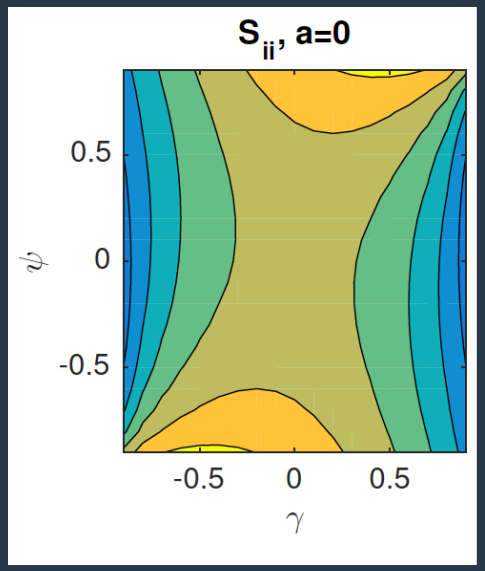


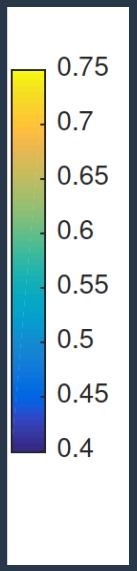
Analysis more accurate at small scales

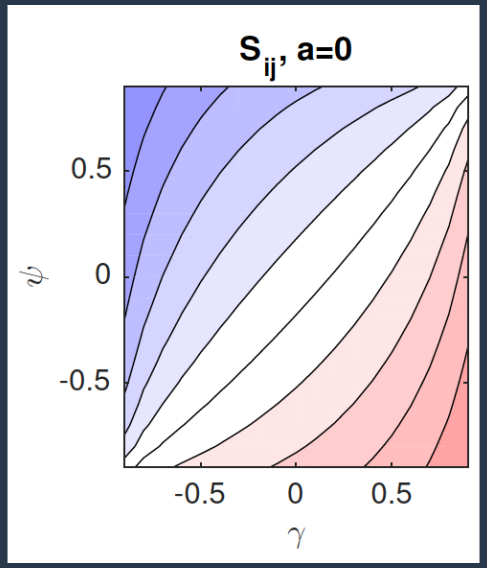
Analysis more accurate at large scales

γ Background error correlations
Ψ Observation error correlations

Analysis sensitivity, $\beta=2, \rho=1$





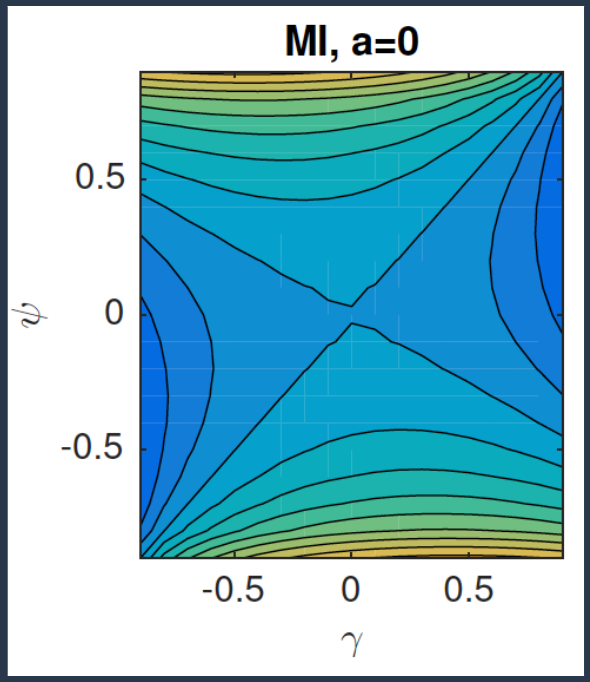


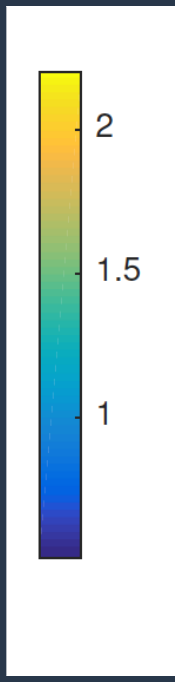
Analysis at large scales more sensitive to observations

Analysis at small scales more sensitive to observations

γ Background error correlations
Ψ Observation error correlations

Mutual information, $\beta=2, \rho=1$





Note: this looks very different to the pattern seen for the self-sensitivities => using dfs as a measure of information content would lead to different conclusions

Data thinning

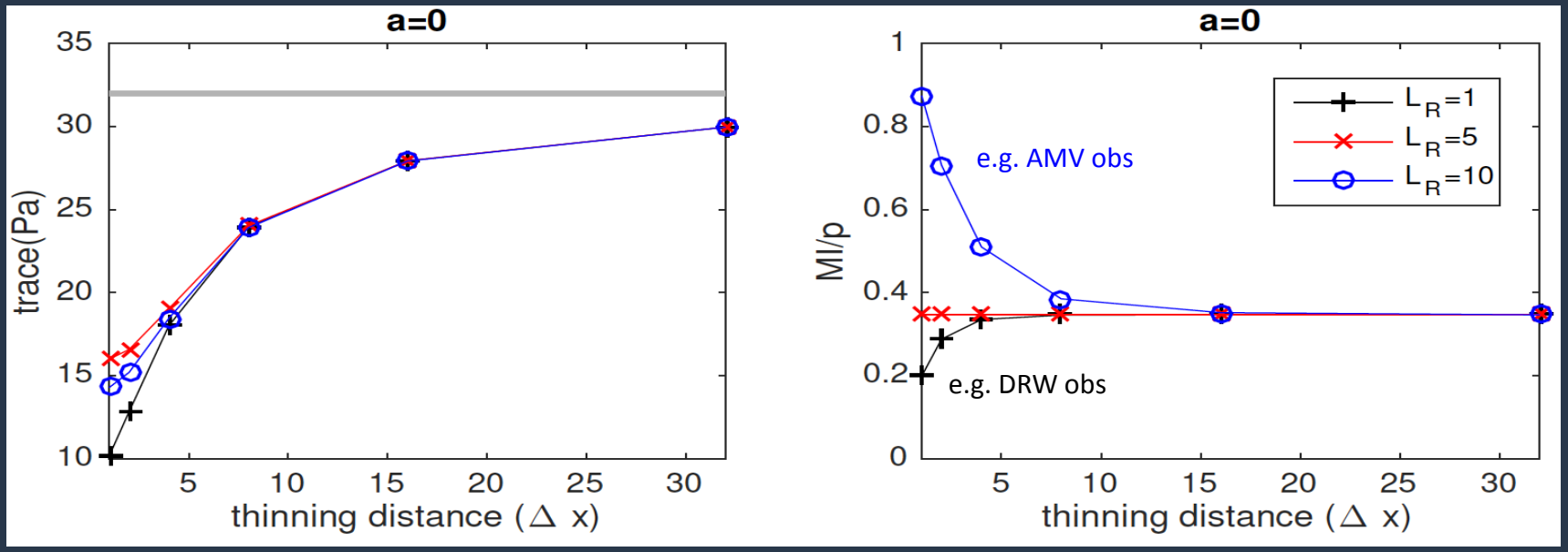
- It was suggested by Miyoshi et al. 2013, who noted the dependence of the observation impact on both the OECs and H, that this could be used in the design of instruments.
- However, in practice this may not be feasible as often the OECs and H are related.
- An easier way to exploit these results is in the choice of the density of the observations.

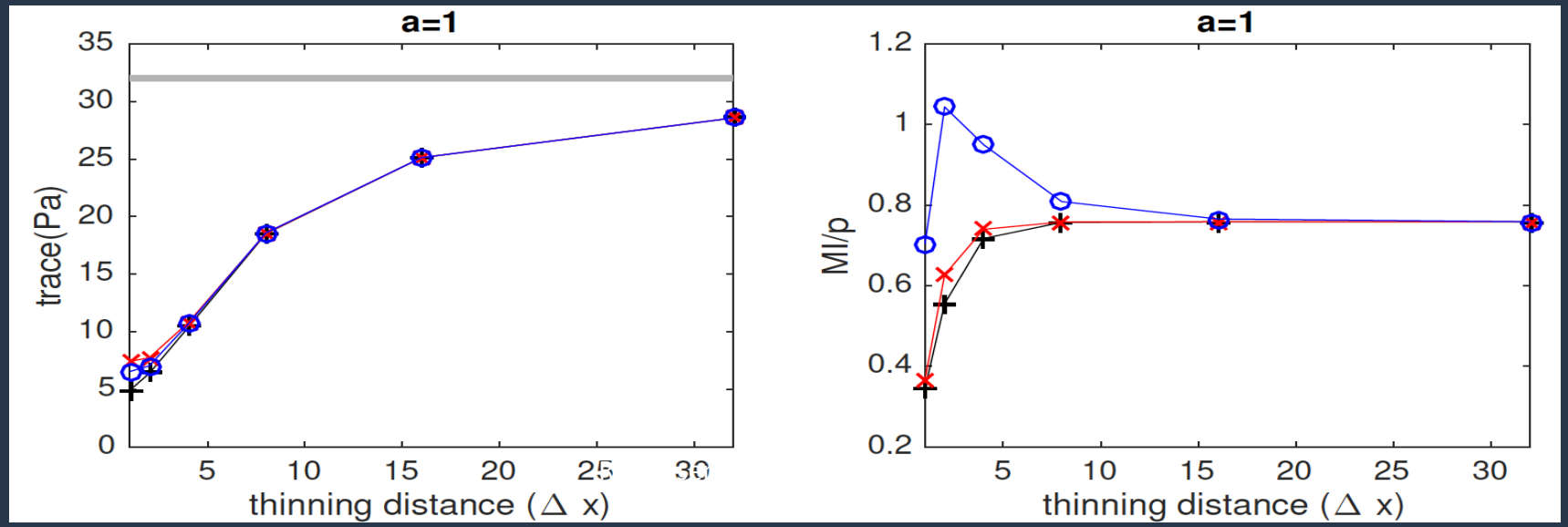
Data thinning

- Results shown assume circular domain discretized into 32 gridpoints.
- R and B are given by circulant matrices with correlations given by the SOAR function $c_k = (1 + r_k/L)e^{-r_k/L}$
- The background error correlation lengthscale is given by L_b=5. The observation error correlation lengthscale L_o is allowed to vary.
- The observation and background error variances are both set to 1.
- $\mathbf{H}$ assumes observations are made regularly and are a weighted combination of the state variables

$$h(x_i) = \sum_{j=i-a}^{i+a} \frac{a+1-|j-i|}{a+1} x_j$$

Data thinning L_B=5





Data thinning

Key findings

- When the correlation lengthscales in the likelihood and prior are similar there is less benefit, in terms of the analysis error, in increasing the density of the observations compared to if the lengthscales are very different.
- If OECs are correctly included in the DA systems then denser observations are beneficial (in terms of MI/p) only if the lengthscales in $\mathbf{R}$ are much larger than $\mathbf{B}$.
- If the observations have overlapping weighting functions then this dominates the effect of the OECs.

Conclusions

- The impact of observations with correlated errors cannot be considered in isolation of the background error correlations or the observation operator.
  - This is particularly true if interested in the impact on analysis errors or the spread in information
- These results could be used in the design of new instruments and observing networks.
  - For example for choosing the optimal density of the data, to provide the most efficient choice of observations
  - In an LETKF this will change as the correlation lengthscales described by the ensemble evolve.
  - These ideas should be investigated with an OSSE study.

THANK YOU FOR LISTENING