Measures of observation impact in non-Gaussian data assimilation

Alison Fowler and Peter Jan van Leeuwen

DA Meeting, 28th March 2012

Introduction

- Gaussian data assimilation has proven to be a powerful tool in NWP, however…
- There are many different potential sources of non-Gaussianity in data assimilation.
  - Non-Gaussian prior, $p(x)$, errors may result from a non-linear forecast model or be an intrinsic property of the state variable e.g. bounds on physical values.
  - Non-Gaussian likelihood , $p(y|x)$, may be due to a non-linear observation operator or due to characteristics of the instrument. (Quality control).
- Many different methods for dealing with these sources of non-Gaussianity in DA
  - from explicitly reformulating the cost function in terms of a given distribution
  - to avoiding making any assumptions and allowing the non-linearity of the models to generate the distributions implicitly (e.g. particle filter)

Outline of talk

- Observation impact in Gaussian data assimilation.
  - Introduction to different measures
- The influence of a non-Gaussian statistics on observation impact.
  - PART I: the non-Gaussian prior
  - PART II: the non-Gaussian likelihood
- Future work

Outline of talk

- Observation impact in Gaussian data assimilation.
  - Introduction to different measures
- The influence of a non-Gaussian statistics on observation impact.
  - PART I: the non-Gaussian prior
  - PART II: the non-Gaussian likelihood
- Future work

Measures of observation impact in ***Gaussian*** data assimilation

- Many different measures exist
- used for targeted observations, design of new observing systems, data thinning, monitoring the DA scheme etc.
- I will concentrate on 3:

Measures of observation impact in ***Gaussian*** data assimilation

- Many different measures exist
- used for targeted observations, design of new observing systems, data thinning, monitoring the DA scheme etc.
- I will concentrate on 3:
  - The sensitivity of the analysis to the observations

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}}$$

$\mathbf{x}_a$ is the analysis vector
$\mathbf{y}$ is observation vector
$\mathbf{H}$ is the linearised ob operator (normally about the analysis)

Measures of observation impact in ***Gaussian*** data assimilation

- Many different measures exist
- used for targeted observations, design of new observing systems, data thinning, monitoring the DA scheme etc.
- I will concentrate on 3:
  - The sensitivity of the analysis to the observations

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}}$$

$\mathbf{x}_a$ is the analysis vector
$\mathbf{y}$ is observation vector
$\mathbf{H}$ is the linearised ob operator (normally about the analysis)

  - Mutual information

$$\mathrm{MI} = -\int p(x) ln p(x) dx + \int p(y) \int p(x|y) ln p(x|y) dx dy$$

Measures of observation impact in ***Gaussian*** data assimilation

- Many different measures exist
- used for targeted observations, design of new observing systems, data thinning, monitoring the DA scheme etc.
- I will concentrate on 3:
  - The sensitivity of the analysis to the observations

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}}$$

$\mathbf{x}_a$ is the analysis vector
$\mathbf{y}$ is observation vector
$\mathbf{H}$ is the linearised ob operator (normally about the analysis)

  - Mutual information
    $\mathrm{MI} = -\int p(x) ln p(x) dx + \int p(y) \int p(x|y) ln p(x|y) dx dy$
  - Relative entropy $RE = \int p(x|y) ln \frac{p(x|y)}{p(x)} dx$

Measures of observation impact in ***Gaussian*** data assimilation

- Many different measures exist
- used for targeted observations, design of new observing systems, data thinning, monitoring the DA scheme etc.
- I will concentrate on 3:
  - The sensitivity of the analysis to the observations

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}}$$

$\mathbf{x}_a$ is the analysis vector
$\mathbf{y}$ is observation vector
$\mathbf{H}$ is the linearised ob operator (normally about the analysis)

  - Mutual information
    $\mathrm{MI} = -\int p(x) ln p(x) dx + \int p(y) \int p(x|y) ln p(x|y) dx dy$
  - Relative entropy $RE = \int p(x|y) ln \frac{p(x|y)}{p(x)} dx$
    - NOTE: $MI = \int p(y) RE dy$

Measures of observation impact in ***Gaussian*** data assimilation

- Many different measures exist
- used for targeted observations, design of new observing systems, data thinning, monitoring the DA scheme etc.
- I will concentrate on 3:
  - The sensitivity of the analysis to the observations

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}}$$

$\mathbf{x}_a$ is the analysis vector
$\mathbf{y}$ is observation vector
$\mathbf{H}$ is the linearised ob operator (normally about the analysis)

  - Mutual information
    $\mathrm{MI} = -\int p(x) ln p(x) dx + \int p(y) \int p(x|y) ln p(x|y) dx dy$
  - Relative entropy $RE = \int p(x|y) ln \frac{p(x|y)}{p(x)} dx$
    - NOTE: $MI = \int p(y) RE dy$
- These can all be explicitly derived in the case of Gaussian DA.

Measures of observation impact in ***Gaussian*** data assimilation

- Recal in Gaussian DA
  - $\mathbf{x}_\mathrm{a} = \mathbf{x}_\mathrm{b} + \mathbf{K}\big(\mathbf{y} - h(\mathbf{x}_\mathrm{b})\big),$
    - where $\mathbf{K} = \mathbf{P}_\mathrm{a}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}$
    - and $\mathrm{P_a} = \left(\mathrm{H^T R^{-1} H + B^{-1}}\right)^{-1}$

$\mathbf{x}_\mathrm{a}$ is the analysis vector
$\mathbf{x}_\mathrm{b}$ is the background (first guess, prior) vector
$\mathbf{y}$ is observation vector
h is the observation operator
$\mathbf{P}_\mathrm{a}$ is the analysis error cov matrix
$\mathbf{H}$ is the linearised ob operator (normally about the analysis)
$\mathbf{R}$ is the ob error cov matrix
$\mathbf{B}$ is the prior error cov matrix

- Therefore the sensitivity of the analysis to observations can be expressed as
  - $\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_\mathrm{a}}{\partial \mathbf{y}} = \mathrm{HK{=}HP_a H^T R^{-1}}$ (Cardinali et al. 2004).
- The analysis is most sensitive to accurate observations which give information about regions of state space for which there is little prior knowledge.

Measures of observation impact in ***Gaussian*** data assimilation

- Mutual information in Gaussian DA can be shown to be
  - $MI = \frac{1}{2} ln\left|\mathbf{B}\mathbf{P}_a{}^{-1}\right|$ (Rodgers, 2000).
    - Like the sensitivity this depends on $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{H}$ only

Measures of observation impact in ***Gaussian*** data assimilation

- Mutual information in Gaussian DA can be shown to be
  - $MI = \frac{1}{2} ln\left|\mathbf{B}\mathbf{P}_a{}^{-1}\right|$ (Rodgers, 2000).
    - Like the sensitivity this depends on $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{H}$ only
    - $MI = -\frac{1}{2}\Sigma ln(1-\lambda_i)$, where λ_i is the i^{th} eigenvalue of $\mathbf{S}$.

Measures of observation impact in ***Gaussian*** data assimilation

- Mutual information in Gaussian DA can be shown to be
  - $MI = \frac{1}{2} ln\left|\mathbf{BP}_a{}^{-1}\right|$ (Rodgers, 2000).
    - Like the sensitivity this depends on $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{H}$ only
    - $MI = -\frac{1}{2}\Sigma ln(1-\lambda_i)$, where λ_i is the i^{th} eigenvalue of $\mathbf{S}$.

- Relative entropy in Gaussian DA can be shown to be.
  - $RE =$
    $\frac{1}{2}(\mathbf{x}_\mathrm{a}-\mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{B}^{-1}(\mathbf{x}_\mathrm{a}-\mathbf{x}_\mathrm{b})+\frac{1}{2} ln\left|\mathbf{BP}_\mathrm{a}{}^{-1}\right|+\frac{1}{2}trace(\mathbf{B}^{-1}\mathbf{P}_\mathrm{a})-\frac{n}{2}.$
    - Where n is the size of the state vector.

Measures of observation impact in ***Gaussian*** data assimilation

- Mutual information in Gaussian DA can be shown to be
  - $MI = \frac{1}{2} ln\left|\mathbf{BP}_a{}^{-1}\right|$ (Rodgers, 2000).
    - Like the sensitivity this depends on $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{H}$ only
    - $MI = -\frac{1}{2}\Sigma ln(1-\lambda_i)$, where λ_i is the i^{th} eigenvalue of $\mathbf{S}$.

- Relative entropy in Gaussian DA can be shown to be.
  - $RE =$
    $\frac{1}{2}(\mathbf{x}_\mathrm{a}-\mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{B}^{-1}(\mathbf{x}_\mathrm{a}-\mathbf{x}_\mathrm{b}) + \frac{1}{2} ln\left|\mathbf{BP}_\mathrm{a}{}^{-1}\right| + \frac{1}{2} trace(\mathbf{B}^{-1}\mathbf{P}_\mathrm{a}) - \frac{n}{2}$.
    - Where n is the size of the state vector.
    - $RE = \frac{1}{2}(\mathbf{x}_\mathrm{a}-\mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{B}^{-1}(\mathbf{x}_\mathrm{a}-\mathbf{x}_\mathrm{b}) + MI - \frac{1}{2} trace(\mathbf{S})$.

Measures of observation impact in ***Gaussian*** data assimilation

- Note that the sensitivity and mutual information both depend solely on the error covariance of the background and obs and $\mathbf{H}$ (the linearised relationship between the state and ob space).
- Relative entropy is a quadratic function of $\mathbf{y}$- so this cannot be calculated before the value of the assimilated observation is known.
- However a study by Xu et al. (2009) found that for defining the optimal radar scan configurement it did not matter which measure was used.

Outline of talk

- Observation impact in Gaussian data assimilation.
  - Introduction to different measures
- The influence of a non-Gaussian statistics on observation impact.
  - PART I: the non-Gaussian prior
  - PART II: the non-Gaussian likelihood
- Future work

PART I: ***Non-Gaussian Prior***

- When the likelihood is Gaussian (and h is linear: $h(\mathbf{x}) = \mathbf{H}\mathbf{x}$) and the prior distribution is arbitrary, it is possible to derive the sensitivity of the mean of the posterior, $\boldsymbol{\mu}_a$, to the mean of likelihood, $\boldsymbol{\mu}_y$, analytically.
- $$\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \frac{\int \mathrm{Hxp(x)}\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathrm{dx}}{\int \mathrm{p(x)p}(\mathbf{y}|\mathbf{x})\mathrm{dx}} - \mathbf{H}\boldsymbol{\mu}_\mathbf{a}\frac{\int \mathrm{p}(\mathbf{x})\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathbf{dx}}{\int \mathrm{p}(\mathbf{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathbf{dx}}.$$
- Know $\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}} = -\mathrm{p}(\mathbf{y}|\mathbf{x})(\boldsymbol{\mu}_\mathbf{y} - \mathbf{H}(\mathbf{x}))^\mathbf{T}\mathbf{R}^{-1}$

PART I: ***Non-Gaussian Prior***

- When the likelihood is Gaussian (and h is linear: $h(\mathbf{x}) = \mathbf{Hx}$) and the prior distribution is arbitrary, it is possible to derive the sensitivity of the mean of the posterior, $\boldsymbol{\mu}_a$, to the mean of likelihood, $\boldsymbol{\mu}_y$, analytically.
- $\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \frac{\int \mathrm{Hxp(x)}\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathrm{dx}}{\int \mathrm{p(x)p}(\mathbf{y}|\mathbf{x})\mathrm{dx}} - \mathbf{H}\boldsymbol{\mu}_\mathbf{a}\frac{\int \mathrm{p}(\mathbf{x})\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathbf{dx}}{\int \mathrm{p}(\mathbf{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathbf{dx}}.$
- Know $\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}} = -\mathrm{p}(\mathbf{y}|\mathbf{x})(\boldsymbol{\mu}_\mathbf{y} - \mathbf{H}(\mathbf{x}))^\mathbf{T}\mathbf{R}^{-1}$
- $\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \mathbf{HP}_a\mathbf{H}^\mathbf{T}\mathbf{R}^{-1}.$

PART I: ***Non-Gaussian Prior***

- When the likelihood is Gaussian (and h is linear: $h(\mathbf{x}) = \mathbf{Hx}$) and the prior distribution is arbitrary, it is possible to derive the sensitivity of the mean of the posterior, $\boldsymbol{\mu}_a$, to the mean of likelihood, $\boldsymbol{\mu}_y$, analytically.
- $\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \frac{\int \mathrm{Hxp(x)}\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathrm{dx}}{\int \mathrm{p(x)p}(\mathbf{y}|\mathbf{x})\mathrm{dx}} - \mathbf{H}\boldsymbol{\mu}_\mathbf{a}\frac{\int \mathrm{p}(\mathbf{x})\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathbf{dx}}{\int \mathrm{p}(\mathbf{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathbf{dx}}.$
- Know $\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}} = -\mathrm{p}(\mathbf{y}|\mathbf{x})(\boldsymbol{\mu}_\mathbf{y} - \mathbf{H}(\mathbf{x}))^\mathbf{T}\mathbf{R}^{-1}$
- $\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \mathbf{HP}_a\mathbf{H}^\mathbf{T}\mathbf{R}^{-1}.$
- When prior is non-Gaussian, $\mathbf{P}_\mathrm{a}$ becomes a function of the observation value.

RECALL: Gaussian case

$\mathbf{P}_a = \left(\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}\right)^{-1}$

PART I: ***Non-Gaussian Prior***

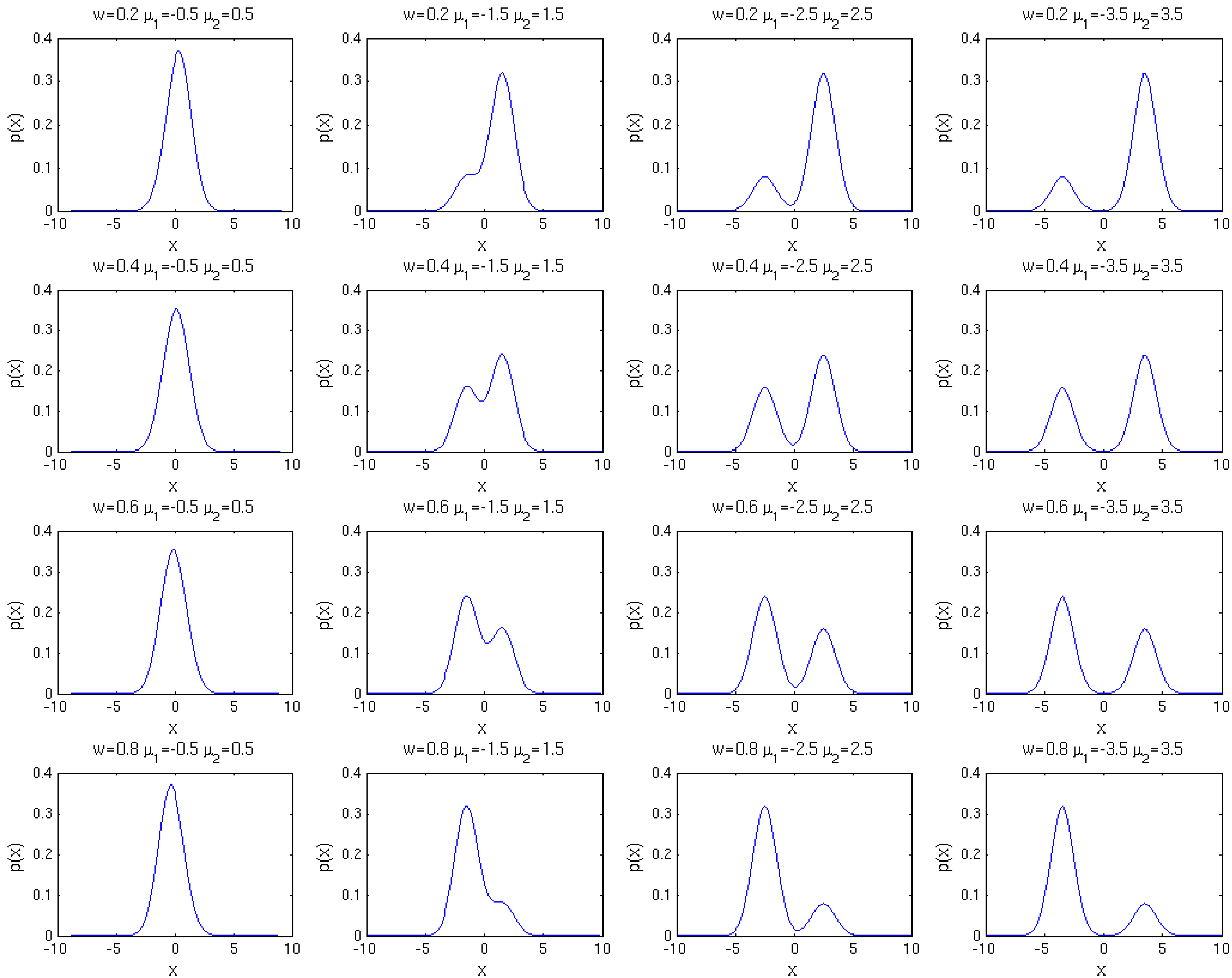
- When the likelihood is Gaussian (and h is linear: $h(\mathbf{x}) = \mathbf{Hx}$) and the prior distribution is arbitrary, it is possible to derive the sensitivity of the mean of the posterior, $\boldsymbol{\mu}_a$, to the mean of likelihood, $\boldsymbol{\mu}_y$, analytically.
- $$\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \frac{\int \mathrm{Hxp(x)}\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathrm{dx}}{\int \mathrm{p(x)p}(\mathbf{y}|\mathbf{x})\mathrm{dx}} - \mathbf{H}\boldsymbol{\mu}_\mathbf{a}\frac{\int \mathrm{p}(\mathbf{x})\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathbf{dx}}{\int \mathrm{p}(\mathbf{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathbf{dx}}.$$
- Know $\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}} = -\mathrm{p}(\mathbf{y}|\mathbf{x})(\boldsymbol{\mu}_\mathbf{y} - \mathbf{H}(\mathbf{x}))^\mathbf{T}\mathbf{R}^{-1}$
- $$\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \mathbf{HP}_a\mathbf{H}^\mathbf{T}\mathbf{R}^{-1}.$$

> RECALL: Gaussian case
> $\mathbf{P}_a = \left(\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}\right)^{-1}$

- When prior is non-Gaussian, $\mathbf{P}_\mathrm{a}$ becomes a function of the observation value.
- The realisation of observation error which results in the greatest analysis error variance also results in the greatest analysis sensitivity.

PART I: ***non-Gaussian prior***

- 1D example:
  - prior is given by a 2 component Gaussian mixture with identical variances
    - $p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
    - 4 parameters: $\sigma, w_1 (w_2 = 1 - w_1), \mu_1, \mu_2$.



w=0.2 μ1=-0.5 μ2=0.5
w=0.2 μ1=-1.5 μ2=1.5
w=0.2 μ1=-2.5 μ2=2.5
w=0.2 μ1=-3.5 μ2=3.5
w=0.4 μ1=-0.5 μ2=0.5
w=0.4 μ1=-1.5 μ2=1.5
w=0.4 μ1=-2.5 μ2=2.5
w=0.4 μ1=-3.5 μ2=3.5
w=0.6 μ1=-0.5 μ2=0.5
w=0.6 μ1=-1.5 μ2=1.5
w=0.6 μ1=-2.5 μ2=2.5
w=0.6 μ1=-3.5 μ2=3.5
w=0.8 μ1=-0.5 μ2=0.5
w=0.8 μ1=-1.5 μ2=1.5
w=0.8 μ1=-2.5 μ2=2.5
w=0.8 μ1=-3.5 μ2=3.5
p(x)
x

PART I: ***non-Gaussian prior***

- prior $p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
- Likelihood given by $N(\mu_y, k\sigma^2)$
- $S = \frac{1}{k+1} + \frac{kw(1-w)(\mu_1-\mu_2)^2 e^{-a_1-a_2}}{(1+k)^2\sigma^2(we^{-a_1}+(1-w)e^{-a_2})^2}$
  - Where $a_i = ((\mu_y - \mu_i)^2)/(2(1+k)\sigma^2)$

PART I: ***non-Gaussian prior***

- prior $p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
- Likelihood given by $N\left(\mu_y, k\sigma^2\right)$
- $S = \frac{1}{k+1} + \frac{kw(1-w)(\mu_1-\mu_2)^2 e^{-a_1-a_2}}{(1+k)^2\sigma^2(we^{-a_1}+(1-w)e^{-a_2})^2}$
  - Where $a_i = ((\mu_y - \mu_i)^2)/(2(1+k)\sigma^2)$
- S is bounded below by $\frac{1}{k+1}$ and has no upper bound.

PART I: ***non-Gaussian prior***

- prior $p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
- Likelihood given by $N(\mu_y, k\sigma^2)$
- $S = \frac{1}{k+1} + \frac{kw(1-w)(\mu_1-\mu_2)^2 e^{-a_1-a_2}}{(1+k)^2\sigma^2(we^{-a_1}+(1-w)e^{-a_2})^2}$
  - Where $a_i = ((\mu_y - \mu_i)^2)/(2(1+k)\sigma^2)$
- S is bounded below by $\frac{1}{k+1}$ and has no upper bound.
  - Therefore, because $S = {\sigma_a}^2/{\sigma_y}^2$, it is possible for ${\sigma_a}^2 > {\sigma_y}^2$ when the prior describes two highly probably but distinct regimes.

PART I: ***non-Gaussian prior***

- prior $p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
- Likelihood given by $N(\mu_y, k\sigma^2)$
- $S = \frac{1}{k+1} + \frac{kw(1-w)(\mu_1-\mu_2)^2 e^{-a_1-a_2}}{(1+k)^2\sigma^2(we^{-a_1}+(1-w)e^{-a_2})^2}$
  - Where $a_i = ((\mu_y - \mu_i)^2)/(2(1+k)\sigma^2)$
- S is bounded below by $\frac{1}{k+1}$ and has no upper bound.
  - Therefore, because $S = {\sigma_a}^2/{\sigma_y}^2$, it is possible for ${\sigma_a}^2 > {\sigma_y}^2$ when the prior describes two highly probably but distinct regimes.
- S is at a maximum when ${\sigma_a}^2$ is at a maximum, i.e. the posterior is symmetric.

PART I: ***non-Gaussian prior***

- prior $p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
- Likelihood given by $N(\mu_y, k\sigma^2)$
- $S = \frac{1}{k+1} + \frac{kw(1-w)(\mu_1-\mu_2)^2 e^{-a_1-a_2}}{(1+k)^2\sigma^2(we^{-a_1}+(1-w)e^{-a_2})^2}$
  - Where $a_i = ((\mu_y - \mu_i)^2)/(2(1+k)\sigma^2)$
- S is bounded below by $\frac{1}{k+1}$ and has no upper bound.
  - Therefore, because $S = {\sigma_a}^2/{\sigma_y}^2$, it is possible for ${\sigma_a}^2 > {\sigma_y}^2$ when the prior describes two highly probably but distinct regimes.
- S is at a maximum when ${\sigma_a}^2$ is at a maximum, i.e. the posterior is symmetric.
- Away from this value of μ_y, S asymptotes to $\frac{1}{k+1}$.

PART I: ***non-Gaussian prior***

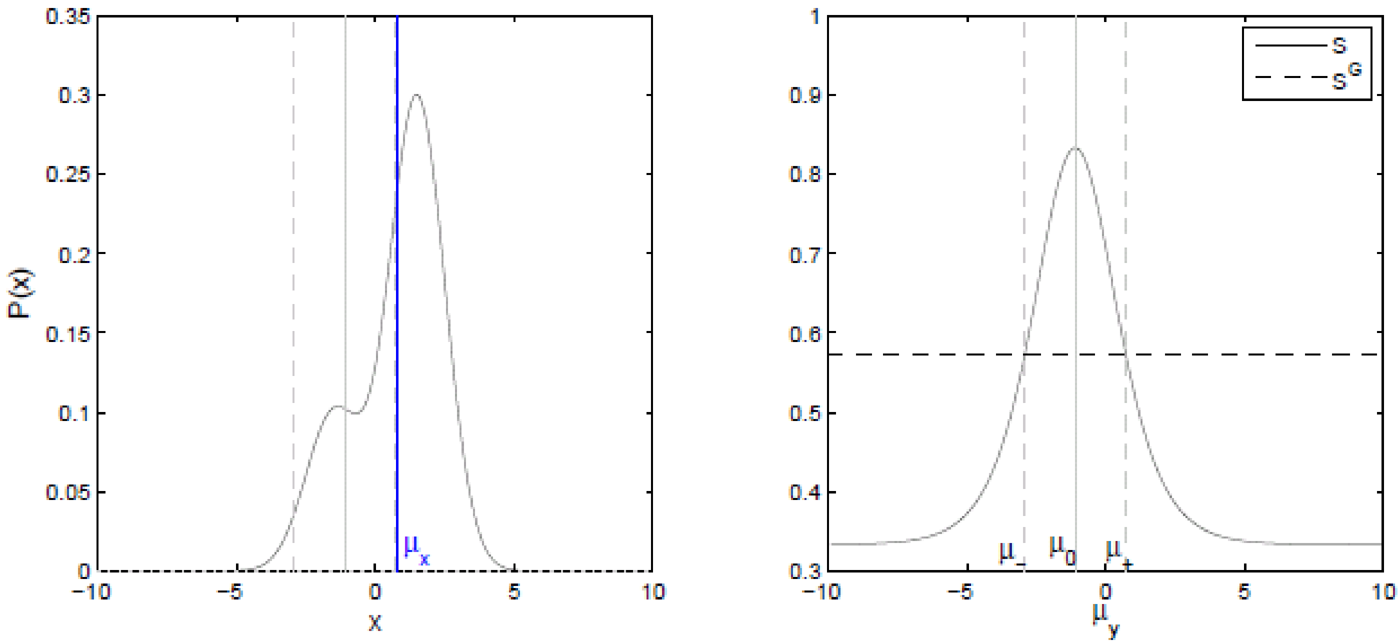


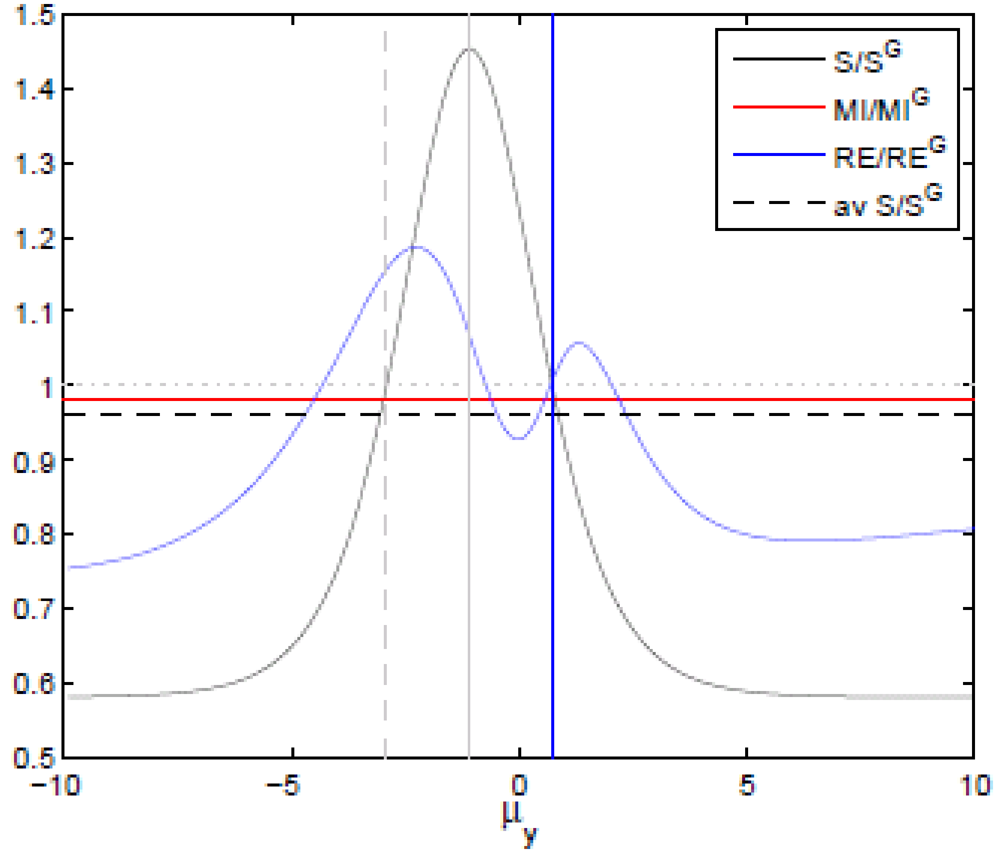
Figure 1: Left: The prior distribution. The vertical blue line shows the prior mean, μ_x. Right: $\frac{\partial \mu_a}{\partial \mu_y}$ (solid) and the Gaussian approximation (dashed) for $k = 2$, $\sigma^2 = 1$, $w = 0.25$, $\mu_1 = -1.5$, $\mu_2 = 1.5$. μ_-, μ_0 and μ_+ explained within the text.

Fowler and van Leeuwen (2012) submitted to Tellus A.

PART I: ***non-Gaussian prior***

- Can compare the sensitivity, in this case, to mutual information and relative entropy.

In this figure all the measures are normalised by their Gaussian approximations.



PART I: ***non-Gaussian prior***

- Application to the Lorenz 63 system using the particle filter (PF).

$$\frac{d\chi_1}{dt} = \sigma(\chi_2 - \chi_1)$$

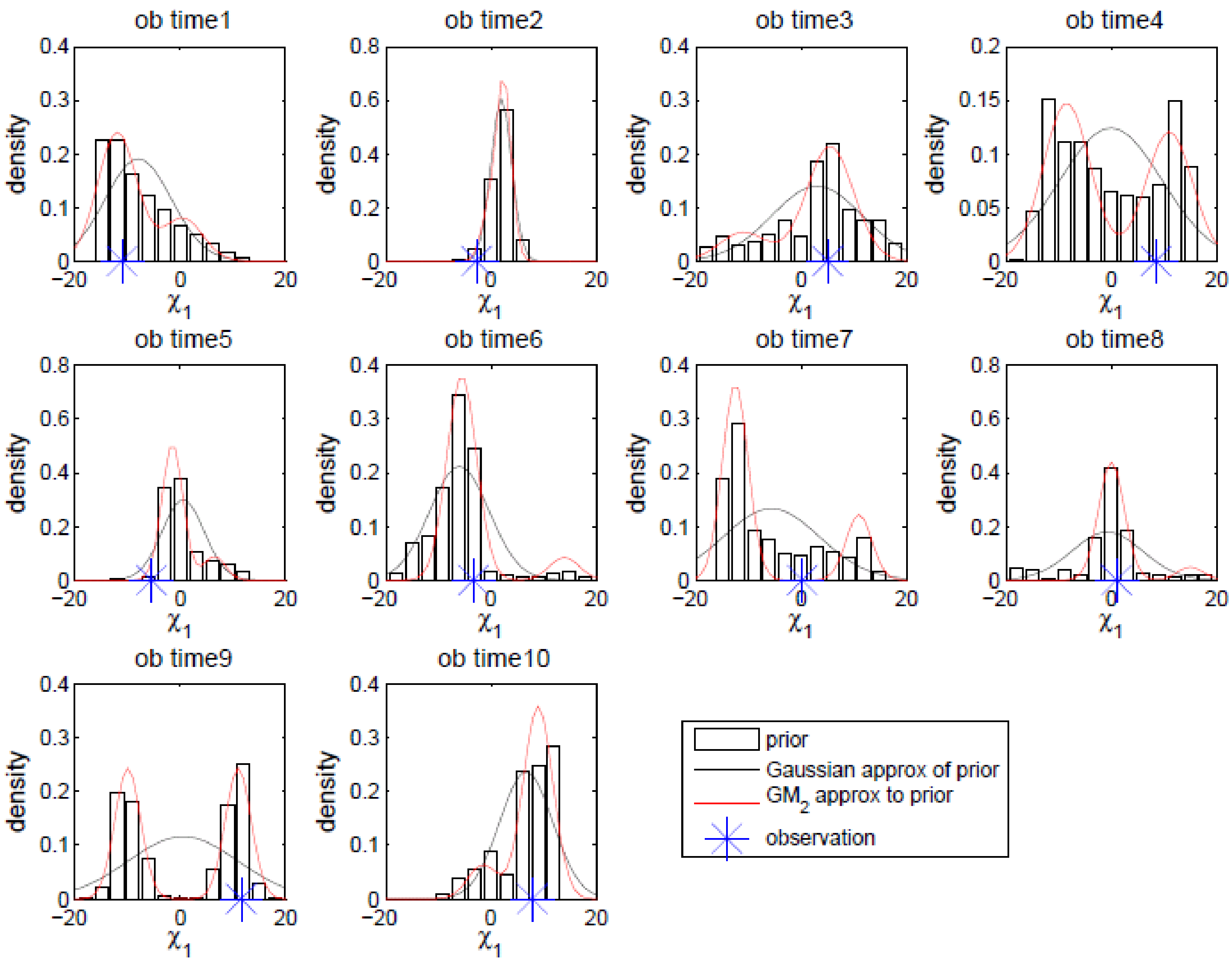
$$\frac{d\chi_2}{dt} = -\chi_1\chi_3 + \rho\chi_1 - \chi_2$$

$$\frac{d\chi_3}{dt} = \chi_1\chi_2 - \beta\chi_3.$$

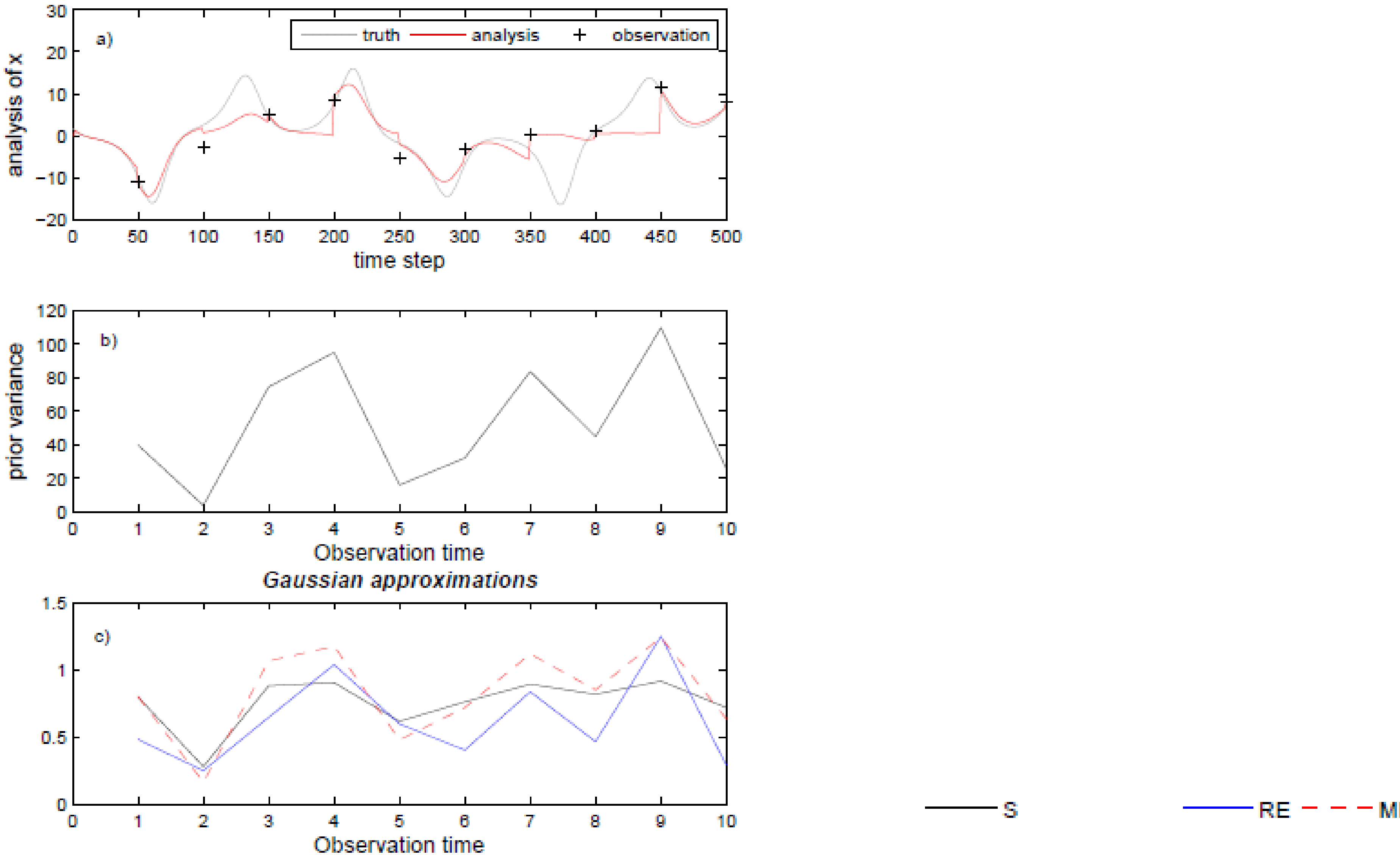
- $\sigma = 10, \rho = 28, \beta = \frac{8}{3}$.
- Represent a prior Gaussian distribution at the initial time by 1000 particles to avoid sampling issues.
- Allow the model to evolve each of the particles forward in time until an observation is available. Then change the weights of the particles to take into account the observed value.
- The timestep is 0.01.
- In this experiment I am only observing χ_1 at every 50 timesteps. Assume the observation error is Gaussian with an error variance of 10.

PART I: ***non-Gaussian prior***

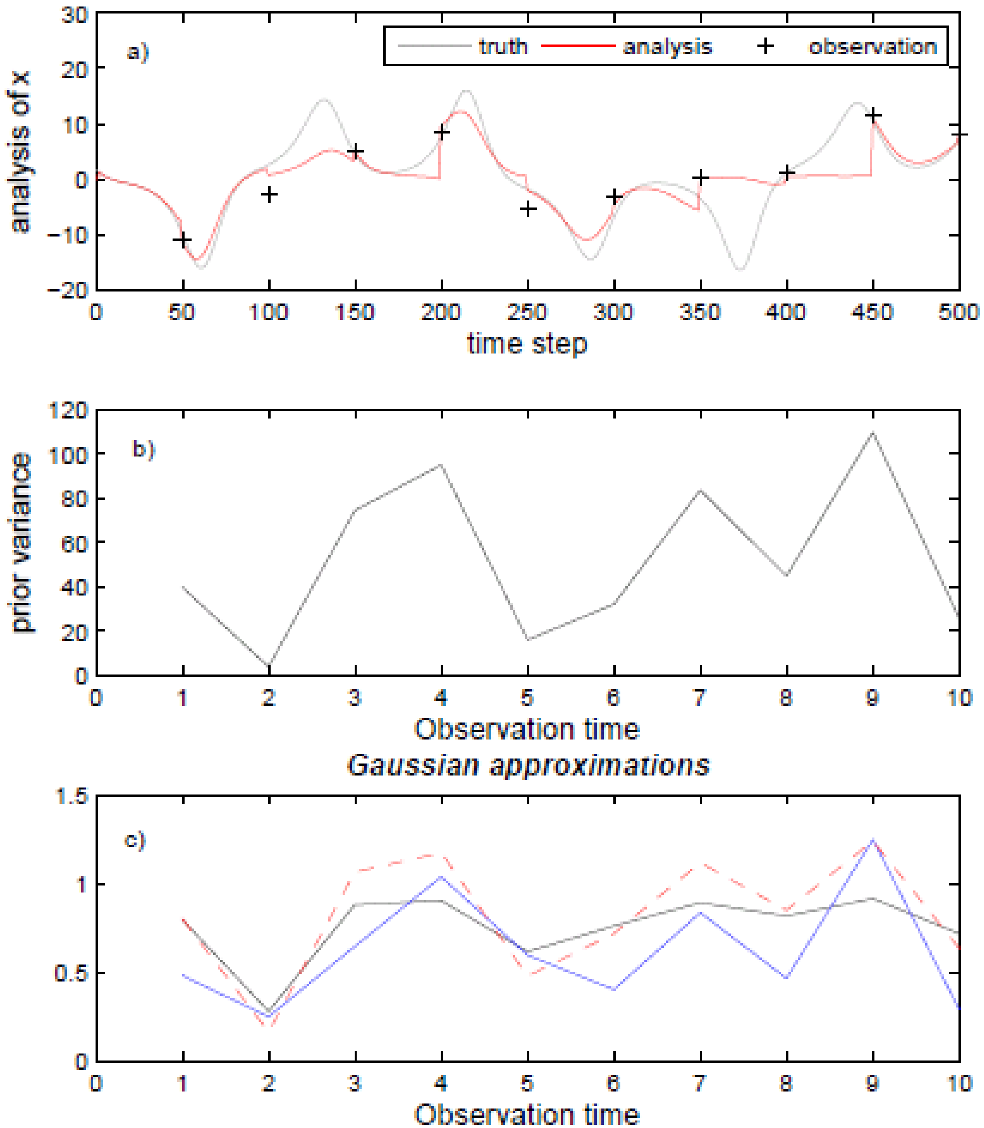
The prior distribution of χ_1 at the first 10 observation times

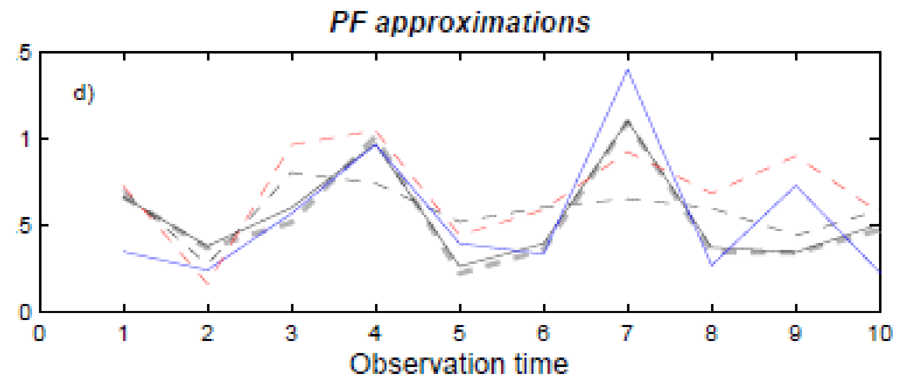


PART I: ***non-Gaussian prior***

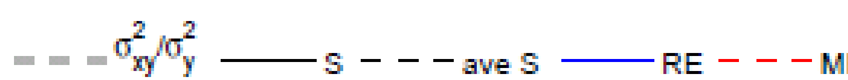


PART I: ***non-Gaussian prior***



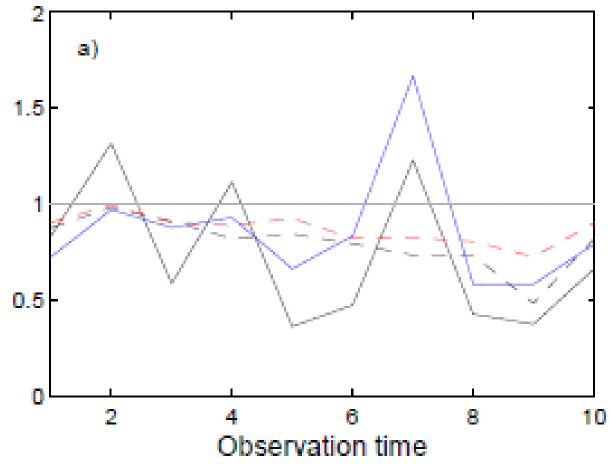


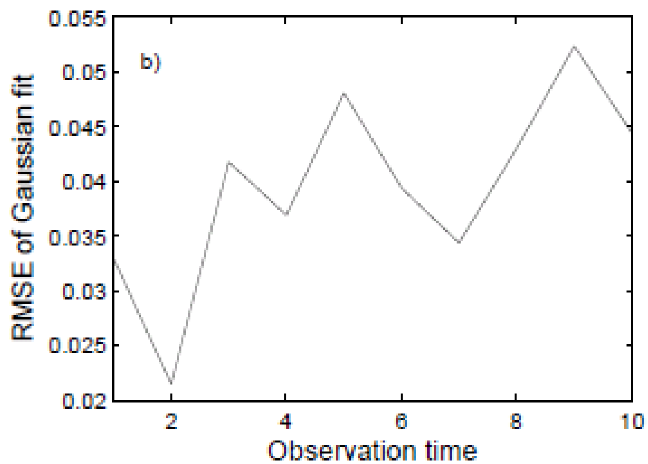
When the full prior is used to calculate observation impact there is no longer a close agreement between the different measures and the prior variance.



PART I: ***non-Gaussian prior***

Top: The ratio of the PF approx to the Gaussian approx for the 4 measures of observation impact

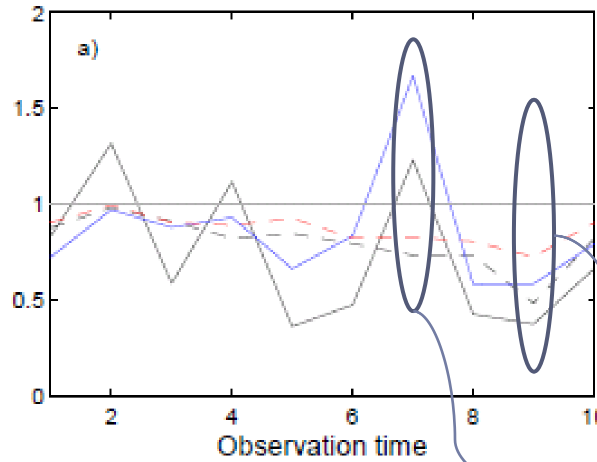


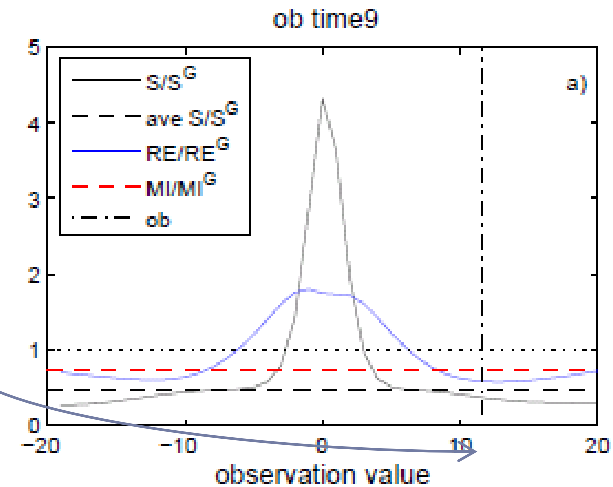


S ave S RE MI

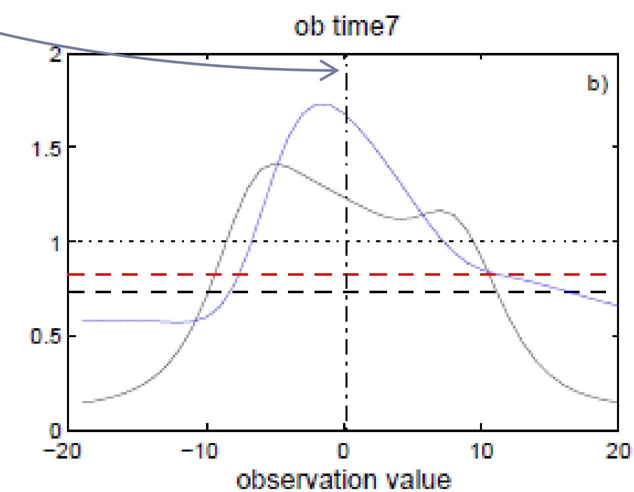
PART I: ***non-Gaussian prior***

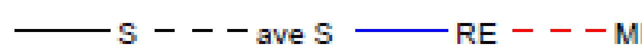
Instead of looking at one realisation of the observation error can plot the measures as a function of observation value for a particular observation time.





The vertical dash dot line gives the realisation of the observation used by the particle filter.





PART I: ***non-Gaussian prior***

- CONCLUSIONS:
  - For any prior the sensitivity of the analysis to the observations can be shown to equal $\mathbf{HP}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$ when the likelihood is Gaussian.
  - The analysis error covariances and hence the sensitivity of the analysis can be a strong function of observation value.
    - Therefore the error in approximating the sensitivity assuming Gaussian statistics is also a strong function of the observation value.
  - The error in approximating relative entropy with the assumption of Gaussian statistics is also a strong function of the observation value.
    - But a different function to the error in the sensitivity!
  - Mutual information is independent of the value of the observation. The Gaussian approximation results in a relatively small overestimate of observation impact.

Outline of talk

- Observation impact in Gaussian data assimilation.
  - Introduction to different measures
- The influence of a non-Gaussian statistics on observation impact.
  - PART I: the non-Gaussian prior
  - PART II: the non-Gaussian likelihood
- Future work

PART II: ***Non-Gaussian likelihood***

- When the prior is Gaussian and the likelihood function is arbitrary, it is possible to derive the sensitivity of the mean of the posterior, $\boldsymbol{\mu}_a$, to the mean of likelihood, $\boldsymbol{\mu}_y$, analytically.
  - $\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \frac{\int \mathrm{H}x p(x) \frac{\partial p(y|x)}{\partial \mu_y} dx}{\int p(x) p(\mathbf{y}|\mathbf{x}) dx} - \mathbf{H}\boldsymbol{\mu}_\mathbf{a} \frac{\int p(\mathbf{x}) \frac{\partial p(y|x)}{\partial \boldsymbol{\mu}_\mathbf{y}} \mathbf{dx}}{\int p(\mathbf{x}) p(\mathbf{y}|\mathbf{x}) \mathbf{dx}}.$
  - In this case $\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathrm{y}}$ is unknown.
  - Use integration by parts and the fact that

$$p(x, \mu + \Delta) = p(x - \Delta, \mu)$$

PART II: ***Non-Gaussian likelihood***

- When the prior is Gaussian and the likelihood function is arbitrary, it is possible to derive the sensitivity of the mean of the posterior, $\boldsymbol{\mu}_a$, to the mean of likelihood, $\boldsymbol{\mu}_y$, analytically.
  - $\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \frac{\int \mathrm{H}\mathrm{x}\mathrm{p}(\mathrm{x})\frac{\partial \mathrm{p}(\mathrm{y}|\mathrm{x})}{\partial \boldsymbol{\mu}_\mathrm{y}}\mathrm{dx}}{\int \mathrm{p}(\mathrm{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathrm{dx}} - \mathbf{H}\boldsymbol{\mu}_\mathbf{a}\frac{\int \mathrm{p}(\mathbf{x})\frac{\partial \mathrm{p}(\mathrm{y}|\mathrm{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathbf{dx}}{\int \mathrm{p}(\mathbf{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathbf{dx}}.$
  - In this case $\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathrm{y}}$ is unknown.
  - Use integration by parts and the fact that

$$p(x, \mu + \Delta) = p(x - \Delta, \mu)$$

- $\frac{\partial \boldsymbol{H\mu}_\mathrm{a}}{\partial \boldsymbol{\mu}_\mathrm{y}} = \mathbf{I}_{\boldsymbol{p}} - \mathbf{HP}_{\boldsymbol{a}}\mathbf{B}^{-1}\mathbf{H}^\mathrm{T}\left(\mathbf{HH}^\mathrm{T}\right)^{-1}$

PART II: ***Non-Gaussian likelihood***

- When the prior is Gaussian and the likelihood function is arbitrary, it is possible to derive the sensitivity of the mean of the posterior, $\boldsymbol{\mu}_a$, to the mean of likelihood, $\boldsymbol{\mu}_y$, analytically.
  - $\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \frac{\int \mathrm{Hx}\,\mathrm{p(x)} \frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathrm{y}} \mathrm{dx}}{\int \mathrm{p(x)p}(\mathbf{y}|\mathbf{x})\mathrm{dx}} - \mathbf{H}\boldsymbol{\mu}_\mathbf{a} \frac{\int \mathrm{p}(\mathbf{x}) \frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathbf{y}} \mathbf{dx}}{\int \mathrm{p}(\mathbf{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathbf{dx}}.$
  - In this case $\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathrm{y}}$ is unknown.
  - Use integration by parts and the fact that
$$p(x, \mu + \Delta) = p(x - \Delta, \mu)$$
- $\frac{\partial \boldsymbol{H\mu}_\mathrm{a}}{\partial \boldsymbol{\mu}_\mathrm{y}} = \mathbf{I}_{\boldsymbol{p}} - \mathbf{H}\mathbf{P}_{\boldsymbol{a}}\mathbf{B}^{-1}\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{H}^\mathrm{T}\right)^{-1}$
  - This is equivalent to $\mathbf{H}\mathbf{P}_{\boldsymbol{a}}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}$ when the statistics are Gaussian.

PART II: ***Non-Gaussian likelihood***

- When the prior is Gaussian and the likelihood function is arbitrary, it is possible to derive the sensitivity of the mean of the posterior, $\boldsymbol{\mu}_a$, to the mean of likelihood, $\boldsymbol{\mu}_y$, analytically.
  - $\frac{\partial \mathbf{H}\boldsymbol{\mu}_\mathbf{a}}{\partial \boldsymbol{\mu}_\mathbf{y}} = \frac{\int \mathrm{H}\mathrm{x}\mathrm{p}(\mathrm{x})\frac{\partial \mathrm{p}(\mathrm{y}|\mathrm{x})}{\partial \boldsymbol{\mu}_\mathrm{y}}\mathrm{dx}}{\int \mathrm{p}(\mathrm{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathrm{dx}} - \mathbf{H}\boldsymbol{\mu}_\mathbf{a}\frac{\int \mathrm{p}(\mathbf{x})\frac{\partial \mathrm{p}(\mathrm{y}|\mathrm{x})}{\partial \boldsymbol{\mu}_\mathbf{y}}\mathbf{dx}}{\int \mathrm{p}(\mathbf{x})\mathrm{p}(\mathbf{y}|\mathbf{x})\mathbf{dx}}.$
  - In this case $\frac{\partial \mathrm{p}(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_\mathrm{y}}$ is unknown.
  - Use integration by parts and the fact that
  $$p(x, \mu + \Delta) = p(x - \Delta, \mu)$$
- $\frac{\partial \boldsymbol{H\mu}_\mathrm{a}}{\partial \boldsymbol{\mu}_\mathrm{y}} = \mathbf{I}_{\boldsymbol{p}} - \mathbf{HP}_{\boldsymbol{a}}\mathbf{B}^{-1}\mathbf{H}^\mathrm{T}\left(\mathbf{HH}^\mathrm{T}\right)^{-1}$
  - This is equivalent to $\mathbf{HP}_a\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}$ when the statistics are Gaussian.
  - The sensitivity is now greatest when $\mathbf{P}_a$ is at a minimum.

PART II: ***non-Gaussian likelihood***

1D example:

- Likelihood $p(y|x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
  - $\mu_y = w\mu_1 + (1-w)\mu_2$
- Prior given by $N(\mu_x, \kappa\sigma^2)$

PART II: ***non-Gaussian likelihood***

1D example:

- Likelihood $p(y|x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
  - $\mu_y = w\mu_1 + (1-w)\mu_2$
- Prior given by $N(\mu_x, \kappa\sigma^2)$
- $S = \frac{\kappa}{\kappa+1} - \frac{\kappa w(1-w)(\mu_1-\mu_2)^2 e^{-\alpha_1-\alpha_2}}{(1+\kappa)^2\sigma^2(we^{-\alpha_1}+(1-w)e^{-\alpha_2})^2}$
  - Where $\alpha_i = ((\mu_x - \mu_i)^2)/(2(1+\kappa)\sigma^2)$

PART II: ***non-Gaussian likelihood***

1D example:

- Likelihood $p(y|x) = (2\pi\sigma^2)^{-\frac{1}{2}} \sum_{i=1}^{2} w_i \exp\left\{-\frac{1}{2\sigma^2}(x-\mu_i)^2\right\}$
  - $\mu_y = w\mu_1 + (1-w)\mu_2$
- Prior given by $N(\mu_x, \kappa\sigma^2)$
- $S = \frac{\kappa}{\kappa+1} - \frac{\kappa w(1-w)(\mu_1-\mu_2)^2 e^{-\alpha_1-\alpha_2}}{(1+\kappa)^2\sigma^2(we^{-\alpha_1}+(1-w)e^{-\alpha_2})^2}$
  - Where $\alpha_i = ((\mu_x - \mu_i)^2)/(2(1+\kappa)\sigma^2)$

RECALL in non-gaussian prior case:

$$S = \frac{1}{k+1} + \frac{kw(1-w)(\mu_1-\mu_2)^2 e^{-a_1-a_2}}{(1+k)^2\sigma^2(we^{-a_1}+(1-w)e^{-a_2})^2}$$

Where $a_i = ((\mu_y - \mu_i)^2)/(2(1+k)\sigma^2)$

PART II: ***non-Gaussian likelihood***

- $S = \frac{\kappa}{\kappa+1} - \frac{\kappa w(1-w)(\mu_1-\mu_2)^2 e^{-\alpha_1-\alpha_2}}{(1+\kappa)^2\sigma^2(we^{-\alpha_1}+(1-w)e^{-\alpha_2})^2}$
  - Where $\alpha_i = ((\mu_x - \mu_i)^2)/(2(1+\kappa)\sigma^2)$

PART II: ***non-Gaussian likelihood***

- $S = \frac{\kappa}{\kappa+1} - \frac{\kappa w(1-w)(\mu_1-\mu_2)^2 e^{-\alpha_1-\alpha_2}}{(1+\kappa)^2\sigma^2(we^{-\alpha_1}+(1-w)e^{-\alpha_2})^2}$
  - Where $\alpha_i = ((\mu_x - \mu_i)^2)/(2(1+\kappa)\sigma^2)$
- S is bounded above by $\frac{\kappa}{\kappa+1}$ and has no lower bound.

PART II: ***non-Gaussian likelihood***

- $S = \frac{\kappa}{\kappa+1} - \frac{\kappa w(1-w)(\mu_1-\mu_2)^2 e^{-\alpha_1-\alpha_2}}{(1+\kappa)^2\sigma^2(we^{-\alpha_1}+(1-w)e^{-\alpha_2})^2}$
  - Where $\alpha_i = ((\mu_x - \mu_i)^2)/(2(1+\kappa)\sigma^2)$
- S is bounded above by $\frac{\kappa}{\kappa+1}$ and has no lower bound.
  - Therefore, because $S = 1 - {\sigma_a}^2/{\sigma_x}^2$, it is possible for ${\sigma_a}^2 > {\sigma_x}^2$ when the likelihood describes two highly probably but distinct regimes.

PART II: ***non-Gaussian likelihood***

- $S = \frac{\kappa}{\kappa+1} - \frac{\kappa w(1-w)(\mu_1-\mu_2)^2 e^{-\alpha_1-\alpha_2}}{(1+\kappa)^2\sigma^2(we^{-\alpha_1}+(1-w)e^{-\alpha_2})^2}$
  - Where $\alpha_i = ((\mu_x - \mu_i)^2)/(2(1+\kappa)\sigma^2)$
- S is bounded above by $\frac{\kappa}{\kappa+1}$ and has no lower bound.
  - Therefore, because $S = 1 - {\sigma_a}^2/{\sigma_x}^2$, it is possible for ${\sigma_a}^2 > {\sigma_x}^2$ when the likelihood describes two highly probably but distinct regimes.
- S is at a minimum when ${\sigma_a}^2$ is at a maximum, i.e. the posterior is symmetric.

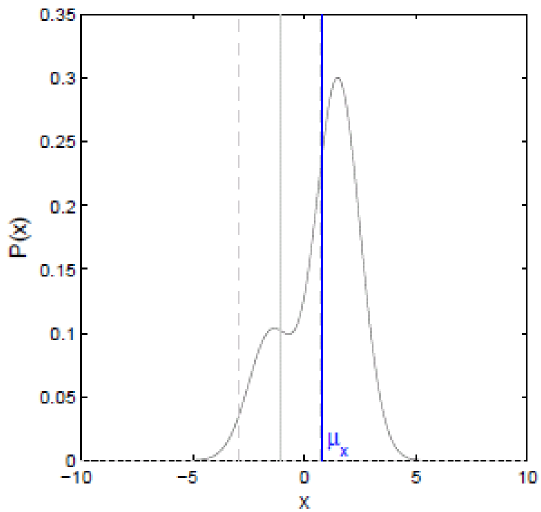
PART II: ***non-Gaussian likelihood***

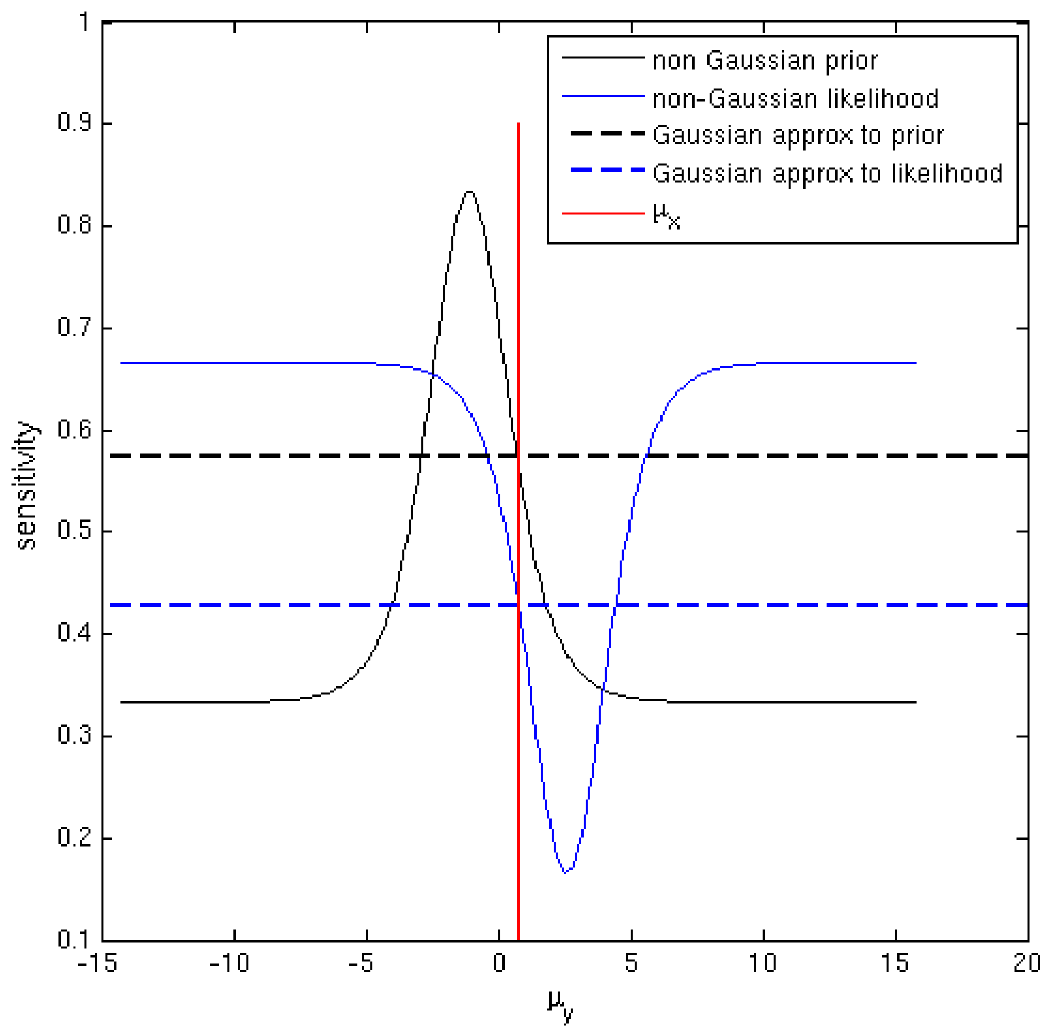
- $S = \frac{\kappa}{\kappa+1} - \frac{\kappa w(1-w)(\mu_1-\mu_2)^2 e^{-\alpha_1-\alpha_2}}{(1+\kappa)^2\sigma^2(we^{-\alpha_1}+(1-w)e^{-\alpha_2})^2}$
  - Where $\alpha_i = ((\mu_x - \mu_i)^2)/(2(1+\kappa)\sigma^2)$
- S is bounded above by $\frac{\kappa}{\kappa+1}$ and has no lower bound.
  - Therefore, because $S = 1 - {\sigma_a}^2/{\sigma_x}^2$, it is possible for ${\sigma_a}^2 > {\sigma_x}^2$ when the likelihood describes two highly probably but distinct regimes.
- S is at a minimum when ${\sigma_a}^2$ is at a maximum, i.e. the posterior is symmetric.
- Away from this value of μ_y, S asymptotes to $\frac{\kappa}{\kappa+1}$.

PART II: ***non-Gaussian likelihood***

- Comparison to non-Gaussian prior case

  - k=2 as in previous figure, $\kappa = 2$. In both cases $\mu_x = \frac{3}{4}$

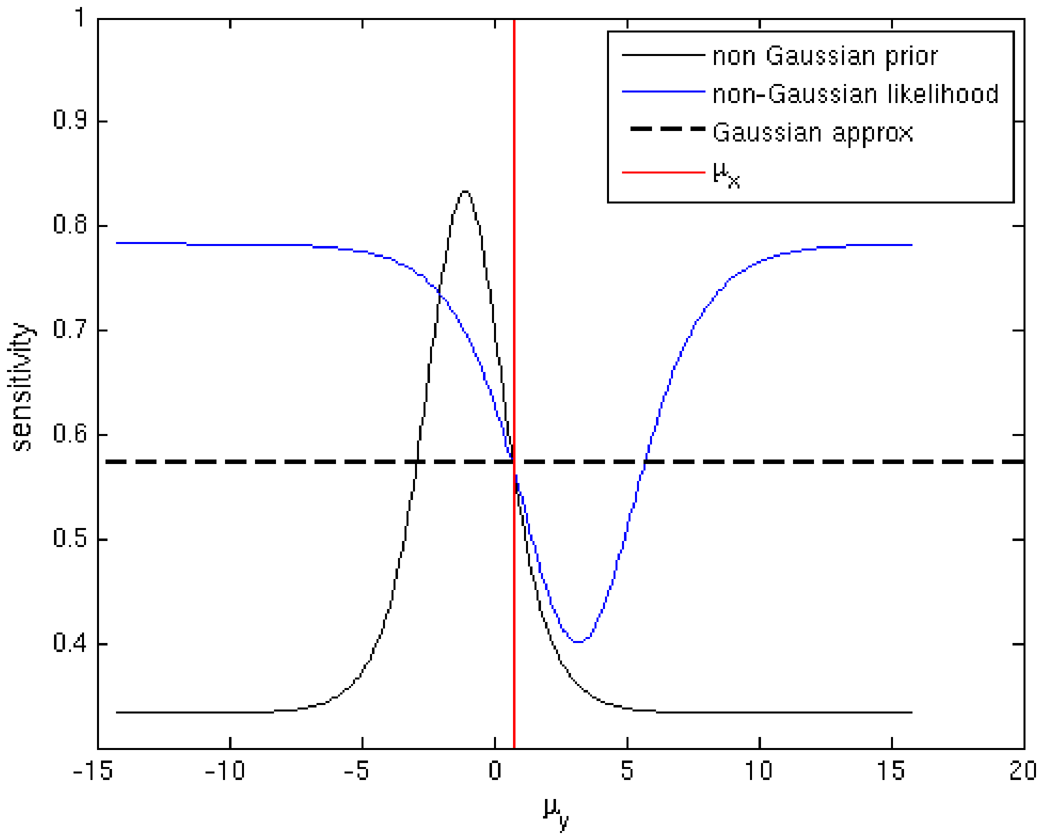




PART II: ***non-Gaussian likelihood***

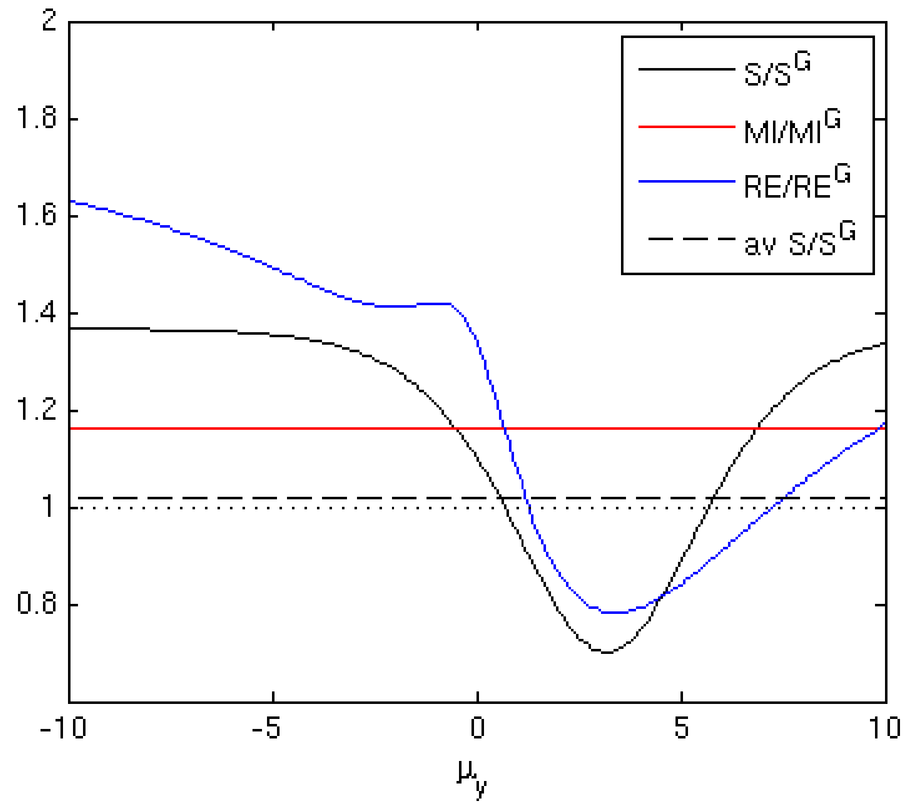
- Comparison to non-Gaussian prior case

  - k=2 as in previous figure, $\kappa = \frac{1849}{512}$, so that the Gaussian approximation to the sensitivity is the same in both cases.



PART II: ***non-Gaussian likelihood***

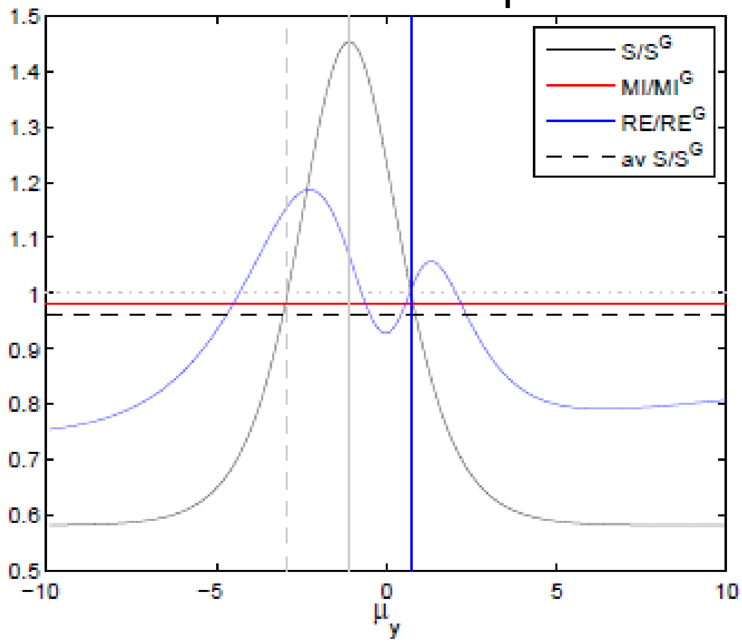
- Can compare the sensitivity, in this case, to mutual information and relative entropy.

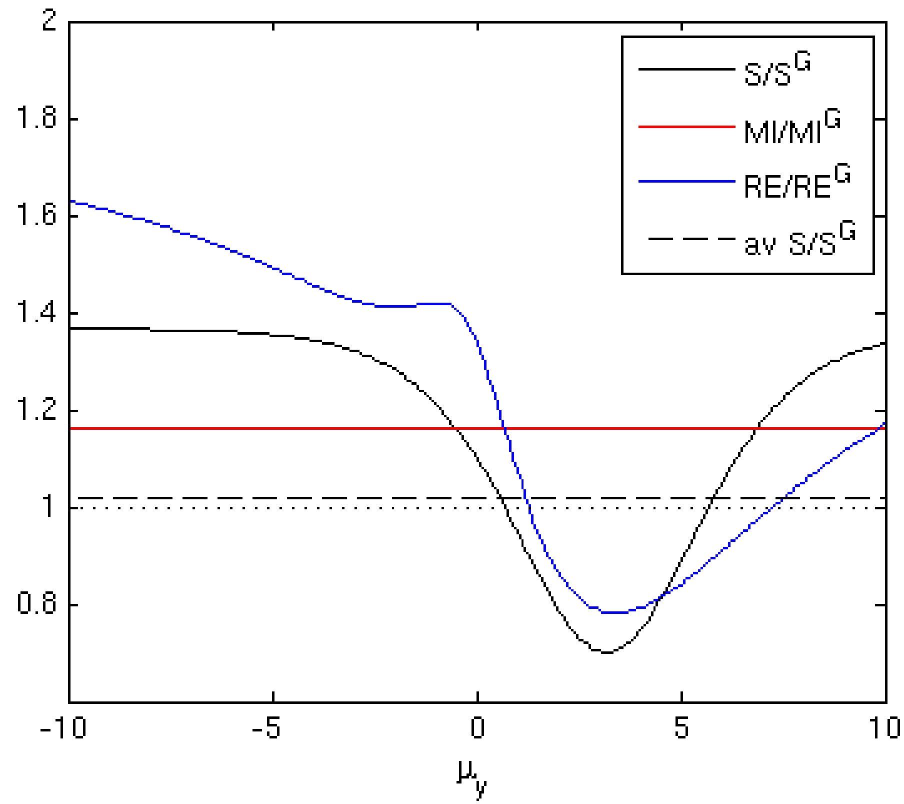


PART II: ***non-Gaussian likelihood***

- Can compare the sensitivity, in this case, to mutual information and relative entropy.

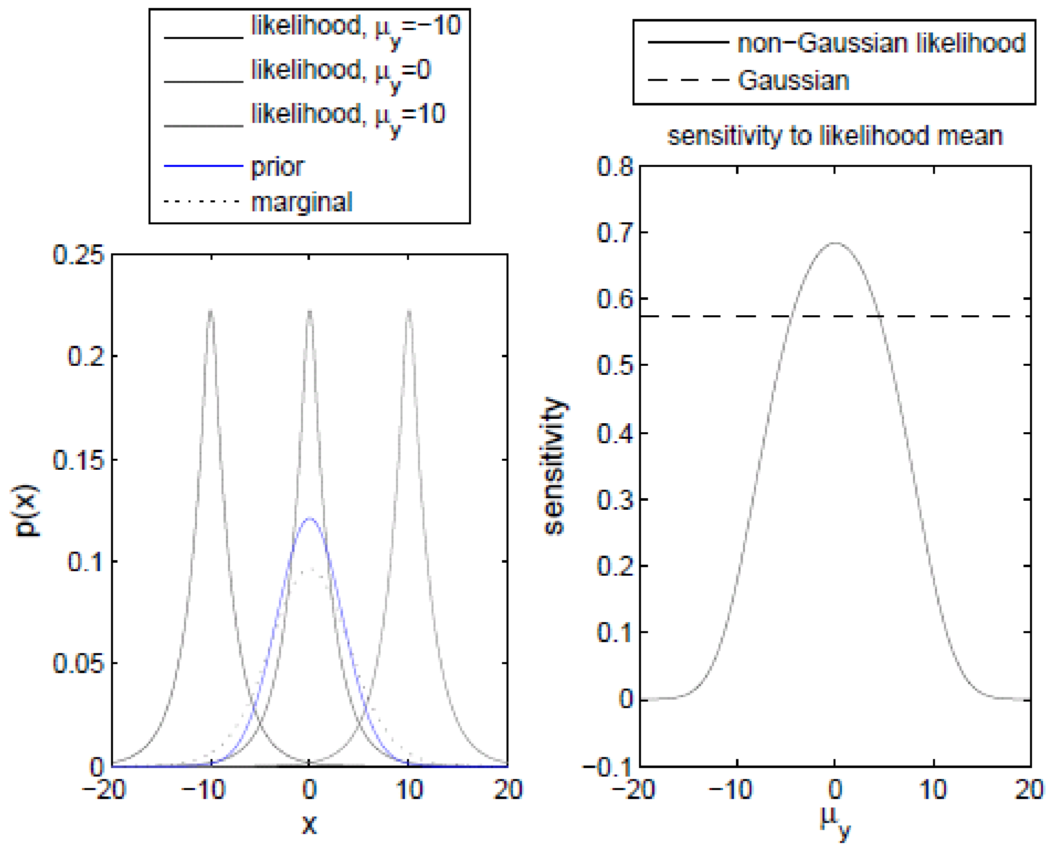
RECALL non-Gaussian prior case:





PART II: ***non-Gaussian likelihood***

- I have also looked at a Huber norm likelihood which has a very different structure to my simplified Gaussian mixture.



PART II: ***non-Gaussian likelihood***

- I have also looked at a Huber norm likelihood which has a very different structure to my simplified Gaussian mixture.

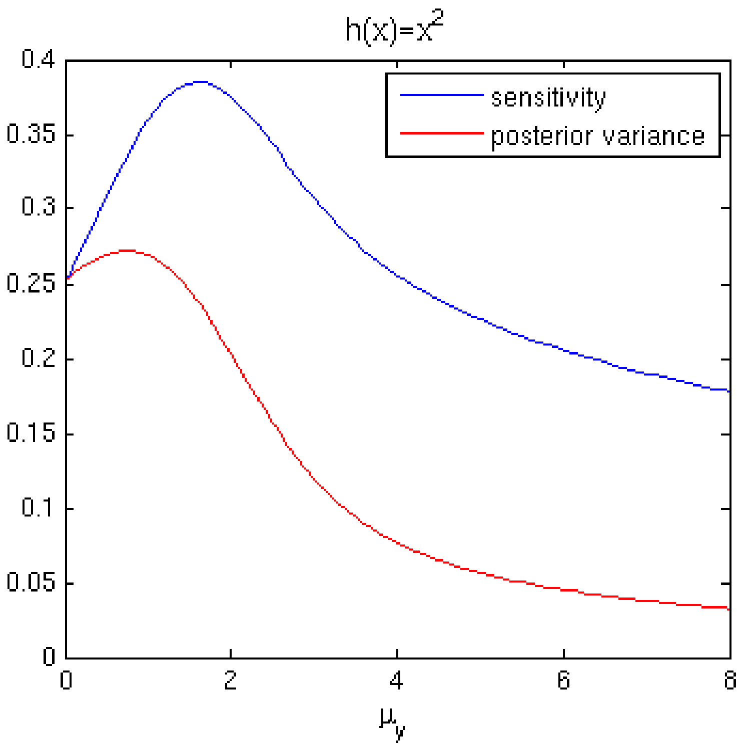
  - Similar results:
    - error in Gaussian approximation to sensitivity and relative entropy very variable with observation value. Variation in error of RE greater.
    - Gaussian estimate to average observation impact always underestimates.

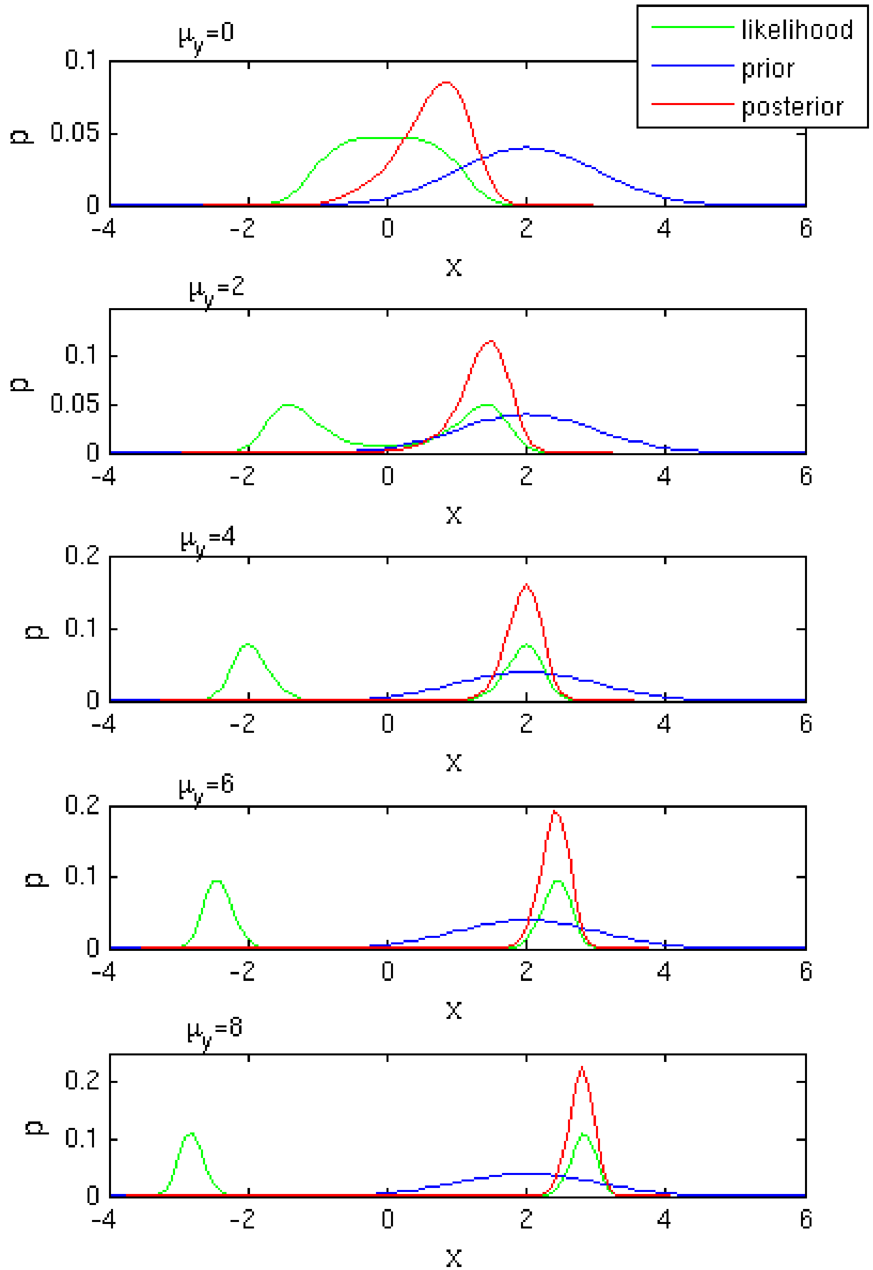
Future work

- I would like to look at the impact of a non-linear observation operator on observation impact.
  - Unlike the examples I have shown so far the shape of the non-Gaussian distribution is not fixed as the observation value changes.
  - The simple relationship between sensitivity and analysis error variance will no longer hold.
  - E.g. if prior and measurement error are Gaussian
    - $\frac{\partial \mu_a}{\partial \mu_y} = {\sigma_y}^{-2}\left[\int xh(x)p(x|y)dx \; - \mu_a \int xh(x)p(x|y)dx\right]$

Future work

e.g. $h(x) = x^2$





Future work

- I would also like to move to larger and more realistic systems with the aim of giving a full critique of the different measures of observation impact in non-Gaussian DA.

Some References

- Observation impact in Gaussian DA
  - Cardinali et al., 2004: Influence-matrix diagnostics of a data assimilation system, *Q. J. R. Met. Soc.*, **130**, 2767-2786.
  - Rodgers, 2000: Inverse methods for atmospheric sounding.
  - Xu, et al., 2009: Measuring information content from observations for data assimilation: connection between different measures and application to radar scan design. *Tellus*, **61A**, 144-153.
- Observation impact in non-Gaussian DA
  - Fowler and van Leeuwen, 2012: Measures of observation impact in non-Gaussian data assimilation. *Accepted by Tellus.*
  - Fowler and van Leeuwen, 2012: Measures of observation impact in data assimilation: the effect of a non-Gaussian likelihood. *In Preparation.*