U.S. NAVAL RESEARCH LABORATORY

Episodic, non-linear and non-Gaussian: ensemble data assimilation for bounded semi-positive definite variables like clouds

Improvements to the Gamma Inverse-Gamma (GIG) variation on the EnKF

Craig H. Bishop[1]

[1]Marine Meteorology Division, Naval Research Laboratory, Monterey, USA

Background: A typical EnKF serial observation assimilation scheme

for $j = 1 : p;$ % where p is the number of observations

Step 1: Do univariate Gaussian assimilation of y to obtain $y_{ji}^{a},\ i = 1, 2, ..., K$

Step 2: Find corresponding analysis ensemble for observations and model variables

$$y_{ki}^{a} = y_{ki}^{f} + \frac{\mathrm{covar}\left(y_k^f, y_j^f\right)}{\mathrm{var}\left(y_j^f\right)}\left(y_{ji}^{a} - y_{ji}^{f}\right),\ \text{for } k = 1, 2, ..., p;\ i = 1, 2, ..., K$$

$$x_{\mu i}^{a} = x_{\mu i}^{f} + \frac{\mathrm{covar}\left(x_\mu^f, y_j^f\right)}{\mathrm{var}\left(y_j^f\right)}\left(y_{ji}^{a} - y_{ji}^{f}\right),\ \text{for } \mu = 1, 2, ..., n;\ i = 1, 2, ..., K$$

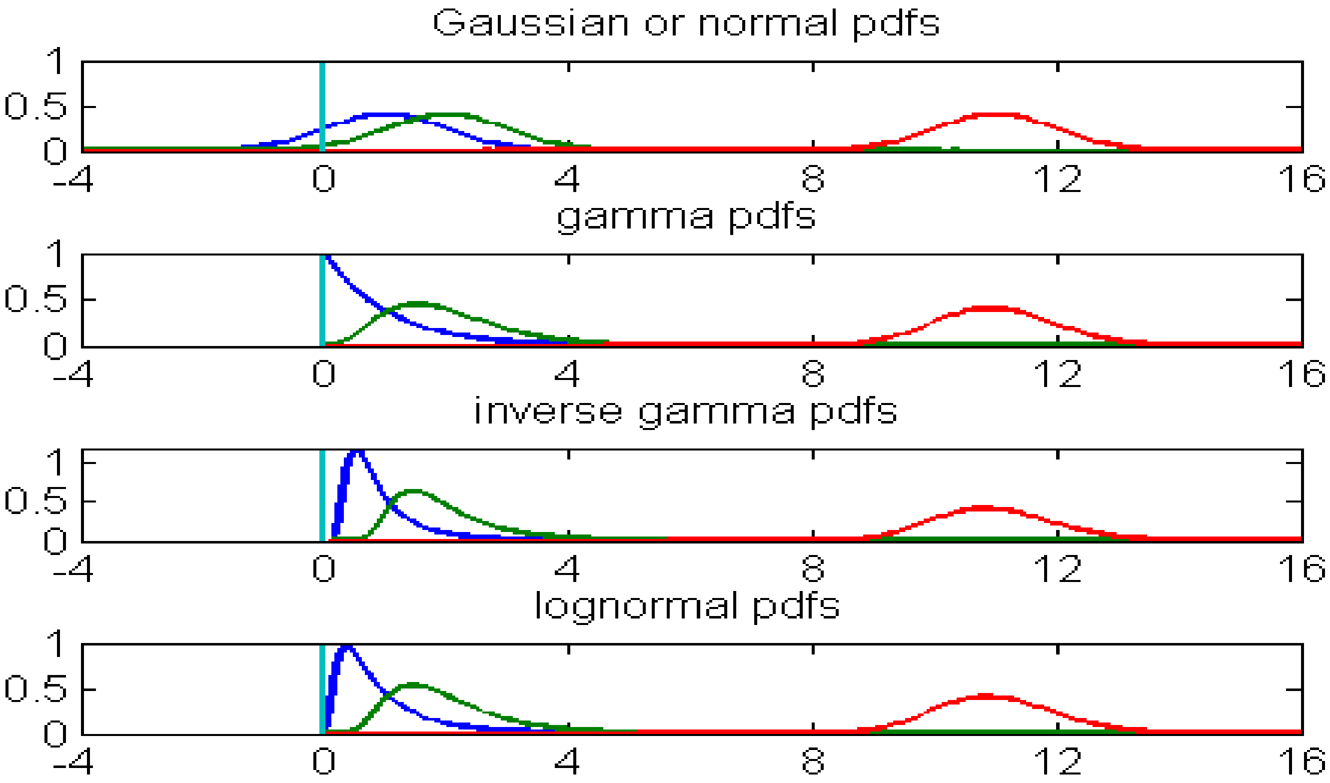
Step 3: Let the analysis ensemble be the prior ensemble for the next observation

$$y_{ki}^{f} = y_{ki}^{a},\ \text{for } k = 1, 2, ..., p;\ i = 1, 2, ..., K$$

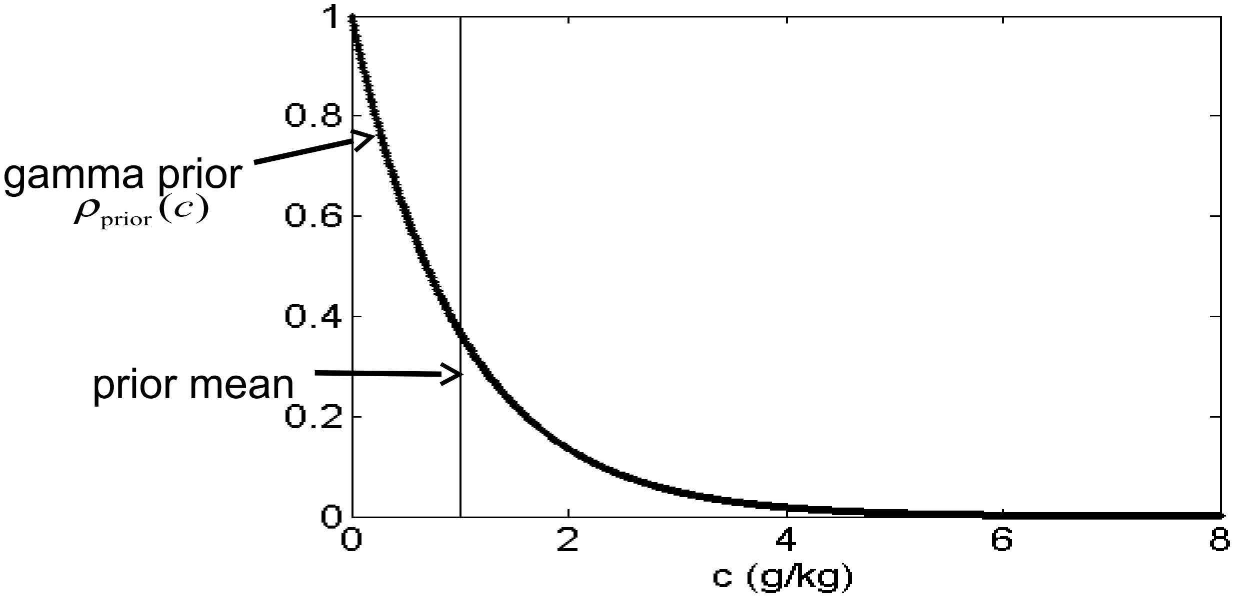
$$x_{\mu i}^{f} = x_{\mu i}^{a},\ \text{for } \mu = 1, 2, ..., n;\ i = 1, 2, ..., K$$

end

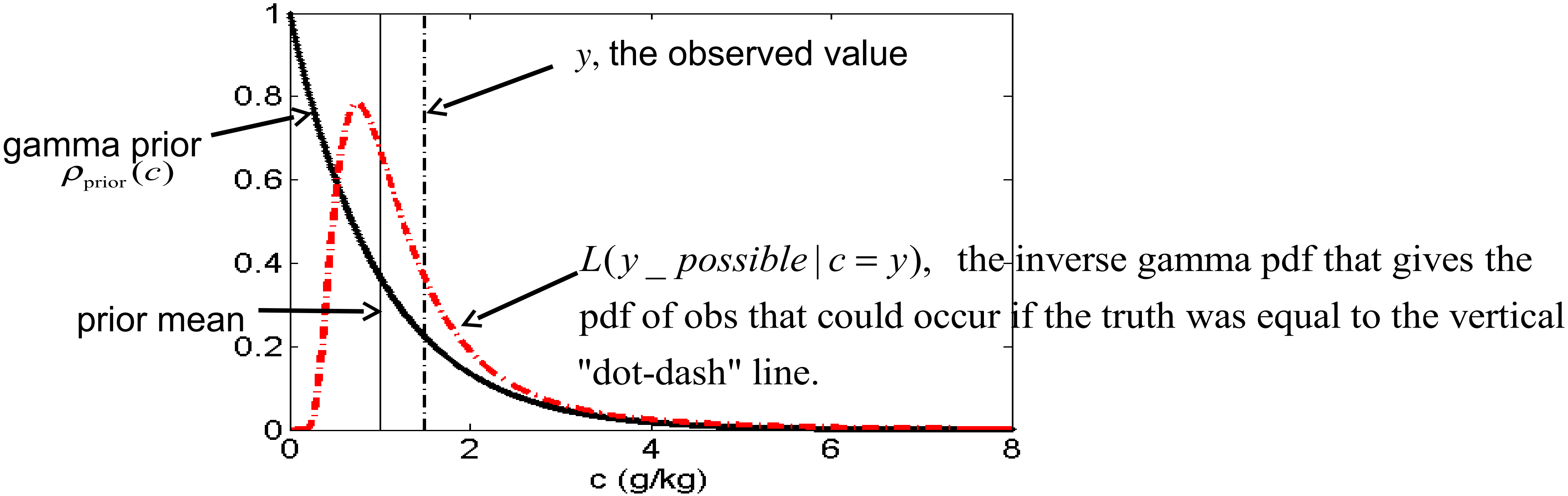
Background: Gaussian pdfs versus bounded pdfs



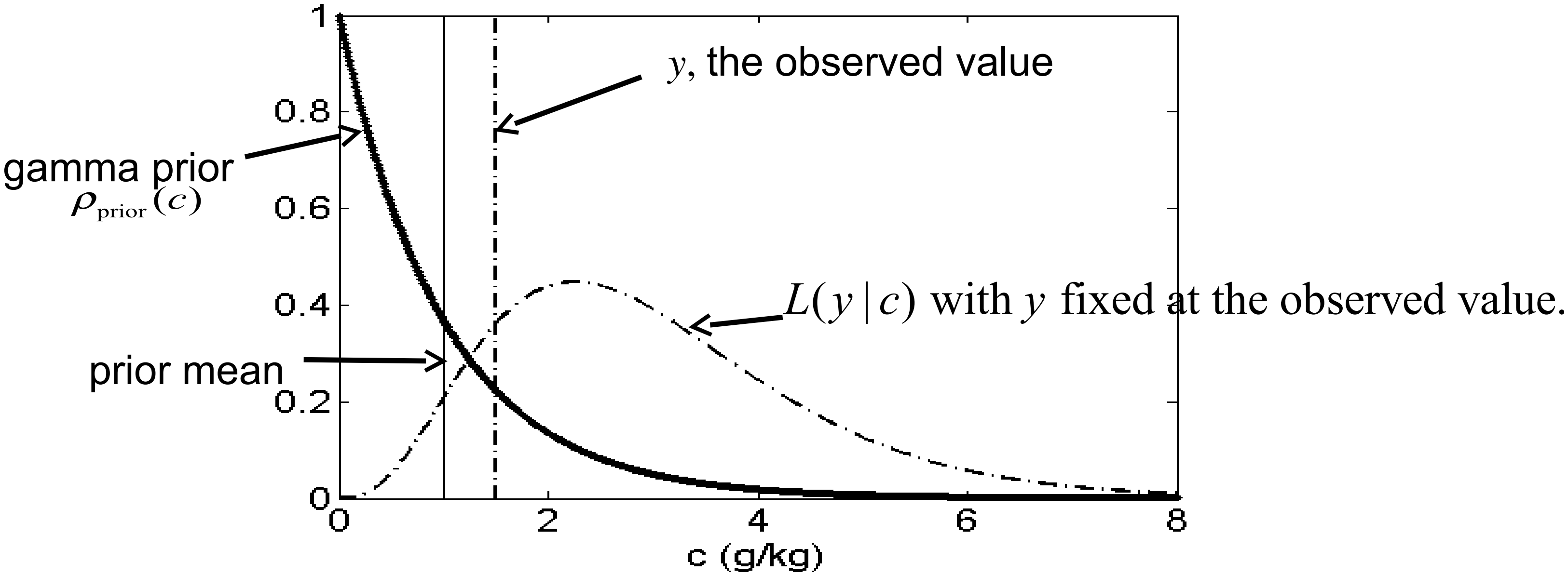
A prior Gamma pdf



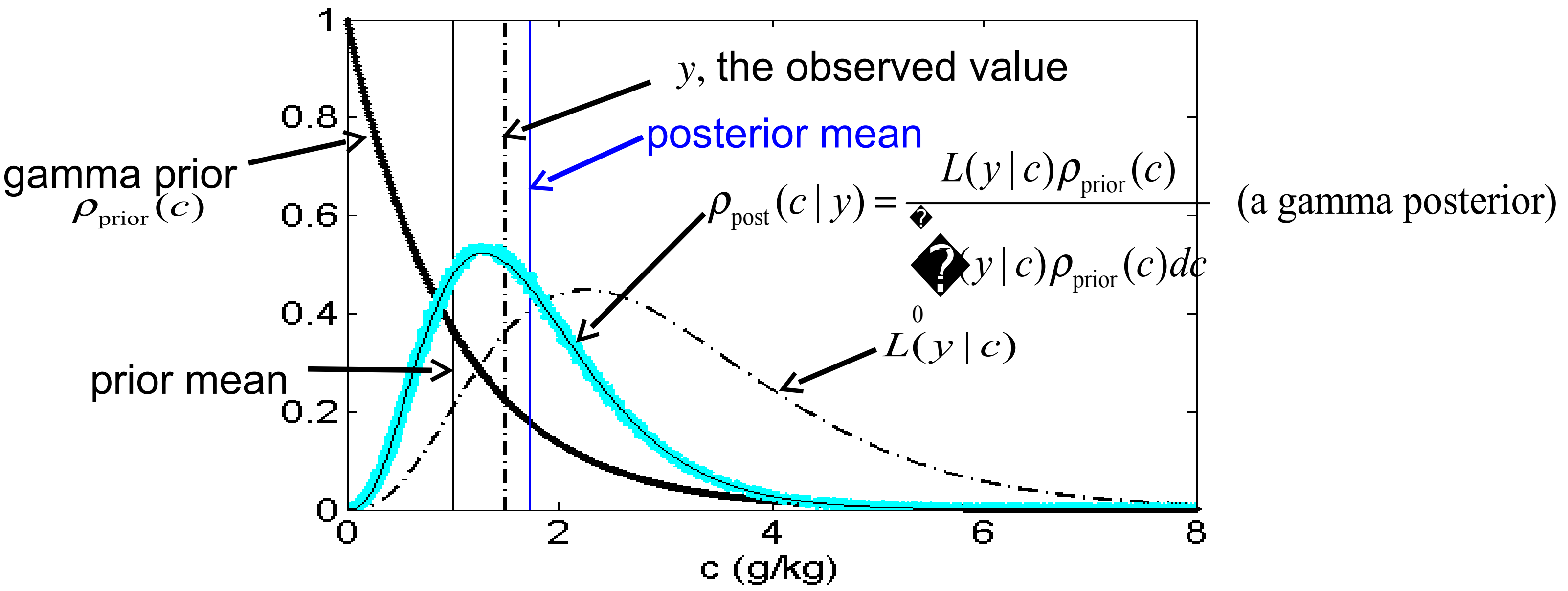
Inverse-Gamma pdf of obs given truth



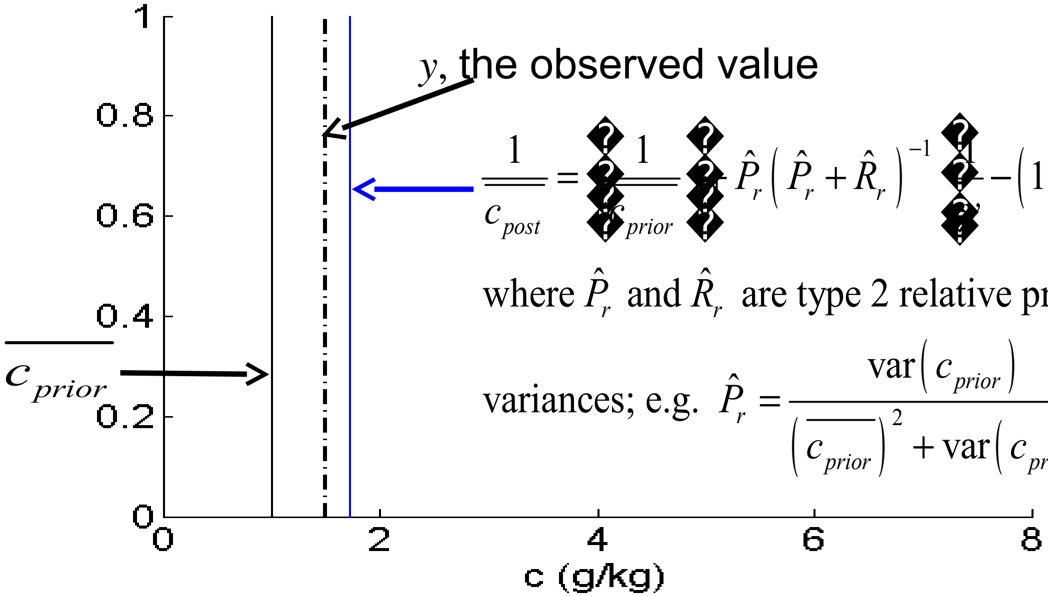
The likelihood function



The posterior pdf is then a gamma



Equation for posterior mean



$$\frac{1}{\overline{c}_{post}} = \left[\frac{1}{\overline{c}_{prior}}\right] + \hat{P}_r\left(\hat{P}_r+\hat{R}_r\right)^{-1}\left[\frac{1}{y} - \left(1+\hat{R}_r\right)\frac{1}{\overline{c}_{prior}}\right]$$

where $\hat{P}_r$ and $\hat{R}_r$ are type 2 relative prior and observation variances; e.g. $\hat{P}_r = \dfrac{\mathrm{var}\left(c_{prior}\right)}{\left(\overline{c_{prior}}\right)^2 + \mathrm{var}\left(c_{prior}\right)} = \dfrac{\sigma^2}{\mu^2+\sigma^2}$

Posterior mean equation has Kalman like gain but everything else is inverted !

Background: The GIGG-EnKF serial observation assimilation scheme with *linear* regression

for $j = 1 : p$;   % where p is the number of observations

Step 1: Decide whether forecast and observation uncertainty associated with y_j^o is best approximated by GIG-delta, GIG, IGG or Gaussian assumptions.

Step 2: if (GIG-delta) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$;

else if (GIG) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$; →

Note: GIG posterior mean eq *very* different to EnKF eq

$$\frac{1}{\overline{\overline{y^a}}} = \left(\frac{1}{\overline{\overline{y^f}}}\right) + \tilde{P}_r \left(\tilde{P}_r + \tilde{R}_r\right)^{-1} \left[\frac{1}{y} - \left(1 + \tilde{R}_r\right)\left(\frac{1}{\overline{\overline{y^f}}}\right)\right].$$

else if (IGG) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$;

else if (Gaussian) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$ (EAKF/EnSRF/EnKF)

Step 3: Find corresponding analysis ensemble for observations and model variables

$$y_{ki}^a = y_{ki}^f + \frac{\text{covar}\left(y_k^f, y_j^f\right)}{\text{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } k = 1, 2, ..., p;\ i = 1, 2, ..., K$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\text{covar}\left(x_\mu^f, y_j^f\right)}{\text{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } \mu = 1, 2, ..., n;\ i = 1, 2, ..., K$$

Step 4: Let the analysis ensemble be the prior ensemble for the next observation

$y_{ki}^f = y_{ki}^a$, for $k = 1, 2, ..., p$; $i = 1, 2, ..., K$

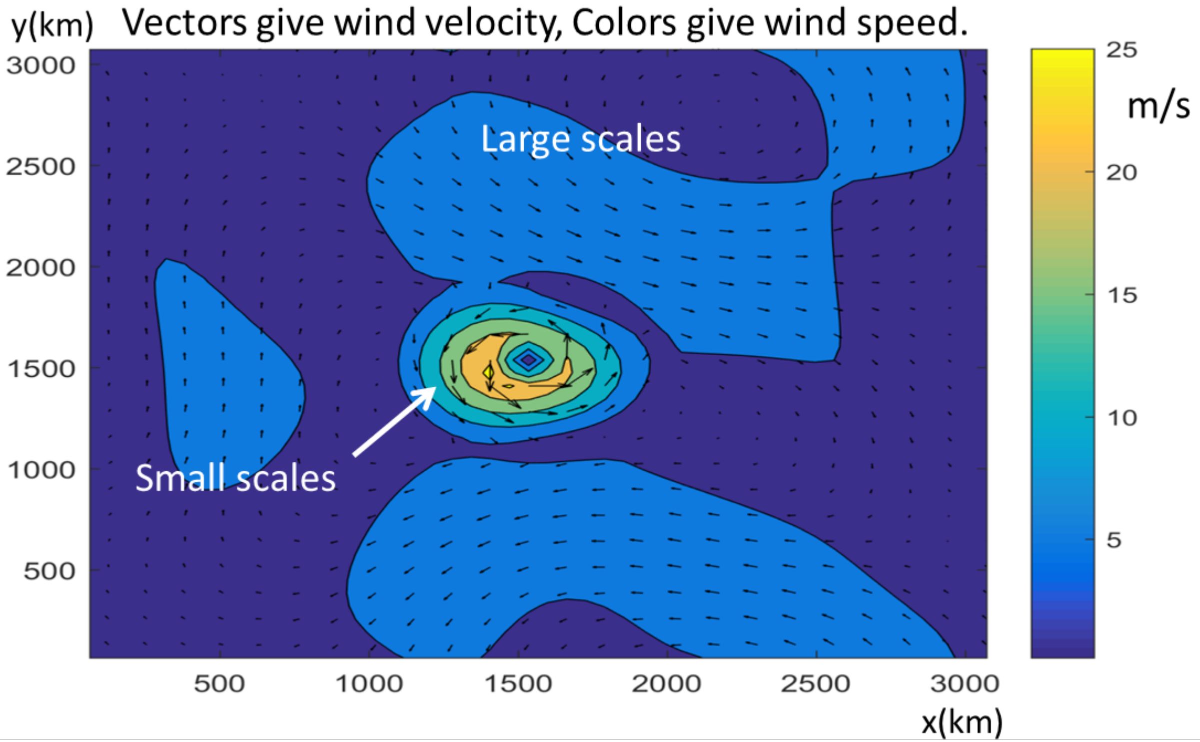
$x_{\mu i}^f = x_{\mu i}^a$, for $\mu = 1, 2, ..., n$; $i = 1, 2, ..., K$

end

Overview

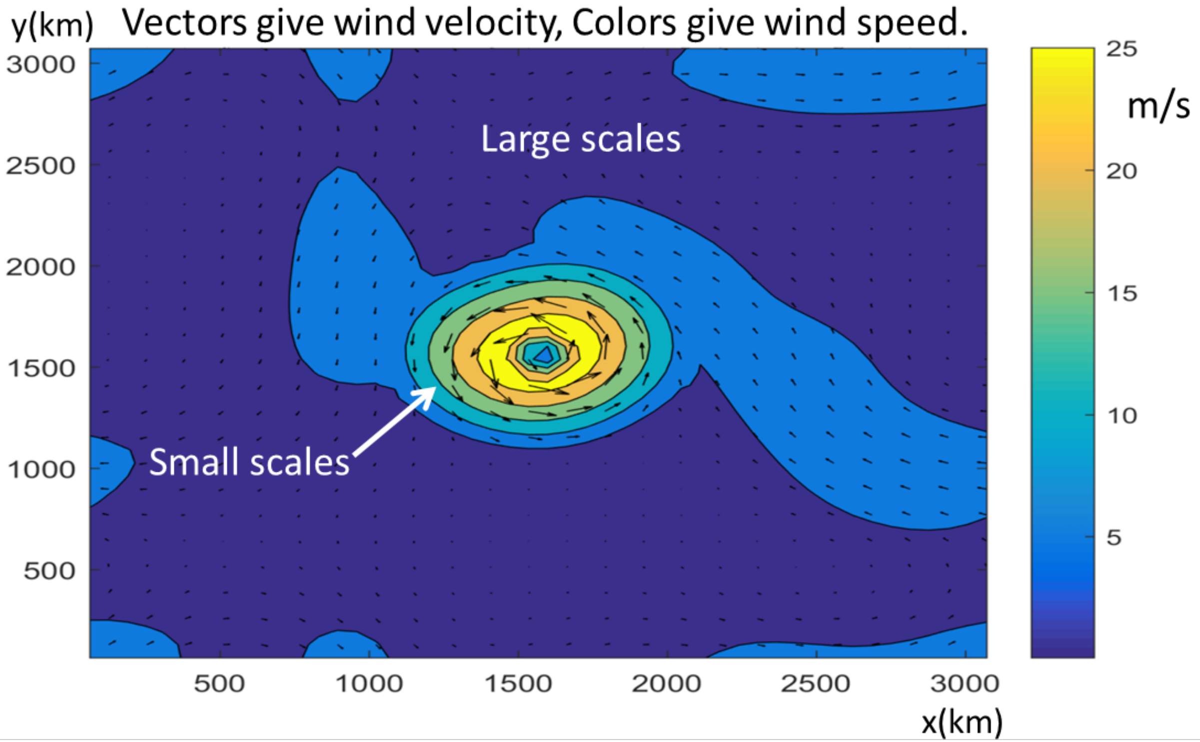
- Aim: Outline some improvements to GIGG-EnKF (Bishop, 2016, QJRMS).
  1. Background
  2. Solution to Bayes' theorem for gamma prior and inverse-gamma-likelihood is now precise as K=>infinity - previously just approximate. ***Significance: Rigorous basis for GIG***
  3. Test of standard GIG for tropical cyclone surface wind energy assimilation problem: ***Significance: Standard GIG better than EAKF/EnKF for this problem.***
  4. Local iterative regression to account for non-linearity in observation operator. ***Significance: Greatly reduces analysis error.***
  5. Rigorous approach for dealing with on-off variables (rain, cloud, fire, etc) with gamma based delta function. ***Significance: Justifies ignoring dry members when rain is observed.***

Simple DA testbed for TC like surface winds



A random draw from a TC relevant pdf

Simple DA testbed for TC like surface winds

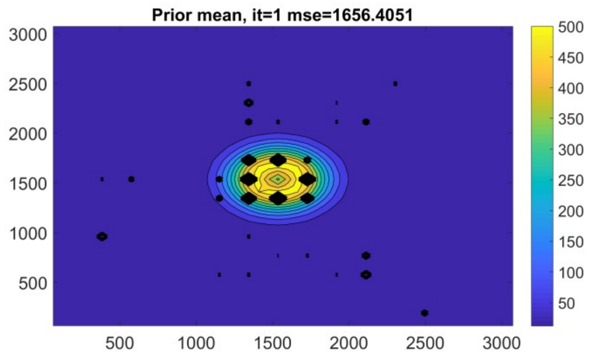


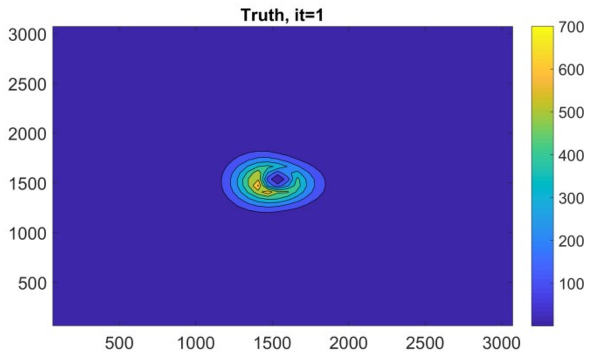
Another random draw from the simple testbed's multi-scale pdf

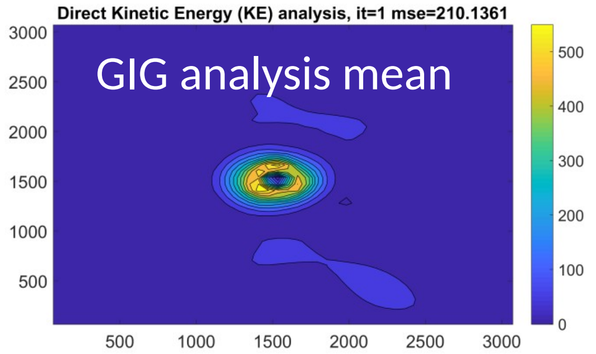
Simple DA testbed for TC like surface winds

- Model states defined by random, multi-scale TC like (u,v) wind field.
- **Let observations be non-linear functions of u and v; e.g. Kinetic Energy, $KE=(u^2+v^2)/2$, tanh(*KE*) or Heaviside(*KE-constant*).**

Prior mean, obs, truth and GIG analysis using a 3000 member ensemble (no localization required).

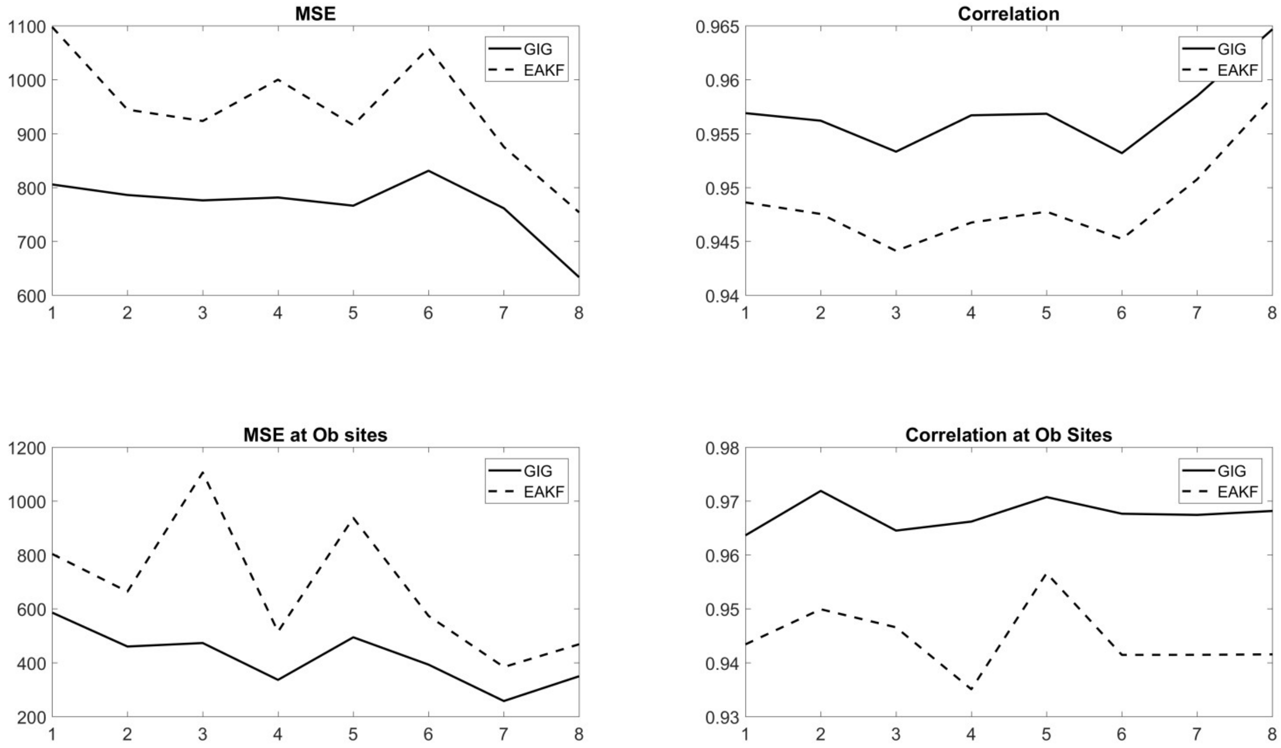






Observed variable is KE=0.5(u^2+v^2). Distribution of random observations given truth is an inverse gamma pdf with a relative variance of 0.25.

Does the GIG variation on the EAKF improve the KE analysis?



The GIG-EnKF outperforms the EAKF under all metrics in all 8 independent sets of 50 trials.

The only difference between EAKF and GIG code is the univariate ensemble update. Linear regression code is identical.

Overview

- Aim: Outline some improvements to GIGG-EnKF (Bishop, 2016, QJRMS).
  1. Background
  2. Solution to Bayes' theorem for gamma prior and inverse-gamma-likelihood is now precise as K=>infinity - previously just approximate. ***Significance: Rigorous basis for GIG***
  3. Test of standard GIG for tropical cyclone surface wind energy assimilation problem: ***Significance: Standard GIG better than EAKF/EnKF for this problem.***
  4. Local iterative regression to account for non-linearity in observation operator. ***Significance: Greatly reduces analysis error.***
  5. Rigorous approach for dealing with on-off variables (rain, cloud, fire, etc) with gamma based delta function. ***Significance: Justifies ignoring dry members when rain is observed.***

Background: The GIGG-EnKF serial observation assimilation scheme with *linear* regression

for $j = 1 : p$;   % where p is the number of observations

Step 1: Decide whether forecast and observation uncertainty associated with y_j^o is best approximated by GIG-delta, GIG, IGG or Gaussian assumptions.

Step 2: if (GIG-delta) then use ... to obtain y_{ji}^a, $i = 1, 2, \ldots, K$;

else if (GIG) then use ... to obtain y_{ji}^a, $i = 1, 2, \ldots, K$; →

Note: GIG posterior mean eq *very* different to EnKF eq

$$\frac{1}{\overline{\overline{y^a}}} = \left(\frac{1}{\overline{\overline{y^f}}}\right) + \tilde{P}_r \left(\tilde{P}_r + \tilde{R}_r\right)^{-1} \left[\frac{1}{y} - \left(1 + \tilde{R}_r\right)\left(\frac{1}{\overline{\overline{y^f}}}\right)\right].$$

else if (IGG) then use ... to obtain y_{ji}^a, $i = 1, 2, \ldots, K$;

else if (Gaussian) then use ... to obtain y_{ji}^a, $i = 1, 2, \ldots, K$ (EAKF/EnSRF/EnKF)

Step 3: Find corresponding analysis ensemble for observations and model variables

$$y_{ki}^a = y_{ki}^f + \frac{\operatorname{covar}\left(y_k^f, y_j^f\right)}{\operatorname{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } k = 1, 2, \ldots, p;\ i = 1, 2, \ldots, K$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\operatorname{covar}\left(x_\mu^f, y_j^f\right)}{\operatorname{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } \mu = 1, 2, \ldots, n;\ i = 1, 2, \ldots, K$$

Step 4: Let the analysis ensemble be the prior ensemble for the next observation

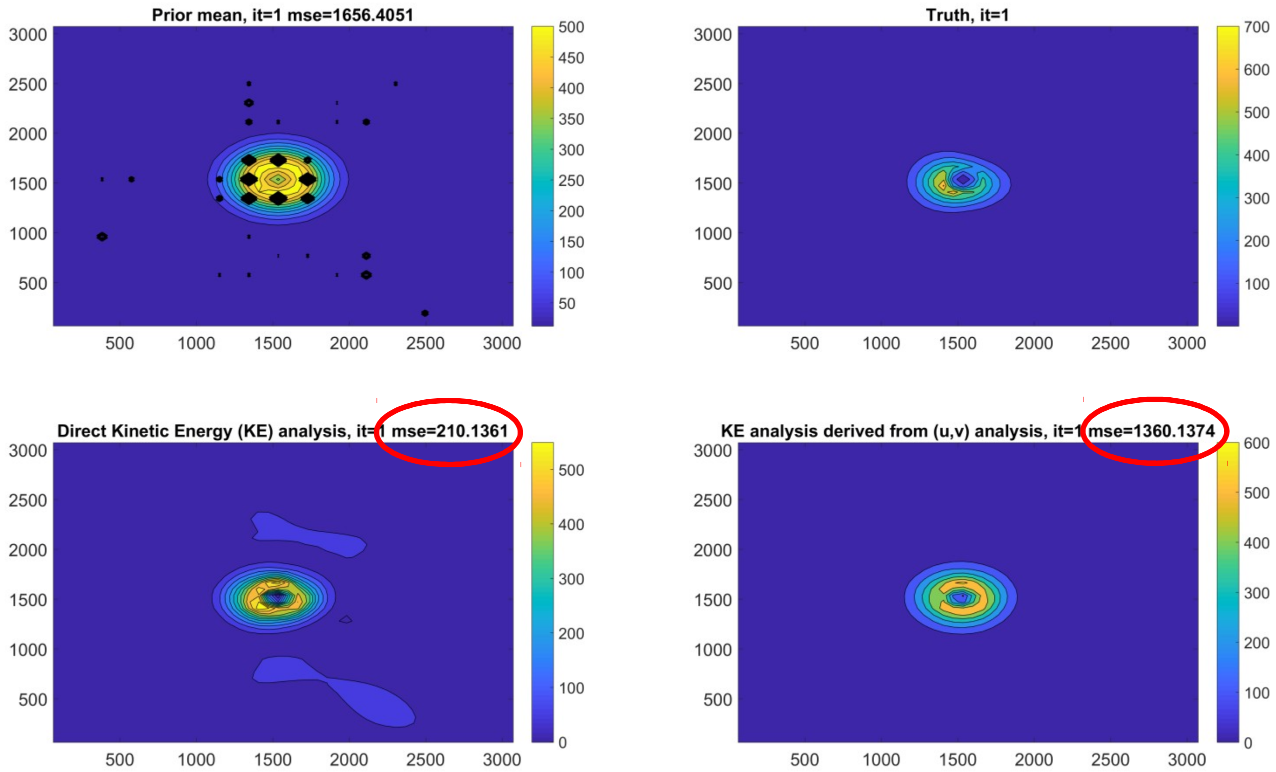
$y_{ki}^f = y_{ki}^a$, for $k = 1, 2, \ldots, p$; $i = 1, 2, \ldots, K$

$x_{\mu i}^f = x_{\mu i}^a$, for $\mu = 1, 2, \ldots, n$; $i = 1, 2, \ldots, K$

end

U.S. NAVAL RESEARCH LABORATORY

Treatment of non-linearity of Kinetic Energy ob operator. Linear regression from ob to model space yields inconsistencies!



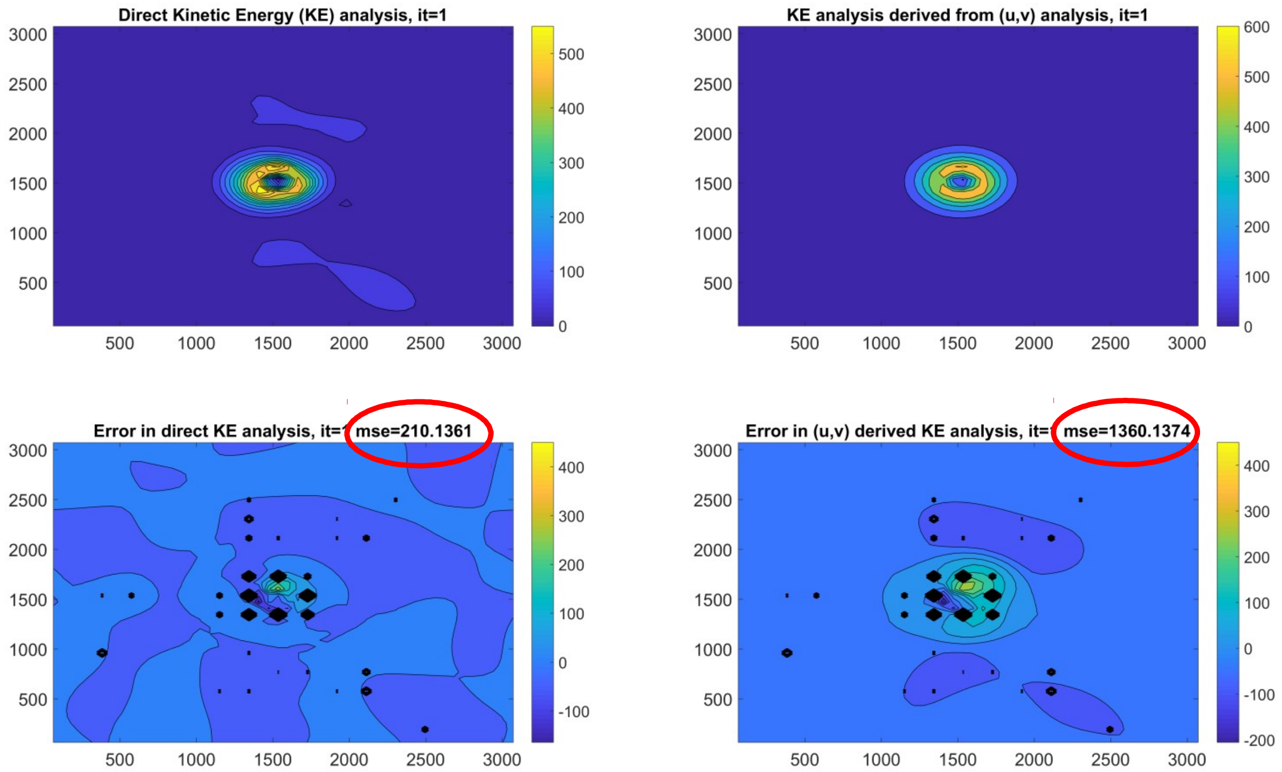
Observed variable is $KE=0.5(u^2+v^2)$. Standard GIG/EAKF uses linear regression to give an inconsistent analysis of (u^a,v^a) and $(KE)^a$. Bottom left panel gives $(KE)^a$. Bottom right gives,

$$KE\left(u^a,v^a\right)=\frac{1}{2}\left(u^{a2}+v^{a2}\right)\neq\left(KE\right)^a$$

- which is far less accurate than $(KE)^a$.

U.S. NAVAL RESEARCH LABORATORY

Treatment of non-linearity of Kinetic Energy ob operator (linear regression from ob to model space yields inconsistencies)



Observed variable is KE=0.5(u^2+v^2). Standard GIG/EAKF uses linear regression to give an inconsistent analysis of (u^a,v^a) and $(KE)^a$. Bottom left panel gives $(KE)^a$. Bottom right gives,

$$KE\left(u^a, v^a\right) = \frac{1}{2}\left(u^{a2} + v^{a2}\right) \neq \left(KE\right)^a$$

- which is far less accurate than $(KE)^a$.

Background: The GIGG-EnKF serial observation assimilation scheme

for $j = 1 : p$;   % where p is the number of observations

Step 1: Decide whether forecast and observation uncertainty associated with y_j^o is best approximated by GIG-delta, GIG, IGG or Gaussian assumptions.

Step 2: if (GIG-delta) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$;

else if (GIG) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$;

else if (IGG) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$;

else if (Gaussian) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$ (EAKF/EnSRF/EnKF)

Step 3: Find corresponding analysis ensemble for observations and model variables

$$y_{ki}^a = y_{ki}^f + \frac{\operatorname{covar}\left(y_k^f, y_j^f\right)}{\operatorname{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } k = 1, 2, ..., p;\ i = 1, 2, ..., K$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\operatorname{covar}\left(x_\mu^f, y_j^f\right)}{\operatorname{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } \mu = 1, 2, ..., n;\ i = 1, 2, ..., K$$

Need to replace the linear regression step with something better!

Step 4: Let the analysis ensemble be the prior ensemble for the next observation

$$y_{ki}^f = y_{ki}^a, \text{ for } k = 1, 2, ..., p;\ i = 1, 2, ..., K$$

$$x_{\mu i}^f = x_{\mu i}^a, \text{ for } \mu = 1, 2, ..., n;\ i = 1, 2, ..., K$$

end

New method to account for non-linearity in ob-operator: The observation to model space consistency iteration

3.1: Define minimum list of variables $\left[(\mathbf{x}_i)_{h_j}\right]$ required to predict the ob y_j that was just assimilated;

for example, in the KE example $(\mathbf{x}_i)_{h_j} = \begin{bmatrix} u_j \\ v_j \end{bmatrix}$ where (u_j, v_j) are the wind components required

to predict the KE of the model state at y_j.

3.2: Find the usual GIG-EnKF model-space analysis $\left[(\mathbf{x}_i^a)_{h_j}\right]_{un}$.

3.3: Starting with $(\mathbf{x}_i)_{h_j} = \left[(\mathbf{x}_i^a)_{h_j}\right]_{un}$ minimize $J\left[(\mathbf{x}_i)_{h_j}\right] = \frac{1}{2}\left\{y_{ji}^a - h_j^{local}\left[(\mathbf{x}_i)_{h_j}\right]\right\}^2$

using ensemble-space constrained Newton iteration on gradient to obtain $(\mathbf{x}_i^a)_{h_j}$ (the minimizer).

New method to account for non-linearity in ob-operator: The observation to model space consistency iteration

3.4: Update the rest of the model state using $\left(\mathbf{x}_i^a\right)_{h_j}$ and linear **multivariate** regression

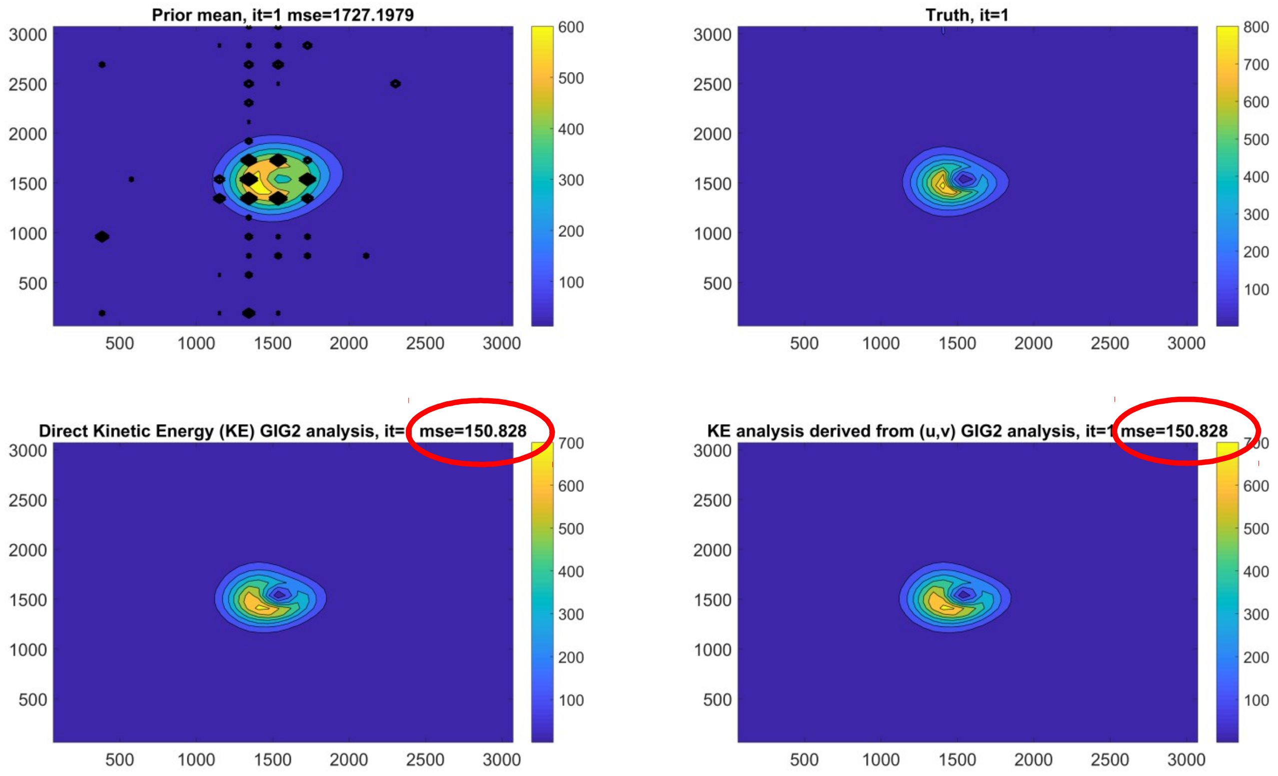
$$x_{\mu i}^a = x_{\mu i}^f + \mathrm{covar}\left[x_{\mu i}^f, \left(\mathbf{x}^f\right)_{h_j}\right]\left[\mathrm{covar}\left(\left(\mathbf{x}^f\right)_{h_j}, \left(\mathbf{x}^f\right)_{h_j}\right)\right]^{-1}\left[\left(\mathbf{x}_i^a\right)_{h_j} - \left(\mathbf{x}_i^f\right)_{h_j}\right] \text{ for } \mu = 1,2,\ldots,n;\ i = 1,2,\ldots,K$$

$$\left(y_{ki}^a\right)_{lin} = y_{ki}^f + \frac{\mathrm{covar}\left(y_k^f, y_j^f\right)}{\mathrm{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } k = 1,2,\ldots,p;\ i = 1,2,\ldots,K$$

$$y_{ki}^a = \begin{cases} 0.5\left[\left(y_{ki}^a\right)_{lin} + h_k^{local}\left(\left(\mathbf{x}_i^a\right)_{h_k}\right)\right] & \text{if } \left(y_{ki}^a\right)_{lin} \geq 0 \\ h_k^{local}\left(\left(\mathbf{x}_i^a\right)_{h_k}\right) & \text{if } \left(y_{ki}^a\right)_{lin} < 0 \end{cases}, \quad \text{for } k = 1,2,\ldots,p;\ i = 1,2,\ldots,K$$

The observation to model space consistency iteration. Test in 2D model

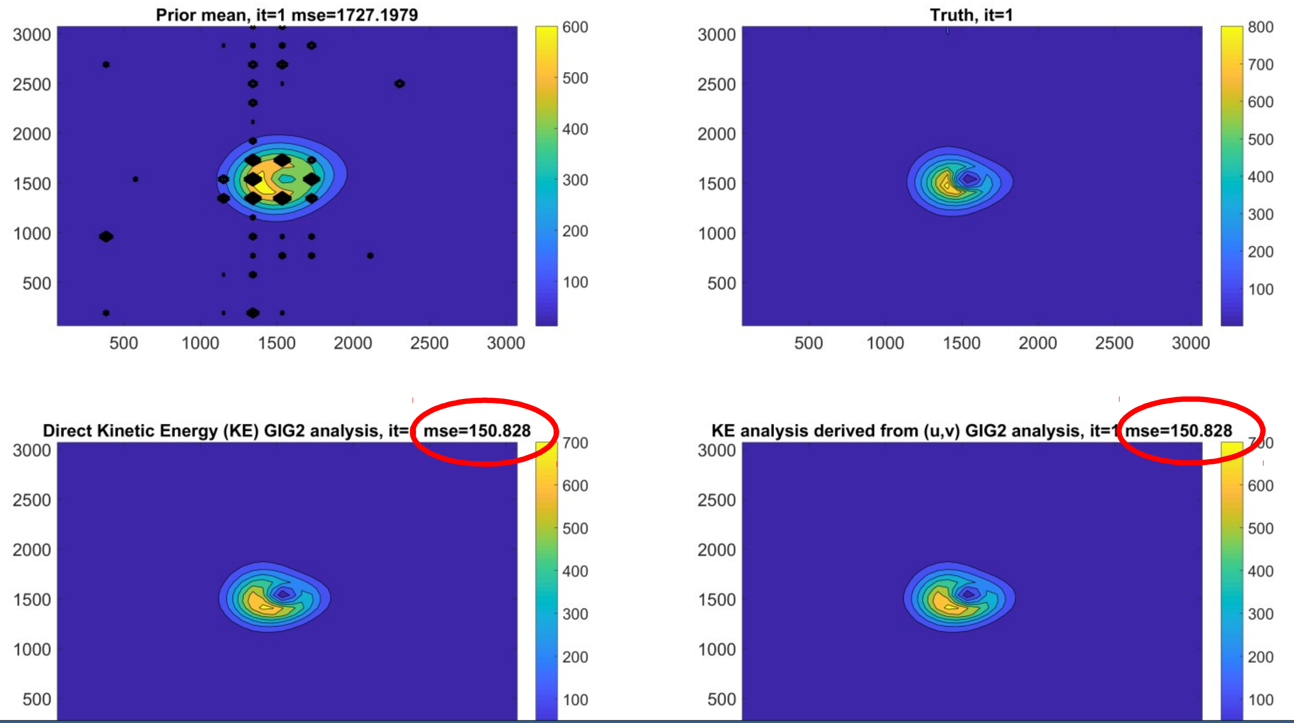
U.S. NAVAL RESEARCH LABORATORY



Observed variable is KE=0.5(u^2+v^2). Linear regression plus consistency iteration improves consistency of (u^a,v^a) and $(KE)^a$. Bottom left panel gives $(KE)^a$. Bottom right gives,

$$KE\left(u^a, v^a\right) = \frac{1}{2}\left(u^{a2} + v^{a2}\right) \approx (KE)^a$$

The observation to model space consistency iteration. Test in 2D model

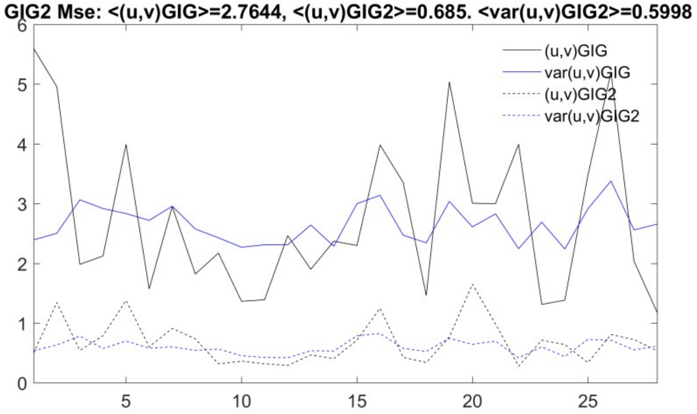


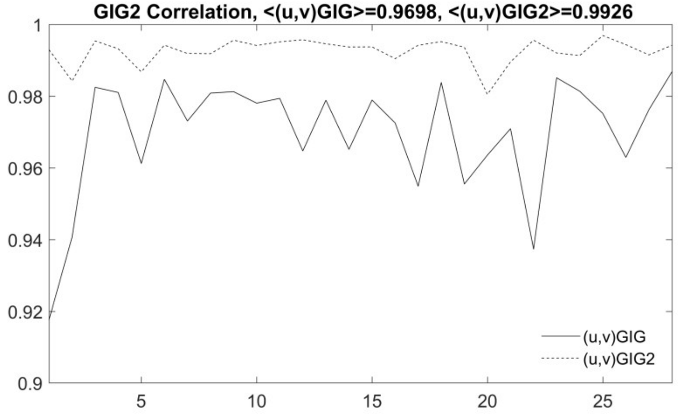
Observed variable is KE=0.5(u^2+v^2). Linear regression plus consistency iteration improves consistency of (u^a,v^a) and $(KE)^a$. Bottom left panel gives $(KE)^a$. Bottom right gives,

$$KE\left(u^a, v^a\right) = \frac{1}{2}\left(u^{a2} + v^{a2}\right) \approx (KE)^a$$

Accuracy of direct and derived KE analyses are now the same

The observation to model space consistency iteration. 28 independent tests in 2D model



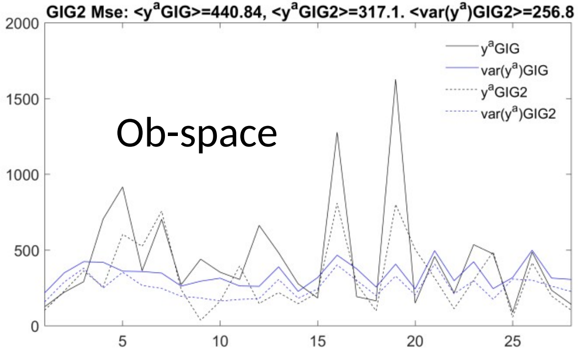


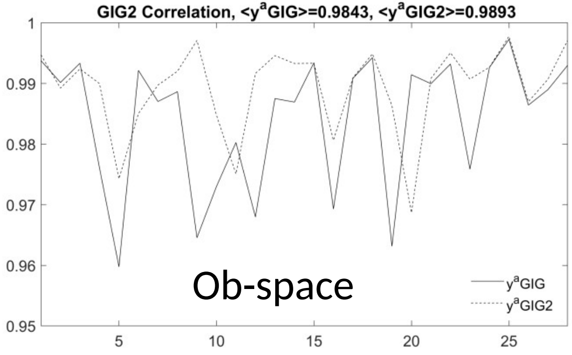
Ob-to-model space consistency iteration reduces mse in (u,v) field by 75%;
i.e. standard deviation of analysis error is halved
Chance of getting 28 wins (as above) by pure chance is 1 in 2.8×10^8 .

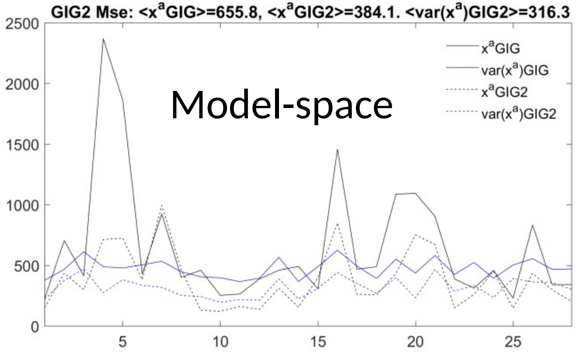
Conclusions

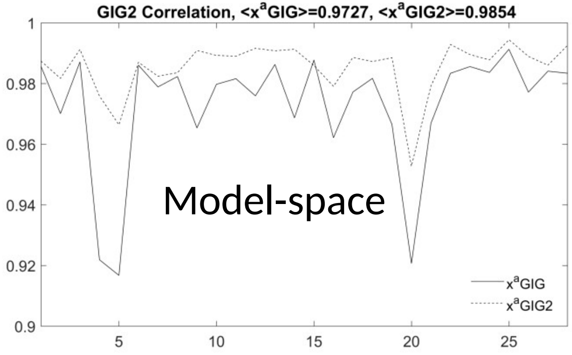
1. Solution to Bayes' theorem for gamma prior and inverse-gamma-likelihood is now precise as K=>infinity - previously just approximate. ***Significance: Rigorous basis for GIG***
2. In 1D experiments, the new local ob-space to model space iteration procedure gave fairly accurate multi-modal posteriors.
3. **In idealized TC surface wind energy assimilation experiments, GIG soundly beat EAKF even using *linear* regression. The newly introduced ob-space to model space *non-linear* regression iteration:**
   i. **Gave consistent model and ob space analyses**
   ii. **Greatly reduced mean square error (mse).**
   iii. **Gave an analysis ensemble variance approximately equal to mse.**
4. Rigorous approach for dealing with on-off variables (rain, cloud, fire, etc) with gamma based delta function. ***Significance: Justifies ignoring dry members when rain is observed.***

The observation to model space consistency iteration. 28 independent tests in 2D model



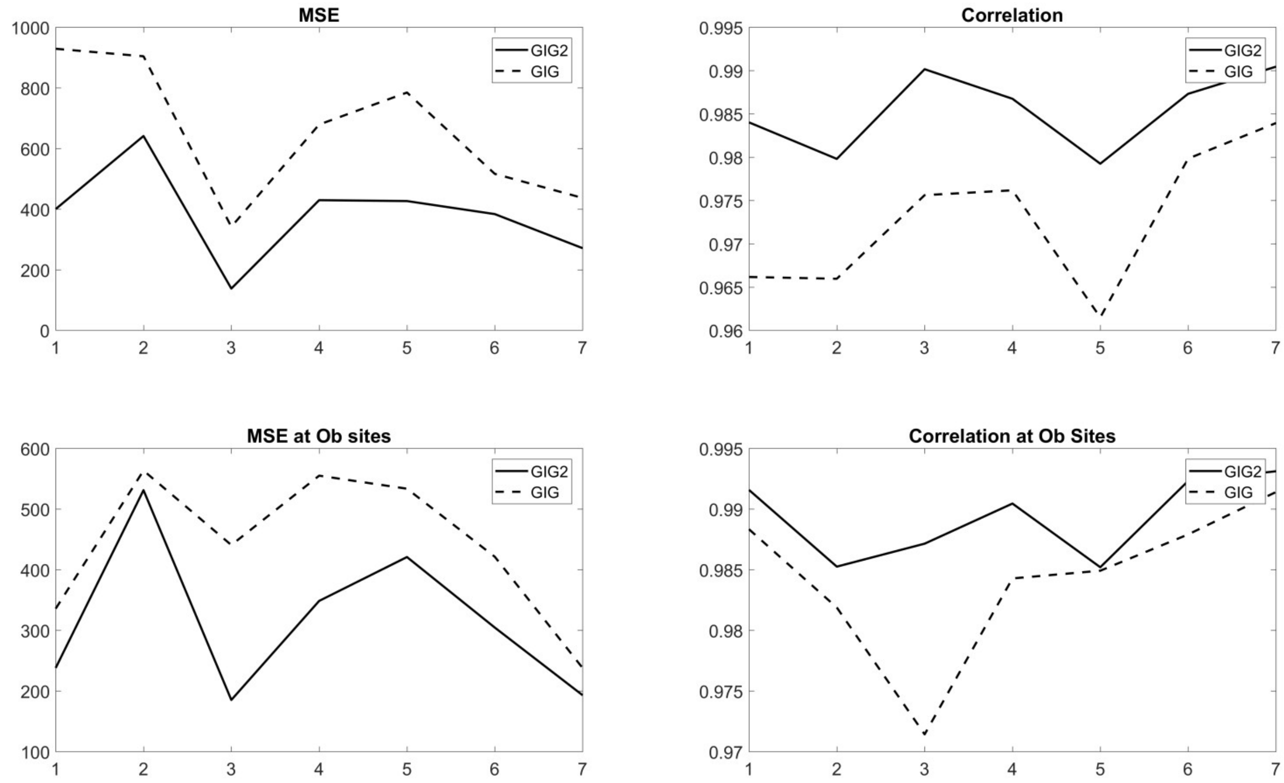






Ob-to-model space consistency iteration reduces mse in KE field by 41%

Ob-to-model-space consistency iteration helps! (Result for Kinetic Energy below, 7x4 independent trials)

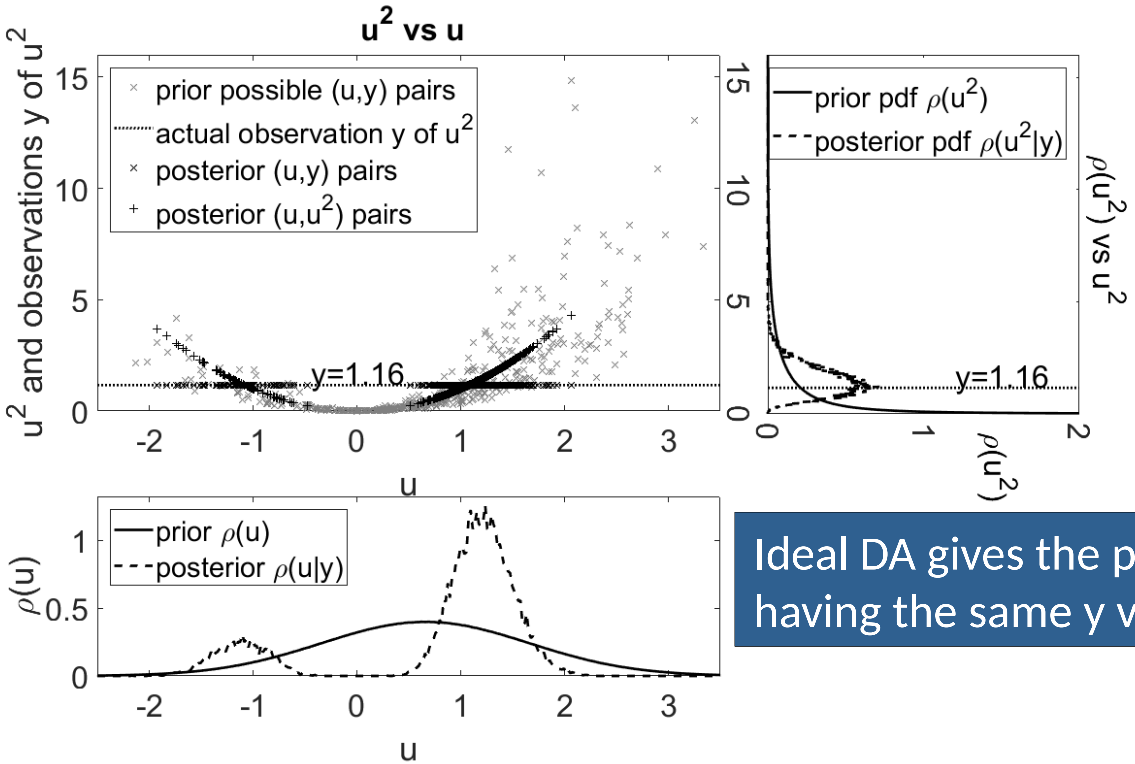


Ob-to-model space consistency iteration reduces mse in KE field by 41%

Chance of getting 7 wins by pure chance is 1 in 128.

Data assimilation for clouds and high-resolution models

Ideal Data Assimilation (DA) in a simple model

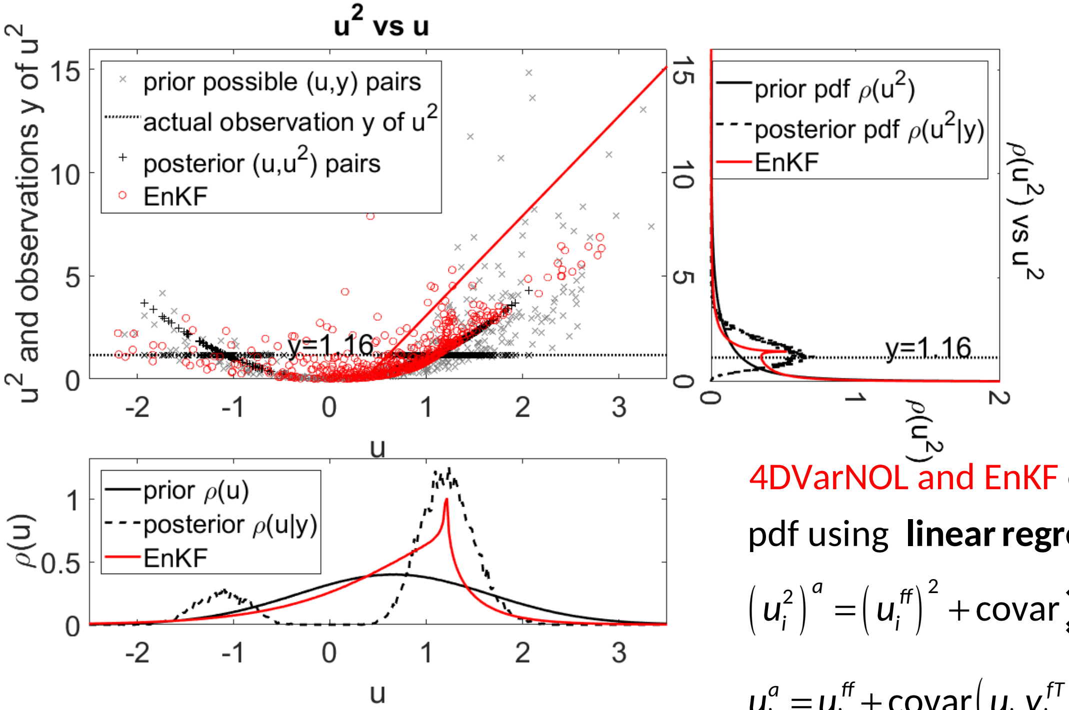


Prior ***and posterior*** pdf of u^2 are like gamma pdfs. Highly non-Gaussian.

Ideal DA gives the posterior pdf of replicate Earths having the same y value as our Earth’s y value.

Ideal posterior pdf is bi-modal. Bi-modality caused by non-linearity

Current DA: 4DVar-No-Outer-Loop (US Navy) and EnKF (DWD)



EnKF & 4DvarNoOuterLoop (4DVarNOL) ***posterior*** pdf of u^2 is highly inaccurate. Also, analyzed u^2 values are not equal to the square of analyzed u values

EnKF/4DVarNOL ***posterior*** pdf of u is very poor.

4DVarNOL and EnKF estimate the posterior pdf using **linear regression** and **perturbed observations**

$$\left(u_i^2\right)^a = \left(u_i^{ff}\right)^2 + \mathrm{covar}\left(\left(u_i\right)^2 y_i^{fT}\right)\left(\mathrm{covar}\left(y_i^{ff} y_i^{T}\right)\right)^{-1}\left(\left(y+\varepsilon_i^o\right)-\left(u_i^f\right)^2\right)$$

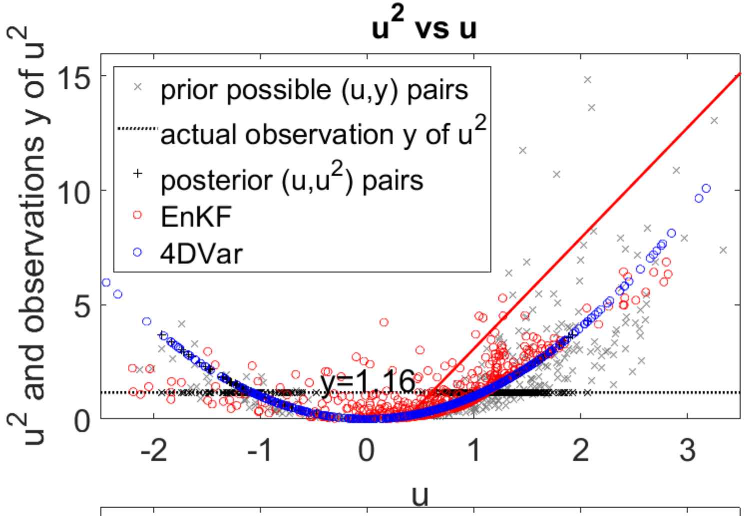
$$u_i^a = u_i^{ff} + \mathrm{covar}\left(u_i\, y_i^{fT}\right)\left(\mathrm{covar}\left(y_i^{ff} y_i^{T}\right)\right)^{-1}\left(\left(y+\varepsilon_i^o\right)-\left(u_i^f\right)^2\right)$$

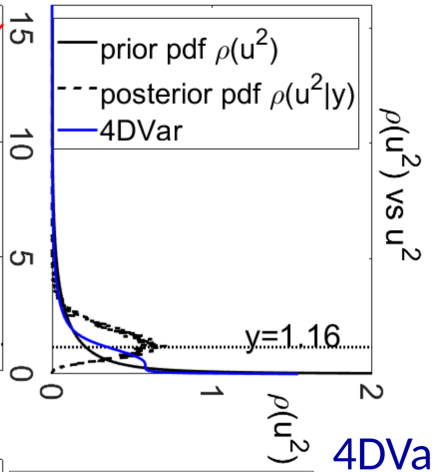
Fails due to linear, Gaussian assumptions

U.S. NAVAL RESEARCH LABORATORY

Data assimilation for clouds and high-resolution models

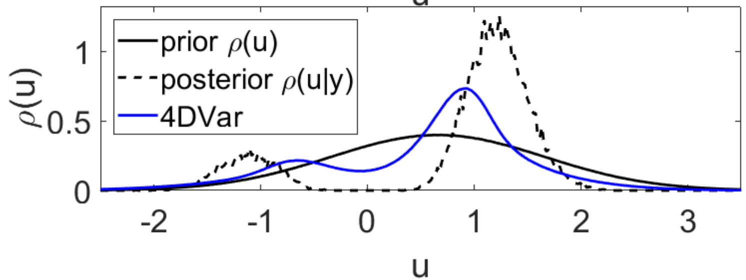
Current DA: Incremental 4DVar (4DVar-with-outer-loop)





4DVar posterior pdf of u^2 is highly inaccurate.

However, analyzed u^2 values are now equal to the square of analyzed u values



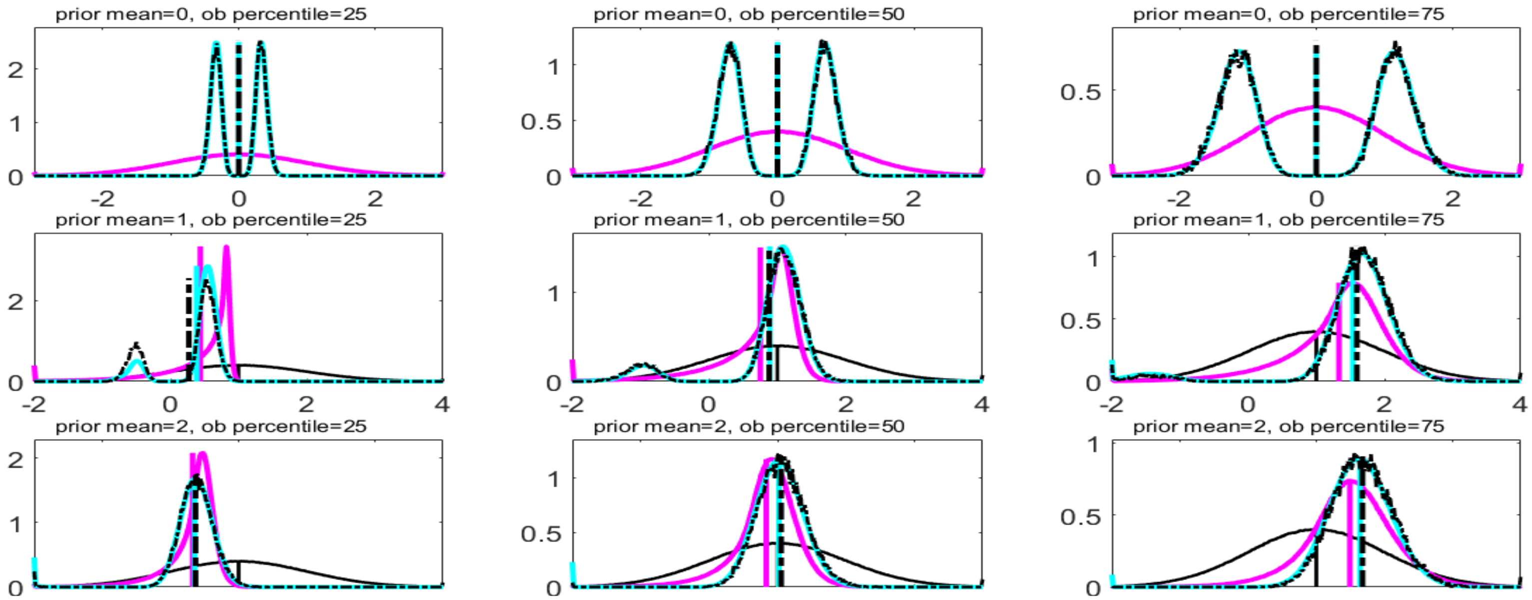
4DVar ***posterior*** pdf of u is still poor.

4DVar uses perturbed observations. Each posterior member is a local extreme value of a Gaussian approximation to the true posterior pdf.

Fails due to Gaussian assumption and the presence of multiple extrema (non-linearity)

The observation to model space consistency iteration. Test in 1D model

U.S. NAVAL RESEARCH LABORATORY



Solid black line gives prior pdf of zonal wind (***u***) field

u^2 is observed at 25th, 50th or 75th percentile of prior (left to right)

Dashed black line gives true posterior pdf of u field

Solid mauve line is GIGG posterior pdf with no outer loop

Problem: No rain in ensemble forecast but rain is observed

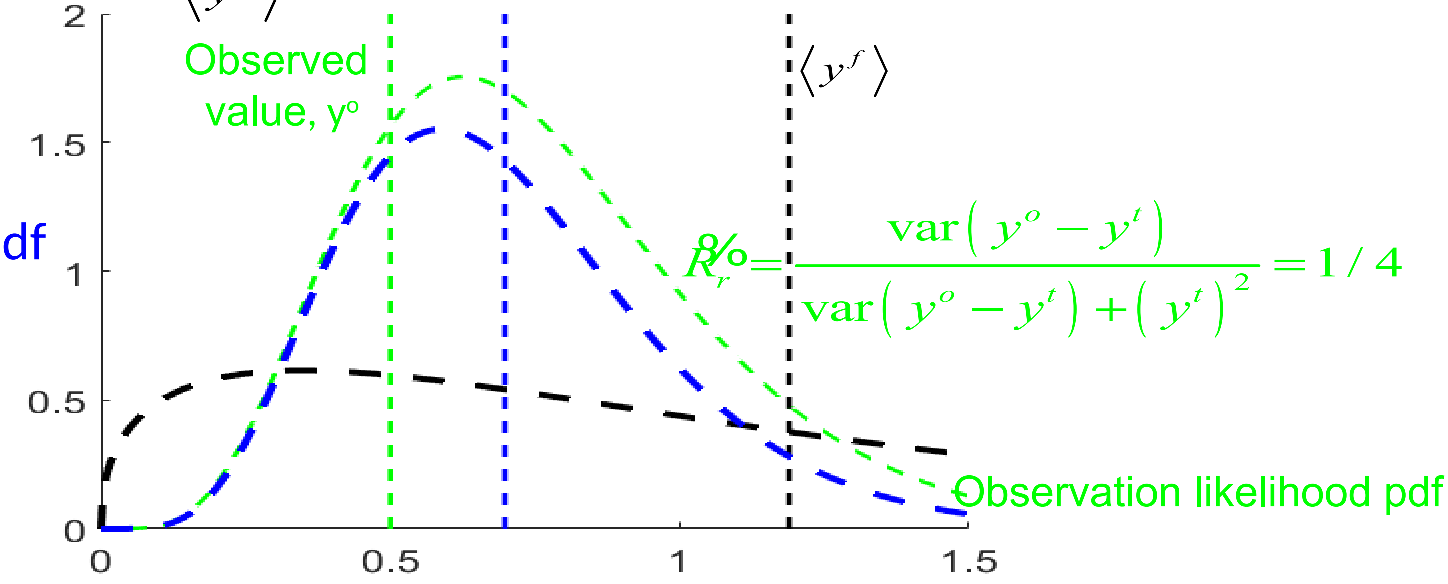
- EnKFs, 4DVAR, Particle filters, etc, all fail in this case.
- How would Bayes' theorem be used in this case?
- Might an adaptation of the GIGG filter better deal with this problem?

No rain forecast as a gamma function limit

Dashed black lines pertain to prior/forecast pdf with

$$P^r = \frac{\mathrm{var}\left(y^f\right)}{\left\langle y^f \right\rangle^2} = 0.707 \text{ and } \left\langle y^f \right\rangle = \left(P^r\right)^{-1/2}$$

Dashed blue lines pertain to posterior/analysis pdf

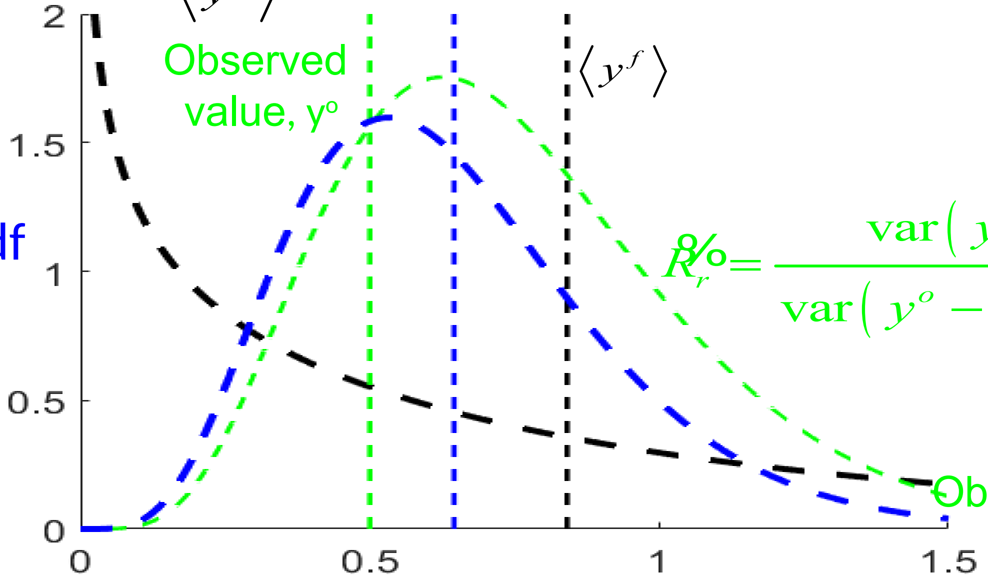


No rain forecast as a gamma function limit

Dashed black lines pertain to prior/forecast pdf with

$$P^r = \frac{\mathrm{var}\left(y^f\right)}{\left\langle y^f \right\rangle^2} = 1.41 \text{ and } \left\langle y^f \right\rangle = \left(P^r\right)^{-1/2}$$

Dashed blue lines pertain to posterior/analysis pdf



$$\tilde{R}^o_r = \frac{\mathrm{var}\left(y^o - y^t\right)}{\mathrm{var}\left(y^o - y^t\right) + \left(y^t\right)^2} = 1/4$$

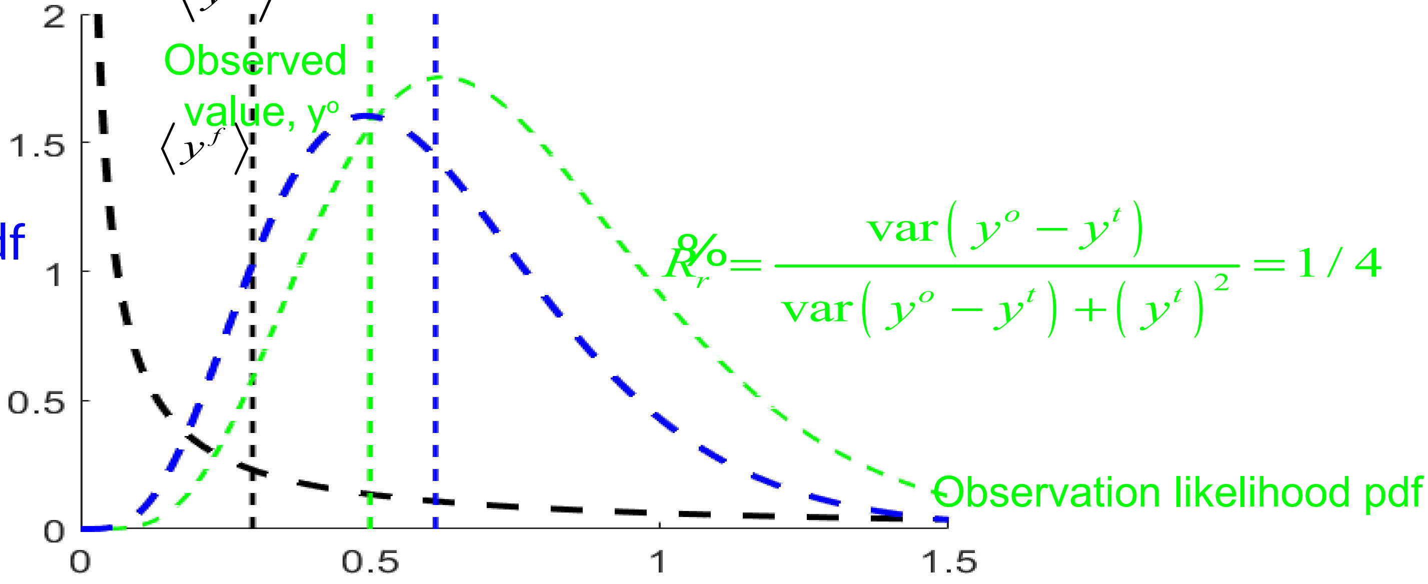
Observation likelihood pdf

No rain forecast as a gamma function limit

Dashed black lines pertain to prior/forecast pdf with

$$P^r = \frac{\mathrm{var}\left(y^f\right)}{\left\langle y^f \right\rangle^2} = 11.3 \text{ and } \left\langle y^f \right\rangle = \left(P^r\right)^{-1/2}$$

Dashed blue lines pertain to posterior/analysis pdf



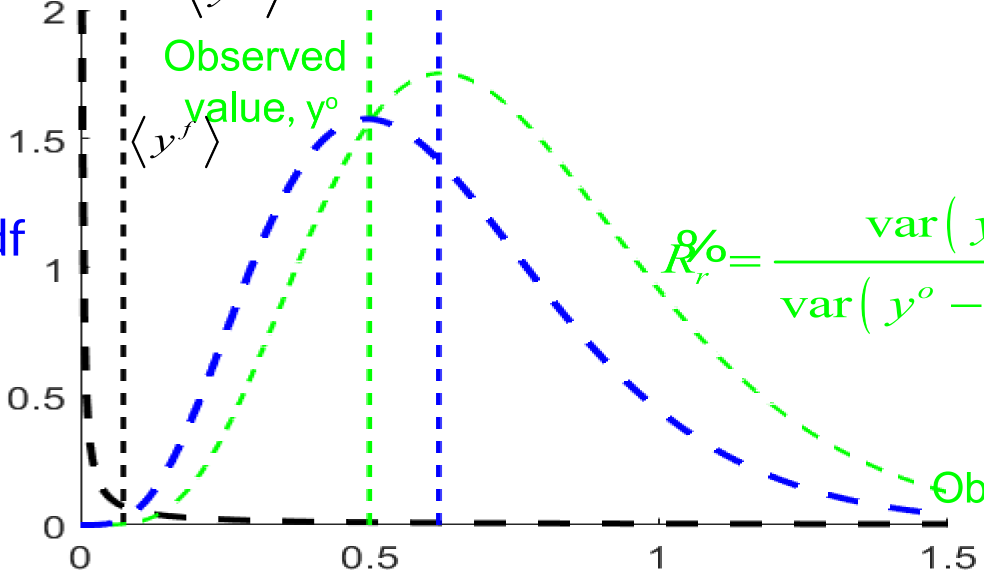
$$\tilde{R}_r^o = \frac{\mathrm{var}\left(y^o - y^t\right)}{\mathrm{var}\left(y^o - y^t\right) + \left(y^t\right)^2} = 1/4$$

No rain forecast as a gamma function limit

Dashed black lines pertain to prior/forecast pdf with

$$P^r = \frac{\mathrm{var}\left(y^f\right)}{\left\langle y^f \right\rangle^2} = 181 \text{ and } \left\langle y^f \right\rangle = \left(P^r\right)^{-1/2}$$

Dashed blue lines pertain to posterior/analysis pdf



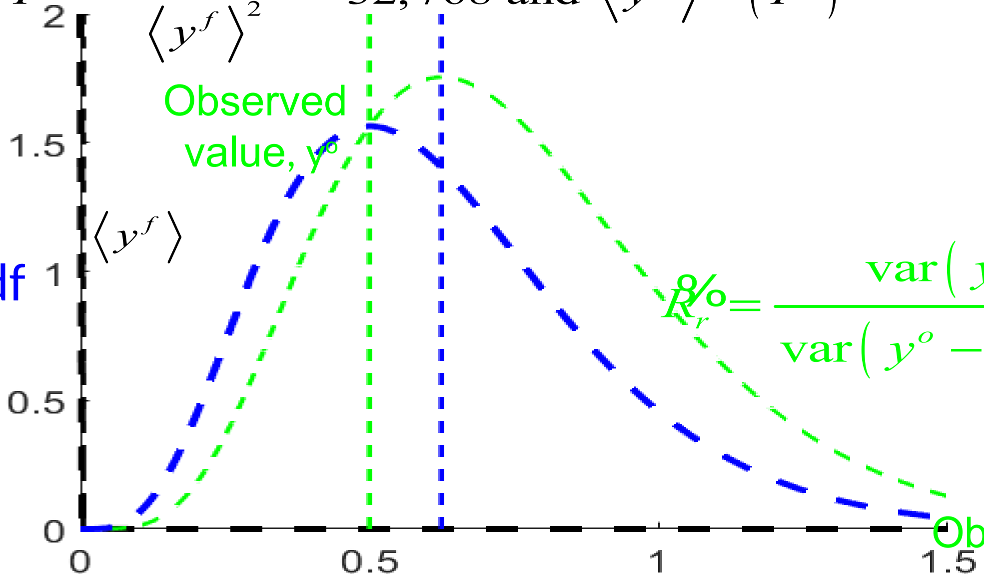
$$R^o_r = \frac{\mathrm{var}\left(y^o - y^t\right)}{\mathrm{var}\left(y^o - y^t\right) + \left(y^t\right)^2} = 1/4$$

No rain forecast as a gamma function limit: a gamma delta function

Dashed black lines pertain to prior/forecast pdf with

$$P^r = \frac{\mathrm{var}\left(y^f\right)}{\left\langle y^f \right\rangle^2} = 32,768 \text{ and } \left\langle y^f \right\rangle = \left(P^r\right)^{-1/2}$$

Dashed blue lines pertain to posterior/analysis pdf



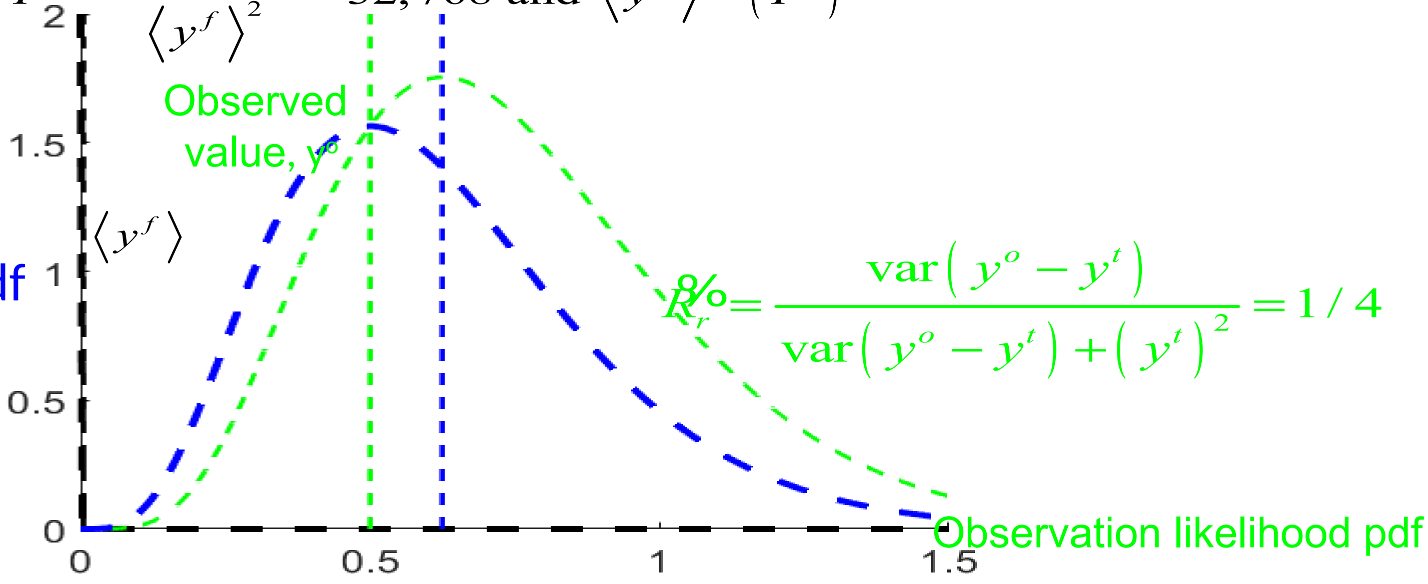
Note that (i) posterior mode is equal to the observed value, and (ii) posterior mean is equal to the mode of ob-likelihood function.

No rain forecast as a gamma function limit: a gamma delta function

Dashed black lines pertain to prior/forecast pdf with

$$P^r = \frac{\mathrm{var}\left(y^f\right)}{\left\langle y^f \right\rangle^2} = 32,768 \text{ and } \left\langle y^f \right\rangle = \left(P^r\right)^{-1/2}$$

Dashed blue lines pertain to posterior/analysis pdf

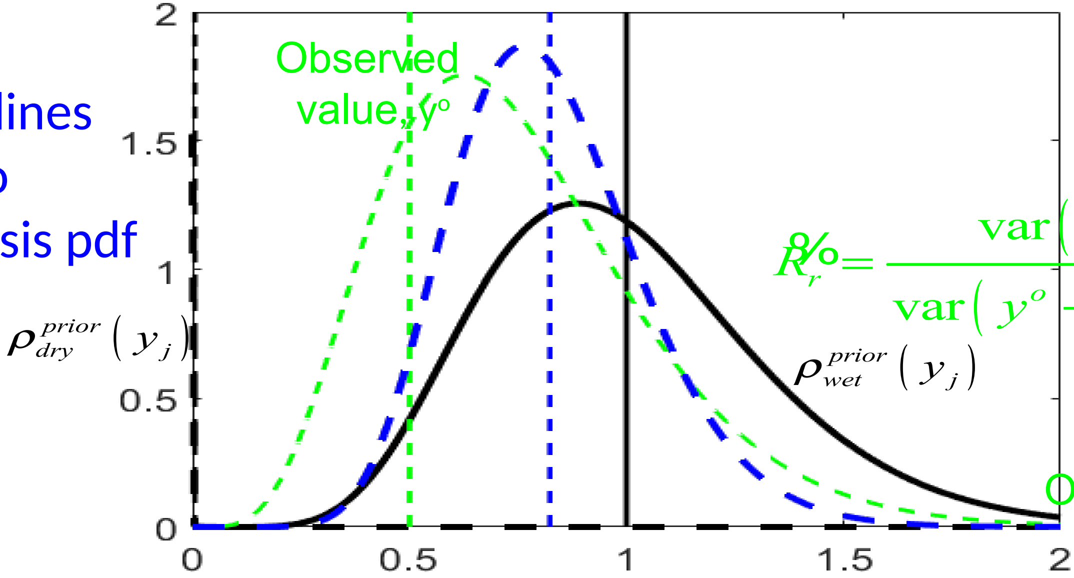


Using gamma delta function to represent the zero-rain-prior pdf makes Bayes' theorem give a plausible posterior pdf.

gamma-delta + gamma pdf for case when some members dry and some wet

$$\rho_{precip}\left(y_j\right) = w_0 \rho_{dry}^{prior}\left(y_j\right) + \left(1 - w_0\right) \rho_{wet}^{prior}\left(y_j\right)$$

Dashed blue lines pertain to posterior/analysis pdf



$$\tilde{R}_r^o = \frac{\mathrm{var}\left(y^o - y^t\right)}{\mathrm{var}\left(y^o - y^t\right) + \left(y^t\right)^2} = 1/4$$

Observation likelihood pdf

In this case, only the mean and variance of the wet members determine the mean and variance of the posterior. Dry members ignored!

Multi-variate GIGG-Delta filter also fits seamlessly in DART

for $j = 1 : p$;   % where p is the number of observations

Step 1: Decide whether forecast and observation uncertainty associated with y_j^o is best approximated by GIG-delta, GIG, IGG or Gaussian assumptions.

Step 2: if (GIG-delta) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$;

else if (GIG) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$;

else if (IGG) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$;

else if (Gaussian) then use ... to obtain y_{ji}^a, $i = 1, 2, ..., K$

Step 3: Find corresponding analysis ensemble for observations and model variables

$$y_{ki}^a = y_{ki}^f + \frac{\text{covar}\left(y_k^f, y_j^f\right)}{\text{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } k = 1, 2, ..., p;\ i = 1, 2, ..., K$$

$$x_{\mu i}^a = x_{\mu i}^f + \frac{\text{covar}\left(x_\mu^f, y_j^f\right)}{\text{var}\left(y_j^f\right)}\left(y_{ji}^a - y_{ji}^f\right), \text{ for } \mu = 1, 2, ..., n;\ i = 1, 2, ..., K$$

Step 4: Let the analysis ensemble be the prior ensemble for the next observation

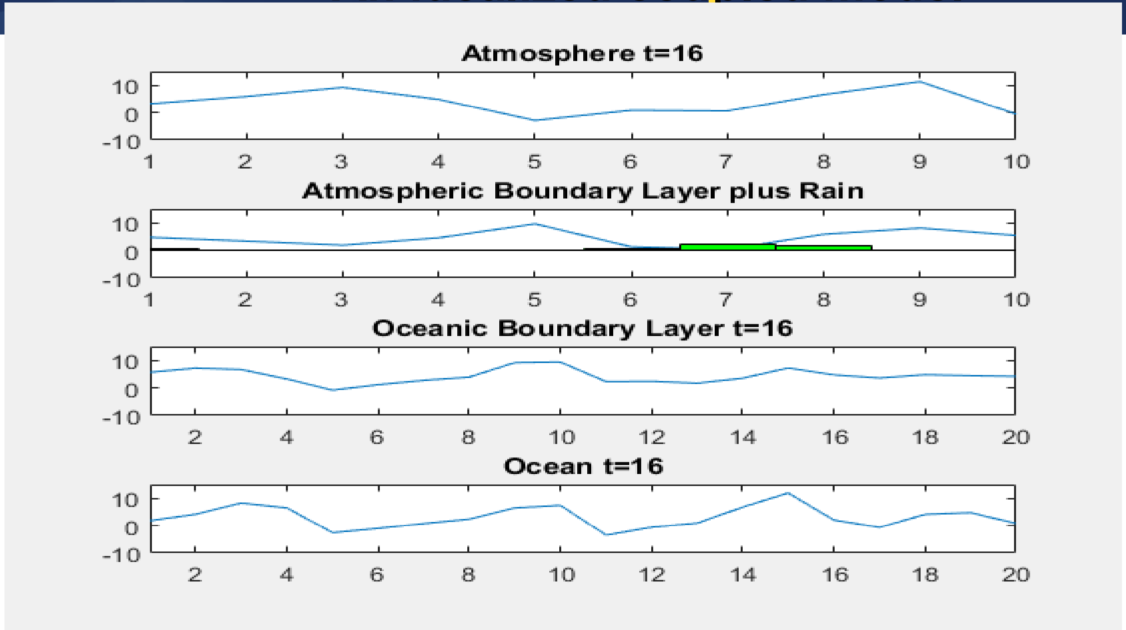
$y_{ki}^f = y_{ki}^a$, for $k = 1, 2, ..., p$; $i = 1, 2, ..., K$

$x_{\mu i}^f = x_{\mu i}^a$, for $\mu = 1, 2, ..., n$; $i = 1, 2, ..., K$

end

GIGG-delta for coupled model DA:

An idealized coupled model

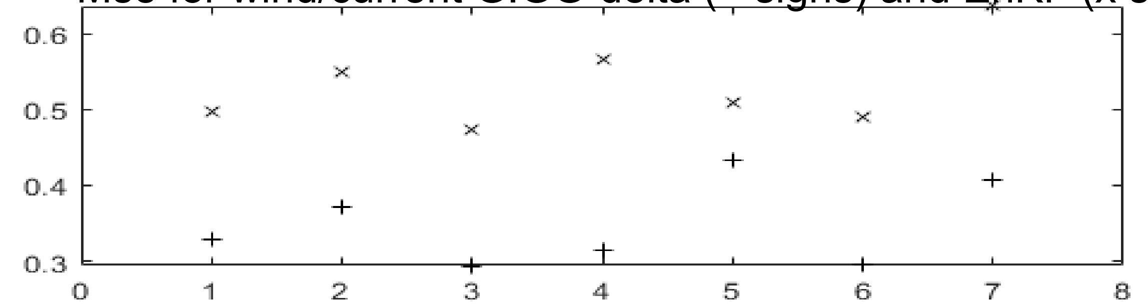


Evolution based on Lorenz 96 model plus relaxation to adjacent levels. The blue line variable is analogous to zonal wind/current. Green bars give rainfall which only occurs when upper level divergence exceeds a small threshold . Rain magnitude is proportional to product of square of the surface wind's deviation from climatological mean and the square root of upper level divergence. Rain increases flux of momentum from upper levels to lower levels.

7 independent 30 day DA cycles

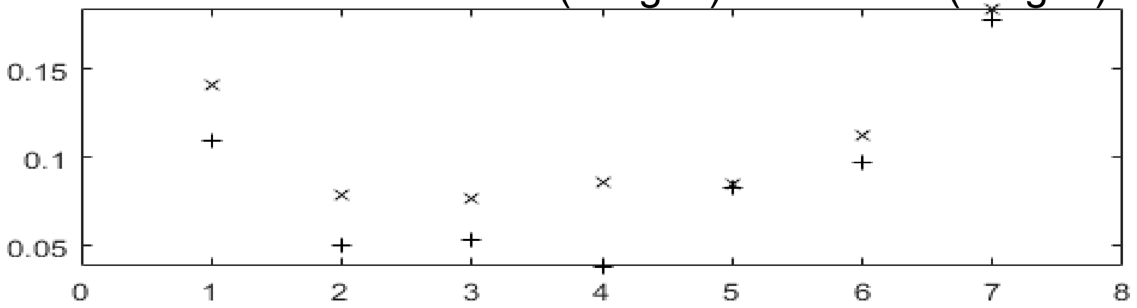
GIGG-delta vs EnKF mse in case when EnKF has stabilizing inflation factor

Mse for wind/current GIGG-delta (+ signs) and EnKF (x signs)



Forecasts

Mse for rain GIGG-delta (+ signs) and EnKF (x signs)



GIGG-delta much better than EnKF in this “stable-system versus stable-system” comparison.

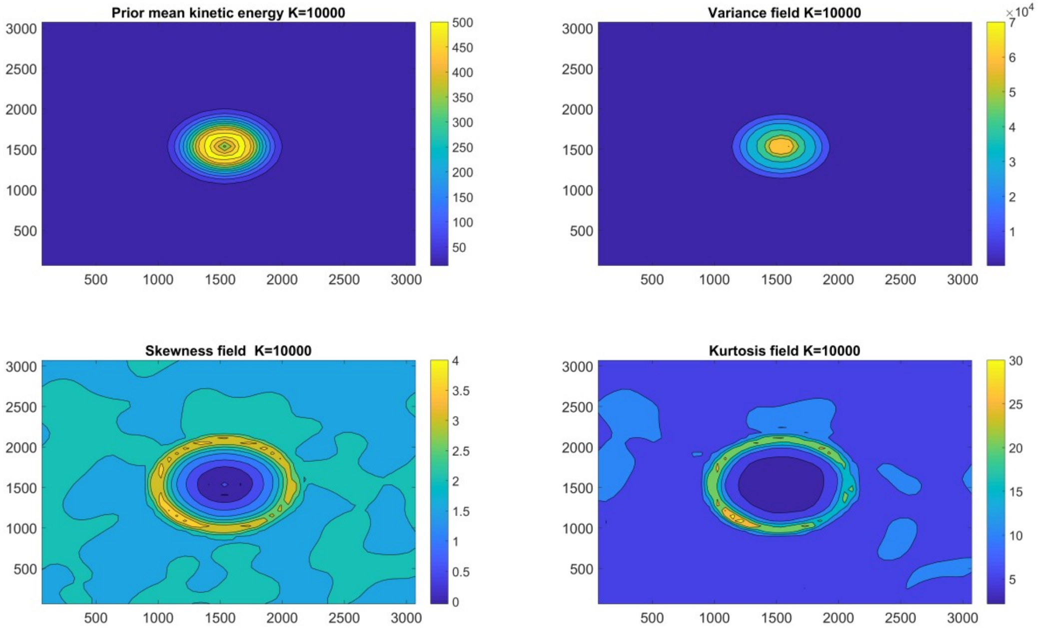
Summary for GIGG-Delta filter

- Theory for GIGG-delta presents a compelling solution to the DA conundrum of observed rain but no rain in prior
- Extension of univariate to multivariate same as Anderson 2003,2007
- Increase in inflation required to stabilize EnKF renders it much less accurate than the GIGG-delta filter.

Multi-variate "all-at-once" assimilation of parameterized non-Gaussian pdfs not so easy

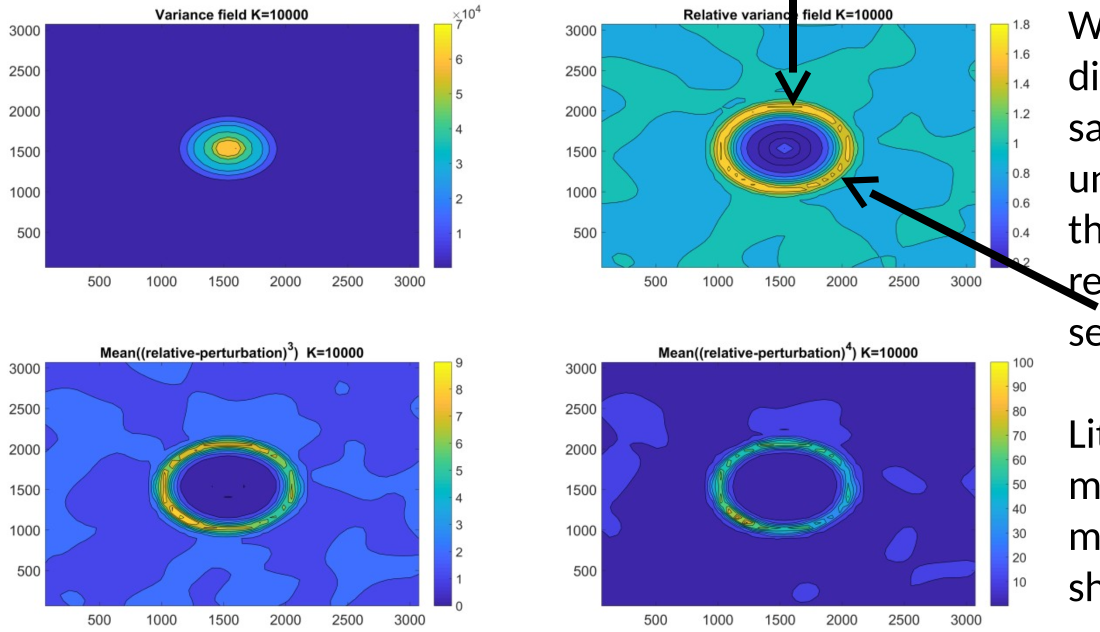
- To do this, one first needs to find a compelling multi-variate statistical model of the moments of the prior.
- … not so easy in the multi-variate case …
- Multi-variate Wishart, Gaussian and log-normal all fail

First four moments of actual prior



Mean of powers of relative perturbation [(x-<x>)/<x>]

Variance almost twice the size of the mean here

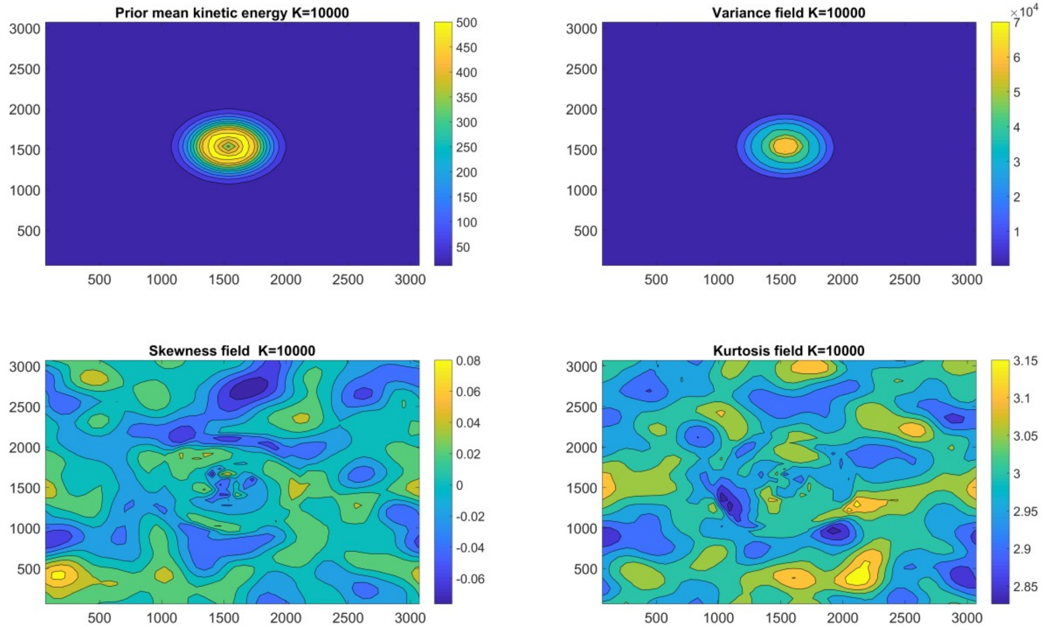


Wishart distribution and other such distributions based on sample covariances of samples from normal pdfs produce spatially uniform relative variances of 2/(N-1) where N is the sample size. Hence, they are incapable of representing the variation in relative variance seen here.

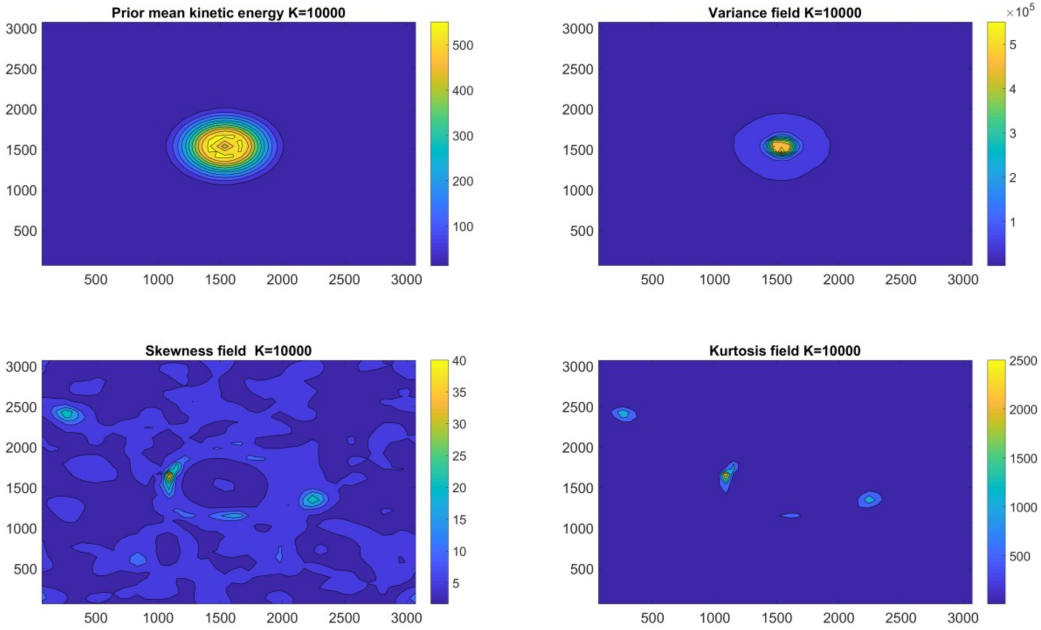
Literature search failed to find gamma-like multi-variate pdf capable of producing the mean of powers of relative perturbations shown here.

U.S. NAVAL RESEARCH LABORATORY

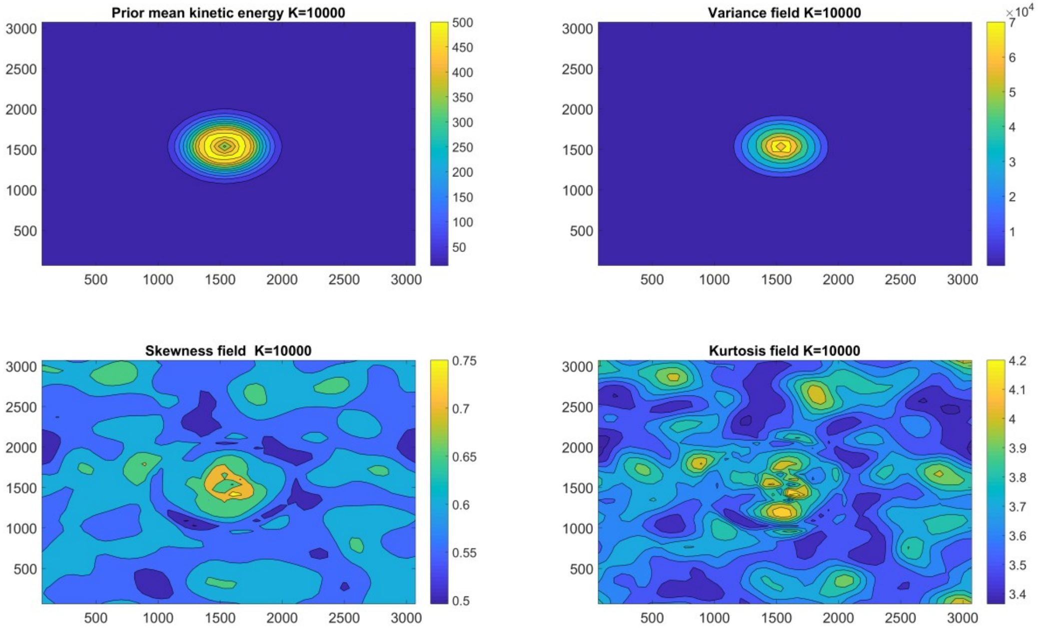
First four moments of Gaussian model of prior



First four moments of log-normal [1+x] model of prior



First four moments of log-normal [f(x)] model of prior



Multi-variate "all-at-once" assimilation of parameterized non-Gaussian pdfs not so easy

- To do this, one first needs to find a compelling multi-variate statistical model of the moments of the prior.
- … not so easy in the multi-variate case …
- Multi-variate Wishart, Gaussian and log-normal all fail

Serial univariate assimilation as in EnKF, EAKF and EnSRF seems to be a way around this