NonGaussian filters

Douglas Nychka
Geophysical Statistics Project
National Center for Atmospheric Research

- A mixture filter
- A local- local filter
- A hybrid
- Something completely different

Supported by the National Science Foundation DMS

Overview

Data Assimilation

Combining predictions made by a numerical model with observed data to estimate the state of a system, $\boldsymbol{x}$. This is also called a *filter*.

The statistical foundation is Bayes Theorem and the uncertainty in the state of the system is represented by a probability distribution.

Overview

Data Assimilation
Combining predictions made by a numerical model with observed data to estimate the state of a system, $\boldsymbol{x}$. This is also called a *filter*.

The statistical foundation is Bayes Theorem and the uncertainty in the state of the system is represented by a probability distribution.

$$PRIOR\ for\ \boldsymbol{x} + observations \rightarrow POSTERIOR\ for\ \boldsymbol{x}$$

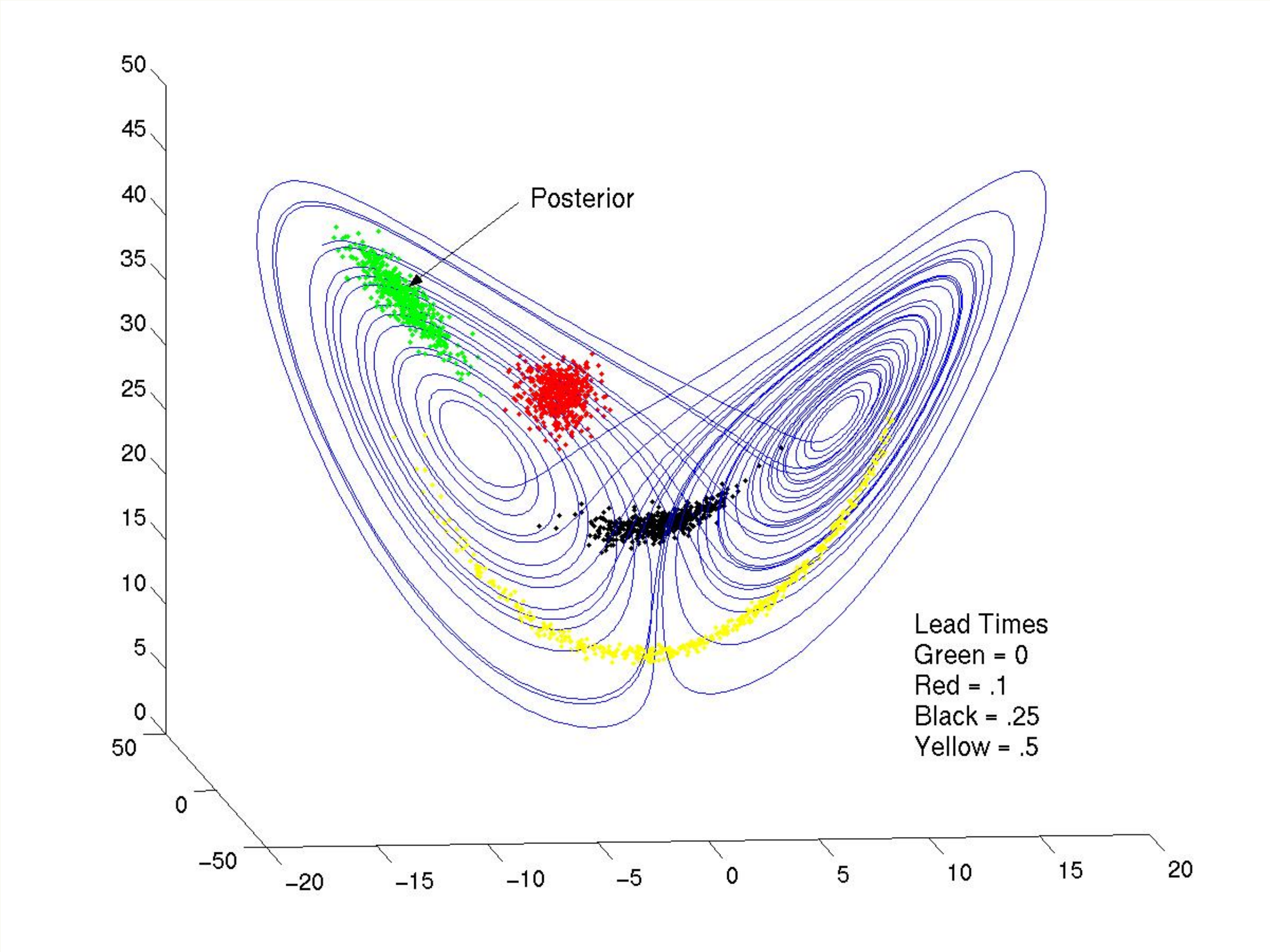
Effects of nonlinear dynamics

Nonlinear systems can generate assimilation problems where distributions of priors are not well represented by multivariate normals.

Lorenz '63 a simple three dimensional system

Easy to visualize, large distortion occurs for trajectories near the origin.
Ensembles are often distinctly non-Gaussian.

Following an ensemble through state space



The Bayes cycle

Observations at time t

$$\boldsymbol{y}_t = H\boldsymbol{x}_t + \text{measurement error}$$

System dynamics:

$\boldsymbol{x}_{t+1} = g(\boldsymbol{x}_t)$ (deterministic)

$$p(\mathbf{x}_t),\ \mathbf{y}_t \xrightarrow{\text{Bayes}} p(\mathbf{x}_t|\mathbf{y}_t) \xrightarrow{g(.)} p(\mathbf{x}_{t+1}|\mathbf{y}_t) = p(\mathbf{x}_{t+1}),\ \mathbf{y}_{t+1}$$

data update forecast new data

Yesterday's posterior becomes today's prior!

Some key ideas:

- Represent a continuous distribution by a discrete sample.
- Non-Gaussian update for components of the state vector *local* to the observation locations and close to the observation value.
- Parts of state vector far from the observation are updated using Ensemble Kalman Filter.

Standard Kalman Filter/ conditional multivariate normal distributions

This is easy in closed form if everything is $multivariate\ normal$ and $linear$.

$Observation\ Model$

$$\boldsymbol{y} = H\boldsymbol{x}_t + \boldsymbol{e} \quad \text{with} \quad \boldsymbol{e} \sim MN(0, R)$$

$Prior$

$$\boldsymbol{x}_t \sim MN(\boldsymbol{\mu}_t, \boldsymbol{P})$$

Standard Kalman Filter/ conditional multivariate normal distributions

This is easy in closed form if everything is *multivariate normal* and *linear*.

Observation Model

$$\boldsymbol{y} = H\boldsymbol{x}_t + \boldsymbol{e} \quad \text{with} \quad \boldsymbol{e} \sim MN(0, R)$$

Prior

$$\boldsymbol{x}_t \sim MN(\boldsymbol{\mu}_t, \boldsymbol{P})$$

Kalman update for state

$$\hat{\boldsymbol{x}}_t = E(\boldsymbol{x}_t|\boldsymbol{y}) = \boldsymbol{\mu}_t + \boldsymbol{P}H^T(H\boldsymbol{P}H^T + R)^{-1}(\boldsymbol{y} - H\boldsymbol{\mu}_t)$$

The Ensemble KF

- All means and covariances in Kalman Filter are replaced by sample quantities found from the ensemble.
- The sample covariance matrix from the ensemble is tapered spatially to regularized the estimate.
- Essentially the ensemble encodes a low rank approximation to the mean and covariance following the $exact$ calculation under the assumption that everything is multivariate normal.

Forecast step with EnKF

In place of

$$p(\boldsymbol{x}_t|y_t) \rightarrow p(g(\boldsymbol{x}_t)|y_t)$$

propagate each ensemble member.

$$\begin{array}{ccc} \boldsymbol{x}_{t,1} & & g(\boldsymbol{x}_{t,1}) = \boldsymbol{x}_{t+1,1} \\ \boldsymbol{x}_{t,2} & g \longrightarrow & g(\boldsymbol{x}_{t,2}) = \boldsymbol{x}_{t+1,2} \\ \vdots & & \vdots \\ \boldsymbol{x}_{t,M} & & g(\boldsymbol{x}_{t,M}) = \boldsymbol{x}_{t+1,M} \end{array}$$

Note:

With no model error, the relationship among state vectors is preserved correctly. The dynamics generates information ...

Ensemble Kalman filter and regression

Suppose we observe just the Jth state component with noise

$$Y = x_J + e$$

Update the i^{th} ensemble member:

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + COV(\boldsymbol{x}, x_J)VAR(Y)^{-1}(Y - x_{i,J}^f - \text{perturbation})$$

Ensemble Kalman filter and regression

Suppose we observe just the Jth state component with noise

$$Y = x_J + e$$

Update the i^{th} ensemble member:

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + COV(\boldsymbol{x}, x_J) VAR(Y)^{-1}(Y - x_{i,J}^f - \text{perturbation})$$

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + \frac{COV(\boldsymbol{x}, x_J)}{VAR(x_J)} \frac{VAR(x_J)}{VAR(Y)} \times (Y - x_{i,J}^f - \text{perturbation})$$

Ensemble Kalman filter and regression

Suppose we observe just the Jth state component with noise

$$Y = x_J + e$$

Update the i^{th} ensemble member:

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + COV(\boldsymbol{x}, x_J) VAR(Y)^{-1} (Y - x_{i,J}^f - \text{perturbation})$$

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + \frac{COV(\boldsymbol{x}, x_J)}{VAR(x_J)} \frac{VAR(x_J)}{VAR(Y)} \times (Y - x_{i,J}^f - \text{perturbation})$$

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + \frac{COV(\boldsymbol{x}, x_J)}{VAR(x_J)} \quad \times (\text{ draw from } \quad p(x_J | Y) - x_{i,J}^f)$$

Ensemble Kalman filter and regression

Suppose we observe just the Jth state component with noise

$$Y = x_J + e$$

Update the i^{th} ensemble member:

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + COV(\boldsymbol{x}, x_J) VAR(Y)^{-1}(Y - x_{i,J}^f - \text{perturbation})$$

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + \frac{COV(\boldsymbol{x}, x_J)}{VAR(x_J)} \frac{VAR(x_J)}{VAR(Y)} \times (Y - x_{i,J}^f - \text{perturbation})$$

$$\boldsymbol{x}_i^u = \boldsymbol{x}_i^f + \frac{COV(\boldsymbol{x}, x_J)}{VAR(x_J)} \quad \times(\text{ draw from } \quad p(x_J|Y) - x_{i,J}^f)$$

Now substitute sample COV and sample VAR to get a linear regression.

Non-Gaussian distributions

Represent the prior distributions as mixtures of multivariate normals
$p(\mathbf{x}_t) = \Sigma_{i=1}^{k} p_i \mathsf{MN}(\boldsymbol{\mu}_i, \mathbf{P}_i)$

The posterior distribution is also a mixture:
$p(\mathbf{x}_t|\mathbf{y}_t) = \Sigma_{i=1}^{k} p_i^\star \mathsf{MN}(\boldsymbol{\mu}_i^\star, \mathbf{P}_i^\star)$

Each component is just the usual KF or EnKF update!

Non-Gaussian distributions

Represent the prior distributions as mixtures of multivariate normals
$p(\mathbf{x}_t) = \Sigma_{i=1}^{k} p_i \mathsf{MN}(\boldsymbol{\mu}_i, \mathbf{P}_i)$

The posterior distribution is also a mixture:
$p(\mathbf{x}_t|\mathbf{y}_t) = \Sigma_{i=1}^{k} p_i^\star \mathsf{MN}(\boldsymbol{\mu}_i^\star, \mathbf{P}_i^\star)$

Each component is just the usual KF or EnKF update!

Key aspect is the update of the mixture probability

$$p_i^\star \sim \frac{e^{-(y-H\mu_i)^T (H\mathbf{P}_i H^T + R)^{-1} (y-H\mu_i)}}{|(H\mathbf{P}_i H^T + R)|^{1/2}}$$

p_i is large if $H\mu_i$ is close to the observed value.

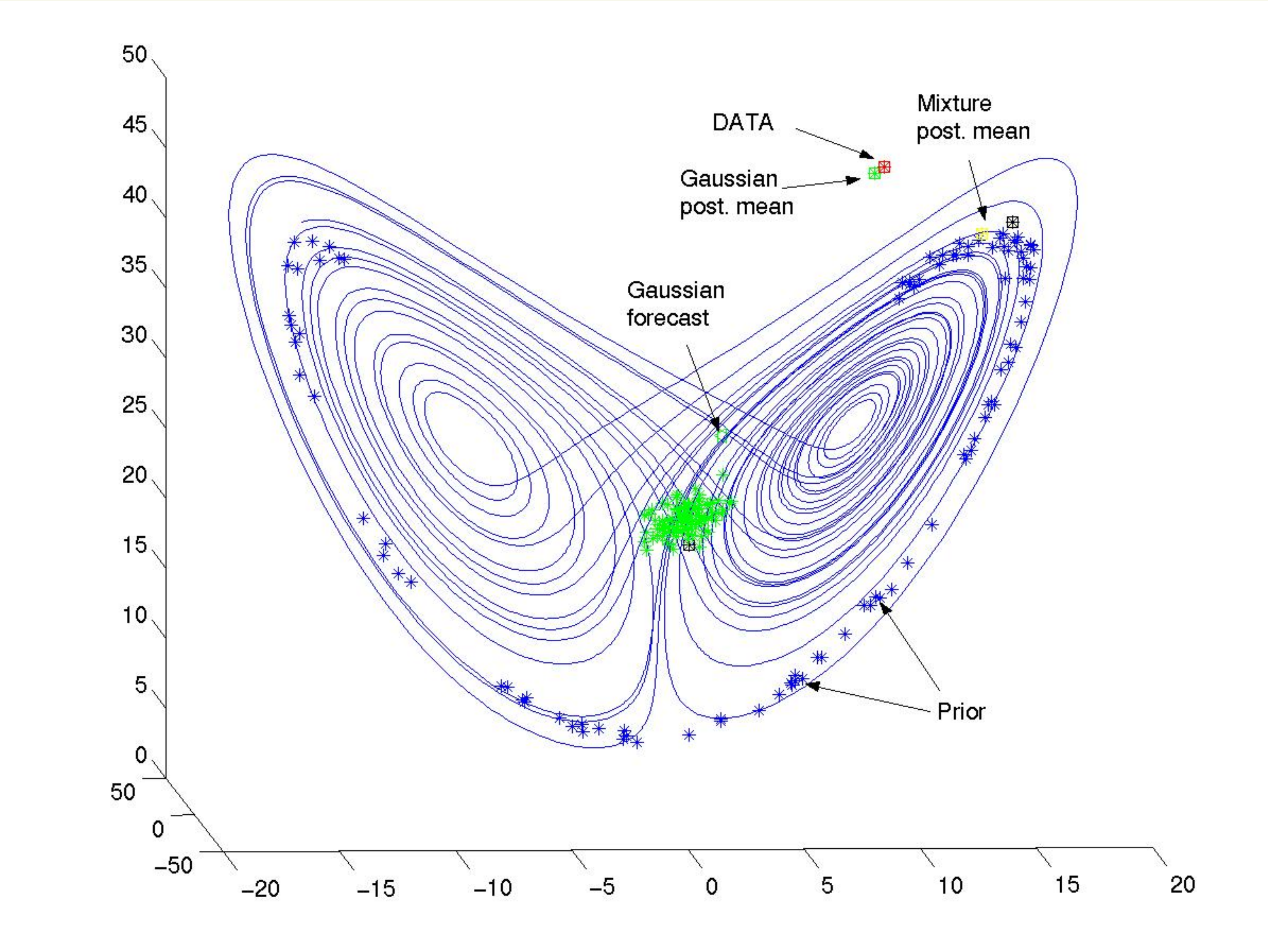
Ensembles as a mixture distribution:

A random subset of ensemble members are the centers of the mixture. ($\boldsymbol{\mu}_k$).

The sample covariance of nearest neighbors about the center is the covariance ($\boldsymbol{P}_k$).

The posterior probabilities look like weights based on a normal kernel. The use of neighborhoods to find the covariance results in a local linear regression.

A non-Gaussian example



Results for Lorenz '63

With a time step of .5 and observation standard deviation of 2 in the saddle region of the state space
(400 ensemble members, 100 centers, 25 Nearest neighbors.)

RMSE Ensemble KF filter $= 1.64$
RMSE Mixture filter $= .73$

Extensions to larger state spaces

The mixture filter breaks down as the dimension of the state vector increases.

e.g. the posterior probability concentrates on a single member or is small

The Local-Local Filter

LOCAL in physical space:
Only update components close to the observation location.

Call these local state components $\boldsymbol{x}_L$

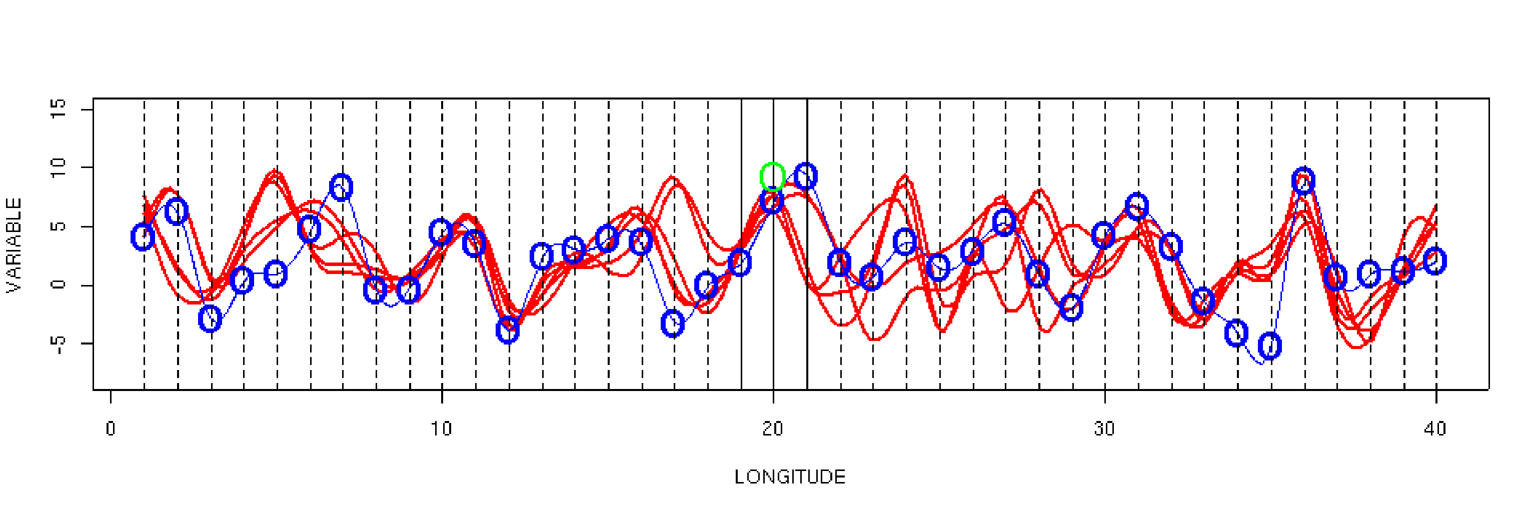
LOCAL in state space:
only use ensemble members that are "close" to the observed value.

Hybrid filter

Update remaining components using EnKF.

Call these remaining (global) components $\boldsymbol{x}_G$

A 40 dimensional system: Lorenz '96



True state

Single Observation at 20

5 ensemble members (as lines)

Local state components: $19, 20, 21$

Global components of state: the rest!

How do we splice two types of solutions together?

Our first idea

Find the global part conditional on first finding the local posterior.

$$p(\boldsymbol{x}|y) = p(\boldsymbol{x}_G|\boldsymbol{x}_L)p(\boldsymbol{x}_L|y)$$

$p(\boldsymbol{x}_L|y)$ from local-local (non-Gaussian) filter

$p(\boldsymbol{x}_G|\boldsymbol{x}_L)$ assuming Gaussian distributions.

How do we splice two types of solutions together?

Our first idea

Find the global part conditional on first finding the local posterior.

$$p(\boldsymbol{x}|y) = p(\boldsymbol{x}_G|\boldsymbol{x}_L)p(\boldsymbol{x}_L|y)$$

$p(\boldsymbol{x}_L|y)$ from local-local (non-Gaussian) filter

$p(\boldsymbol{x}_G|\boldsymbol{x}_L)$ assuming Gaussian distributions.

This did not work ...

What worked as a hybrid filter.

Use posterior means from $p(\boldsymbol{x}_G|\boldsymbol{x}_L)$ and $p(\boldsymbol{x}_L|y)$

The non Gaussian gives good point predictions.

What worked as a hybrid filter.

Use posterior means from $p(\boldsymbol{x}_G|\boldsymbol{x}_L)$ and $p(\boldsymbol{x}_L|y)$
The non Gaussian gives good point predictions.

Reverse conditioning! $[\boldsymbol{x}_L|\boldsymbol{x}_G]$
Posterior for $\boldsymbol{x}_G$ is just $p(\boldsymbol{x}_G|Y)$

Posterior draw for $\boldsymbol{x}_L$ is (up to the mean)

$$E[\boldsymbol{x}_L|\boldsymbol{x}_G] + [A\boldsymbol{x}_L|y]$$

The matrix A is chosen so that complete ensemble agrees with posterior covariance for EnKF.

Summary

Results

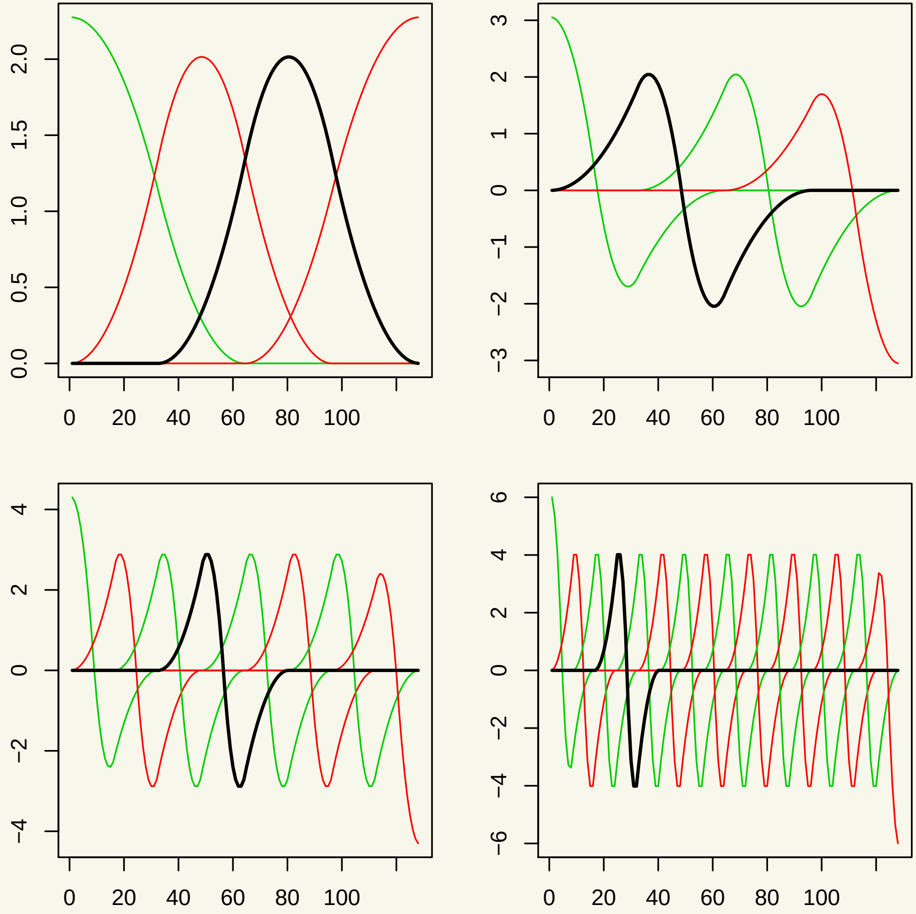
- We have some evidence that the practical version of the EnKF actually handles non-Gaussian distributions better than an exact Kalman filter.
- The Local-Local filter clearly out performs EnKF in a simple 3-d system especially in places where g is very nonlinear.
- A version of the L-L filter also performs better than EnKF with about 5% improvement (without any extensive tuning) for the 40 variable model.

Issues

- The local linear fitting seems important but it is hard to beat the EnKF
- The spread of the ensemble may have nothing to do with Bayes; All that is important is to generate a good regression relationship.
- RMSE as a criterion has little to do with non-Gaussian distributions.
- Are components far away from observations more amenable to a Gaussian update?

Better models for covariance functions

A set of 32 wavelet basis functions



Sparseness with wavelets

Decomposition of a 1-d covariance matrix