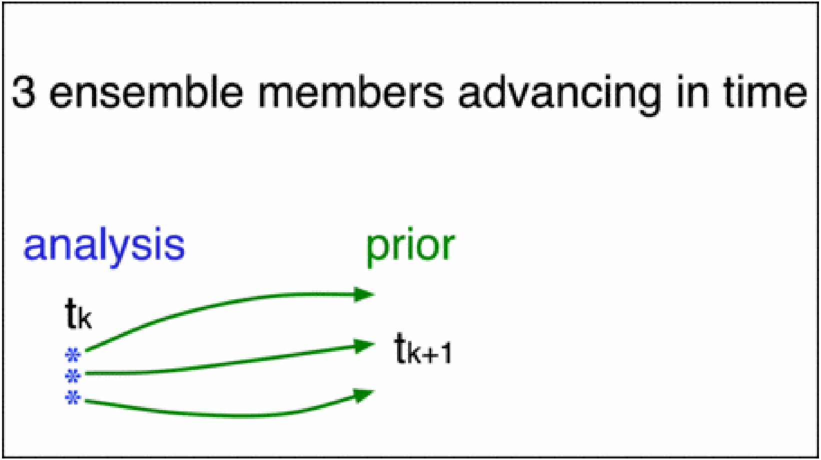
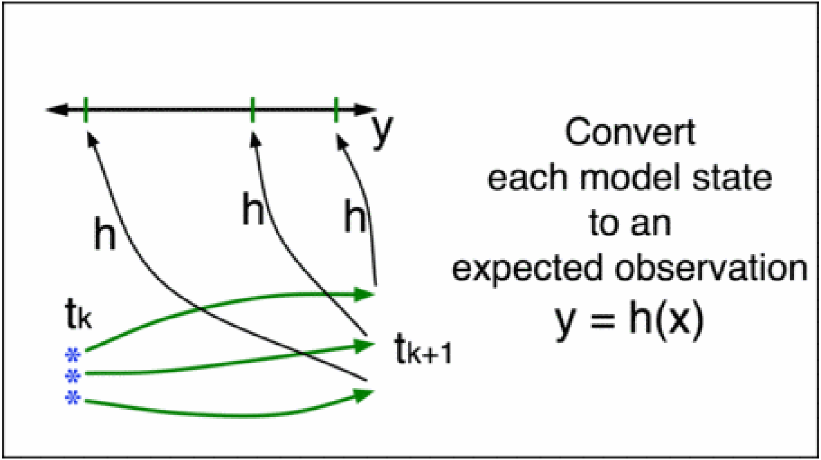
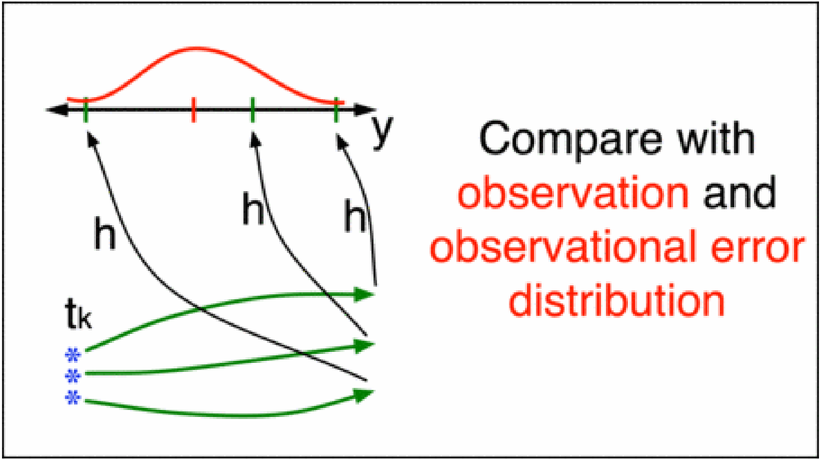
How does the regression step in the two-step EnKF connect to Bayesian estimation?

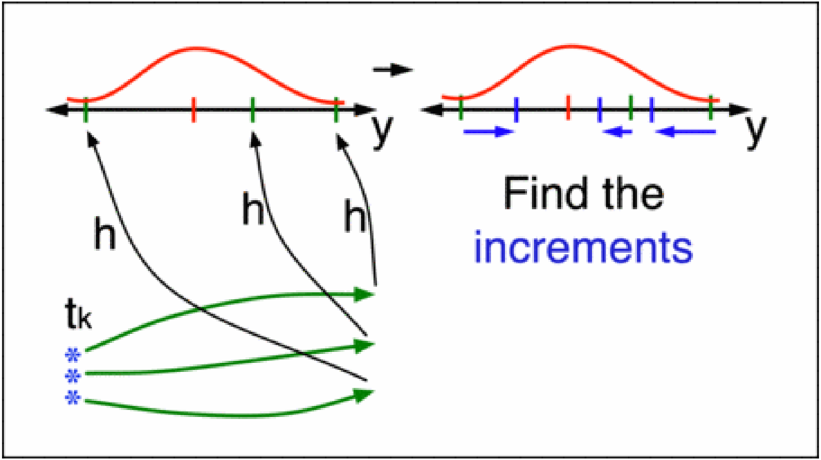
Ian Grooms

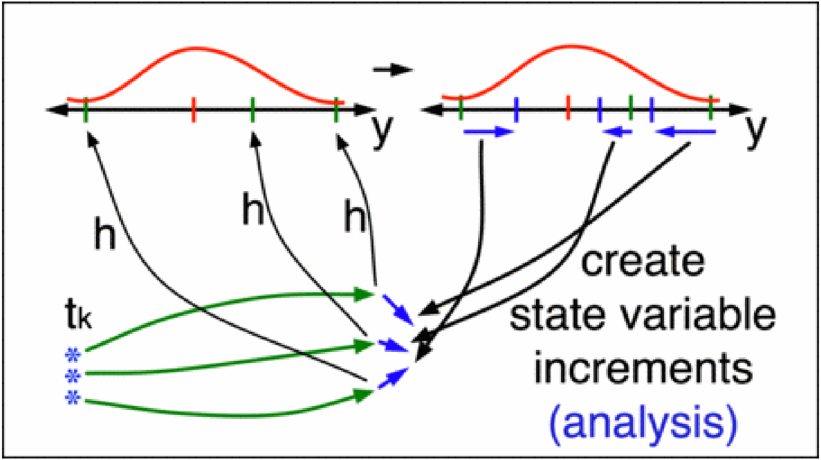
EnKF Workshop 2022

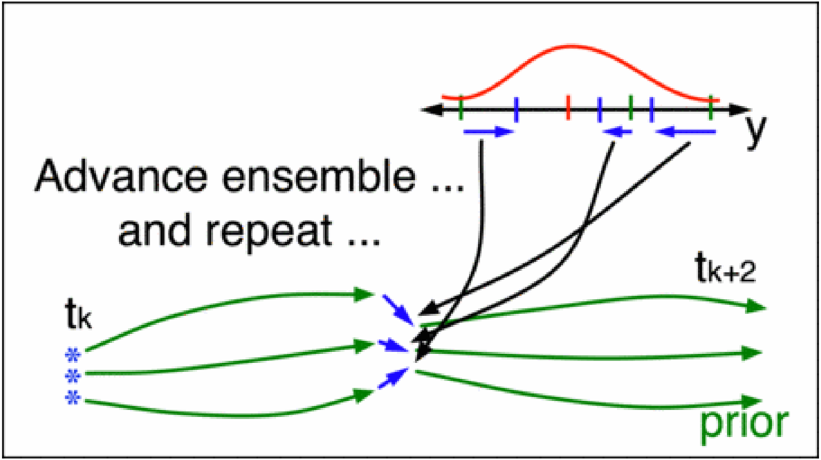

3 ensemble members advancing in time
analysis
prior
t_k
t_{k+1}


y
h
h
h
tk
tk+1
Convert
each model state
to an
expected observation
y = h(x)


y
h
h
h
tk
Compare with
observation and
observational error
distribution


y
y
h
h
h
t_k
Find the
increments


h
h
h
y
y
tk
create
state variable
increments
(analysis)


y
Advance ensemble ...
and repeat ...
tk
tk+2
prior

If

- The update in observation space is Gaussian and
- The regression is linear and estimated using ordinary least squares (OLS)

Then this two-step is equivalent to a one-step EnKF where

$$\mathbf{BH}^T = \frac{1}{N-1}\sum_{n=1}^{N}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{H}(\boldsymbol{x}_i) - \overline{\boldsymbol{H}(\boldsymbol{x})})^T$$

$$\mathbf{HBH}^T = \frac{1}{N-1}\sum_{n=1}^{N}(\boldsymbol{H}(\boldsymbol{x}_i) - \overline{\boldsymbol{H}(\boldsymbol{x})})(\boldsymbol{H}(\boldsymbol{x}_i) - \overline{\boldsymbol{H}(\boldsymbol{x})})^T$$

Is this just a different way of implementing an EnKF?

If we replace EAKF or linear regression, do we lose the connection to Bayesian estimation?

We have two random variables X and Y whose joint pdf is denoted

$$[\,x,\,y\,]$$

Our goal is to sample from the conditional distribution

$$[\,x\,|\,Y = y_o\,]$$

which can also be written as a Bayesian posterior

$$[\,x\,|\,Y = y_o\,] = \frac{[\,Y = y_o\,|\,x\,]}{[Y = y_o]}[\,x\,]$$

(y_o is the actual value of the observation; y denotes any value that the random variable Y could take.)

Next, introduce a new random variable $\boldsymbol{Z}$. With this new random variable we have a new joint distribution

$$[\boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}\,].$$

The old one is just the marginal

$$[\boldsymbol{x},\, \boldsymbol{y}\,] = \int [\boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}\,]\mathrm{d}\boldsymbol{z}.$$

The posterior that we care about is just the marginal of a new posterior:

$$[\boldsymbol{x} \,|\, \boldsymbol{Y} = \boldsymbol{y}_o\,] = \int [\boldsymbol{x},\, \boldsymbol{z} \,|\, \boldsymbol{Y} = \boldsymbol{y}_o\,]\mathrm{d}\boldsymbol{z} = \int \frac{[\,\boldsymbol{Y} = \boldsymbol{y}_o \,|\, \boldsymbol{x},\, \boldsymbol{z}\,]}{[\boldsymbol{Y} = \boldsymbol{y}_o]}[\boldsymbol{x},\, \boldsymbol{z}\,]\mathrm{d}\boldsymbol{z}.$$

For the general two-step framework we must make the following assumption about $\boldsymbol{Z}$:

$$[\,\boldsymbol{Y} = \boldsymbol{y}_o \,|\, \boldsymbol{x},\, \boldsymbol{z}\,] = [\,\boldsymbol{Y} = \boldsymbol{y}_o \,|\, \boldsymbol{z}].$$

In the standard DART two-step, we have $\boldsymbol{Y} = \boldsymbol{H}(\boldsymbol{X}) + \boldsymbol{\epsilon}$ and usually set $\boldsymbol{Z} = \boldsymbol{H}(\boldsymbol{X})$.

We **do not** need an observation model of the form $\boldsymbol{Y} = \boldsymbol{H}(\boldsymbol{X}) + \boldsymbol{\epsilon}$. It's just mentioned here for illustration.

Use Bayes' rule to expand the posterior inside the integral, and use our one assumption about $\boldsymbol{Z}$:

$$\begin{aligned}[\,\boldsymbol{x}\,|\,\boldsymbol{Y}=\boldsymbol{y}_o\,] &= \int \frac{[\,\boldsymbol{Y}=\boldsymbol{y}_o\,|\,\boldsymbol{x},\,\boldsymbol{z}\,]}{[\boldsymbol{Y}=\boldsymbol{y}_o]}[\boldsymbol{x},\,\boldsymbol{z}\,]\mathrm{d}\boldsymbol{z} \\ &= \int \frac{[\,\boldsymbol{Y}=\boldsymbol{y}_o\,|\,\boldsymbol{z}\,]}{[\boldsymbol{Y}=\boldsymbol{y}_o]}[\,\boldsymbol{x},\,\boldsymbol{z}\,]\mathrm{d}\boldsymbol{z}.\end{aligned}$$

Now expand the prior $[\,\boldsymbol{x},\,\boldsymbol{z}\,]$ as marginal times conditional

$$[\,\boldsymbol{x}\,|\,\boldsymbol{Y}=\boldsymbol{y}_o\,] = \int \left(\frac{[\,\boldsymbol{Y}=\boldsymbol{y}_o\,|\,\boldsymbol{z}\,]}{[\boldsymbol{Y}=\boldsymbol{y}_o]}[\,\boldsymbol{z}\,]\right)[\,\boldsymbol{x}\,|\,\boldsymbol{z}\,]\mathrm{d}\boldsymbol{z}.$$

$$[\, \boldsymbol{x} \,|\, \boldsymbol{Y} = \boldsymbol{y}_o \,] = \int \left(\frac{[\, \boldsymbol{Y} = \boldsymbol{y}_o \,|\, \boldsymbol{z} \,]}{[\boldsymbol{Y} = \boldsymbol{y}_o]} [\, \boldsymbol{z} \,] \right) [\, \boldsymbol{x} \,|\, \boldsymbol{z} \,] \mathrm{d}\boldsymbol{z}.$$

Notice that the quantity in parenthesis is a posterior distribution

$$\frac{[\, \boldsymbol{Y} = \boldsymbol{y}_o \,|\, \boldsymbol{z} \,]}{[\boldsymbol{Y} = \boldsymbol{y}_o]} [\, \boldsymbol{z} \,] = [\, \boldsymbol{z} \,|\, \boldsymbol{Y} = \boldsymbol{y}_o \,]$$

so

$$[\, \boldsymbol{x} \,|\, \boldsymbol{Y} = \boldsymbol{y}_o \,] = \int [\, \boldsymbol{x} \,|\, \boldsymbol{z} \,][\, \boldsymbol{z} \,|\, \boldsymbol{Y} = \boldsymbol{y}_o \,] \mathrm{d}\boldsymbol{z}.$$

How can we draw samples $\{\boldsymbol{x}_i^+\}_{i=1}^N$ from a distribution of this form?

Consider the following analogy. Suppose we have the following dynamics

$$\boldsymbol{X}^{k+1} = \boldsymbol{M}\left(\boldsymbol{X}^{k}\right) + \boldsymbol{W}^{k}$$

where $\boldsymbol{W}^k$ is a random variable with pdf $[\,\boldsymbol{w}\,]$. We know how to sample from $[\,\boldsymbol{X}^{k+1}\,]$.

First draw an ensemble $\{\boldsymbol{x}_i^k\}_{i=1}^N$ of samples of $\boldsymbol{X}^k$.

Then apply the dynamics $\boldsymbol{M}$ to each $\boldsymbol{x}_i^k$ and finally add a sample from the noise $\boldsymbol{w}_i^k$.

The pdf of $\boldsymbol{X}^{k+1}$ is the marginal distribution

$$[\,\boldsymbol{x}^{k+1}\,] = \int [\,\boldsymbol{x}^{k+1}, \boldsymbol{x}^{k}\,]\mathrm{d}\boldsymbol{x}^{k} = \int [\,\boldsymbol{x}^{k+1} \,|\, \boldsymbol{x}^{k}\,][\,\boldsymbol{x}^{k}\,]\mathrm{d}\boldsymbol{x}^{k}$$

The first step in sampling from this distribution is to sample from $[\,\boldsymbol{x}^{k}\,]$.

Then sample from $[\,\boldsymbol{x}^{k+1} \,|\, \boldsymbol{X}^{k} = \boldsymbol{x}^{k}_{i}\,]$.

GENERAL TWO-STEP SUMMARY

The two-step Bayesian update has exactly the same form. Recall that we want to sample from

$$[\,\boldsymbol{x}\,|\,\boldsymbol{Y}=\boldsymbol{y}_o\,] = \int [\,\boldsymbol{x}\,|\,\boldsymbol{z}\,][\,\boldsymbol{z}\,|\,\boldsymbol{Y}=\boldsymbol{y}_o\,]\mathrm{d}\boldsymbol{z}.$$

- Step 1: Generate an ensemble $\{\boldsymbol{z}_i^+\}_{i=1}^N$ from the posterior $[\,\boldsymbol{z}\,|\,\boldsymbol{Y}=\boldsymbol{y}_o\,]$.
- Step 2: Sample $\boldsymbol{x}_i^+$ from pdf $[\,\boldsymbol{x}\,|\,\boldsymbol{Z}=\boldsymbol{z}_i^+\,]$.

The difference between this case and the analogy is that in the analogy we know what the dynamics are, so we know how to sample from the conditional distribution.

This is where regression comes in. Propose a linear model of the form

$$\boldsymbol{X} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 Z + \boldsymbol{\eta}.$$

(Assume for the moment that Z is scalar to make the exposition easier.)

If $\boldsymbol{\eta}$ is Gaussian with zero mean and covariance $\boldsymbol{\Sigma}$, then we are saying

$$\boldsymbol{X}|Z = z \sim \mathcal{N}\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z, \boldsymbol{\Sigma}\right)$$

We can estimate the regression coefficients using OLS.

2-STEP ENKF SUMMARY

- Generate a prior ensemble $\{\boldsymbol{x}_i^-\}_{i=1}^N$.
- Generate a prior ensemble $\{\boldsymbol{z}_i^-\}_{i=1}^N$.
- Generate a posterior ensemble $\{\boldsymbol{z}_i^+\}_{i=1}^N$ using an EnKF.
- Estimate the regression coefficients as the OLS solution of the following system

$$\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_i^- = \boldsymbol{x}_i^-, \quad \boldsymbol{i} = 1, \ldots, N$$

- Generate samples $\boldsymbol{\eta}_i = \boldsymbol{x}_i^- - \boldsymbol{\beta}_0 - \boldsymbol{\beta}_1 \boldsymbol{z}_i^-$
- Set $\boldsymbol{x}_i^+ = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{z}_i^+ + \boldsymbol{\eta}_i$.

You can write this in incremental form as: $\Delta \boldsymbol{x}_i = \boldsymbol{\beta}_1 \Delta \boldsymbol{z}_i$.

Non-Gaussian Generalizations

The nice thing about the first step is that it is typically low-dimensional, so we can use methods to sample from $[\, z \,|\, Y = y_o \,]$ that work for non-Gaussian distributions but that might be impractical in higher dimensions. E.g.

- Particle Filters
- Gaussian Mixture Methods
- RHF: Anderson MWR 2010
- Gamma/Inverse-Gamma/Gaussian: Bishop QJRMS 2016
- Etc

Non-Gaussian Generalizations

Now that we know how the second step connects to the Bayesian problem, we can use advanced regression models

- General Linear Model $\boldsymbol{X} = \boldsymbol{\eta} + \sum_{j=1}^{J} \boldsymbol{\beta}_j \phi_j(\boldsymbol{Z})$
- Generalized Linear Model $\boldsymbol{g}(\boldsymbol{X}) = \boldsymbol{\eta} + \sum_{j=1}^{J} \boldsymbol{\beta}_j \phi_j(\boldsymbol{Z})$. (E.g. Anderson Rank Regression 2019)
- Nonlinear Models (e.g. neural nets) $\boldsymbol{g}(\boldsymbol{X};\, \boldsymbol{\beta}_g) = \boldsymbol{\eta} + \boldsymbol{f}(\boldsymbol{Z}; \boldsymbol{\beta}_f)$

The assumptions about $\boldsymbol{\eta}$ determine the form of the conditional distribution $[\,\boldsymbol{x} \,|\, \boldsymbol{z}\,]$, which also indicates the form of the objective function that must be minimized to find the unknown parameters.

I set up Lorenz-96 observing *all* variables every 0.05 model time units, with three different observation models:

$$\text{Linear: } \boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{\epsilon}$$

$$\text{Logit-Normal: } \boldsymbol{Y} = \frac{1}{1 + \exp\{0.5 \times (\boldsymbol{X} - 2.5) + \boldsymbol{\epsilon}\}}$$

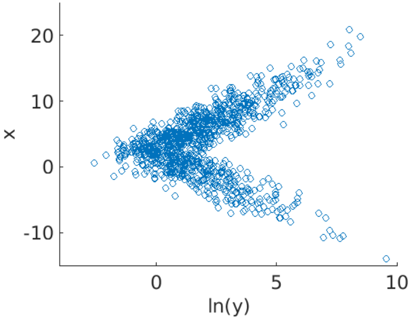
$$\text{Log-Normal: } \boldsymbol{Y} = \exp\left\{0.5 \times |\boldsymbol{X} - 2.5| + \boldsymbol{\epsilon}\right\}$$

where $\boldsymbol{\epsilon}$ are standard normal.

I won't show results for the linear obs case.

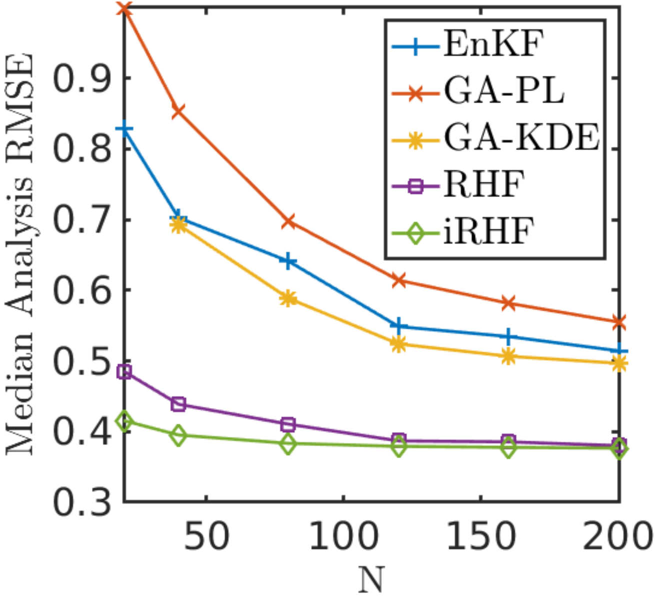
I used multiplicative prior inflation and localization, both tuned to produce optimal results.

For the log-normal observations the likelihood is bimodal. In the standard approach ($Z = H(\boldsymbol{X})$) the second step requires fitting a regression to this kind of data

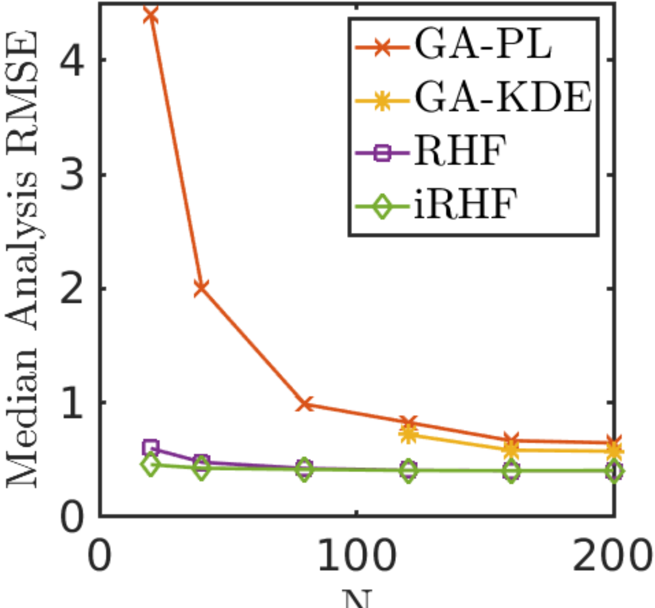


By choosing $Z = X$, I don't have to do regression through this kind of scatterplot. All the nonlinearity/bimodality is handled in the first step.

Logit-Normal Results



LOG-NORMAL RESULTS



More details about iRHF along with a comparative discussion of Gaussian Anamorphosis methods can be found in

Grooms, "A comparison of nonlinear extensions to the ensemble Kalman filter" Computational Geosciences, 2022.