Introduction to Data Assimilation

Ronald M. Errico

Goddard Earth Sciences Technology and Research Center
Morgan State University

Global Modeling and Assimilation Office
GSFC NASA

A presentation about concepts rather than techniques

Some characteristics of NWP

1. It is an initial value problem: $\mathbf{x}_t = M(\mathbf{x}_0, \mathbf{b}, \boldsymbol{\beta})$.
2. The representation **x** is generally incomplete, and sometimes unclear.
3. Errors in model formulation ("model error") or in model input generally create forecast error.
4. Errors can change shape, magnitude, or field as time progresses.
5. Nonlinearity implies the possibility of chaos (Lorenz).

Due to chaotic dynamics and physics, small differences in model input generally, eventually, grow in time until the original and perturbed forecasts are as different as two randomly chosen states from the possible states of the system (with caveats).

6. Information is required to estimate the initial conditions

Available Information

1. Observations ($\mathbf{y}^{o}$)
   a. incomplete in time and space
   b. indirect
   c. imperfect
2. Physics and dynamics
   a. diagnostic constraints between fields (e.g., balances; $G(\mathbf{x})=0$)
   b. relationships between what is observed and what is analyzed ($\mathbf{y}=H(\mathbf{x})$)
   c. connect states in time
   d. imperfect
3. Prior information ($\mathbf{x}^{f}$)
   a. generally available on analysis grid
   b. generally provided by a short-term forecast (temporal extrapolation of previously analyzed observations
   c. imperfect (although generally a best available estimate)
   d. errors have significant correlations
   e. mathematically similar to observations, albeit with different properties

Goal of data assimilation

1. Determine a “best” estimate of the (instantaneous) state of the system
2. Consider the relative inadequacies of the available useful information.
3. Consider that realizations of errors are unknown.
4. Consider available estimates of statistics of input information
5. Consider that the analysis will also be imperfect
6. Minimize some measure of the analysis error or maximize some probability that the estimated state is correct.

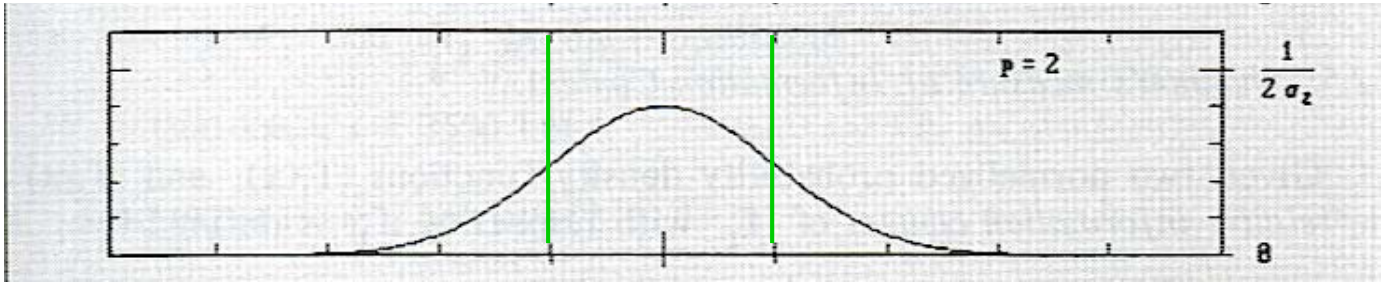
Data Assimilation is a fundamentally statistical problem!

"The most general way of describing information:"

As a probability density function (PDF or pdf)

Probabilities described by PDFs ρ:

$$P(w_1 \leq w \leq w_2 \,|z) = \int_{w_1}^{w_2} \rho(w \,|z)\; dw$$



Some basic probability relationships and Bayes Theorem (1763)

$$\rho(x,y) \;=\; \rho(x\;|y)\;\rho(y) \;=\; \rho(y\;|x)\;\rho(x)$$

$$\rho(y) = \int_x \rho(y|x)\;\rho(x)\;dx$$

$$\rho(x|y) = \frac{\rho(y|x)\;\rho(x)}{\int_x \rho(y|x)\;\rho(x)\;dx}$$

$$\int_x \rho(x|y)\;dx = 1$$

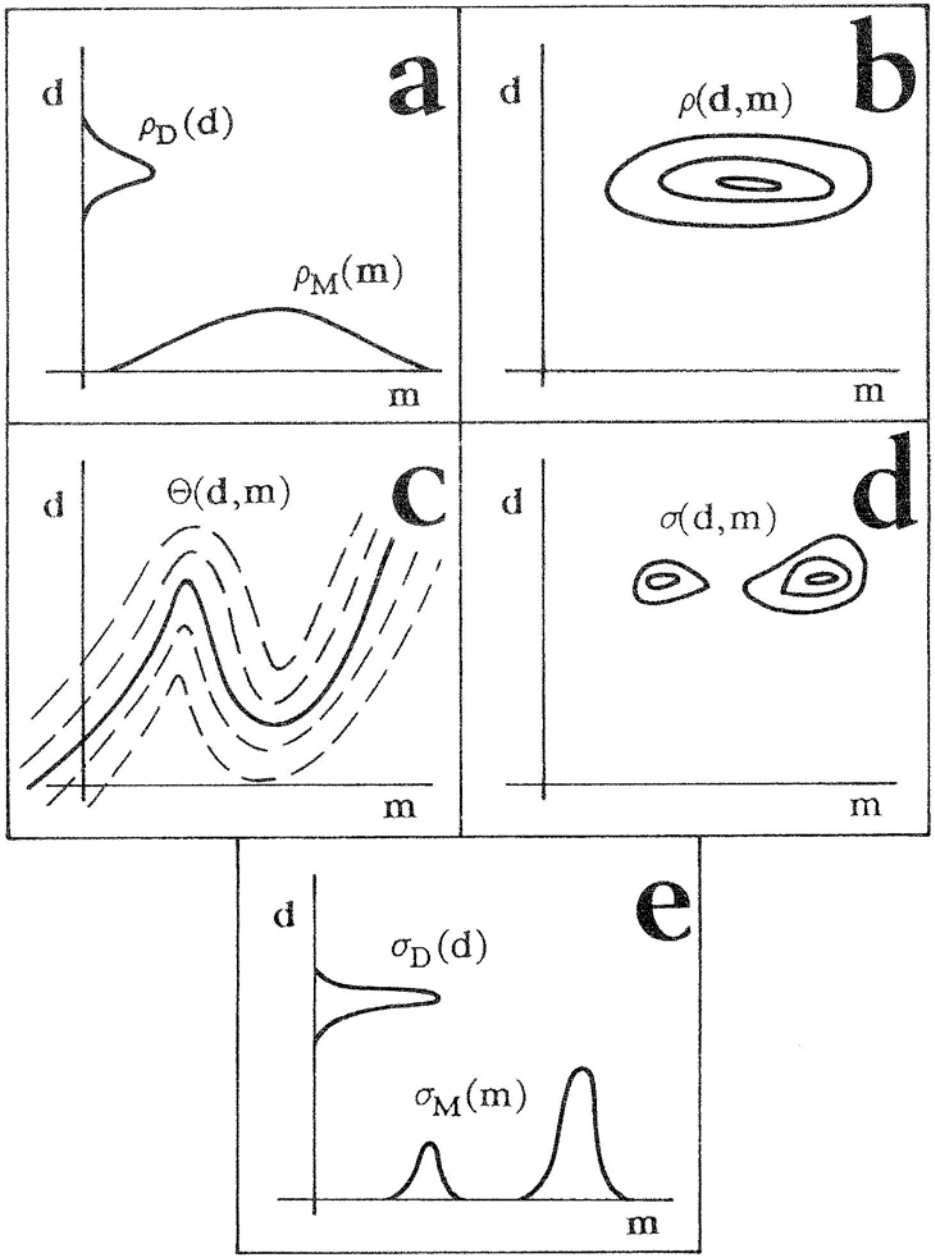
PDFs of the Information in our DA problem

$\mathbf{y}$	data ("true" in contrast to observed values)
$\mathbf{y}^o$	observations
$\mathbf{x}$	gridded fields to be analyzed
$\mathbf{x}^p$	prior estimates of gridded fields
$H(\mathbf{x})$	imperfect model relating $\mathbf{x}$ to $\mathbf{y}$

Information from observations:	$\rho_o(\mathbf{y}^o\|\mathbf{y})$
Information from models:	$\rho_H(\mathbf{y}\|\mathbf{x})$
Information from prior:	$\rho_p(\mathbf{x})$

$$\rho_d(\mathbf{y}^o|\mathbf{x}) = \int_Y \rho_o(\mathbf{y}^o|\mathbf{y})\ \rho_H(\mathbf{y}|\mathbf{x})\ d\mathbf{y}$$

$$\begin{aligned}\rho_a(\mathbf{x}|\mathbf{y}^o) &= \text{const} \times \rho_p(\mathbf{x})\ \rho_d(\mathbf{y}^o|\mathbf{x}) \\ &= \rho(\mathbf{x}; \mathbf{y}^o, \mathbf{x}^p, H(\mathbf{x}))\end{aligned}$$



Illustrative example from
Tarantola 1987
(Fig 1.10, page 54)

His d = our y
His m = our x
His Θ = our H

Implications of the Bayesian Approach

1. Unless the underlying distributions are simple, the problem is computationally intractable for large problems.
2. We see how the different information should be optimally combined.
3. We see what statistical knowledge is required as input.
4. Results may depend on shapes of distributions, not only their means and variances.
5. We see that selection of a “best” analysis can be somewhat ambiguous.
6. Multi-modality of the PDF can occur, particularly due to model non-linearity.
7. Any analysis has associated error statistics.
8. While an explicit Bayesian approach may be impractical, the Bayesian implications of other techniques should be considered.

Some comments about DA

1. Data assimilation serves as a filter (of errors) and as a smoother (interpolator) of information

2. Presumably, it is called “data assimilation” because data is being assimilated into a model, but this thinking has lead to confusion.

3. All observations do not improve an analysis: Even good data can be spread poorly (both theoretical and numerical experimentation suggest that maybe 46-48% of good observations make the analysis worse).

4. Computational practicality is a critical constraint.

For Gaussian PDFs

For $\mathbf{x}$ of size n and $\mathbf{y}$ of size m

$$\rho_o(\mathbf{y}^o|\mathbf{y}) = \left[(2\pi)^m \,\mathrm{Det}\,(\mathbf{E})\right]^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{y}^o - \mathbf{y}\right)^T \mathbf{E}^{-1}\left(\mathbf{y}^o - \mathbf{y}\right)\right]$$

$$\rho_b(\mathbf{x}) = \left[(2\pi)^n \,\mathrm{Det}\,(\mathbf{B})\right]^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{x} - \mathbf{x}^f\right)^T \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^f\right)\right]$$

$$\rho_H(\mathbf{y}|\mathbf{x}) = \left[(2\pi)^m \,\mathrm{Det}\,(\mathbf{F})\right]^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{y} - H(\mathbf{x})\right)^T \mathbf{F}^{-1}\left(\mathbf{y} - H(\mathbf{x})\right)\right]$$

$$\rho_d(\mathbf{y}^o|\mathbf{x}) = \int_Y \rho_o(\mathbf{y}^o|\mathbf{y})\rho_H(\mathbf{y}|\mathbf{x})\; d\mathbf{y}$$

$$= \left[(2\pi)^m \,\mathrm{Det}\,(\mathbf{R})\right]^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\mathbf{y}^o - H(\mathbf{x})\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^o - H(\mathbf{x})\right)\right]$$

$$\mathbf{R} = \mathbf{E} + \mathbf{F}$$

What is Representativeness Error ?

Representativeness error is a realization of the uncertainty in our estimation (or ability to represent) an observation (a component of **y**) given our representation **x** of the state *x*, i.e., due to the fact that $\rho_H(\mathbf{y}|\mathbf{x})$ is not a delta function.

Representativess error has two primary sources:

1. Our representation **x** of the state *x* does not fully describe all we need to know about the atmosphere to determine **y** without probable error (resolution is one such issue here).

2. Our representation of the relationship between some observations (e.g., concerning radiances or precipitation) and the fields being analyzed (e.g., *T*, *q*, etc.) is imperfect.

Both of these together can be called representation or modelization errors if interpolation is essentially considered a model of spatial relationships.

A more common description of rep. error

$$y^t - H(\mathbf{x}^t) = e$$

1. This explicitly refers to a model *H* rather than a more general pdf relating *y* to **x.**
2. Mathematically, this is essentially a transformation of variables, with a pdf now required to describe the errors.
3. Note that *y* here refers to a perfectly measured value of an observation so that *e* does not include instrument error

Central Limit Theorem

When means of random samples are determined, those means will tend toward a normal (Gaussian) distribution as the sample sizes increase, even if the sample values themselves are not normally distributed, except for some unusual cases.

Furthermore, the variance of the distribution of the sample means is inversely proportional to the sample size.

Common Approximations

Data are either unbiased or their biases can be estimated and removed.
Correlations of errors for some data types can be ignored.
Balance can either be ignored or simply imposed.
Simple or linearized observation operators are adequate.
Multi-modality can be ignored, even when nonlinearity is present
After QC is applied, error distributions are Gaussian.
Error statistics are static.
Small ensembles are sufficient to estimate uncertainty.
Model error can be ignored.

Some are utilized because they are reasonable.
Some are motivated by the fact that we do not know the required statistics to consider them appropriately.
Others are motivated more by their computational efficiency.

Solution to the analysis problem for Gaussian errors and linear H

$$\rho_a(\mathbf{x}|\mathbf{x}^f, \mathbf{y}^o, H) = \text{const} \times \rho_{\mathrm{b}}(\mathbf{x}|\mathbf{x}^{\mathrm{f}})\ \rho_{\mathrm{d}}(\mathbf{y}^{\mathrm{o}}|\mathrm{H}(\mathbf{x}))$$

$$= \left[(2\pi)^n \,\mathrm{Det}\,(\mathbf{A})\right]^{\frac{1}{2}} \times$$

$$\exp\left\{-\frac{1}{2}\left(\left[\mathbf{x}-\mathbf{x}^f\right]^{\mathrm{T}}\mathbf{B}^{-1}\left[\mathbf{x}-\mathbf{x}^f\right] + \left[\mathbf{y}^o - H(\mathbf{x})\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y}^o - H(\mathbf{x})\right]\right)\right\}$$

The mean, median, and max. liklihood, and least squares estimate:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left[\mathbf{y}^o - \mathbf{H}\mathbf{x}^f\right]$$

where the Kalman Gain and analysis error covariance are

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

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

3DVAR

$$J(\mathbf{x}) = -\frac{1}{2}\left(\left[\mathbf{x}-\mathbf{x}^f\right]^{\mathrm{T}}\mathbf{B}^{-1}\left[\mathbf{x}-\mathbf{x}^f\right] + \left[\mathbf{y}^o - H(\mathbf{x})\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y}^o - H(\mathbf{x})\right]\right)$$
$$+ \; C(\mathbf{x}) + \ldots$$
$$J(\delta\mathbf{x}) = -\frac{1}{2}\left(\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \left[\mathbf{d}-\mathbf{H}\delta\mathbf{x}\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{d}-\mathbf{H}\delta\mathbf{x}\right]\right) + C'(\delta\mathbf{x}) + \ldots$$

where

$$\begin{aligned}
\delta\mathbf{x} &= \mathbf{x}-\mathbf{x}^f \\
H\mathbf{x} &= H(\mathbf{x}+\delta\mathbf{x}) \approx H(\mathbf{x}^f) + \mathbf{H}\delta\mathbf{x} \\
\mathbf{H} &= \left[\partial H(\mathbf{x})/\partial\mathbf{x}\right]_{\mathbf{x}^f} \\
\mathbf{d} &= \mathbf{y}^o - H(\mathbf{x}^f)
\end{aligned}$$

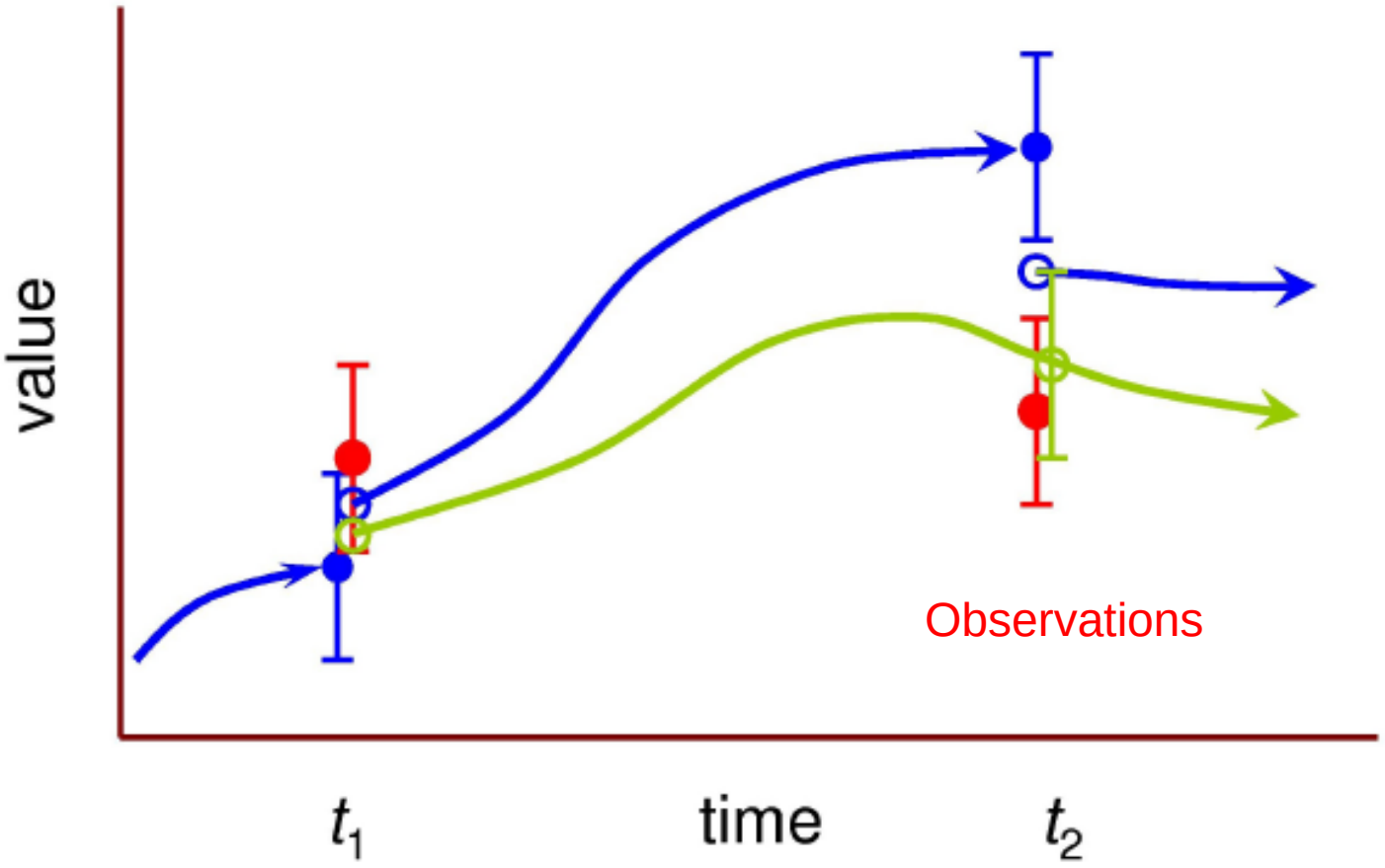
$J(\delta\mathbf{x})$ is a minimum when $\partial J/\partial\delta\mathbf{x} = \mathbf{0}$

$$\frac{\partial J}{\partial\delta\mathbf{x}} = -\mathbf{B}^{-1}\delta\mathbf{x} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{d}-\mathbf{H}\delta\mathbf{x}\right] + \frac{1}{2}\frac{\partial C'}{\partial\delta\mathbf{x}} + \ldots$$

In 4DVAR:
H=HM

3DVAR and 4DVAR

Forecast Trajectories



Picture from Jeff Kepert

The Kalman Filter

For linear $H(\mathbf{x}) = \mathbf{H}\mathbf{x}$, $\rho(\mathbf{x})$ is a maximum when $J(\mathbf{x})$ is a minimum:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T \left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1} \left[\mathbf{y}^o - \mathbf{H}\mathbf{x}^b\right]$$

$$\rho(\mathbf{x}) = \text{ constant } \times \text{ exp}\left[-\frac{1}{2}\left(\mathbf{x} - \mathbf{x}^a\right)^T \mathbf{A}^{-1} \left(\mathbf{x} - \mathbf{x}^a\right)\right]$$

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$$

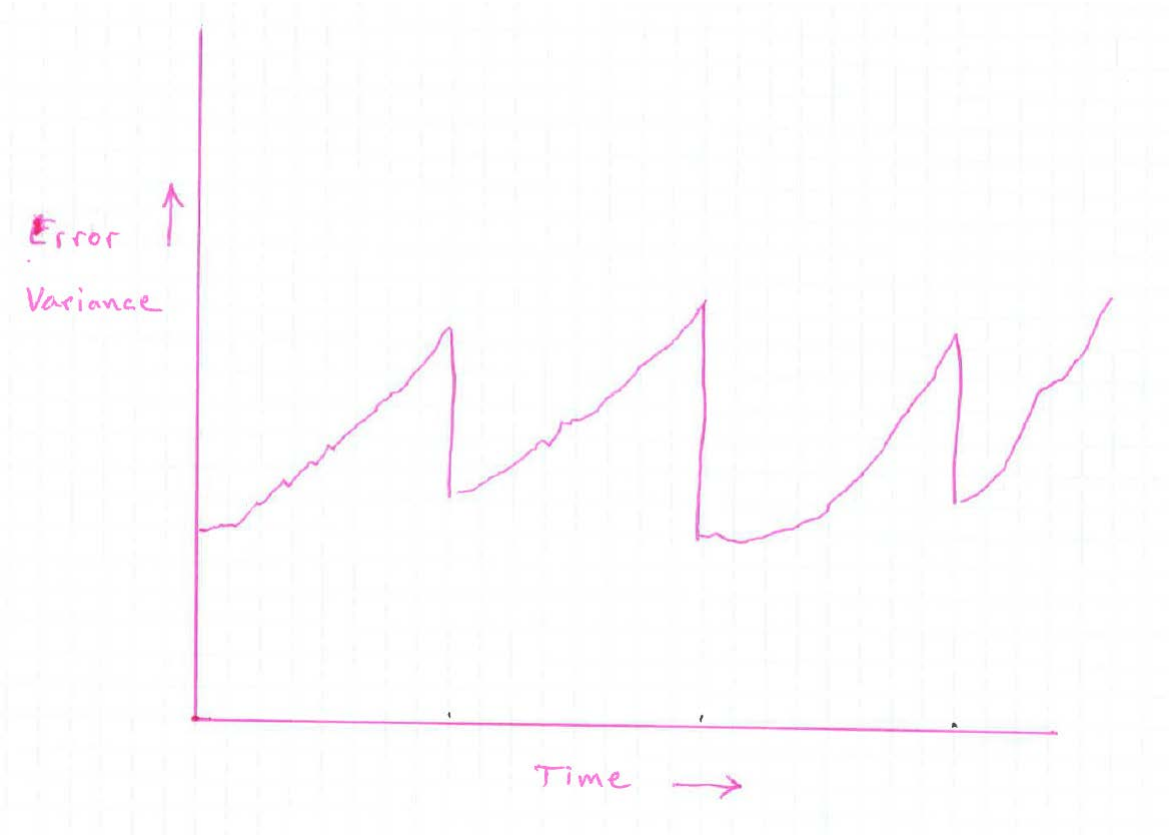
For a linear, unbiased, imperfect, model $\mathbf{M}\mathbf{x}$ that propagates information in time:

$$\mathbf{x}^b(t) = \mathbf{M}\mathbf{x}^a(t - \Delta t)$$
$$\mathbf{B}(t) = \mathbf{M}\mathbf{A}(t - \Delta t)\mathbf{M}^T + \mathbf{Q}$$

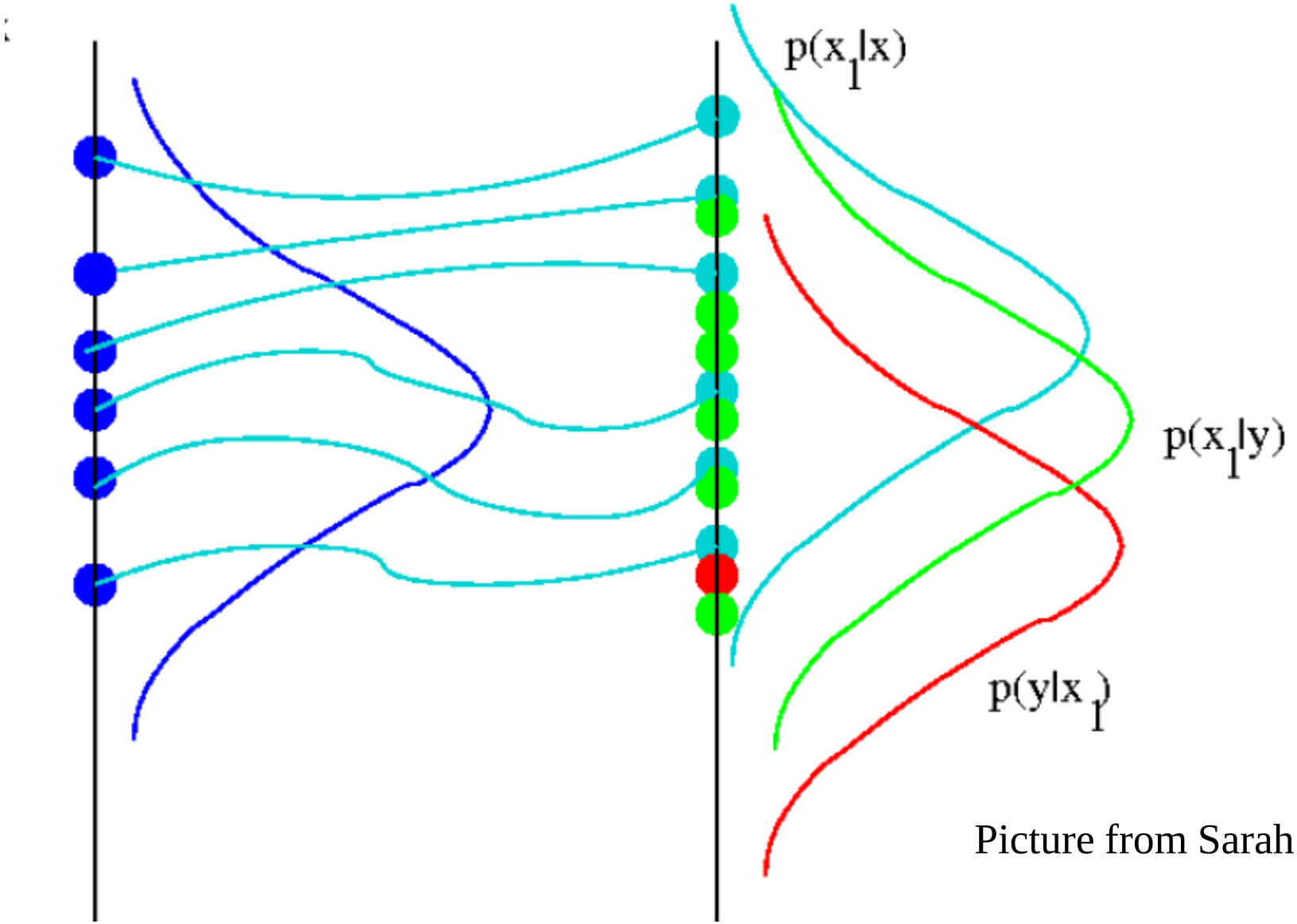
More generally:

$$\mathbf{B}(t) = E\left\{\left[\mathbf{x}(t) - \mathbf{x}_{\text{true}}(t)\right]\left[\mathbf{x}(t) - \mathbf{x}_{\text{true}}(t)\right]^T\right\}$$

Schematic of error in evolution of sequential KF



Ensemble Kalman Filter (ExKF)



Picture from Sarah Dance

Character of the Problem

1. A well-developed body of theory exists.
   Control theory, Inverse modeling, Bayesian analysis
   It is fundamental and foundational.

2. This theory is currently insufficient.
   The computational demand can be overwhelming.
   The required input statistics are not well known.

3. Gross approximations or unsupported assumptions may be required
   Although “wrong,” they can be useful.
   Sometimes they create confusion.

4. Many techniques are available
   Most are similar in a very general sense
   Results are affected by details

Basics

1. Fundamentals are foundational.
2. Statistical theory is critical.
3. Quality control is critical.
4. Consideration of covariances is critical.
5. Consideration of dynamic balance is critical.
6. Model error is not negligible.
7. Much model physics is not linear.
8. Model error is probably not white noise.
9. Experience counts!
10. Data assimilation is as much art as science.

Warnings

1. Most types of observations have potentially gross errors that must be detected and excluded by Quality Control (QC).
2. Computational (speed, data storage) requirement constrain what DA recipes can be reasonably employed.
3. Be aware that some assumptions even routinely employed may be motivated by their computational utility rather than reasonableness.
4. Do not confuse what is useful with what is correct.
5. Do not misinterpret metrics.

The nature of scientific questions

1. “Does using XXXX improve forecast accuracy?”

   This question cannot be answered without further specificity.

   So we change the question:

2. “Does using WWW in such and such a way reduce the metric XXX determined by validation using YYY as truth compared with the same metric applied to forecasts produced by using ZZZZ.

   This question can now be answered but is no longer the same as first expressed

   A misinterpretation occurs when we generalize (2) as though it is the answer to (1).

3. I encounter such confusion monthly!

Use of “Toy” models

1. Simple analytical examples
2. Lorenz-type models (illustrate behaviors)
3. Thesis by Nancy Baker (A sequence of simple models)
4. Shallow water (examine some balance issues)
5. Quasi-geostrophic (no balance but exceed Lorenz)
6. Others
7. OSSEs (most realistic, but also very complex)

Daley, R., 1991: *Atmospheric Data Analysis*, Cambridge University Press. 420 pp.

Tarantola, A., 1987: Inverse problem theory: Methods for data fitting and model parameter estimation. Elsevier Science B.V., 629 pp.

Ghil, M., K. Ide, A. Bennett, P. Courtier, M. Kimoto, M. Nagata, M. Saiki, and N. Sato, Eds., 1997: *Data Assimilation in Meteorology and Oceanography: Theory and Practice.* Meteorological Society of Japan. 386 pp.

Baker, N., 2000: Observation adjoint sensitivity and the adaptive observation-targeting problem. *Thesis*, Naval Post–Graduate School.

Daley, R., and E. Barker, 2000: NRL atmospheric variational data assimilation system source book. NRL publication, NRL/PU/7530–00–418, 153pp.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Q. J. Roy. Met. Soc.*, **112**, 1177–1194.

Lorenc, A.C. and O. Hammon, 1988: Objective quality control of observations using Bayesian methods: Theory, and a practical implementation. *Q. J. Roy. Met. Soc.*, **114**, 515–543.