# Probability in Variational Data Assimilation

David Livings

DARC Seminar
9 March 2016

# Motivation

- I've always taken a Bayesian approach to DA.
- This deals in probabilities such as $p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})$.
- In variational DA we might make a Gaussian background assumption:

$$\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}} \sim N(\mathbf{x}_k^{\mathrm{b}}, \mathbf{B}_k)$$

- The truth is a random variable.
- The background is a parameter of its distribution.
- The usual cost function has a probabilistic interpretation.

# Motivation (cont)

- But there's another approach.
- For example, from ECMWF Technical Memo 383 (Hólm et al, 2002):

> *Consider two forecasts of the truth $x$, $x_1^b$ and $x_2^b$, where*
>
> $$x_i^b = x + b^b(x) + \varepsilon_i^b$$
>
> *with $b^b$ the bias and $\varepsilon^b$ the stochastic error.*

- The background is the random variable.
- The truth is an unknown constant.
- Can we still interpret the cost function in probabilistic terms?

## Outline

# Bayes and Price (1763)

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

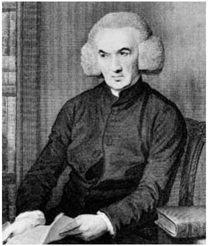
Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

He had, you know, the honour of being a member of that illuſtrious Society, and was much eſteemed by many in it as a very able mathematician. In an introduction which he has writ to this Eſſay, he ſays, that his deſign at firſt in thinking on the ſubject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumſtances, upon ſuppoſition that we know nothing concerning it but that, under the ſame circum-

?

Thomas Bayes (1702–1761)



Richard Price (1723–1791)

# Bayesian, Sense 1

- Using a subjective interpretation of probability

> 5. The *probability of any event* is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon it's happening.

- If I paid £10 for an insurance policy that pays £100,000 if I die this year, then the probability that I shall die this year is $10/100,000 = 10^{-4}$.
- Some say this is a subjective definition of probability based on betting.
- I say that the word 'ought' suggests Bayes considered there to be a single objective probability, based on (say) actuarial data.

# Bayesian, Sense 2

- Using Bayes's theorem

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

PROP. 3.

The probability that two ſubſequent events will both happen is a ratio compounded of the probability of the 1ſt, and the probability of the 2d on ſuppoſition the 1ſt happens.

$$P(AB) = P(A)P(B|A)$$

PROP. 5.

If there be two ſubſequent events, the probability of the 2d $\frac{b}{N}$ and the probability of both together $\frac{P}{N}$, and it being 1ſt diſcovered that the 2d event has happened, from hence I gueſs that the 1ſt event has alſo happened, the probability I am in the right is $\frac{P}{b}$†.

$$P(A|B) = \frac{P(AB)}{P(B)}$$

- Bayes never explicitly puts the two pieces together.

# What I Mean by Bayesian

- By a Bayesian approach to DA I mean Bayesian in Sense 2 (using Bayes's theorem).
- Independent of subjective Bayesian interpretation of probability.
- In modern axiomatic approach to probability, Bayes's theorem is trivial consequence of definition of conditional probability.
- Theorem applies to all interpretations of probability.
- There is a frequentist interpretation of probabilities such as $p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})$.
- But not insisting on this: everything that follows follows from the axioms and applies to any interpretation.

# Observations

- In the DA literature observations are customarily expressed as

$$\mathbf{y}_k^{\mathrm{o}} = H_k(\mathbf{x}_k^{\mathrm{t}}) + \varepsilon_k^{\mathrm{o}}$$

- In general, error distribution will depend on truth: $p(\varepsilon_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})$.
- For much of the theory we only need $p(\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})$.
- Independence of observation errors at different times is expressed through the conditional independence assumption:

$$p(\mathbf{y}_{1:k}^{\mathrm{o}}|\mathbf{x}_{1:k}^{\mathrm{t}}) = p(\mathbf{y}_1^{\mathrm{o}}|\mathbf{x}_1^{\mathrm{t}}) \dots p(\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})$$

# Observations (cont)

- When the observation is known, $p(\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})$ becomes a likelihood function on state space.
- It is this function (not just $\mathbf{y}_k^{\mathrm{o}}$) that we are assimilating.
- Sometimes we make the Gaussian observation assumption:

$$\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}} \sim N(H_k(\mathbf{x}_k^{\mathrm{t}}), \mathbf{R}_k)$$

# Bayesian Approach to 3D-Var

- By the conditional independence assumption, Bayes's theorem can be written

$$p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k}^{\mathrm{o}}) = \frac{p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})p(\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})}{p(\mathbf{y}_k^{\mathrm{o}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})}$$

- 3D-Var finds the analysis by maximising $p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k}^{\mathrm{o}}) \propto p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})p(\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})$.
- It is a maximum a posteriori (MAP) method.
- If we make the Gaussian observation assumption and the Gaussian background assumption (Bayesian version)

$$\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}} \sim N(\mathbf{x}_k^{\mathrm{b}}, \mathbf{B}_k)$$

3D-Var reduces to minimising the cost function

$$J(\mathbf{x}_k) = (\mathbf{x}_k - \mathbf{x}_k^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}_k^{-1}(\mathbf{x}_k - \mathbf{x}_k^{\mathrm{b}}) + (\mathbf{y}_k^{\mathrm{o}} - H_k(\mathbf{x}_k))^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{y}_k^{\mathrm{o}} - H_k(\mathbf{x}_k))$$

# Flow Dependence in the Bayesian Approach

- In the Gaussian background assumption

$$\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}} \sim N(\mathbf{x}_k^{\mathrm{b}}, \mathbf{B}_k)$$

  $\mathbf{x}_k^{\mathrm{b}}$ is a function of $\mathbf{y}_{1:k-1}^{\mathrm{o}}$.
- $\mathbf{B}_k$ may be a function of $\mathbf{y}_{1:k-1}^{\mathrm{o}}$ too.
- A special case is $\mathbf{B}_k(\mathbf{x}_k^{\mathrm{b}})$. This is one way of bringing flow-dependency into variational DA.
- In either case, $\mathbf{B}_k$ is fixed during the minimisation of $J(\mathbf{x}_k)$ because $\mathbf{y}_{1:k-1}^{\mathrm{o}}$ are fixed.
- In the Gaussian observation assumption

$$\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}} \sim N(H_k(\mathbf{x}_k^{\mathrm{t}}), \mathbf{R}_k)$$

  $\mathbf{R}_k$ may be a function of $\mathbf{x}_k^{\mathrm{t}}$.
- In this case, $\mathbf{R}_k(\mathbf{x}_k)$ varies during the minimisation of $J(\mathbf{x}_k)$.

# Towards the Alternative Approach

- $\mathbf{x}_k^{\mathrm{b}}$ from the Bayesian approach is a function of $\mathbf{y}_{1:k-1}^{\mathrm{o}}$.
- If we give up conditioning on $\mathbf{y}_{1:k-1}^{\mathrm{o}}$, then $\mathbf{x}_k^{\mathrm{b}}$ becomes a random variable (because $\mathbf{y}_{1:k-1}^{\mathrm{o}}$ are) and we can consider $p(\mathbf{x}_k^{\mathrm{b}}|\mathbf{x}_k^{\mathrm{t}})$.
- We can generalise this to an arbitrary estimate $\mathbf{x}_k^{\mathrm{b}}(\mathbf{y}_{1:k-1}^{\mathrm{o}})$.
- The background is now a random variable and the truth is a parameter of its distribution.
- In this approach the background is rather like an observation.
- At the start of the assimilation cycle, $\mathbf{x}_k^{\mathrm{b}}$ is known and $p(\mathbf{x}_k^{\mathrm{b}}|\mathbf{x}_k^{\mathrm{t}})$ is a likelihood function on state space.
- I therefore call this the likelihood approach.

# Likelihood Approach to 3D-Var

- By the conditional independence assumption

$$p(\mathbf{x}_k^{\mathrm{b}}, \mathbf{y}_k^{\mathrm{o}} | \mathbf{x}_k^{\mathrm{t}}) = p(\mathbf{x}_k^{\mathrm{b}} | \mathbf{x}_k^{\mathrm{t}}) p(\mathbf{y}_k^{\mathrm{o}} | \mathbf{x}_k^{\mathrm{t}})$$

- 3D-Var finds the analysis by maximising $p(\mathbf{x}_k^{\mathrm{b}}, \mathbf{y}_k^{\mathrm{o}} | \mathbf{x}_k^{\mathrm{t}})$.
- It is a maximum likelihood (ML) method.
- If we make the Gaussian observation assumption and the Gaussian background assumption (likelihood version)

$$\mathbf{x}_k^{\mathrm{b}} | \mathbf{x}_k^{\mathrm{t}} \sim N(\mathbf{x}_k^{\mathrm{t}}, \mathbf{B}_k)$$

  3D-Var reduces to minimising the cost function

$$J(\mathbf{x}_k) = (\mathbf{x}_k^{\mathrm{b}} - \mathbf{x}_k)^{\mathrm{T}} \mathbf{B}_k^{-1} (\mathbf{x}_k^{\mathrm{b}} - \mathbf{x}_k) + (\mathbf{y}_k^{\mathrm{o}} - H_k(\mathbf{x}_k))^{\mathrm{T}} \mathbf{R}_k^{-1} (\mathbf{y}_k^{\mathrm{o}} - H_k(\mathbf{x}_k))$$

# Flow Dependence in the Likelihood Approach

- In the Gaussian background assumption

$$\mathbf{x}_k^{\mathrm{b}} | \mathbf{x}_k^{\mathrm{t}} \sim N(\mathbf{x}_k^{\mathrm{t}}, \mathbf{B}_k)$$

  $\mathbf{B}_k$ may be a function of $\mathbf{x}_k^{\mathrm{t}}$.
- In this case, $\mathbf{B}_k(\mathbf{x}_k)$ varies during the minimisation of $J(\mathbf{x}_k)$.
- This differs from the Bayesian approach, where $\mathbf{B}_k$ was fixed.
- As in the Bayesian approach, we can have $\mathbf{R}_k(\mathbf{x}_k)$ varying during the minimisation of $J(\mathbf{x}_k)$.

# Strong Constraint 4D-Var in Brief

- Strong constraint 4D-Var is a lot like 3D-Var.
- The Bayesian approach is an MAP method.
- The likelihood approach is an ML method.
- With Gaussian assumptions both approaches reduce to minimisation of the usual cost function.
- $\mathbf{B}_k$ can vary in the minimisation in the likelihood approach, but not in the Bayesian approach.
- $\mathbf{R}_l$ can vary in the minimisation in both approaches.

# Elements Common to Both Approaches

- The system evolves according to a Markov process:

$$p(\mathbf{x}^{\mathrm{t}}_k|\mathbf{x}^{\mathrm{t}}_{1:k-1}) = p(\mathbf{x}^{\mathrm{t}}_k|\mathbf{x}^{\mathrm{t}}_{k-1})$$

- The process is specified by the transition densities $p(\mathbf{x}^{\mathrm{t}}_k|\mathbf{x}^{\mathrm{t}}_{k-1})$.
- Sometimes we assume a Gauss-Markov process:

$$\mathbf{x}^{\mathrm{t}}_k|\mathbf{x}^{\mathrm{t}}_{k-1} \sim N(M_k(\mathbf{x}^{\mathrm{t}}_{k-1}), \mathbf{Q}_k)$$

- $\mathbf{Q}_k$ may be a function of $\mathbf{x}^{\mathrm{t}}_{k-1}$.
- We have a sequence of observations $\mathbf{y}^{\mathrm{o}}_{k:k+K}$ to assimilate.
- We seek an estimate of the trajectory $\mathbf{x}^{\mathrm{t}}_{k:k+K}$ over the entire assimilation window.

# Bayesian Approach to Weak Constraint 4D-Var

- Start with the prior distribution $p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})$.
- By conditional independence, Bayes's theorem can be written

$$p(\mathbf{x}_{k:k+K}^{\mathrm{t}}|\mathbf{y}_{1:k+K}^{\mathrm{o}}) = \frac{p(\mathbf{x}_{k:k+K}^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})p(\mathbf{y}_{k:k+K}^{\mathrm{o}}|\mathbf{x}_{k:k+K}^{\mathrm{t}})}{p(\mathbf{y}_{k:k+K}^{\mathrm{o}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})}$$

- By conditional independence and the Markov condition

$$p(\mathbf{x}_{k:k+K}^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}}) = p(\mathbf{x}_k^{\mathrm{t}}|\mathbf{y}_{1:k-1}^{\mathrm{o}})p(\mathbf{x}_{k+1}^{\mathrm{t}}|\mathbf{x}_k^{\mathrm{t}})\dots p(\mathbf{x}_{k+K}^{\mathrm{t}}|\mathbf{x}_{k+K-1}^{\mathrm{t}})$$

- By conditional independence

$$p(\mathbf{y}_{k:k+K}^{\mathrm{o}}|\mathbf{x}_{k:k+K}^{\mathrm{t}}) = p(\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})\dots p(\mathbf{y}_{k+K}^{\mathrm{o}}|\mathbf{x}_{k+K}^{\mathrm{t}}).$$

- Weak constraint 4D-Var finds the analysis trajectory by maximising $p(\mathbf{x}_{k:k+K}^{\mathrm{t}}|\mathbf{y}_{1:k+K}^{\mathrm{o}})$.
- It is an MAP method.

# Bayesian Approach to Weak Constraint 4D-Var (cont)

- If we make the Gaussian background assumption, the Gauss-Markov assumption, and the Gaussian observation assumption, then weak constraint 4D-Var reduces to minimising the cost function

$$
\begin{aligned}
J(\mathbf{x}_{k:k+K}) \; = \; & (\mathbf{x}_k - \mathbf{x}_k^{\mathrm{b}})^{\mathrm{T}} \mathbf{B}_k^{-1} (\mathbf{x}_k - \mathbf{x}_k^{\mathrm{b}}) \\
& + \sum_{l=k+1}^{k+K} (\mathbf{x}_l - M_l(\mathbf{x}_{l-1}))^{\mathrm{T}} \mathbf{Q}_l^{-1} (\mathbf{x}_l - M_l(\mathbf{x}_{l-1})) \\
& + \sum_{l=k}^{k+K} (\mathbf{y}_l^{\mathrm{o}} - H_l(\mathbf{x}_l))^{\mathrm{T}} \mathbf{R}_l^{-1} (\mathbf{y}_l^{\mathrm{o}} - H_l(\mathbf{x}_l))
\end{aligned}
$$

- $\mathbf{B}_k$ may be a function of $\mathbf{y}_{1:k-1}^{\mathrm{o}}$ (or $\mathbf{x}_k^{\mathrm{b}}$), but is fixed during the minimisation.
- $\mathbf{Q}_l(\mathbf{x}_{l-1})$ and $\mathbf{R}_l(\mathbf{x}_l)$ may vary during the minimisation.

# Likelihood Approach to Weak Constraint 4D-Var

- Start with an estimate $\mathbf{x}_k^{\mathrm{b}}(\mathbf{y}_{1:k-1}^{\mathrm{o}})$ with likelihood $p(\mathbf{x}_k^{\mathrm{b}}|\mathbf{x}_k^{\mathrm{t}})$.
- By conditional independence and the Markov condition

$$p(\mathbf{x}_k^{\mathrm{b}}, \mathbf{x}_{k+1:k+K}^{\mathrm{t}}, \mathbf{y}_{k:k+K}^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}}) = p(\mathbf{x}_k^{\mathrm{b}}|\mathbf{x}_k^{\mathrm{t}})p(\mathbf{x}_{k+1:k+K}^{\mathrm{t}}|\mathbf{x}_k^{\mathrm{t}})p(\mathbf{y}_{k:k+K}^{\mathrm{o}}|\mathbf{x}_{k:k+K}^{\mathrm{t}})$$

- By the Markov condition

$$p(\mathbf{x}_{k+1:k+K}^{\mathrm{t}}|\mathbf{x}_k^{\mathrm{t}}) = p(\mathbf{x}_{k+1}^{\mathrm{t}}|\mathbf{x}_k^{\mathrm{t}})p(\mathbf{x}_{k+2}^{\mathrm{t}}|\mathbf{x}_{k+1}^{\mathrm{t}})\dots p(\mathbf{x}_{k+K}^{\mathrm{t}}|\mathbf{x}_{k+K-1}^{\mathrm{t}})$$

- By conditional independence

$$p(\mathbf{y}_{k:k+K}^{\mathrm{o}}|\mathbf{x}_{k:k+K}^{\mathrm{t}}) = p(\mathbf{y}_k^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})\dots p(\mathbf{y}_{k+K}^{\mathrm{o}}|\mathbf{x}_{k+K}^{\mathrm{t}})$$

- Weak constraint 4D-Var finds the analysis trajectory by maximising $p(\mathbf{x}_k^{\mathrm{b}}, \mathbf{x}_{k+1:k+K}^{\mathrm{t}}, \mathbf{y}_{k:k+K}^{\mathrm{o}}|\mathbf{x}_k^{\mathrm{t}})$.
- It is a mixed ML/MAP method.

# Likelihood Approach to Weak Constraint 4D-Var (cont)

- If we make the Gaussian background assumption, the Gauss-Markov assumption, and the Gaussian observation assumption, then weak constraint 4D-Var reduces to minimising the cost function

$$
\begin{aligned}
J(\mathbf{x}_{k:k+K}) &= (\mathbf{x}_k^{\mathrm{b}} - \mathbf{x}_k)^{\mathrm{T}} \mathbf{B}_k^{-1} (\mathbf{x}_k^{\mathrm{b}} - \mathbf{x}_k) \\
&\quad + \sum_{l=k+1}^{k+K} (\mathbf{x}_l - M_l(\mathbf{x}_{l-1}))^{\mathrm{T}} \mathbf{Q}_l^{-1} (\mathbf{x}_l - M_l(\mathbf{x}_{l-1})) \\
&\quad + \sum_{l=k}^{k+K} (\mathbf{y}_l^{\mathrm{o}} - H_l(\mathbf{x}_l))^{\mathrm{T}} \mathbf{R}_l^{-1} (\mathbf{y}_l^{\mathrm{o}} - H_l(\mathbf{x}_l))
\end{aligned}
$$

- $\mathbf{B}_k(\mathbf{x}_k)$, $\mathbf{Q}_l(\mathbf{x}_{l-1})$, and $\mathbf{R}_l(\mathbf{x}_l)$ may all vary during the minimisation.

# Summary

- There are (at least) two approaches to variational DA: the Bayesian approach and the likelihood approach.
- By 'Bayesian' I mean merely using Bayes's theorem.
- In the Bayesian approach, all three variants of Var are MAP methods.
- In the likelihood approach, 3D-Var and strong constraint 4D-Var are ML methods, but weak constraint 4D-Var is a mixed ML/MAP method.
- The usual cost functions are obtained by making Gaussian assumptions.
- No linear assumptions are necessary.
- In the Bayesian approach:
  - $\mathbf{B}_k$ may be a function of $\mathbf{y}^{\mathrm{o}}_{1:k-1}$ (or $\mathbf{x}^{\mathrm{b}}_k$), but is held fixed during the minimisation of the cost function.
  - $\mathbf{Q}_l(\mathbf{x}_{l-1})$ and $\mathbf{R}_l(\mathbf{x}_l)$ may vary during the minimisation.
- In the likelihood approach, $\mathbf{B}_k(\mathbf{x}_k)$, $\mathbf{Q}_l(\mathbf{x}_{l-1})$, and $\mathbf{R}_l(\mathbf{x}_l)$ may all vary during the minimisation.