Overview of Data Assimilation Methods

Massimo Bonavita
ECMWF
Massimo.Bonavita@ecmwf.int

Outline

- What is data assimilation and how does it work?
- Data assimilation ingredients: Observations and models
- Blending observations and model: the Bayes perspective
- A whirlwind introduction to DA methods in the geophysical sciences:
  - Particle Filters
  - Kalman Filters
  - Variational methods
  - Hybrid methods – Machine Learning

Data Assimilation

The goal of Data Assimilation is:

"*Estimate the probability distribution function (pdf) of the Atmosphere (Earth system) at the initial time*"

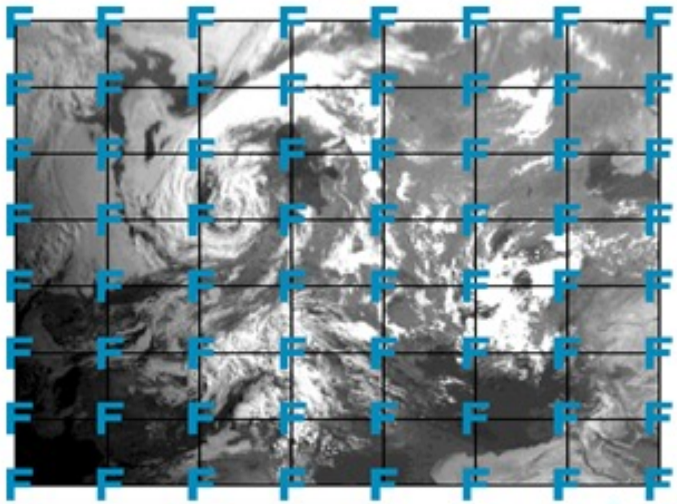
The initial state pdf is typically sampled (Ensemble DA) and is usually summarised in terms of its central moment (the "analysis") and its uncertainty (the covariance around the central estimate).

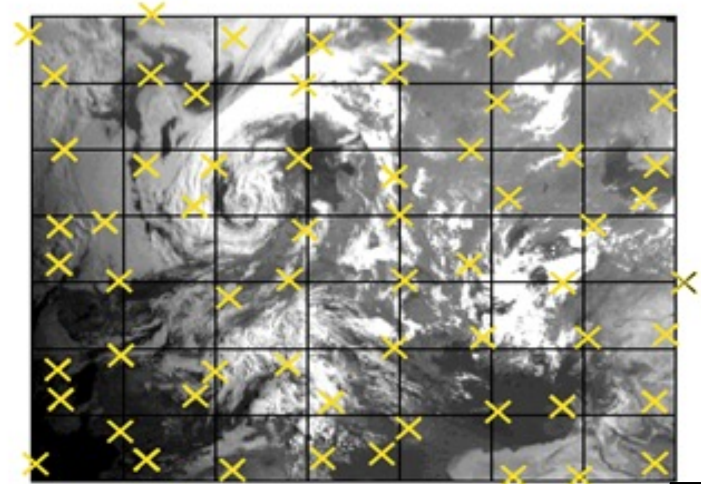
This representation of the initial pdf in terms of its first two moments (mean and covariance) is adequate for ~ Gaussian (or at least unimodal) error distributions, less informative for multimodal error distributions.

In realistic geophysical applications of DA we need to make assumptions to make the problem computationally tractable

Model Forecast (with errors)

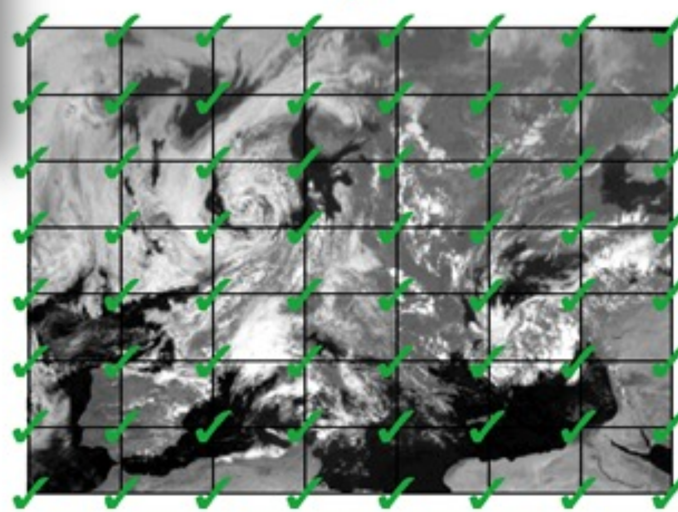
Observations (with errors)



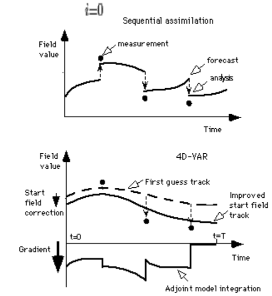


People

Computer (with a lot of CPUs/GPUs/TPUs)

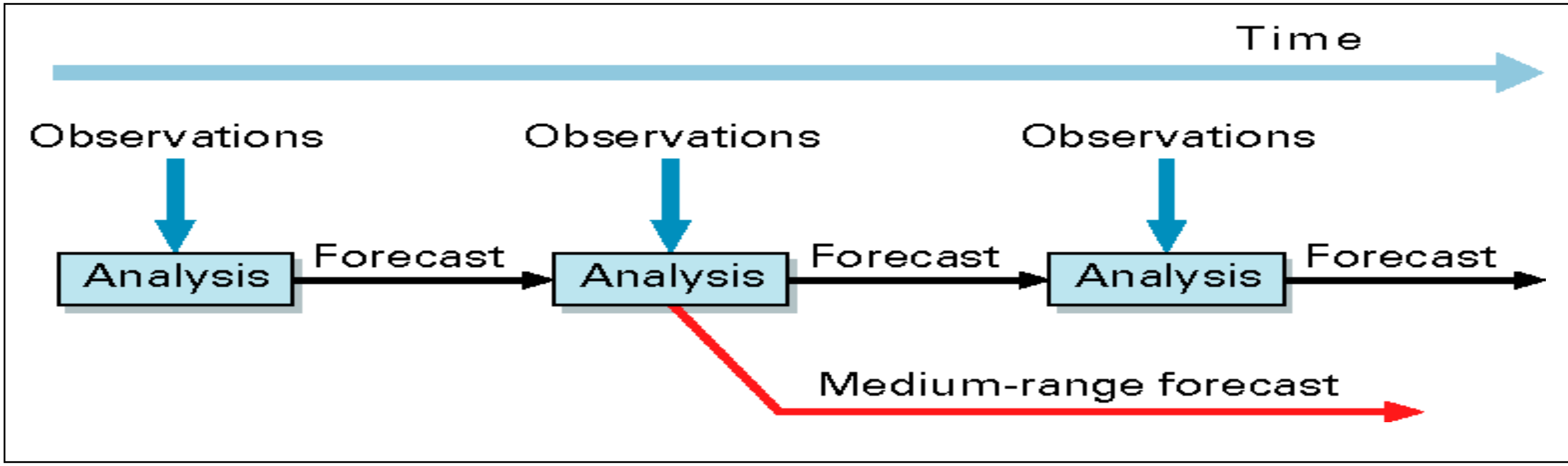


$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \sum_{i=0}^{n}(\mathbf{y}_i - H_i[\mathbf{x}_i])^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i - H_i[\mathbf{x}_i])$$



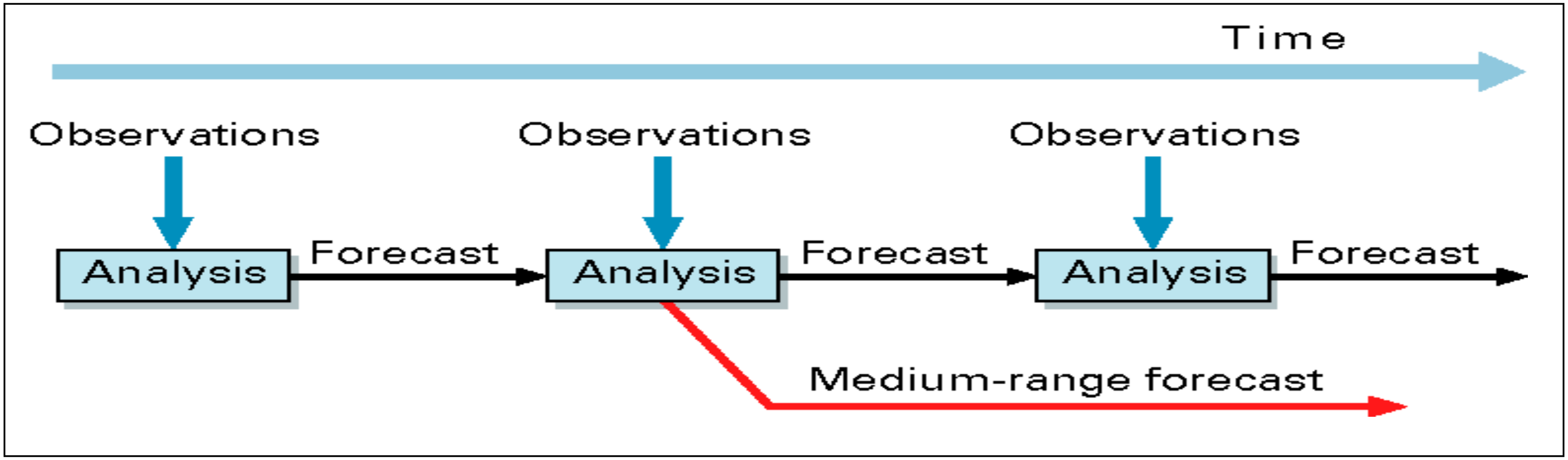
Analysis (with - smaller – errors)

The Data assimilation cycle



- An analysis is not produced by observations alone!
- The observations are used to correct errors in the short forecast from the previous analysis time (every 12 hours at ECMWF, 3-6 hourly in other global NWP systems; 1-3 hourly for higher resolution, limited area models).
- The short-range forecast carries information from past observations into the current analysis update

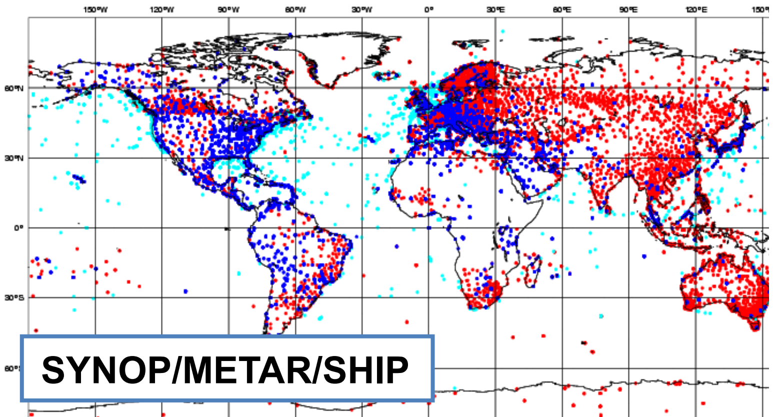
The Data assimilation cycle

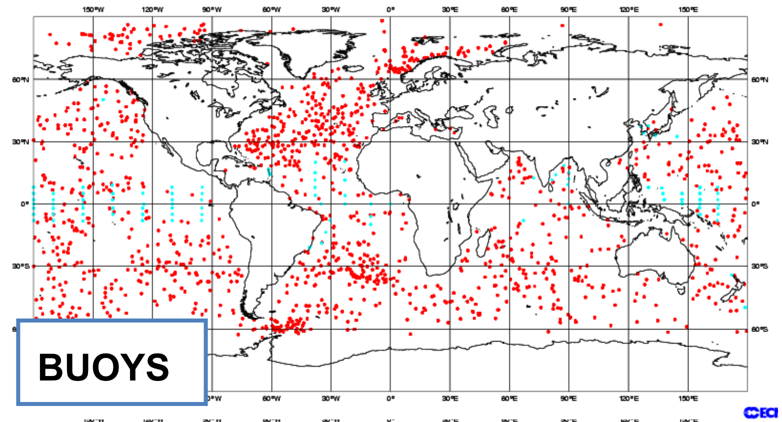


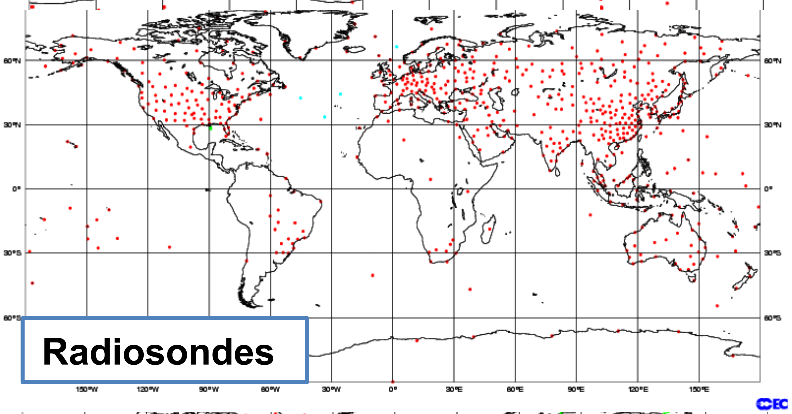
- At ECMWF, twice a day about 40,000,000 observations are used to correct the O(10^9) variables that define the analysis state.
- This is done by a 4-dimensional adjustment in space and time based on the available observations (4D-Var); 4D-Var is computer intensive (approx. as much as the 10-day forecast)

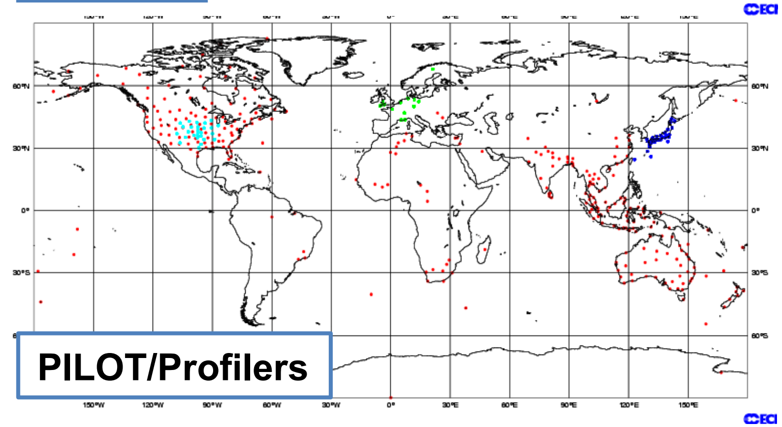
Observations

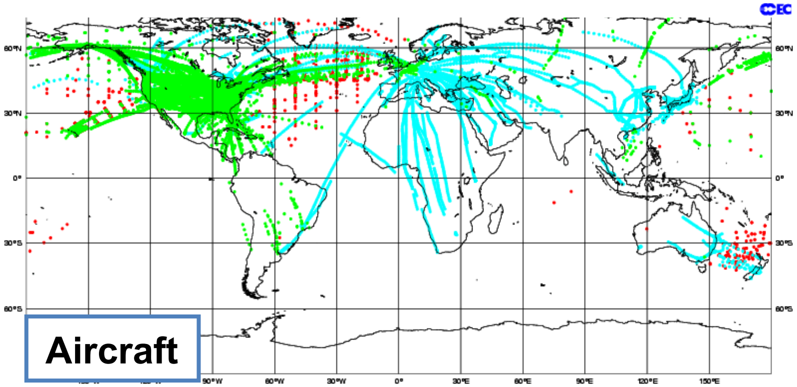
Distribution of in situ observations







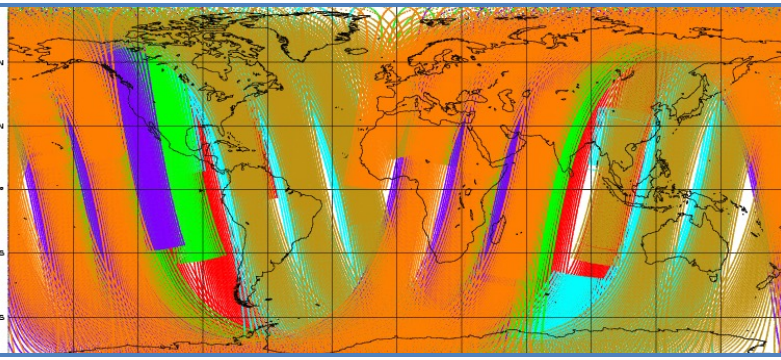




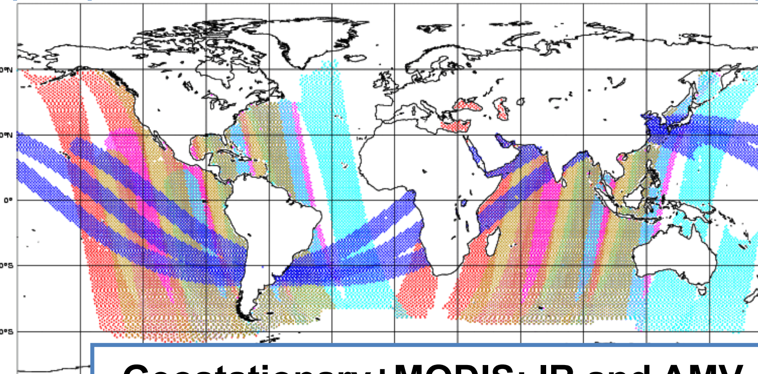
https://www.ecmwf.int/en/forecasts/quality-our-forecasts/monitoring-observing-system

Satellite data sources used by ECMWF's analysis

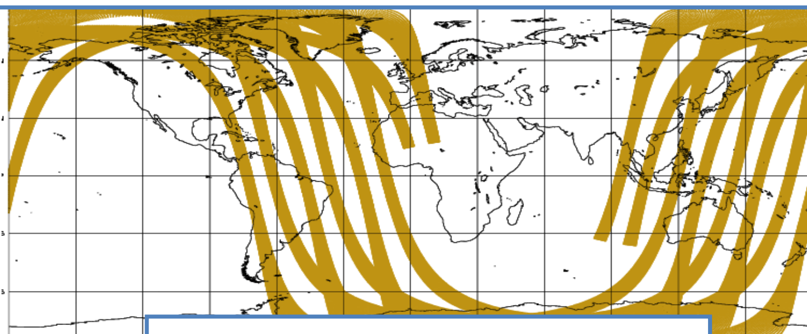
Sounders: NOAA AMSU-A/B, HIRS, AIRS, IASI, MHS



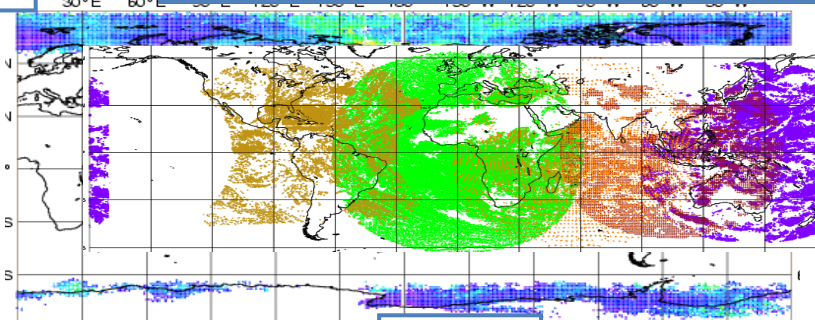
Imagers: SSMI, SSMIS, AMSR-E,



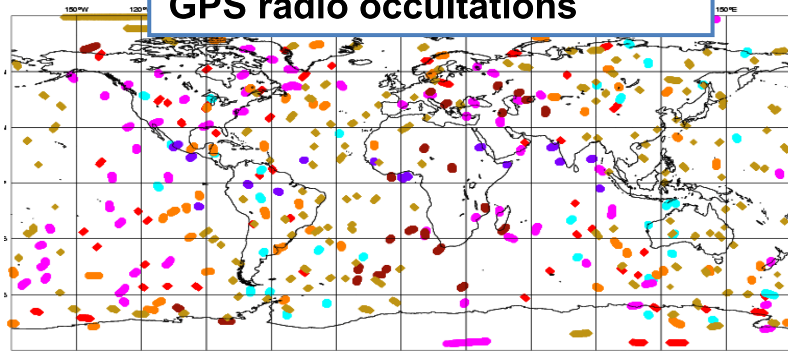
Scatterometer ocean low-level winds: ASCAT



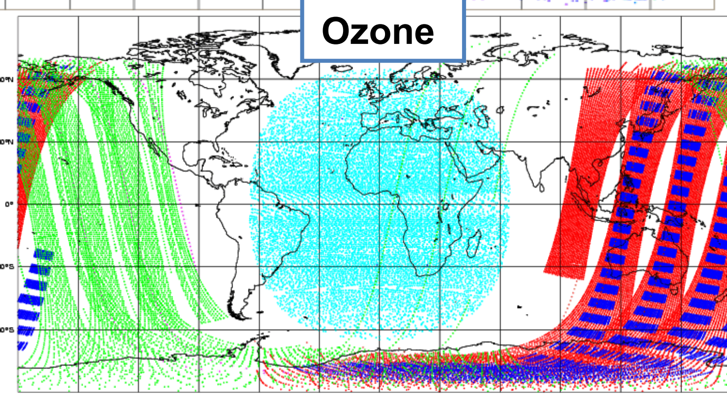
Geostationary+MODIS: IR and AMV



GPS radio occultations



Ozone



Observations

- The current global observing system is not able to observe the atmosphere completely (data voids): from a mathematical standpoint data assimilation is an under-determined problem
- Most satellite observations (e.g. radiances) are only indirectly related to the quantities of interest (i.e., grid point values of T,u,v,q,O3,…)
- Satellite observations have coarse vertical and/or horizontal resolution
- Observations measure quantities not located at grid points

Observations

In order to compare observations ($\mathbf{y}$) and model ($\mathbf{x}$) we need to perform spatial and temporal interpolations of the model fields and (for satellite observations) transform model fields into the quantities observed by the satellite sensors (radiances, bending angles, back-scattered radiation, etc)

We call this set of operations the observation operator ($\mathcal{H}$):

$$\mathbf{y} = \mathcal{H}(\mathbf{x})$$

Note: we project model fields into observed quantities, not observations into model variables (this second operation is called a "retrieval": $\mathbf{x} = \mathcal{H}^{-1}(\mathbf{y})$, and it is usually ill-posed -> there are many/infinite $\mathbf{x}$ corresponding to the same $\mathbf{y}$)

Observation errors

- Observations are affected by different types of errors
- Denoting $\mathbf{y}^*$ as the observations of the true model state ($\mathbf{y}^* = \mathcal{H}(\mathbf{x}^*)$):

$$\mathbf{y} - \mathbf{y}^* = \varepsilon_o = \varepsilon_G + \varepsilon_M + \varepsilon_R + \varepsilon_H$$

ε_G = Gross errors (incorrect coding of observation, duplicates, incorrect location, wrong cloud clearing, etc.).

ε_M = Measurement errors (instrument noise)

ε_R = Representativity errors (e.g., in situ observations compared to grid point model value)

ε_H = Observation operator (Forward model) errors (e.g., errors in the radiative transfer model, interpolation errors, etc.)

Observation errors

$$\mathbf{y} - \mathbf{y}^* = \varepsilon_o = \varepsilon_G + \varepsilon_M + \varepsilon_R + \varepsilon_H$$

- ε_G (gross errors) are dealt with by Observation Quality Control techniques (to be discussed later this week); Some of these checks are applied before ingesting the observations (Climatological checks, physical consistency checks, first guess checks), others are part of the analysis algorithm itself (buddy checks, Variational Quality Control)

- Observations are assumed to be un-biased:

$$\langle \varepsilon_o \rangle = 0$$

- Biases are dealt with specific Bias Correction techniques (to be discussed this week), which can be part of the analysis algorithm itself (e.g., Variational Bias Correction)

- The covariance matrix of the observation errors is denoted as $\mathbf{R}$:

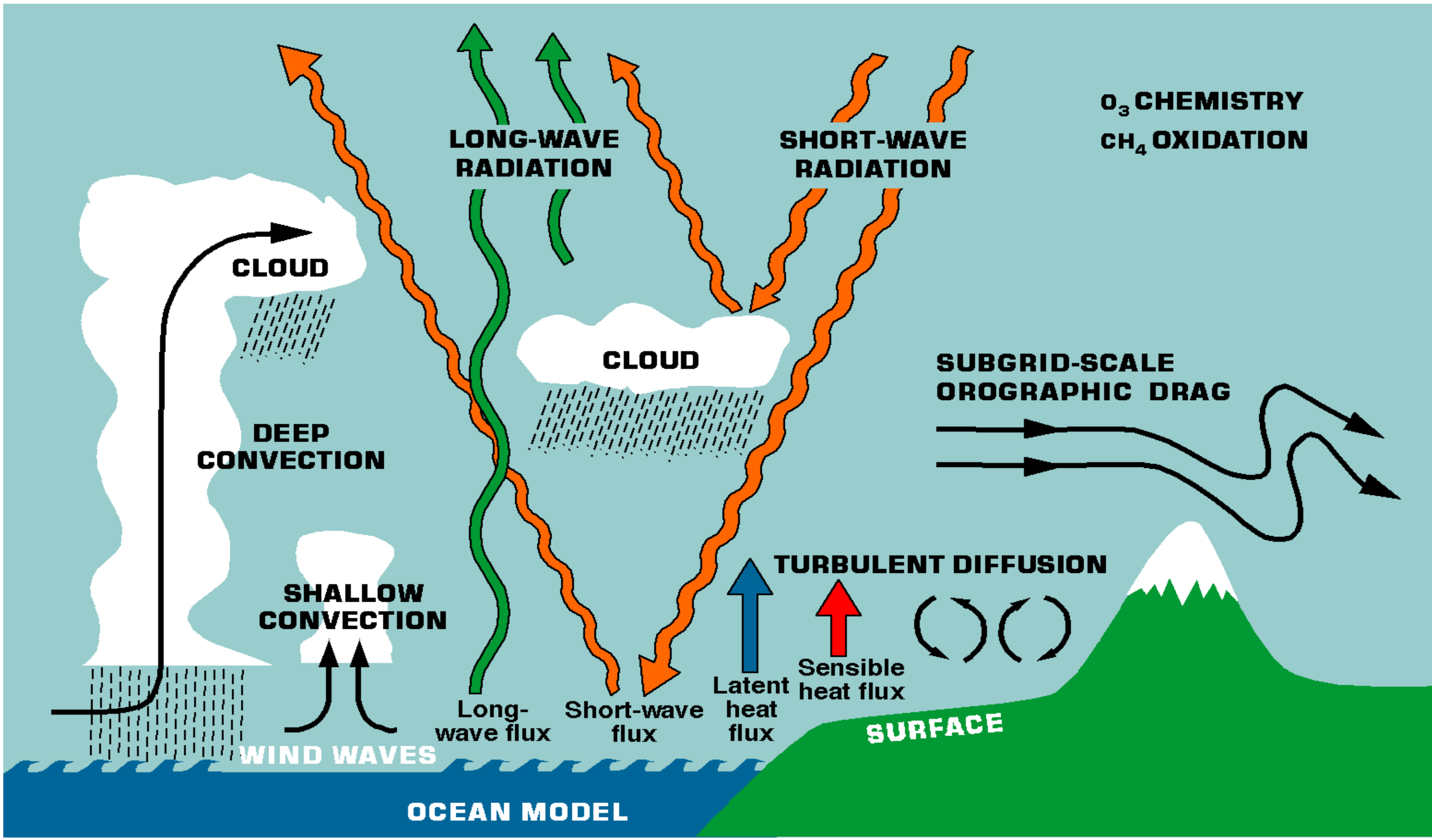
$$\langle \varepsilon_o \varepsilon_o^T \rangle = \mathbf{R}$$

The forecast model

The forecast model is a very important part of the data assimilation system

- The short-range forecast connecting successive analysis updates carries information from past observations to the current analysis time (this is called the "background"): the better the model the more accurate the background state
- A good model starting from accurate previous analysis will produce an accurate background ⇒ the analysis will make only small corrections to the background
- When the analysis makes large corrections to the background state is usually a sign that something interesting is happening… (e.g., rapid development not present in the forecast; suspect observations)
- In modern data assimilation methods (4D-Var, EnKF, PF) the analysed state is constructed so as to respect the physical and dynamical balances of the model ⇒ the model is an integral part of the analysis algorithm

The forecast model is a very important part of the data assimilation system



Physical processes in the ECMWF model

Model errors

- Despite their increasing complexity and sophistication models are not perfect!
- Sources of model error include: missing physical processes, errors in parametrizations of physical processes, discretisation errors (from continuous PDEs to discrete formulation), error in the forcing fields, etc.,
- We define model error as (* denotes true state, $\boldsymbol{i}$ is the time index):

$$\boldsymbol{x}_{\boldsymbol{i}}^{*} = \mathcal{M}(\boldsymbol{x}_{\boldsymbol{i-1}}^{*}) + \boldsymbol{\eta}_i$$

- Model error can in general have non zero mean:

$$\langle \boldsymbol{\eta}_i \rangle \neq 0$$

- The covariance matrix of the model errors is denoted as $\mathbf{Q}$:

$$\langle \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \rangle = \mathbf{Q}_i$$

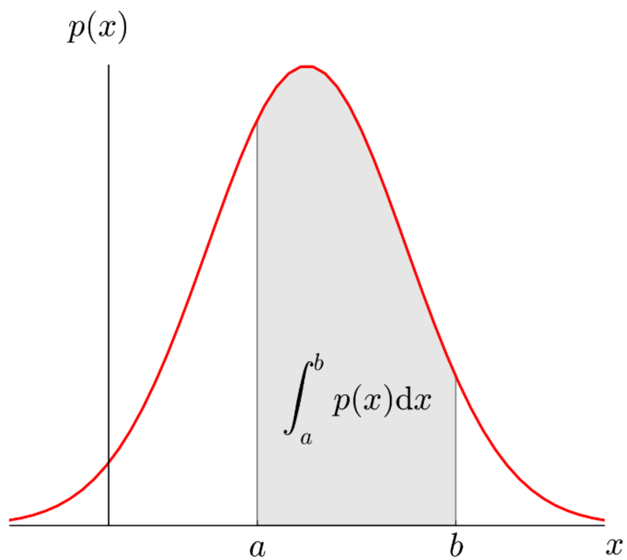
- The treatment of model error in DA will be discussed in a dedicated lecture later this week

Blending observations and model information: the Bayes perspective

The Bayes perspective

- Both observations and models are affected by random errors*
- This means that they should be described as random variables
- All we can/need to know about random variables are their probability distribution functions:

$$\Pr[a \leq X \leq b] = \int_a^b p(x)dx$$



* Assume here that systematic errors have been corrected before

The Bayes perspective

- Bayes law descends directly by the definition of conditional probabilities:

$$p(A,B) = p(A|B)p(B) = p(B|A)p(A)$$

$$\Rightarrow$$

$$\boxed{p(A|B) = \frac{p(B|A)p(A)}{p(B)}}$$

Where:

$p(A,B) = probability\ of\ events\ A\ and\ B\ both\ happening\ (joint\ prob.distribution)$

$p(A|B) = probability\ of\ event\ A\ given\ that\ B\ is\ true\ (conditional\ prob.distribution)$

$p(A), p(B) = probability\ of\ event\ A\ (B)\ happening\ (marginal\ prob.distribution)$

The Bayes perspective

- An illustration (http:en.wikipedia.org/wiki/Base_rate_fallacy):

The police have been issued with breathalysers which never fail to detect a drunk person but have a 5% rate of false positives. Prior campaigns have shown that, on average, one in one thousand drivers drives drunk. If the police stop a driver at random, and he/she results positive to the breathalyser, what is the probability that he/she is actually drunk?

Event A: being a drunk driver. Probability of being a drunk driver, before being tested: $p(A) = 0.001$

Event B: testing positive to the breathalyser. The probability of testing positive is 1 for the drunken subset of the drivers (0.001) and 0.05 for the sober subset of the drivers (0.999):
$p(B) = (1 * 0.001) + (0.999 * 0.05) = 0.05095$

Probability of testing positive to the breathalyser when drunk: $p(B|A) = 1$

Probability of being drunk after testing positive to the breathalyser, $p(A|B)$:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{1 * 0.001}{0.05095} = 0.0198$$

In words: out of 1000 people stopped by the police, about 51 will result positive, but the probability that anyone of them is actually drunk is less than 2%. (This shows how Bayesian thinking can be useful even beyond data assimilation!)

The Bayes perspective

- Another illustration: the Monty Hall problem (https://en.wikipedia.org/wiki/Monty_Hall_problem):

This is a brain teaser inspired by the American quiz show *Let's Make a Deal* and named after its original host, Monty Hall. The version of the problem that appeared in the Parade mag1990 read:

"*Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say #1, and the host, who knows what's behind the doors, opens another door, say #3, which has a goat. He says to you, "Do you want to pick door #2?" Is it to your advantage to switch your choice of doors?*"

Let us use our Bayesian tools to see what the savvy game show participants should do! Let usindicate with p(1), p(2), p(3) the probability that the car is behind door 1,2,3. Initially p(1)=p(2)=p(3)=1/3. To check whether it is a good idea to switch door, we are interested in is the probability of the car being behind door=2 after the host has chosen door=3 and we (the guest) have chosen door=1, in symbols:

$$p(2|H=3, G=1)$$

The Bayes perspective

- Another illustration: the Monty Hall problem (continued)

$$p(2|H = 3, G = 1) = \frac{p(2,H=3,G=1)}{p(H=3,G=1)}$$ (from the definition of conditional prob.)

$$\frac{p(2,H=3,G=1)}{p(H=3,G=1)} = \frac{p(H=3 \mid 2,G=1)\, p(2,G=1)}{p(H=3,G=1)}$$ (again from def.of conditional prob.)

Now note that a) the probability that the host chooses door=3, given the car is in 2 and we have chosen door=1 is 1; and b) the probability that the car is behind door=2 is independent of our choice of door=1:

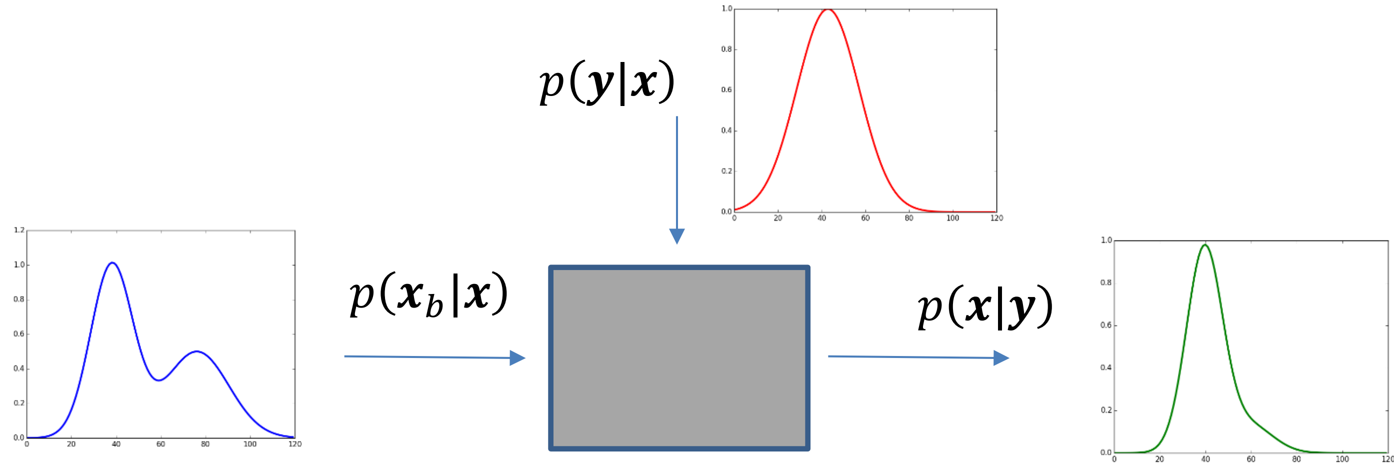
$$\frac{p(H=3 \mid 2,G=1)\, p(2,G=1)}{p(H=3,G=1)} = \frac{p(2)\, p(G=1)}{p(H=3 \mid G=1) p(G=1)} = \frac{p(2)}{p(H=3 \mid G=1)} = \frac{1/3}{1/2} = \frac{2}{3}$$

It does make sense to switch our choice to door=2!

This is another example of how the probability of the new piece of information (the Host's choice of door=3) has modified the a-priory probability of where the car might be.

The Bayes perspective

- At an abstract level, we can think of the analysis process as updating our prior knowledge about the state, represented by a background forecast and its pdf, with new observations, represented by their values and their respective pdfs:



$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}_b|\boldsymbol{x})}{p(\boldsymbol{y})} \propto \; p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}_b|\boldsymbol{x})$$

- $p(\boldsymbol{x}_b|\boldsymbol{x})$ = prior pdf (encapsulate our knowledge about the state before new observations)
- $p(\boldsymbol{y}|\boldsymbol{x})$ = observations likelihood (pdf of the observations conditioned on the state)
- $p(\boldsymbol{x}|\boldsymbol{y})$= posterior pdf (updated pdf of the state after the analysis)
- $p(\boldsymbol{y})$= marginal pdf of the observations (does not depend on $\boldsymbol{x}$: normalising constant in Bayes' law)

Particle Filters

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}_b|\boldsymbol{x}) \qquad (1)$$

- In principle an analysis update requires being able to compute the product pdf of the random variables $\boldsymbol{y}$, $\boldsymbol{x}_b$. This is usually not possible to do explicitly unless we choose very specific functional forms for the pdfs

- We thus need to make approximations… (This is where Data Assimilation starts!)

- One idea is to use Monte Carlo methods to sample and propagate the pdfs in (1): Particle Filters

- In Particle Filters, pdfs are sampled by a collection of "particles" (i.e., model states) with assigned weights:

$$p(\boldsymbol{x}) \sim \sum_{i=1,N} w_i \delta(\boldsymbol{x} - \boldsymbol{x}_i) \qquad (2)$$

Particle Filters

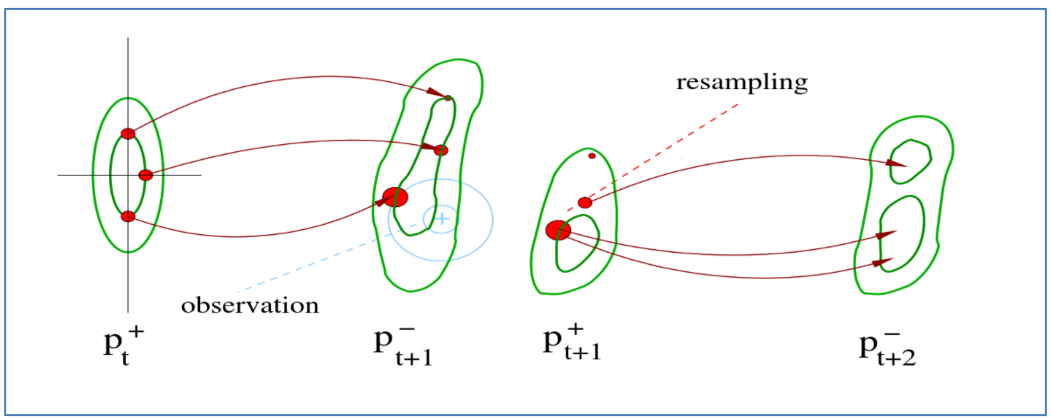
- The pdf is propagated in time by integrating the different particles with the model:

$$p(\boldsymbol{x}_b|\boldsymbol{x}) \sim \sum_{i=1,N} w_i \delta\big(\boldsymbol{x} - M(\boldsymbol{x}_i)\big) \qquad (3)$$

- In the analysis update the weights of the particles are updated according to the observations' likelihood:

$$w_i^a \propto w_i p(\boldsymbol{y}|\boldsymbol{x}_i)$$

- The ensemble of particles is resampled, i.e. high-weight particles close to the observed state are duplicated and low-weight particles discarded
- The Particle Filter described here is one of the most basic implementations (Bootstrap Particle Filter, Gordon et al., 1993)



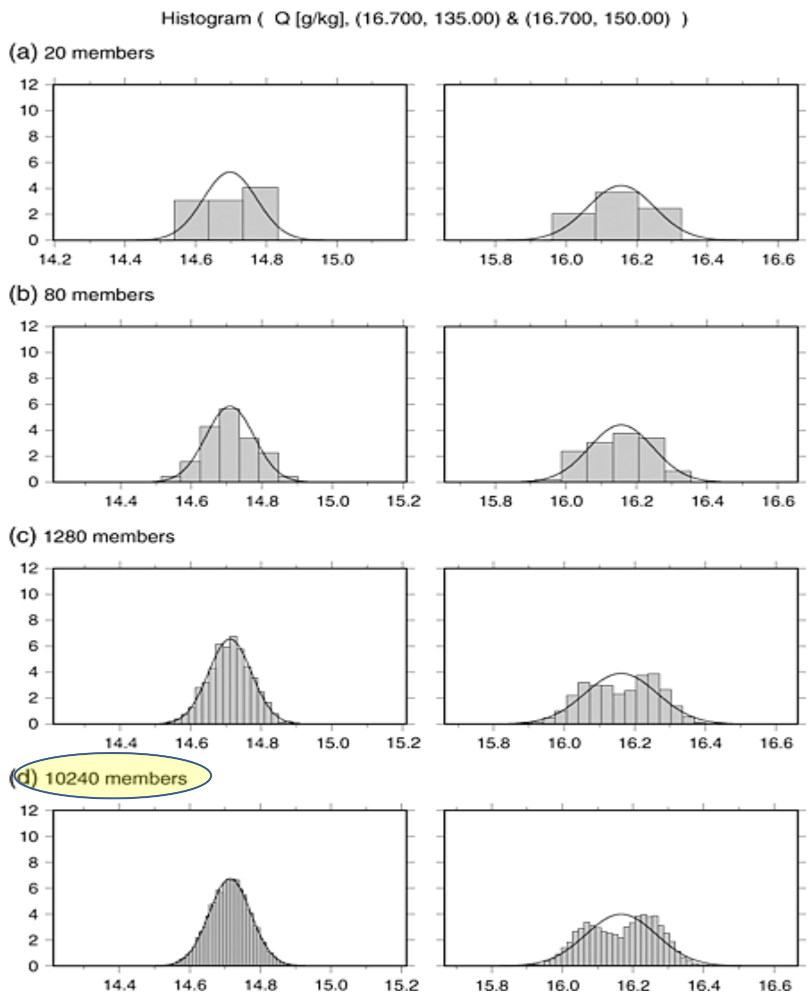
From M. Bocquet

Particle Filters

- The Particle Filter idea is attractive because it does not involve approximations on the Bayes update
- Particle Filters work well for very small state space sizes **and** observation sizes (N~10 to 100)
- For larger state space and/or observation sizes the required number of particles increases exponentially (Snyder et al., 2015)
- A large body of research has focussed on reducing the computational demands of particle filters for high dimensional systems
- One of the main themes of PF research is how to prevent the particles from diverging from the true state and becoming too unlikely, i.e. uninformative about the true state
- Many ideas: use observations to "guide" the particles' evolution from $t=t_{n-1}$ to $t=t_n$; many variants possible (Ades and van Leeuwen, 2014); introduce some form of localisation in the PF (similar to what is done for the EnKF): see Farchi and Bocquet, 2018, for a review
- Latest iteration is called Particle Flow Filter (Pulido and van Leeuwen, 2019; Hu et al., 2024)

Particle Filters

- Regardless of the assimilation algorithm the number of particles (ensemble members) needed to reliably resolve non-Gaussian pdfs is very high:



Histograms of a 6 h ensemble forecast for specific humidity ($g\ kg^{-1}$) for a intermediate AGCM. Miyoshi et al., 2014

The Gaussian approximation

- Not making assumptions on the shape of the prior and the likelihood pdf makes the Bayesian problem difficult (i.e., analytically <u>and</u> computationally intractable)
- Standard choice is to assume a Gaussian distribution for the both the observations' likelihood and the prior pdf of the background forecast
- Why Gaussian?
    1. Mathematically tractable problem;
    2. Full distribution characteristics defined by only its first two moments (mean and covariance);
    3. Supported by the Central Limit Theorem;
    4. Least committed distribution for given first and second moments (i.e., we are making the least amount of hypotheses on the shape of the pdf for a given sample variance)

The Gaussian approximation

- Usual choice is to assume a Gaussian distribution for the both the observations' likelihood and the prior pdf

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\mathbf{R}|^{1/2}} exp\left(-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x})\right)^{T}(\mathbf{R})^{-1}\left(\boldsymbol{x}_b-\boldsymbol{H}(\boldsymbol{x})\right)\right)$$

$$p(\boldsymbol{x}_b|\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\mathbf{P}_B|^{1/2}} exp\left(-\frac{1}{2}(\boldsymbol{x}_b-\boldsymbol{x})^{T}(\mathbf{P}_B)^{-1}(\boldsymbol{x}_b-\boldsymbol{x})\right)$$

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}_b|\boldsymbol{x}) \propto exp\left(-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x})\right)^{T}(\mathbf{R})^{-1}\left(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x})\right)-\frac{1}{2}(\boldsymbol{x}_b-\boldsymbol{x})^{T}(\mathbf{P}_B)^{-1}(\boldsymbol{x}_b-\boldsymbol{x})\right)$$

- where $\langle \varepsilon_o \varepsilon_o^T \rangle = \mathbf{R}$ and $\langle \varepsilon_b \varepsilon_b^T \rangle = \mathbf{P}_B$ are the covariances of the errors of the observations and of the prior (background forecast)

- Under this assumption the posterior (analysis) distribution $p(\boldsymbol{x}|\boldsymbol{y})$ can also be expressed as a Gaussian distribution

Kalman Filter methods

- Once we know (at least in principle!) the form of the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$ we have a choice:

  1) Either we can solve for the **mean** and the **covariance** of the posterior distribution:

$$\boldsymbol{x}_a = \int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x}$$

$$\mathbf{P}_a = \int (\boldsymbol{x} - \boldsymbol{x}_a)(\boldsymbol{x} - \boldsymbol{x}_a)^T\, p(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x}$$

Methods based on this approach include Optimum Interpolation (O.I.), Kalman Filter, Ensemble Kalman Filter (EnKF). These will all be discussed this week. The analysis found through this approach is called the minimum variance solution or the Best Linear Unbiased Estimate (BLUE).

Note: Kalman Filter based methods can be derived without making any assumption about the Gaussianity of the errors. However, only if all error distributions are Gaussian will the KF provide the correct posterior distribution (i.e. Bayes posterior pdf).

Variational methods

2) Alternatively, we might choose to estimate the **mode** of the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$, i.e. find the analysis $\boldsymbol{x}_a$ as the state that corresponds to the maximum of the posterior distribution (=> the most probable state):

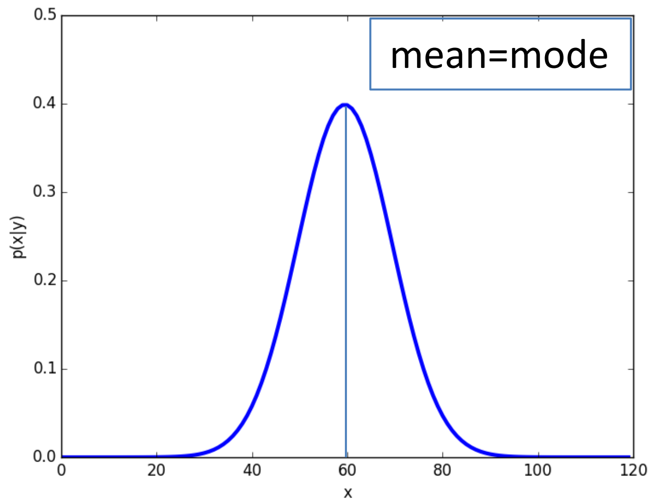
$$\boldsymbol{x}_a = \arg\max_{\boldsymbol{x}}\big(p(\boldsymbol{x}|\boldsymbol{y})\big)$$

This way of attacking the problem leads to the variational approach (3D-Var, 4D-Var). They will be covered extensively in this week's lectures. The solution found in this way is called the Maximum A-posteriori Probability state (MAP).

In the variational framework the linear and Gaussian assumptions can be relaxed, i.e. the full nonlinear analysis problem can be decomposed into a series of linear Gaussian problems (incremental 4D-Var, to be discussed later this week). However there is no guarantee of convergence!

Kalman Filter vs Variational methods

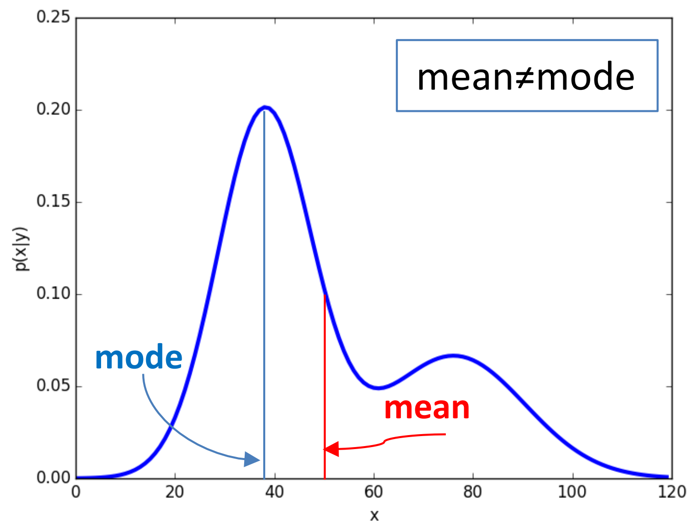
- For a Gaussian pdf the mean and the mode of the distribution coincide:



- Thus if all the error statistics are Gaussian the minimum variance and maximum a-posteriori solutions coincide!

Kalman Filter vs Variational methods

- For non-Gaussian pdfs the mean and the mode of the distribution generally differ:



- In non-Gaussian assimilation problems the minimum variance and maximum likelihood solutions will differ
- Which solution is better is problem dependent
- The more non-Gaussian the problem the more one needs information about the whole posterior pdf, not only its first two moments!

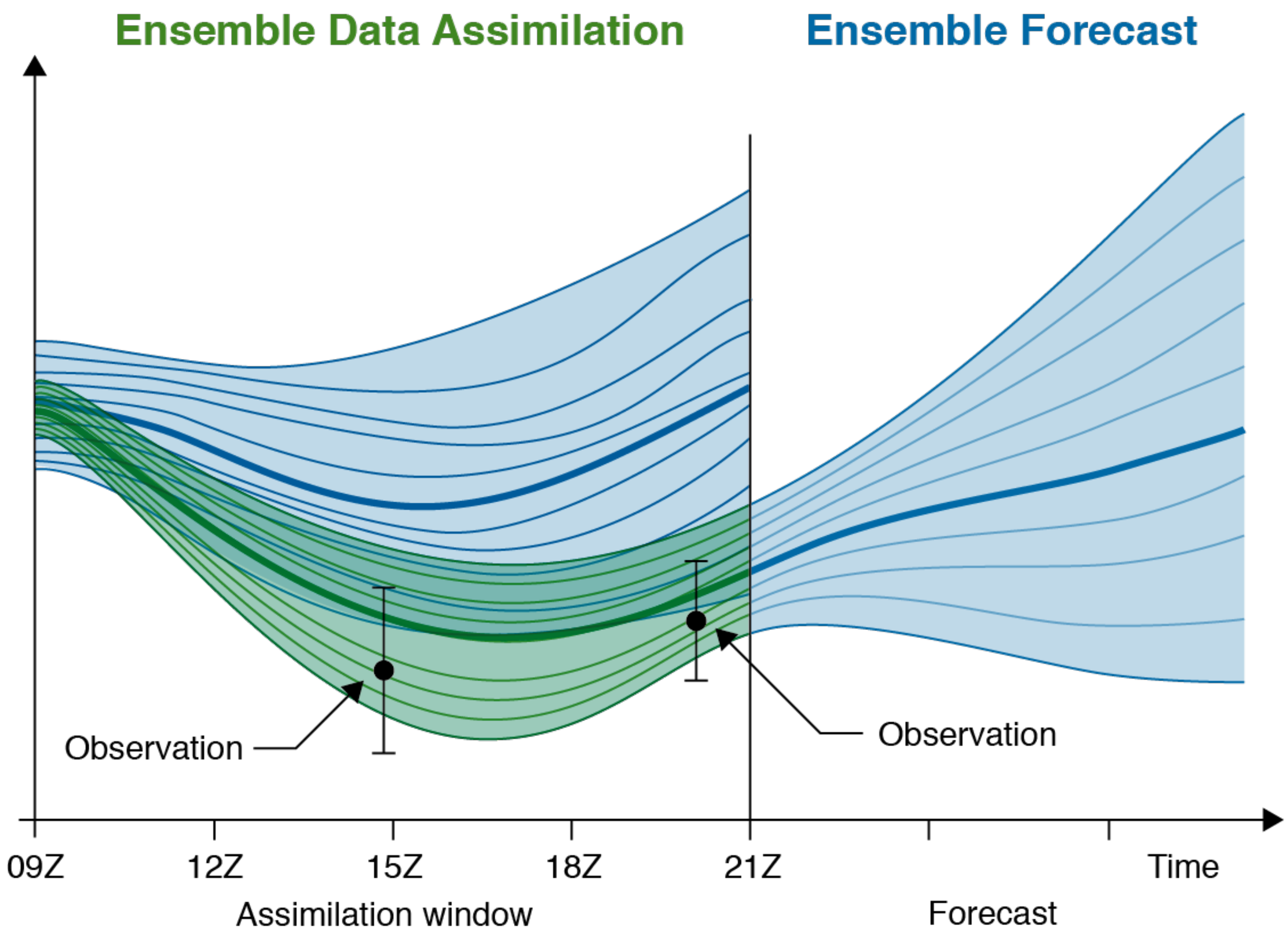
Hybrid DA methods

- Both Variational and Kalman Filter based analysis methods require estimates of the background state and its error covariances ($p(\boldsymbol{x}_b|\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{x}_b, \mathbf{P}_B)$)
- The background state is usually provided by an integration of the forecast model started from the previous analysis:

$$\boldsymbol{x}_b^t = \mathcal{M}(\boldsymbol{x}_a^{t-1})$$

- Hybrid DA methods: background (and analysis) error statistics are sampled with an Ensemble DA component (EnKF, Ensemble of 4DVars, EnsVar, etc)
- All major global NWP Centres run some form of hybrid data assimilation: a variational analysis cycle to estimate the mean/mode of the analysis pdf coupled with an ensemble data assimilation system to give a flow-dependent estimate of the second moments (covariances) of the error distributions.
- The ensemble DA component not only serves the purpose of estimating the background errors used in the analysis update, but it also provides a Monte Carlo sampling of the analysis pdf from which ensemble forecasts can be initialised

Hybrid DA methods



Hybrid DA-ML

- A more recent meaning of hybrid DA has emerged from the rapidly growing field which tries to merge DA and Machine Learning concepts: Hybrid DA-ML
- Basic idea:

"Use DA framework to improve forecast model **together with** estimate of initial state"

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}_b|\boldsymbol{x}) \;\; => \;\; p(\boldsymbol{x},\boldsymbol{w}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}_b|\boldsymbol{x})p(\boldsymbol{w}_b|\boldsymbol{w})$$

Where w are a set of physics model parameters or weights/biases for a ML model

- Various possible ways of doing this:

  1. Model parameter estimation in physics-based models
  2. Hybrid data-driven physical modelling
  3. ML online error correction
  4. …

- From a Bayesian (and also a technical) perspective variational DA and ML are solving similar problems with similar tools! More on this later this week…

Summary

- Data assimilation in NWP aims to optimally blend information from observations and model to produce an accurate and physically consistent estimate of the initial state of the atmosphere and of the other components of the Earth System
- Both observations and models are affected by systematic and random errors: these need to be evaluated and taken into account in order to produce a statistically optimal analysis
- The Bayesian approach provides a unified theoretical framework for data assimilation
- Particle Filters provide a Monte Carlo implementation of the Bayes' Law in data assimilation. Asymptotically correct for $N_{ens} \to \infty$, but unaffordable
- A Gaussian assumption on the error statistics is usually made to make the problem tractable in realistic geophysical DA
- Kalman Filter type methods and Variational methods can both be derived from Bayes' Law under these assumptions: they lead to the same solution for linear, Gaussian problems
- Hybrid data assimilation methods currently used in global NWP combine a variational analysis system with an ensemble data assimilation component for error estimation
- Hybrid DA-ML: The new kid on the block!

Bibliography

Ades M., Van Leeuwen P., 2014: The equivalent weights particle filter in a high dimensional system. Q. J. R. Meteorol. Soc.,141: 484–503.

Bocquet M., Pires C. A., Wu L., 2010: Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation. Mon. Wea. Rev., 138, 2997–3023, doi: 10.1175/2010MWR3164.1.

Farchi, A. and M. Bocquet, 2018: Comparison of local particle filters and new implementations. Nonlin. Processes Geophys., 25, 765–807. doi:10.5194/npg-25-765-2018

Geer A.J., 2021: Learning earth system models from observations: machine learning or data assimilation?Phil. Trans. R. Soc. A.3792020008920200089

Gordon, N.J., D.J. Salmond and A.F.M. Smith, 1993: A novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: IEE Proceedings on Radar and Signal Processing. Vol. 140. pp. 107-113

Hu, C., P. J. van Leeuwen, and J. L. Anderson, 2024: An Implementation of the Particle Flow Filter in an Atmospheric Model. *Mon. Wea. Rev.*, **152**, 2247–2264

Miyoshi, T., K. Kondo, and T. Imamura, 2014: The 10,240-member ensemble Kalman filtering with an intermediate AGCM, Geophys. Res. Lett., 41, 5264–5271, doi:10.1002/2014GL060863.

Pulido, M. and van Leeuwen, P.J. (2019) Sequential Monte Carlo with kernel embedded mappings: the mapping particle filter. *Journal of Computational Physics*, 396, 400–415.

Snyder, C., T. Bengtsson, and M. Morzfeld, 2015: Performance Bounds for Particle Filters Using the Optimal Proposal. Mon. Wea. Rev., 143, 4750–4761, doi: 10.1175/MWR-D-15-0144.1.

Wikle, C. K., and M. Berliner, 2007: A Bayesian tutorial for data assimilation. Physica D, 230, 1-16, doi:10.1016/j.physd.2006.09.017