Data assimilation techniques: The Kalman Filter

Henk Eskes
KNMI, the Netherlands

Notation

E.g. Ide et al, J.Met.Soc.Japan, 1997

$\mathbf{x}^t(t_i)$	The unknown "true" state vector of the system (discrete in space and time, e.g. an appropriate grid-box average, at a time t_i , of the true continuum state of the atmosphere). Dimension n.
$\mathbf{x}^f(t_i)$	The forecast of the state vector, obtained from a (non-linear) model, $\mathbf{x}^f(t_{i+1}) = M_i[\mathbf{x}^f(t_i)]$
$\mathbf{y}_i^0$	A vector of observations at time t_i , dimension p_i
$\mathbf{x}^a(t_i)$	The analysis of the state vector, after including the information from the observations

Normal distributions

A normal (or Gaussian) probability distribution for a random variable x is fully determined by the mean x_m and standard deviation | |

$$N(x_m, \square^2) \sim \exp[-\frac{(x - x_m)^2}{2\sqcap^2}]$$

This can be extended to vector quantities, with covariance matrix **P**

$$N(\mathbf{x}_m, \mathbf{P}) \sim \exp[-\frac{1}{2}(\mathbf{x} - \mathbf{x}_m)^T \mathbf{P}^{-1} (\mathbf{x} - \mathbf{x}_m)]$$

Normal distributions

The default assumption in data assimilation is to assume that the *a-priori* probability density functions (PDF) are normal distributions.

This is a convenient choice:

- Normal PDF's are described by the mean and covariance only: no need for higher-order moments
- The square in the exponent is easy to work with
- A Gaussian PDF remains Gaussian after linear operations
- Assimilation: when the *a-priori* PDFs are normal, and for linear operators, the *a-posteriori* PDF is also normal

Normal distributions

The default assumption in data assimilation is to assume that the *a-priori* probability density functions (PDF) are normal distributions.

This is also the most obvious choice:
- Because of the Central Limit Theorem

Central limit theorem

E.g. Grimmett and Stirzaker, Probability and random processes

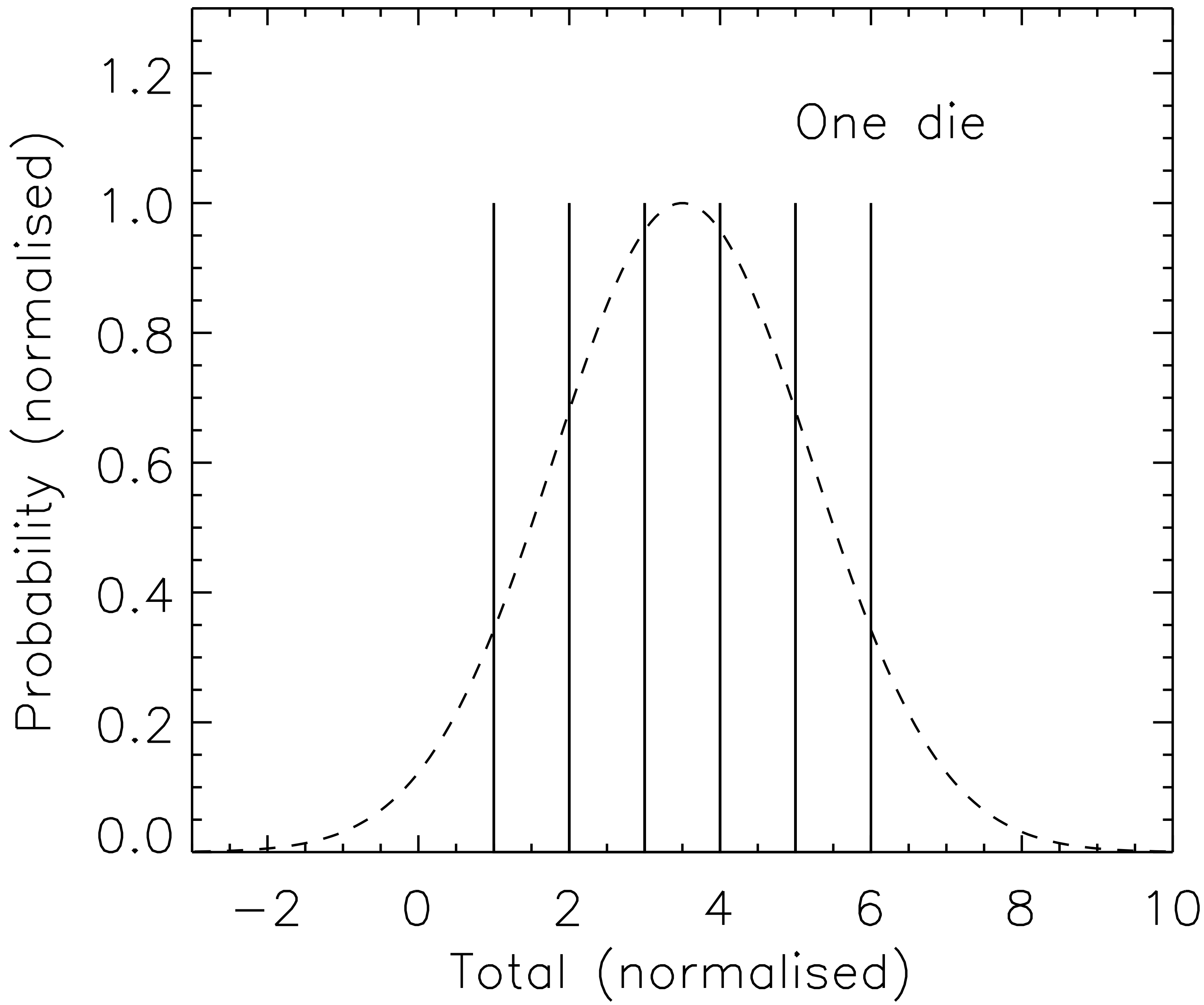
Let $X_1, X_2 \ldots$ be a sequence of independent identically distributed random variables with finite means μ and finite non-zero variance $\square^2$, and let

$$S_n = X_1 + X_2 + \ldots + X_n$$

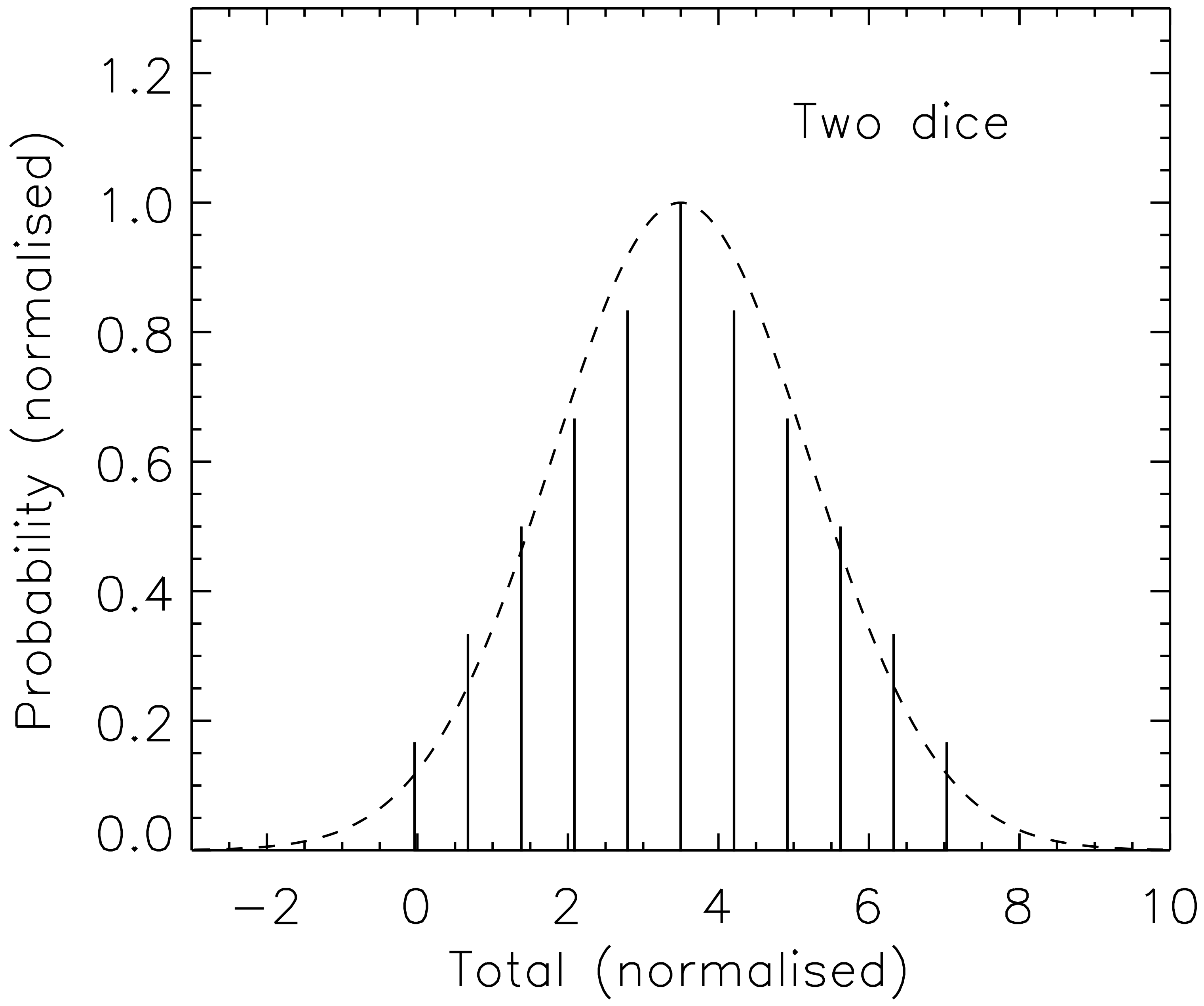
Then

$$\frac{S_n - n\mu}{\sqrt{n\square^2}} \rightarrow N(0,1) \quad \text{as} \quad n \rightarrow$$

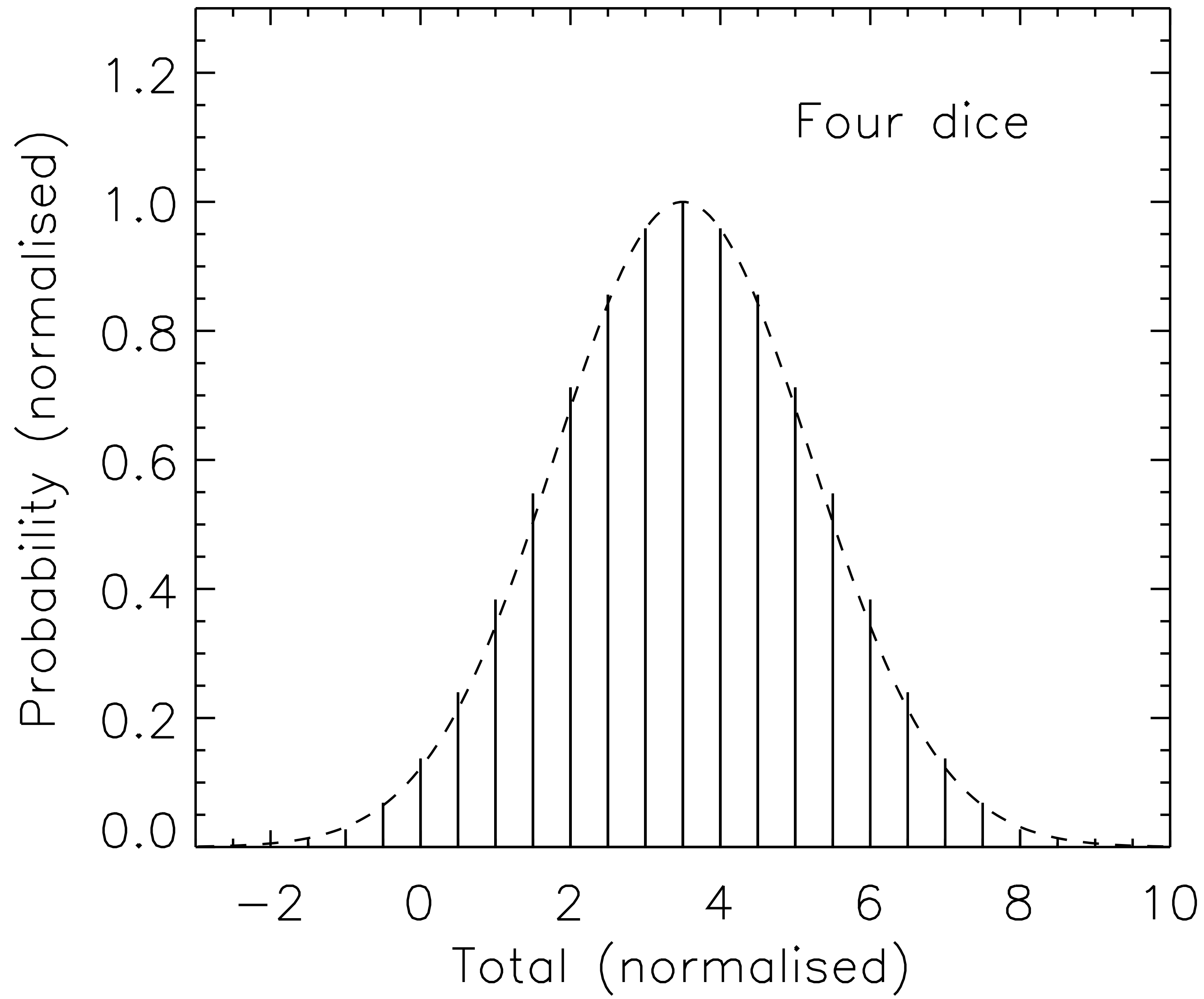
Central limit theorem: throwing dice



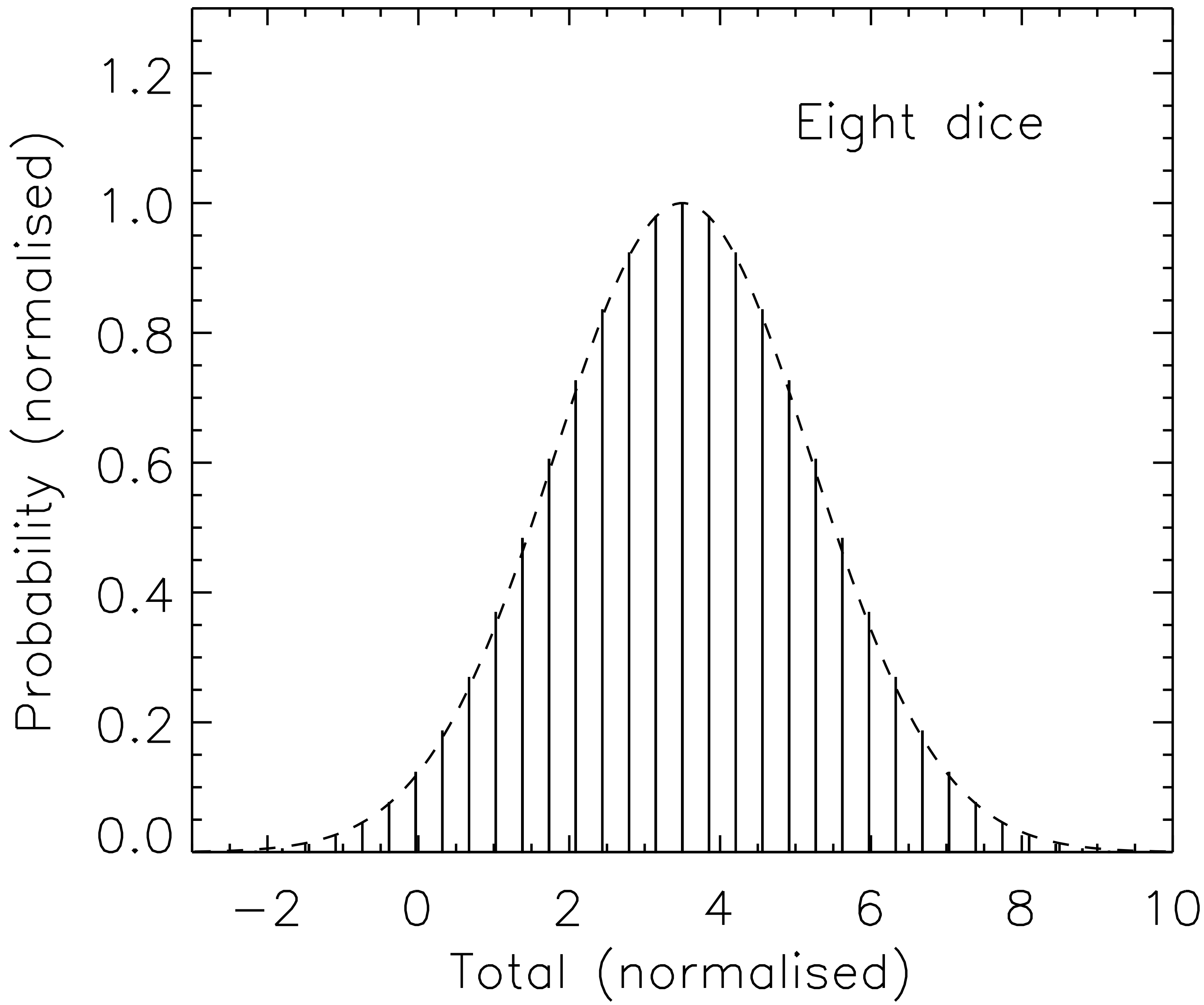
Central limit theorem: throwing dice



Central limit theorem: throwing dice



Central limit theorem: throwing dice



Conclusion:

The envelope of the probability distribution of the sum of a few dice is rapidly approaching a Gaussian

Central limit theorem

In words:

When the uncertainty of a quantity is the result of many independent random processes (error sources), then the probability distribution function (PDF) of the quantity will be approximately Gaussian (normal)

or

Without further knowledge a Gaussian distribution is the most natural candidate for the PDF

Kalman filter: starting points

Model and model error:

The model M describes the evolution of the state,

$$\mathbf{x}^f(t_{i+1}) = M[\mathbf{x}^f(t_i)]$$

The model will have errors,

$$\mathbf{x}^t(t_{i+1}) = M[\mathbf{x}^t(t_i)] + \square(t_i)$$

which will be assumed random, normally distributed, with mean zero and covariance $\mathbf{Q}(t_i)$

$$\langle \square \rangle = 0 \qquad \mathbf{Q} = \langle \square(t_i)\square(t_i)^T \rangle$$

For linear models $M_i = \mathbf{M}_i$, a matrix.

For weakly non-linear models a linearization may be performed about the trajectory $\mathbf{x}^f(t_i)$

Kalman filter: starting points

Observation and observation operator (1):

Observations that are available at time *t* are related to the true state vector by the observation operator,

$$\mathbf{y}_i^o = H_i[\mathbf{x}^t(t_i)] + \Box_i$$

The observation operator H_i may range from a simple linear interpolation to the position of the observation, to a complicated non-linear full radiative transfer model in the case of radiance observations.

Remote sensing observations generally involve the retrieval averaging kernel matrix $\mathbf{A}$ and retrieval *a-priori* states,

$$\mathbf{y}^o - \mathbf{y}^{o,ap} = \mathbf{A}\left(\mathbf{x}^t - \mathbf{x}^{ap}\right)$$

Kalman filter: starting points

Observation and observation operator (2):

$$\mathbf{y}_i^o = H_i[\mathbf{x}^t(t_i)] + \epsilon_i$$

The noise process is again assumed to be normal, with mean zero and covariance $\mathbf{R}$, combining errors of different origin,

- Instrumental and retrieval errors
- Averaging kernel errors
- Interpolation / representativeness errors

State vector covariance:

The error covariance associated with $\mathbf{x}^f$ is $\mathbf{P}^f$

$$\mathbf{P}^f(t_i) = \langle [\mathbf{x}^f(t_i) - \mathbf{x}^t(t_i)][\mathbf{x}^f(t_i) - \mathbf{x}^t(t_i)]^T \rangle$$

Example 1

One observation, one grid point

$$x^f = x^t + \square \qquad y = x^t + \square$$

$$\langle \square^2 \rangle = P \qquad \langle \square^2 \rangle = R \qquad \langle \square\square \rangle = 0$$

We choose as estimate of *x* a value in between the forecast and observation,

$$\hat{x} = (1-k)x^f + ky \qquad 0 \le k \le 1$$

Exercise:

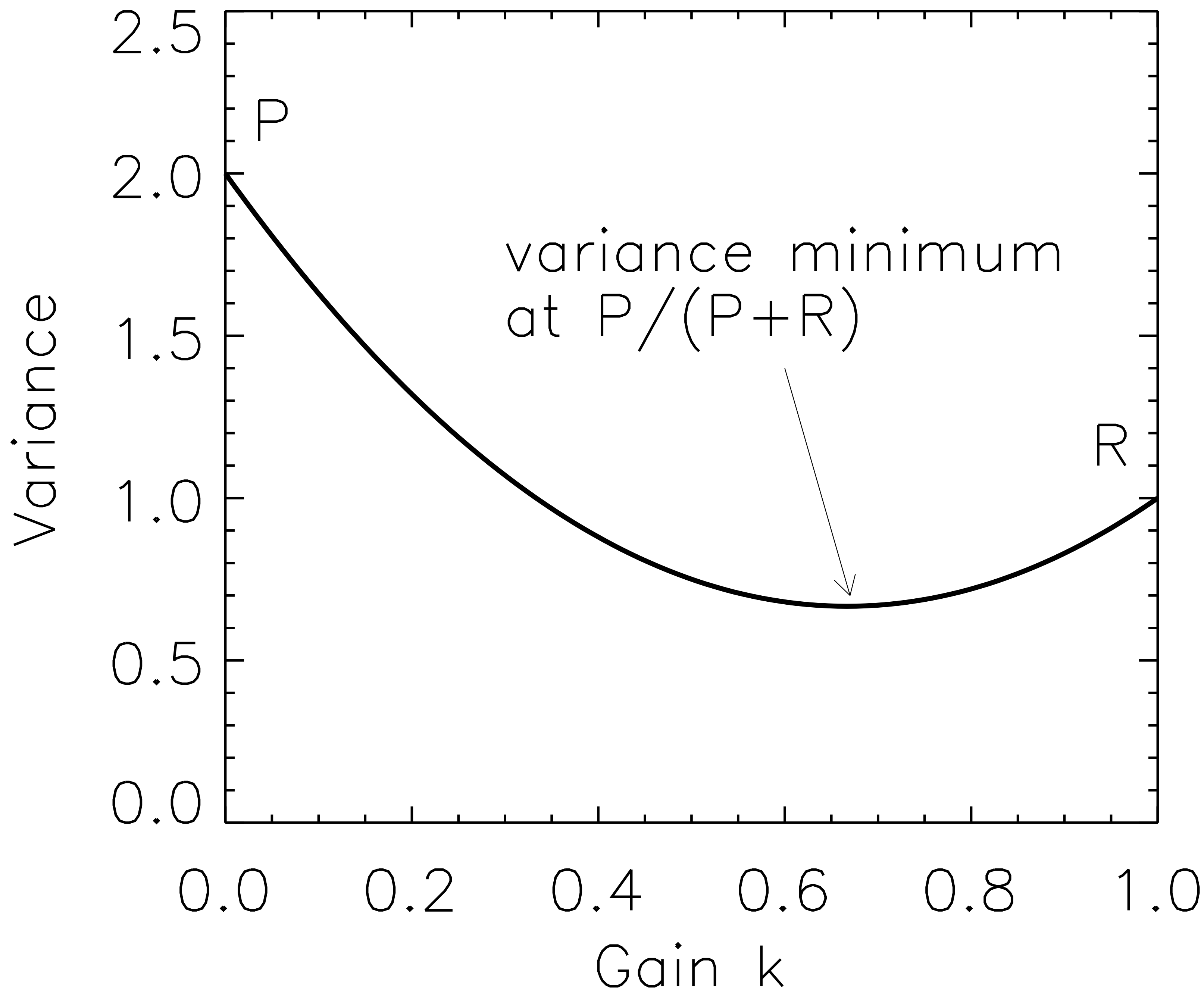
Show that the *a-posteriori* variance V is minimal for

$$k = P/(P+R) \qquad V = PR/(P+R)$$

Hint:

$$V = \langle (\hat{x} - x^t)^2 \rangle \quad \text{and solve} \quad \frac{\partial}{\partial k} V = 0$$

Example 1



2.5
2.0
1.5
1.0
0.5
0.0
Variance
P
R
variance minimum
at P/(P+R)
0.0
0.2
0.4
0.6
0.8
1.0
Gain k

Example 1, using Bayes rule

Conditional probability

$P_{x|y}$ What we want to know: the a-posteriori PDF of x, given that a measurement returns a value y

$P_{y|x}$ The conditional PDF of the measurement given that the state has a value x. $P_{y|x} \sim N(x, R)$

P_x The PDF of the state x. $P_x \sim N(x^f, P)$

P_y The *a-priori* PDF of the observation. This is just a normalisation factor.

$$P_y = \int P_{y|x} \, P_x \, dx$$

Example 1, using Bayes rule

Exercise:
Use Bayes rule

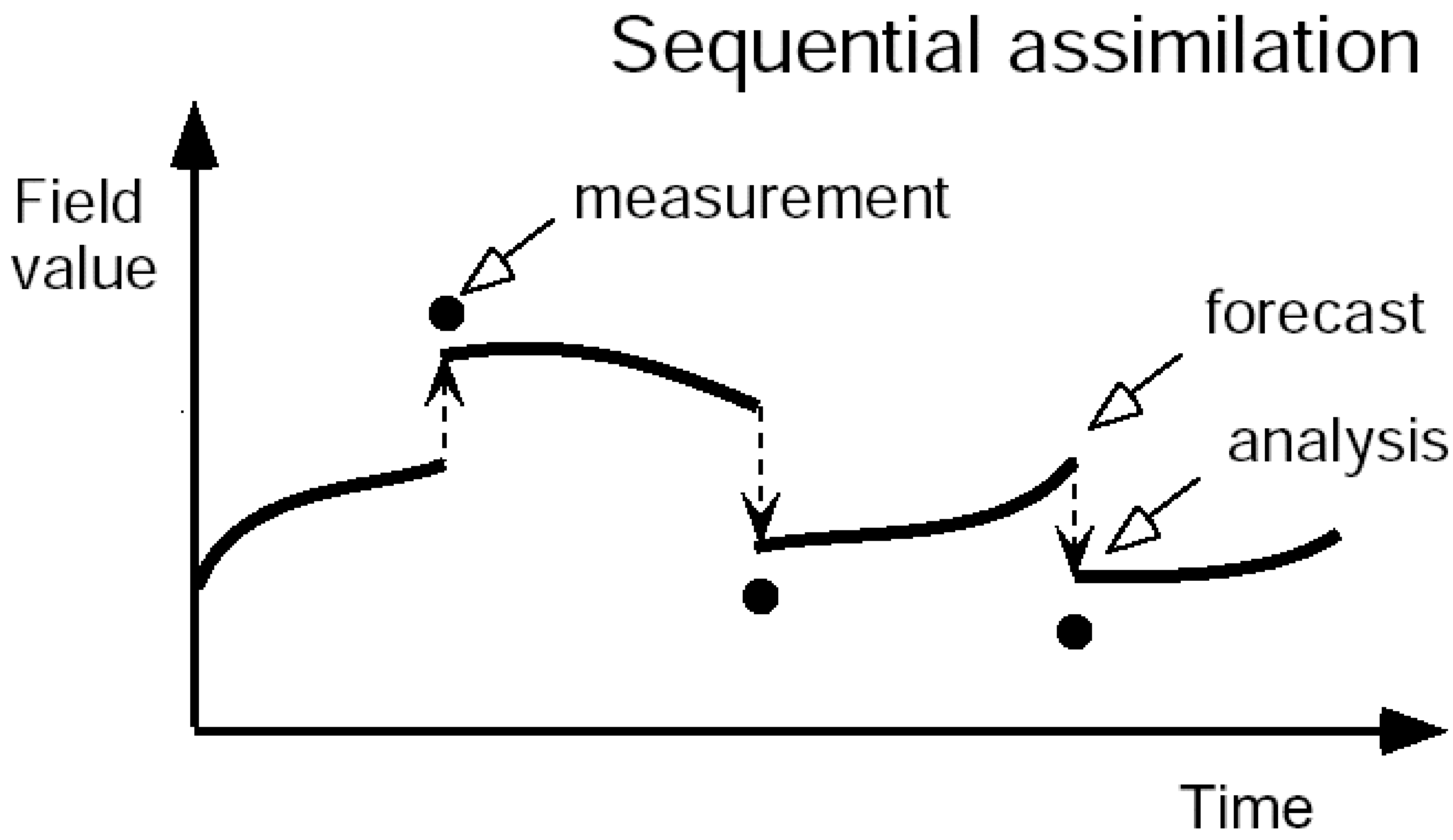
$$P_{x|y} = \frac{P_{y|x} P_x}{P_y}$$

to show that the *a-posteriori* PDF is equal to

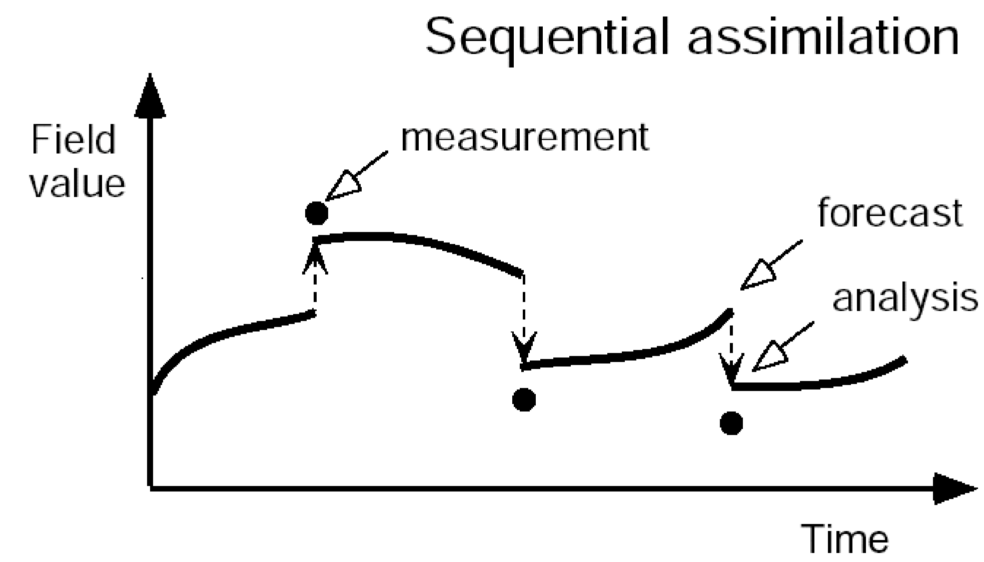
$$P_{x|y} \sim N\left[x^f + \frac{P}{P+R}(y - x^f), \frac{PR}{P+R}\right]$$

Note: the minimum variance estimate and maximum probability solutions are identical.
This result is quite general, related to the use of normal PDF's and linear operators.

Sequential assimilation: Kalman filter



Sequential assimilation: Kalman filter



Construct the optimal state (analysis) and *a-posteriori* covariance matrix by including observations step by step

- At time t_i the analysis is based on all previous observations, at times $t_j\,;\quad j \leq i$ The information from previous time steps is accumulated in the covariance matrix.

One Kalman cycle consists of

- Propagation of the state vector and covariance in time
- Analysis of the state vector and covariance, based on the observations available at that time

Kalman filter: forecast step

Eq. 1 extended Kalman filter: State vector forecast

$$\mathbf{x}^f(t_{i+1}) = M_i[\mathbf{x}^a(t_i)]$$

Eq. 2 extended Kalman filter: Error covariance forecast

$$\mathbf{P}^f(t_{i+1}) = \mathbf{M}_i\mathbf{P}^a(t_i)\mathbf{M}_i^T + \mathbf{Q}(t_i)$$

The error covariance is propagated in time in the same way as the state vector, namely through the model. **P** increases with time due to the model error covariance which is added every time step.

Kalman filter: covariance forecast

Exercise:

Derive the second Kalman equation

Hint:

$$\mathbf{P}^f(t_{i+1}) = \left\langle [\mathbf{x}^f(t_{i+1}) - \mathbf{x}^t(t_{i+1})][\mathbf{x}^f(t_{i+1}) - \mathbf{x}^t(t_{i+1})]^T \right\rangle$$

and use the linear model to express $\mathbf{x}^f(t_{i+1})$ in terms of $\mathbf{x}^f(t_i)$

Kalman filter: covariance forecast

Example: passive tracer transport

Lagrangian approach: define the model on a set of trajectories instead of a fixed grid. The model for the passive tracer is now a simple unity matrix.

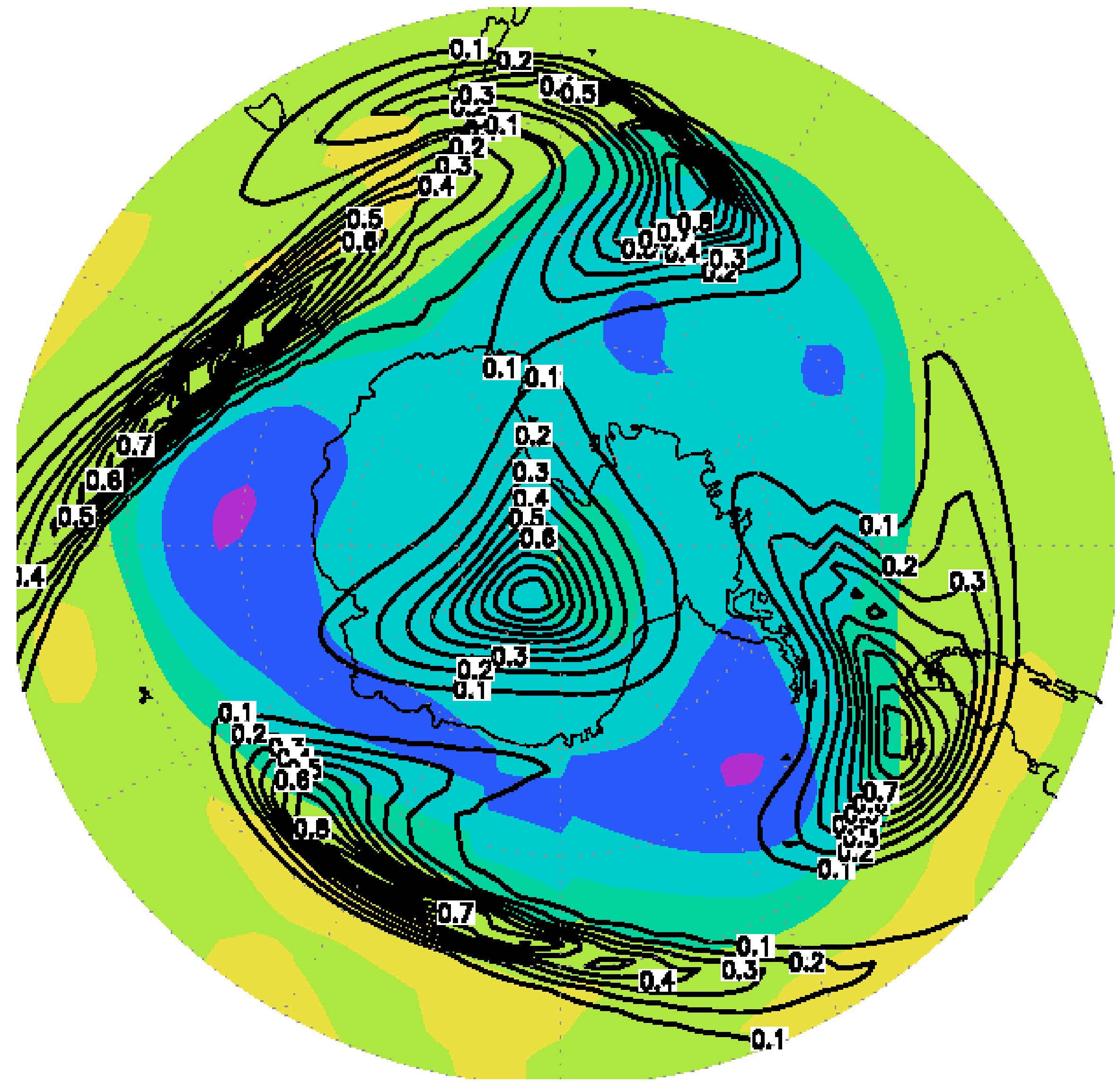
$$\mathbf{M}_i = \mathbf{I} \qquad \mathbf{P}^f(t_{i+1}) = \mathbf{P}^a(t_i) + \mathbf{Q}(t_i)$$

Or: the passive tracer variance of air parcels, and the correlations between parcels are conserved in time in the absence of observations and for a perfect model $\mathbf{Q}(t_i) = 0$

Kalman filter: covariance forecast

Example:
passive tracer transport

The image shows several rows of the covariance matrix after 24 h of 2D advection, starting from a simple homogeneous isotropic correlation matrix.



Source: Kris Wargan, NASA

Kalman filter: analysis step

Kalman gain matrix

$$\mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^T \left[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right]^{-1}$$

Eq. 3 extended Kalman filter: State vector analysis

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}_i\left(\mathbf{y}_i^o - H_i[\mathbf{x}^f(t_i)]\right)$$

Eq. 4 extended Kalman filter: Error covariance analysis

$$\mathbf{P}^a(t_i) = (I - \mathbf{K}_i\mathbf{H}_i)\,\mathbf{P}^f(t_i)$$

Kalman filter: analysis step

Derivation of Kalman equations 3 and 4 (linear operators)

The derivation follows Bayes rule (see the example)

$$-2\ln P_{x|y} = [y_i^o - \mathbf{H}_i\mathbf{x}(t_i)]^T \mathbf{R}_i^{-1} [y_i^o - \mathbf{H}_i\mathbf{x}(t_i)] + \left[\mathbf{x}(t_i) - \mathbf{x}^f(t_i)\right]^T \mathbf{P}^f(t_i)^{-1} \left[\mathbf{x}(t_i) - \mathbf{x}^f(t_i)\right] + c_1$$

The sum of quadratic terms is also quadratic, so this can be written as

$$-2\ln P_{x|y} = [\mathbf{x}(t_i) - \mathbf{x}^a(t_i)]^T \mathbf{P}^a(t_i)^{-1} [\mathbf{x}(t_i) - \mathbf{x}^a(t_i)] + c_2$$

These two equations define $\mathbf{x}^a(t_i)$ and $\mathbf{P}^a(t_i)$

Exercise :-)

Warning: this equivalence will lead to matrix expressions that look different from, but are equivalent to, the analysis Kalman equations

Kalman filter: analysis step

State analysis interpretation:

The Kalman gain matrix controls how much the analysis is forced to the observations

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}_i\left(\mathbf{y}_i^o - H_i[\mathbf{x}^f(t_i)]\right)$$

- When **K** is small, the analysis will approach the forecast
- When **K** is "large", the analysis will reproduce the observations as much as possible

[remember the example, where $k = P/(P+R)$]

Kalman filter: analysis step

Covariance analysis:

The covariance analysis equation can be written as

$$\mathbf{P}^a(t_i) = \mathbf{P}^f(t_i) - \mathbf{P}^f(t_i)\mathbf{H}_i^T\left[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right]^{-1}\mathbf{H}_i\mathbf{P}^f(t_i)$$

State analysis:

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{P}^f(t_i)\mathbf{H}_i^T\left[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right]^{-1}\left(\mathbf{y}_i^o - H_i[\mathbf{x}^f(t_i)]\right)$$

The * indicate the dimension of the space of the matrices:

* State space, dimension n

* Observation space, dimension p_i

Kalman filter: analysis step

One observation:

$\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T$	The variance at the observation, write P_{oo}^f
$\mathbf{R}_i$	Now a number, write R
$\mathbf{H}_i$	Now a vector of dimension *n*. Suppose this is a simple interpolation, e.g. one at the gridbox with the observation, zero elsewhere

The analysis equation for the variance in grid box l :

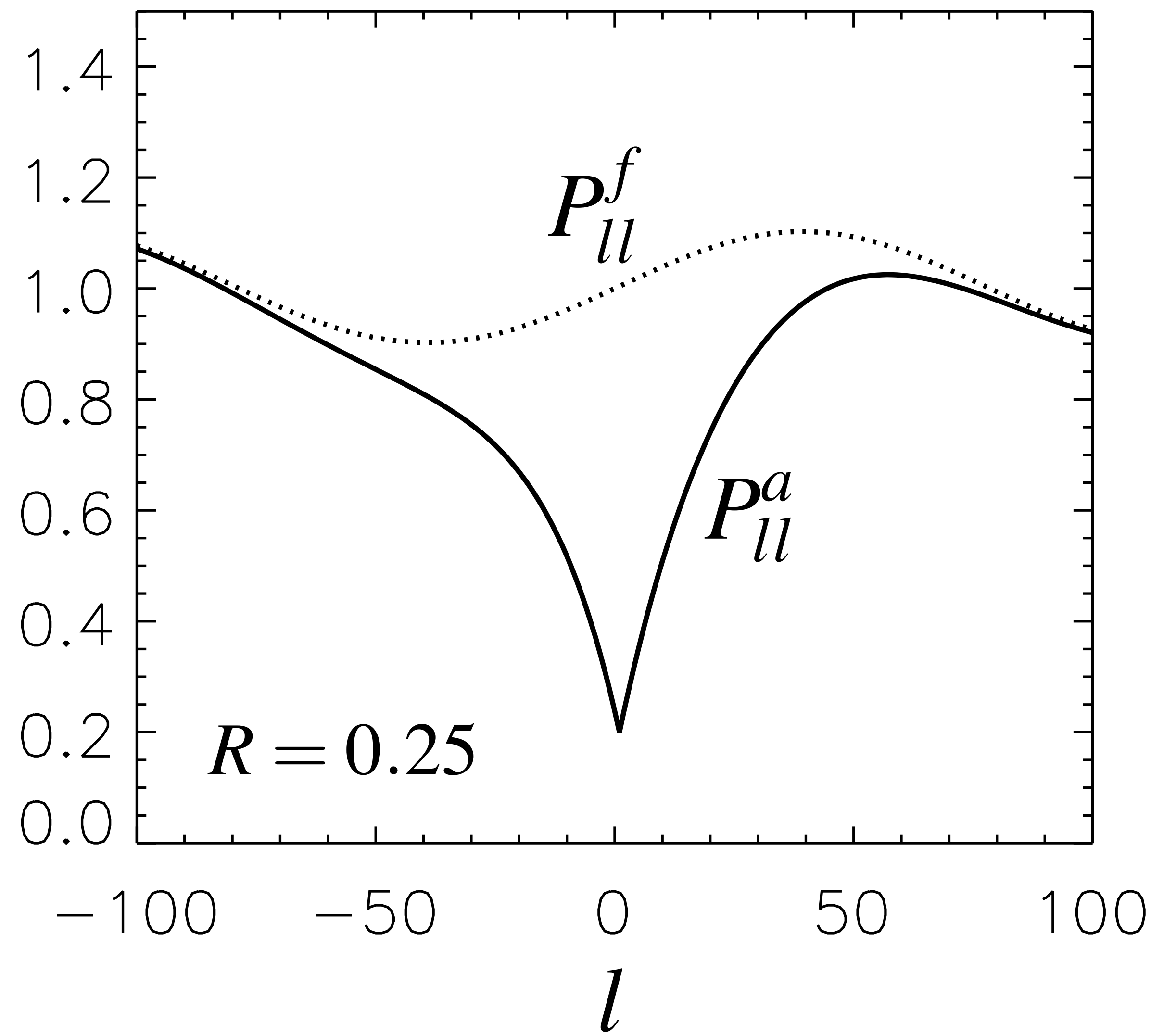
$$P_{ll}^a = P_{ll}^f - \frac{P_{lo}^f P_{ol}^f}{P_{oo}^f + R}$$

Kalman filter: covariance analysis

One observation

Correlation model:

$$P^f_{lo} = \square_l \square_o e^{-|l|/L} \quad ; L = 40$$



Note:

- **P** reduced at the observations with a factor $R/(P+R)$
- **P** also reduced in the neighbourhood of the observation. Influence radius determined by the correlation length L.

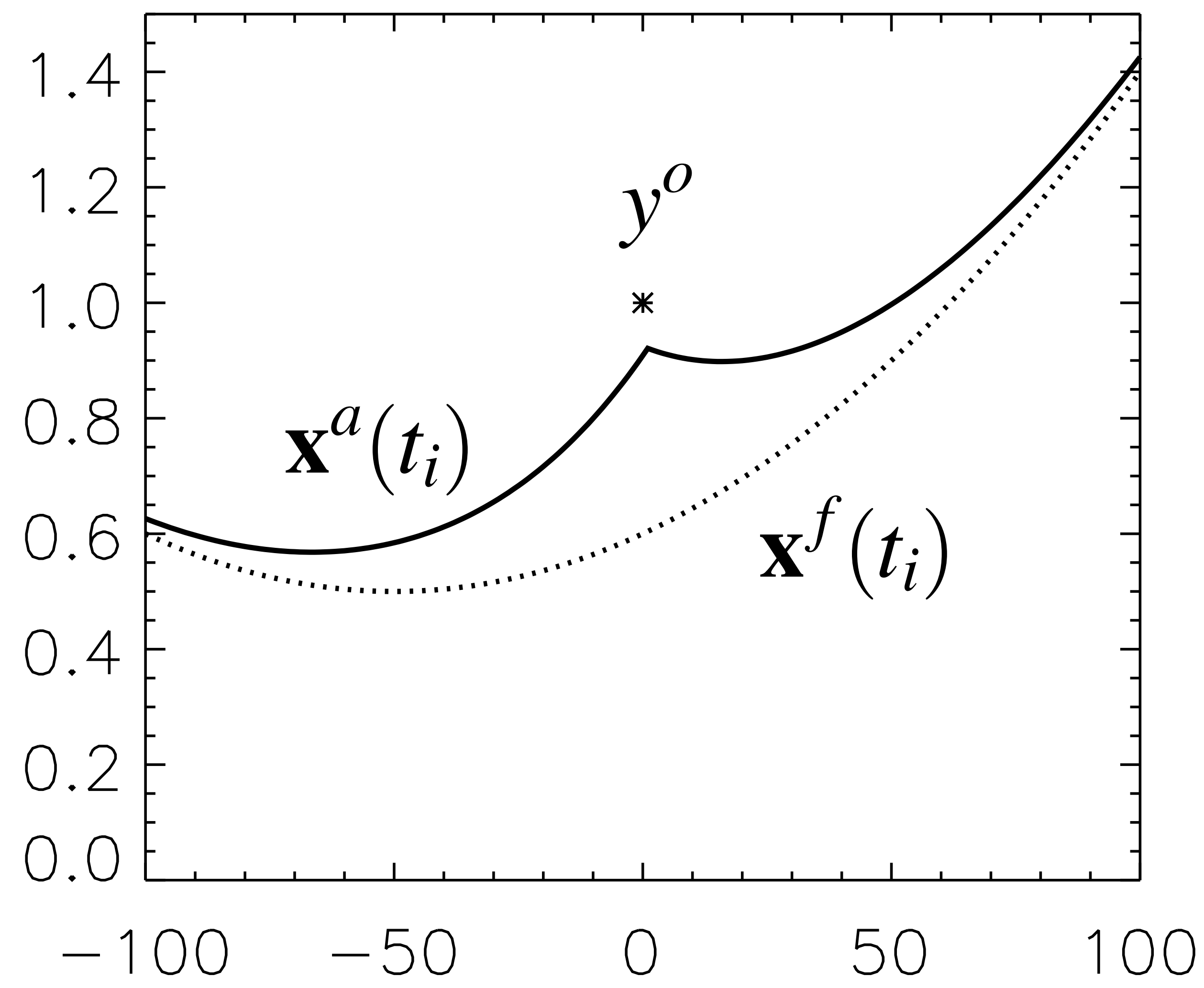
Kalman filter: covariance analysis

One observation

The corresponding state analysis

Note:

- Again the information is used in an area determined by the length L

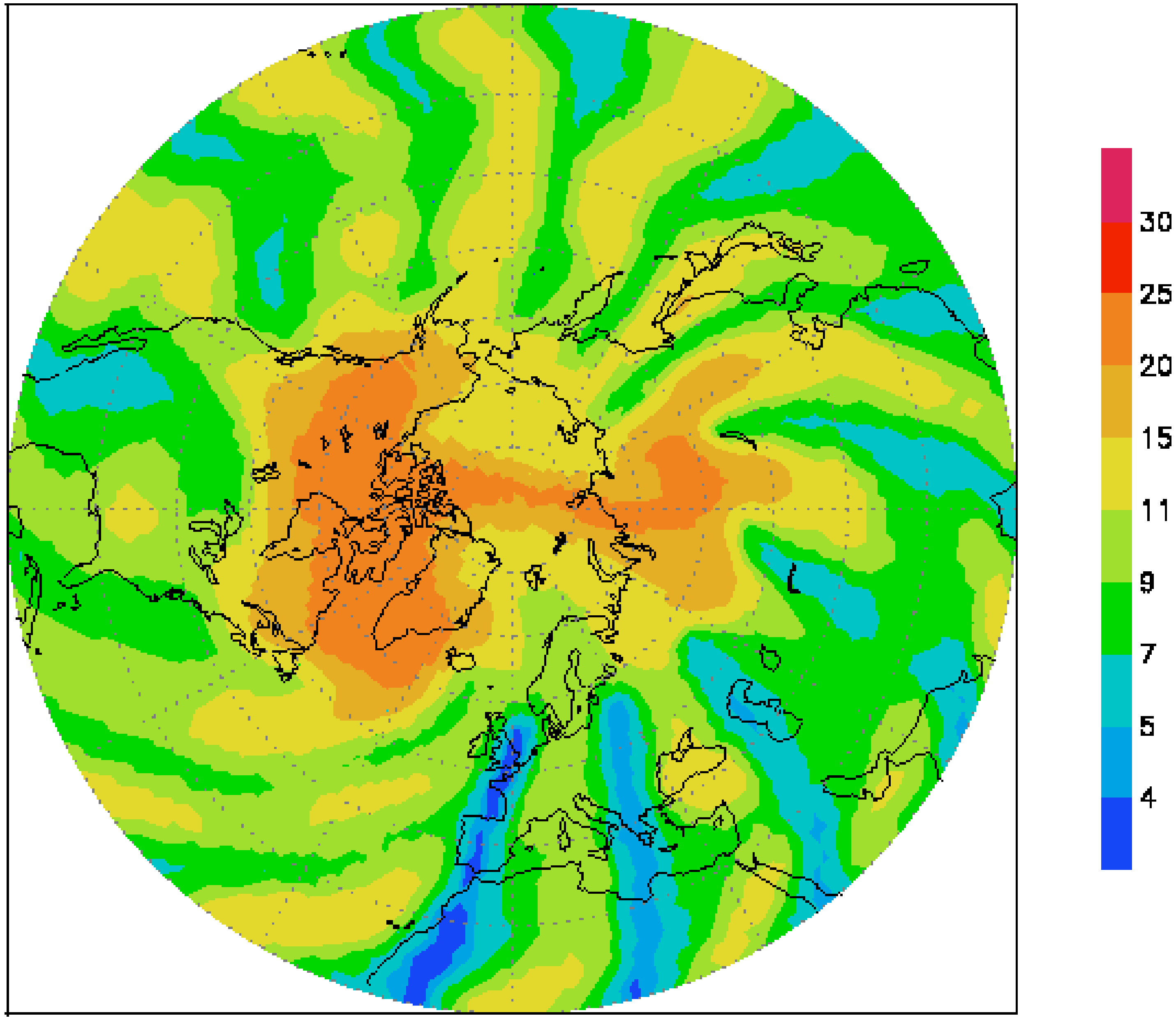


Kalman filter: covariance

Example:
ozone column assimilation

The plot demonstrates key aspects of the Kalman filter covariance evolution:

- Reduction at observations
- Model error
  (error growth)
- Covariance advection



Total ozone standard deviation (DU)

Correlations

Importance of correlations

- Information spread over a region with radius given by the model error covariance correlation length
- Efficient removal of model biases with "few" observations
- Avoids spurious "spikes" in analysis at observations

Significance of correlation length

- Acts as a low-pass filter for the observations:
  - The model is strongly forced towards the observational information which varies slowly w.r.t. correlation length
  - Nearby observations: the analysis adopts the mean of the observations and the variability of the observations has only a minor influence on the analysed state

Covariance matrices: many unknowns

The Kalman filter is optimal only when the *a-priori* covariance matrices are realistic

Major problem:

How to choose all the matrix elements of **Q** and **R** ?

Recipes:

- Simple model of **Q** and **R**, with just a couple of parameters, to be determined from the observation-minus-forecast statistics
- "NMC method", for time independent **P** :
  use the differences between the analyses and forecast fields as a measure of the covariance (diagonal, correlations)

Kalman filter: computational problem

For practical atmosphere/ocean applications the Kalman filter is far too expensive:

- Only state vectors with ≤ 1000 elements are practical, but a typical state-of-the-art model has 10^6 elements
- Example: if applying the model takes 1 min for $n = 10^6$ then propagation of the variance will take $2n$ times as long, i.e. about two years ! Storage of the complete covariance matrix is also enormous.

Conclusion:

- Efficient approximations are needed for large problems

Kalman filter: practical aspects

A few practical problems:

- Covariance matrices are positive definite (needed to calculate the inverse). Truncations and rounding may easily cause negative eigenvalues.
- The model error term $\mathbf{Q}$ should be large enough to explain the observed $\mathbf{y}_i^o - H_i[\mathbf{x}^f(t_i)]$
  Filter divergence: occurs if a simple choice of $\mathbf{Q}$ leads to values of P which are unrealistically small in parts of the state space. The model will drift away from the observations

$\square^2$ test: e.g. Menard, 2000

$$\left\langle \left(\mathbf{y}_i^o - H_i[\mathbf{x}^f(t_i)]\right)^T \left[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right]^{-1} \left(\mathbf{y}_i^o - H_i[\mathbf{x}^f(t_i)]\right) \right\rangle \approx 1$$

Optimal (statistical) interpolation

Until recently, the OI sub-optimal filter was the most widespread scheme for numerical weather prediction

OI approximation:

Replace the covariance matrix $\mathbf{P}$ by a prescribed, time-independent "background" covariance $\mathbf{B}$. The Kalman filter reduces to

$$\mathbf{x}^f(t_{i+1}) = M_i[\mathbf{x}^a(t_i)]$$
$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{B}\mathbf{H}_i^T\left[\mathbf{H}_i\mathbf{B}\mathbf{H}_i^T + \mathbf{R}_i\right]^{-1}\left(\mathbf{y}_i^o - H_i[\mathbf{x}^f(t_i)]\right)$$

The expensive covariance forecast and analysis equations are avoided

Sub-optimal Kalman filter

Several fundamental Kalman filter properties can be maintained by expressing the covariance as a product of a time-dependent diagonal matrix and a time-independent correlation matrix.

$$\mathbf{B} = \mathbf{D}^{1/2}\mathbf{C}\mathbf{D}^{1/2}$$

$$\mathbf{D}^f(t_{i+1}) = N\left[\mathbf{D}^a(t_i)\right]$$

$$\mathbf{B}^a(t_i) = \mathbf{B}^f(t_i) - \mathbf{B}^f(t_i)\mathbf{H}_i^T\left[\mathbf{H}_i\mathbf{B}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right]^{-1}\mathbf{H}_i\mathbf{B}^f(t_i)$$

$$\mathbf{D}^a(t_i) = \mathrm{diag}\left[\mathbf{B}^a(t_i)\right]$$

A variance propagation model has been introduced here

Low rank Kalman filters

Idea:
Use a finite subset of vectors (leading eigenvectors, random ensemble) to describe the covariance matrix

- Reduced rank square root filter (Verlaan & Heemink, 1997)
- Ensemble Kalman filter (Evensen, 1994)
- Singular evolutive extended/interpolated Kalman (SEEK/SEIK; Pham, 1998; Verron, 1999)
- Error subspace statistical estimation (Lermusiaux, 1999)
- ECMWF (M. Fisher)

Literature

Books:

- Daley, *Atmospheric Data Analysis*, Cambridge, 1991
- Ghil & Malanotte-Rizzoli, *Data assimilation in meteorology and oceanography,* Advances in Geophysics, Academic Press, 1991.
- Rodgers, *Inverse methods for atmospheric sounding - theory and practice*, World Scientific, 2000.
- Jazwinski, *Stochastic processes and filtering theory*, Academic Press, 1970.
- Grimmett and Stirzaker, *Probability and random processes,* Clarendon Press, 1992.

Literature

Papers:

- Ide et al, *Unified notation for data assimilation: operational, sequential and variational*, J.Met.Soc.Japan, 75, 181, 1997
- Lorenc, *Analysis methods for NWP*, QJRMS, 112, 1177, 1986
- Heemink, *Filtertheorie,* Lecture notes, TUDelft.
- Cohn, *An introduction to estimation theory,* J.Met.Soc.Japan, 75, 257, 1997.
- Todling, Estimation theory and atmospheric data assimilation, Workshop on inverse methods in global biogeochemical cycles, Crete, Greece 1998.
- ECMWF Meteorological Training Course, Lecture notes by Bouttier, Undén, Eyre, Courtier, Holm, ...

Literature

Papers:

- Menard, *Assimilation of stratospheric chemical tracer observations using a Kalman filter,* MWR 128, 2654, 2000.

Low rank Kalman filter papers:

- Evensen, *Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, JGR 99, 10143, 1994
- Verlaan & Heemink, *Reduced rank square root filters for large-scale data assimilation problems,* Second Intl. Symp. on Assimilation of Observations in Meteorology and Oceanography, p247, WMO, 1995

Literature

Low rank Kalman filter papers (cont):

- Pham et al, *A singular Evolutive Extended Kalman filter for data assimilation in oceanography,* J. of Marine Systems, 16, 323, 1998.
- Verron et al, *An extended Kalman filter to assimilate satellite altimeter data into a non-linear numerical model of the tropical pacific ocean,* JGR 104, 5441, 1999
- Lermusiaux & Robinson, *Data Assimilation via error subspace statistics,* MWR 127, 1385, 1999