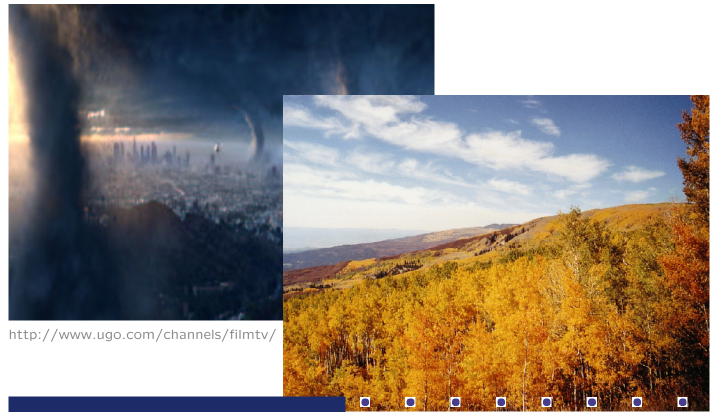
“Ensembles” in Numerical Weather Prediction & the Ensemble Kalman Filter



http://www.ugo.com/channels/filmtv/

Overview

- Aspects of Numerical Weather prediction
- (Ensemble) Kalman filter
- Statistical approach to prevent filter divergence

Thomas Bengtsson, Jeff Anderson, Doug Nychka

http://www.cgd.ucar.edu/∼furrer

Numerical Weather Prediction

Forecasting weather by the use of numerical models and observations, run on high speed computers.

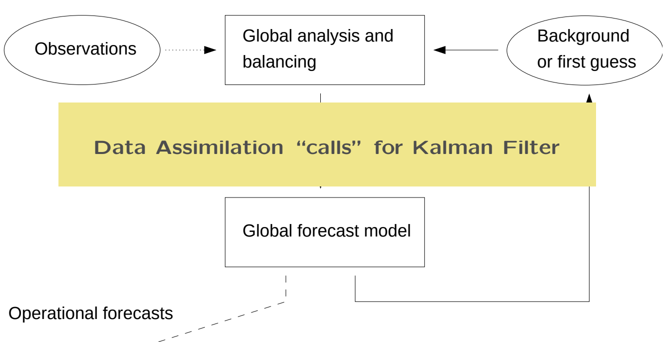
Numerical Models

- Physical laws of flow
- Simplification
- Discretization
- Parameterization

Observations

- Surface (towers, ships)
- Altitude (planes, ballons)
- Radar
- Satellite data

NWP Scheme



Kalman Filter

State and space equations:

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{R}_t)$$
$$\mathbf{x}_{t+1} = \mathbf{G}_t\mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{Q}_t)$$

where

$\mathbf{y}_t$: observations — $\mathbf{H}_t$: measurement operator
$\mathbf{x}_t$: state — $\mathbf{G}_t$: system operator

$\mathbf{w}_t$: measurement error — $\mathbf{R}_t$: measurement covariance matrix
$\mathbf{v}_t$: state error — $\mathbf{Q}_t$: state error covariance matrix

Kalman Filter

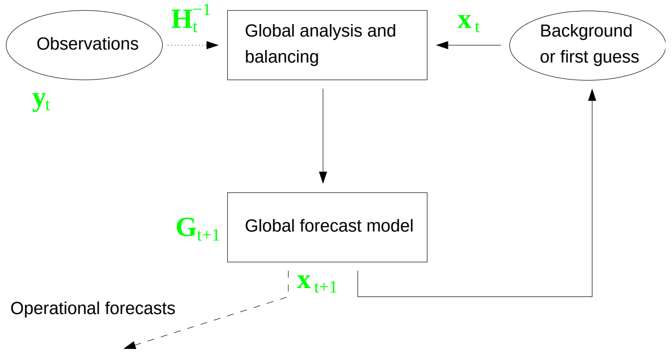
State and space equations:

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{w}_t \qquad \mathbf{w}_t \sim \mathcal{N}_r(\mathbf{0}, \mathbf{R}_t)$$
$$\mathbf{x}_{t+1} = \mathbf{G}_t\mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathcal{N}_q(\mathbf{0}, \mathbf{Q}_t)$$

We aim to update our knowledge of the state $\mathbf{x}_t$ given the new data $\mathbf{y}_t$.

The Kalman filter is an iterative procedure based on
– Bayes approach
– conditional expectation
– . . .

NWP within KF Scheme



Kalman Filter

The standard Kalman filter uses the initial knowledge

$$p(\mathbf{x}_t \,|\, \mathcal{Y}_{t-1}) \sim \mathcal{N}_q(\mathbf{x}_t^f, \mathbf{P}_t^f) \qquad \mathcal{Y}_{t-1} = \{\mathbf{y}_{t-1}, \dots, \mathbf{y}_0\}$$

then

$$p(\mathbf{x}_t \,|\, \mathcal{Y}_t) \sim \mathcal{N}_q(\mathbf{x}_t^a, \mathbf{P}_t^a) \qquad \begin{aligned} \mathbf{x}_t^a &= \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^f) \\ \mathbf{P}_t^a &= (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_t^f \\ \mathbf{K}_t &= \mathbf{P}_t^f\mathbf{H}_t^\mathsf{T}(\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^\mathsf{T} + \mathbf{R}_t)^{-1} \end{aligned}$$

$$p(\mathbf{x}_{t+1} \,|\, \mathcal{Y}_t) \sim \mathcal{N}_q(\mathbf{x}_{t+1}^f, \mathbf{P}_{t+1}^f) \qquad \begin{aligned} \mathbf{x}_{t+1}^f &= \mathbf{G}_t\mathbf{x}_t^a \\ \mathbf{P}_{t+1}^f &= \mathbf{G}_t\mathbf{P}_t^a\mathbf{G}_t^\mathsf{T} + \mathbf{Q}_t \end{aligned}$$

Kalman Filter

What if we have:

- nonlinear state $\mathbf{G}_t(\mathbf{x}_t)$
  $\rightsquigarrow$ linearize the state via Taylor (Extended KF)
- non Gaussian errors
  $\rightsquigarrow$ distributions do not have a closed form
  $\rightsquigarrow$ KF is still the best linear predictor
- outliers, . . .
  $\rightsquigarrow$ work for statisticians. . .

"High" Dimensions

KF in one and higher dimensions have similar properties.

If "high" gets really high, say 10^6, 10^7:
$\longrightarrow$ Infeasible to calculate the densities

Ensemble Kalman filter (EnsKF):
propagation of the forecast distribution with a small ensemble.

Unscented Kalman filter:
description of the moments with a small specific sample.

EnsKF

Assume a sample $\{\mathbf{x}_{t,i}^f\}$ for the forecast distribution $p(\mathbf{x}_t \,|\, \mathcal{Y}_{t-1})$.
The update is performed according to

$$\mathbf{x}_{t,i}^a = \mathbf{x}_{t,i}^f + \mathbf{K}_t\big(\mathbf{y}_t + \varepsilon_{t,i} - \mathbf{H}_t\mathbf{x}_{t,i}^f\big) \qquad \{\varepsilon_{t,i}\} \overset{\text{iid}}{\sim} \mathcal{N}_r(\mathbf{0}, \mathbf{R}_t)$$

Simplifications:

- $\mathbf{K}_t$ is replaced by $\widehat{\mathbf{K}}_t = \widehat{\mathbf{P}}_t^f\mathbf{H}_t{}^\mathsf{T}(\mathbf{H}_t\widehat{\mathbf{P}}_t^f\mathbf{H}_t{}^\mathsf{T} + \mathbf{R}_t)^{-1}$ for some estimate $\widehat{\mathbf{P}}_t^f$ of $\mathbf{P}_t^f$.
- $\mathbf{R}_t = r_t\mathbf{I}$, $\mathbf{Q}_t = \mathbf{0}$: serial computations of "inverses".
- Square root filter:
  deterministic version, no need to draw samples.

EnsKF: $+/-$

- Is feasible with observation vector is of size $\mathcal{O}(10^6)$ and state vector is of size $\mathcal{O}(10^7)$.
- Works well even for very small ensemble sizes: $n = 10, 40, \dots$ **if** we take a well chosen estimate of $\mathbf{P}_t^f$.
- Some "tuning" is needed to prevent filter divergence.

Operational EnsKF in NWP

EnsKF has essentially three goals:

- improve the forecast by ensemble averaging
- provide an indication of the reliability of the forecast
- provide a quantitative basis for probabilistic forecasting

Operational EnsKF in NWP

Practically, EnsKF may be performed by:

- choosing different initial or boundary conditions
- using different numerical techniques to solve the PDEs
- using different parameterizations

Statistical Questions

We need to choose an estimate or approximation of $\mathbf{P}_t^f$.

- What estimate or approximation should we choose?
- What is the optimal ensemble size n?
- What is the optimal "tuning"?

Approximations of $\widetilde{\mathbf{P}}_t^f$

Typical examples for approximations of $\widetilde{\mathbf{P}}_t^f$:

1. $\mathbf{P}_t^f$ is estimated with n ensembles
   $\rightsquigarrow \quad \widetilde{\mathbf{P}}_t^f = \widehat{\mathbf{P}}_t^f$.
2. $\mathbf{P}_t^f$ is estimated with n ensembles and tapered with $\mathbf{C}$
   $\rightsquigarrow \quad \widetilde{\mathbf{P}}_t^f = \widehat{\mathbf{P}}_t^f \circ \mathbf{C}$.
3. $\mathbf{P}_t^f$ is inflated/boosted with $\rho > 1$
   $\rightsquigarrow \quad \widetilde{\mathbf{P}}_t^f = \rho \widehat{\mathbf{P}}_t^f$ or $\widetilde{\mathbf{P}}_t^f = \rho \widehat{\mathbf{P}}_t^f \circ \mathbf{C}$.

Goal of the Study

What is the dependence of the error with

$$\| \mathbf{P}^f - \widetilde{\mathbf{P}}^f \| \qquad \| \mathbf{P}^a - \widetilde{\mathbf{P}}^a \| \qquad \| \mathbf{K} - \widetilde{\mathbf{K}} \|$$

where

$$\widetilde{\mathbf{P}}^f = \widehat{\mathbf{P}}^f \quad \text{or} \quad \widetilde{\mathbf{P}}^f = \widehat{\mathbf{P}}^f \circ \mathbf{C} \quad \text{or} \quad \dots$$

on ensemble size n, state dimension q and eigenvalues of λ_i of $\mathbf{P}^f$?

Associated questions:

- how big has the ensemble size n to be?
- what is an optimal taper matrix $\mathbf{C}$?
- what is the optimal boosting factor ρ?

EnsKF with $\widetilde{\mathbf{P}}^f = \widehat{\mathbf{P}}^f$

Straightforward, but "tedious" analysis is used to obtain

forecast covariance matrix:	exact expression
update covariance matrix:	lower bound
Kalman gain matrix:	approximation

EnsKF with tapering

What is the error when using $\widetilde{\mathbf{P}}^f = \widehat{\mathbf{P}}^f \circ \mathbf{C}$, for a positive definite taper matrix $\mathbf{C}$?

The expressions involve q, n , $\{\lambda_i\}$ and the eigenvectors of $\mathbf{P}^f$.

For $\widehat{\mathbf{P}}^a$ and $\widehat{\mathbf{K}}$, the expressions are long and are only approximate.

We want to minimize the error for the forecast covariance matrix with respect to all positive definite matrices $\mathbf{C}$.

For diagonal $\mathbf{P}^f$ we have simple solutions.

Optimal Taper C

A naive approach is to minimize component-wise

$$(\mathbf{C}_{\text{min}})_{ij} = \frac{p_{ij}^2}{p_{ij}^2 + (p_{ij}^2 + p_{ii}p_{jj})/n}$$

But $\mathbf{C}_{\text{min}}$ is

- not a correlation matrix: $(\mathbf{C}_{\text{min}})_{ii} = n/(n+2)$
- not always positive definite: depends on n

As $n \to \infty$, $(\mathbf{C}_{\text{min}})_{ij} = \begin{cases} 1 & \text{if } p_{ij} \neq 0 \\ 0 & \text{if } p_{ij} = 0 \end{cases}$

Optimal Taper with isotropic $\mathbf{P}^f$

Suppose we have an "isotropic" forecast matrix, i.e. the covariance depends only on the distance of the locations.

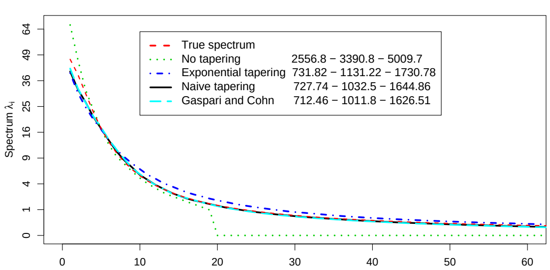
It is natural to minimize within "isotropic" taper matrices.

We get analytic solutions for particular cases.

Solutions in the mean squared sense are feasible.

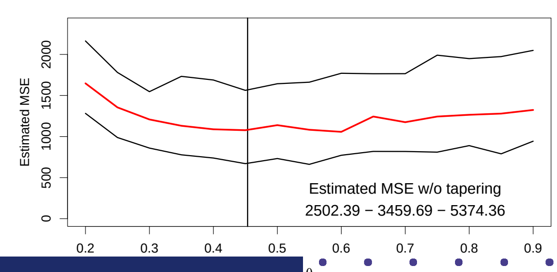
Illustration

Suppose a regular transect with $q = 250$ in $[0, 1]$ and exponentially decaying forecast covariance structure.
We consider ensemble size $n = 20$.
For 100 MC samples we calculated the spectrum.



Illustration

Suppose a regular transect with $q = 250$ in $[0, 1]$ and exponentially decaying forecast covariance structure.
We taper with Gaspari and Cohn's taper for ensemble size $n = 20$.
For 100 MC samples, calculate the MSE of the spectrum in function of taper length θ.



Discussion

- Tapering the forecast covariance matrix reduces significantly the error due to sampling variability.
- There exists an optimal taper matrix:
  - which is not always practical, i.e. is not always positive definite,
  - the system is not sensitive with respect to the taper choice.
- Statistical tool to determine optimal taper matrices, "boosting factors", . . .
- Discrepancy between theory and practice

Further Research

- Find optimal taper matrices for specific cases.
- Apply to "more"-step forecasts.
- Generalize to other matrices $\mathbf{H}_t$ and $\mathbf{R}_t$.
- Apply to high dimensional problems, i.e. numerical applications, DART.