Numerical weather prediction and statistics

Hans R. Künsch

Seminar for Statistics
ETH Zurich

DAGStat, Göttingen, March 18, 2016

Based on joint work with Fabio Sigrist, Marco Frei and Sylvain Robert

Why care about weather forecasting?

- Disaster prevention
- Agriculture, energy, transport: economic use of resources
- Interest in topics like big data, prediction vs. explanation, or uncertainty quantification
- Curiosity about surprising physical phenomena in our immediate surroundings

Remarkable progress has been made in weather forecasts, in contrast to other areas like earthquake or economic forecasts.

Overview

Physics, numerical analysis and computing are the main players in weather forecasting, not statistics. Deterministic methods dominate. In this talk I would like to

- Give some reasons for this situation

Overview

Physics, numerical analysis and computing are the main players in weather forecasting, not statistics. Deterministic methods dominate.
In this talk I would like to

- Give some reasons for this situation
- Explain 3 instances where stochastic methods and concepts are important to improve current weather forecast methods
  - Statistical postprocessing: Quantify uncertainty and correct biases of deterministic forecasts
  - Data assimilation: Combine observations and forecasts to obtain a better starting point for the next forecast
  - Stochastic parametrization: Construct stochastic dynamics in order to remove deficiencies of deterministic models

Overview

Physics, numerical analysis and computing are the main players in weather forecasting, not statistics. Deterministic methods dominate. In this talk I would like to

- Give some reasons for this situation
- Explain 3 instances where stochastic methods and concepts are important to improve current weather forecast methods
  - Statistical postprocessing: Quantify uncertainty and correct biases of deterministic forecasts
  - Data assimilation: Combine observations and forecasts to obtain a better starting point for the next forecast
  - Stochastic parametrization: Construct stochastic dynamics in order to remove deficiencies of deterministic models
- Introduce you to an area that is unfamiliar to many statisticians

A (very) short history of weather forecasting

- 19th century: Forecasts are based on same day weather maps (mainly of pressure), made possible by the invention of the telegraph. Empirical, essentially based on statistical methods

A (very) short history of weather forecasting

- 19th century: Forecasts are based on same day weather maps (mainly of pressure), made possible by the invention of the telegraph. Empirical, essentially based on statistical methods
- 1903: Vilhelm Bjerknes argues to base weather prediction purely on physical laws of fluid motion and thermodynamics
- 1910-1920: Lewis Fry Richardson develops numerical methods to solve the equations of physics. It took 6 weeks to compute a 6 hour forecast, and the results were disastrous
- 1947-1955: John von Neumann and coworkers show that numerical weather prediction is possible using the newly invented computers

A (very) short history of weather forecasting

- 19th century: Forecasts are based on same day weather maps (mainly of pressure), made possible by the invention of the telegraph. Empirical, essentially based on statistical methods
- 1903: Vilhelm Bjerknes argues to base weather prediction purely on physical laws of fluid motion and thermodynamics
- 1910-1920: Lewis Fry Richardson develops numerical methods to solve the equations of physics. It took 6 weeks to compute a 6 hour forecast, and the results were disastrous
- 1947-1955: John von Neumann and coworkers show that numerical weather prediction is possible using the newly invented computers
- 1963: Ed Lorenz discovers the chaotic nature of atmospheric circulation
- 21th century: Ensemble methods become used operationally

Outline

1. Introduction

2. **Statistical Postprocessing**

3. Data Assimilation (aka Filtering)

4. Stochastic Parametrizations

Subjectivity in weather forecasting

Weather forecasting based on weather maps was more an art than science. It relied strongly on the experience and intuition of the forecaster and there was little formalization.

In the period 1940 - 1955, there was a massive effort to make forecasting more objective, using statistical methods like regression. It ended in disappointment “[No one] would claim that the [statistical] method merits further serious study”.

Subjectivity in weather forecasting

Weather forecasting based on weather maps was more an art than science. It relied strongly on the experience and intuition of the forecaster and there was little formalization.

In the period 1940 - 1955, there was a massive effort to make forecasting more objective, using statistical methods like regression. It ended in disappointment “[No one] would claim that the [statistical] method merits further serious study”.

Still, humans are able to improve the accuracy of precipitation forecasts by a constant factor of about 25%, irrespective of the progress of numerical forecasts. Statistical postprocessing attempts to formalize these improvements.

Goals of statistical postprocessing

- Correct biases of numerical forecasts
- Quantify uncertainty of numerical forecasts
- Account correctly for spatial and temporal dependence
- Obtain predictive distributions which are calibrated and sharp

Many developments of statistical postprocessing are due to Tilmann Gneiting and coauthors. Here I show the example of Sigrist, K. and Stahel, JRSS B (2015).

A statistical model for postprocessing precipitation

Deterministic forecasts on a dense spatial grid over northern Switzerland are issued each day at midnight for lead times 3h, 6h, ..., 24h. We build a regression type model with these forecasts as predictors.

A statistical model for postprocessing precipitation

Deterministic forecasts on a dense spatial grid over northern Switzerland are issued each day at midnight for lead times 3h, 6h, ..., 24h. We build a regression type model with these forecasts as predictors.

Let $y(t, s)$ and $y_F(t, s)$ be the true and forecasted precipitation, respectively, at time t and position s. In order to deal with skewness and the atom at zero of the precipitation, we use a “potential precipitation” $w(t, s)$ which is Gaussian. Negative w means zero precipitation, and skewness is obtained via Box-Cox transformation.

A statistical model for postprocessing precipitation

Deterministic forecasts on a dense spatial grid over northern Switzerland are issued each day at midnight for lead times 3h, 6h, ..., 24h. We build a regression type model with these forecasts as predictors.

Let $y(t, s)$ and $y_F(t, s)$ be the true and forecasted precipitation, respectively, at time t and position s. In order to deal with skewness and the atom at zero of the precipitation, we use a "potential precipitation" $w(t, s)$ which is Gaussian. Negative w means zero precipitation, and skewness is obtained via Box-Cox transformation.

This means we use the model

$$\begin{aligned} y(t,s) &= w(t,s)^{\lambda}\, \mathbf{1}_{\{w(t,s)>0\}} \\ w(t,s) &= \beta_1 y_F(t,s)^{1/\tilde{\lambda}} + \beta_2 \mathbf{1}_{\{y_F(t,s)=0\}} + \varepsilon(t,s) \end{aligned}$$

with a Gaussian error field ε, correlated in space and time.

Space-time correlations

Use past forecasts and past station data to estimate parameters and the error field ε at time 00:00 of day t. Then the statistical prediction for day t follows from the model and the deterministic predictions.

Space-time correlations

Use past forecasts and past station data to estimate parameters and the error field ε at time 00:00 of day t. Then the statistical prediction for day t follows from the model and the deterministic predictions.

For this the choice of the covariance function of ε is crucial. Instead of choosing a member from the “zoo” of available space-time covariances, we define it implicitly by a stochastic advection-diffusion equation. Advantages are a parsimonious parametrization with interpretable parameters and a Markovian time evolution.

Space-time correlations

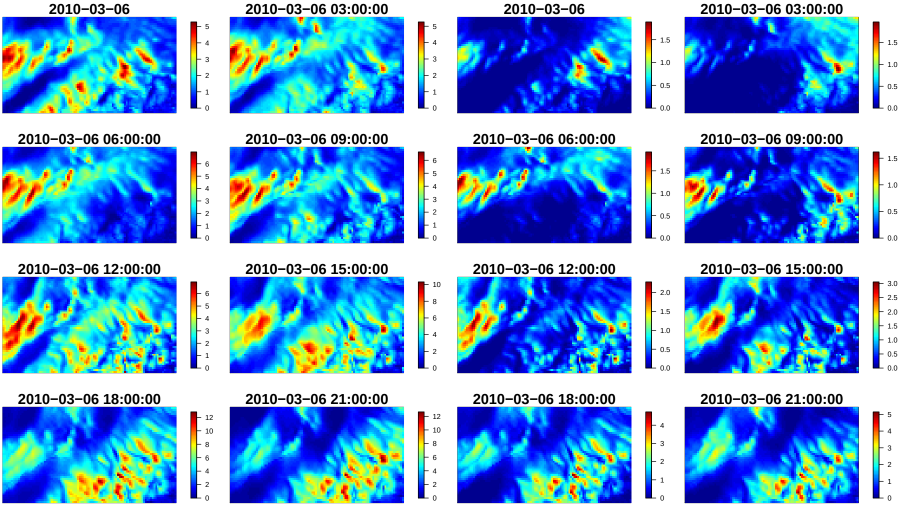
Use past forecasts and past station data to estimate parameters and the error field ε at time 00:00 of day t. Then the statistical prediction for day t follows from the model and the deterministic predictions.

For this the choice of the covariance function of ε is crucial. Instead of choosing a member from the “zoo” of available space-time covariances, we define it implicitly by a stochastic advection-diffusion equation. Advantages are a parsimonious parametrization with interpretable parameters and a Markovian time evolution.

After some approximations, we end up with a vector autoregression for the Fourier coefficients of $\varepsilon(t, .)$. Matrices have a 2×2 block structure which is essential for large grids. State space methods and Gibbs iterations are used to fit the model under the Bayesian paradigm.

Example of postprocessed forecasts

Left: Deterministic forecast. Right: Median of predictive distribution.



Outline

1. Introduction

2. Statistical Postprocessing

3. **Data Assimilation (aka Filtering)**

4. Stochastic Parametrizations

What is data assimilation?

Laws of physics are formulated as differential equations whose solutions are determined completely by initial conditions. A perfect forecast would be possible for all lead times if

- initial condition is known precisely at some time point
- solutions can be computed accurately
- laws of physics are sufficient

What is data assimilation?

Laws of physics are formulated as differential equations whose solutions are determined completely by initial conditions. A perfect forecast would be possible for all lead times if

- initial condition is known precisely at some time point
- solutions can be computed accurately
- laws of physics are sufficient

However, initial conditions are never known precisely because of measurement errors and a limited number of measurements, in particular for the upper parts of the atmosphere. Because solutions are sensitive to initial conditions, there is a limit to predictability.

What is data assimilation?

Laws of physics are formulated as differential equations whose solutions are determined completely by initial conditions. A perfect forecast would be possible for all lead times if

- initial condition is known precisely at some time point
- solutions can be computed accurately
- laws of physics are sufficient

However, initial conditions are never known precisely because of measurement errors and a limited number of measurements, in particular for the upper parts of the atmosphere. Because solutions are sensitive to initial conditions, there is a limit to predictability.

Therefore, we need to update initial conditions frequently. Data assimilation does this by combining a forecast for time t with the observations at time t. Updated initial conditions at time t become the basis for the forecast for time $t + \Delta t$. Would like to take uncertainty of initial conditions into account.

State space models

The time series of the state of the atmosphere and the sequence of observations about it form a state space model. State space models are useful in other applications of time series, e.g. economics, finance, ecology, systems biology. In state space models, data assimilation is usually called filtering.

State space models

The time series of the state of the atmosphere and the sequence of observations about it form a state space model. State space models are useful in other applications of time series, e.g. economics, finance, ecology, systems biology. In state space models, data assimilation is usually called filtering.

A state space model consists of a pair $((X_t), (Y_i))$ where X_t is the state of a system at time t and Y_i are partial and noisy observations of X_{t_i} at some discrete time points t_i.

X_t contains a complete description of the system, so it is not fully observable. Its dynamics is a Markov process in discrete or continuous time (Differential equations are deterministic Markov processes).

Observations Y_i are conditionally independent given the state, and Y_i depends only on X_{t_i}.

Graphical representation of state space models

For simplicity, assume observations are at integer times and write t instead of t_i.

The dependence between the variables and implied conditional independence relations of a state space model can be represented by the following directed acyclic graph

$$\begin{tikzcd}
\cdots \arrow[r] & X_{t-1} \arrow[r] \arrow[d] & X_t \arrow[r] \arrow[d] & X_{t+1} \arrow[r] \arrow[d] & \ldots \\
\cdots & Y_{t-1} & Y_t & Y_{t+1} & \ldots
\end{tikzcd}$$

In particular, (Y_t) is not a Markov chain. Weather maps of the current day are not sufficient.

Basics of data assimilation/filtering

Uncertainty about the state represented by distributions:
Prediction π_t^p = conditional distribution of X_t given $Y_{1:t-1} = y_{1:t-1}$.
Filter π_t^f = conditional distribution of X_t given $Y_{1:t} = y_{1:t}$.
Here $y_{1:t}$ is shorthand for $(y_1, y_2, \ldots, y_t)$.

Basics of data assimilation/filtering

Uncertainty about the state represented by distributions:
Prediction π_t^p = conditional distribution of X_t given $Y_{1:t-1} = y_{1:t-1}$.
Filter π_t^f = conditional distribution of X_t given $Y_{1:t} = y_{1:t}$.
Here $y_{1:t}$ is shorthand for $(y_1, y_2, \ldots, y_t)$.

Recursive scheme:
Propagation $\pi_{t-1}^f \longrightarrow \pi_t^p$ based on the dynamics of the state

$$\pi_t^p(dx) = \int \pi_{t-1}^f(dx')\mathbb{P}(X_t \in dx \mid X_{t-1} = x')$$

Update $\pi_t^p \longrightarrow \pi_t^f$ based on the likelihood of the new observation

$$\pi_t^f(dx) \propto \pi_t^p(dx)p(y_t \mid X_t = x)$$

Hence prediction π_t^p is used as the prior for X_t.

Monte Carlo filters

Recursions from the previous page typically cannot be computed analytically or numerically, except in the linear Gaussian case. Moreover, transition distribution $\mathbb{P}(X_t \in dx \mid X_{t-1} = x')$ is often intractable, but for given starting point x' one can draw from it. In the case of differential equations, this means computing the solution for starting point x'.

Monte Carlo filters

Recursions from the previous page typically cannot be computed analytically or numerically, except in the linear Gaussian case. Moreover, transition distribution $\mathbb{P}(X_t \in dx \mid X_{t-1} = x')$ is often intractable, but for given starting point x' one can draw from it. In the case of differential equations, this means computing the solution for starting point x'.

Monte Carlo filters approximate π_t^p and π_t^f by samples or ensembles of "particles" $x_t^{p,j}$ and $x_t^{f,j}$ ($j = 1, 2, \ldots, N$).

Monte Carlo filters

Recursions from the previous page typically cannot be computed analytically or numerically, except in the linear Gaussian case. Moreover, transition distribution $\mathbb{P}(X_t \in dx \mid X_{t-1} = x')$ is often intractable, but for given starting point x' one can draw from it. In the case of differential equations, this means computing the solution for starting point x'.

Monte Carlo filters approximate π_t^p and π_t^f by samples or ensembles of “particles” $x_t^{p,j}$ and $x_t^{f,j}$ ($j = 1, 2, \ldots, N$).

In the propagation step, filter particles move forward according to the dynamics of the state to become the next prediction particles.

Updating converts prediction particles into filter particles, thereby correcting the prediction error. Several methods for this have been proposed.

The update step for ensembles of particles

- Two main methods are the Particle Filter (PF) and the Ensemble Kalman Filter (EnKF)
- PF originated in statistics (Gordon et al., 1993), EnKF in geophysics (Evensen, 1994)

The update step for ensembles of particles

- Two main methods are the Particle Filter (PF) and the Ensemble Kalman Filter (EnKF)
- PF originated in statistics (Gordon et al., 1993), EnKF in geophysics (Evensen, 1994)
- PF works for arbitrary likelihoods $p(y_t \mid X_t = x)$ and uses weighting and resampling
- PF is consistent under very weak assumptions, but degenerates easily, in particular in high dimensions

The update step for ensembles of particles

- Two main methods are the Particle Filter (PF) and the Ensemble Kalman Filter (EnKF)
- PF originated in statistics (Gordon et al., 1993), EnKF in geophysics (Evensen, 1994)
- PF works for arbitrary likelihoods $p(y_t \mid X_t = x)$ and uses weighting and resampling
- PF is consistent under very weak assumptions, but degenerates easily, in particular in high dimensions
- EnKF needs a Gaussian likelihood $p(y_t \mid X_t = x) = \varphi(y_t; Hx, R)$ and moves the particles towards the observations
- EnKF needs a Gaussian prediction distribution π_t^p for consistency, but is extremely robust in practice.

Particle filter update

Drop from now on the time index t.

Basic particle filter resamples the j-th prediction particle $x^{p,j}$ N_j times where

$$\mathbb{E}\left[N_j\right] \propto N\, p(y \mid x^{p,j}), \quad \sum_j N_j = N$$

Particles with a poor fit to the observation die, those with an excellent fit have children. Balanced resampling minimizes Monte Carlo error: N_j differs by less than 1 from its expectation.

Particle filter update

Drop from now on the time index t.

Basic particle filter resamples the j-th prediction particle $x^{p,j}$ N_j times where

$$\mathbb{E}\left[N_j\right] \propto N\, p(y \mid x^{p,j}), \quad \sum_j N_j = N$$

Particles with a poor fit to the observation die, those with an excellent fit have children. Balanced resampling minimizes Monte Carlo error: N_j differs by less than 1 from its expectation.

In high dimensions, filter ensemble is often grossly overconfident, because too few prediction particles survive. Analysed theoretically by Bickel et al. (2008).

Ensemble Kalman filter update

Based on the following standard result:

If $X \sim \mathcal{N}(\mu^p, P^p)$ and $Y \mid X = x \sim \mathcal{N}(Hx, R)$ then $X \mid Y = y \sim N(\mu^f, P^f)$ where

$$\mu^f = \mu^p + K(y - H\mu^p), \quad P^f = P^p - KHP^p$$

where K is the Kalman gain

$$K = Cov(X, Y)Cov(Y, Y)^{-1} = P^p H^T (HP^p H^T + R)^{-1}$$

Ensemble Kalman filter update

Based on the following standard result:

If $X \sim \mathcal{N}(\mu^p, P^p)$ and $Y \mid X = x \sim \mathcal{N}(Hx, R)$ then $X \mid Y = y \sim N(\mu^f, P^f)$ where

$$\mu^f = \mu^p + K(y - H\mu^p), \quad P^f = P^p - KHP^p$$

where K is the Kalman gain

$$K = Cov(X, Y)Cov(Y, Y)^{-1} = P^p H^T (HP^p H^T + R)^{-1}$$

EnKF estimates μ^p and P^p from $(x^{p,j})$ and K, μ^f and P^f by plug-in. Finally $(x^{p,j})$ is converted into a sample with mean $\hat{\mu}^f$ and covariance $\hat{P}^f$. There are different algorithms to achieve this.

Two versions of the Ensemble Kalman filter

Stochastic: Let ε^j be i.i.d. $\sim N(0, R)$ and set

$$x^{f,j} = x^{p,j} + \hat{K}(y + \varepsilon^j - Hx^{p,j})$$

Mean update of each particle with a perturbed observation. Can be considered as a balanced sample from

$$\frac{1}{N}\sum_{j=1}^{N}\mathcal{N}(x^{p,j} + \hat{K}(y - Hx^{p,j}), \hat{K}R\hat{K}^T)$$

Two versions of the Ensemble Kalman filter

Stochastic: Let ε^j be i.i.d. $\sim N(0, R)$ and set

$$x^{f,j} = x^{p,j} + \hat{K}(y + \varepsilon^j - Hx^{p,j})$$

Mean update of each particle with a perturbed observation.
Can be considered as a balanced sample from

$$\frac{1}{N} \sum_{j=1}^{N} \mathcal{N}(x^{p,j} + \hat{K}(y - Hx^{p,j}), \hat{K} R \hat{K}^T)$$

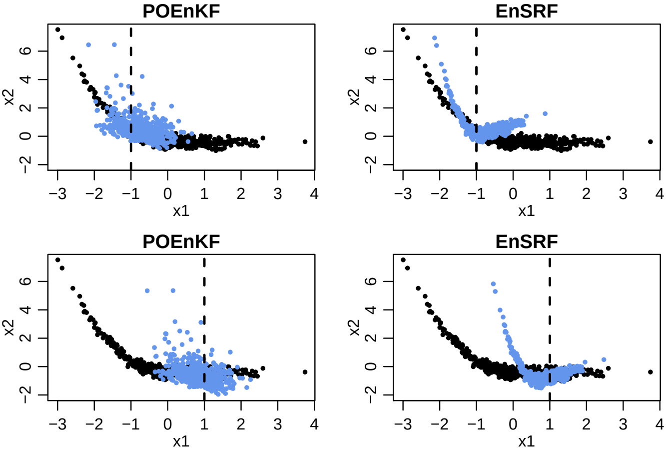
Deterministic: First update the mean

$$\bar{x}^f = \bar{x}^p + \hat{K}(y - H\bar{x}^p)$$

and transform the residuals $x^{p,j} - \bar{x}^p$ linearly so that the empirical covariance is $\hat{P}^f$. Involves matrix square roots.

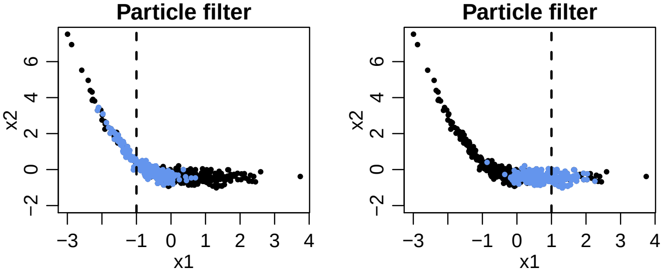
Banana-shaped example

Black: prediction ensemble. Observation $Y \sim \mathcal{N}(x_1, 0.5^2)$.
Blue: EnKF updates for two values $y = \pm 1$. Left: stochastic (perturbed observations), Right: deterministic (square-root).



Banana-shaped example II

Particle filter update for the same situation.



Bridging the Particle and the Ensemble Kalman filter

EnKF works surprisingly well in weather forecasting where the dimension of the state and of the observations are of the order $10^6 - 10^8$ and the ensemble size is usually less than 100, provided updates are localized.

Can we explain this success by some theory? Can we improve the ability of the EnKF to deal with non-Gaussian features in π^p without loosing the stability?

Bridging the Particle and the Ensemble Kalman filter

EnKF works surprisingly well in weather forecasting where the dimension of the state and of the observations are of the order $10^6 - 10^8$ and the ensemble size is usually less than 100, provided updates are localized.

Can we explain this success by some theory? Can we improve the ability of the EnKF to deal with non-Gaussian features in π^p without loosing the stability?

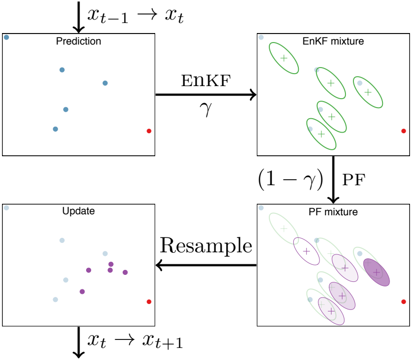
EnKPF (Frei and K., Biometrika, 2013) uses progressive update

$$\pi^p(dx) \xrightarrow{\text{EnKF}} \pi^{f,\gamma} \propto \pi^p(dx)p(y|x)^\gamma \xrightarrow{\text{PF}} \pi^f(dx) \propto \pi^{f,\gamma}(dx)p(y|x)^{1-\gamma}.$$

Interpolates continuously between PF ($\gamma = 0$) and EnKF ($\gamma = 1$).

Illustration of EnKPF

Using the Gaussian mixture interpretation of the EnKF, the second step can be done exactly. It gives a another Gaussian mixture with unequal weights. So sampling can be postponed until the end:



Does the EnKPF beat the EnKF for weather forecasts?

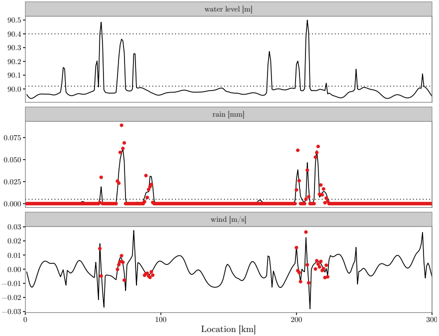
We are trying to find out in a collaboration with MeteoSwiss and German Weather Service. Localization of the EnKPF is delicate. Moreover, the complexity of an actual weather forecast system limits the methods that we can implement.

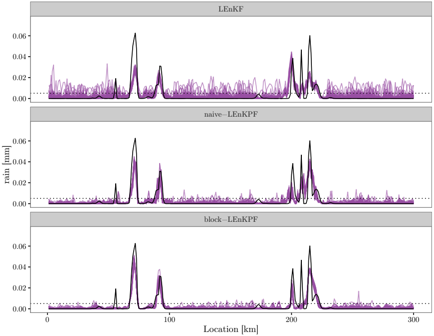
Does the EnKPF beat the EnKF for weather forecasts?

We are trying to find out in a collaboration with MeteoSwiss and German Weather Service. Localization of the EnKPF is delicate. Moreover, the complexity of an actual weather forecast system limits the methods that we can implement.

We obtain some improvements over EnKF in toy examples. The one closest to an actual weather model is a modified shallow water equation. It is a one-dimensional system with 3 variables, wind, rain and height of wet air. Wet air raises, and if a certain height is reached, rain starts. Height is unobserved, and wind is only observed where there is rain.

A modified shallow water equation



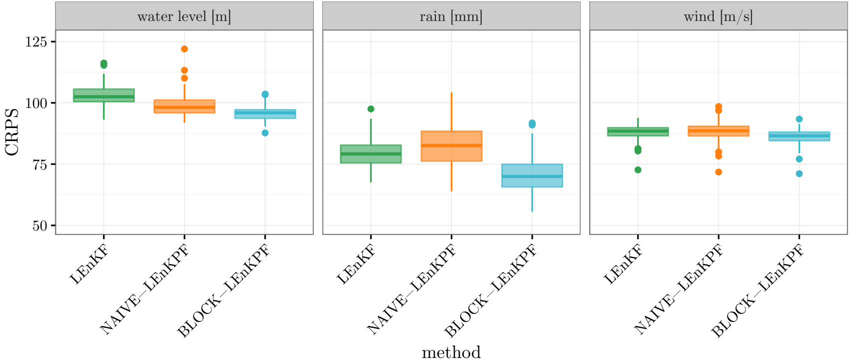


A solution of the equation at some fixed time.

Rain component of ensemble members at the same time. Top: localized EnKF, Middle and Bottom: 2 versions of localized EnKPF.

Calibration and sharpness of filters

Continuously ranked probability score for 3 ensemble filters in a simulation with many iterations.



Outline

1. Introduction
2. Statistical Postprocessing
3. Data Assimilation (aka Filtering)
4. **Stochastic Parametrizations**

What is parametrization?

Another condition for a perfect forecast was that solutions can be computed accurately. But processes on small scales are not resolved by numerical methods, and these have feedback effects on larger scales that are resolved.

Examples of unresolved processes in numerical weather modeling are turbulent processes in the atmospheric boundary layer that links the free atmosphere to the surface, or different types of clouds.

What is parametrization?

Another condition for a perfect forecast was that solutions can be computed accurately. But processes on small scales are not resolved by numerical methods, and these have feedback effects on larger scales that are resolved.

Examples of unresolved processes in numerical weather modeling are turbulent processes in the atmospheric boundary layer that links the free atmosphere to the surface, or different types of clouds.

A parametrization approximates the effect of unresolved variables by a function of resolved variables. This function is determined empirically and has usually unknown parameters that are estimated from data.

What is parametrization?

Another condition for a perfect forecast was that solutions can be computed accurately. But processes on small scales are not resolved by numerical methods, and these have feedback effects on larger scales that are resolved.

Examples of unresolved processes in numerical weather modeling are turbulent processes in the atmospheric boundary layer that links the free atmosphere to the surface, or different types of clouds.

A parametrization approximates the effect of unresolved variables by a function of resolved variables. This function is determined empirically and has usually unknown parameters that are estimated from data.

Stochastic parametrizations allow in addition for a random error.
“I believe that the ultimate climatic models ... will be stochastic, i.e., random numbers will appear somewhere in the time derivatives.”
(Ed Lorenz)

A toy example (Arnold et al., Phil. Trans. Royal Soc. A, 2013)

Resolved process $x_t = (x_{t,1}, \ldots, x_{t,K})$ (e.g. averages in K large boxes). Unresolved processes in box k: $y_{t,k} = (y_{t,k,1}, \ldots, y_{t,k,J})$. Both processes are coupled:

$$\frac{d}{dt} x_{t,k} = F_k(x_t) - const \frac{1}{J} \sum_{j=1}^{J} y_{t,k,j}$$

and the derivative of $y_{t,k,j}$ depends not only on other Y-variables, but also on $x_{t,k}$.

A toy example (Arnold et al., Phil. Trans. Royal Soc. A, 2013)

Resolved process $x_t = (x_{t,1}, \ldots, x_{t,K})$ (e.g. averages in K large boxes). Unresolved processes in box k: $y_{t,k} = (y_{t,k,1}, \ldots, y_{t,k,J})$. Both processes are coupled:

$$\frac{d}{dt} x_{t,k} = F_k(x_t) - const \frac{1}{J} \sum_{j=1}^{J} y_{t,k,j}$$

and the derivative of $y_{t,k,j}$ depends not only on other Y-variables, but also on $x_{t,k}$.

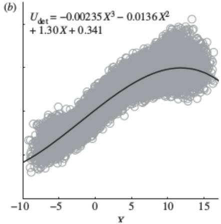
Parametrization approximates the differential equation by

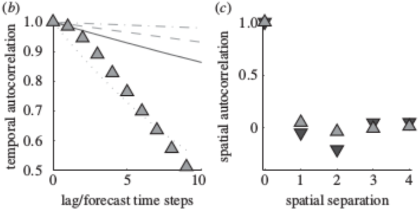
$$\frac{d}{dt} x_{t,k} = F_k(x_t) - U(x_{t,k})$$

for some function U, $U(x_{t,k}) \approx F_k(x_t) - \frac{d}{dt} x_{t,k}$.

Estimation of the parametrizing function *U*

Estimate x_t and their derivatives from data (or simulations in this example) and use regression:





Obviously, the relation is not deterministic. However, errors are strongly correlated in time and show some heteroscedasticity. If these errors are modeled as autoregressive and added to the state variables, one has a stochastic state space model and can do data assimilation.

Summary and Conclusion

- Weather forecasting presents challenges and opportunities for statisticians who are prepared to cooperate with atmospheric physicists and applied mathematicians
- There is more room for stochastic models and methods to complement the deterministic ones
- State space models and Monte Carlo methods for filtering should belong to the toolbox of applied statisticians

Thank you for your attention!