Ensemble Methods in Atmospheric Data Assimilation

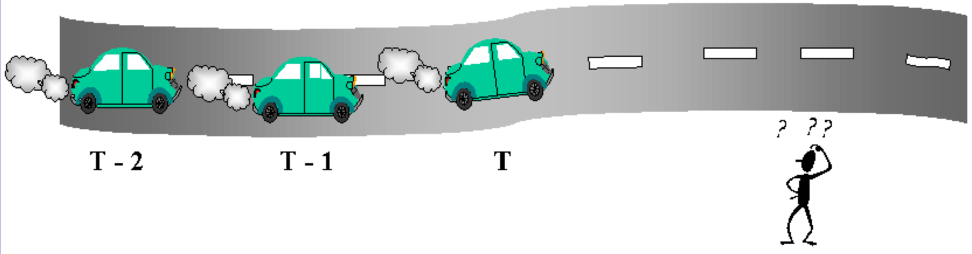
Harald Anlauf

Research and Development, Data Assimilation Section
Deutscher Wetterdienst, Offenbach, Germany

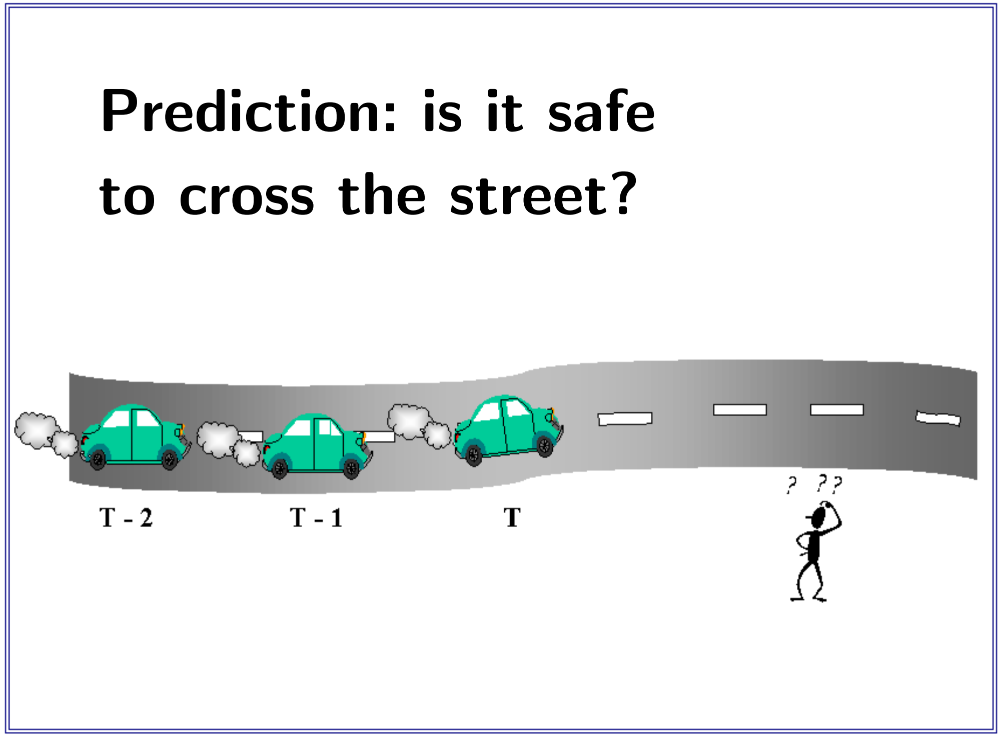
DESY Monte Carlo Workshop, DESY, Hamburg
19.-21. February 2013

Data assimilation in everyday life

Prediction: is it safe to cross the street?

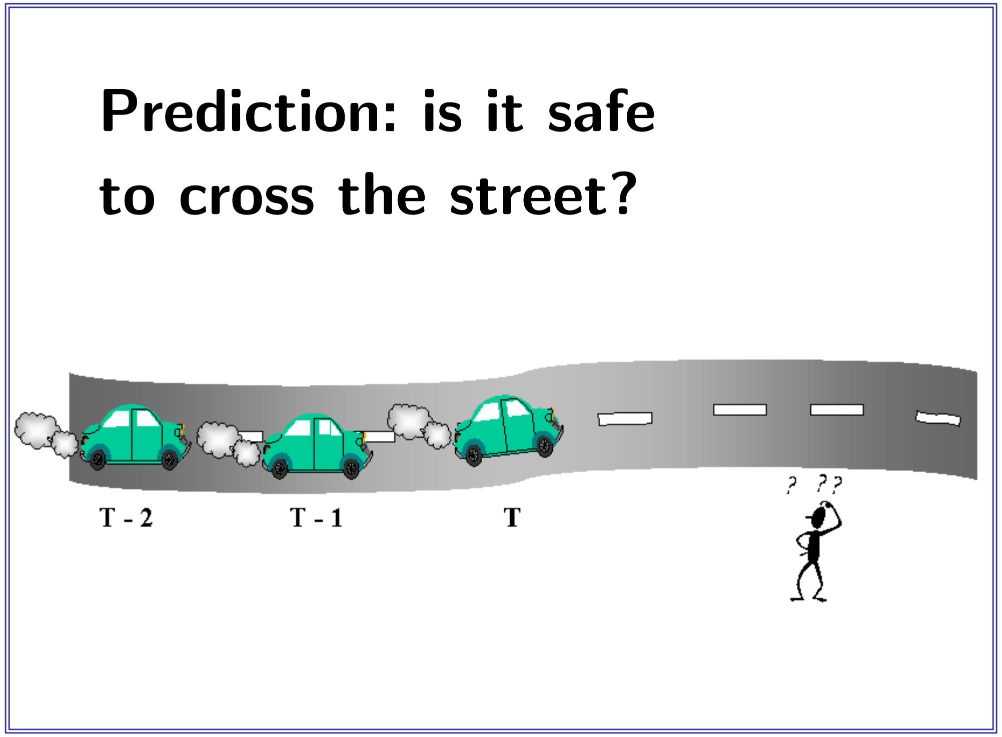


Data assimilation in everyday life



- **Used information:**
  - Observations
  - Knowledge about cars, roads
  - Experience ($\Longrightarrow$ statistics)

Data assimilation in everyday life



- **Used information:**
  - Observations
  - Knowledge about cars, roads
  - Experience ($\Longrightarrow$ statistics)
- **Forecast error** due to:
  - Observation error
  - Model error (icy road)
  - Current conditions not represented by statistics

Operational NWP Models at DWD

GME

Global model, hydrostatic

Triangular grid, mesh size: 20 km
60 levels (top: 5 hPa, ≈ 36 km)
(1474562×60 grid points)

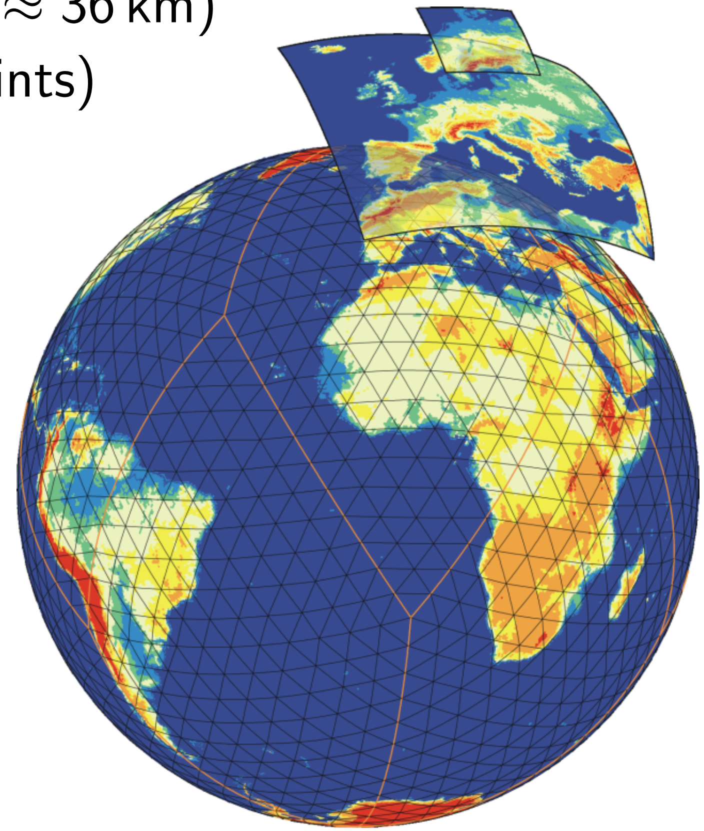
Forecast times:
174h from 00Z, 12Z;
48h from 06Z, 18Z

COSMO-EU

Non-hydrostatic

Mesh size: 7 km
40 levels to ≈ 23 km

Forecast times:
78h from 00Z, 12Z;
48h from 06Z, 18Z



COSMO-DE

Non-hydrostatic,
"convection permitting"

Mesh size: 2.8 km
50 levels to ≈ 22 km

Forecast times:
21h from 00Z, 03Z, . . . , 21Z

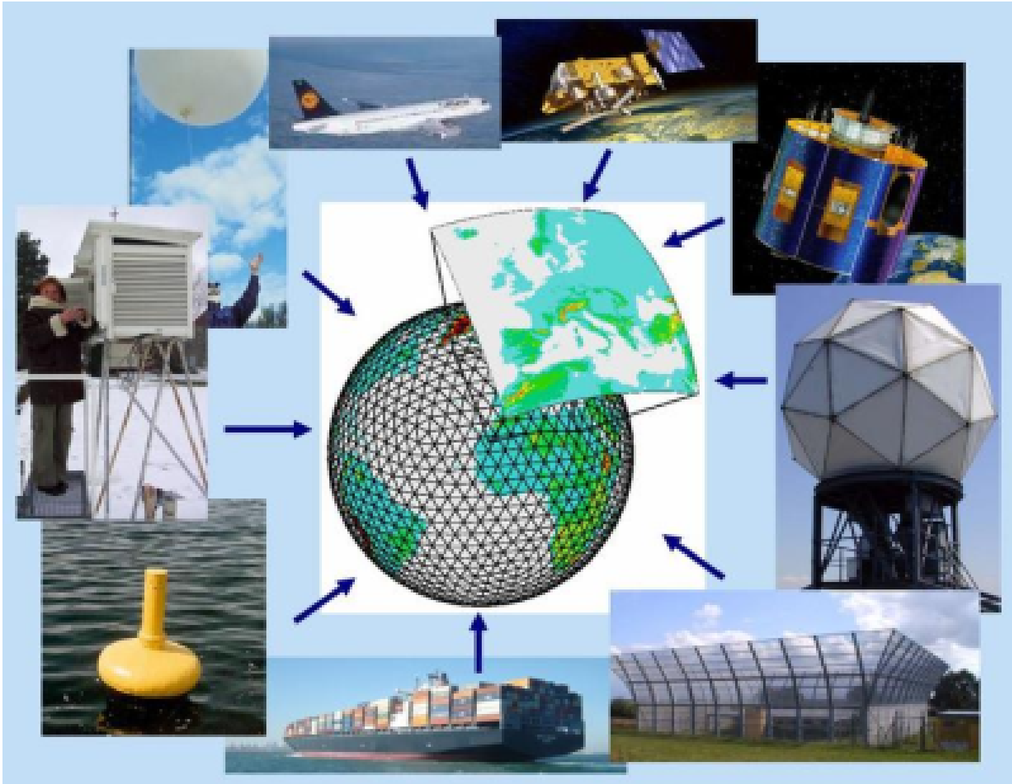
COSMO-DE-EPS

Ensemble prediction system
20 ensemble members
(in 2013: 40 members)

Forecast times:
21h from 00Z, 03Z, . . . , 21Z

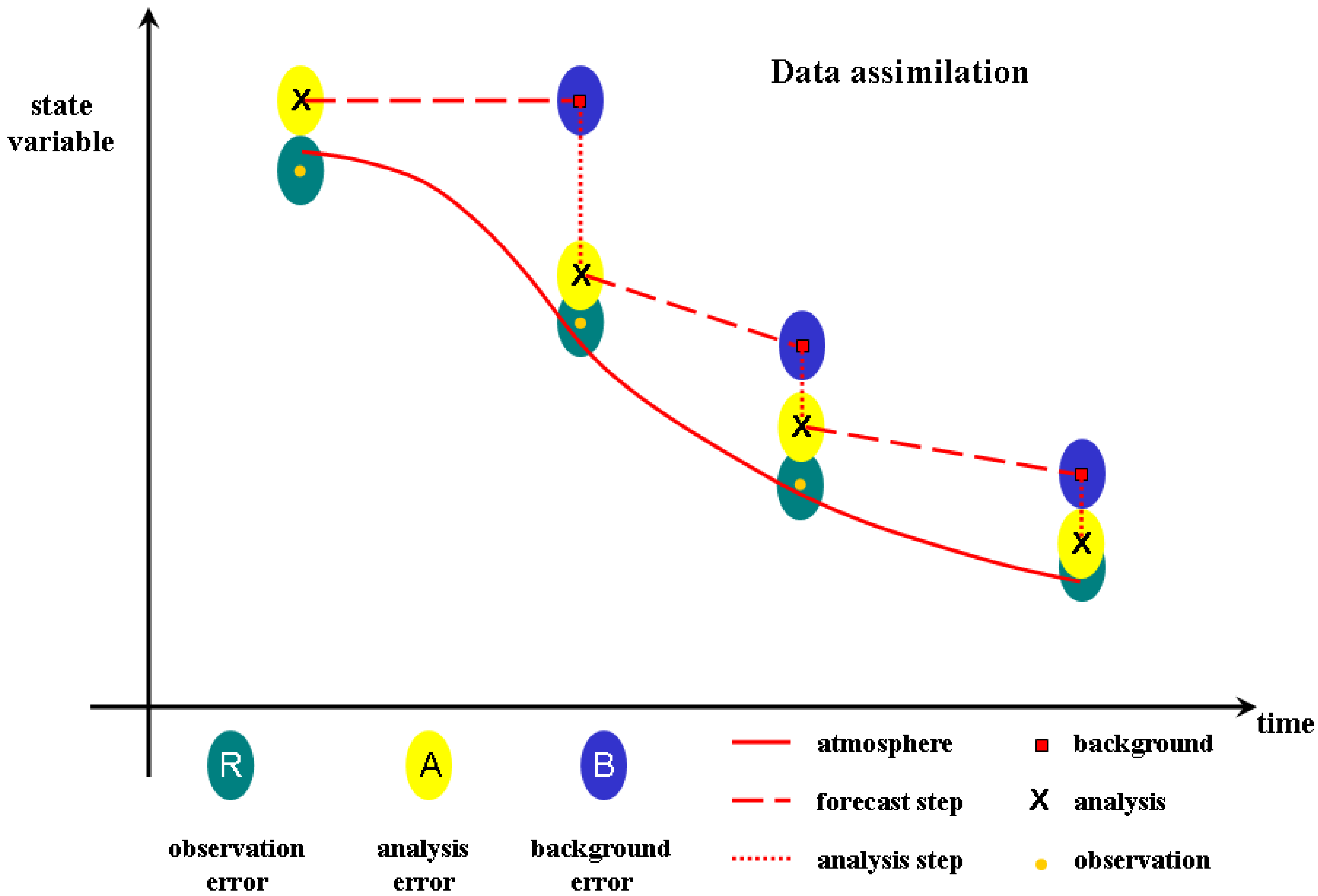
The Global Observing System

- In-Situ Observations
  - Surface observations (land, ship, buoys)
  - Radiosondes
  - Aircraft reports
  - ...



- Remote Sensing Observations
  - Satellite based
    - geostationary
    - polar orbiting
    - LEOs (low-earth orbit), partly research satellites
  - RADAR
  - LIDAR
  - GNSS (GPS) ground observations
  - ...
- Network very inhomogeneous, some areas not well observed!

Intermittent (Cycled) Data Assimilation



Outline

Data Assimilation for Dynamical Systems

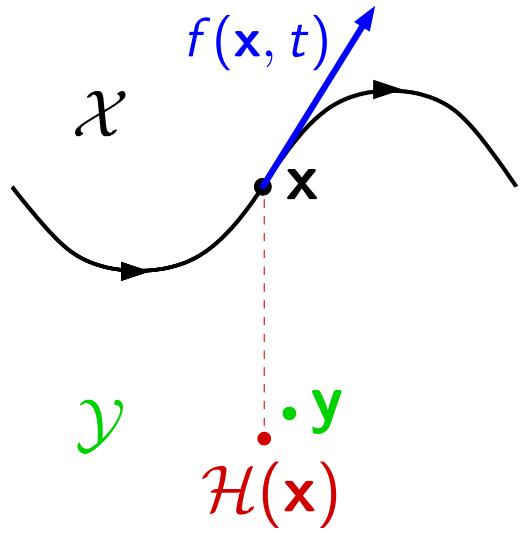
- Dynamical system:
  State space $\mathcal{X}$, model operator (flow):

  $$\mathcal{M} : \mathbb{R} \times \mathcal{X} \to \mathcal{X}$$

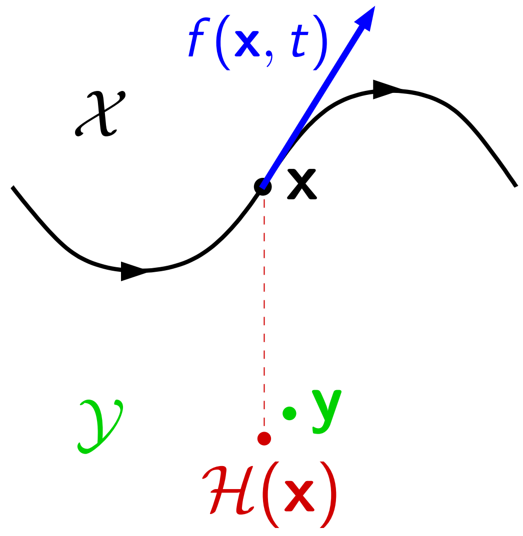
  so that for all $x \in \mathcal{X}$, $s, t \in \mathbb{R}$:

  $$\mathcal{M}(0, x) = x\,, \quad \mathcal{M}(s, \mathcal{M}(t, x)) = \mathcal{M}(t+s, x)$$

- Observation space: $\mathcal{Y}$
- Observation operator $\mathcal{H} : \mathcal{X} \ni x \to y \in \mathcal{Y}$
- Given observations $y_1, \ldots, y_n$, determine $x(t)$!



Data Assimilation for Dynamical Systems



- Dynamical system:
  State space $\mathcal{X}$, model operator (flow):

  $$\mathcal{M} : \mathbb{R} \times \mathcal{X} \to \mathcal{X}$$

  so that for all $x \in \mathcal{X}$, $s, t \in \mathbb{R}$:

  $$\mathcal{M}(0, x) = x\,, \quad \mathcal{M}(s, \mathcal{M}(t, x)) = \mathcal{M}(t+s, x)$$

- Observation space: $\mathcal{Y}$
- Observation operator $\mathcal{H} : \mathcal{X} \ni x \to y \in \mathcal{Y}$
- Given observations $y_1, \ldots, y_n$, determine $x(t)$!
  - Under-determined (ill-posed) **inverse problem**: number of observations much smaller than degrees of freedom (related problem: tomography)
  - Observations and models have errors
  - Model variables may not be directly observed, observation operator (in general) not injective, not invertible
  - **Data assimilation:** combine observations with prior information

Linear Systems: Kalman Filter (I)

- Linear model, linear observation operator: $\mathcal{H}(x) = \mathbf{H}x$
  - Forecast error covariance (x^f: forecast, x_t: true state)

$$\mathbf{P}^f = \mathrm{E}\left\{(x^f - x_t)(x^f - x_t)^\mathsf{T}\right\}$$

  - Observation error covariance (y_t: true observation)

$$\mathbf{R} = \mathrm{E}\left\{(y - y_t)(y - y_t)^\mathsf{T}\right\}$$

- Analysis: weighted linear combination of forecast and observations

$$x^a = x^f + \mathbf{K}(y - \mathbf{H}x^f)$$

- Optimal "Kalman gain"

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

minimizes analysis error:

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f = \left((\mathbf{P}^f)^{-1} + \mathbf{H}^\mathsf{T}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}$$

Kalman Filter (II)

- Forecast step ($\mathbf{M}_i = \mathbf{M}(t_i \to t_{i+1})$):

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

  $\mathbf{Q}$: model error (system noise) covariance
- Kalman filter gives optimal (minimal variance) solution for *linear* models, *linear* observations, and *Gaussian* observation errors

Kalman Filter (II)

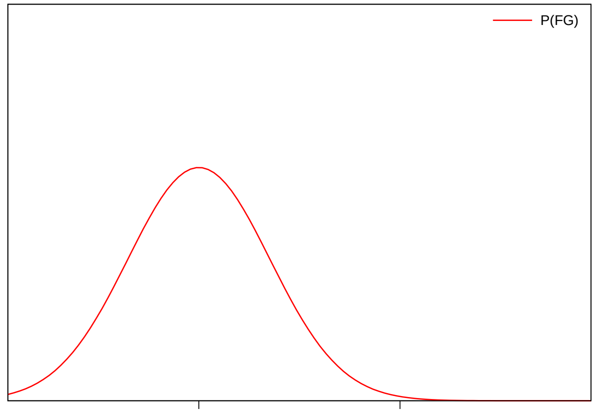
- Forecast step ($\mathbf{M}_i = \mathbf{M}(t_i \to t_{i+1})$):

$$
\begin{aligned}
x^f(t_{i+1}) &= \mathbf{M}_i x^a(t_i) \\
\mathbf{P}^f(t_{i+1}) &= \mathbf{M}_i \mathbf{P}^a(t_i)\mathbf{M}_i^\mathsf{T} + \mathbf{Q}(t_i)
\end{aligned}
$$

  $\mathbf{Q}$: model error (system noise) covariance
- Kalman filter gives optimal (minimal variance) solution for *linear* models, *linear* observations, and *Gaussian* observation errors
- Kalman filter computationally expensive:
  - Solve large linear systems for many observations ($\mathcal{O}(10^5 \ldots 10^6)$)
  - Memory requirement: $\mathbf{P}^f$ has N^2 elements (N: grid points $\times$ d.o.f.)
  - Computation of $\mathbf{P}^a$, forecast of $\mathbf{P}^f$: $\sim N^3$ operations

  $\Longrightarrow$ some approximations will be needed anyway
- Generalization to more realistic, non-linear models required
  - Observation impact in Kalman filter depends only on specified errors, not on model state ("actual weather situation")!

Variational Method and Bayes Method



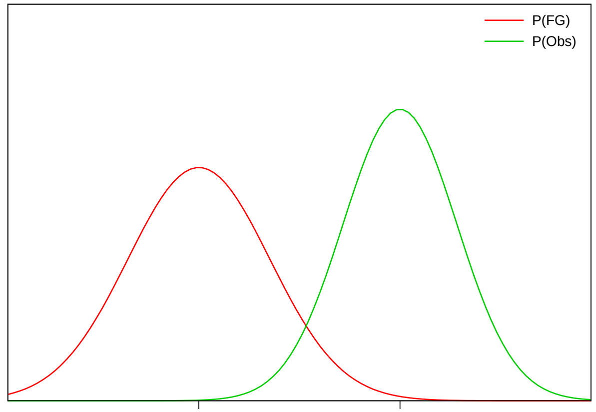
- Maximum-likelihood method
  - Likelihood for forecast

$$L_{\mathbf{P}^f}(x|x^f) \sim \exp\left[-\frac{1}{2}(x^f - x)^{\mathsf{T}} {\mathbf{P}^f}^{-1}(x^f - x)\right]$$

Variational Method and Bayes Method

- Maximum-likelihood method
  - Likelihood for forecast

$$L_{\mathbf{P}^f}(x|x^f) \sim \exp\left[-\frac{1}{2}(x^f - x)^{\mathsf{T}} {\mathbf{P}^f}^{-1}(x^f - x)\right]$$



  - Likelihood for new observations

$$L_{\mathbf{R}}(x|y) \sim \exp\left[-\frac{1}{2}(y - \mathcal{H}(x))^{\mathsf{T}} \mathbf{R}^{-1}(y - \mathcal{H}(x))\right]$$

Variational Method and Bayes Method

- Maximum-likelihood method
  - Likelihood for forecast

$$L_{\mathbf{P}^f}(x|x^f) \sim \exp\left[-\frac{1}{2}(x^f - x)^\mathrm{T}\mathbf{P}^{f^{-1}}(x^f - x)\right]$$

  - Likelihood for new observations

$$L_{\mathbf{R}}(x|y) \sim \exp\left[-\frac{1}{2}(y - \mathcal{H}(x))^\mathrm{T}\mathbf{R}^{-1}(y - \mathcal{H}(x))\right]$$

  - Forecast and observations are independent $\Rightarrow$ Likelihood function

$$L(x|x^f, y) = L_{\mathbf{P}^f}(x|x^f)\, L_{\mathbf{R}}(x|y)$$

  - Bayesian interpretation: maximize a-posteriori probability for given observations y and prior p(x):

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Variational method (3D-Var)

- **Cost function** ($J \equiv -2 \ln p + \text{const.}$):

$$J(x) = (x^f - x)^\mathsf{T} \left(\mathbf{P}^f\right)^{-1} (x^f - x) + (y - \mathcal{H}(x))^\mathsf{T} \mathbf{R}^{-1} (y - \mathcal{H}(x))$$

- Minimization: conjugate-gradient (CG), quasi-Newton methods
  - control variable: x
  - dual formulation (control variable y): "Physical Space Analysis System"

Variational method (3D-Var)

- **Cost function** ($J \equiv -2 \ln p + \text{const.}$):

$$J(x) = (x^f - x)^{\mathsf{T}} \left(\mathbf{P}^f\right)^{-1} (x^f - x) + (y - \mathcal{H}(x))^{\mathsf{T}} \mathbf{R}^{-1} (y - \mathcal{H}(x))$$

- Minimization: conjugate-gradient (CG), quasi-Newton methods
  - control variable: x
  - dual formulation (control variable y): "Physical Space Analysis System"
- Gradient of J:

$$\frac{1}{2} \nabla_x J(x) = \left(\mathbf{P}^f\right)^{-1} (x - x^f) + \mathbf{H}^{\mathsf{T}} \mathbf{R}^{-1} (\mathcal{H}(x) - y)$$

  with $\mathbf{H}$ from linearization of $\mathcal{H}(x)$
- J quadratic near minimum $\Rightarrow$ Hessian:

$$\frac{1}{2} \nabla_x \nabla_x J(x^a) = \left(\mathbf{P}^f\right)^{-1} + \mathbf{H}^{\mathsf{T}} \mathbf{R}^{-1} \mathbf{H} = (\mathbf{P}^a)^{-1}$$

- Solution x is best fit in model state space at $t = t_0$

General variational method (4D-Var)

- Cost function for model trajectory $x(t) = \mathcal{M}(t, x(0))$

$$
\begin{aligned}
J(x) &= (x^f(0) - x(0))^{\mathsf{T}} \mathbf{P}^{f\,-1} (x^f(0) - x(0)) \\
&+ \sum_i (y^{(i)} - \mathcal{H}(x(t_i)))^{\mathsf{T}} \mathbf{R}^{-1} (y^{(i)} - \mathcal{H}(x(t_i))) \\
&+ (\text{terms for model error}) + \ldots
\end{aligned}
$$

General variational method (4D-Var)

- Cost function for model trajectory $x(t) = \mathcal{M}(t, x(0))$

$$
\begin{aligned}
J(x) &= (x^f(0) - x(0))^\mathsf{T} \mathbf{P}^{f\,-1} (x^f(0) - x(0)) \\
&+ \sum_i (y^{(i)} - \mathcal{H}(x(t_i)))^\mathsf{T} \mathbf{R}^{-1} (y^{(i)} - \mathcal{H}(x(t_i))) \\
&+ (\text{terms for model error}) + \ldots
\end{aligned}
$$

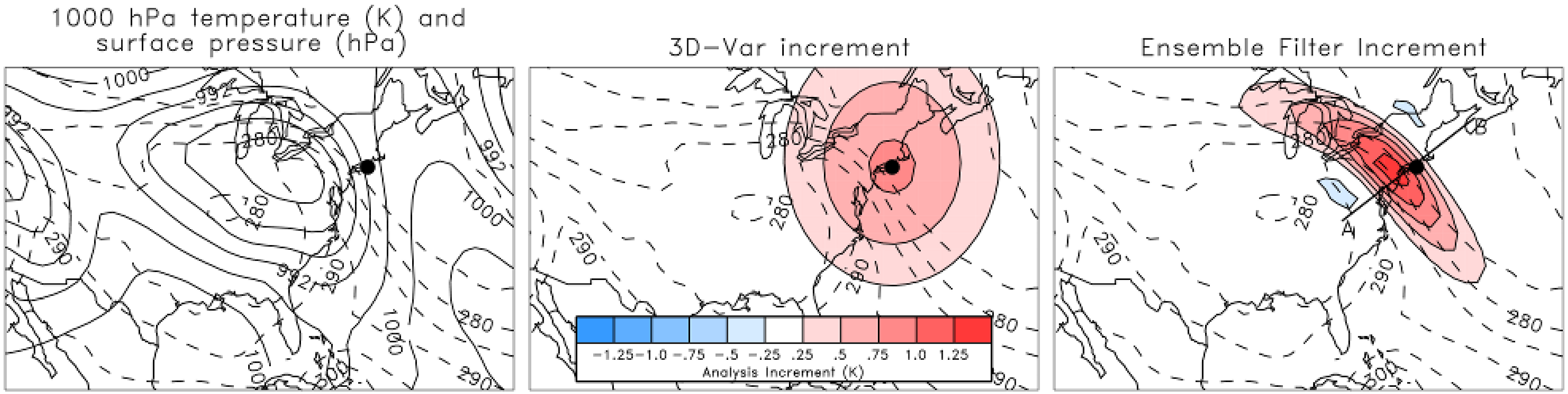
- Remarks
  - ☺ 4D-Var solution: $x(0)$, model trajectory with best fit to observations within time window
  - ☺ Implicit evolution of $\mathbf{P}^f$ matrix (but only in time window)
  - ☺ For linear model mathematically equivalent to Kalman filter/smoother
  - ☹ Huge development efforts necessary (tangent linear and adjoint model)
  - ☹ Large computational ressources required, but poor scaling on modern HPC architectures
  - ☹ Long-window 4D-Var $\longleftrightarrow$ strong nonlinearities at small scales

Ensemble Methods in Data Assimilation

- Why Ensemble Data Assimilation?

  Motivation: Impact of observations (determined by $\mathbf{R}$ and by $\mathbf{P}^f$) should depend on actual weather situation (e.g. near fronts)

  $\Longrightarrow$ explicit or implicit evolution of $\mathbf{P}^f$ essential!



Whitaker (2005)

Ensemble Kalman Filters

- Represent mean forecast, forecast uncertainty by ensemble (size: N) of (equal-weight) model states x_i^f drawn from appropriate p.d.f.
- Gaussian distribution characterized by first two moments:

$$\text{Departures:} \qquad x'^f_i = x_i^f - \bar{x}^f \qquad (\bar{x}^f : \text{Ensemble mean})$$

$$\text{Ensemble covariance:} \qquad \mathbf{P}^f = \frac{1}{N-1} \sum_{i=1}^{N} x'^f_i (x'^f_i)^T$$

- Generalized analysis equations

$$\begin{aligned} \text{Analysis mean:} \qquad \bar{x}^a &= \bar{x}^f + \mathbf{K}(\bar{y} - \mathbf{H}\bar{x}^f) \\ \text{Analysis departures:} \qquad x'^a &= x'^f + \tilde{\mathbf{K}}(y' - \mathbf{H}x'^f) \end{aligned}$$

  Choose $\mathbf{K}$, $\tilde{\mathbf{K}}$ such that $\mathbf{P}^a$ represents true analysis error covariance
- Implement model error $\mathbf{Q}$ as random forcing of forecast model

"The" Ensemble Kalman Filter (EnKF)

- Stochastic EnKF (Evensen, 1994; Burgers et al., 1998)
- $\tilde{\mathbf{K}} = \mathbf{K} = \mathbf{P}^f \mathbf{H}^\mathsf{T} (\mathbf{H}\mathbf{P}^f \mathbf{H}^\mathsf{T})^{-1} \quad \Longleftrightarrow \quad \mathrm{E}\left\{ y' (y')^T \right\} = \mathbf{R}$
  Observations must be randomly perturbed using true error covariance!

“The” Ensemble Kalman Filter (EnKF)

- Stochastic EnKF (Evensen, 1994; Burgers et al., 1998)
- $\tilde{\mathbf{K}} = \mathbf{K} = \mathbf{P}^f\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathsf{T})^{-1} \quad \Longleftrightarrow \quad \mathrm{E}\left\{y'(y')^T\right\} = \mathbf{R}$
  Observations must be randomly perturbed using true error covariance!
- rank($\mathbf{P}^f$)=$N-1$: analysis increments projected onto low-dimensional subspace of state space, leading to underestimation of analysis error and risk of filter collapse
  - Use cross-validation approach by splitting the ensemble
  - Artificially increase forecast error (“covariance inflation”)

“The” Ensemble Kalman Filter (EnKF)

- Stochastic EnKF (Evensen, 1994; Burgers et al., 1998)
- $\tilde{\mathbf{K}} = \mathbf{K} = \mathbf{P}^f \mathbf{H}^\mathsf{T} (\mathbf{H}\mathbf{P}^f \mathbf{H}^\mathsf{T})^{-1} \quad \Longleftrightarrow \quad \mathrm{E}\left\{ y'(y')^T \right\} = \mathbf{R}$
  Observations must be randomly perturbed using true error covariance!
- rank($\mathbf{P}^f$)=$N-1$: analysis increments projected onto low-dimensional subspace of state space, leading to underestimation of analysis error and risk of filter collapse
  - Use cross-validation approach by splitting the ensemble
  - Artificially increase forecast error (“covariance inflation”)
- $\mathbf{P}^f$ has spurious long-range correlations ($\sim N^{-1}$)
  - Suppressing the influence of distant observations requires “localization”:
    Schur product of $\mathbf{P}^f$ with suitably chosen correlation function or local analysis method (c.f. LETKF below)
  - Localization effectively increases the rank of $\mathbf{P}^f$ (and $\mathbf{K}$)
  - Localization is computationally expensive, disturbs model balance
- EnKF operationally used by Canadian Meteorological Service since 2005, now using 192 ensemble members (4 sub-ensembles $\times$ 48)

Deterministic Ensemble Kalman Filters

- Large family of Deterministic EnKF w/o need of perturbation of observations: **Ensemble Square-Root-Filters (EnSRF)** (see Whitaker and Hamill, 2002)
  - Ensemble Adjustment Kalman Filter (EAKF)
  - Ensemble Transform Kalman Filter (ETKF)
  - ...
  - Different localization concepts, differing computational efficiency

Deterministic Ensemble Kalman Filters

- Large family of Deterministic EnKF w/o need of perturbation of observations: **Ensemble Square-Root-Filters (EnSRF)** (see Whitaker and Hamill, 2002)
  - Ensemble Adjustment Kalman Filter (EAKF)
  - Ensemble Transform Kalman Filter (ETKF)
  - ...
  - Different localization concepts, differing computational efficiency
- Atmospheric dynamics locally low-dimensional (Patil et al., 2001)
  - λ_i: Eigenvalues of covariance matrices of (fast-growing) "bred-vectors" in a limited domain (1100 km $\times$ 1100 km)
  - "Bred-vector"-dimension:

$$\psi(\lambda_1, \ldots, \lambda_k) = \frac{(\sum_i \sqrt{\lambda_i})^2}{\sum_i \lambda_i} \quad \ll \quad k$$

- For local covariances, about 40–100 ensemble members needed from storm to global scales (Kalnay et al.)

LETKF: Local Ensemble Transform Kalman-Filter

- LETKF (Ott et al., 2004; Hunt et al., 2004, 2007):
  - Forecast perturbations: $\mathbf{X}^f = [\, x_1 - \bar{x} \,|\, \ldots \,|\, x_N - \bar{x} \,]$
    Obs. in ensemble space: $\mathbf{Y} = [\, y_1 - \bar{y} \,|\, \ldots \,|\, y_N - \bar{y} \,]$, $y_i \equiv \mathcal{H}(x_i)$
  - **Local analysis** in **ensemble space**, at each grid point:

$$\tilde{\mathbf{P}}^f = (N-1)^{-1}\mathbf{I} \,, \quad \tilde{\mathbf{P}}^a = \left[(N-1)\mathbf{I} + \mathbf{Y}^\mathsf{T}\mathbf{R}^{-1}\mathbf{Y}\right]^{-1}$$

using all observations in the local region.

LETKF: Local Ensemble Transform Kalman-Filter

- LETKF (Ott et al., 2004; Hunt et al., 2004, 2007):
  - Forecast perturbations: $\mathbf{X}^f = [\, x_1 - \bar{x} \,|\, \ldots \,|\, x_N - \bar{x} \,]$

    Obs. in ensemble space: $\mathbf{Y} = [\, y_1 - \bar{y} \,|\, \ldots \,|\, y_N - \bar{y} \,]$, $y_i \equiv \mathcal{H}(x_i)$
  - **Local analysis** in **ensemble space**, at each grid point:

    $$\tilde{\mathbf{P}}^f = (N-1)^{-1}\mathbf{I} \,, \quad \tilde{\mathbf{P}}^a = \left[(N-1)\mathbf{I} + \mathbf{Y}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{Y}\right]^{-1}$$

    using all observations in the local region.
  - Analysis mean: $\bar{x}^a = \bar{x}^f + \mathbf{X}^f \bar{w}$, $\bar{w} = \tilde{\mathbf{P}}^a \mathbf{Y} \mathbf{R}^{-1} (y - \bar{y})$

    Analysis ensemble: $\mathbf{X}^a = \mathbf{X}^f \mathbf{W} + \bar{x}^a$, $\mathbf{W} = \left[(N-1)\,\tilde{\mathbf{P}}^a\right]^{1/2}$
  - Localization: apply to (inverse) observation error!
  - Computationally very efficient, well parallelizable:
    symmetric square-roots of $N \times N$ matrices at each grid-point
  - No minimization $\Longrightarrow$ no adjoint of observation operators needed

LETKF: Local Ensemble Transform Kalman-Filter

- LETKF (Ott et al., 2004; Hunt et al., 2004, 2007):
  - Forecast perturbations: $\mathbf{X}^f = [\,x_1 - \bar{x}\,|\ldots|\,x_N - \bar{x}\,]$
    Obs. in ensemble space: $\mathbf{Y} = [\,y_1 - \bar{y}\,|\ldots|\,y_N - \bar{y}\,]\ , \quad y_i \equiv \mathcal{H}(x_i)$
  - **Local analysis** in **ensemble space**, at each grid point:

$$\tilde{\mathbf{P}}^f = (N-1)^{-1}\mathbf{I}\,, \quad \tilde{\mathbf{P}}^a = \left[(N-1)\mathbf{I} + \mathbf{Y}^\mathsf{T}\mathbf{R}^{-1}\mathbf{Y}\right]^{-1}$$

    using all observations in the local region.
  - Analysis mean: $\bar{x}^a = \bar{x}^f + \mathbf{X}^f \bar{w}\,, \quad \bar{w} = \tilde{\mathbf{P}}^a \mathbf{Y}\mathbf{R}^{-1}(y - \bar{y})$
    Analysis ensemble: $\mathbf{X}^a = \mathbf{X}^f\mathbf{W} + \bar{x}^a\,, \quad \mathbf{W} = \left[(N-1)\,\tilde{\mathbf{P}}^a\right]^{1/2}$
  - Localization: apply to (inverse) observation error!
  - Computationally very efficient, well parallelizable:
    symmetric square-roots of $N \times N$ matrices at each grid-point
  - No minimization $\Longrightarrow$ no adjoint of observation operators needed
  - DWD: KENDA (experimental LETKF for COSMO-DE-EPS)
  - Hybrid global 3D-Var/LETKF, with high-resolution deterministic analysis: e.g. GFS at NCEP (operational); GME at DWD (experim.)

Nonlinear filtering & Particle Filter

Nonlinear filtering & Particle Filter

- Bayes theorem

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\mathrm{d}x}$$

Nonlinear filtering & Particle Filter

- Bayes theorem

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)\mathrm{d}x}$$

- Ensemble representation of probability density ("particles")

$$p(x) = \sum_{i=1}^{N} \frac{1}{N}\,\delta(x - x_i)$$

- Posterior density

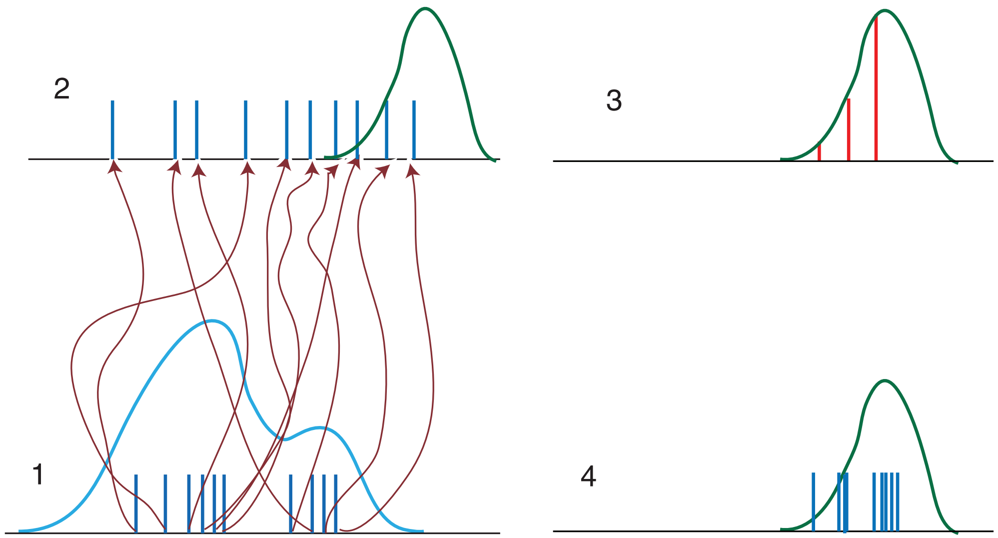
$$p(x|y) = \sum_{i=1}^{N} w_i\,\delta(x - x_i) \qquad \text{with weights} \quad w_i = \frac{p(y|x_i)}{\sum_i p(y|x_i)}$$

- For Gaussian distributed observations

$$w_i \propto \exp\left[-\frac{1}{2}(y - \mathcal{H}(x_i))^{\mathrm{T}}\mathbf{R}^{-1}(y - \mathcal{H}(x_i))\right]$$

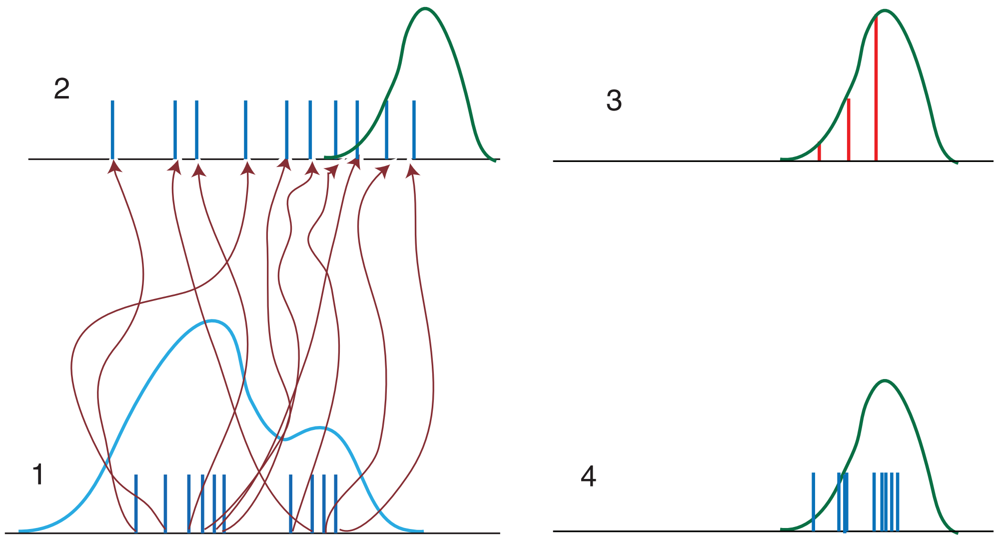
Standard Particle Filter

- Sequential Importance Resampling Filter (van Leeuwen, 2003)
  (1) Generate Ensemble as sample of the probability density, run the model
  (2) Assign weights accordings to probability density given the observations
  (3) Reject particles with small weights and
  (4) replace by particles according to posteriory density (resampling)



Standard Particle Filter

- Sequential Importance Resampling Filter (van Leeuwen, 2003)
  (1) Generate Ensemble as sample of the probability density, run the model
  (2) Assign weights accordings to probability density given the observations
  (3) Reject particles with small weights and
  (4) replace by particles according to posteriory density (resampling)



- Standard PF inefficient: many particles lost for many observations!
- Required ensemble size $\sim$ exp (number of observations)!
- For particle filters to be useful, we need importance sampling!

Bayes Theorem and proposal transition density (I)

- Stochastic model equation (β: model error, e.g. with $\beta \sim N(0, \mathbf{Q})$)

$$x^n = f(x^{n-1}) + \beta^{n-1}$$

- Transition density

$$p(x^n|x^{n-1}) = N(f(x^{n-1}), \mathbf{Q})$$

- Use transition density to derive marginal prior p.d.f. at time n

$$p(x^n) = \int p(x^n|x^{n-1})p(x^{n-1})\,\mathrm{d}x^{n-1}$$

- Rewrite Bayes theorem using (arbitrary) proposal transition density q

$$\begin{aligned} p(x^n|y^n) &= \frac{p(y^n|x^n)p(x^n)}{p(y^n)} \\ &= \frac{p(y^n|x^n)}{p(y^n)} \int p(x^n|x^{n-1})p(x^{n-1})\,\mathrm{d}x^{n-1} \\ &= \frac{p(y^n|x^n)}{p(y^n)} \int \frac{p(x^n|x^{n-1})}{q(x^n|x^{n-1}, y^n)} q(x^n|x^{n-1}, y^n)p(x^{n-1})\,\mathrm{d}x^{n-1} \end{aligned}$$

Proposal transition density (II)

- Starting with equal-weight particles at $n-1$, we have:

$$p(x^n|y^n) = \frac{p(y^n|x^n)}{p(y^n)} \frac{1}{N} \sum_{i=1}^{N} \frac{p(x^n|x_i^{n-1})}{q(x^n|x_i^{n-1}, y^n)} q(x^n|x_i^{n-1}, y^n)$$

- For each particle at $n-1$ draw from proposal transition density q

$$p(x^n|y^n) = \frac{1}{N} \sum_{i=1}^{N} \frac{p(y^n|x_i^n)}{p(y^n)} \frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)} \delta(x^n - x_i^n)$$

- Weights

$$w_i = \underbrace{\frac{p(y^n|x_i^n)}{p(y^n)}}_{\text{Likelihood weight}} \times \underbrace{\frac{p(x_i^n|x_i^{n-1})}{q(x_i^n|x_i^{n-1}, y^n)}}_{\text{Proposal weight}}$$

- The proposal transition density is essentially arbitrary!
  Can we use it to draw the particles closer to the observations?

Implicit particle filter (I)

- For weakly non-linear observation operators ($\mathcal{H}(x_i) \approx \mathbf{H}x_i$)

$$w_i \cdot q \quad \propto \quad \exp\left[-\frac{1}{2}(y - \mathbf{H}x_i)^\mathsf{T}\mathbf{R}^{-1}(y - \mathbf{H}x_i) - \frac{1}{2}\left(x_i^n - f(x_i^{n-1})\right)^\mathsf{T}\mathbf{Q}^{-1}\left(x_i^n - f(x_i^{n-1})\right)\right]$$

  This is a quadratic function in $(x_i^n - \mu_i)$ with Hessian H.
- The minimum μ_i can be determined by e.g. variational methods. Choose proposal density as:

$$q(x_i^n|x_i^{n-1}, y^n) = N(\mu_i, H^{-1})$$

- Resulting weights:

$$w_i \propto \exp\left[-\frac{1}{2}\left(y^n - \mathbf{H}f(x_i^{n-1})\right)^\mathsf{T}\left(\mathbf{HQH}^\mathsf{T} + \mathbf{R}\right)^{-1}\left(y^n - \mathbf{H}f(x_i^{n-1})\right)\right]$$

Implicit particle filter (II)

- Implicit particle filter: run special 4D-Var for each particle to obtain μ_i
- Determine (approximate) Hessian H
- Generate proposal particles from

$$q(x_i^n | x_i^{n-1}, y^n) = N(\mu_i, H^{-1})$$

- Calculate weights
  (Note: no explicit expression for weights for non-linear models).
- Resample

- See Atkins et al., 2013, and references for a review and the connection between variational methods and implicit particle methods.

Equivalent-Weights Particle Filter

- Van Leeuwen, 2010:
  - Assume that model needs several time steps between observations
  - Use simple proposal at each time step, e.g. corresponding to nudging:

$$q(x^n|x^{n-1}, y^n) = N\left(f(x^{n-1}) + S\left(y^n - \mathcal{H}(x^{n-1})\right), \mathbf{Q}\right)$$

  - Use **different proposal at final time step** to achieve similar weights. Determine maximum achievable weight ($w_i^{\max}$) during last time step(s) and choose target weight w^{target}. Then set:

$$q(x^n|x^{n-1}, y^n) = \begin{cases} q_1(x^n|x^{n-1}, y^n) & \text{for } w_i^{\max} > w^{\text{target}} \\ q_2(x^n|x^{n-1}, y^n) & \text{for } w_i^{\max} < w^{\text{target}} \end{cases}$$

    For particles that cannot reach the target weight, one uses:

$$q_2(x^n|x^{n-1}, y^n) = N\left(f(x^{n-1}), \mathbf{Q}\right)$$

    Choose "special move" (different forcing) for high-weight particles
  - Calculate weights, resample "lost" particles

Comments on Particle Filters

- Particle Filters do not need state covariances
- No assumptions on linearity, Gaussianity, . . . , needed
- Perfectly scalable (number of particles, dimensionality of problem)
- Proposal transition density may solve the degeneracy problem connected with many observations, apparently highly flexible
- In practice, particle filters have only been shown yet to work with toy models (Lorenz '63, Lorenz '96, barotropic vorticity equation, . . .), no experience with general circulation models:
  - Arbitrary forcing terms may destroy model balances
  - How to resample efficiently?
  - Some kind of localization similar to Ensemble Kalman Filters may be beneficial, but it is not clear how to do this (cannot linearly combine particles!)
- Many interesting ideas are being investigated . . .

Concluding Remarks

- Ensemble methods have been established as a valuable method in data assimilation to estimate the analysis uncertainties
  - Several competitive methods which are also efficient on modern massively parallel computer architectures (EnKF, LETKF, hybrid Var/EnKF methods, ...)
  - Ensemble DA already operationally used at some meteorological centers, more centers to follow
  - Key challenges have shifted: from poorly known $\mathbf{P}^f$ in deterministic DA to poorly understood model error ($\mathbf{Q}$) in ensemble DA
    $\longrightarrow$ many ideas, but mostly ad-hoc solutions and lots of tuning

Concluding Remarks

- Ensemble methods have been established as a valuable method in data assimilation to estimate the analysis uncertainties
  - Several competitive methods which are also efficient on modern massively parallel computer architectures (EnKF, LETKF, hybrid Var/EnKF methods, ...)
  - Ensemble DA already operationally used at some meteorological centers, more centers to follow
  - Key challenges have shifted: from poorly known $\mathbf{P}^f$ in deterministic DA to poorly understood model error $(\mathbf{Q})$ in ensemble DA
    $\longrightarrow$ many ideas, but mostly ad-hoc solutions and lots of tuning
- While Ensemble Kalman Filters work well on the synoptic scale, DA for the convective scale is still a subject of current research
  - Ensemble methods are essential to capture the flow dependence of uncertainties
  - Strong non-linearities in models and observation operators pose challenges to all methods
  - Particle filters are a promising candidate, but efficient importance sampling is a challenge to prevent ensemble collapse, still a lot to learn!

Selected References

- Ensemble Kalman Filter (EnKF)
  - G. Burgers et al., 1998. On the analysis scheme of the Ensemble Kalman Filter, Monthly Weather Rev., 126, 1719–1724.
  - Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10,143–10,162.
  - P. L. Houtekamer, et al., 2005. Atmospheric data assimilation with an ensemble Kalman filter: results with real observations, Mon. Wea. Rev. 133, 604–620.
- Ensemble Square Root Filters, Ensemble Transform Kalman Filters
  - Hunt et al., 2004. Four-dimensional ensemble Kalman filtering, Tellus, 56A, 273–277.
  - Ott et al., 2004. A local ensemble Kalman filter for atmospheric data assimilation, Tellus, 56A, 415–428.
  - Patil et al., 2001. Local Low Dimensionality of Atmospheric Dynamics, Phys. Rev. Lett., 86, 5878–5881
  - Whitaker, J.S. and Hamill, T.H., 2002. Ensemble Data Assimilation without perturbed observations, Mon. Wea. Rev. 130, 1913–1924.
- Particle Filter
  - Van Leeuwen, P.J., 2010. Nonlinear data assimilation in geosciences: an extremely efficient particle filter, Q. J. R. Meteorol. Soc. 136: 1991–1999. DOI:10.1002/qj.699
  - Snyder et al., 2008. Obstacles to high-dimensional particle filtering, Mon. Wea. Rev. 136, 4629–4640.
  - Atkins et al., 2013. Implicit particle methods and their connection with variational data assimilation, Mon. Wea. Rev., in print.

Questions?

Localization

- Let $\mathbf{P}$ be symmetric positive semi-definite, and $\mathbf{C}$ (the localization) be a correlation function. The Schur product

  $$\widetilde{\mathbf{P}} \equiv \mathbf{P} \odot \mathbf{C}$$

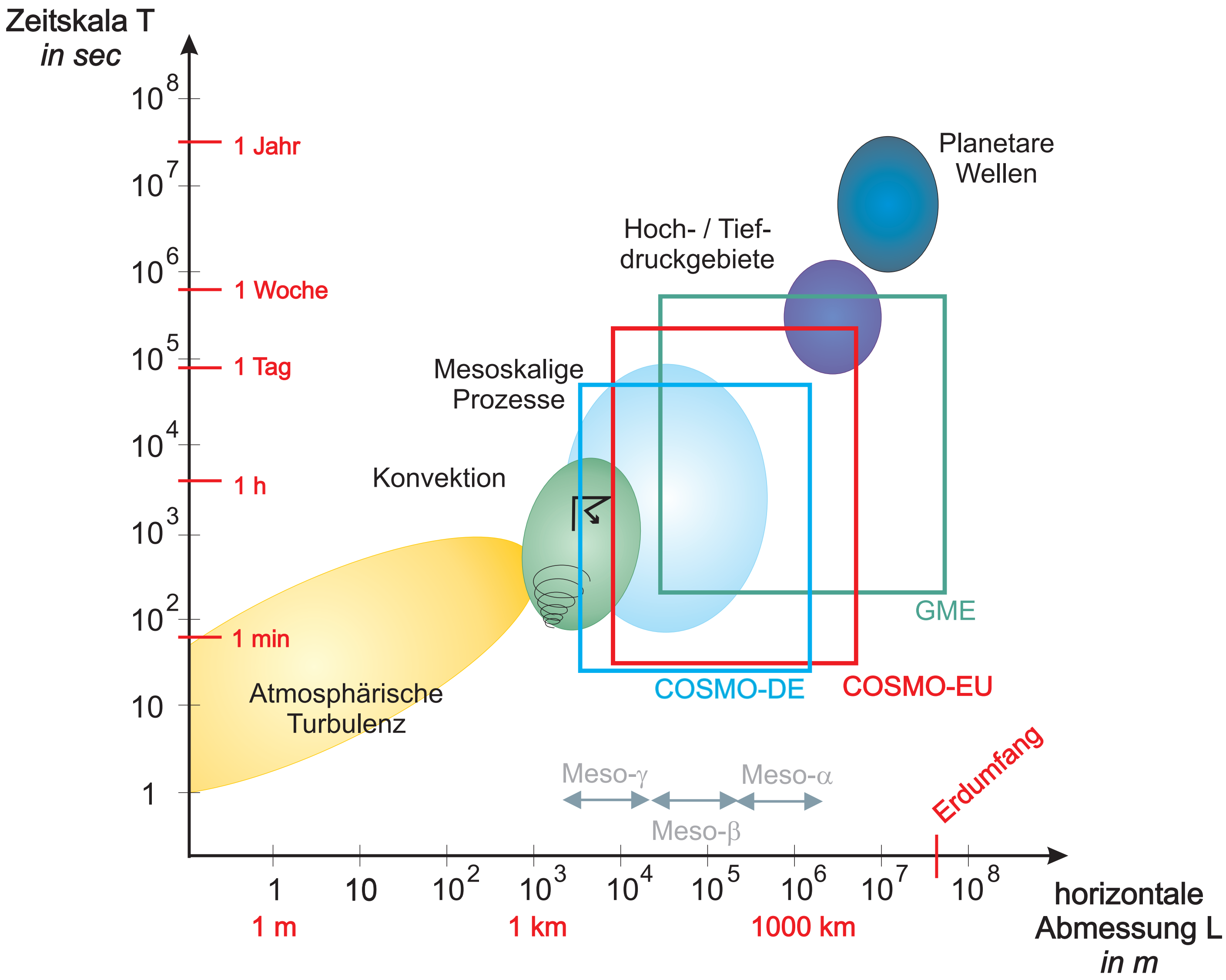
  defines the localized version of $\mathbf{P}$

  $$(\mathbf{P} \odot \mathbf{C})(x, y) \equiv P(x, y)C(x, y)$$

  which is also positive semi-definite.

  (See Gaspari & Cohn, 1999, for a treatment of localization on the sphere.)
- In practice, localization of $\mathbf{P}^{\mathbf{f}}$ is too expensive. Many authors apply localization to $(\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathrm{T})$ and $(\mathbf{P}^f\mathbf{H}^\mathrm{T})$, which leads to small but systematic errors.

Length- and time-scales in the atmosphere



Global Observing System

Observation usage of DWD's global data assimilation on 2011-02-01

Observation type	variable	Total	used	monitored
TEMP	T, rh, u, v	57 258	5.4 %	199 973
PILOT	u, v	4 200	0.4 %	15 122
SYNOP	p_s, u, v	112 647	10.7 %	114 558
DRIBU	p_s, u, v	6 353	0.6 %	6 643
Aircraft	T, u, v	225 947	21.4 %	250 578
AMV geo	u, v	86 620	8.2 %	99 844
AMV polar	u, v	25 332	2.4 %	25 384
SCATT	u_{10}, v_{10}	188 774	17.9 %	224 412
AMSU-A	T_b	287 950	27.3 %	18 838 147
GPSRO	bending angle	60 659	5.7 %	65 052
All data		1 055 740	100.0 %	19 839 713
Polar satellites		562 715	53.3 %	19 152 995