Treatment of model and observation error in the ensemble data assimilation system

Tijana Janic-Pfander, Andreas Rhodin
Jaison Ambadan, Hendrik Reich

Stochastic Physics week DWD Offenbach

2013-11-26

1. Ensemble Data Assimilation Algorithm
   - KF – Kalman Filter
   - EnKF – Ensemble Kalman Filter
2. Error Covariances
3. EnDA at DWD
   - Purpose of EnDA
   - LETKF setup for ICON and COSMO
   - Model error approaches currently used
4. Some Preliminary Results
   - additive 3D-Var **B**
   - Soil Moisture – 2m Temperature Correlations
   - SKEB in GME
   - Adaptive inflation in COSMO

Ensemble Data Assimilation Algorithms

Based on the Ensemble Kalman Filter Algorithm (as to be used at DWD) the role of model and observational error will be discussed.

There are other approaches as running independent assimilation systems with perturbed observations (4D-Var, ECMWF). The role of model and observational error remains the same.

Algorithm: KF – Kalman Filter

Data Assimilation:

- Provide "optimal" initial state and information on its uncertainty. From observations $\mathbf{o}$, previous forecasts $\mathbf{x}^{(b)}$ (and their uncertainties)

(Extended) Kalman Filter:

- For a linear system $\mathbf{M}$ and Gaussian errors the KF provides the exact pdf (error covariance matrix $\mathbf{P}$) for forecast$^{(b)}$ and analysis$^{(a)}$. For a nonlinear system $\mathcal{M}$ (Extended KF) this holds approximately.

	Model state $\mathbf{x}$	Error Covariances $\mathbf{P}$
Forecast	$\mathbf{x}^{(b)} = \mathcal{M}(\mathbf{x}^{(a)})$	$\mathbf{P}^{(b)} = \mathbf{M}\mathbf{P}^{(\mathbf{a})}\mathbf{M}^{\mathbf{T}} + \mathbf{Q}$
Analysis	$\mathbf{x}^{(a)} = \mathbf{x}^{(b)} + \mathbf{K}(\mathbf{o} - \mathbf{H}\mathbf{x}^{(b)})$	$\mathbf{P}^{(a)} = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{P}^{(\mathbf{b})}$

$$\text{with } \mathbf{K} = \mathbf{P}^{(\mathbf{b})}\mathbf{H}^t(\mathbf{H}\mathbf{P}^{(\mathbf{b})}\mathbf{H}^t + \mathbf{R})^{-1}$$

- The KF requires specification of model and observation error covariances $\mathbf{Q}$, $\mathbf{R}$.
- $\mathbf{P}^{(a)}$ is not prescribed but depends on $\mathbf{Q}$, $\mathbf{R}$, observation coverage

Algorithm: EnKF – Ensemble Kalman Filter

- The KF is not applicable to large problems.
  - $\mathbf{MP^{(a)}M^T} + \mathbf{Q}$ requires $O(N)$ integrations of the tangent linear Model $\mathbf{M}$.
    (N = number of degrees of freedom of the model)
- EnKF: Probabilistic approach:
  - approximate $\mathbf{P}^{(b)}$ and $\mathbf{P}^{(a)}$ by ensembles of model states $\mathbf{x}_i^{(b)}$, $\mathbf{x}_i^{(a)}$:
    $\mathbf{P}^{(b)} = \mathbf{X}^{(b)}\mathbf{X}^{(b)\,\mathbf{T}}$
    $\mathbf{X}^{(b)}$ = vector of ensemble deviations from the mean.
  - approach requires L integrations of the nonlinear model:
    $\mathbf{x}_i^{(b)} = \mathcal{M}(\mathbf{x}_i^{(a)}) + \mathbf{q}$.
    $i = 1 \ldots L \quad$ (L =ensemble size).
  - error in $\mathbf{P}^{(b)}$ scales with $1/\sqrt{L}$.
  - requires representation of model error covariances $\mathbf{Q}$ by a stochastic process $\mathbf{q}$.
    (with reasonable amplitude, flow dependence, correlation length scales)

Error Covariances

- The EnKF requires suitable ensemble forecast covariances $\mathbf{P}^{(b)}$ (both variances and correlations).
- DA systems rely on forecast error (cross-)correlations to update unobserved variables.
  Example: slide 2
  - The ensemble must be of sufficient size to provide correlations with statistical significance (40 . . . 200).
  - $\mathbf{q}$ should have reasonable characteristics:
    - flow dependence
    - amplitute
    - spatial correlations
    - temporal correlations
- Currently a simple approach is used for representing $\mathbf{q}$
  - Statistical noise consistent with the balance conditions of the 3D-Var $\mathbf{P}^{(b)}$ matrix: $\mathbf{q}_i = {\mathbf{P}^{(b)}_{3dvar}}^{1/2} \epsilon_i$
  - Inflation: $\mathbf{X} \to \gamma\mathbf{X}$, $\gamma > 1$.

Summary: Stochastic Representation of Model Error

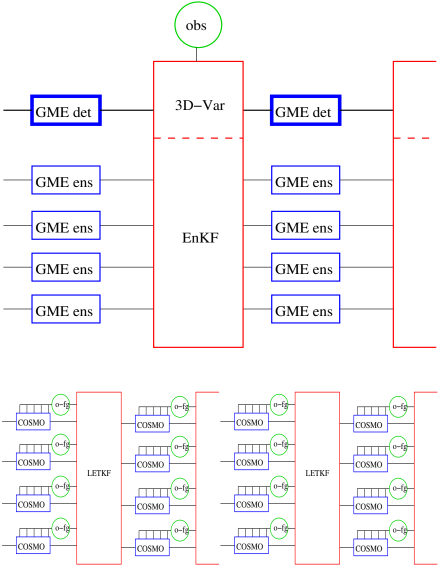
- Current DA schemes aim to model
  the flow dependent evolution of forecast errors
  in order to use it in the analysis step.
- Key ingredients are observational errorrs **R** and model errors **Q**.
- Due to the size of the problem we have to use ensemble methods
- Ensemble DA algorithms require a stochastic process **q**
  to represent **Q**.

EnKF at DWD

Purpose of the Ensemble Data Assimilation System

- Provide an optimal analysis
  - Uses flow dependent **B** to cycled DA system
  - Requires reasonable spread/skill relationship
  - Requires reasonable cross-correlations
  - Requires ensemble size of $O(40 \ldots 200)$
  - Requires reasonable model error representation
- Provide BC for local area model
- Provide initial conditions for ensemble prediction system

LETKF setup for ICON and COSMO



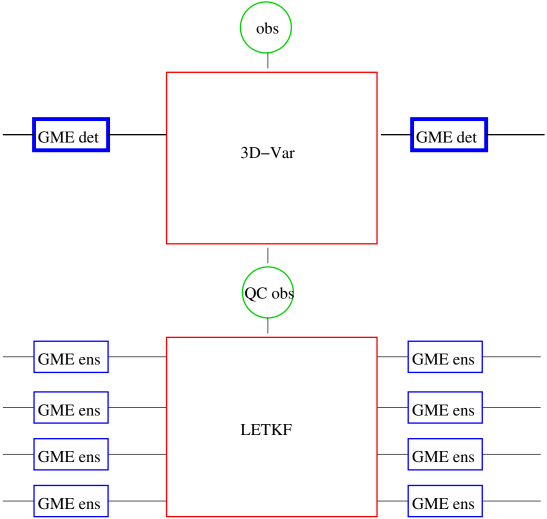
ICON/GME:

- High resolution deterministic analysis (currently 3D-Var)
- Low resolution ensemble analysis (currently LETKF)
  - First guess calculated in the analysis scheme

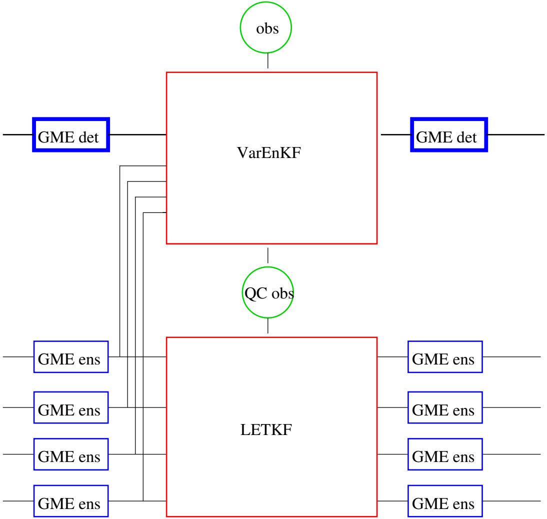
COSMO:

- Convection resolving ensemble analysis (4D-LETKF)
  - First guess at appropriate time calculated in the model
- (Not shown:) Deterministic analysis using LETKF gain matrix

GME/ICON EnKF – Currently: 3D-Var + LETKF



GME/ICON EnKF – Under development: VarEnKF



Model error approaches currently used

	GME/ICON	COSMO/KENDA
(adaptive) inflation	X	X
additive 3D-Var **B**	X	
SPPT		X[1]
SKEB	X[2]	X

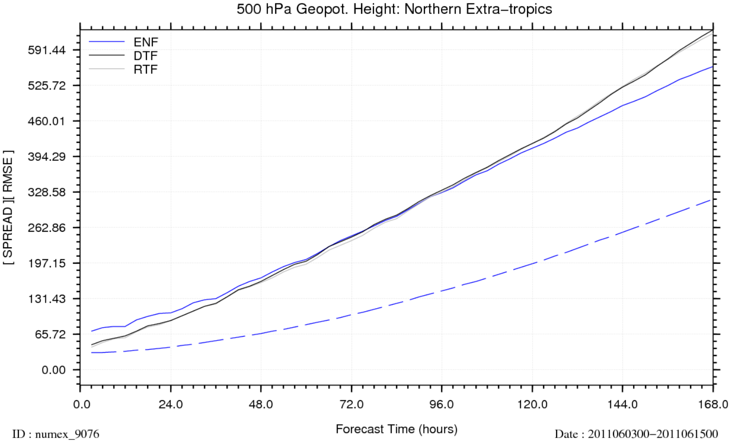
[1] Recently implemented in COSMO, cf. talk of Lucio Torrisi

[2] Implemented in GME by Jaison Ambadan

Some Preliminary Results

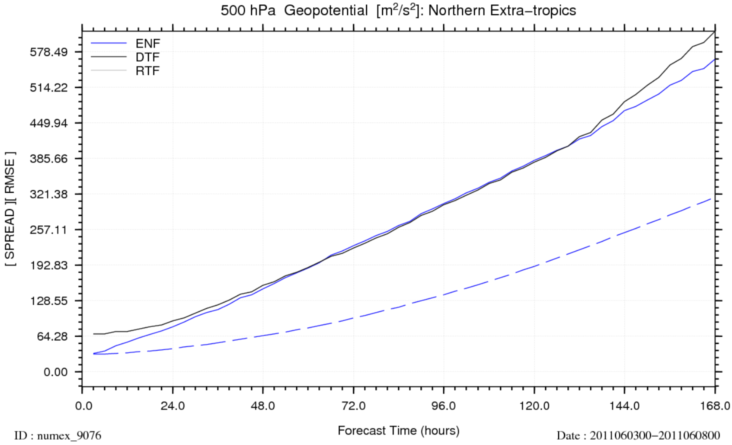
additive 3D-Var **B** (GME LETKF)

Verification of Ensemble Mean (vs. operational analysis)



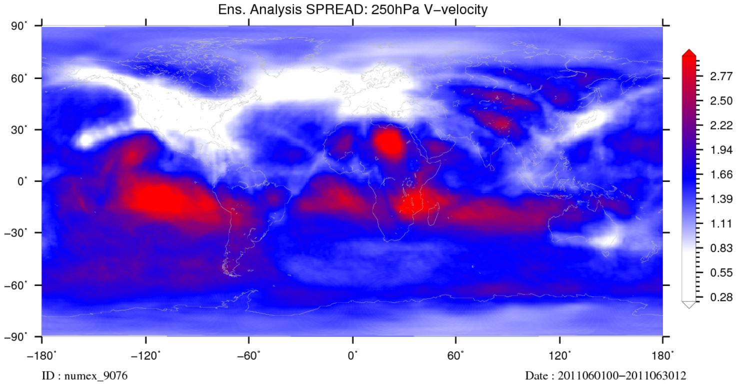
DTF=3D-Var ENF=LETKF line=RMSE broken=SPREAD

Verification of Ensemble Mean (vs. its own analysis)

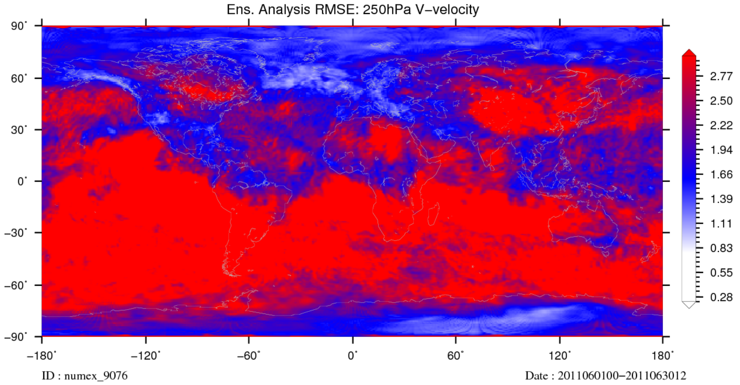


DTF=3D-Var ENF=LETKF line=RMSE broken=SPREAD

Background Ensemble SPREAD



Background Ensemble Mean RMSE (vs.operational analysis)



Known Deficiencies, Ongoing Calibration

LETKF Ensemble vs. 3D-Var Routine

- RMSE generally larger until 24 h
- Comparable or better $>$ 48 h

Known Deficiencies

- Spread generally too small
- .. much too small in well observed areas
- RMSE too large at model top
- Humidity bias

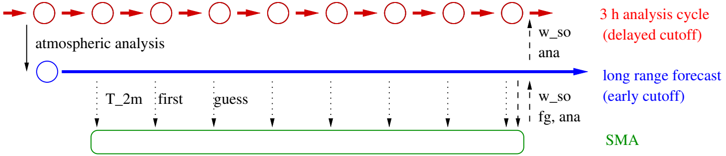
Calibration

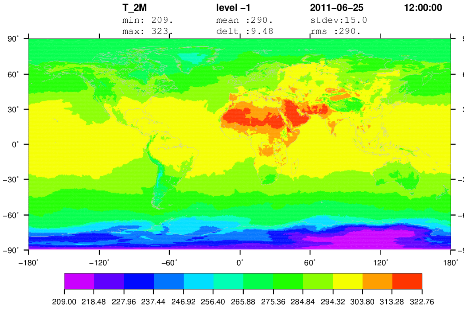
- adjust inflation, localisation, model & observational error
- use adaptive methods

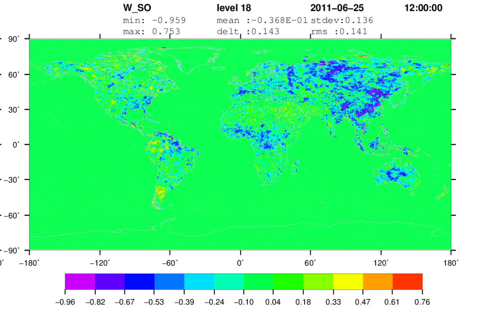
Soil Moisture – 2m Temperature Correlations

- Soil moisture (and soil moisture spread) is not analysed/updated in the EnKF so far.
- spread of the (unobserved) soil moisture and correlations with 2m tenperature are reasonable, caused by interaction with the atmosphere.
- correlations with 2m temperature may be utilized for soil moisture assimilation

2m Temperature – Soil Moisture Correlations at 12 UT (Perspective for SMA in the EnKF framework)

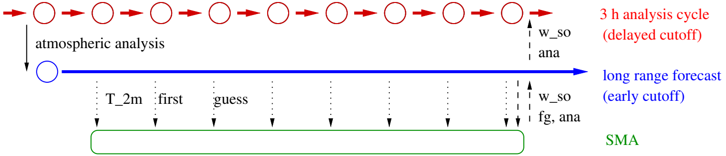


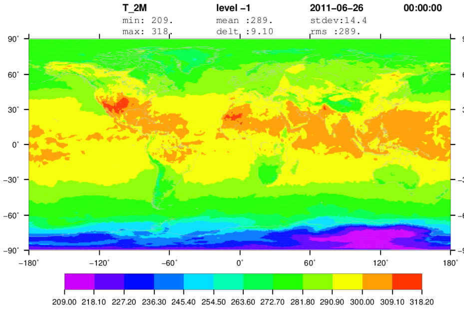


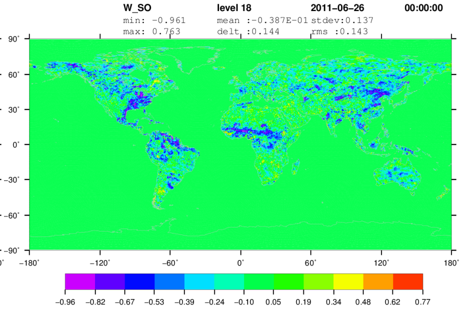


12 UT 2m mean temperature and correlation with 24 UT 18cm soil moisture

2m Temperature – Soil Moisture Correlations at 24 UT (Perspective for SMA in the EnKF framework)





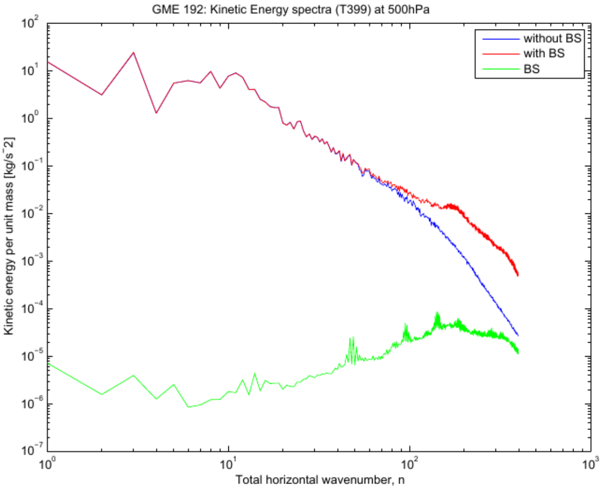


24 UT 2m mean temperature and correlation with 24 UT 18cm soil moisture

SKEB in GME

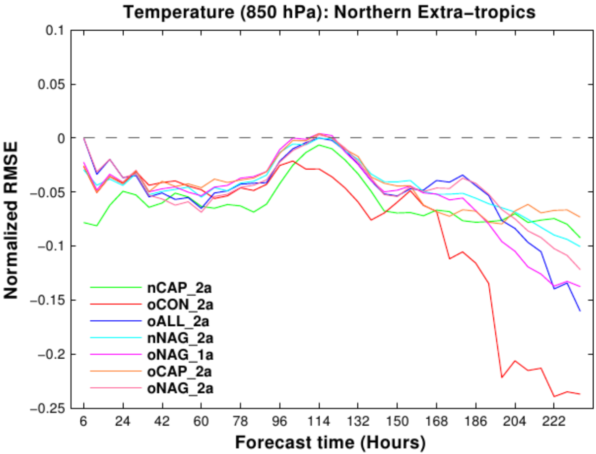
- SKEB scheme implemented in GME (J. Ambadan)
- improved scores in GME ensemble prediction
- no benefit in the assimilation cycle so far

SKEB in GME



GME Kinetic Energy Spectrum (J.Ambadan)

SKEB in GME

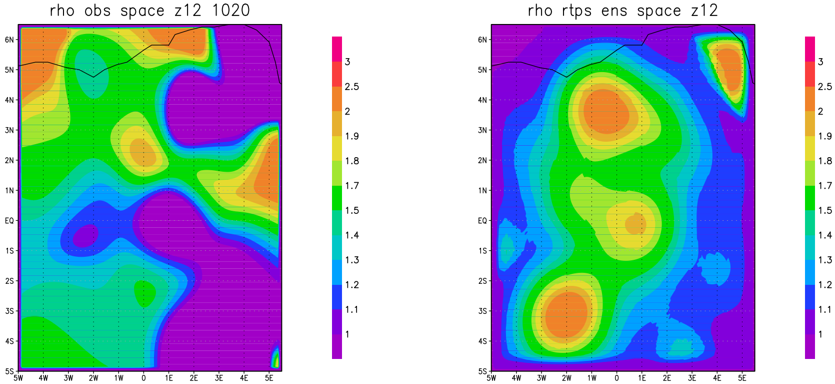


Normalized RMSE differences for various configurations of SKEP compared to deterministic forecast

Adaptive inflation in COSMO

- lack of spread is (partly) due to model error which is not accounted for so far
- there may be are other causes for lack of spread:
  - insufficient ensemble size
  - inappropriate localisation radius
  - inappropriate observation error and correlations
- one (simple) method to increase spread is multiplicative covariance inflation:
  - $\mathbf{X}_{ens} \to \rho \mathbf{X}_{ens}$ with $\rho > 1$.
- adaptive method to estimate ρ preferable
  two methods have been used:
  - Desroziers et al.: Compare observed statistics with expected ones: $\langle (y - H(x_b))(y - H(x_b))^T \rangle = \mathbf{R} + \rho \mathbf{H}\mathbf{P}_b\mathbf{H}^T$
  - relaxaxation to prior spread (RTPS): $\rho = \sqrt{\alpha \frac{\sigma_b - \sigma_a}{\sigma_a} + 1}, \quad \alpha > 1$

Adaptive inflation in COSMO



Inflation factors estimated by Desroziers method (left) or RTPS right.