PDAF Parallel Data Assimilation Framework

ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

A hybrid nonlinear-Kalman ensemble transform filter with regulation following the system nonlinearity

Lars Nerger

Alfred Wegener Institute
Helmholtz Center for Polar and Marine Research
Bremerhaven, Germany

Invited Seminar at Meteorological Research Institute, Tsukuba, Japan, November 20, 2025

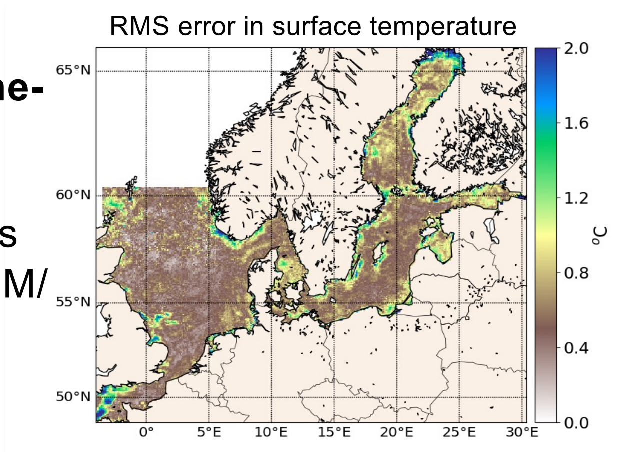
Overview

- Linear and nonlinear ensemble filters
- Hybrid nonlinear-Kalman filter
- Application tests

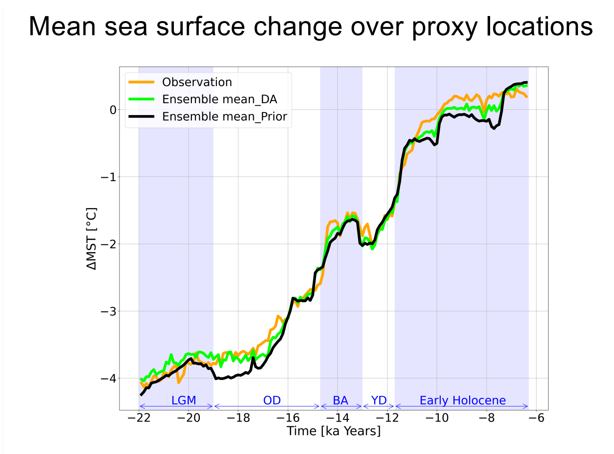
Application Examples (different models utilizing PDAF)

PDAF Parallel Data Assimilation Framework

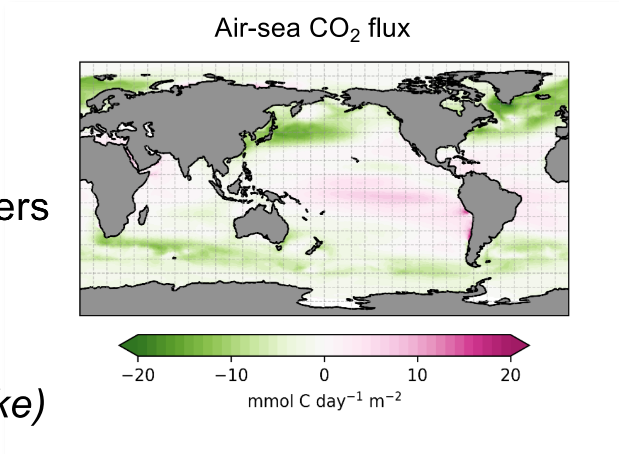
Coastal coupled physics/biogeochemistry DA: CMEMS/BSH - Improving forecasts with NEMO-ERGOM/ HBM-ERGOM: *(S. Vliegen, A. Sathanarayanan)*



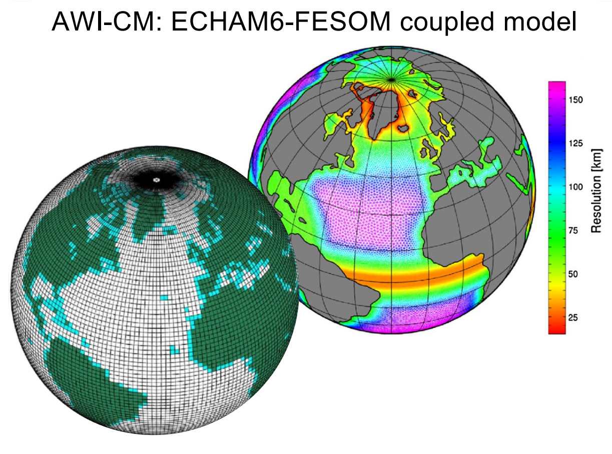
Paleo-climate DA: improve simulation of last deglaciation with CLIMBER-X *(A. Masoum)*



Coupled physics/biogeochemistry DA: Improving parameters & carbon flux in REcoM *(N. Mamnun, F. Bunsen, A. Broschke)*



Coupled ocean-atmosphere DA: Assimilate ocean observations into atmosphere with AWI-CM *(Q. Tang)*



Motivation

Data Assimilation (DA) into non-linear models (e.g. nonhydrostatic dynamics)

- Nonlinearity in DA → non-Gaussianity of error distributions
- For Gaussian distributions: (Ensemble) Kalman filter is optimal
- For non-Gaussian distributions:
  - → Kalman filter suboptimal or failing
  - → Non-linear DA methods (e.g. particle filters): possibly better estimates, but higher sampling errors than KFs

Non-linear model dynamics lead to different degree of non-Gaussianity

→ Aim for DA method that

- adapts to non-Gaussianity
- Allows to utilize optimality of KF for Gaussian distributions

→ **Utilize hybrid combination of KF and nonlinear filter**

Linear and Nonlinear Ensemble Filters

Linear and Nonlinear Ensemble Filters

- Represent state and its error by ensemble $\mathbf{X}$ of N states (use ensemble perturbation matrix $\mathbf{X}' = \mathbf{X} - \bar{\mathbf{X}}$)
- **Forecast**:
  - Integrate ensemble with numerical model
- **Analysis step:**
  - update ensemble mean $\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}'^f \tilde{\mathbf{w}}$
  - update ensemble perturbations $\mathbf{X}'^a = \mathbf{X}'^f \mathbf{W}$

  (both can be combined in a single step)

- Ensemble Kalman & nonlinear filters: Different definitions of
  - weight vector $\tilde{\mathbf{w}}$ (dimension N)
  - Transform matrix $\mathbf{W}$ (dimension $N \times N$)

The Kalman Filter

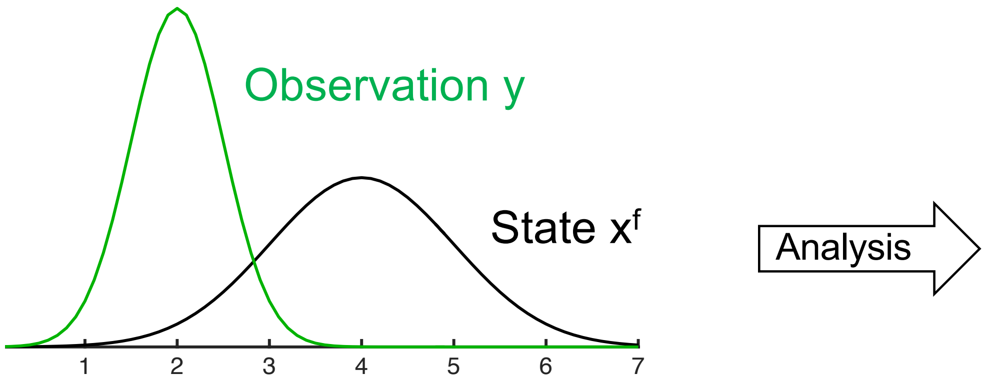
Assume Gaussian distributions
fully described by

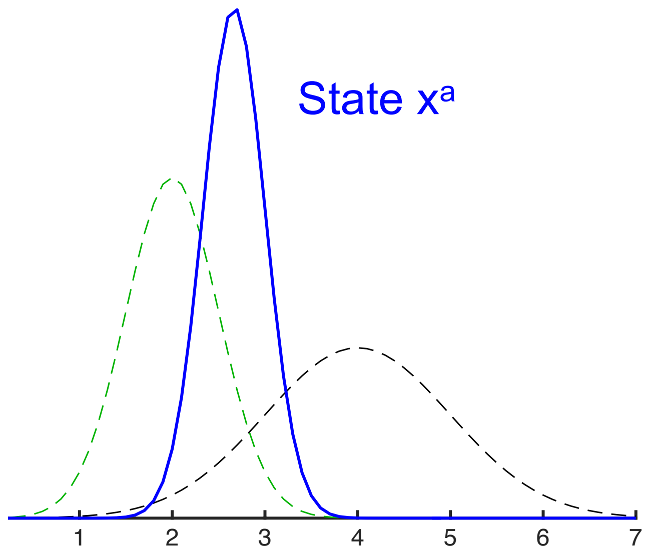
- mean state estimate
- covariance matrix

→ Strong simplification of estimation problem

Analysis is combination auf two Gaussian distributions computed as

- Correction of state estimate
- Update of covariance matrix





ETKF (Bishop et al., 2001)

- Ensemble Transform Kalman filter
  - Assume Gaussian distributions
  - Transform matrix

$$\mathbf{A}^{-1} = (N-1)\mathbf{I} + (\mathbf{H}\mathbf{X}'^f)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}'^f$$

  - Mean update weight vector

$$\tilde{\mathbf{w}} = \mathbf{A}(\mathbf{H}\mathbf{X}'^f)^T\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}^f}\right)$$

(depends linearly on observation vector $\mathbf{y}$)

  - Transformation of ensemble perturbations

$$\mathbf{W} = \sqrt{N-1}\ \mathbf{A}^{1/2}\mathbf{\Lambda}$$

$\mathbf{\Lambda}$: mean-preserving random matrix or identity

Note: $\mathbf{W}$ depends only on $\mathbf{R}$, not observation $\mathbf{y}$

Optimality of the Kalman Filter

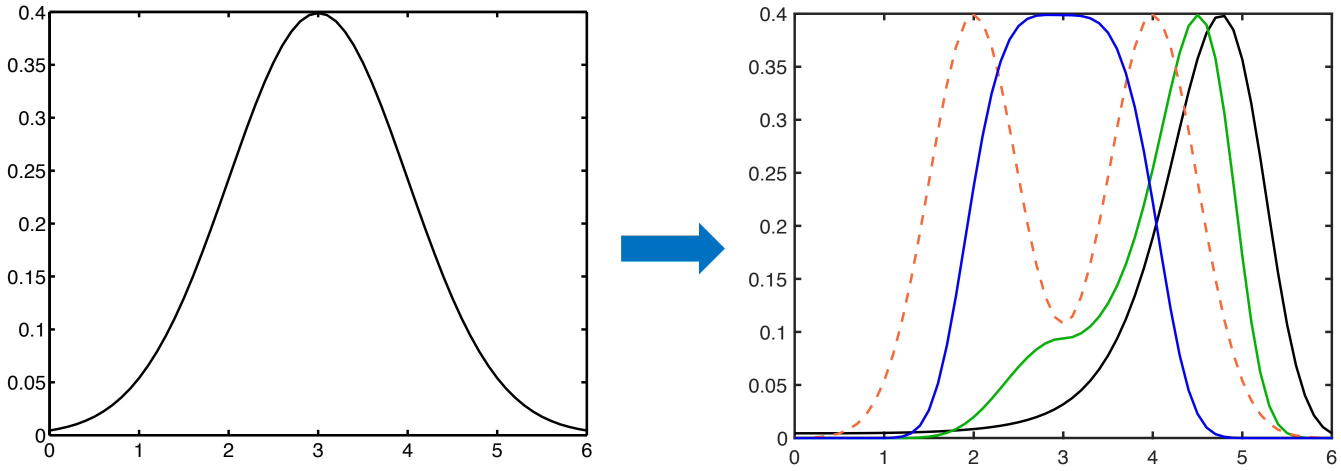
Kalman filter was derived to minimize variance

Kalman filter is optimal only if

- Covariance matrices are known
  (they are not in high-dimensional systems)
- Errors have normal distribution

With a nonlinear model

- Initial Gaussianity not preserved by nonlinear transformation



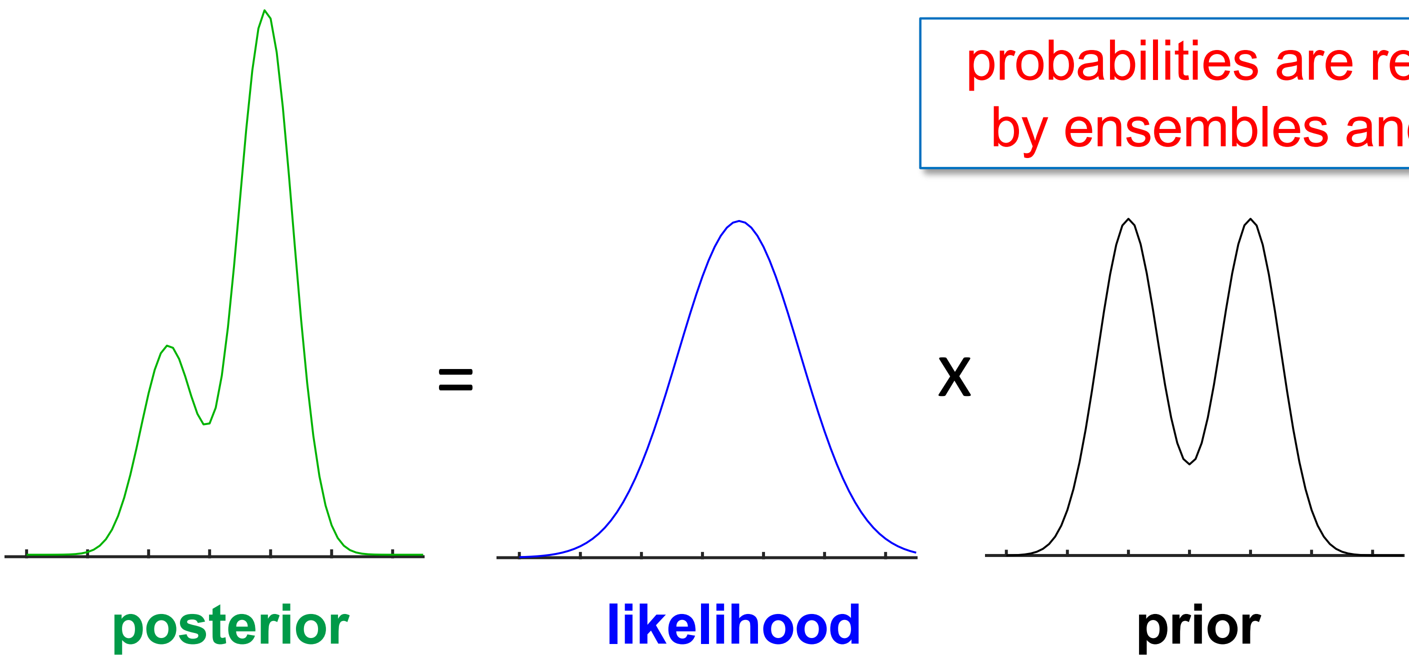
Particle filters: fully nonlinear

Just multiply probabilities:

$$p(x|y) \propto p(y|x)p(x)$$

… and normalize

probabilities are represented by ensembles and weights



Alternative uses of Bayes law

Bayes law: Probability density of $\mathbf{x}$ given $\mathbf{y}$

$$p\left(\mathbf{x}_i|\mathbf{y}_i\right) = \frac{p\left(\mathbf{y}_i|\mathbf{x}_i\right) p\left(\mathbf{x}_i\right)}{p\left(\mathbf{y}_i\right)}$$

Represent $p(\mathbf{x}_i)$ by ensemble: $p(\mathbf{x}_i) = \frac{1}{N}\sum_{j=1}^{N}\delta(\mathbf{x}_i - \mathbf{x}_i^{(j)})$

$$p(\mathbf{x}_i|\mathbf{y}_i) = \sum_{j=1}^{N}\delta(\mathbf{x}_i - \mathbf{x}_i^{(j)})\,\frac{p(\mathbf{y}_i|\mathbf{x}_i^{(j)})}{p(\mathbf{y}_i)}$$

Kalman filter:
assume normal distributions
compute new ensemble states

$$\mathbf{x}_i^{a(j)}; j = 1, \ldots, N$$

Particle Filter:
keep ensemble states
but assign weights

$$w^{(j)} = \frac{p(\mathbf{y}_i|\mathbf{x}_i^{(j)})}{p(\mathbf{y}_i)}$$

NETF (Tödter & Ahrens, 2015)

- Nonlinear Ensemble Transform Filter
  - Mean update from Particle Filter weights:
    for Gaussian observation errors for all particles *i*

$$\tilde{w}^i \sim \exp\left(-0.5(\mathbf{y}-\mathbf{H}\mathbf{x}_i^f)^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x}_i^f)\right)$$

(nonlinear function of observations $\mathbf{y}$)

  - Ensemble update
    - Transform ensemble to fulfill analysis covariance
      (like ETKF, but not assuming Gaussianity)
    - Derivation gives

$$\mathbf{W} = \sqrt{N}\left[\mathrm{diag}(\tilde{\mathbf{w}}) - \tilde{\mathbf{w}}\tilde{\mathbf{w}}^T\right]^{1/2}\Lambda$$

($\boldsymbol{\Lambda}$: mean-preserving random matrix; useful for stability)

NETF is a second-order exact particle filter

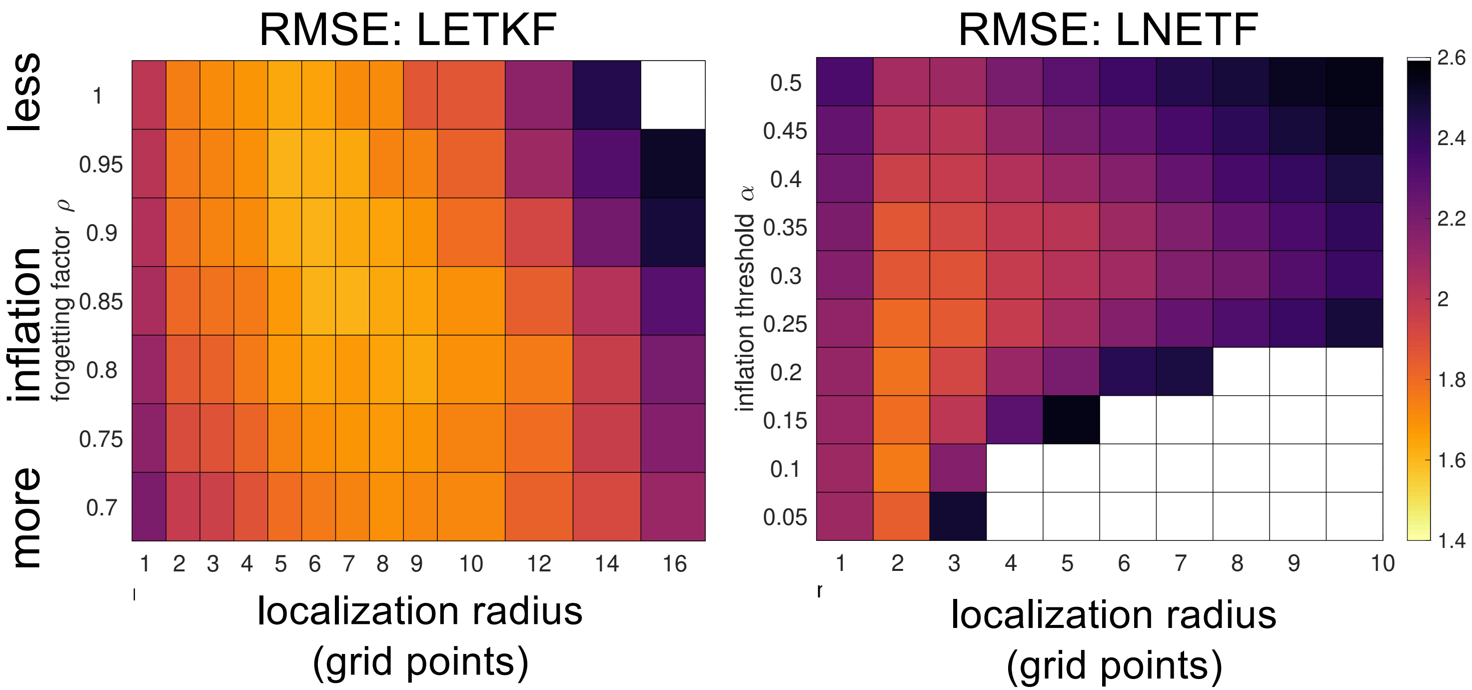
Tödter, J. and Ahrens, B. (2015) *Mon. Wea. Rev.* **143**,1347–1367

ETKF & NETF

- Analogous update schemes
- Both filters can be localized (LETKF and LNETF)

- But
  - ETKF
    - very stable, even in nonlinear cases
    - Optimal for Gaussian / sub-optimal for nonlinear cases
  - NETF
    - accounts for nonlinearity (non-Gaussianity)
    - higher sampling errors than LETKF
    - needs very small localization radii

Test with Lorenz-96 model

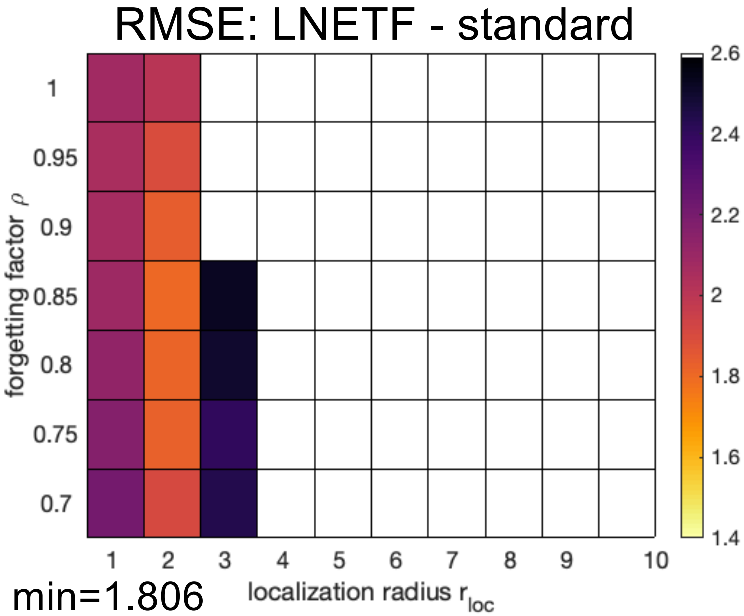


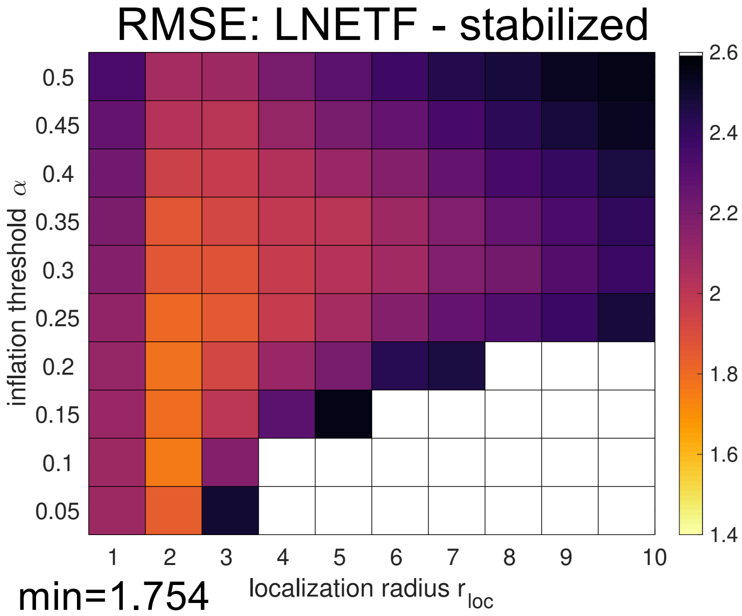
RMSE min=1.606

RMSE min=1.754

- State dimension 40
- Ensemble size N_{ens}=15
- Forecast: 8 time steps (strongly nonlinear DA case)
- 20 observations, obs. error 1.0
- LETKF better than LNETF at this N_{ens}

Stabilizing LNETF – Lorenz-96 model





- State dimension 40
- Ensemble size 15
- Forecast: 8 time steps
- 20 observations

- Standard LNETF needs very small localization radius ('curse of dimensionality' also holds with localization)

Inflation:
- forgetting factor

Inflation:
- fixed inflation (ρ=0.85)
- Minimum effective sample size

$$N_{eff} = \sum \frac{1}{(w^i)^2} > \alpha$$

Hybrid – LETKF-LNETF

Combine the stability of LETKF with nonlinear features of LNETF

ETKF-NETF – Hybrid Filter Variants

Factorize the likelihood: ('tempering') $p(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}|\mathbf{x})^{\gamma} p(\mathbf{y}|\mathbf{x})^{(1-\gamma)}$

1-step update (*HSync*)

$$\mathbf{X}^a_{HSync} = \overline{\mathbf{X}}^f + (1-\gamma)\Delta\mathbf{X}_{NETF} + \gamma\Delta\mathbf{X}_{ETKF}$$

- $\Delta\mathbf{X}$: assimilation increment of a filter
- γ: hybrid weight (between 0 and 1; 1 for fully ETKF)

2-step updates

Variant 1 (*HNK*): NETF followed by ETKF

$$\tilde{\mathbf{X}}^a_{HNK} = \mathbf{X}^a_{NETF}[\mathbf{X}^f, (1-\gamma)\mathbf{R}^{-1}]$$

$$\mathbf{X}^a_{HNK} = \mathbf{X}^a_{ETKF}[\tilde{\mathbf{X}}^a_{HNK}, \gamma\mathbf{R}^{-1}]$$

- Both steps computed with increased $\mathbf{R}$ according to γ

Variant 2 (*HKN*): ETKF followed by NETF

Choosing hybrid weight γ

- Hybrid weight shifts filter behavior

Some possibilities:

- Fixed value
- Adaptive - According to which condition?

  - Frei & Künsch (2013) suggested using effective sample size $N_{eff} = \sum \frac{1}{(w^i)^2}$ (Usual choice for 'tempering')
    - γ_α : Choose γ so that N_{eff} is as small as possible but above minimum limit α (done iteratively)
  - Adaptive alternative

$$\gamma_{lin} = 1 - \frac{N_{eff}}{N_e}$$

(close to 1 if N_{eff} small; no iterations)

Issue: Using N_{eff}
- only ensures non-collapsing ensemble
- does not ensure good analysis result

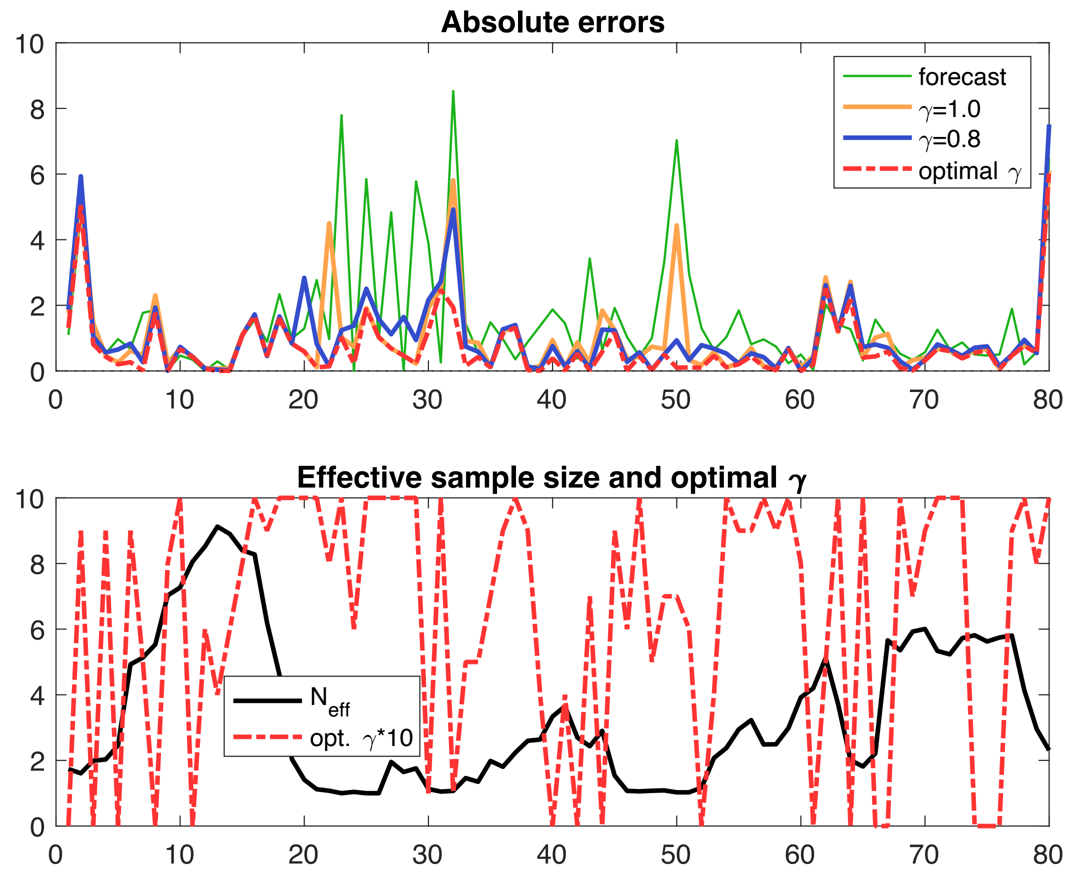
Effect of hybrid weight γ

- Lorenz-96 model, size 80
- Examine single analysis step

1. Run 33 analysis steps with γ=1 (LETKF)
2. Run analysis step 34 with one of
   a) γ=1
   b) γ=0.8
3. Examine N_{eff} and analysis errors

Additional experiment:

c) Adjust γ at each grid point to get minimum error



No obvious relation between N_{eff} and γ!

Other works using hybrid filters

Frei & Künsch, 2013 (*FK13*)

- Derived combined KF with resampling PF; KF before PF (with covariance localization)
- Hybrid weight from $N_{eff}/N_e \geq a$

Chustagulprom/Reich/Reinhardt, 2016

- Ensemble square-root filter (ESRF) + ensemble transform PF solving linear transport problem
- KF before PF, PF before KF (better results if PF before KF); hybrid weight from *FK13*

Robert/Leuenberger/Künsch, 2018

- Update of FK13 using ETKF; application to convective-scale setup of COSMO
- KF before PF; hybrid weight from *FK13*

Grooms & Robinson, 2021

- EnSRF + SIR-PF without localization; PF before KF; hybrid weight from *FK13*

Poterjoy, 2022; Kurosawa & Poterjoy, 2023

- Multi-step (tempered) Local PF + final EnKF step
- Poterjoy(2022): Hybrid weight using *FK13*; Kurosawa/Poterjoy(2023): Shapiro-Wilk test

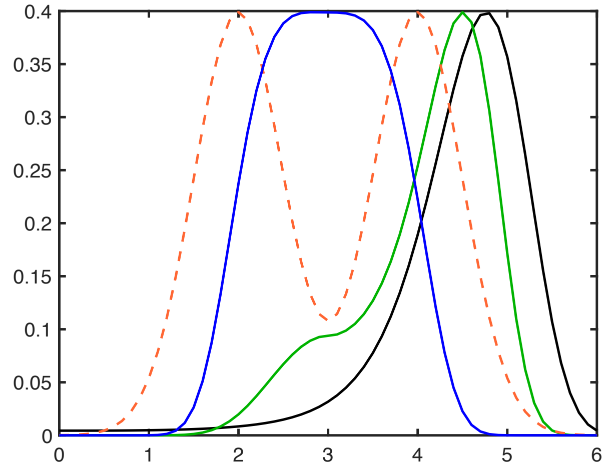
Account for non-Gaussianity: Skewness and Kurtosis

- **Mean** – 1st moment
- **Variance** – 2nd moment
- **Skewness** – 3rd moment

$$skew = \frac{\frac{1}{N_e}\sum_{i=1}^{N_e}\left(\mathbf{x}^i - \overline{\mathbf{x}}\right)^3}{\left[\frac{1}{(N_e-1)}\sum_{i=1}^{N_e}\left(\mathbf{x}^i - \overline{\mathbf{x}}\right)^2\right]^{3/2}}$$

- **Kurtosis** – 4th moment

$$kurt = \frac{\frac{1}{N_e}\sum_{i=1}^{N_e}\left(\mathbf{x}^i - \overline{\mathbf{x}}\right)^4}{\left[\frac{1}{(N_e)}\sum_{i=1}^{N_e}\left(\mathbf{x}^i - \overline{\mathbf{x}}\right)^2\right]^{2}} - 3$$

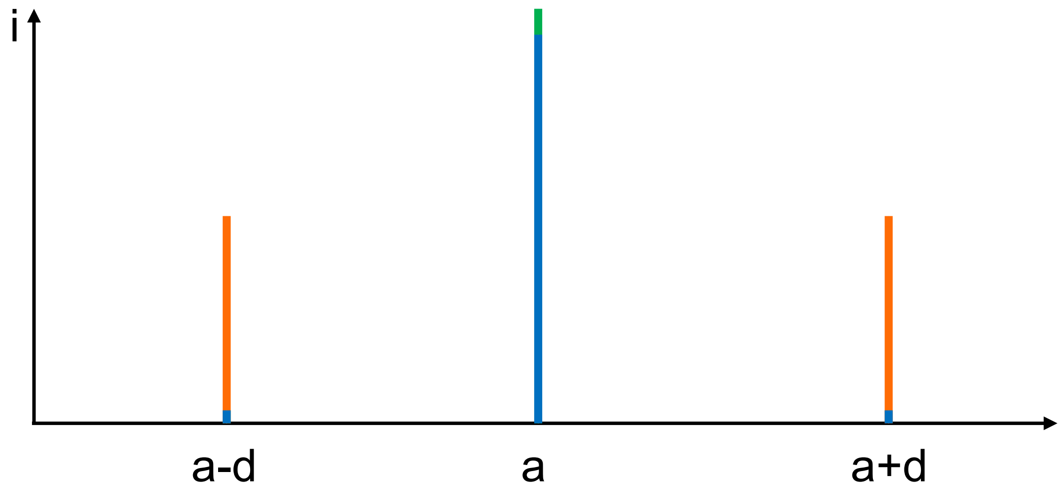


- Skewness and kurtosis
  - generally not bounded
    → but limits depend on ensemble size

Finding bounds: Asymptotic properties of skewness and kurtosis

- Bounds of skewness and kurtosis depend on ensemble size
- Assess extreme cases

Case	Values	*skew* limit	*kurt* limit
max. skew	$x^{(1)} = a - d, \quad x^{(i)} = a,\ i = 2, \ldots, N_e$	$\sqrt{N_e}$	N_e
max. kurt	$x^{(1)} = a - d, \quad x^{(2)} = a + d, \quad x^{(i)} = a,\ i = 3, \ldots, N_e$	0	-2
min. kurt	$x^{(i)} = a - d,\ i = 1, \ldots N_e/2; \quad x^{(j)} = a + d,\ j = N_e/2 + 1, \ldots, N_e$	0	$N_e/2$



Using skewness and kurtosis to define hybrid weight γ

- Sampling errors are larger in NETF than ETKF
  → Always use ETKF for Gaussian (linear) cases
- Skewness and kurtosis describe deviation from Gaussianity
- mean absolute skewness (mas) and kurtosis (mak) of observed ensemble (with localization: use locally assimilated observations)
- Use normalized means:

$$nmas = \frac{1}{\sqrt{\kappa}} mas \qquad nmak = \frac{1}{\kappa} mak$$

standard value:
$\kappa = N_e$

Now define

$$\gamma_{sk,\alpha} = \max\left[\min(1 - nmak,\ 1 - nmas),\ \gamma_\alpha\right]$$
$$\gamma_{sk,lin} = \max\left[\min(1 - nmak,\ 1 - nmas),\ \gamma_{lin}\right]$$

stronger influence of $nmas$ and $nmak$ limited by N_{eff}

Note: There are sampling errors, e.g. for skewness $\sigma_{skew} \sim \sqrt{6/N_e}$
→ For N_e=25: ~10% error in γ

Numerical Experiments

PDAF – Parallel Data Assimilation Framework

PDAF Parallel Data Assimilation Framework

A unified tool for interdisciplinary data assimilation …

- provide support for parallel ensemble forecasts
- provide DA methods (EnKFs, smoothers, PFs, 3D-Var) - fully-implemented & parallelized
- provide tools for observation handling and for diagnostics
- easy implementation with (probably) any numerical model (<1 month)
- a program library (PDAF-core) plus additional functions & templates
- run from notebooks to supercomputers (Fortran, MPI & OpenMP – model compatibility)
- ensure separation of concerns (model – DA method – observations – covariances)
- first release in year 2004; continuous further development

Open source:
Code, documentation, and tutorial available at

https://pdaf.awi.de

https://github.com/PDAF/PDAF

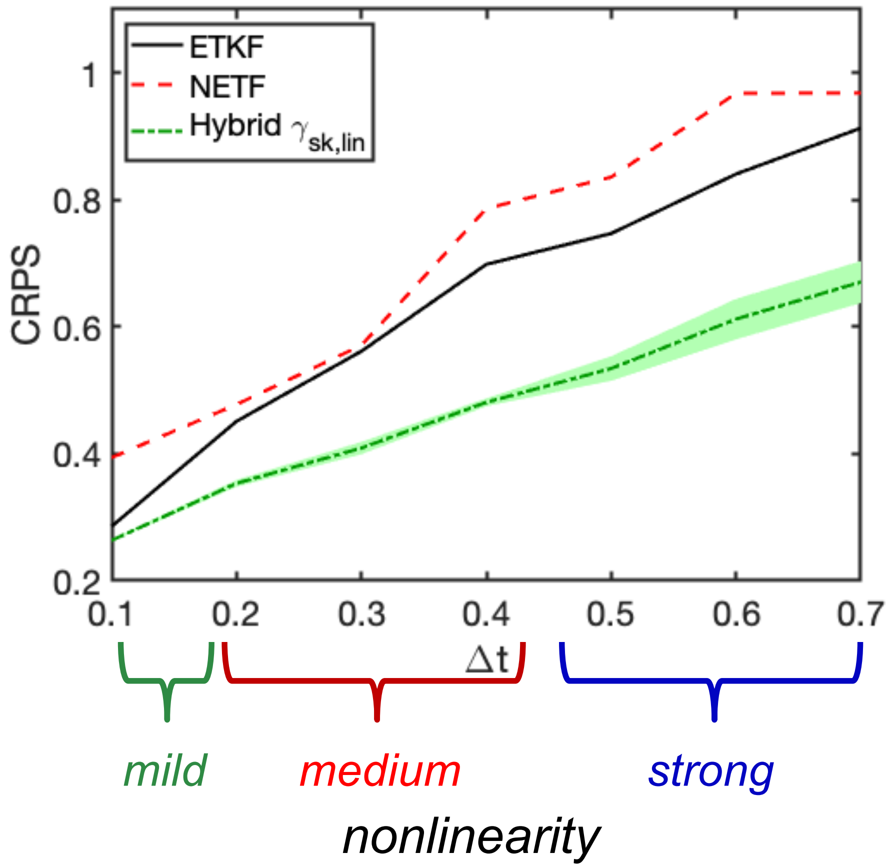
L. Nerger, W. Hiller, Computers & Geosciences 55 (2013) 110-118

AWI

Assimilation with Lorenz-63 model

- Observe full state
- Time step size 0.05
- Vary forecast duration *Δt* to vary nonlinearity
- Ensemble size N_e =25
- HNK filter variant (nonlinear before Kalman)
- Implemented with PDAF

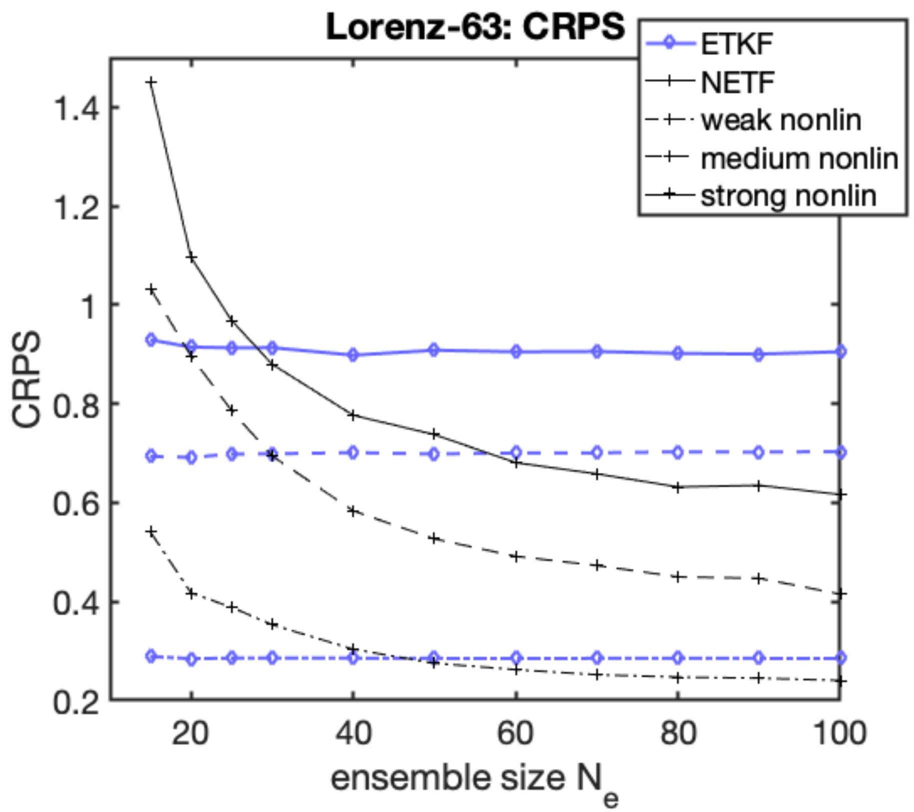
- Error of NETF > ETKF due to sampling errors
- Effect of hybrid filter grows with nonlinearity of assimilation problem (forecast length)
- Hybrid weight $\gamma_{sk,lin}$ yield smallest errors without any tuning
  → Errors are reduced up to 28%
- Note: Hybrid weight $\gamma_{sk,\alpha}$ is suboptimal unless optimally tuned



EKTF & NETF with Lorenz-63 model

Dependence on ensemble size

- NETF yields smaller errors than ETKF if ensemble size large enough
  - → Size limit decreases for larger nonlinearity
  - → Improvement by NETF stronger for higher nonlinearity

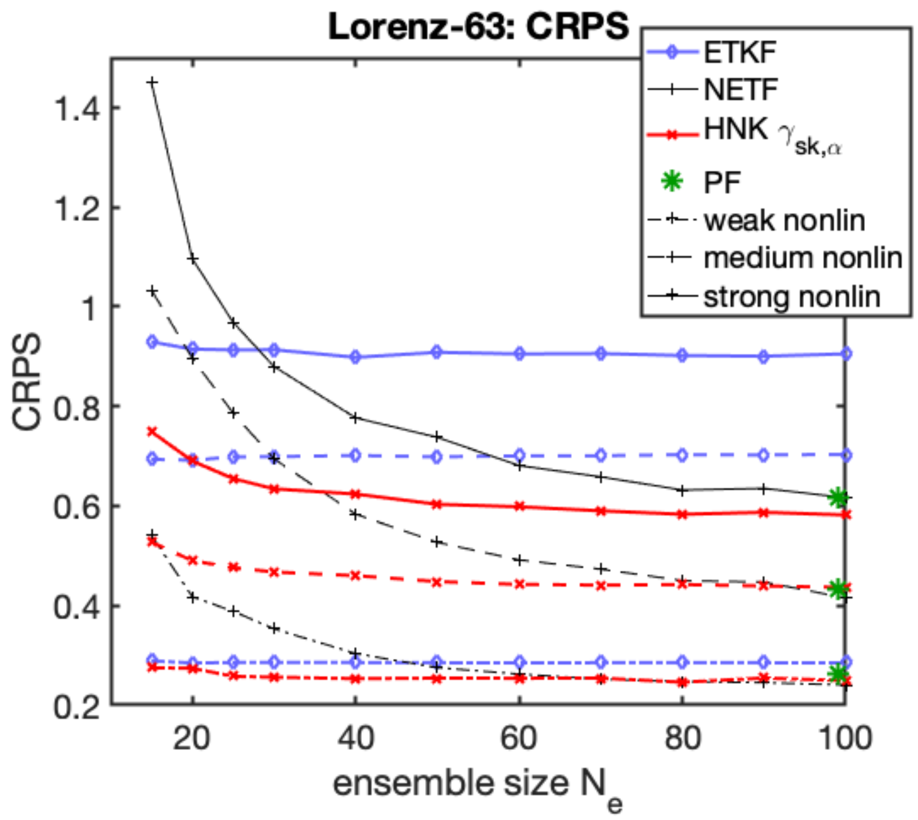


ETKF and NETF for
3 different nonlinearities

(weak Δt=0.1, medium Δt=0.4, strong Δt=0.7)

Assimilation with Lorenz-63 model

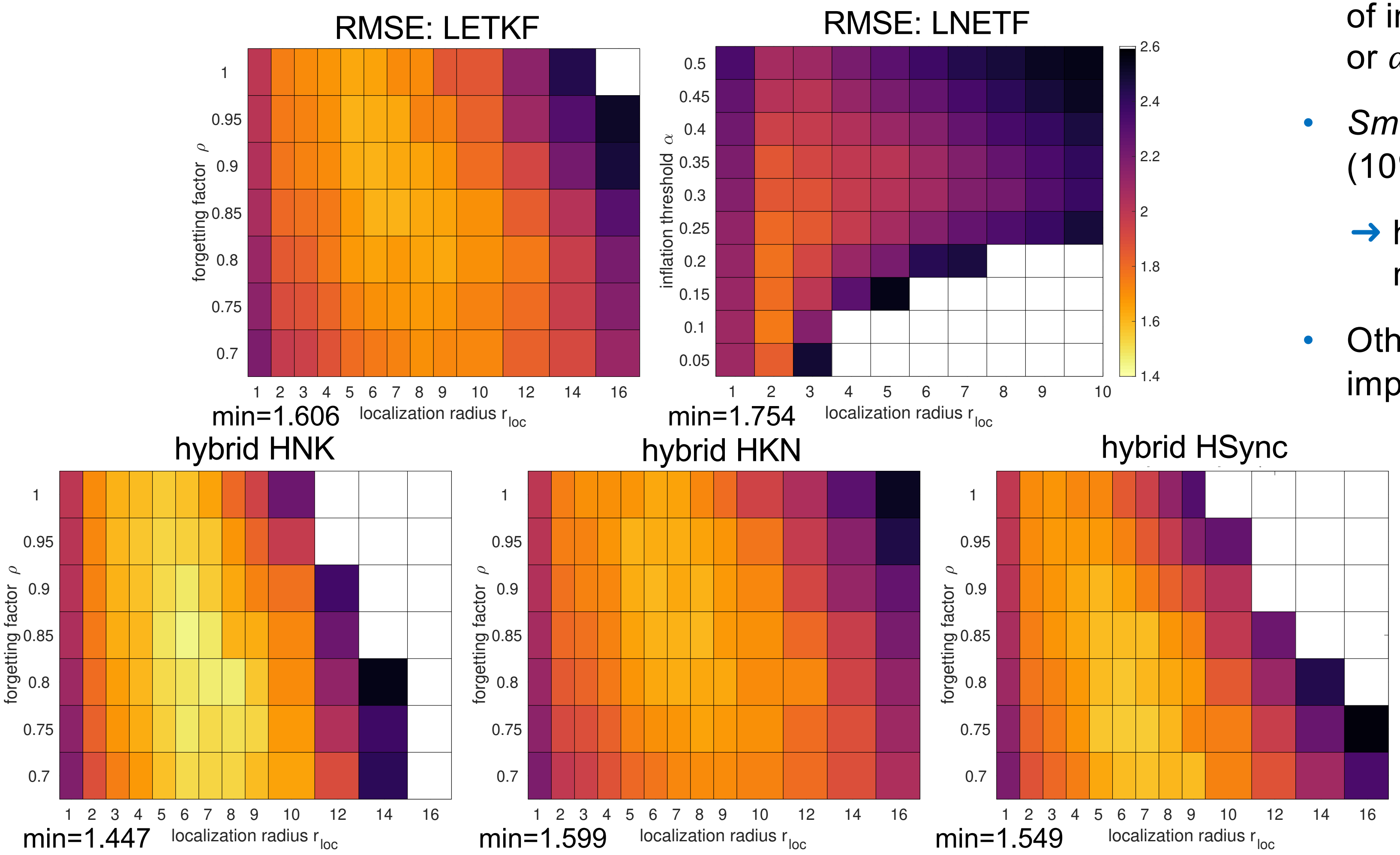
- Hybrid filter HNK
  - particular strong effect for small N_e
  - CRPS from NETF and HNK converge for large N_e
  - errors reduced up to 28%
- Particle Filter
  - comparable CRPS for large N_e
    - PF expected to be superior if N_e sufficiently large (the full nonlinear filter)



- Note: Easy to use large ensemble for Lorenz-63, difficult for higher dimensional models

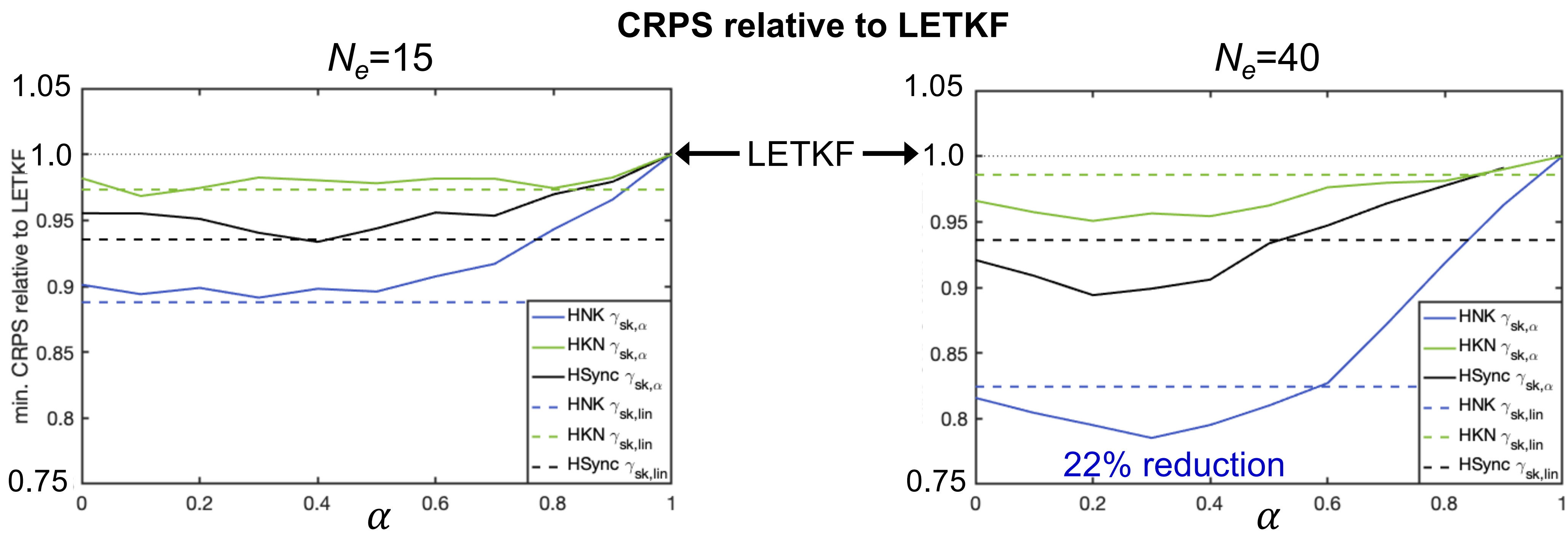
Test with Lorenz-96 model

Ensemble size 15; Forecast length: 8 time steps; 20 observations



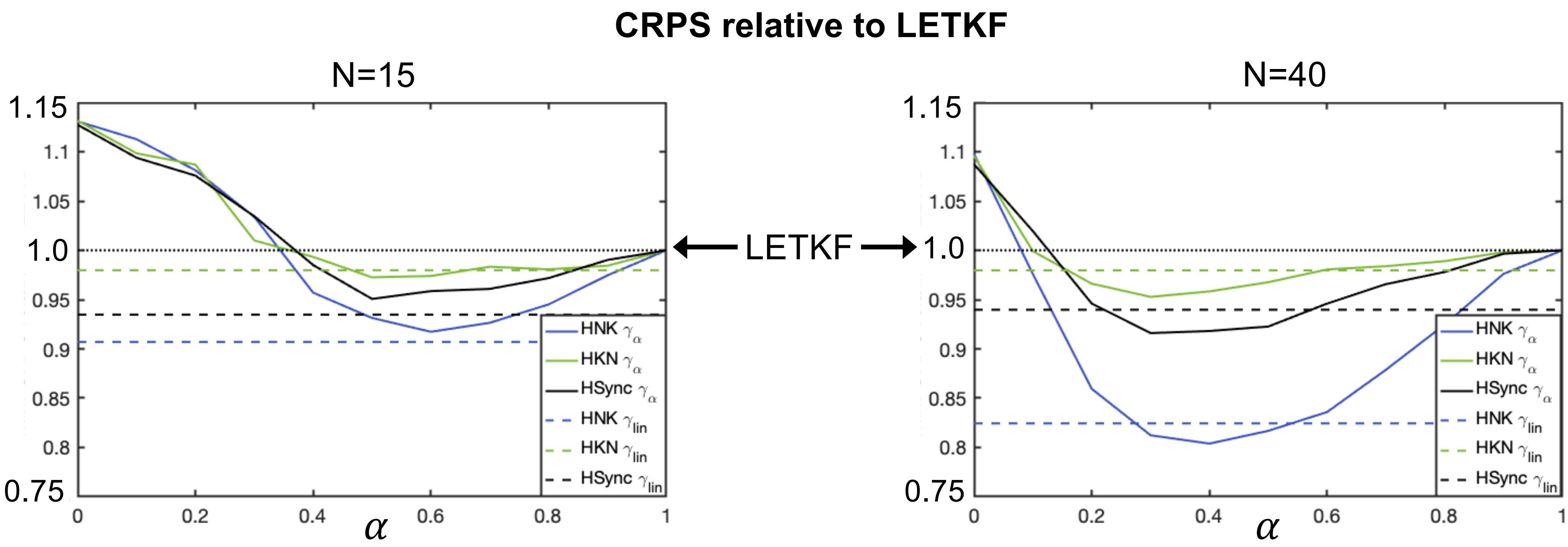
- Show RMS errors as function of inflation (forgetting factor or α) and localization radius
- *Smallest errors:* Hybrid HNK (10% error reduction)
  - → hybrid filter able to utlize non-Gaussian information
- Other hybrid variants also improve the state estimate

Lorenz-96: Influence of γ – using skewness and kurtosis



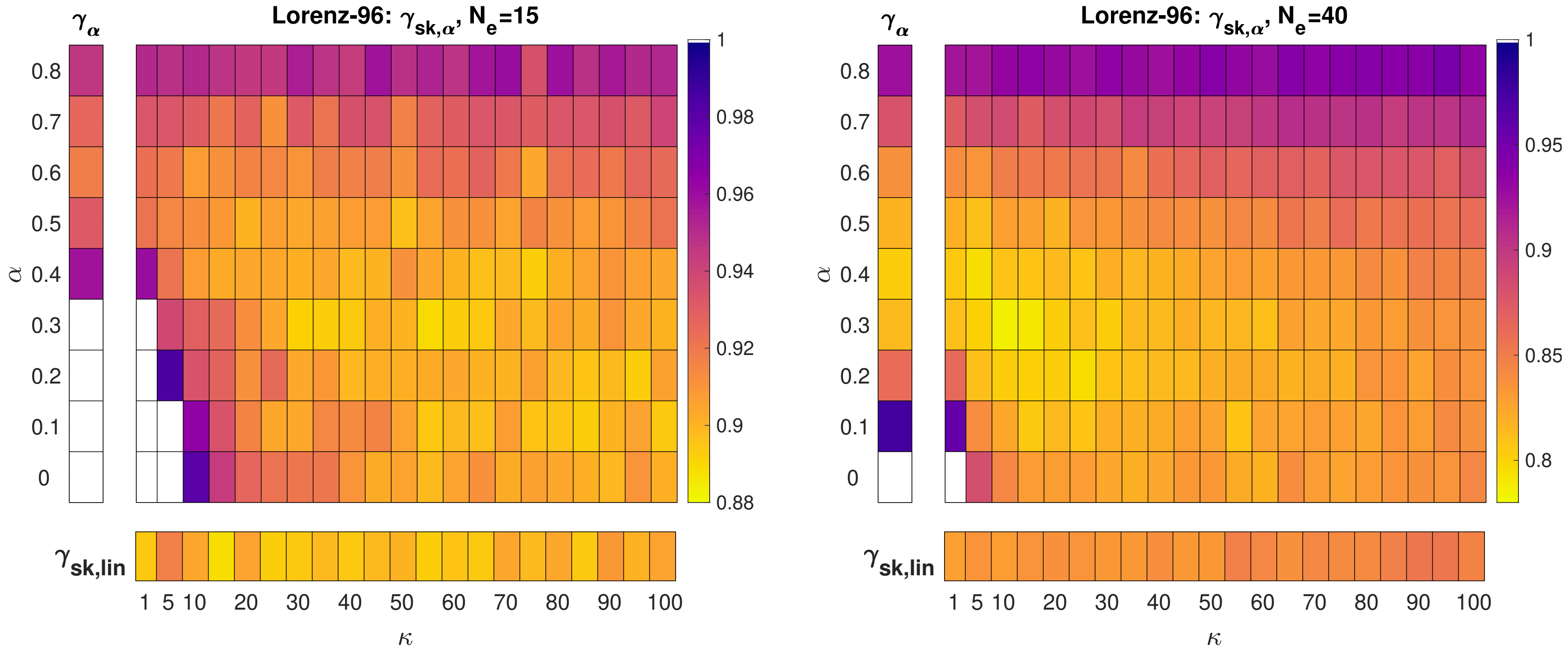
- When accounting for skewness/kurtosis filter is more stable
- $\gamma_{sk,lin}$ yields smallest (N_e=15) or nearly smallest (N_e=40) errors
- smallest errors with $\gamma_{sk,\alpha}$ for optimal tuning

Lorenz-96: Influence of γ – cases γ_α and γ_{lin} (account only for N_{eff})



- Stronger effect of hybrid filter for N_e=40
- γ_{lin} yields optimal (N=15) or nearly optimal (N=40) errors
- γ_α requires tuning; increased errors for small α compared to γ_{sk}

Lorenz-96, Hybrid HNK, dependence on κ



CRPS relative to LETKF

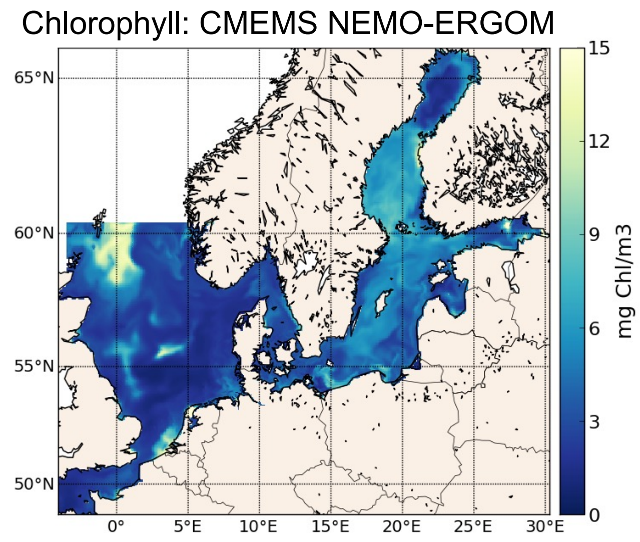
κ can be chosen dependent on ensemble size

- Limits of skewness and kurtosis depend on N_e
  - but actual skewness and kurtosis do not depend on system, not on N_e
- Standard value $=N_e$, but smaller large large N_e

Application example

PDAF Parallel Data Assimilation Framework

- Ocean-biogeochemical model:
    - NEMO + ERGOM
- Configuration: NORDIC 2.0
    - 1.8km resolution, 56 layers, 90s time step
    - North Sea & Baltic Sea
    - Operational use in CMEMS for the Baltic Sea
- DA implementation
    - augment NEMO-ERGOM with DA functionality by PDAF (online-coupling in memory)
    - State vector:
        - physics + biogeochemistry
          State vector size ~153 million
    - Assimilate satellite chlorophyll data



ocean and biogeochemical dynamics are nonlinear and distributions non-Gaussian

SEAMLESS

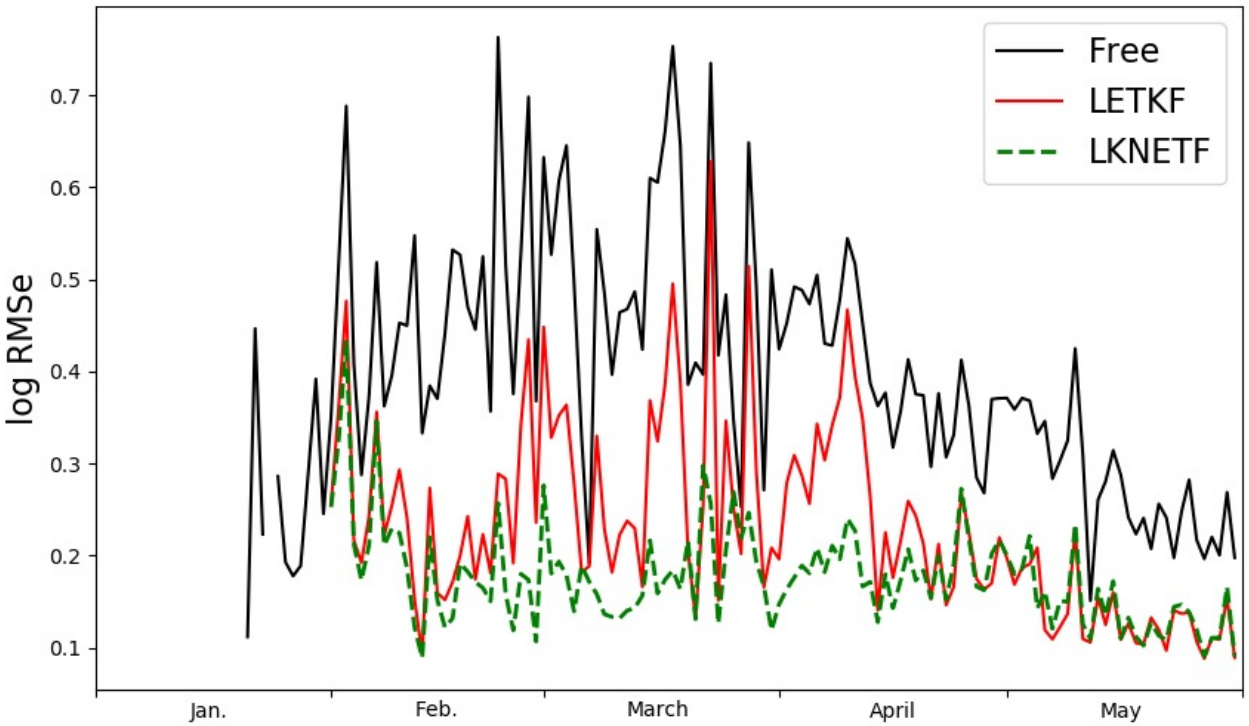
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776480

AWI

Effect of hybrid filter in high-dimensional application

Assimilation using rule $\gamma_{sk,\alpha}$

RMS deviation for log-Chlorophyll



Regional model setup

Only assimilate chlorophyll observations

Stronger assimilation effect of LKNETF

We still don't know optimal choice of rule for γ

Summary

PDAF Parallel Data Assimilation Framework

Introduced hybrid nonlinear-Kalman ensemble transform filter

- Combine LETKF and LNETF methods
- hybrid weight γ shifts filter behavior
- Cost of analysis step ~2x LETKF

Experiments with Lorenz models

- Hybrid filter successfully reduces errors compared to LETKF and LNETF
- Best results for variant HNK: LNETF applied before LETKF
- Can compute γ from skewness and kurtosis
  - → allows to control nonlinearity of filter based on non-Gaussianity
  - → Improved stability & reduced errors compared to tempering rule on N_{eff}

Nerger, Q. J. Meteorol. Soc., 148 (2022) 620-640, doi:10.1002/qj.4221

AWI

Next steps

PDAF Parallel Data Assimilation Framework

Need to

- improve understanding of effect of γ
  - mathematical basis
  - Are skewness & kurtosis good choices?
  - Is linear dependence of skewness & kurtosis right?

- asses for which nonlinear cases hybrid filter is superior
  - only 3% lower errors in test with ocean physics at 0.25° resolution

Nerger, Q. J. Meteorol. Soc., 148 (2022) 620-640, doi:10.1002/qj.4221

AWI