

A hybrid nonlinear-Kalman ensemble transform filter for data assimilation in systems with different degrees of nonlinearity

Lars Nerger

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

EnKF-Workshop, Norheimsund, May 2023



Overview

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

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
$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{X}'^f \tilde{\mathbf{w}}$$
 - 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 $N \times N$)

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

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

(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 $\mathbf{y}$

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

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

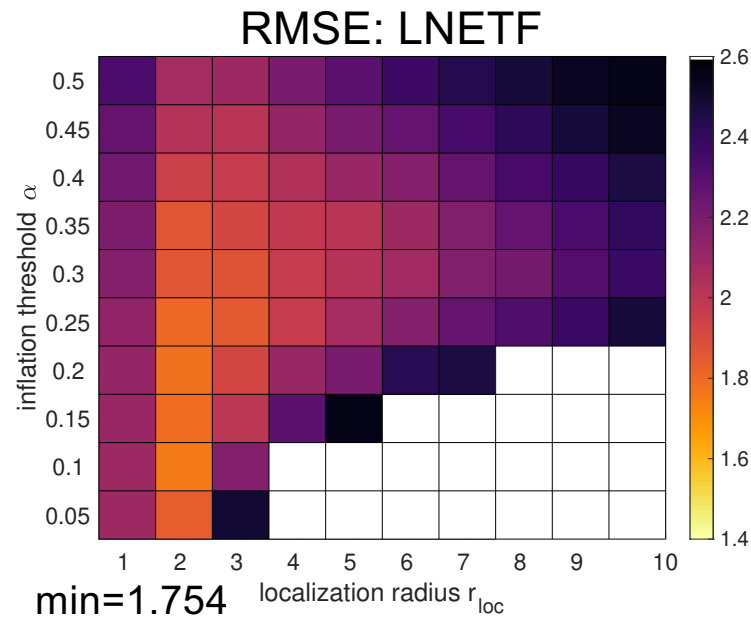
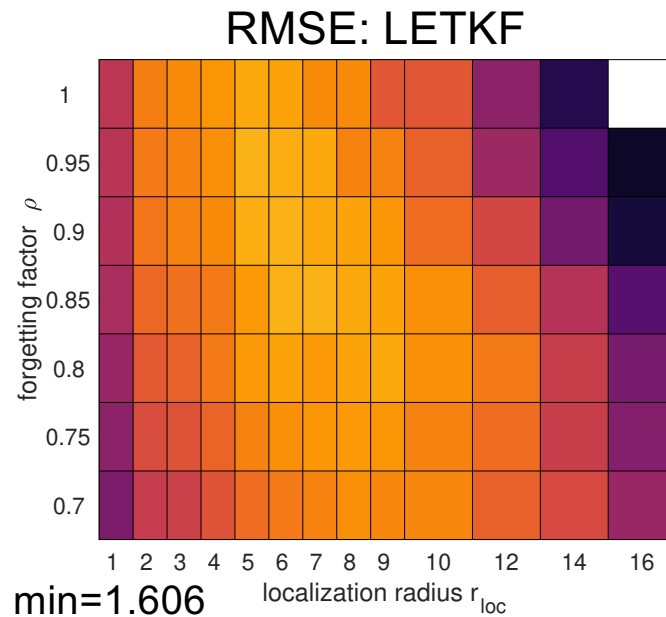
(Λ : mean-preserving random matrix; useful for stability)

NETF is a second-order exact particle filter

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



- State dimension 40
- Ensemble size 15
- Forecast: 8 time steps
- 20 observations
- Show RMS errors as function of inflation (forgetting factor or inflation threshold α) and localization radius
- LNETF worse than LETKF

Hybrid – LETKF-LNETF

Combine the stability of LETKF with nonlinear features of LNETF

ETKF-NETF – Hybrid Filter Variants

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

1-step update (*HSync*)

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

- $\Delta\mathbf{X}$: 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

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

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

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

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

Lars NERGER – hybrid nonlinear-Kalman filter

Choosing hybrid weight γ

- Hybrid weight shifts filter behavior

Some possibilities:

- Fixed value
- Adaptive - According to which condition?

- Frei & Kuensch (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

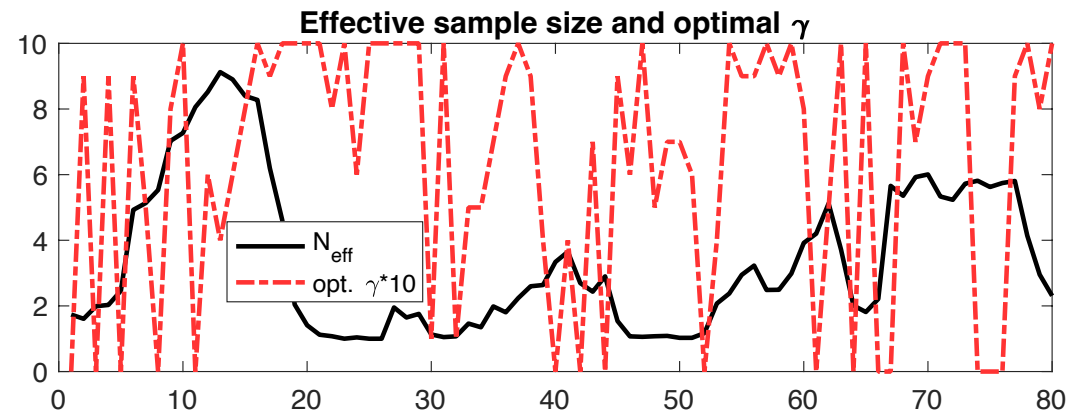
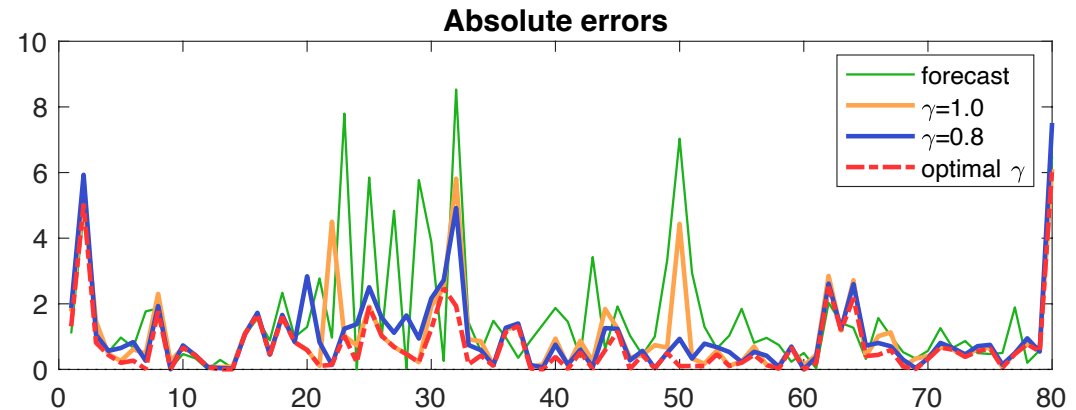
Effect of hybrid weight γ

- Lorenz-96 model, size 80
 - Examine single analysis step
1. Run 33 analysis steps with $\gamma=1$ (LETKF)
 2. Run analysis step 34 with one of
 - a) $\gamma=1$
 - b) $\gamma=0.8$
 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 !



Account for non-Gaussianity: Skewness and Kurtosis

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

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

- **Kurtosis** – 4th moment

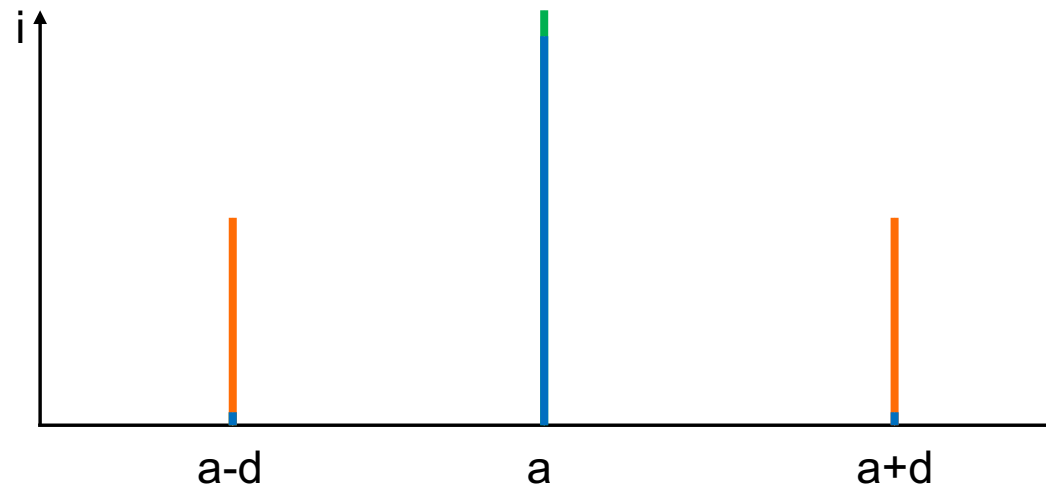
$$kurt = \frac{\frac{1}{N_e} \sum_{i=1}^{N_e} (\mathbf{x}^i - \bar{\mathbf{x}})^4}{\left[\frac{1}{(N_e)} \sum_{i=1}^{N_e} (\mathbf{x}^i - \bar{\mathbf{x}})^2 \right]^2} - 3$$

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

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



Lars Nerger – hybrid nonlinear-Kalman filter

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

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

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

A unified tool for interdisciplinary data assimilation ...

- provide support for parallel ensemble forecasts
- provide assimilation methods (solvers) - fully-implemented & parallelized
- provide tools for observation handling and for diagnostics
- easily useable with (probably) any numerical model
- a program library (PDAF-core) plus additional functions
- run from notebooks to supercomputers ([Fortran](#), [MPI](#) & [OpenMP](#))
- ensure separation of concerns (model – DA method – observations – covariances)

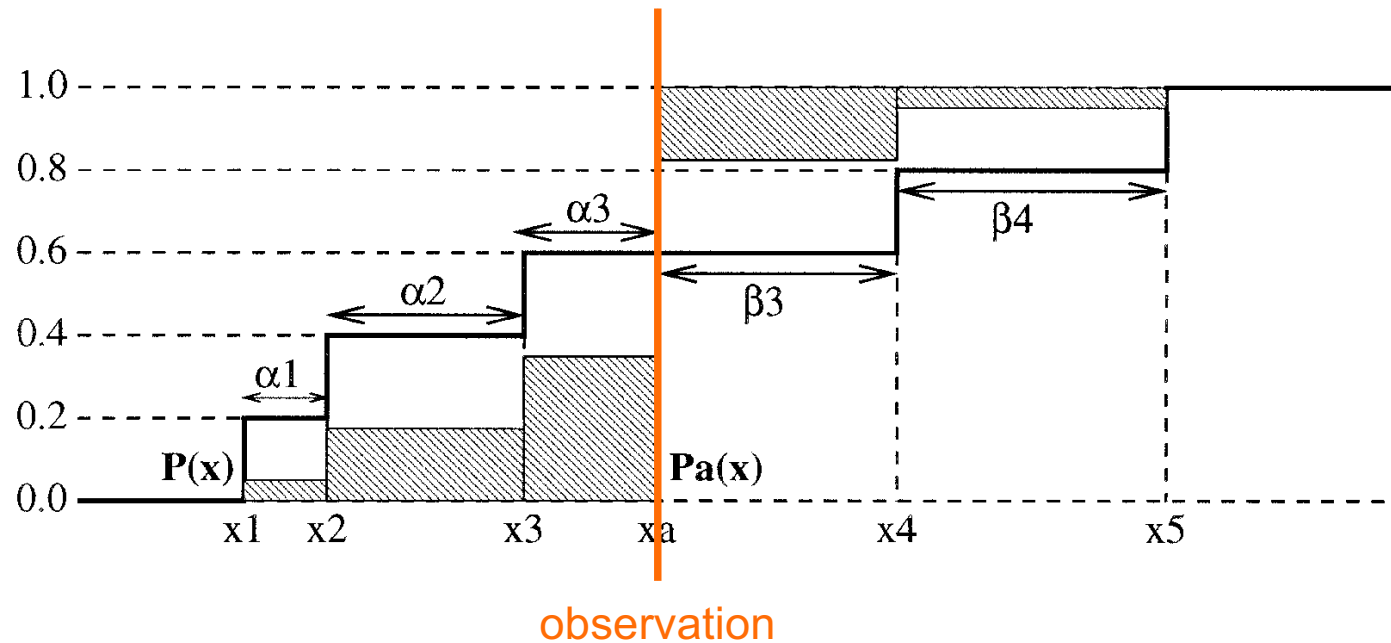
Code also at
github.com/PDAF

Open source:
Code, documentation, and tutorial available at
<http://pdaf.awi.de>



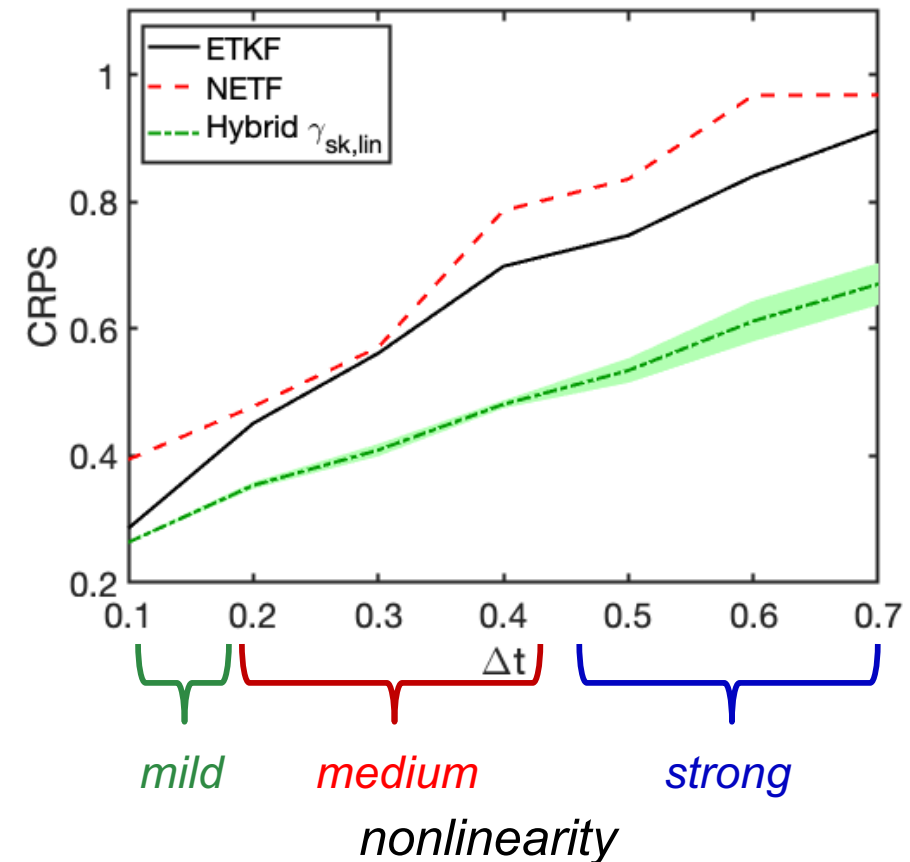
CRPS – Continuous Ranked Probability Score

- Measure deviation of all ensemble members from observation
- No assumption on ensemble distribution
- Example for 5 ensemble states
 - CRPS = shaded area



Assimilation with Lorenz-63 model

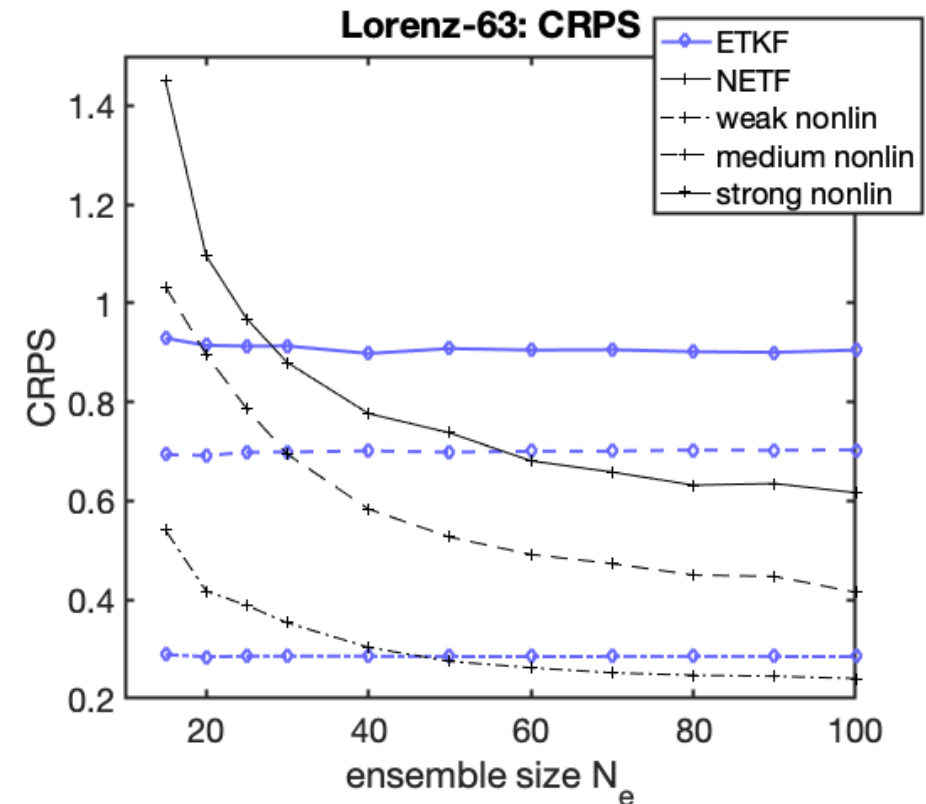
- Observe full state
 - Time step size 0.05
 - Vary forecast duration Δt to vary nonlinearity
 - Ensemble size $N=25$
 - HNK filter variant (nonlinear before Kalman)
 - Implemented with PDAF
-
- Error of NETF > ETKF due to sampling errors
 - Effect of hybrid filter grows with nonlinearity of assimilation 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

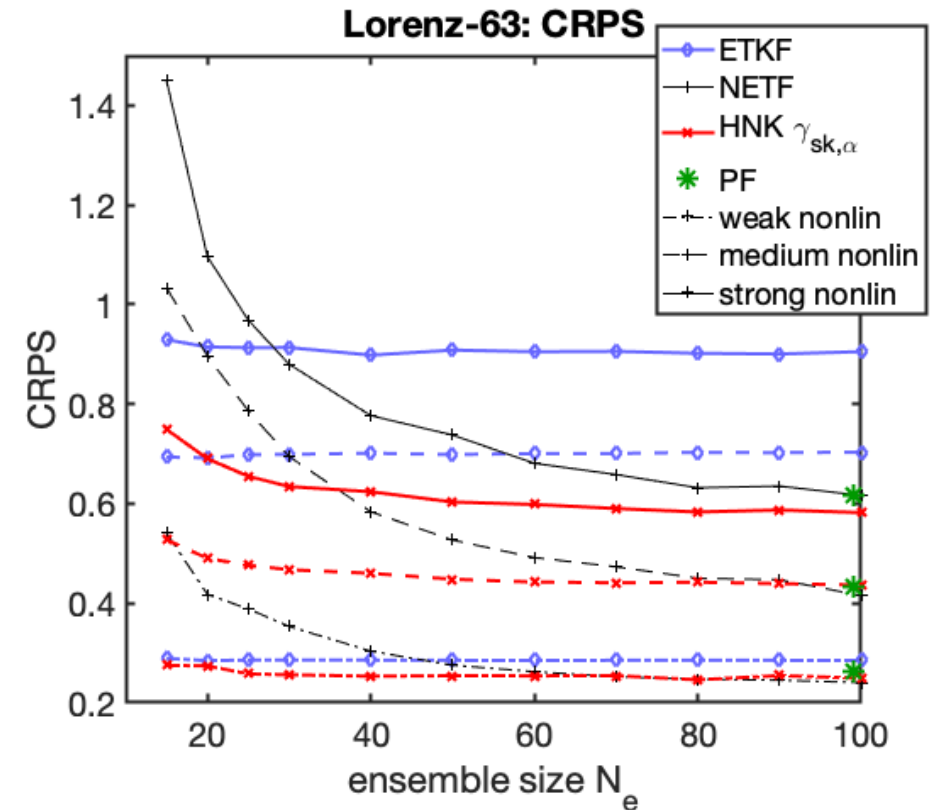


ETKF and NETF for
3 different nonlinearities

(weak $\Delta t=0.1$, medium $\Delta t=0.4$, strong $\Delta t=0.7$)

Assimilation with Lorenz-63 model

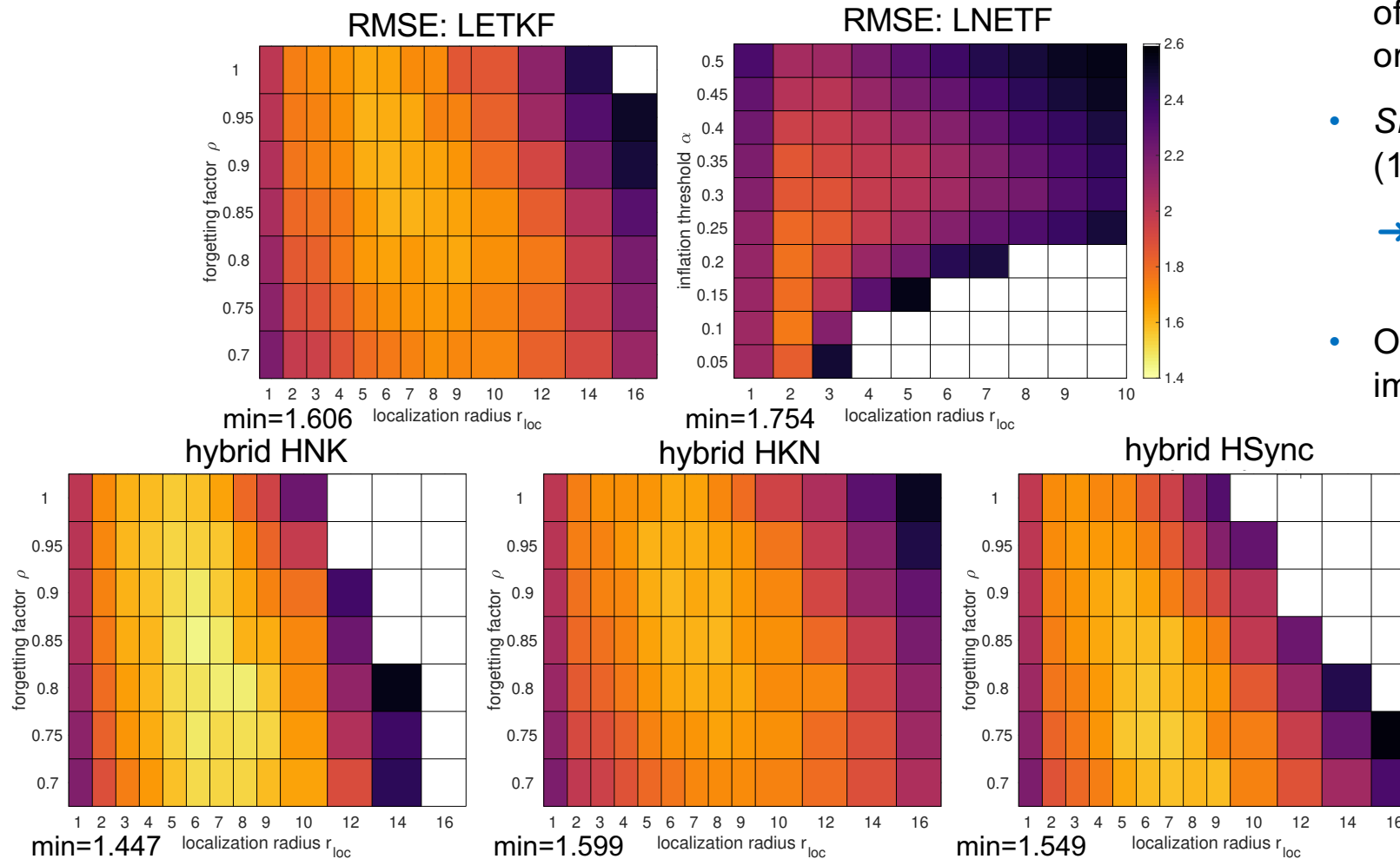
- Hybrid filter HNK
 - particular strong effect for small N_e
 - CRPS from NETF and HNK converge for large N_e
 - errors reduced up to 28%
- Particle Filter
 - comparable CRPS for large N_e
 - PF expected to be superior if N_e sufficiently large (the full nonlinear filter)
- Note: Easy to use large ensemble for Lorenz-63, difficult for higher dimensional models



Test with Lorenz-96 model

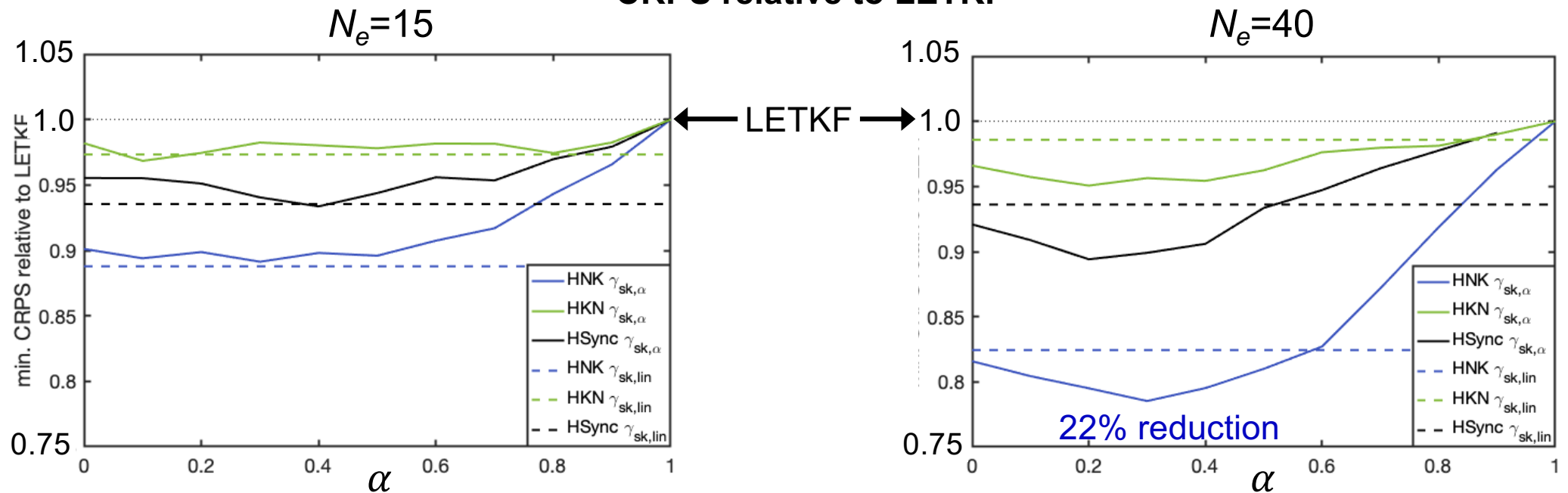
Ensemble size 15; Forecast length: 8 time steps; 20 observations

- 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 utilize non-Gaussian information
- Other hybrid variants also improve the state estimate



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

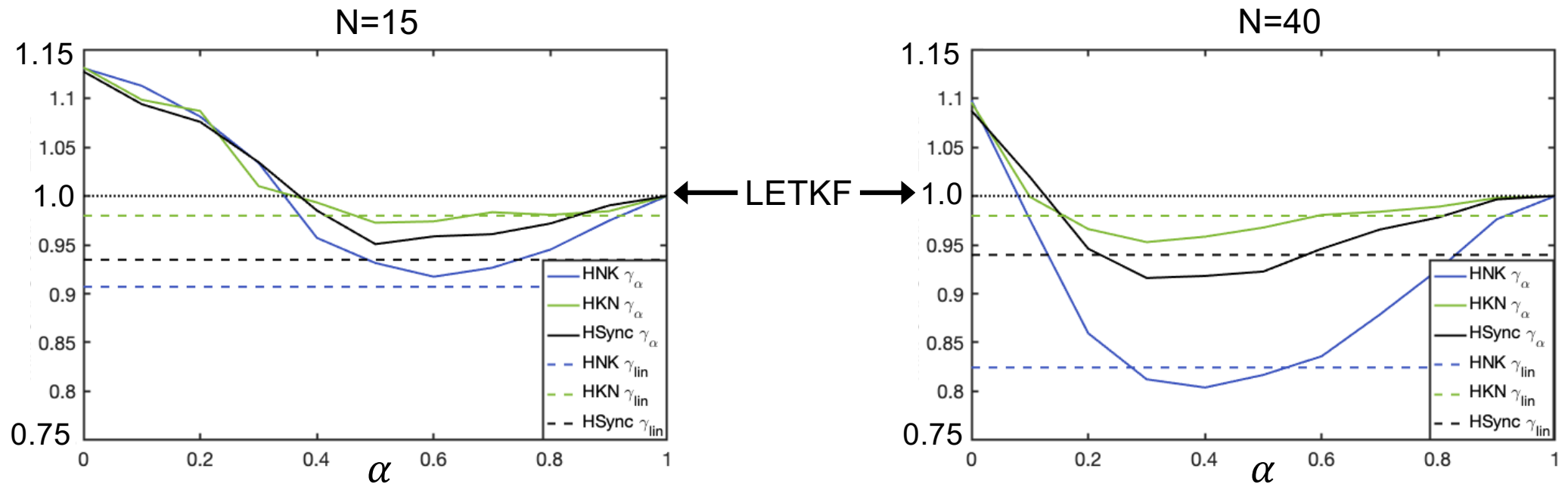
CRPS relative to LETKF



- 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})

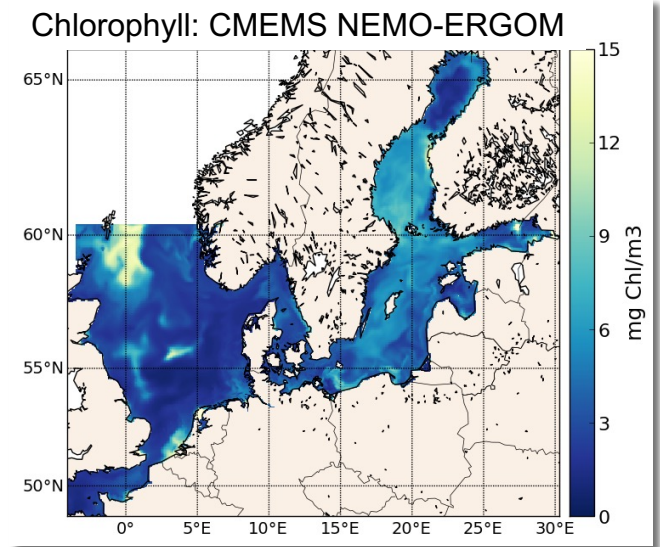
CRPS relative to LETKF



- 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}

Application example

- 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



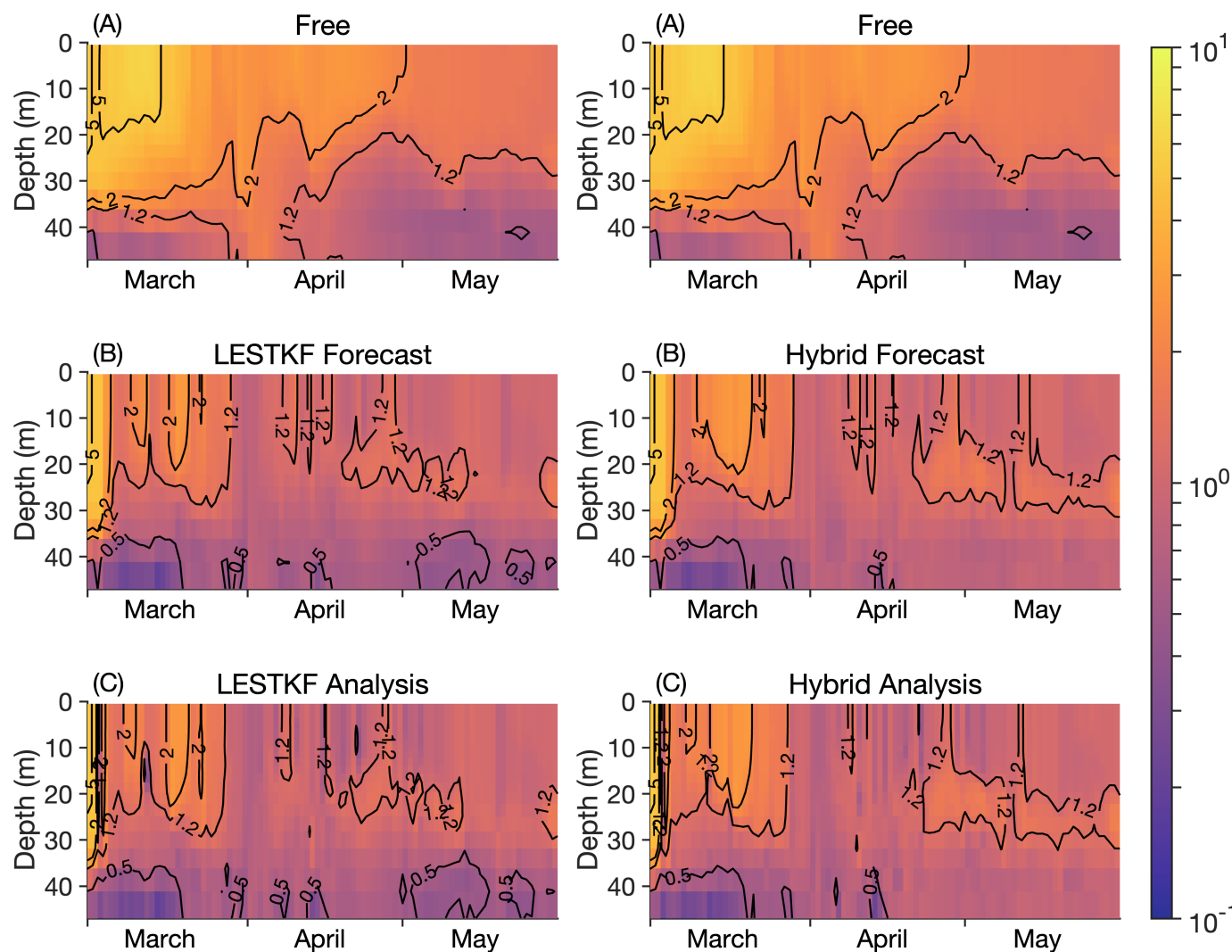
ocean and biogeochemical
dynamics are nonlinear and
distributions non-Gaussian



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



Application in the North Sea & Baltic Sea



- Plots: depth-profile of chlorophyll over time
- Case $\gamma_{sk,\alpha}$ with $N_{\text{eff}} > 0.4$
- LKNETF can be applied
- Small differences between LESTKF and LKNETF
- Not clear yet if LKNETF is better

Summary

Introduced hybrid nonlinear-Kalman ensemble transform filter

- Combine LETKF and LNETF methods
- hybrid weight γ shifts filter behavior
- Cost of analysis step $\sim 2\times$ 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}



Next steps

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

