SINTEF

Handling Sparse Observations in Ensemble-based Filtering

With an Application to Drift Trajectory Forecasting

Florian Beiser
Håvard Heitlo Holm
Jo Eidsvik

EnKF Workshop 2022 at Balestrand

Francesco Silva yesterday evening in front of the hotel

Motivation: Search-and-rescue at sea

Photo: Tom Kausanrød, Redningsselskapet

- Define search areas
  - Forecast drift trajectory and associated uncertainty
  - Efficient computational models
- In-situ buoy data
  - Becoming available during missions
  - **Spatially very sparse observations**
- Reduce uncertainty by data assimilation
  - Updating and re-running forecasts
  - Efficient and tailored ensemble methods

Handling Sparse Observations in Ensemble-based Filtering

1. Simplified Ocean Models
2. Data Assimilation Methods
   - Tailored localization for sparse observations (LETKF)
   - Implicit equal-weight particle filter (IEWPF)
3. Comparison
   - Benchmark experiment
   - Skill score assessment

Florian Beiser[1,2,*], Håvard Heitlo Holm[1], Jo Eidsvik[2]

[1] SINTEF Digital, Norway

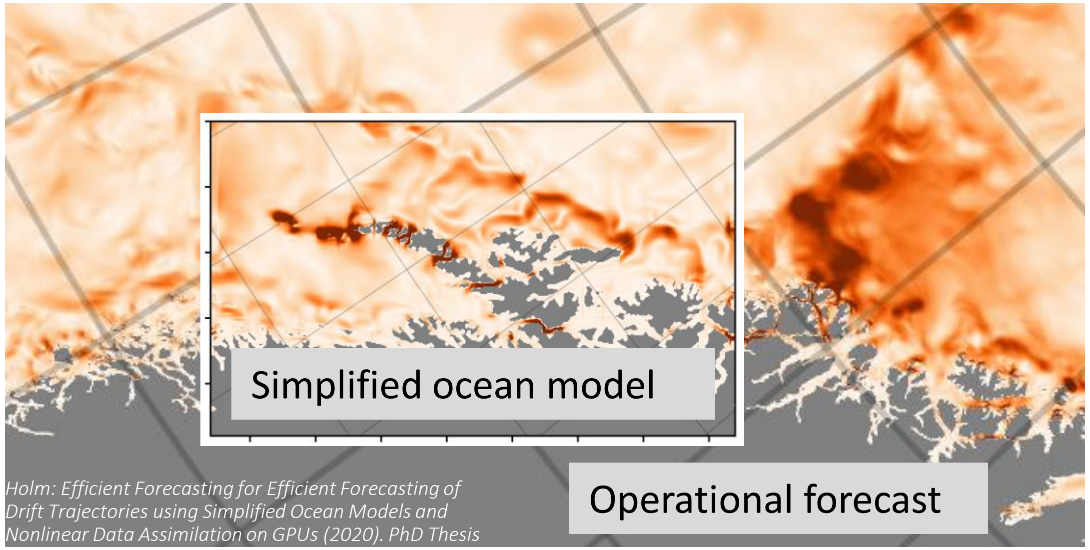
[2] Norwegian University of Science and Technology, Norway

* florian.beiser@sintef.no

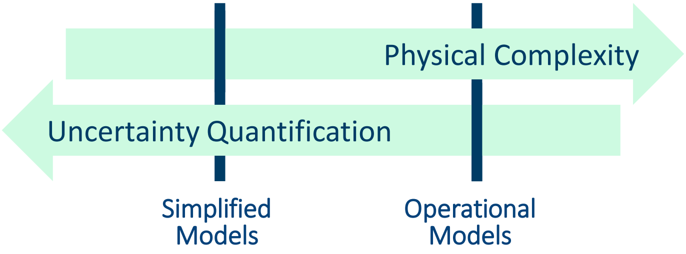
This work is supported by the Research Council of Norway (RCN) under the *Havvarsel* project, grant number 310515.

Simplified Ocean Model

- Operational forecast machinery
  - Complex physical models
  - Multiple types of observations
- Complementary approach
  - Modelling short term physics
  - Initialised from operational forecasts
- Probabilistic forecasts
  - Enabling larger ensembles
  - Improved statistical explanatory power



Holm: Efficient Forecasting for Efficient Forecasting of Drift Trajectories using Simplified Ocean Models and Nonlinear Data Assimilation on GPUs (2020). PhD Thesis

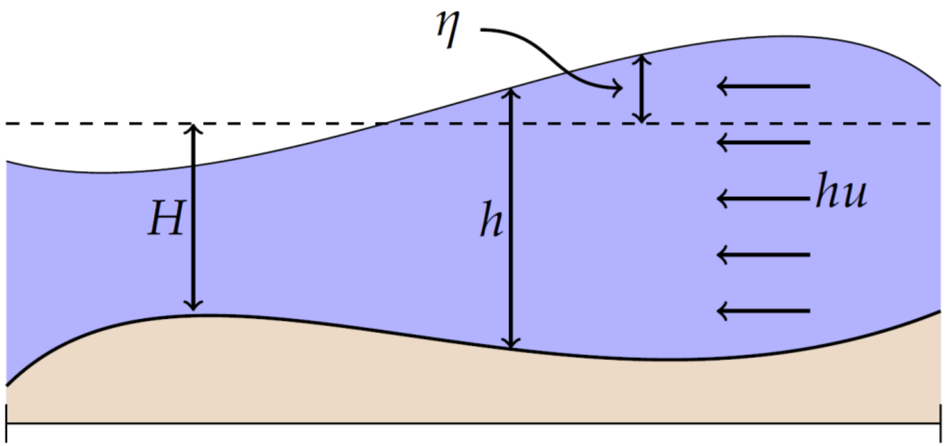


Simplified Ocean Model

Rotational Shallow Water Equation

$$\begin{bmatrix} \eta \\ hu \\ hv \end{bmatrix}_t + \begin{bmatrix} hu \\ hu^2 + 1/2\, gh^2 \\ huv \end{bmatrix}_x + \begin{bmatrix} hv \\ huv \\ hv^2 + 1/2\, gh^2 \end{bmatrix}_y = \begin{bmatrix} 0 \\ fhv \\ -fhu \end{bmatrix} + \begin{bmatrix} 0 \\ ghB_x \\ ghB_y \end{bmatrix}$$

- Assumptions
  - Depth-integrated quantities
  - Barotropic dynamics
- Hyperbolic conservation law
  → Conserved $\boldsymbol{x} = [\eta, hu, hv]^T$
- Parallelised numerics using FVM on GPU
  → Computationally highly efficient model



Holm: Efficient Forecasting for Efficient Forecasting of Drift Trajectories using Simplified Ocean Models and Nonlinear Data Assimilation on GPUs (2020). PhD Thesis

Data Assimilation Problem

Spatiotemporal State

$$\boldsymbol{x}^n = \boldsymbol{x}(t^n, \boldsymbol{s})$$

- Very high-dimensional N_x
- Ocean current and sea surface level

Model

$$\boldsymbol{x}^{n,f} = \mathcal{M}\left(\boldsymbol{x}^{n-1}\right) + \boldsymbol{\nu}^n$$

- Non-linear model $\mathcal{M}$
- Model error $\boldsymbol{\nu}^n \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$
- Simplified ocean model

Observation

$$\boldsymbol{y}^n = \boldsymbol{H}(\boldsymbol{x}^n) + \boldsymbol{\varepsilon}^n$$

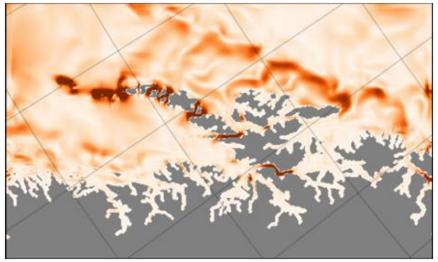
- Very low dimensional N_y
  $N_y \ll N_x$
- Observation error $\boldsymbol{\varepsilon}^n \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$
- Buoy measurements

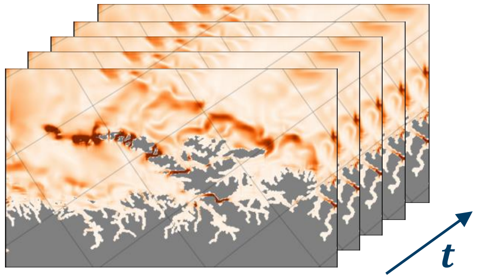
Ensemble Approach

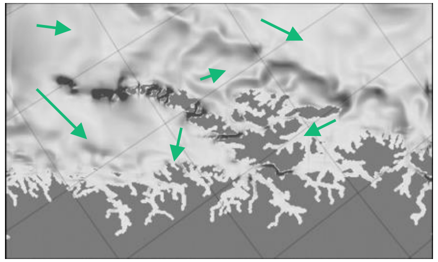
$$p(\boldsymbol{x}^n | \boldsymbol{y}^{1:n})$$

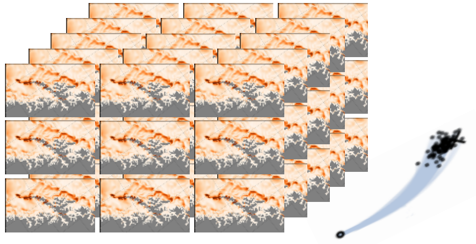
- Estimated by a set of realisations

$$\boldsymbol{x}_e^n, \qquad e = 1, \dots, N_e$$









Local Ensemble Transform Kalman Filter (LETKF)

- LETKF common in operational numerical weather forecasting
- ETKF linear data assimilation method
- Small ensemble sizes can lead to spurious correlations
- Localisation statistically and physically motivated

Classical Observation Localisation[1]

The analysis of a certain point in space is only influenced by the observations in its neighbourhood.

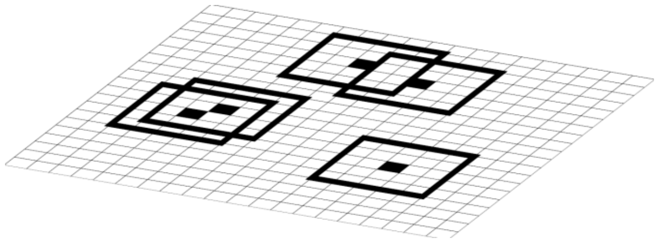
↓

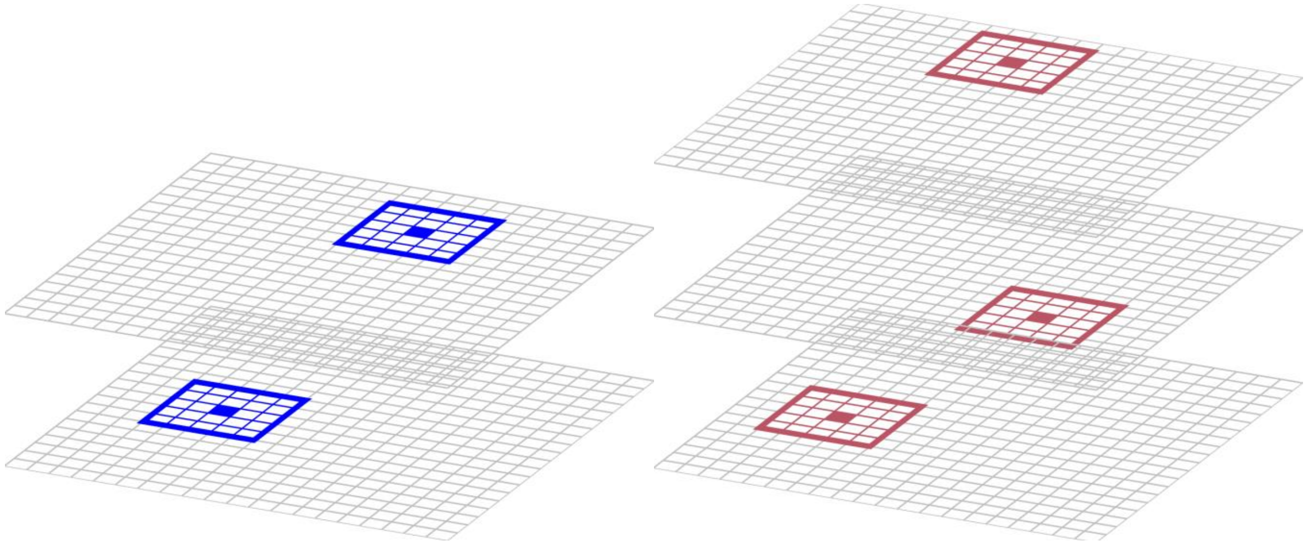
Localisation for Sparse Observation

A certain observation influences only the analyses of the points in space within its neighbourhood.

[1] Hunt, Kostelich, and Szunyogh (2007). Efficient Data Assimilation for Spatiotemporal Chaos: A Local Ensemble Transform Kalman Filter. *PhysicaD: Nonlinear Phenomena*, 230.

LETKF: Local Analysis for Sparse Observations





- Forecast $\boldsymbol{x}_e^{n,f}$ in global domain
- Batches of "uncorrelated" observations

- ETKF analyses $\boldsymbol{x}_e^{n,a}$ in local domains independently
- Serial processing of observation batches

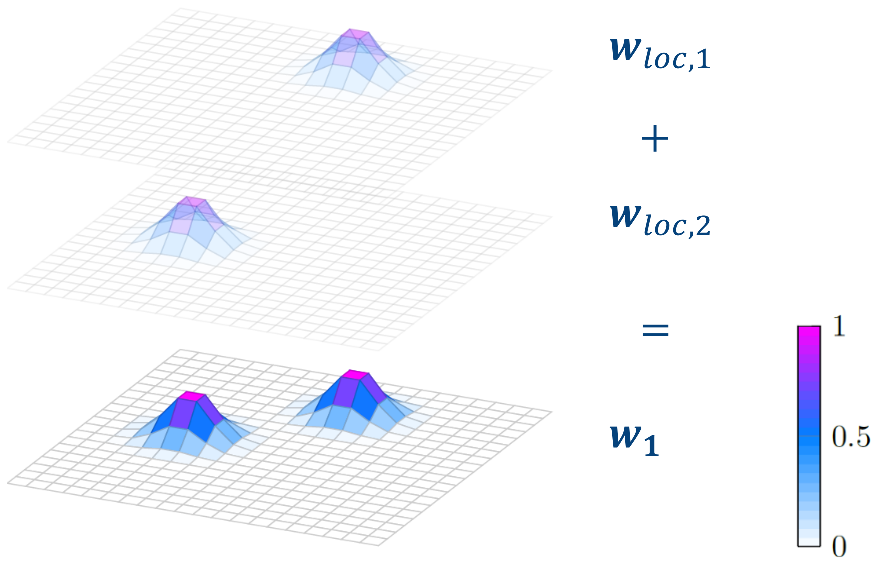
LETKF: From Local to Global by Weighting

Reconstructing Global Analysis State

Intermediate global analysis

$$\boldsymbol{x}_e^{n,a,1} = (1 - \boldsymbol{w}_1)\, \boldsymbol{x}_e^{n,f} + \boldsymbol{w}_1 \left(\sum_{j \in B_1} \boldsymbol{x}_e^{n,a}(j) \right)$$

where $\boldsymbol{w}_1$ constructed by local weights $\boldsymbol{w}_{loc,i}$ around each observation in batch B_1



Gaspari-Cohn function as $\boldsymbol{w}_{loc,i}$

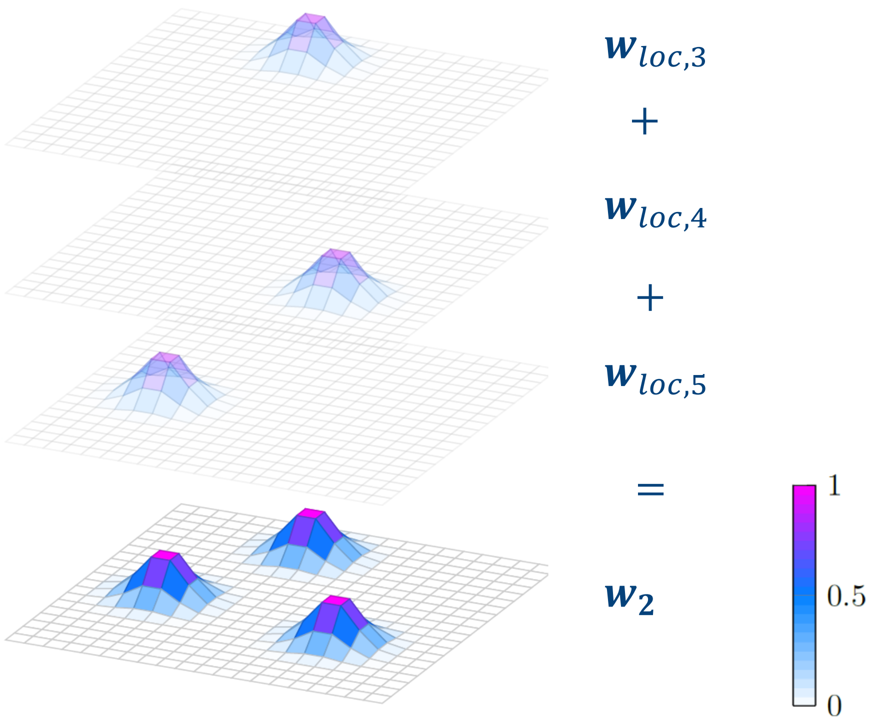
LETKF: From Local to Global by Weighting

Reconstructing Global Analysis State

Final global analysis

$$x_e^{n,a} = (1 - \boldsymbol{w_2})\ x_e^{n,a,1} + \boldsymbol{w}_2 \left(\sum_{i \in B_2} x_e^{n,a}(j) \right)$$

where $\boldsymbol{w_2}$ constructed by local weights $\boldsymbol{w}_{loc,i}$ around each observation in batch B_2



Gaspari-Cohn function as $\boldsymbol{w}_{loc,i}$

LETKF: Inflation through Weighting

- Counteract overfitting by inflation
- E.g., when data very very sparse
- Keeping more spread in analysis
- Introduction of parameter $\phi \in [0,1]$

Inflated Weights

$$\boldsymbol{w}_b^{\text{infl}} = \sum_{j \in \mathcal{B}_b} \phi \, \boldsymbol{w}_{\text{loc},j}$$

- $\phi = 0$: Monte-Carlo
- $\phi = 1$: Scheme without localisation

Implicit equal-weight particle filter (IEWPF)[2]

- Non-linear data assimilation method
- Actively uses model error term to steer ensemble towards observation
  - Computes $\mathbf{QH}^\top$
- Ensures equal weights throughout time
- IEWPF avoids filter degeneracy
  - Applicable to high-dimensional systems

Classical Particle Filter

IEWPF

[2] Skauvold, Eidsvik, van Leeuwen, and Amezcua (2019). A revised implicit equal-weights particle filter. *Quarterly Journal of the Royal Meteorological Society: 145 (721)*.

IEWPF: Localisation and Sparse Observation

Computing $\mathbf{QH}^\top$

- No ensemble approximation in optimal proposal pull
  - No spurious correlations
- Highly dependent on the structure of $\mathbf{Q}$

Localisation

- Correcting only in correlation radius of $\mathbf{Q}$
- Built-in localisation

Sparse Observations

- If correlation ranges of observations do not overlap, then updates independent

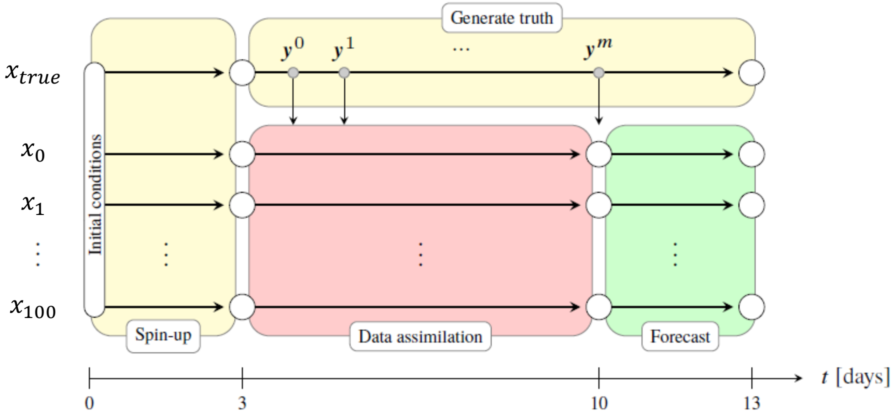
Benchmark Experiment[3]

Set-Up

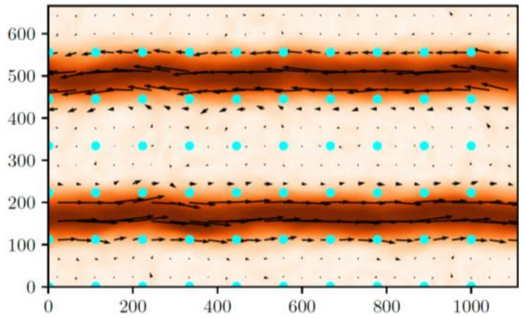
- Identical twin
- Synthetic truth

→ Perfect model with known ground truth

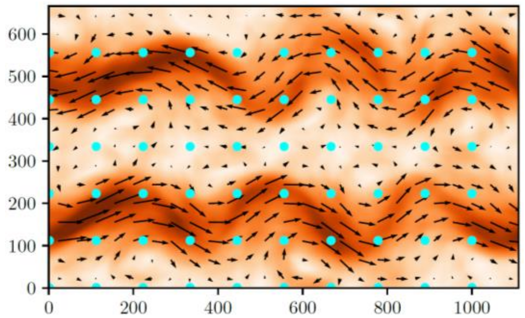
- Observing u and v



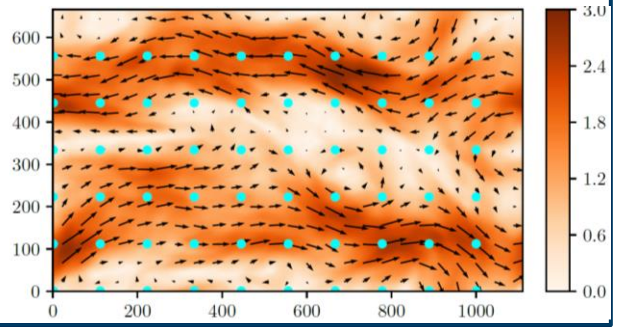
Truth at day 3



Truth at day 6

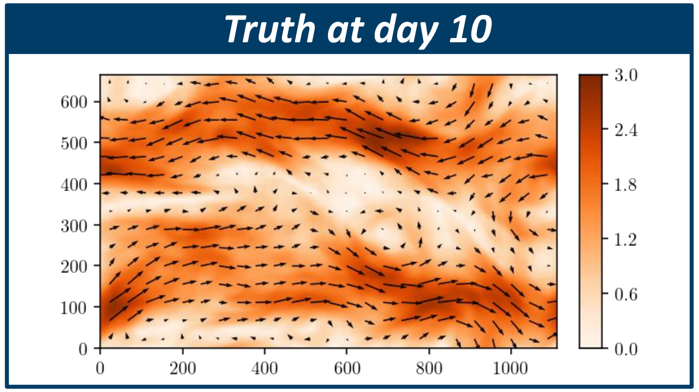


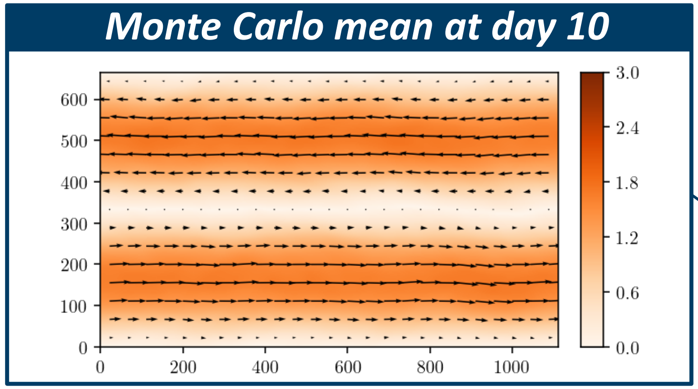
Truth at day 10

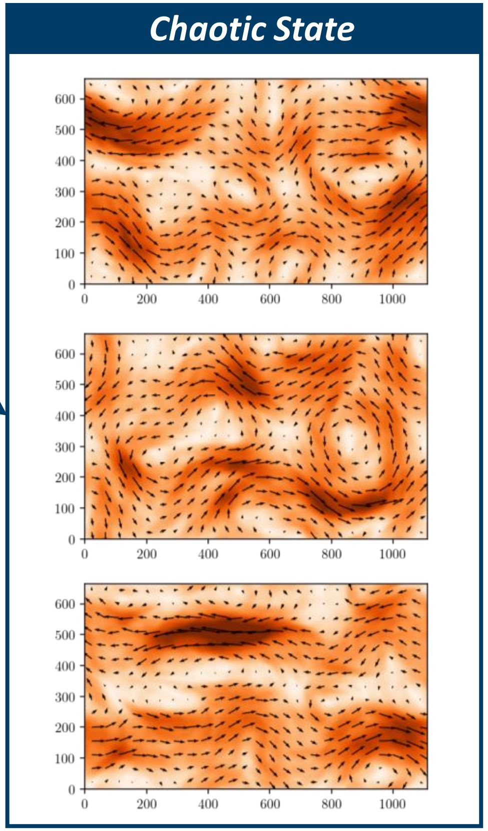


[3] Holm, Sætra, and van Leeuwen (2020). Massively parallel implicit equal-weights particle filter for ocean drift trajectory forecasting. *Journal of Computational Physics: X*, 6, 100053.

Numerical Results







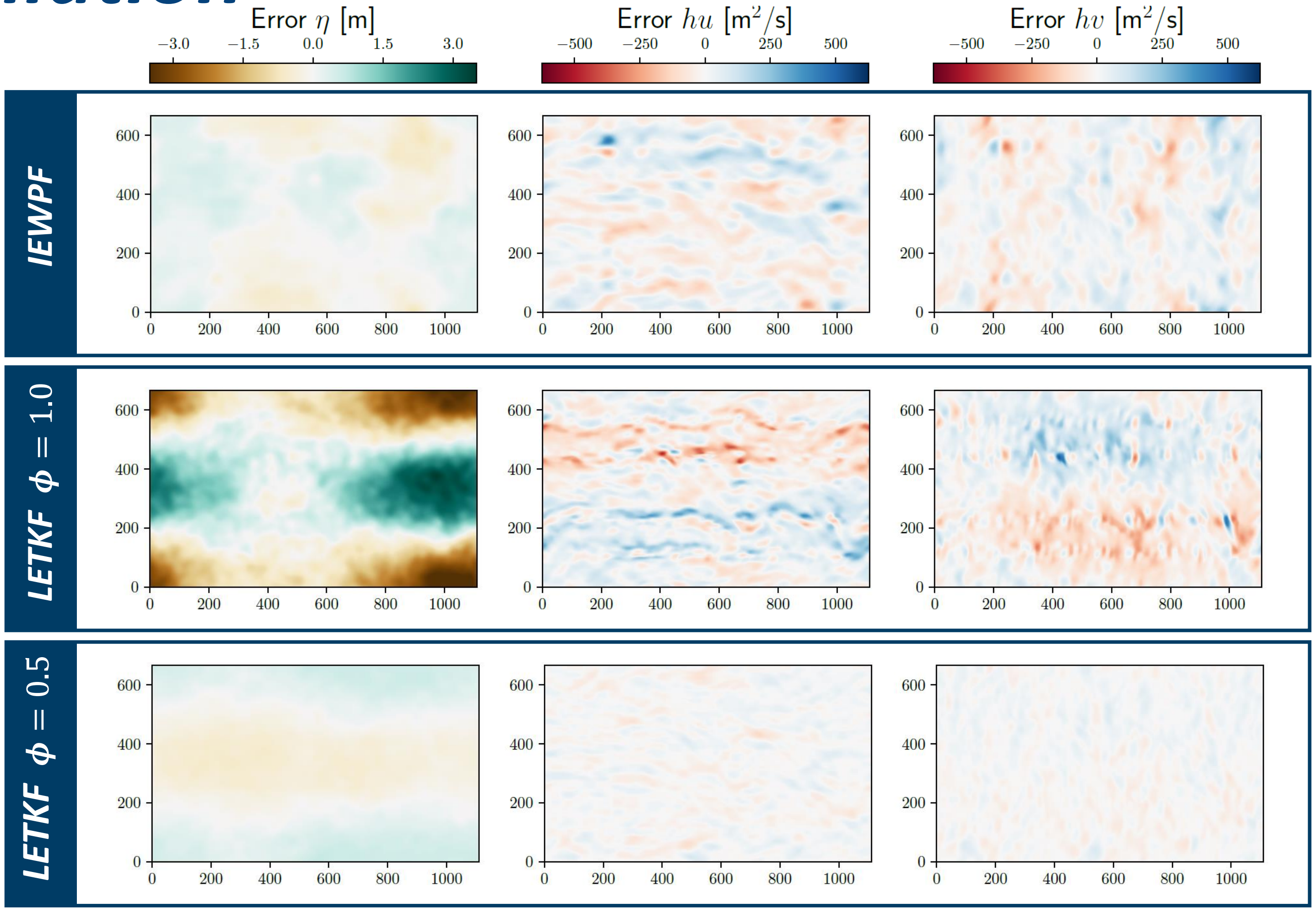
- Chaos develops after few simulation days
- 120 measured quantities per observation to assimilate into 450.000 state variables

Numerical Results
State Assimilation

Error

$$\bar{\boldsymbol{x}} - \boldsymbol{x}_{true}$$

- Evaluated after day 10
- Assesses calibration of ensemble mean

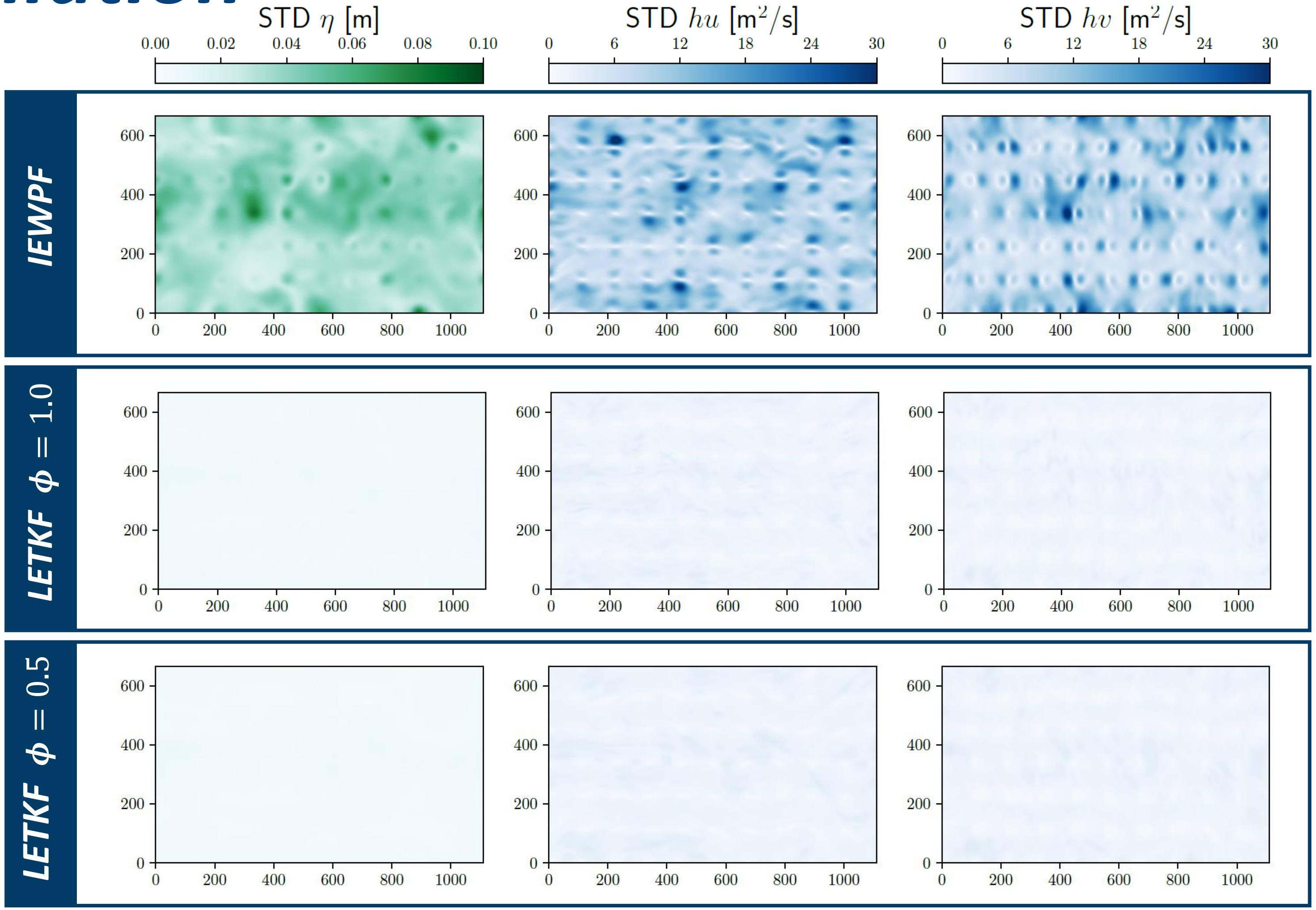


Numerical Results
State Assimilation

Standard Deviation

$$\frac{1}{N_e - 1}\sqrt{\sum(\overline{x} - x_e)^2}$$

- Evaluated after day 10
- Assesses ensemble spread around its mean



Skill Scores

Skill Score s

- Quantitative assessment of the performance of data assimilation methods
- Evaluate how good ensemble can forecast next observations

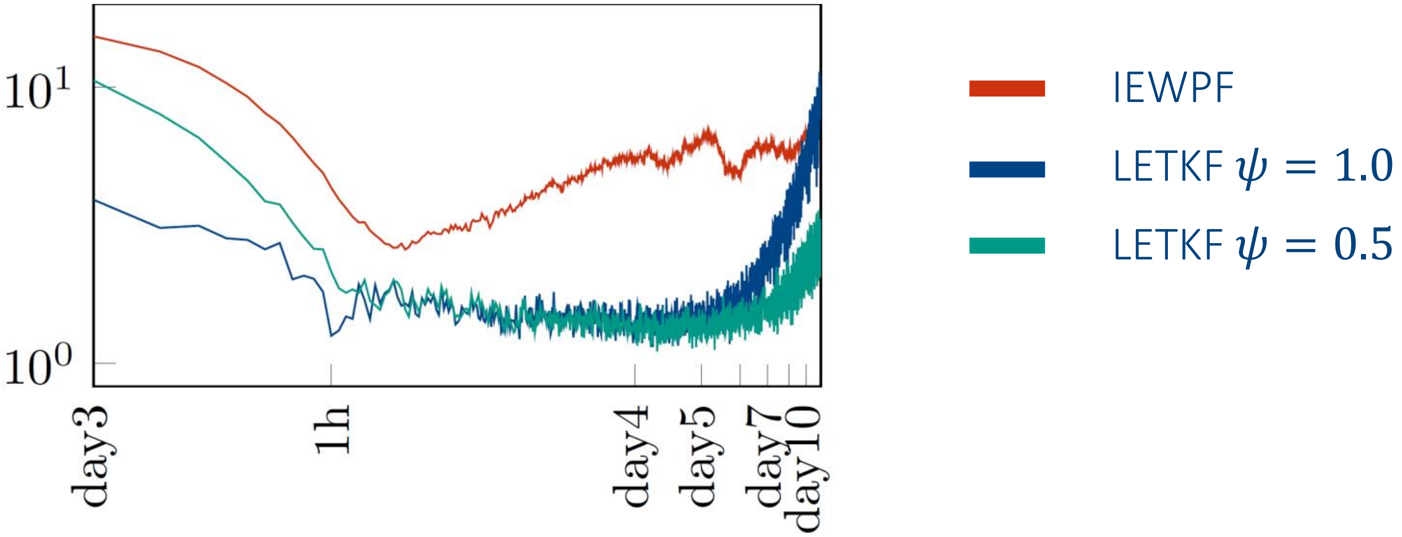
$$s\left(\boldsymbol{x}_e^{n,f}, \boldsymbol{y}^n\right) \in \mathbb{R}$$

- Revealing properties of the ensemble

1. Continuous ranked probability score
2. Bias

Skill Scores

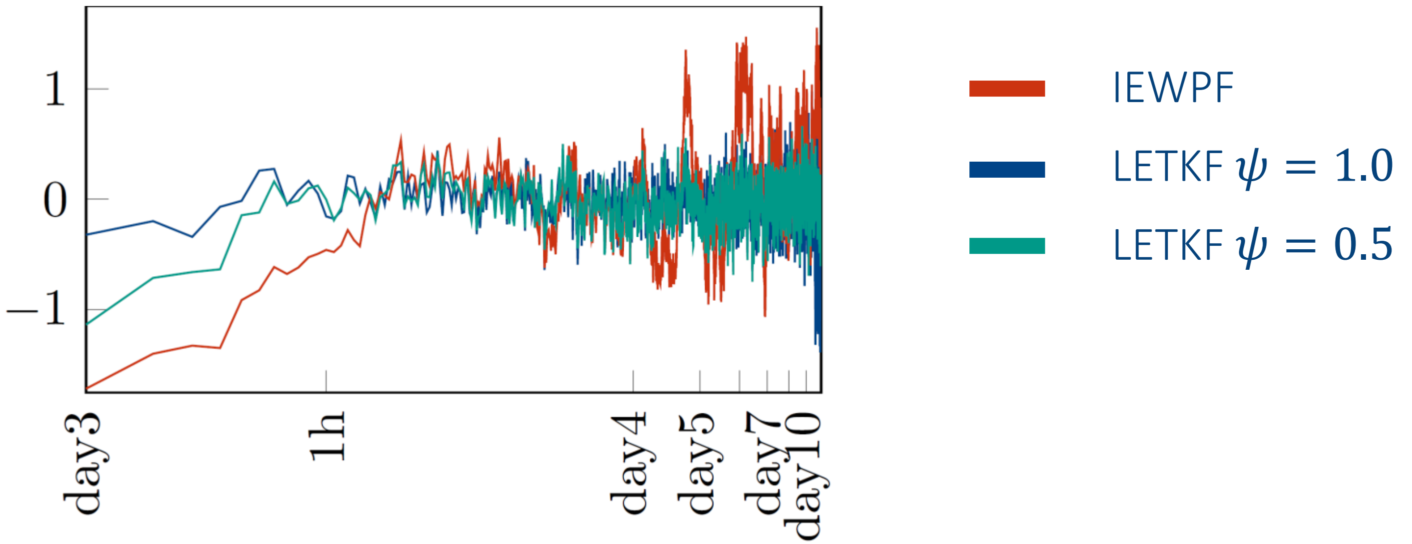
CRPS



- IEWPF corrects less in every update and stays high
- The more inflation in LETKF, the later it diverges
- Contributions to CRPS may originate from bias or high spread

Skill Scores

Bias



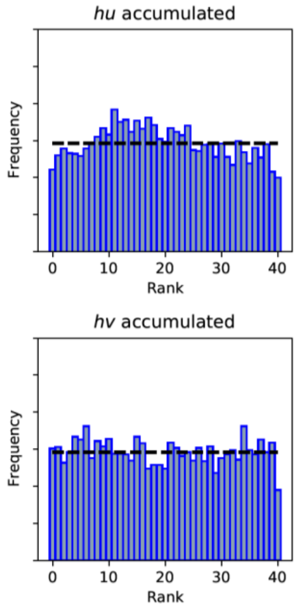
- All biases rather small in the beginning
  - For IEWPF and LETKF $\psi = 1.0$ gets bigger in the end
- Bias stays smallest for LETKF with inflation $\psi = 0.5$

Skill Scores

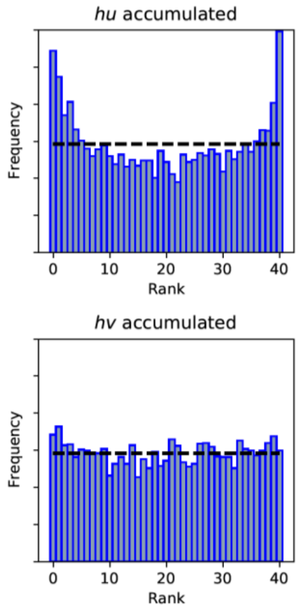
Rank Histograms

- Repeating the first hour of data assimilation in the previous experimental set-up with $N_e = 40$
- Keep record of rank of truth within ensemble at a set of spatial positions

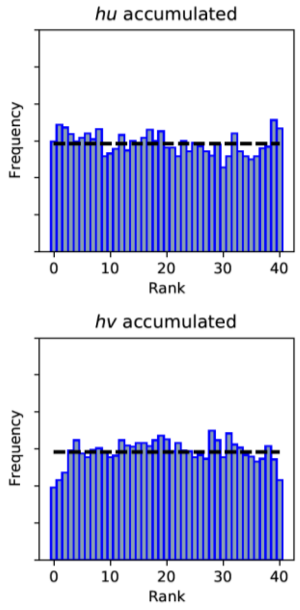
IEWPF



LETKF $\phi = 1.0$



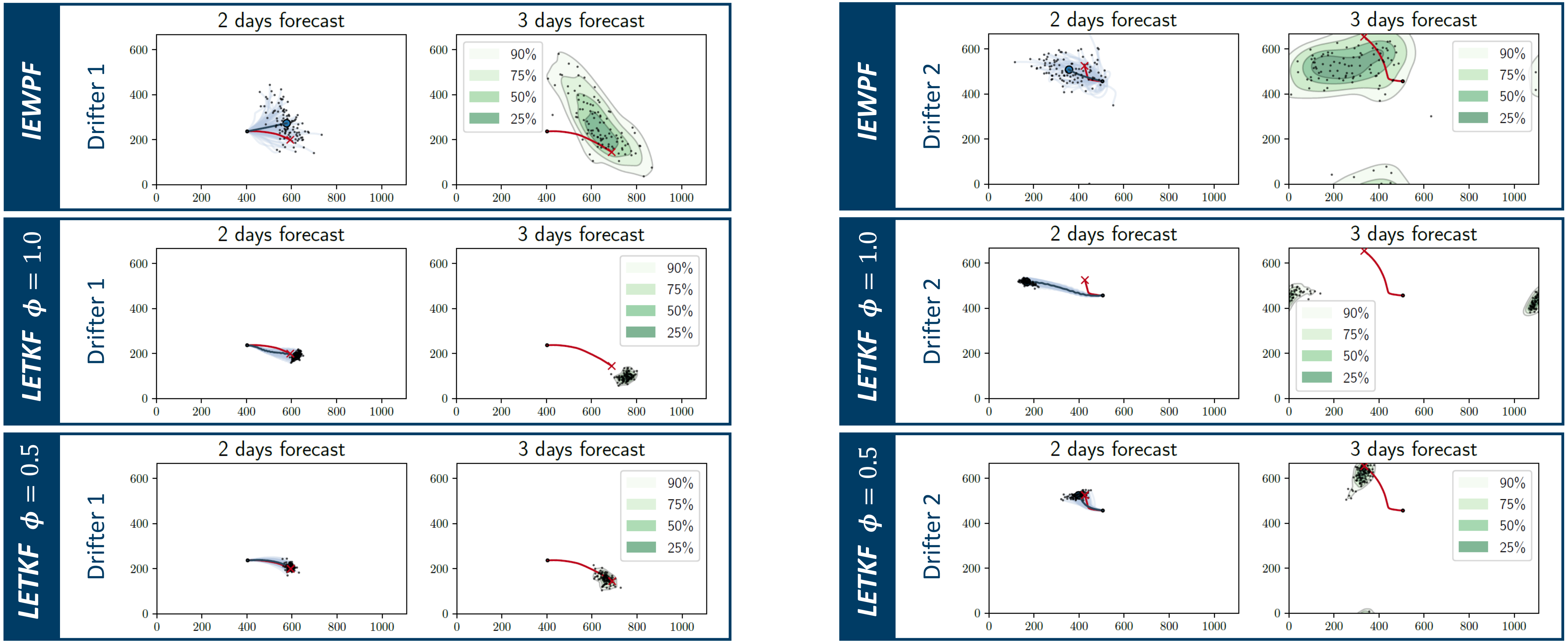
LETKF $\phi = 0.5$



SINTEF

Numerical Results
Drift Trajectory Forecasting

Truth
Ensemble member
Ensemble mean



Closing Remarks[4]

- Localisation for the LETKF is efficient method to assimilate very sparse point data
- Strengthen argument that IEWPF is applicable to high-dimensional applications, but heavily depending on structure of model error covariance matrix
- Broader range of skill scores reveals deep insight

- Evaluate drift trajectories with real-world data
- Employ localisation with other EnKF versions

[4] Beiser, Holm, and Eidsvik (2022). Comparison of Ensemble-Based Data Assimilation Methods for Sparse Oceanographic Data. *In review*

Technology for a
better society