Rank Histograms Estimators for Multi-Level Data Assimilation

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

[1] SINTEF Digital, Norway
[2] Norwegian University of Science and Technology, Norway
* florian.beiser@sintef.no

Motivation: Search-and-rescue at sea

Photo: Tom Kausanrød, Redningsselskapet

- Time-critical definition of search areas
  - Drift trajectories based on a simplified ocean model
  - Refined predictions by data assimilation
- Ensemble methods
  - Forecast together with associated uncertainty
  - EnKF approach for data assimilation
- Multi-level methods
  - Hierarchy of grid resolutions
  - Improving computational efficiency

Rank Histograms Estimators for Multi-Level Data Assimilation

1. Multi-Level Methods
   - Introduction of the basics
2. Rank Histograms
   - Overview of classical rank histograms
   - Multi-level estimation
3. Application
   - Multi-level data assimilation for simplified ocean model

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

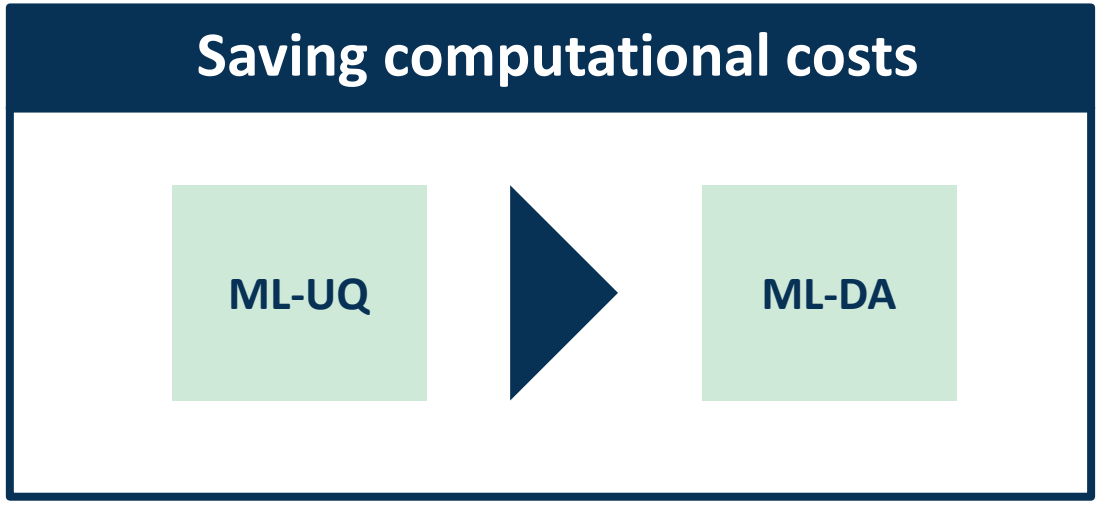
Multi-Level Methods
Motivation

Uncertainty quantification

- Established tool
- Lower computational cost than single-level Monte Carlo at same accuracy

Multi-level data assimilation

- Recent efforts
- New challenges



Multi-Level Methods
Ensemble

Hierarchy of grids with $N_x^0 < \cdots < N_x^L$

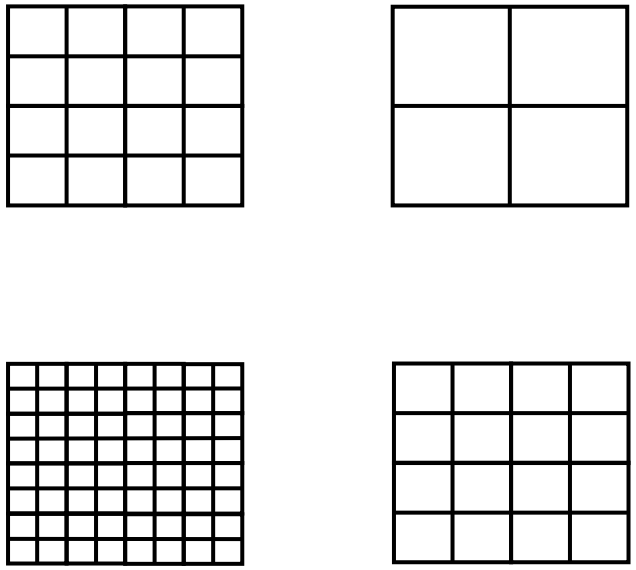
- Coarsest level

$$(x_n^0)_{n=1}^{N^0}$$

- Finer levels

$$\left(x_n^{l+}, x_n^{l-}\right)_{n=1}^{N^l}$$

Same stochastic realisation
on consecutive grids

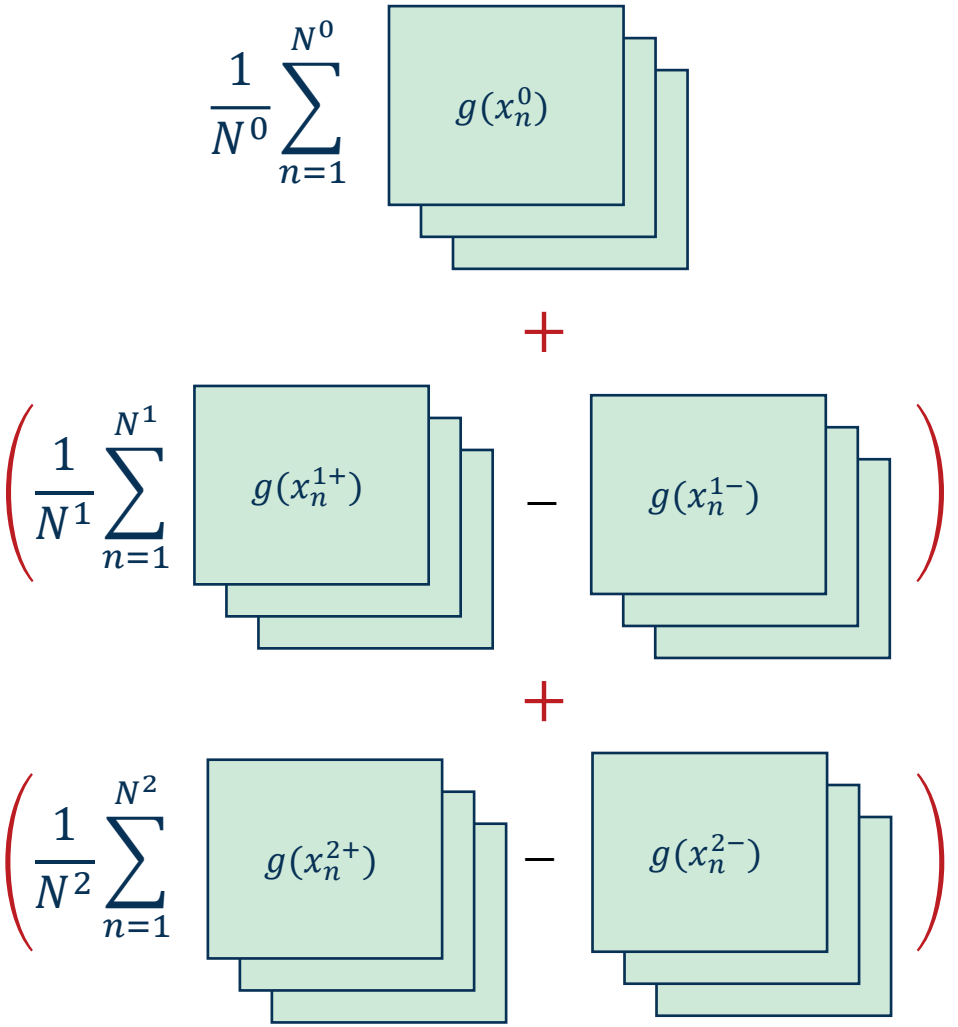


Multi-Level Methods
Estimator

- Smooth function g
- $N^0, \dots, N^L$ optimally chosen for quantity of interest

Multi-Level Estimator

$$\mathbb{E}[g(x^L)] \approx \frac{1}{N^0}\sum_{n=1}^{N^0} g(x_n^0) + \sum_{l=1}^{L}\left(\frac{1}{N^l}\sum_{n=1}^{N^l} g(x_n^{l+}) - g(x_n^{l-})\right)$$



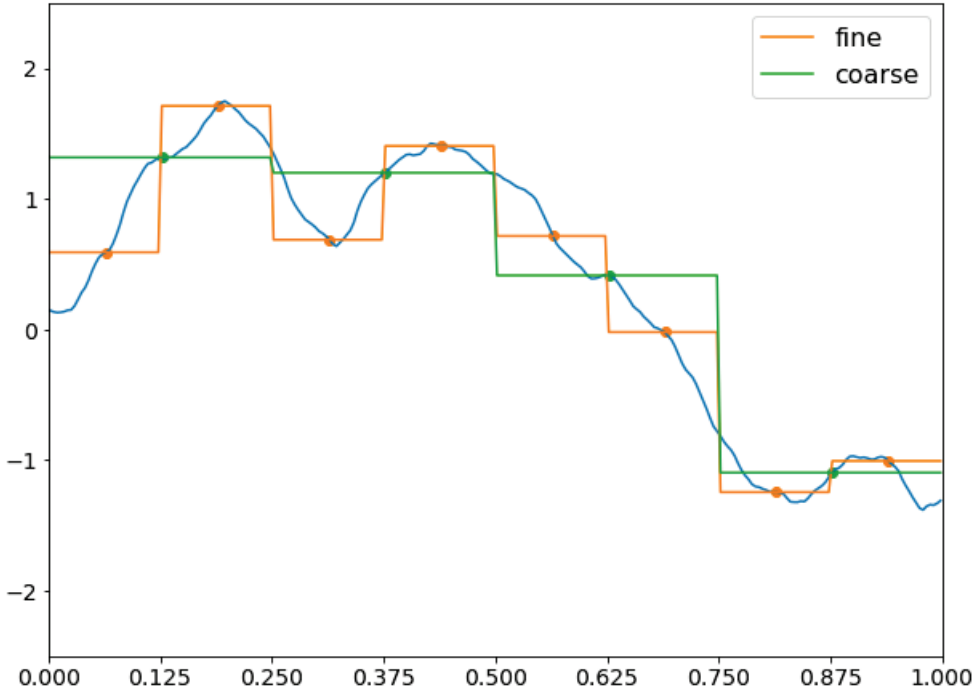
Multi-Level Methods
Gaussian Process

Example on interval $[0,1]$

$$x \sim \mathcal{N}(0, \mathbf{Q})$$

with Matern-like covariance matrix $\mathbf{Q}$

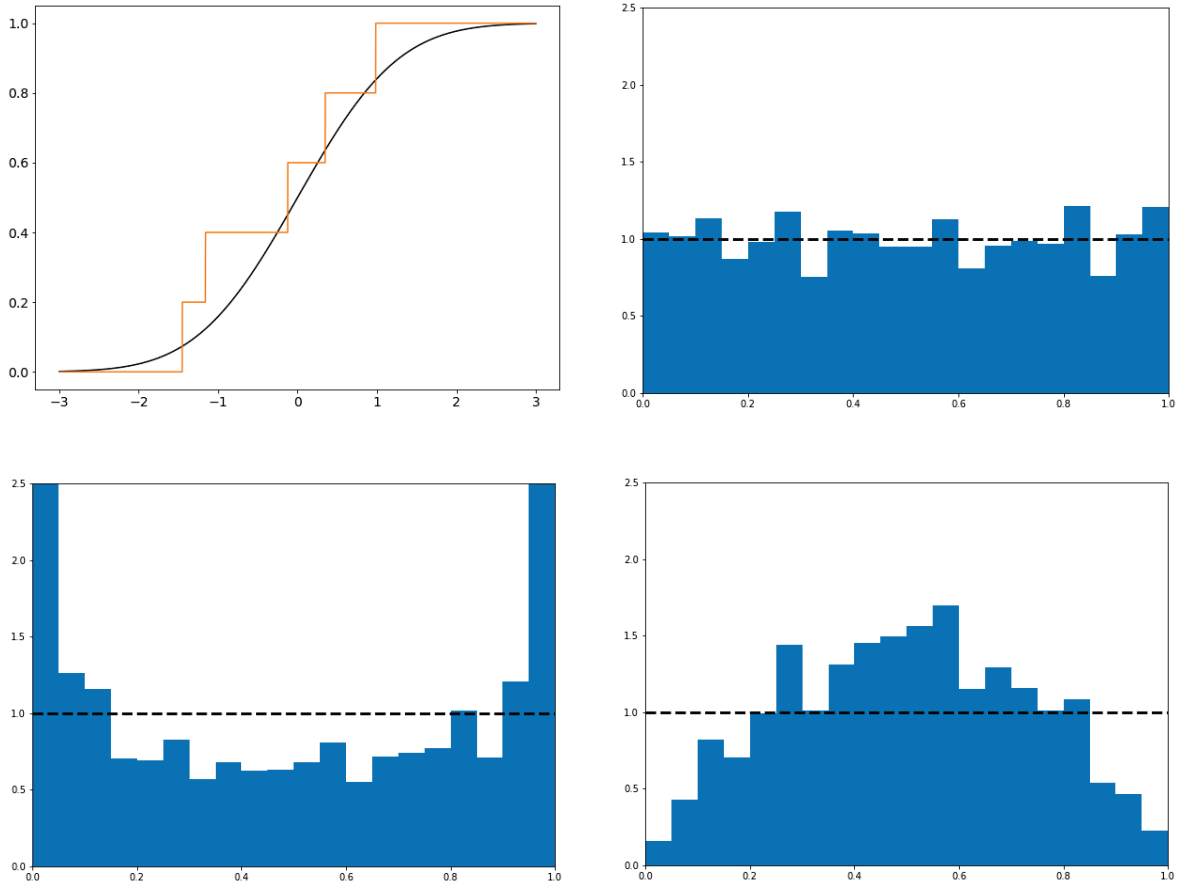
- Levels $l = 0, \ldots, L$
- Values from 2^{4+l} equidistant points



Rank Histograms
Overview

- Assessing "right" spread of ensemble
  - Different skill scores suggested
  - Rank histograms common in practice
- Rank Histograms
  - Tracking rank of truth in sorted ensemble
  - Repeating experiment multiple times
- What rank of true y mathematically is:

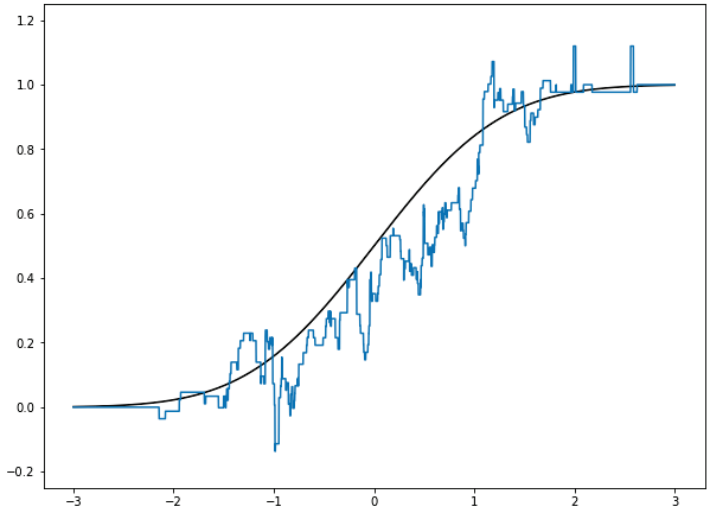
$$\text{rank}(y) = P(X \leq y)$$
$$= \mathbb{E}\left[1_{(-\infty, y]}(X)\right]$$



Rank Histograms
Overview

Representative ensemble size $43, 43, 28, 15, 7$

$$\frac{1}{N^0}\sum_{n=1}^{N^0} 1_{(-\infty,y]}(x_n^0) + \sum_{l=1}^{L}\left(\frac{1}{N^l}\sum_{n=1}^{N^l} 1_{(-\infty,y]}(x_n^{l+}) - 1_{(-\infty,y]}(x_n^{l-})\right)$$



Multi-level comes at a price:

- (Fixed) ensemble size is not good for approximation of indicator function
- Direct evaluation of expected value does not yield feasible CDF!

Empirical CDF

- Direct evaluation of expected value
- Evaluation of expected value with smoothing
- Maximum entropy with polynomial basis

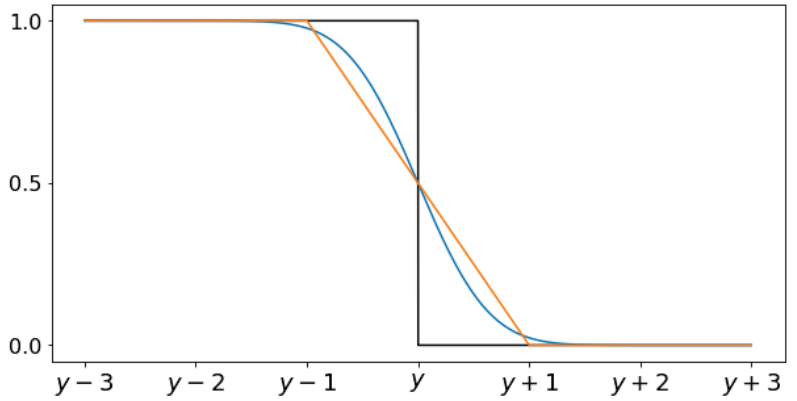
Rank Histograms
Overview

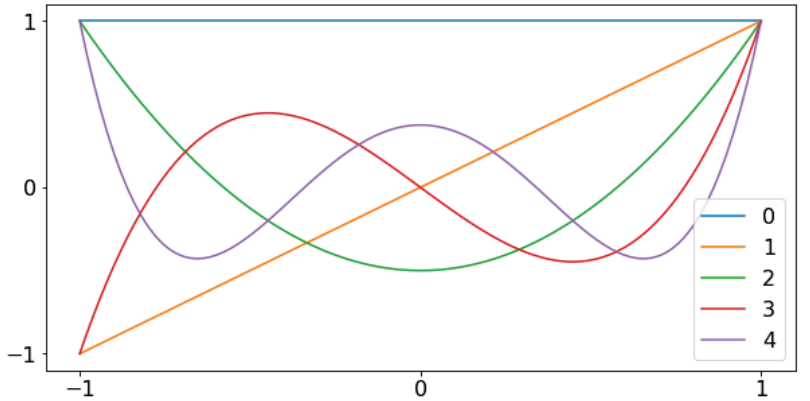
Smoothing of $1_{(-\infty,y]}$

- Kernel with bandwith ρ
  - Gaussian
  - Piecewise linear

Maximum entropy

- For set of basis functions $\phi_0, \ldots, \phi_{N-1}$
- PDF approximation by $p = \exp\left(\sum_{i=0}^{N-1} \lambda_i \phi_i\right)$
- Maximising Shannon entropy
- Constraints $\mathbb{E}[\phi_i] = \int \phi_i p \, \mathrm{dx}$

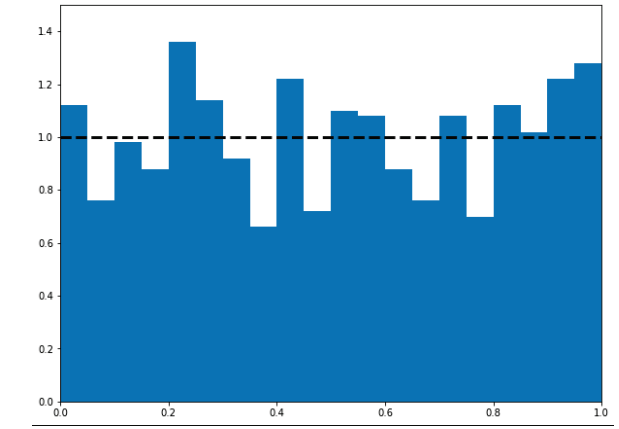


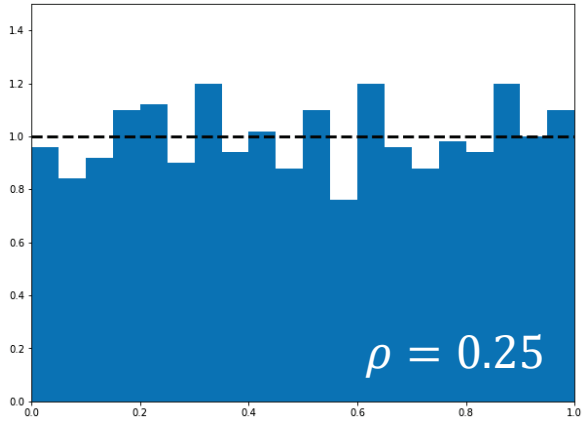


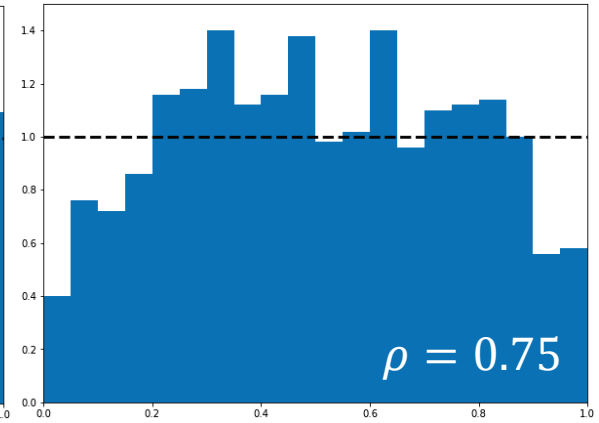
Rank Histograms
Overview

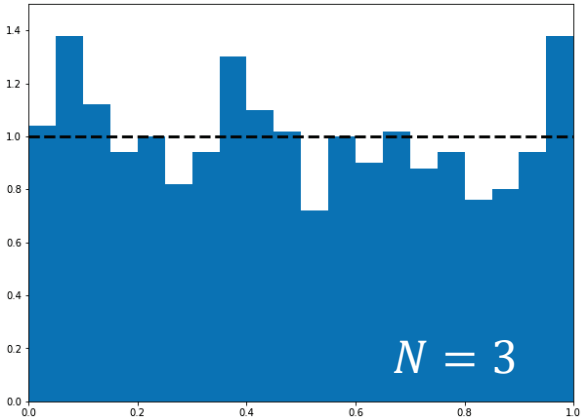
- *Single-level* ensemble with 50 members
- Truth and ensemble from same distribution

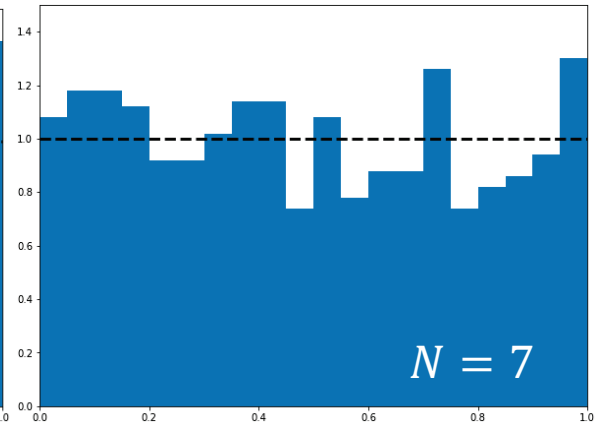
1. Direct evaluation
2. Smoothing
3. Maximum entropy









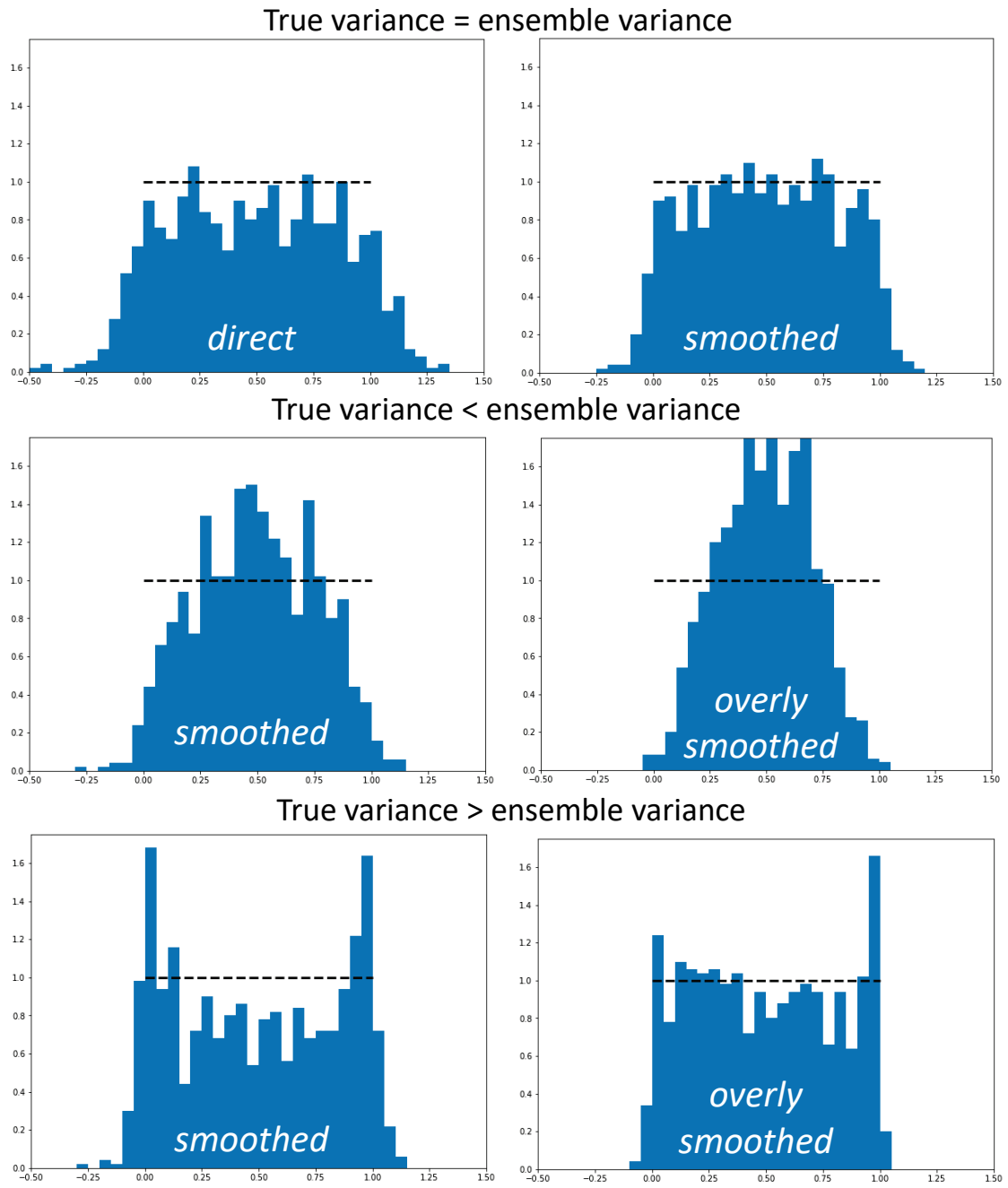


Rank Histograms
Multi-Level

- Multi-level ensemble
- Smoothing:
  The finer level, the bigger bandwidth
- Case with enhanced variance

→ Always applicable

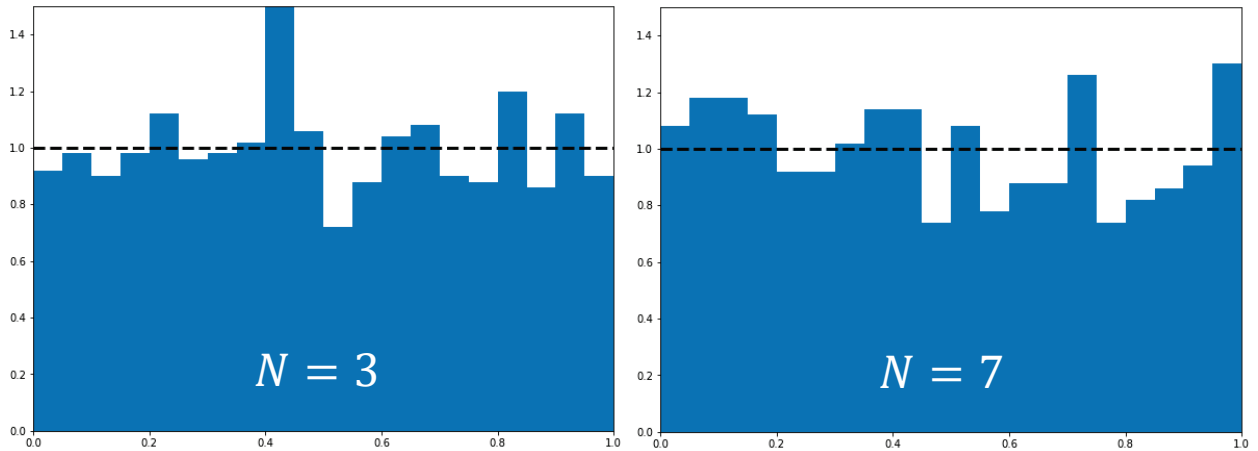
→ Choice of bandwidth must not affect interpretability



Rank Histograms *Multi-Level*

- Multi-level ensemble
- Maximum entropy with Legendre basis

True variance = ensemble variance



→ Few parameters

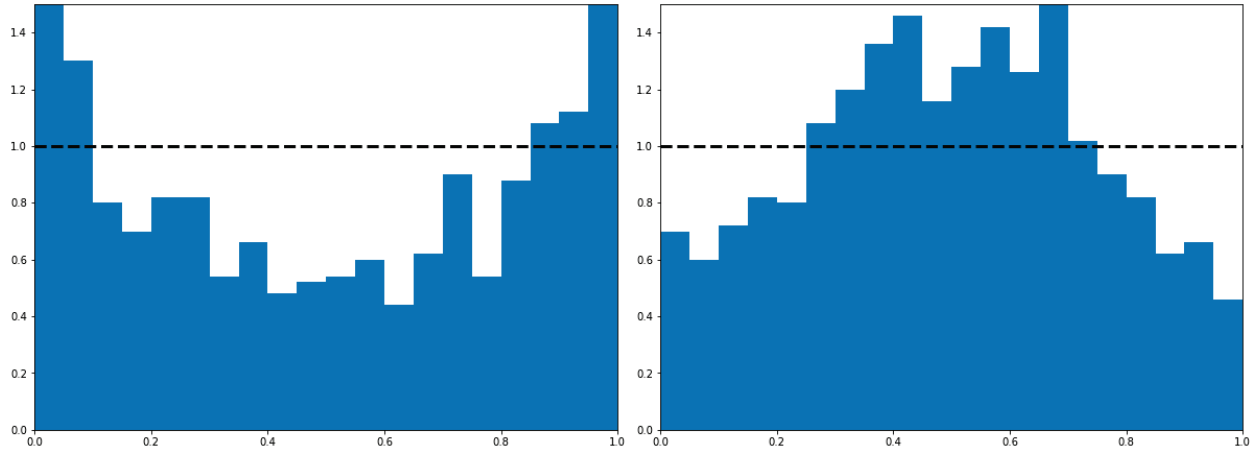
→ Ranks are always $[0,1]$

→ Sensitive to estimation quality of $\mathbb{E}[\phi_i]$

→ Convergence issues with high variance

True variance
>
ensemble variance

True variance
<
ensemble variance

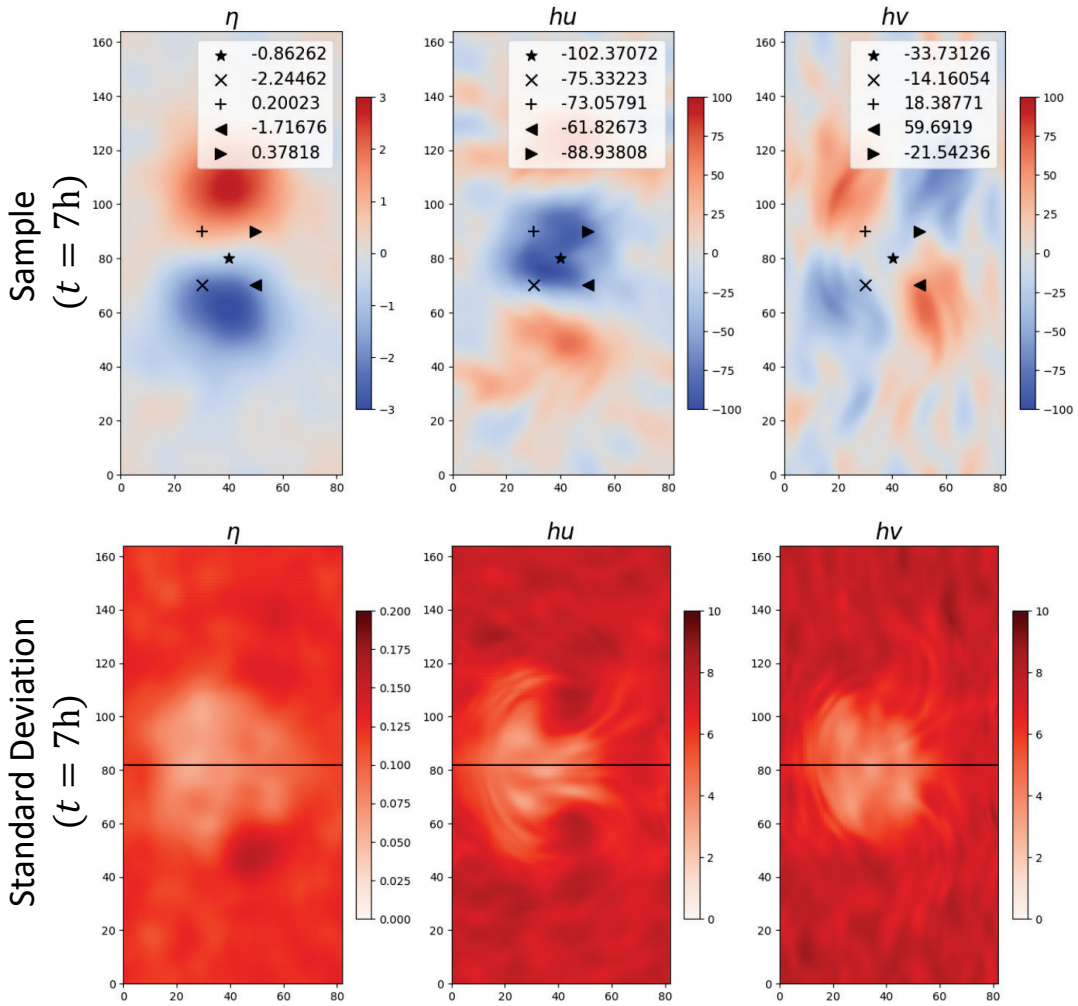


Application
MLEnKF for simplified ocean model

- Rotational shallow-water model
  - h: water column. η: deviation from equilibrium
  - u, v: velocities in x, y-direction
  - Additive Gaussian error

- Data assimilation with MLEnKF*
  - Ensemble size $313, 84, 26, 6$
  - Sparse observations in center
  - 6h simulation with DA every 15min
  - 1h forecast

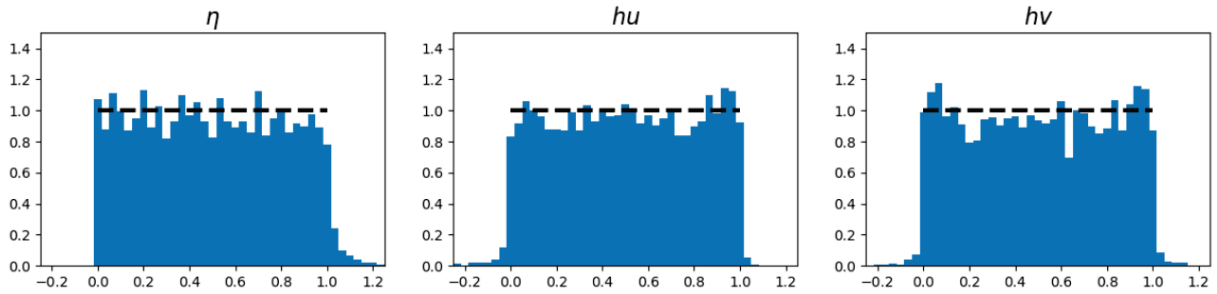
- Rank histograms
  - Observing locations along horizontal line



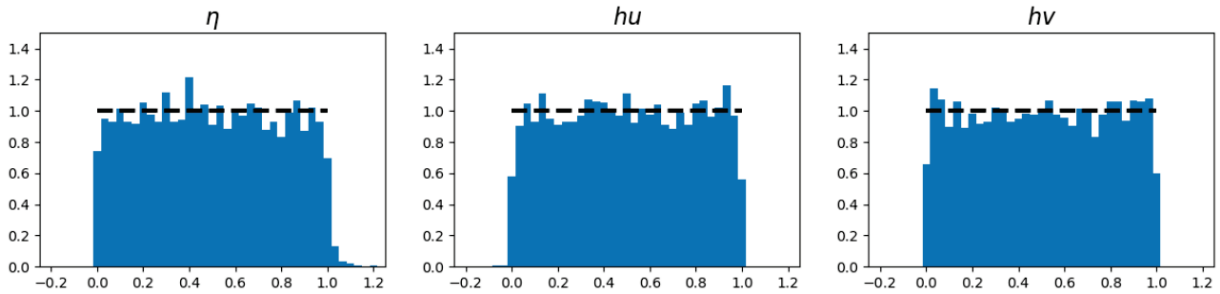
* Chernov, Hoel, Law, Nobile, Tempone: *Multilevel Ensemble Kalman filtering for spatio-temporal processes.* Numerische Mathematik (2021) 147:71-125

Application
MLEnKF for simplified ocean model

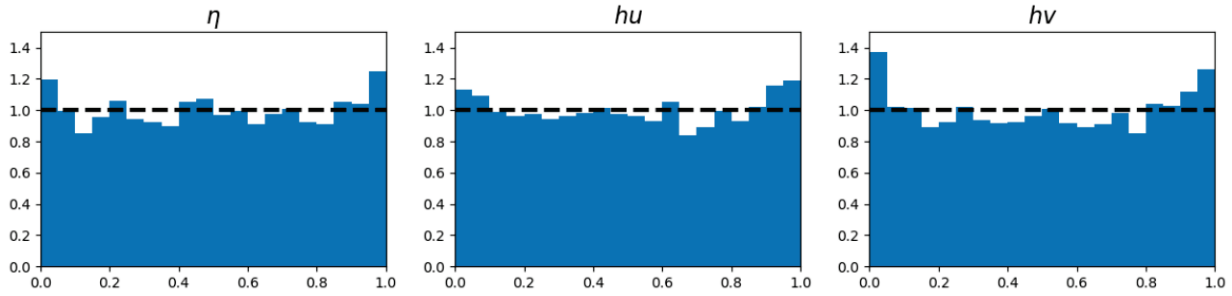
Direct evaluation



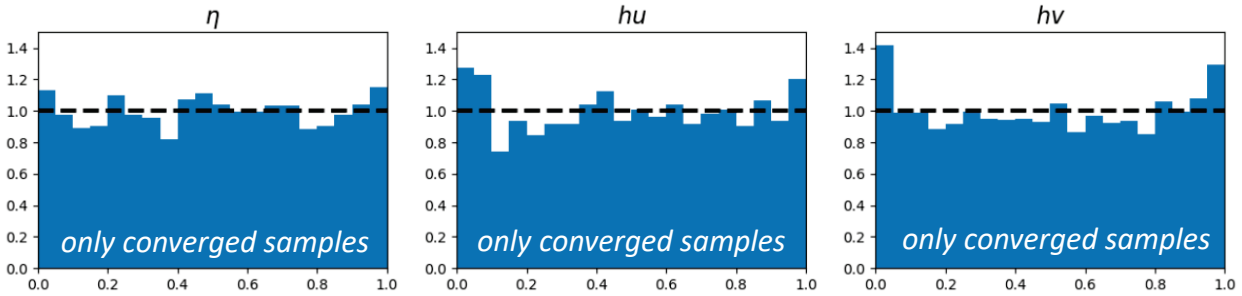
Smoothing with increasing bandwidth



Maximum entropy ($N = 3$)



Maximum entropy ($N = 5$)



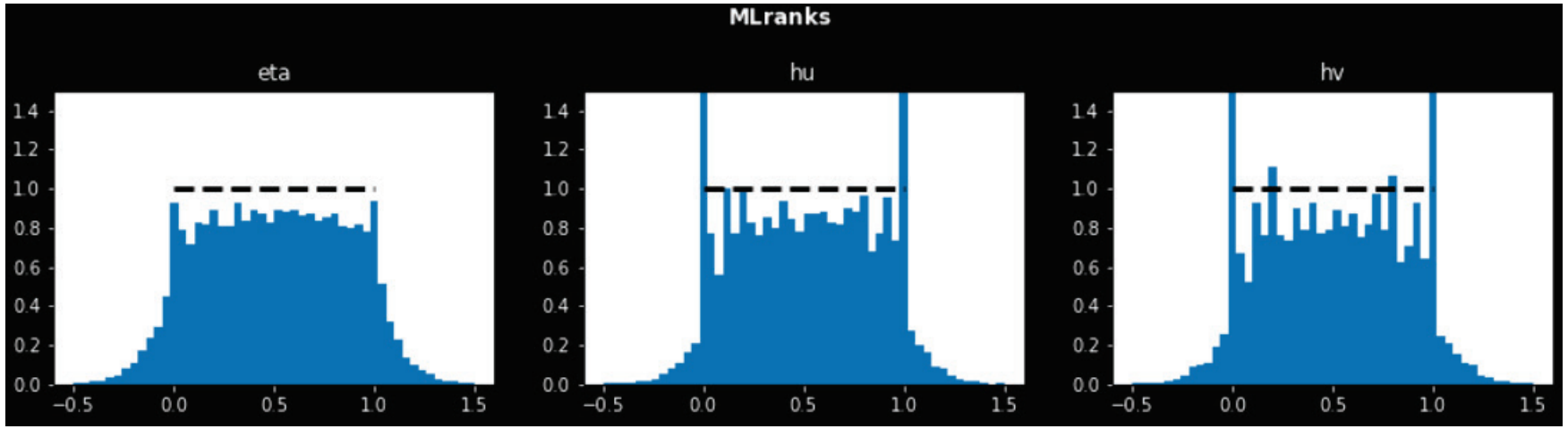
Closing Remarks

- Multi-level ensembles interesting approach for the future
- Still open question how to evaluate ensemble spread
- For small multi-level ensemble size, the direct evaluation of ranks might be infeasible
- Three approaches to calculate rank histograms discussed
- Multi-level rank histograms provide tool to evaluate spread in multi-level DA

SINTEF

Technology for a better society

SWE rank histograms (initial prior)



Ranks per level (GP, different run)