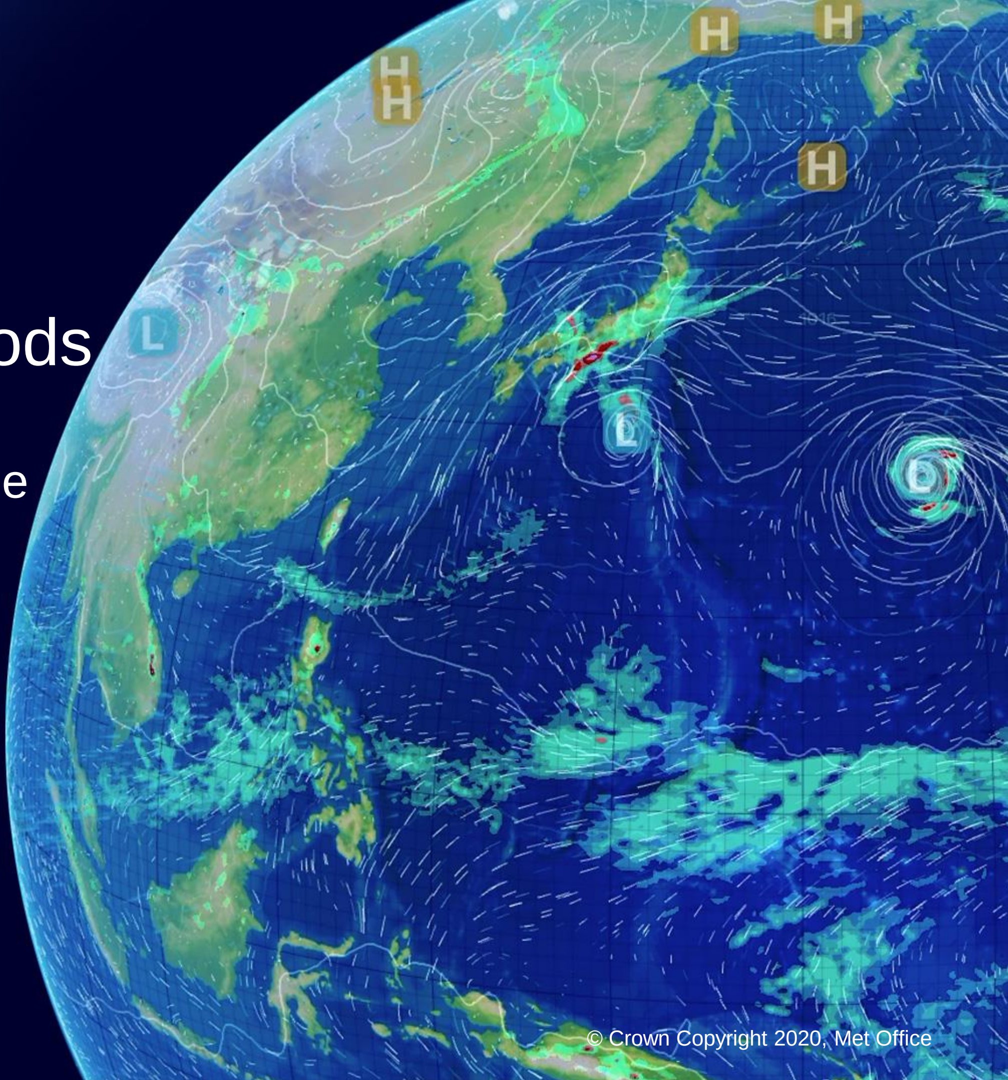
Improving background error covariance estimation with Multilevel Monte Carlo Methods

Mayeul Destouches, Paul Mycek, Selime Gürol, Anthony Weaver, Serge Gratton, Ehouarn Simon, Jérémy Briant

ISDA 2023, Bologna



Outline

- Introduction and theory
- Experiments with a quasi-geostrophic model
- Conclusion

Met Office

Multilevel Monte Carlo (MLMC) and ensemble Data Assimilation

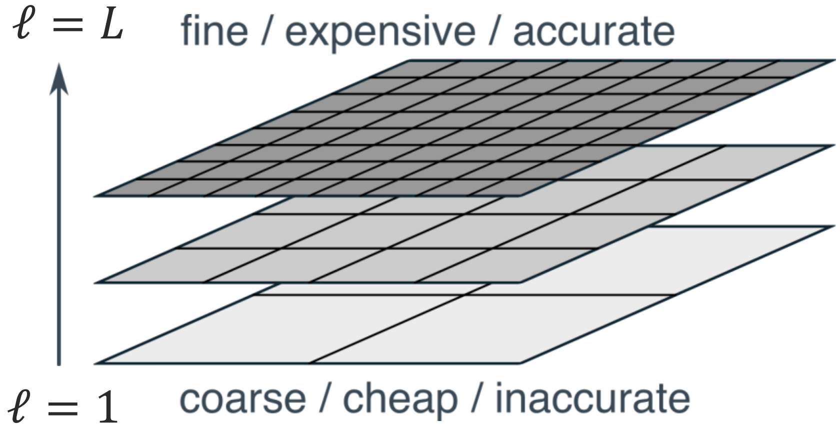
- Already tried with EnKFs and in small dimensions (Hoel et al., 2016)
- Here: Ensemble variational DA based on an Ensemble of Data Assimilations (EDA)

$$\mathbf{x}_{\mathrm{a}} = \min_{\mathbf{x}} \frac{1}{2}(\mathbf{x} - \mathbf{x}_{\mathrm{b}})^t \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_{\mathrm{b}}) + \frac{1}{2}\left(\mathbf{y} - \mathcal{H}(\mathbf{x})\right)^t \mathbf{R}^{-1}\left(\mathbf{y} - \mathcal{H}(\mathbf{x})\right)$$

$$\mathbf{B} = \left\{\frac{1}{N-1}\sum_{k=1}^{N}(\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^t\right\} \circ \mathbf{L} \quad \text{with} \quad \bar{\mathbf{x}} := \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k$$

Ensemble members $\mathbf{x}_1 \ldots \mathbf{x}_N$ from an EDA (N independent perturbed forecast-analysis systems)

Hierarchy of grids



Let $f_\ell\colon \mathbb{R}^n \mapsto \mathbb{R}, 1 \leq \ell \leq L$ be a hierarchy of simulators.

Goal: Estimate the scalar mean $\mu_L \coloneqq \mathbb{E}[f_L(X)]$, for some random vector X

Simple Monte Carlo (MC) estimator

$$\mu_L^N = \frac{1}{N}\sum_{i=1}^{N} f_L\left(X^{(i)}\right)$$

- Bias: $\mathbb{E}[\mu_L^N] - \mu_L = 0$
- Variance of the estimator (sampling noise):

$$\mathbb{V}[\mu_L^N] = \frac{1}{N}\mathbb{V}[f_L(X)]$$

with independent members $f_L\left(X^{(i)}\right)$

MLMC estimation of the mean (Giles, 2008)

Same stochastic inputs for both ensembles in a correction term

$$\mu^{\mathrm{ML}} := \frac{1}{N_1}\sum_{i=1}^{N_1} f_1\left(X_i^{(1)}\right) + \sum_{\ell=2}^{L}\left\{\frac{1}{N_\ell}\sum_{i=1}^{N_\ell} f_\ell\left(X^{(\ell,i)}\right) - \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} f_{\ell-1}\left(X^{(\ell,i)}\right)\right\}$$

Quasi-telescopic sum.
Expectations $\mu_\ell - \mu_{\ell-1}$ cancel out: no bias

Ensemble sizes

$\ell = L$				N_4
$\ell = 3$			N_3	N_4
$\ell = 2$		N_2	N_3	
$\ell = 1$	N_1	N_2		

Giles, M. B. (2008). Multilevel Monte Carlo Path Simulation. *Operations Research*, *56*(3), 607–617. https://doi.org/10.1287/opre.1070.0496

Met Office

Variance of the multilevel estimator

$$\mathbb{V}\left[\mu^{\mathrm{ML}}\right] = \frac{\mathbb{V}[f_1(X)]}{N_1} + \sum_{\ell=2}^{L} \frac{(\mathbb{V}[f_\ell(X) - f_{\ell-1}(X)])}{N_\ell} = \sum_{\ell=1}^{L} \frac{\mathcal{V}_\ell}{N_\ell} \quad \text{(notation)}$$

$$\sum_{\ell=1}^{L} \frac{\mathcal{V}_\ell}{N_\ell} \overset{?}{<} \frac{\mathbb{V}[f_L(X)]}{N}$$

Yes, if either:

- N_ℓ is large
- $\mathcal{V}_\ell$ is small

$$\mathcal{V}_\ell = \mathbb{V}[f_\ell(X) - f_{\ell-1}(X)] = \mathbb{V}[f_\ell(X)] + \mathbb{V}[f_{\ell-1}(X)] - 2\,\mathrm{Cov}[f_\ell(X), f_{\ell-1}(X)]$$

Met Office

Generalization to covariance matrices

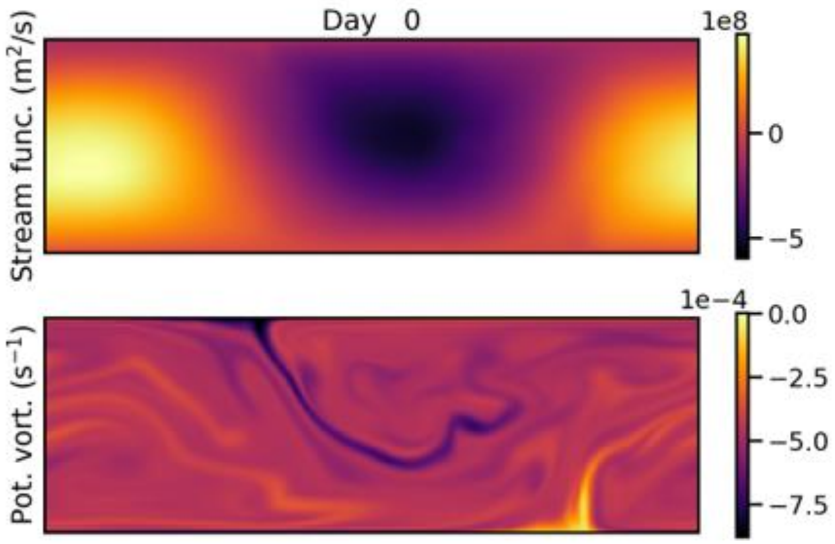
$$\hat{\mathbf{B}} \coloneqq \hat{\mathbf{B}}_1^{(1)} + \sum_{\ell=2}^{L} \left\{ \hat{\mathbf{B}}_\ell^{(\ell)} - \hat{\mathbf{B}}_{\ell-1}^{(\ell)} \right\}$$

Mycek, & De Lozzo (2019). Multilevel Monte Carlo Covariance Estimation for the Computation of Sobol' Indices. *SIAM/ASA Journal on Uncertainty Quantification*, *7*(4), 1323–1348. 10.1137/18M1216389
Destouches, M., Mycek, P., & Gürol, S. (2023). *Multivariate extensions of the Multilevel Best Linear Unbiased Estimator for ensemble-variational data assimilation*. http://arxiv.org/abs/2306.07017

Outline

- Introduction and theory
- **Experiments with a quasi-geostrophic model**
- Conclusion

Experiments with OOPS-JEDI quasi-geostrophic model



- 2-layer model.
- $79 \times 240 \times 2$ grid points
- About 60 positive Lyapunov exponents
- Model variable: **stream function** or potential vorticity.

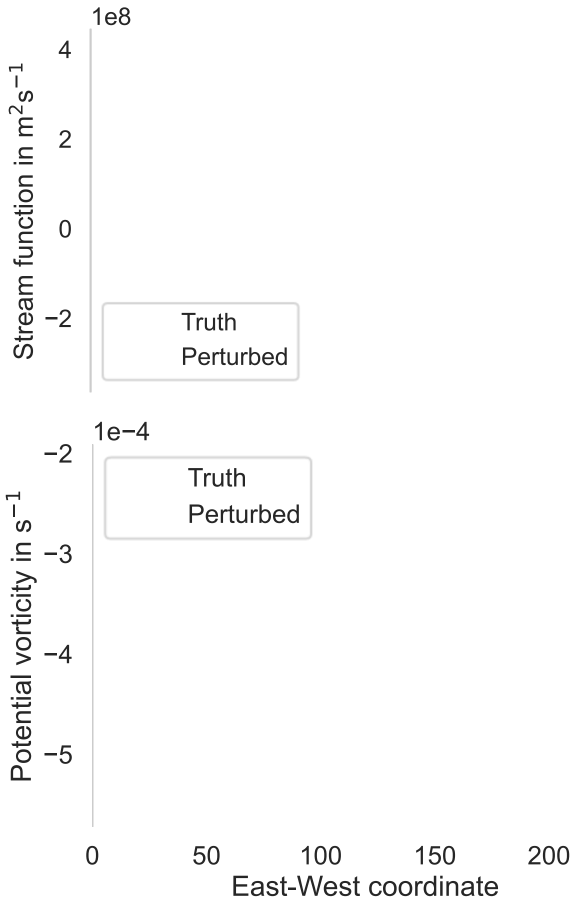
Ensemble generation

- Background generation from truth run:

$$\mathbf{x}_\mathrm{b} = \mathbf{x}_\mathrm{t} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$$

- Ensemble generation from 12h forecasts with perturbed initial conditions:

$$\mathbf{x}_k^\mathrm{f} = \mathcal{M}(\mathbf{x}_\mathrm{b} + \boldsymbol{\epsilon}_k), \qquad \boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$$



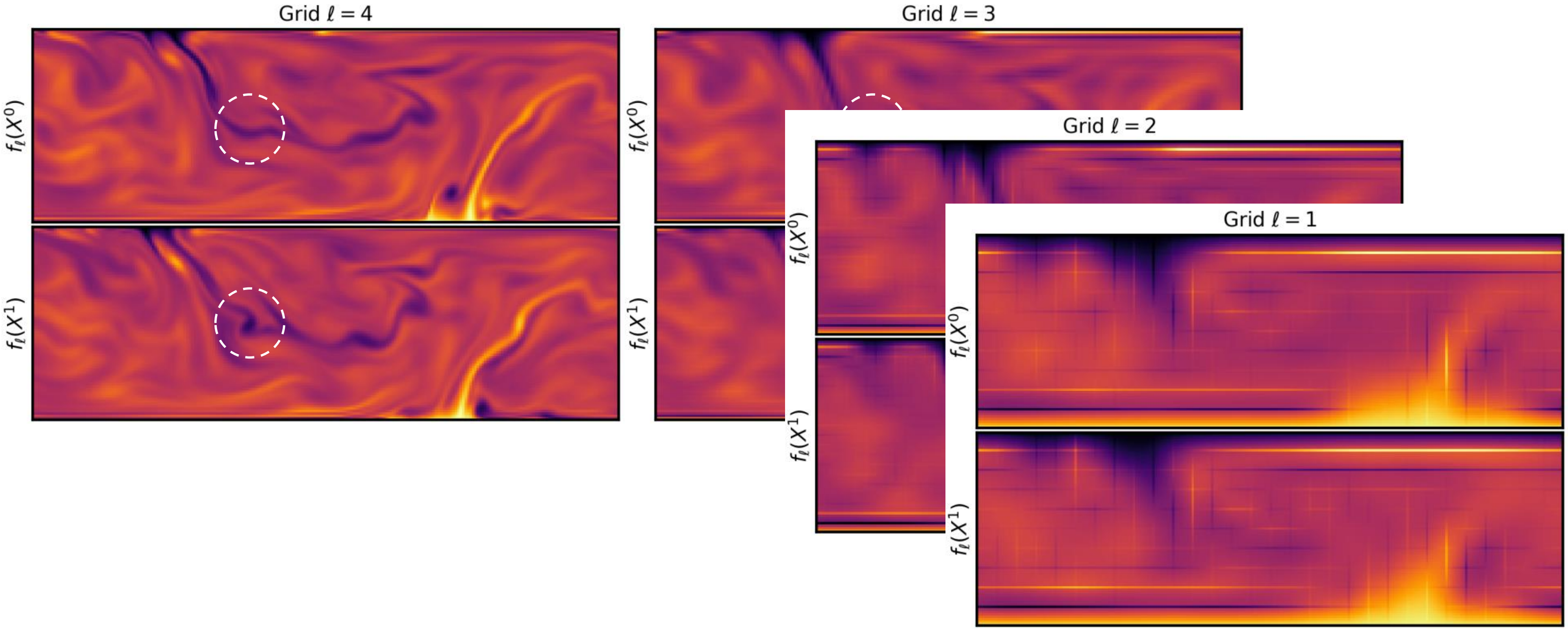
Introducing a hierarchy of low-fidelity models

- Forecasts run on nested grids with half horizontal resolution:
  - $\ell = 4$: 79×240
  - $\ell = 3$: 39×120
  - $\ell = 2$: 19×60
  - $\ell = 1$: 9×30
- Sample generation on grid ℓ:

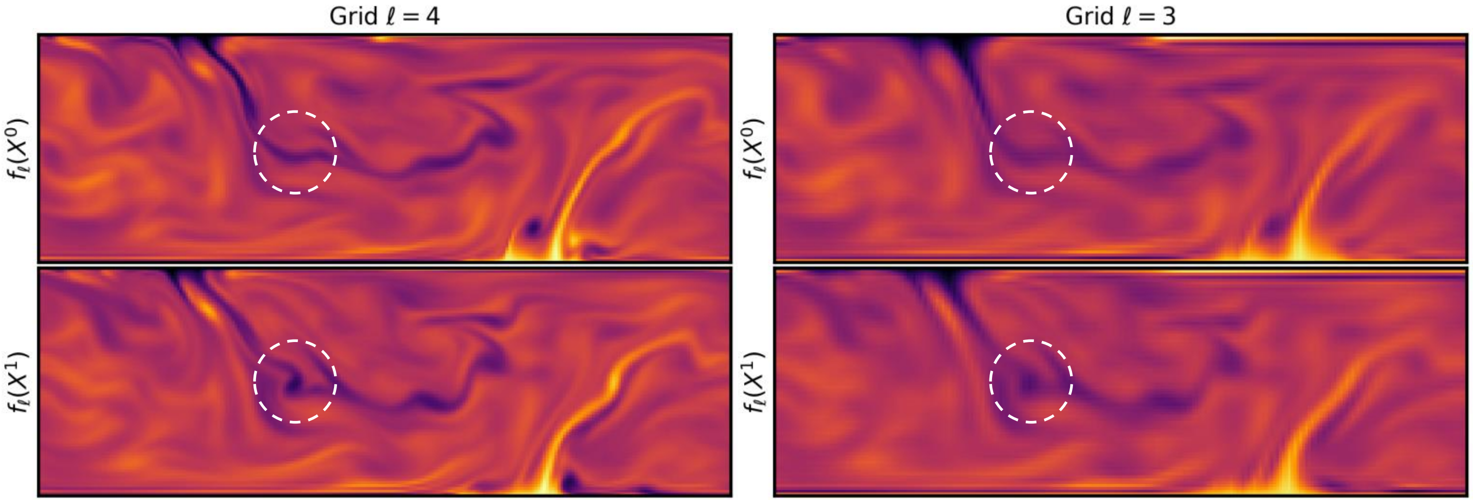
$$\mathbf{T}_{L \leftarrow \ell}\left(\mathcal{M}_{\ell}\left(\mathbf{T}_{\ell \leftarrow L}(\mathbf{x}_{\mathrm{b}} + \boldsymbol{\epsilon}_k)\right)\right)$$

Bicubic interpolation operators

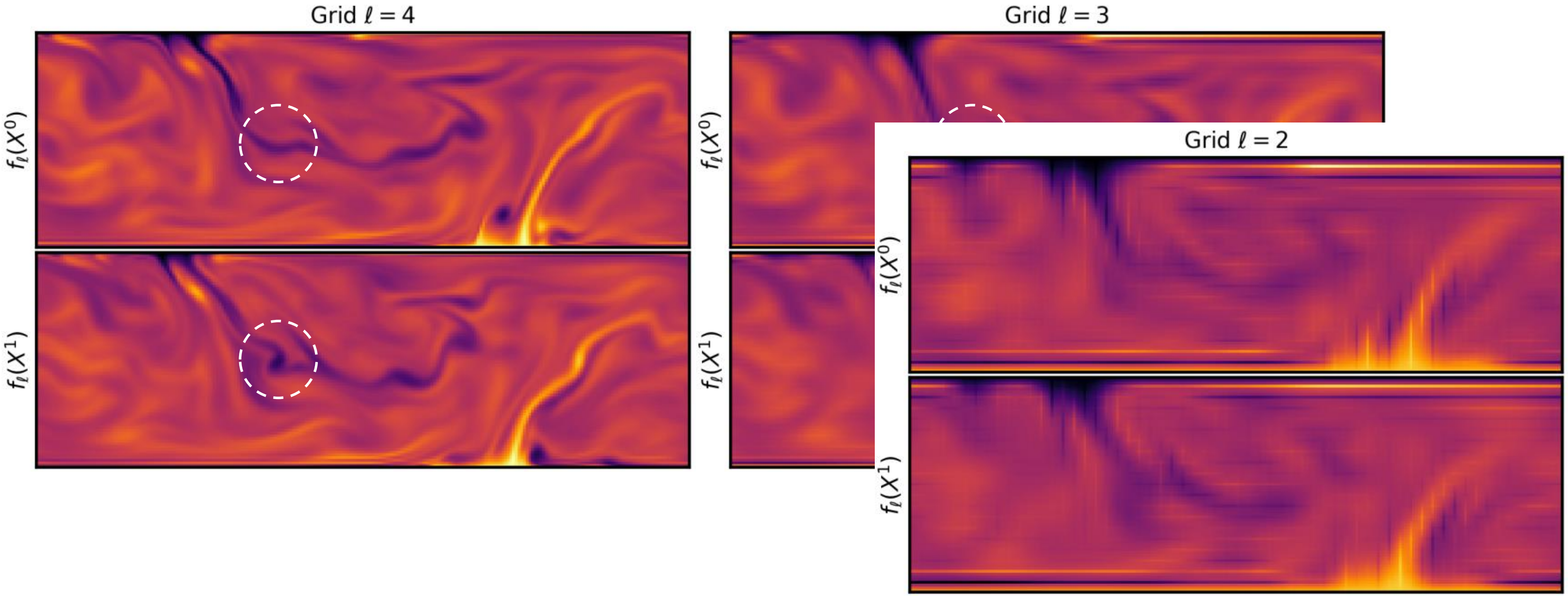
Ensemble spread & inter-level coupling



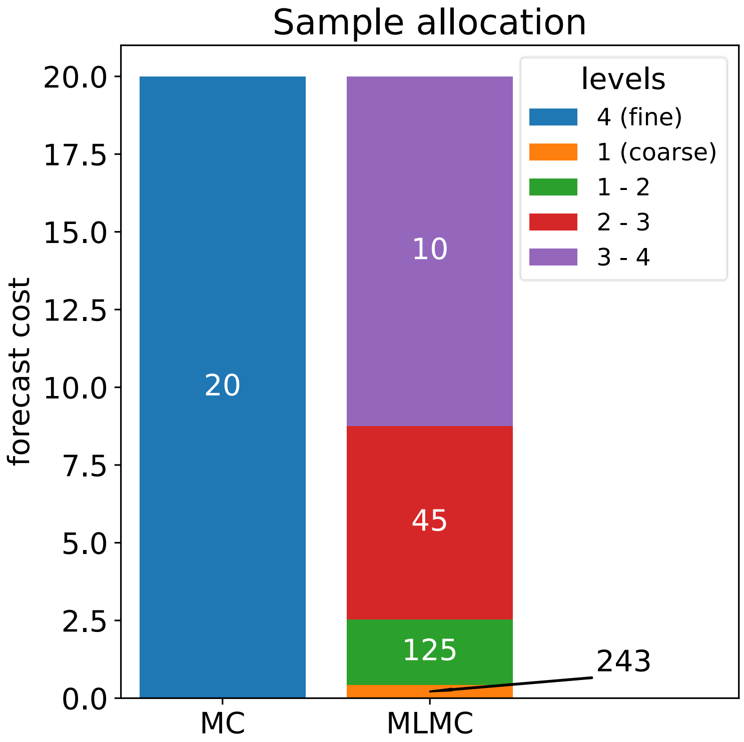
Ensemble spread & inter-level coupling

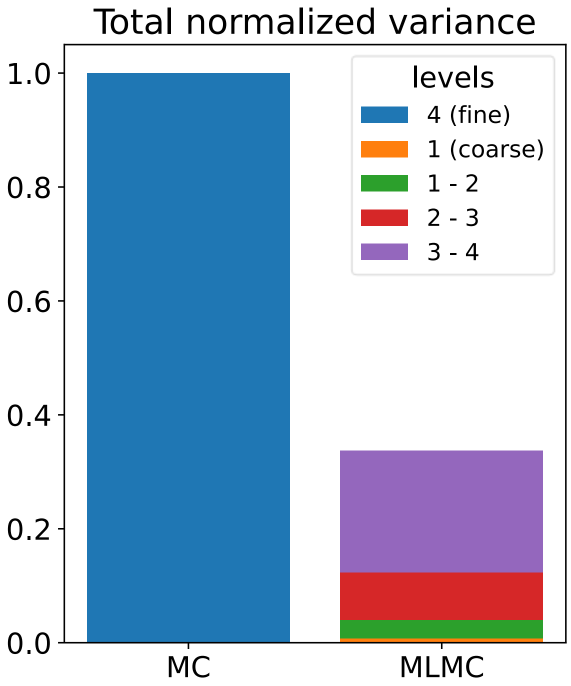


Ensemble spread & inter-level coupling



Optimal member allocation and theoretical gain

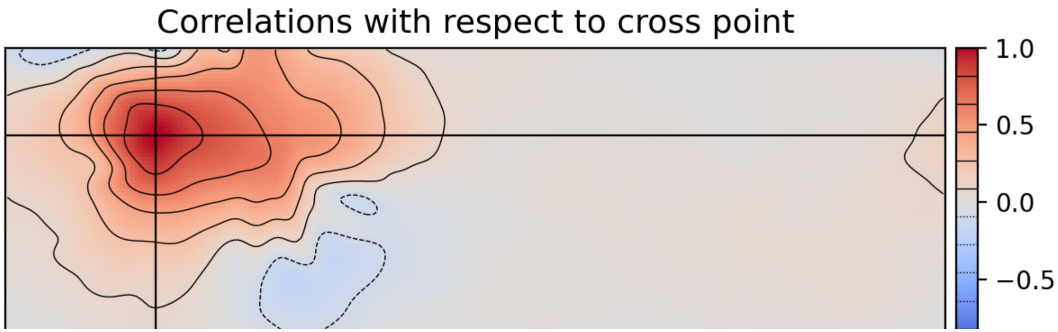


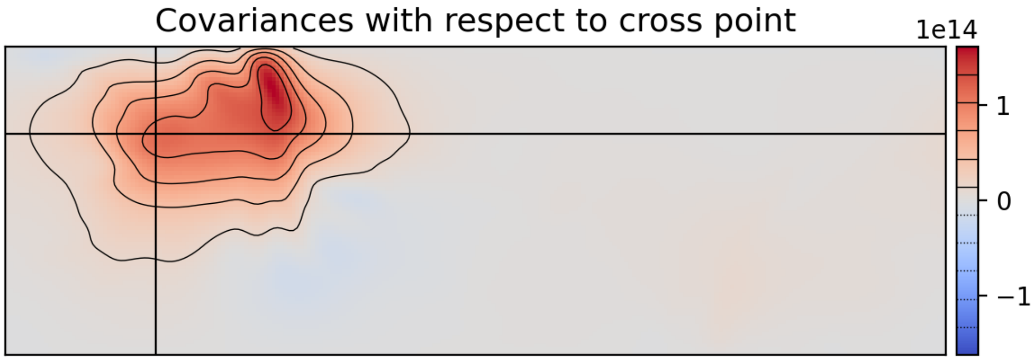


Variance reduced by 3:
Tripled effective ensemble size

Empirical impact on B estimation

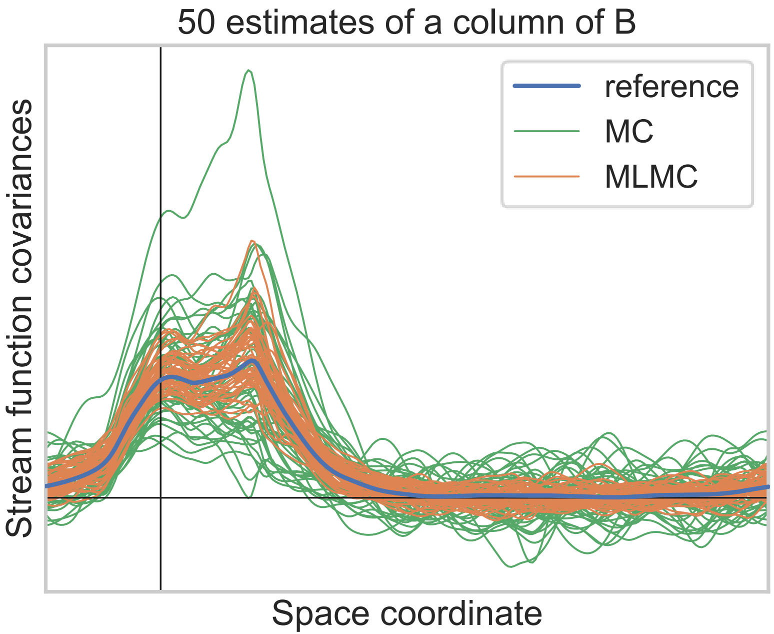
Focus on a single column of B

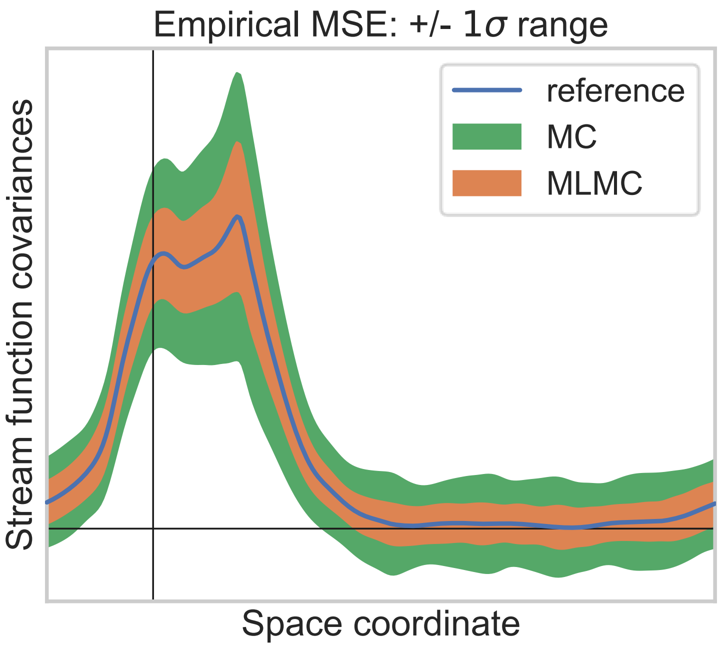




Reference estimated with MC and 10,000 members

MC vs MLMC


50 estimates of a column of B
reference
MC
MLMC
Stream function covariances
Space coordinate


Empirical MSE: +/- 1σ range
reference
MC
MLMC
Stream function covariances
Space coordinate

Empirical impact on the analysis

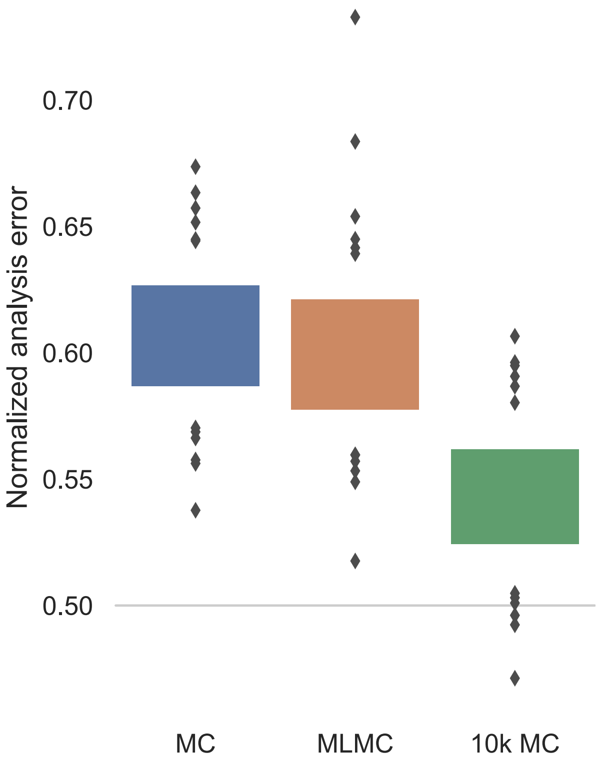
Problem: Multilevel B has negative eigenvalues…

- Possible solutions?
  - Make it positive semidefinite (PSD) by construction? (Maurais *et al.*, 2023)
  - Get nearest PSD matrix by removing negative eigenvalues? (Hoel *et al.*, 2016)
  - Rebuild B with absolute values of the eigenvalues?
  - Hybridization?
  - …
- Can we use a non-PSD B? The cost function is not convex anymore…

Maurais, Alsup, Peherstorfer, & Marzouk (2023). *Multifidelity Covariance Estimation via Regression on the Manifold of Symmetric Positive Definite Matrices*. http://arxiv.org/abs/2307.12438
Hoel, Law & Tempone *Multilevel ensemble Kalman filtering*. https://doi.org/10.1137/15M100955X

Results

- 10 iterations of a **B-Preconditioned Conjugate Gradient, with non-PSD checks for MLMC**
  - Algorithm stops when facing a negative B-“norm”
  - After minimization, back-track to iteration with minimum residual.
- Independent localization tuning for MC and MLMC.
- 200 analyses for each experiment
  - Random observation networks, 1% grid points observed.
  - Random observation errors
  - Random ensembles for MC and MLMC, both with computational cost 20 fine simulations
- Best achievable performance given by (unlocalized) MC with 10,000 members



Outline

- Introduction and theory
- Experiments with a quasi-geostrophic model
- **Conclusion**

Conclusion

- Multilevel Monte Carlo methods provide a way to leverage low-fidelity simulators **without introducing bias**
- Experiments on a QG model show **improved reconstruction** of an ensemble background error **covariance matrix**.
- Using this ensemble background error covariance matrix can improve the **accuracy of the analysis**.

- Non-PSDness of the multilevel covariance matrix still to understand and tackle…
- Experiments were based on stochastic coupling from initial conditions. In a cycled EDA, stochastic coupling would be weaker.
- Impact on subsequent forecasts not tested
- MLMC would impact EPS design

Other possible applications…

- Lower fidelity models could also be ML surrogates, runs at single or half precision, runs with simplified physics, with hydrostatic assumptions…
- Lower fidelity can come from the analysis step: simplified scheme, with less iterations, less observations…

- Applications to other statistics possible: variance field, mean field, quantiles, probability of exceeding a threshold, pdfs…

Correct way to build multi-fidelity estimators:
Schaden & Ullmann (2020). On Multilevel Best Linear Unbiased Estimators. *SIAM/ASA Journal on Uncertainty Quantification*, *8*(2), 601–635. doi:10.1137/19M1263534

mayeul.destouches@metoffice.gov.uk

Backup slides

Adding localization to a multilevel B

- Localization should be applied independently to each term

$$L \circ \hat{B} = L \circ \widehat{B_1}^{(1)} + \left(L \circ \widehat{B_2}^{(2)} - L \circ \widehat{B_1}^{(2)}\right) + \cdots$$

- Cost still reasonable if applied on reduced space, *e.g.*, before interpolation:

$$L_{\text{fine}} \circ \ (PX)(PX)^t \approx P(L_{\text{coarse}} \circ XX^t)P^t$$

- Having different localizations introduces additional bias. But this is what localization does… An optimal localization theory can be devised (Destouches *et al.*, 2023)

Destouches, M., Mycek, P., & Gürol, S. (2023). *Multivariate extensions of the Multilevel Best Linear Unbiased Estimator for ensemble-variational data assimilation*. http://arxiv.org/abs/2306.07017

Allocating samples

- Costs assumed proportional to number of grid points and time steps
- $\mathcal{V}_\ell$ estimated from 100 members coupled across all levels
- Optimal sample allocation with weighted MLMC