

EGU22-336

<https://doi.org/10.5194/egusphere-egu22-336>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Multilevel Monte Carlo estimation of background error covariances in ensemble variational data assimilation

Mayeul Destouches^{1,2}, Paul Mycek^{1,2}, Jérémy Briant^{1,3}, Selime Gürol^{1,2}, Anthony Weaver^{1,2}, Serge Gratton³, and Ehouarn Simon³

¹Cerfacs, Toulouse, France

²CECI-CNRS UMR 5318, Cerfacs, Toulouse, France

³IRIT-CNRS UMR 5505, INP, Université de Toulouse, Toulouse, France

In ensemble variational (EnVar) data assimilation systems, background error covariances are sampled from an ensemble of forecasts evolving with time. One possible way of generating this ensemble is by running an Ensemble of Data Assimilations (EDA) that samples all possible error sources (initial condition errors, boundary condition errors, model errors). Large ensemble sizes are desirable to minimize sampling errors, but generating a single ensemble member is usually expensive due to the cost of integrating the physical model. In practice, ensembles with coarser spatial resolutions are sometimes used, allowing for cheaper generation of individual members, and thus larger ensemble sizes.

Multilevel Monte Carlo (MLMC) methods propose to go beyond this usual trade-off between grid resolution and ensemble size, by expressing a fine-grid estimator as an astute combination of estimators computed on a hierarchy of spatial grids. Starting from a Monte Carlo covariance estimator on a coarse grid but with a large ensemble size, correction terms are added to form a quasi-telescopic sum. The correction terms come from EDAs of increasing spatial resolutions and decreasing ensemble sizes, with a pairwise stochastic coupling between EDAs of two successive resolutions. The expectation of this MLMC estimator is equal to the expectation of the Monte Carlo estimator on the finest grid, so that no bias is introduced by the coarse resolution forecasts. Without increasing the computational cost, MLMC effectively reduces the variance of the covariance estimator, *i.e.* reduces the sampling noise on covariances.

We first present the theoretical basis of MLMC and how it can apply to the estimation of covariance matrices. An illustration with a quasi-geostrophic model is then presented. For a given computational budget, we compare three equal-cost methods to estimate background error covariances: (1) the usual single-resolution ensemble estimate, (2) a combination of estimates of various resolutions based on Bayesian Model Averaging and (3) the MLMC estimate. The methods are compared in terms of mean square error of the covariance estimators, and in terms of quality of the resulting analyses for one assimilation cycle. The role of covariance localization in each case is also briefly discussed.

This work is partially supported by 3IA Artificial and Natural Intelligence Toulouse Institute,

French "Investing for the Future- PIA3" program under the Grant agreement ANR-19-PI3A-0004.

This project has received financial support from the CNRS through the 80Prime program.