FINAL REPORT PROGRAM LEFE

Program: LEFE/MANU	**Project Title:** Mathematical and Algorithmic aspects of Diffusion-based Correlation OPerators (MADCOP)	**Years:** 2017-2019

PI: A. Weaver, weaver@cerfacs.fr, CERFACS/CECI CNRS UMR 5318 **Participating Laboratories**: CNRM UMR 3589; IRIT UMR 5505; INRIA/LJK UMR 5224; ISAE-SUPAERO	**Contribution to**: *OceanPredict; operational activities at ECMWF, UK Met Office, Météo-France* **Other funding sources**: FP7 (ERA-CLIM2); C3S

Context: This project aimed at advancing research and development in diffusion modelling for representing correlations of background and observation error in ensemble-variational data assimilation systems. Methods developed in this project have been implemented and evaluated in global ocean configurations of the NEMOVAR system, which is used operationally at ECMWF and the UK Met Office.

Objectives / scientific questions: This project addressed three topics, which were the subject of separate publications in the *Quarterly Journal of the Royal Meteorological Society* and a PhD thesis. The first topic aimed to improve parallel aspects of the diffusion algorithm (*Weaver et al. 2018*). The second topic aimed to develop methods for estimating normalizing factors for converting diffusion-generated covariances into (unit-amplitude) correlations (*Weaver et al. 2021*). The third topic involved adapting the diffusion technique to unstructured grids, in view of using it to represent spatially correlated observation errror (*Guillet 2010; Guillet et al. 2019*).

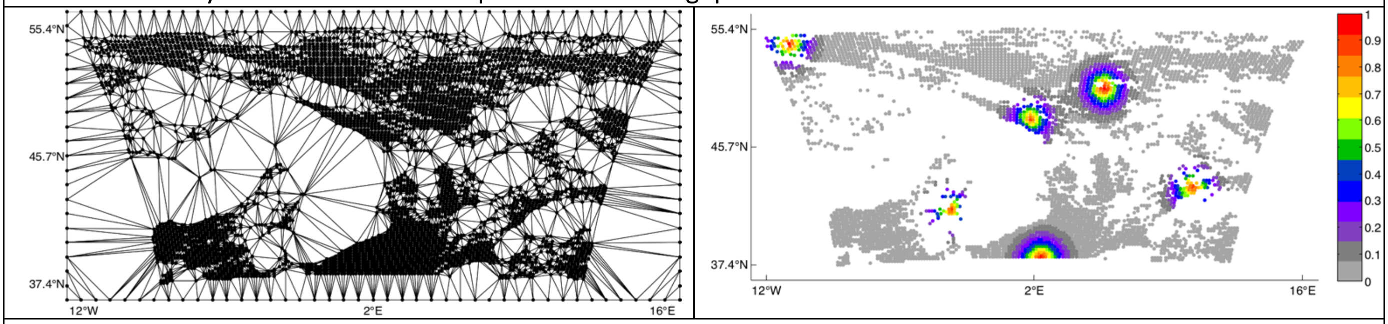
Main results:

We have reproduced below the abstracts from the three articles resulting from this project.

1) ***From Weaver et al. (2018)***. Correlation operators based on the solution of an implicitly formulated diffusion equation can be implemented numerically using the Chebyshev iteration method. The attractive properties of the algorithm for modelling correlation functions on high-performance computers have been discussed in a recent paper. The current paper describes a straightforward variant of that algorithm that allows the matrix–vector products involved in the sequential pseudo-time diffusion process to be performed in parallel. Contrary to the original algorithm, which requires solving, in sequence, linear systems involving a symmetric positive-definite (SPD) matrix, the “time”-parallel algorithm requires solving a single linear system involving a non-symmetric positive-definite (NSPD) matrix. The key information required by the Chebyshev iteration for solving the NSPD problem is an estimate of the extreme eigenvalues of the NSPD matrix. For the problem under consideration, the extreme eigenvalues of the NSPD matrix are the same as those of the original SPD matrix, and can be pre-computed using a Lanczos algorithm applied to the latter. The convergence properties of the algorithm are studied from a theoretical perspective and using numerical experiments with a diffusion-based covariance model in a variational data assimilation system for the global ocean. Results suggest that time-parallelization can reduce the run-time of an implicit diffusion-based correlation operator by greater than a factor of two. It can be implemented practically using a hybrid parallelization approach that combines MPI tasks in the spatial domain with Open Multi-Processing threads spanning the pseudo-time dimension. The sensitivity of the results to preconditioning, to the choice of first guess and to the stopping criterion is discussed.
2) ***From Weaver et al. (2021)***. Developing effective ways to model and cycle the background-error covariance matrix is an active area of research in data assimilation. An important aspect of this problem when using a filter to model the background-error correlations is the computation of normalization factors to ensure that the diagonal elements of the modelled correlation matrix are all equal to one. Updating the parameters of a flow-dependent correlation model on each assimilation cycle requires updating the normalization factors, which is costly using traditional methods such as randomization. In this article, we discuss the normalization problem within the context of a diffusion filter-based covariance model used for background-error modelling in a variational data assimilation system for the global ocean. We evaluate various methods for estimating normalization factors when the diffusion tensor of the correlation model is derived from an ensemble of ocean states. Our results show that estimates produced using inexpensive methods derived from analytical considerations of the diffusion equation can have significant errors, especially near boundaries. Estimates obtained using randomization with a small sample size (∼100) are more accurate in a globally averaged sense but are noisy and can have unacceptably large errors locally. Next, we focus on the specific problem of accounting for flow-dependent correlation parameters in the vertical component of the diffusion operator only, which is especially important near the surface for the assimilation of sea surface temperature observations. Remarkably accurate estimates are obtained by approximating the normalization matrix as a separable

product of two normalization matrices: one computed using randomization with the horizontal diffusion operator only and the other computed using randomization with the vertical diffusion operator only. If the parameters of the horizontal component of the diffusion operator are static, then only the normalization factors of the flow-dependent vertical component need to be recomputed on each cycle. This result is of significant practical interest since the vertical diffusion operator employs an inexpensive direct solver and thus can be applied on each cycle with a large random sample to obtain a good approximation of the normalization matrix.

3) ***From Guillet et al. (2019)***. We propose a method for representing spatially correlated observation errors in variational data assimilation. The method is based on the numerical solution of a diffusion equation, a technique commonly used for representing spatially correlated background errors. The discretization of the pseudo-time derivative of the diffusion equation is done implicitly using a backward Euler scheme. The solution of the resulting elliptic equation can be interpreted as a correlation operator whose kernel is a correlation function from the Matérn family. In order to account for the possibly heterogeneous distribution of observations, a spatial discretization technique based on the finite element method (FEM) is chosen where the observation locations are used to define the nodes of an unstructured mesh on which the diffusion equation is solved. By construction, the method leads to a convenient operator for the inverse of the observation-error correlation matrix, which is an important requirement when applying it with standard minimization algorithms in variational data assimilation. Previous studies have shown that spatially correlated observation errors can also be accounted for by assimilating the observations together with their directional derivatives up to arbitrary order. In the continuous framework, we show that the two approaches are formally equivalent for certain parameter specifications. The FEM provides an appropriate framework for evaluating the derivatives numerically, especially when the observations are heterogeneously distributed. Numerical experiments are performed using a realistic data distribution from the Spinning Enhanced Visible and InfraRed Imager (SEVIRI). Correlations obtained with the FEM-discretized diffusion operator are compared with those obtained using the analytical Matérn correlation model (see Figures). The method is shown to produce an accurate representation of the target Matérn function in regions where the data are densely distributed. The presence of large gaps in the data distribution degrades the quality of the mesh and leads to numerical errors in the representation of the Matérn function. Strategies to improve the accuracy of the method in the presence of such gaps are discussed.



Left panel: Triangular mesh built from the locations of SEVIRI radiance measurements. **Right panel**: Correlations at selected locations on the mesh generated using a diffusion operator solved with the finite element method. From Guillet (2019) and Guillet et al. (2019).

Future of the project*:* The diffusion-based correlation operator is a critical component of ensemble-variational data assimilation applications with NEMOVAR. Continuous work is needed to revise the solution algorithms of the diffusion operator in order to improve computational performance (efficiency and scalability) in response to increased resolution of the ocean model, the use of more sophisticated correlation models, and changes in computer architecture. The work conducted in this project to use the diffusion technique to model spatially correlated observation error has opened up exciting new possibilities for assimilating a much larger number of observations from satellites in both ocean and atmospheric data assimilation.

1. Weaver, A. T., M. Chrust, B. Ménétrier and A. Piacentini, 2021: An evaluation of methods for normalizing diffusion-based covariance operators in variational data assimilation. *Q. J. R. Meteorol. Soc.*, in press. https://doi.org/10.1002/qj.3918
2. Guillet, O., Weaver, A. T., Vasseur, X., Michel, Y., Gratton, S and Gürol, S., 2019: Modelling spatially correlated observation errors in variational data assimilation using a diffusion operator on an unstructured mesh. *Q. J. R. Meteorol. Soc.*, **145**, 1947-1967, https://doi.org/10.1002/qj.3537
3. Guillet, O., 2019. Modélisation des corrélations spatiales d'erreurs d'observation en assimilation de données variationnelle. Étude sur des maillages non structurés. PhD thesis, Université de Toulouse.
4. Weaver, A. T., Gürol, S., Tshimanga, J., M. Chrust and A. Piacentini, 2018: "Time"-parallel diffusion-based correlation operators. *Q. J. R. Meteorol. Soc.*, **144**, 2067–2088, https://doi.org/10.1002/qj.3302