Internship 2024

Accounting for horizontal and vertical observation error correlations in data assimilation

1 Title and keywords

Accounting for horizontal and vertical observation error correlations in data assimilation.

Keywords : Variational data assimilation and methods, observations treatment in meteorology, diffusion-based correlation operators, interchannel correlations, finite element method.

2 General information

Supervisor : Oliver Guillet, researcher at Météo - France / CNRM / GMAP / ASSIM

Laboratory : CNRM UMR 3589

Location : Toulouse, France

Duration : 6 months starting in 2024

Stipend is covered by Météo-France and the LEFE-MANU national program.

Possible accomodation is available on-site at the start of the internship.

3 Technical requirements

Language requirements : The candidate will need a good level in English (C1 on TOEFL score). Although French is appreciated, it is not required for the internship.

Scientific requirements :

- Master level in geoscience or atmospheric science, with notions of data assimilation;
- Solid background in numerical analysis, PDEs discretization and optimization methods;
- Programming in either Python or C++. Development under Linux/Shell.

Recommendation :

- Teamwork and communication skills;
- A certain taste for small mathematical problems.

4 General context

The assimilation of high spatial density data has been identified as one of the major challenges of our decade in improving convective-scale numerical weather prediction models. Thus, the improvement of the limited-area model AROME using massive data has been cited as a priority research axis of Météo France in the 2022 Objectives and Performance Contract.

This challenge goes hand in hand with the inauguration of the MTG-IRS sounder on board the Third Generation Satellites in 2024, which will have 1960 acquisition channels coupled with a maximum horizontal resolution of 4km. The arrival of this new instrument will provide the best initial conditions for numerical weather prediction models. However, it will also be accompanied by a number of challenges related to the storage and analysis of this data.

A major challenge lies in the statistical representation of observation errors in operational data assimilation systems. In particular, the quality and performance of high spatial density data assimilation directly depend on the representation of observation error correlations in our systems. This research theme is the subject of many studies in Europe, with Météo France being one of the pioneers.

To meet the challenge of taking into account observation error correlations in data assimilation for meteorology, we propose to conduct a set of studies around this same theme. Ongoing and upcoming studies include :

- **Theme 1 : Accounting for horizontal and vertical observation error correlations in data assimilation** - available for internship.

- Theme 2 : A cheap version of the square-root diffusion operator for generating observation perturbations - available for internship.

- Theme 3 : Parallelization strategies for the 2d diffusion equation with applications to data assimilation - available for internship.

- Theme 4 : Operational implementation of a bidimensional, diffusion-based observation error correlation model in the limited-are model AROME - Phd starting in september 2023.

5 Scientific introduction

Data assimilation places observations at the heart of weather forecasting. With the use of increasingly advanced instruments, finer scales can be represented in the analysis, leading to improved forecast quality. As such, the ability to assimilate a growing amount of data is a fundamental challenge in data assimilation.

However, this expansion is hindered by the need to accurately represent observation error correlations. Some observations are affected by spatially and spectrally dependent errors, meaning that measurements that are closer together are more correlated. These correlations must be accounted for in data assimilation equations using an appropriate correlation model, or the analysis quality will decrease.

Recent studies have revealed significant observation error correlations in remotely sensed datasets such as AMVs (Bormann, 2002), radar reflectivities and radial winds (Waller et al., 2016a), and hyperspectral infrared sounder radiances (Bormann et al., 2010). These studies highlight the importance of accounting for observation error correlations to improve the use of observations in future data assimilation systems (see Liu and Rabier, 2003).

In the case of radiances, it is common to distinguish between inter-channel correlations, also known as « vertical » correlations, and horizontal correlations. Weston et al. (2014) recently demonstrated that inter-channel correlations can be easily accounted for by using small matrices, typically with a few hundred rows. These matrices can be easily inverted numerically, for example using the Cholesky decomposition. A similar technique was previously used to model temporal correlations (Järvinen et al. (1999)).

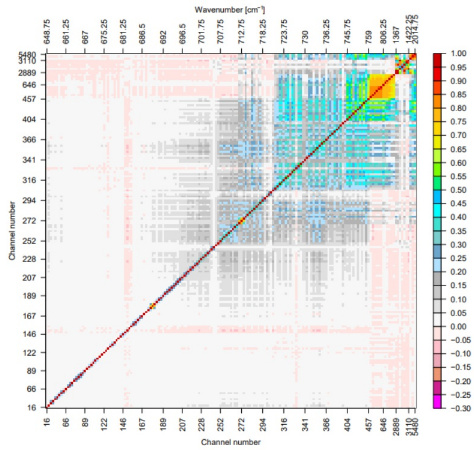

Figure 1. Typical interchannel correlation matrix estimated from the Seviri infrared sounder (Waller et al. 2016).

Representing spatial correlations, on the other hand, is more complex (see Michel (2018) for an in-depth discussion). This is due to the often irregular spatial distribution of data, which greatly complicates the structure of the $\boldsymbol{R}$ matrix (see O. Guillet's thesis). As a result, only a small fraction of the available data (from 1% to 15%) is typically assimilated to avoid correlations. While this results in a diagonal $\boldsymbol{R}$ matrix, it is not an ideal solution as it underutilizes the available observations. However, Guillet et al. (2019) developed a finite element model based on the diffusion equation to model horizontal observation error correlations using data from the Seviri radiometer. Their method allows for direct modeling of the inverse $\boldsymbol{R}$ matrix. This study will serve as a foundation for this project.

6 Internship objectives

The aim of this internship is to advance our understanding of how to model observation error correlations by working with real data. Observations come in many forms, and the sources of error correlations can be complex. As such, the first issue we will address is finding the best statistical model to accurately represent three-dimensional spatial observation error correlations. Building on the work of Guillet et al. (2019) and Weaver and Mirouze (2010), we plan to use a model based on the diffusion equation

$$(1-\ell^2\Delta)^m u = f$$

to represent horizontal correlations.

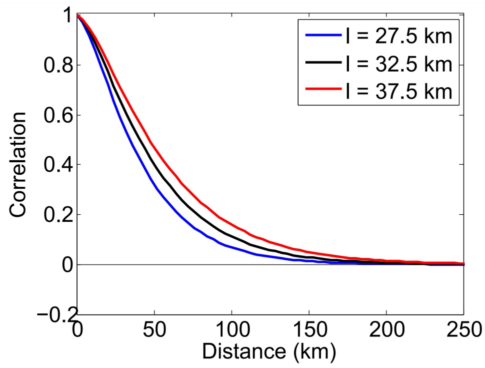

Figure 2. One-dimension correlation function of the Matérn class and their sensitivity to the lengthscale parameter.

We will explore the possibility of extending Guillet et al.'s (2019) model to three dimensions by discretizing the diffusion equation on volumetric meshes. This will allow us to account for both horizontal and vertical correlations without having to separate them. We will compare this approach to the 2D+1D strategy, which treats the vertical dimension separately from the horizontal but requires observations to have a homogeneous vertical structure.

Several types of observations will be considered in this study, including data from the Seviri radiometer on board the geostationary MSG satellite, meteorological radar data, and surface observations. Depending on our progress and results, we may explore some or all of these options.

7 Methodology

The study begins by examining the error correlations present in the observations. The first step is to diagnose the correlations of Seviri and radar, and surface observation data measurements at different resolutions. Then we find a descriptive model that captures their main statistical characteristics. To do this, we will consider Matérn functions, which are the kernels of the diffusion equation. The parameters of the Matérn model will be determined using a diagnosis based on the outputs from AROME (Desroziers et al, 2005).

Next, we will study the geometric properties of observations and the associated meshes for one type of data, starting with an existing python model designed for Seviri data. These meshes will be used to discretize the diffusion equation using the finite element method.

We will consider two cases: first, when the horizontal structure of observations is homogeneous along the vertical (*i.e.*, each latitude/longitude is paired with a vertical column). In this case, it makes sense to separate horizontal (2D) and vertical (1D) correlations. Second, when the quality control results in an observation structure that does not fit the first scenario. In this case, we will need to use a « full-3D » finite element approach by constructing a mesh that covers all vertical levels at once.

To compare the two approaches, we will evaluate how well the numerical model can reproduce correlations diagnosed from the real data.

8 Expected results

Python code for 3D correlations, written essay and public presentation of the results.

9 References

Guillet O, Weaver A T, Vasseur X, Michel Y, Gratton S, Gürol S. 2019. Modelling spatially correlated observation errors in variational data assimilation using a diffusion operator on an unstructured mesh. Q. J. R. Meteorol.

Guillet O. Thèse. 2019. Modélisation des corrélations spatiales d'erreurs d'observation en assimilation de données variationnelle. Etude sur des maillages non structurés.

Michel Y. 2018. Revisiting Fisher's approach to the handling of horizontal spatial correlations of the observation errors in a variational framework. Q. J. R. Meteorol.

Stewart LM, Dance SL, Nichols. 2013. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. Tellus.

Liu ZQ, Rabier F. 2003. The potential of high-density observations for numerical weather prediction: a study with simulated observations. Q. J. R. Meteorol.