Tailoring data assimilation to discontinuous Galerkin models

Article

Published Version

Pasmans, I. ORCID: https://orcid.org/0000-0001-5076-5421, Chen, Y. ORCID: https://orcid.org/0000-0002-2319-6937, Carrassi, A. ORCID: https://orcid.org/0000-0003-0722-5600 and Jones, C. K. R. T. (2024) Tailoring data assimilation to discontinuous Galerkin models. Quarterly Journal of the Royal Meteorological Society. ISSN 0035-9009 doi: 10.1002/qj.4737 Available at https://centaur.reading.ac.uk/116511/

To link to this article DOI: http://dx.doi.org/10.1002/qj.4737

Publisher: Wiley

Received: 4 May 2023 | Revised: 16 February 2024 | Accepted: 8 April 2024

DOI: 10.1002/qj.4737

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Tailoring data assimilation to discontinuous Galerkin models

Ivo Pasmans[1] | **Yumeng Chen**[1] | **Alberto Carrassi**[1,2] | **Chris K. R. T. Jones**[3]

[1]Department of Meteorology, University of Reading, Reading, UK

[2]Department of Physics "Augusto Righi", Universitá di Bologna, Bologna, Italy

[3]Mathematics Department, University of North Carolina, Chapel Hill, North Carolina, USA

Correspondence
Ivo Pasmans, Department of Meteorology, University of Reading, Brian Hoskins Building, Reading, RG6 6ET, UK.
Email: i.c.pasmans@reading.ac.uk

Funding information
Schmidt Sciences

Abstract

In recent years discontinuous Galerkin (DG) methods have received increased interest from the geophysical community. In these methods the solution in each grid cell is approximated as a linear combination of basis functions. Ensemble data assimilation (DA) aims to approximate the true state by combining model outputs with observations using error statistics estimated from an ensemble of model runs. Ensemble data assimilation in geophysical models faces several well-documented issues. In this work we exploit the expansion of the solution in DG basis functions to address some of these issues. Specifically, it is investigated whether a DA–DG combination (a) mitigates the need for observation thinning, (b) reduces errors in the field's gradients, and (c) can be used to set up scale-dependent localisation. Numerical experiments are carried out using stochastically generated ensembles of model states, with different noise properties, and with Legendre polynomials as basis functions. It is found that strong reduction in the analysis error is achieved by using DA–DG and that the benefit increases with increasing DG order. This is especially the case when small scales dominate the background error. The DA improvement in the first derivative is, on the other hand, marginal. We think this to be a counter-effect of the power of DG to fit the observations closely, which can deteriorate the estimates of the derivatives. Applying optimal localisation to the different polynomial orders, thus exploiting their different spatial length, is beneficial: it results in a covariance matrix closer to the true covariance than the matrix obtained using traditional optimal localisation in state space.

KEYWORDS

data assimilation, discontinuous Galerkin, localisation, neXtSIM$_{\mathrm{DG}}$, numerical methods

1 INTRODUCTION

Discontinuous Galerkin (DG) methods produce gridded numerical solvers with a mix of characteristics that are ideally suited to large geophysical fluid models. They combine several advantageous features from other numerical approaches like finite-difference, finite-volume, and finite-element methods. First, they conserve tracer volumes, mass, and momentum (though not necessarily energy). Second, owing to the discontinuity of the solution at the grid cell edges, numerical stencils involve only face neighbours. These stencils are smaller than those in continuous Galerkin methods, in which the stencils can also involve vertex neighbours, and higher order finite-volume methods, which use stencils stretching over multiple grid cells. Consequently, DG methods can be parallelised more efficiently. Third, as discontinuities are allowed at the grid cell face it is relatively straightforward to merge/divide cells (*h*-adaptivity) and/or vary the number of basis functions (*p*-adaptivity) on a cell-by-cell basis (Marras *et al.*, 2016). Fourth, as higher order DG models allow the solution to vary within grid cells, discontinuities at the grid cell edges are smaller than those in finite-volume models. This allows for better approximations of the continuous fields found in geophysical fluid dynamics. Finally, for sufficiently smooth solutions—for example, the solutions one customarily encounters in geophysical models in regions with at least some degree of turbulent dissipation—increasing the polynomial order turns out to be a computationally more efficient way to reach a desired accuracy than reducing the grid cell size (Vos *et al.*, 2010).

The potential benefits of DG solvers have not escaped the attention of the geophysical community. Over the last few years, several numerical geophysical prediction (research) systems have been developed/are under development that use (partial) DG methods in the computational core. Examples of such efforts are the coastal ocean model Thetis (Kärnä *et al.*, 2018), the US Navy/Air Force NUMA atmospheric model (Giraldo, 2011), the soil water flow model DORiE (Riedel *et al.*, 2020), or the sea-ice model neXtSIM$_{\mathrm{DG}}$ being developed as part of the Scale-Aware Sea Ice Project (https://sasip-climate.github.io/). The latter development is the direct motivation for the work laid out in this article. neXtSIM$_{\mathrm{DG}}$ uses the Maxwell-elasto-brittle rheology and thermodynamics from its ancestor, the neXtSIM sea-ice model (Bouillon & Rampal, 2015; Rampal *et al.*, 2016; Samaké *et al.*, 2017), but replaces the original finite-element numeric on an adaptive Lagrangian mesh with a DG numeric on an Eulerian mesh. This eases the embedding in a global climate system, keeps accurate representation of sharp gradients in the physical fields, and facilitates efficient parallelisation of the model to achieve superior computational performance.

In the geosciences and climate enterprise it is common practice to periodically correct the model state towards the truth using observations in a process called data assimilation (DA). A wide range of DA algorithms is available for this purpose. For example, see Kalnay (2002); Navon (2009); Bannister (2017); Carrassi *et al.* (2018); Evensen *et al.* (2022) for an overview. For example, the three-dimensional variational (3DVar) method is used in the sea-ice model of Environment Canada (Buehner *et al.*, 2013), the Arctic Cap Nowcast/Forecast System (Hebert *et al.*, 2015), and the Multivariate Ocean Variational Estimation/Meteorological Research Institute Community Ocean Model (Toyoda *et al.*, 2016). The MITgcm model has been paired with a local error subspace transform Kalman filter (Mu *et al.*, 2018). Other examples, such as the TOPAZ4 (Sakov *et al.*, 2012; Xie *et al.*, 2018), the NEMO-LIM2 (Massonnet *et al.*, 2013), and the neXtSIM model (Cheng *et al.*, 2023; Richter *et al.*, 2023; Sampson *et al.*, 2021), all use different flavours of the ensemble Kalman filter (EnKF) (Evensen *et al.*, 2022). It is planned that DA capabilities will also be added to neXtSIM$_{\mathrm{DG}}$. EnKFs are model agnostic and can in principle directly be used with DG models such as neXtSIM$_{\mathrm{DG}}$. The overarching objective of this study is to investigate an optimal approach that allows the EnKF to directly update the DG decomposition of the physical fields based on information provided in the observations. We believe this will improve the DA performance, stability, and efficiency relative to a comparable approach that is blind to the underlying DG structure.

We hypothesise that it is possible to exploit the structure of a DG solver directly to get better DA performance at little additional computational cost. In particular, we will investigate whether it is possible to (a) exploit the fact that DG basis functions partially resolve the solution at a subcell level to assimilate multiple observations per grid cell; (b) correct not only the field itself but also its derivatives, and (c) profit from the separation of the solution in basis functions inherent to DG to construct a scale-dependent localisation scheme. As a first and explorative study, we take an incremental approach to the problems. In a realistic set-up, DA would be applied in a sequential fashion. That is, the DA would assimilate observations into a model forecast and the results would subsequently be used as initial conditions for the following forecast cycle. Instead, we plan to investigate these hypotheses using a synthetic, univariate one-dimensional (1D) model. That is, a model in which the states are realisations of a statistical distribution and in which a dynamical forward model is absent. Consequently, we will not investigate whether the DA correction gives rise to numerical instabilities when the corrected states are propagated forward in time but will defer this to future work. The primary reason for this choice is that the statistical set-up allows for the generation of arbitrary background fields while maintaining control over their scales. Furthermore, it also limits the number of factors that can affect DA performance. For example, when the state is propagated forward in time between DA corrections using a dynamical model, the model can cause a non-Gaussian error distribution to develop. In that case, the assumptions under which most DA methods are valid are violated and the DA correction is suboptimal. Note that, even though we utilise the DG structure in DA, we will solely focus on the performance of DA instead of the DG scheme itself. In particular, this study is carried out under the assumption that a specific grid resolution and order of DG scheme is given by the modellers, as this is the typical situation faced by DA practitioners.

The first and second goals of this study are motivated by the density of observations bestowed by satellites. The resolution for the latest generation of satellite missions is within $\mathcal{O}(10^1)$–$\mathcal{O}(10^3)$ m (Drüe & Heinemann, 2004; Drusch *et al.*, 2012; Emery *et al.*, 1994; Gong *et al.*, 2013; Gourmelen *et al.*, 2018), which is already smaller than typical grid sizes in numerical models ($\mathcal{O}(10^3)$ m). Consequently, the observed fields contain spatial scales smaller than those that can be resolved with the current generation of finite-volume models. The portion of the model field that cannot be resolved by the model gives rise to representativeness errors. Together with instrumental errors, these errors make up the observational error. The work on the impact of these representativeness errors on DA performance is extensive, and reviews can be found in Cohn (1997); Janjić and Cohn (2006); Janjić *et al.* (2018). These representiveness errors should be accounted for when determining the statistics of the observational errors. This is particularly so in the specification of the observational error covariance. If not accounted for or removed, these errors can degrade the DA analysis (Liu & Rabier, 2002). Instead of explicitly accounting for the representiveness errors in the observational error statistics the issue is usually coped with by data thinning, the process of selecting only a portion of available data, and by superobbing. The latter is the process of averaging all observations from the same observation platform over one (or more) grid cell(s) into a single observation (Oke *et al.*, 2008). Although very popular, as they represent a concrete pragmatic option, superobbing and thinning are not ideal, as information in the small scales is lost in the process. In DG models the solution is (partially) resolved at the subcell scale. We therefore conjecture that the typical one observation per grid cell rule of thumb for superobbing/thinning no longer applies when the structure of DG is used in the DA. An improved capability to assimilate densely spaced data might then open up to the possibility for DA to correct subcell scales, thus possibly improving first-order (the gradient), or even higher order, derivatives of the model fields.

Our third research question, DG-based scale-dependent localisation, tackles a central concern for ensemble-based DA methods in high dimensions. Ensemble-based DA algorithms use an ensemble of model runs to estimate $\mathbf{B}$, the covariance of the errors in the model state. Owing to the finite size of the ensemble, the ensemble estimate is plagued by sampling errors (Houtekamer & Mitchell, 1998). If the absolute error correlation between the entries of the state vector is small (as is usually the case in physical systems where correlation decays exponentially with distance; e.g., Carrassi *et al.*, 2018), sampling errors can dominate the true correlation. The process to suppress this sampling error is referred to as localisation. Two categories of localisation scheme are in common use (Carrassi *et al.*, 2018; Greybush *et al.*, 2011): domain localisation and covariance localisation. Though computationally more efficient, domain localisation is not suitable for observations that are non-local or exhibit spatially correlated observation errors. As such conditions are common when dealing with dense satellite observations, the focus of this study will be on covariance localisation, in which localisation is applied directly to the background error covariance. In its most prevalent form, covariance localisation is applied in state space: it is assumed that the true error correlations tend to zero over long distances, and this behaviour is imposed on the ensemble covariance by tapering it off to zero with increasing distance (Ehrendorfer, 2007; Hamill *et al.*, 2001). The distance at which the sampling noise dominates the true error covariance will depend on the scales in the true covariance: the larger the scale in the covariance, the larger the tapering, or localisation, distance should be. However, distance-based localisation schemes in the state space do not offer this flexibility. Alternatively, the covariance can be localised in Fourier space (Buehner & Charron, 2007). This makes localisation scale dependent, but this dependence cannot be varied with position in the model domain. Localisation in wavelet space combines the best parts of the preceding two approaches (Berre *et al.*, 2015; Chabot *et al.*, 2017; Deckmyn & Berre, 2005; Tangborn, 2004; Varella *et al.*, 2011). As each wavelet represents a limited range of scales for a limited region, the localisation can be adjusted based on both scale and position. Multiple families of wavelets (e.g., orthogonal Meyer and spherical harmonics) have been used. In previous applications the use of wavelets required transformation of model output. As the current generation of geophysical models are formulated in either a physical grid space or spectral space, wavelet-based localisation in an ensemble DA system requires transformation of ensemble members to and from wavelet space. We will show in this study that DG basis functions form a filter bank separating different scales and consequently act, at the subgridscale, much like a wavelet basis, but without the need for wavelet transformations. The third aim of this study is to investigate if this model-intrinsic wavelet space can be used for localisation.

The article is organised as follows. Section 2 describes the general set-up of a DG model like neXtSIM$_{\mathrm{DG}}$. This is followed in Section 3 by a general description of the DA scheme used in this study. Section 4 contains the set-up and results of several univariate 1D twin experiments using synthetic truths and ensembles. With these we look at the benefits of assimilating multiple observations per cell on the model field and on its derivatives. Section 5 describes the framework of our proposed scale-dependent localisation scheme and numerical experiments. Finally, Section 6, summarises our findings, draws the conclusion, and gives forward perspectives of our study, thereby paving the way to the application of DG-informed DA in more realistic settings.

2 | DG METHODS

In recent decades the finite-volume method has become one of the most popular numerical methods to solve partial differential equations in geophysical models (Adcroft *et al.*, 2022; Chen *et al.*, 2006; Harris *et al.*, 2021; Kühnlein *et al.*, 2019; Shchepetkin & McWilliams, 2005) due to the fact that it inherently obeys mass and tracer conservation laws. In the finite-volume method, the unknowns are assumed to be constant within a model grid cell and their changes in time are determined by the flux through the grid cell edges. Equivalently, one can say that the solution in the grid cell is approximated by a multiple of a zeroth-order polynomial. This suggests a straightforward extension to a scheme in which the solution in each grid cell is approximated as a linear combination of higher order polynomials. This gave rise to DG methods (Hesthaven & Warburton, 2007). A short overview of the general structure of univariate DG models is presented in this section. Hereafter, we will use italic symbols to indicate quantities that are piecewise continuous functions of space and we will use bold symbols for quantities that can be represented by finite arrays and, therefore, can be stored in computer memory.

Suppose that the exact solution at position r, time t, to a system of partial differential equations is given as $x^{\mathrm{truth}}(r,t)$. Then the DG approximation to the scalar function $x^{\mathrm{DG}}(r,t)$ is given by

$$x^{\mathrm{DG}}(r,t) = \sum_{l=0}^{L-1} \sum_{m=1}^{M} \mathbf{x}_{lm}^{\mathrm{DG}}(t)\phi_{lm}(r), \tag{1}$$

with M the number of grid cells, L the number of basis polynomials used in the DG approximation, and $\mathbf{x}^{\mathrm{DG}}(t) \in \mathbb{R}^{L\times M}$ a matrix. The (scalar) entries of this matrix, the DG coefficients $\mathbf{x}_{lm}^{\mathrm{DG}}$, depend only on time, whereas the spatial part of the DG solution is encoded in the basis functions

$$\phi_{lm}(r) = \begin{cases} \sqrt{|\det \mathbf{D}\Psi_m^{-1}(r)|}\,\tilde{\phi}_l \circ \Psi_m^{-1}(r), & \text{if } r \in D_m \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where $\Psi_m : \widetilde{D} \rightarrow D_m$ is a coordinate transform from a reference domain $\widetilde{D}$ to the mth grid cell D_m, with $\det \mathbf{D}\Psi_m(r)$ the determinant of its Jacobian, D being the model domain, and $\{\tilde{\phi}_l : 0 \leq l < L-1\}$ is a set of reference functions defined on $\widetilde{D}$. Specifically, each basis function is defined locally within a grid cell and is irrelevant outside of it. The set $\phi = \{\phi_{lm} : 0 \leq l < L,\ 1 \leq m \leq M\}$ endowed with the L^2 inner product defines a Hilbert space of functions denoted here by V.

Several function families provide suitable candidates for the basis functions $\tilde{\phi}_l$. On a top level, bases can be divided between nodal and modal bases (Hesthaven & Warburton, 2007, section 3.1). For nodal basis functions the DG coefficients are equal to the solution evaluated at specific nodal (grid) points; that is, $\mathbf{x}_{lm}^{\mathrm{GP}}(t) = \mathbf{x}^{\mathrm{GP}}(r_{lm},t)$ if $r_{lm} \in D_m$. A 1D example of such a nodal basis is the family of Lagrange polynomials with $\tilde{\phi}_l$, defined as the unique $(L-1)$th-order polynomial that satisfies $\tilde{\phi}_l(r_{l'}) = \delta_{ll'}$ for all $0 \leq l < L$, and δ is the Kronecker delta. In contrast, DG coefficients for a modal basis cannot be directly found by evaluating the function but require projecting the solution onto the modal basis functions. One such example of a modal basis in one dimension, and the one that will be used in this study, is the Legendre basis. This choice is motivated by the fact that this basis is being used in neXtSIM$_{\mathrm{DG}}$ as well as that it allows for point-wise evaluation of model fields within the cells. More esoteric bases that are not pointwise defined and require more advanced observation operators will not be considered as their use is currently rare in geophysical modelling. In this case, $\widetilde{D} = [-1, 1]$,

$$\Psi_m(r) = \frac{r_{m+1} + r_m}{2} + r\frac{r_{m+1} - r_m}{2},$$

with r_m the left side of the mth grid cell, and $\tilde{\phi}_l(r) = P_l(r)$, the lth-order Legendre polynomial. In this case, the DG coefficient $\mathbf{x}_{lm}^{\mathrm{DG}}$ can be found by projecting x^{DG} on the basis polynomial:

$$\int_D \phi_{lm}(r)x^{\mathrm{DG}}(r,t)\,dr$$
$$= \sum_{l'=0}^{L-1}\sum_{m'=1}^{M} \mathbf{x}_{l'm'}^{\mathrm{DG}}(t)\int_D \phi_{lm}(r)\phi_{l'm'}(r)\,dr$$

$$
\begin{aligned}
&= \sum_{l'=0}^{L-1} \mathbf{x}^{\mathrm{DG}}_{l'm}(t) \int_{r_m}^{r_{m+1}} \phi_{lm}(r)\phi_{l'm}(r)\,dr \\
&= \sum_{l'=0}^{L-1} \mathbf{x}^{\mathrm{DG}}_{l'm} \int_{-1}^{1} P_l(r)P_{l'}(r)\,dr = \mathbf{x}^{\mathrm{DG}}_{lm},
\end{aligned}
$$

where x^{DG} has been expanded using Equation (1) and the orthogonality of the Legendre polynomials, $\int_{-1}^{1} P_l(r)P_{l'}(r)\,dr \sim \delta_{ll'}$, has been used.

In its most general form, a nonlinear geophysical model like neXtSIM$_{\mathrm{DG}}$ can be written as

$$
\begin{aligned}
0 = \mathcal{R}(x^{\mathrm{truth}}, t) &\overset{\mathrm{def}}{=} \frac{\partial \mathcal{A}(x^{\mathrm{truth}}, t)}{\partial t} \\
&- [-\nabla \cdot \mathcal{F}(x^{\mathrm{truth}}, t) + \mathcal{V}(x^{\mathrm{truth}}, t)],
\end{aligned} \tag{3}
$$

where $\mathcal{F}$ is the flux and $\mathcal{V}$ is the volume force. $\mathcal{A}$ can be a nonlinear operator, but often is equal to the identity operator (i.e., $\mathcal{A}(x,t) = x$) and $\mathcal{R}$ is the residual operator. In the Bubnov–Galerkin method (Bellman, 1970), an approximation $x \in V$ is sought such that the residual in Equation (3) is orthogonal to V; that is, for all $0 \leq l \leq L-1, 1 \leq m \leq M$, it must hold that

$$
\begin{aligned}
0 = \int_D \phi_{lm}\mathcal{R}(x,t)\,dr &= \frac{\partial}{\partial t}\int_D \phi_{lm}\mathcal{A}(x,t)\,dr \\
&+ \int_D \phi_{lm}\nabla\cdot\mathcal{F}(x,t)\,dr - \int_D \phi_{lm}\mathcal{V}(x,t)\,dr \\
&= \frac{\partial}{\partial t}\int_{D_m} \phi_{lm}\mathcal{A}(x,t)\,dr + \int_{\partial D_m} \phi_{lm}\mathcal{F}(x,t)\cdot dS \\
&- \int_{D_m} (\nabla\phi_{lm})\cdot\mathcal{F}(x,t)\,dr - \int_{D_m} \phi_{lm}\mathcal{V}(x,t)\,dr \\
&\approx \frac{\partial}{\partial t}\int_{D_m} \phi_{lm}\mathcal{A}(x,t)\,dr + \int_{\partial D_m} \phi_{lm}\mathcal{F}^*(x,t)\cdot dS \\
&- \int_{D_m} (\nabla\phi_{lm})\cdot\mathcal{F}(x,t)\,dr - \int_{D_m} \phi_{lm}\mathcal{V}(x,t)\,dr,
\end{aligned} \tag{4}
$$

where ∂D_m is the boundary of the mth cell and dS the boundary area element; $\mathcal{F}^*$ is the numerical flux, which will be discussed later. In the second line of the equation, integration by parts has been used in combination with the information on the local support of ϕ_{lm} given in Equation (2).

The flux $\mathcal{F}(x,t)$ is not uniquely defined on the grid cell boundaries since ϕ_{lm} exhibits discontinuities here. The fundamental ansatz made in DG methods is that this flux can be replaced by a numerical flux $\mathcal{F}^* = \mathcal{F}^*(x_-, x_+, t)$. Here, x_- is the approximation of x on one side of the cell boundary and x_+ on the other side. In the case the cell boundary is part of the domain's outer boundary, x_+ is provided by the boundary conditions. The numerical flux is problem specific and has to be chosen with care, as the stability of the DG scheme depends on it (Hesthaven & Warburton, 2007, section 4.4).

After integration, Equation (4) can be represented as a system of ordinary differential equations for $\mathbf{x}^{\mathrm{DG}}_{lm}$ with $0 \leq l < L, 1 \leq m \leq M$:

$$
\frac{d}{dt}\tilde{\mathcal{A}}_{lm}(\mathbf{x}^{\mathrm{DG}}, t) = -\tilde{\mathcal{S}}_{lm}(\mathbf{x}^{\mathrm{DG}}, t) + \tilde{\mathcal{M}}_{lm}(\mathbf{x}^{\mathrm{DG}}, t), \tag{5}
$$

with $\mathbf{x}^{\mathrm{DG}}$ a tensor having $\mathbf{x}^{\mathrm{DG}}_{lm}$ as its entries and

$$
\tilde{\mathcal{A}}_{lm}(\mathbf{x}^{\mathrm{DG}}, t) = \int_D \phi_{lm}(r)\mathcal{A}\left(\sum_{l'm'} \mathbf{x}^{\mathrm{DG}}_{l'm'}(t)\phi_{l'm'}(r), t\right) dr,
$$

$$
\begin{aligned}
&\tilde{\mathcal{M}}_{lm}(\mathbf{x}^{\mathrm{DG}}, t) \\
&\quad = \int_D \nabla\phi_{lm}(r,t)\cdot\mathcal{F}\left(\sum_{l'm'} \mathbf{x}^{\mathrm{DG}}_{l'm'}(t)\phi_{l'm'}(r), t\right) dr \\
&\quad + \int_D \phi_{lm}(r)\mathcal{V}\left(\sum_{l'm'} \mathbf{x}^{\mathrm{DG}}_{l'm'}(t)\phi_{l'm'}(r), t\right) dr,
\end{aligned}
$$

$$
\begin{aligned}
&\mathcal{S}_{lm}(\mathbf{x}^{\mathrm{DG}}, t) = \sum_{m'<n'} \int_{\partial D_{m'} \cap \partial D_{n'}} \\
&\quad \mathcal{F}^*\left(\sum_{l'} \mathbf{x}^{\mathrm{DG}}_{l'm'}\phi_{l'm'}(r,t), \sum_{l'} \mathbf{x}^{\mathrm{DG}}_{l'n'}\phi_{l'n'}(r,t)\right)\cdot dS \\
&\quad + \sum_{m'} \int_{\partial D_{m'} \cap \partial D} \\
&\quad \mathcal{F}^*\left(\sum_{l'} \mathbf{x}^{\mathrm{DG}}_{l'm'}\phi_{l'm'}(r,t), \sum_{l'} \mathbf{x}^{\mathrm{DG}}_{l'm'}\phi_{l'm'}(\mathbf{r},t)\right)\cdot dS.
\end{aligned}
$$

The preceding discussion can easily be extended to multivariate models by expanding each model field separately—for example, see Hesthaven and Warburton (2007, section 5.9) for a multivariate example. In that case, $\mathbf{x}^{\mathrm{DG}}$ would contain coefficients for different model fields, and consequently $\mathcal{A}(\mathbf{x}^{\mathrm{DG}}, t)$, $\mathcal{V}(\mathbf{x}^{\mathrm{DG}}, t)$, and $\mathcal{F}(\mathbf{x}^{\mathrm{DG}})$ can contain terms that depend on different fields. Furthermore, the aforementioned procedure can trivially be expanded to diagnostics equations (i.e., equations without time derivative) that constrain the solution to a specific submanifold by setting the left-hand side of Equation (5) to zero.

3 | DA WITH DG-BASED MODELS

DA combines a prior guess of the model state, also known as the background, with observations in order to produce an improved estimate of the true state of the system.

A plethora of DA methods have been developed over the past decades; though different in their details, all share some basic components. Among these are a list of observations and the specification of a probability distribution for the observational errors, the specification of the model state as a finite array, together with an estimate of their a priori error distribution, the background error, and an observation operator that takes the model state as input and outputs predictions for the observed quantities.

We shall study the implications, advantages, and drawbacks of adapting DA to DG models. As a prototype of a DA method, we will work with the deterministic ensemble-3DVar (D-E3DVar): an ensemble-variational method based on the deterministic EnKF (Sakov & Oke, 2008). In the version of D-E3DVar used here, it is assumed that observational error consists solely of the instrumental error, thus neglecting the representativeness error, that the statistics are Gaussian, static in time, and that there is no correlation between these instrumental errors in the different observations, while background errors are assumed to be Gaussian with a varying covariance estimated from an ensemble. Our choice for a method as D-E3DVar is motivated by the fact that it is an ensemble method and thus requires localisation. It is therefore amenable to study scale-aware localisations, as we shall discuss in Section 5. As it is a variational method, the implementation of covariance localisation in the linear solver is straightforward, certainly when compared with an EnKF method. Furthermore, its results are easily extendable to an ensemble of four-dimensional variational systems, which is an approach currently used in a number of four-dimensional variational implementations (Bonavita *et al.*, 2016; Ngodock *et al.*, 2020; Pasmans *et al.*, 2020; Pasmans & Kurapov, 2019; Zhu *et al.*, 2022) and theoretically can deal with nonlinear observation operators.

D-E3DVar DA updates the forecast state by minimising the cost function

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}\left[\mathbf{y} - H(\overline{\mathbf{x}}^{\mathrm{f}} + \delta\mathbf{x})\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y} - H(\overline{\mathbf{x}}^{\mathrm{f}} + \delta\mathbf{x})\right], \quad (6)$$

where $\overline{\mathbf{x}}^{\mathrm{f}}$ is the ensemble mean of the forecast states, $\delta\mathbf{x}$ the DA correction (the analysis increment) to the forecast mean, $\mathbf{y}$ the vector containing the different observations, H the (possibly nonlinear) observation operator mapping the model state into predictions for the observations, and $\mathbf{R}$ the observation error covariance matrix, which we assume to be diagonal and constant in time. The background error covariance $\mathbf{B}$ is estimated from an ensemble of N independent model realisations as

$$\mathbf{B} = \frac{1}{N-1}\sum_{n=1}^{N}\mathbf{a}^{(n)} \otimes \left(\mathbf{a}^{(n)}\right)^{\mathrm{T}},$$

with $\mathbf{a}^{(n)} = \mathbf{x}^{(n)} - \overline{\mathbf{x}}^{\mathrm{f}}$ the forecast ensemble anomaly of ensemble member n.

For the purpose of this study we conveniently consider point observations of the system's state vector; therefore, H is linear and we can write

$$\mathbf{y} - H(\overline{\mathbf{x}}^{\mathrm{f}} + \delta\mathbf{x}) = \mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^{\mathrm{f}} - \mathbf{H}\delta\mathbf{x} \overset{\mathrm{def}}{=} \mathbf{d} - \mathbf{H}\delta\mathbf{x}. \quad (7)$$

The post-DA model state (i.e., the analysis) produced by minimising the cost function in Equation (6) is given by

$$\overline{\mathbf{x}}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{f}} + \hat{\delta\mathbf{x}} = \overline{\mathbf{x}}^{\mathrm{f}} + \mathbf{K}\mathbf{d}, \quad (8)$$

with $\hat{\delta\mathbf{x}}$ the value for which J attains its minimum. Similar to Sakov and Oke *et al.* (2008), the members of the analysis ensemble can be approximated as

$$\mathbf{x}^{\mathrm{a},(n)} = \overline{\mathbf{x}}^{\mathrm{a}} + \mathbf{a}^{(n)} - \frac{1}{2}\mathbf{K}\mathbf{H}\mathbf{a}^{(n)}, \quad (9)$$

where $\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$ is the Kalman gain.

The values $\mathbf{K}\mathbf{d}$ and $\mathbf{K}\mathbf{H}\mathbf{a}^{(n)}$ are given as $\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1/2}\chi^{(0)}$ and $\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1/2}\chi^{(n)}$ respectively, where $\chi^{(n)}$ $0 \le n \le N$ are found by iteratively solving

$$\begin{aligned}(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1/2} + \mathbf{I})\left[\chi^{(0)}, \chi^{(1)}, \ldots, \chi^{(N)}\right] \\ = [\mathbf{R}^{-1/2}\mathbf{d}, \mathbf{R}^{-1/2}\mathbf{H}\mathbf{a}^{(1)}, \ldots, \mathbf{R}^{-1/2}\mathbf{H}\mathbf{a}^{(N)}]\end{aligned} \quad (10)$$

using the reduced conjugate gradient method (Gürol *et al.*, 2014) and a block-diagonal Krylov algorithm (Aulígné *et al.*, 2016; Mercier *et al.*, 2019).

With the DA scheme in hand, here the D-E3DVar (Equations 6–9), one can then decide whether the DA correction should act on the model physical quantities at grid nodes $\mathbf{x}^{\mathrm{GP}}$ or on the coefficients obtained by projecting the field on the Legendre basis functions. Given the former approach, the approximation of the field is completely determined by the field's values at a set of grid points. We will refer to this approach as the "grid-point approach." In the latter case, the field is a piecewise Legendre polynomial of order $< L$. Because of this, it will be referred to as the "DG space." The grid point and DG spaces are isomorphic with $\mathbb{R}^{M}$ and $\mathbb{R}^{L\cdot M}$ respectively, and in this work we will apply the term grid-point space to the function space and its isomorphic equivalent in $\mathbb{R}^{M}$ interchangeably. Similarly, we will apply the term DG space to

TABLE 1 Interpolated solutions (zeroth-order derivative) and their first and second derivatives, for the discontinuous Galerkin (DG) and grid point (GP) representation for the model field at point r.

Order	DG	Grid point
0	$\sum_{l=0}^{L-1}\sum_{m=1}^{M}\mathbf{x}_{lm}^{\mathrm{DG}}\phi_{lm}(r)=\mathbf{H}_r^{\mathrm{DG}}$	$\sum_{m=1}^{M}\mathbf{x}_m^{\mathrm{GP}}\chi_m(r)=\mathbf{H}_r^{\mathrm{GP}}$
1	$\sum_{l=0}^{L-1}\sum_{m=1}^{M}\mathbf{x}_{lm}^{\mathrm{DG}}\frac{d\phi_{lm}}{dr}(r)$ $+\sum_{l=0}^{L-1}\sum_{m=1}^{M}\sum_{m'=1}^{M}\phi_{lm}(r)\left([[\phi_{lm}]]_{m'}\{\{x^{\mathrm{DG}}\}\}_{m'}-[[\phi_{lm}x^{\mathrm{DG}}]]_{m'}\right)$	$\sum_{m=1}^{M}\frac{\mathbf{x}_{m+1}^{\mathrm{GP}}-\mathbf{x}_m^{\mathrm{GP}}}{\Delta r}\chi_m(r-\frac{1}{2}\Delta r)$
2	$\sum_{l=0}^{L-1}\sum_{m=1}^{M}\mathbf{x}_{lm}^{\mathrm{DG}}\frac{d^2\phi_{lm}}{dr^2}(r)$ $+\sum_{l=0}^{L-1}\sum_{m=1}^{M}\sum_{m'=1}^{M}\phi_{lm}(r)\left([[\phi_{lm}]]_{m'}\{\{\frac{dx^{\mathrm{DG}}}{dr}\}\}_{m'}-[[\phi_{lm}\frac{dx^{\mathrm{DG}}}{dr}]]_{m'}\right)$	$\sum_{m=1}^{M}\frac{\mathbf{x}_{m+1}^{\mathrm{GP}}+\mathbf{x}_{m-1}^{\mathrm{GP}}-2\mathbf{x}_m^{\mathrm{GP}}}{\Delta r^2}\chi_m(r)$

Note: Δr is the grid cell width, $\chi_m(r)=\max(0,1-\frac{|r_m-r|}{\Delta r})$, x^{DG} (dx^{DG}/dr) is the polynomial DG approximation in the first (second) row of the table, $[[\cdot]]$ is the difference between the left and right limits towards the point separating grid cells m and $m+1$, and $\{\{\cdot\}\}_m$ is the average of the two limits. Basis functions in second and third rows are assumed to be orthonormal.

the piecewise-polynomial function space and its isomorphic equivalent in $\mathbb{R}^{L\cdot M}$ interchangeably. In particular, a specific field in grid-point space (DG space) and its representation as $\mathbb{R}^M$ ($\mathbb{R}^{L\cdot M}$) vector of values on grid points (vector of projection coefficients) are treated equivalently in this article. Assimilating in the DG or in the grid-point space implies defining different observation operators. In this work, the observations are point observations. For a DG model, predictions for the observations are determined by evaluating the polynomial DG approximation at the observation points. In the grid-point approach, model values at the left side of cell m are stored in $\mathbf{x}_m^{\mathrm{GP}}$ and linear interpolation is used to find the field value at an arbitrary observation position r. The pointwise interpolation error of this scheme being $\mathcal{O}(\Delta r^2)$, with Δr the cell size (Quarteroni *et al.*, 2007, section 8.1). The equations for the observation operator for the DG and grid-point models can be found in the first row of Table 1. Also listed in this table are the approximations for the first and second derivatives of the field used in the calculation of the root-mean square error (RMSE) ratios in Section 4.2 for the case in which the basis functions are orthonormal; they are thus only relevant as an evaluation metric for the performance of the methods, and not in the methods workflow. The DG coefficients for the first-order (second-order) derivative are found using the strong DG formulation obtained by applying integration by parts twice to $\int_D \phi_{lm}(r)(dx^{\mathrm{DG}}/dr(r))\mathbf{d}r$ ($\int_D \phi_{lm}(r)(d^2x^{\mathrm{DG}}/dr^2(r))\mathbf{d}r$). The resulting expression consists of two parts. The first part is the first (second) derivative of the polynomial DG approximation found in row 1 of Table 1. The second part accounts for discontinuities at the grid cell boundaries. In the grid-point approach, values for the first (second) derivative of the model field are obtained by first approximating the values of the first-order (second-order) derivative in the grid cell centres (grid cell vertices) using a central finite-difference method. These values at the grid cell centres (grid cell vertices) are then interpolated to the observation location using linear interpolation. As the error in both the central finite-difference method and linear interpolation is $\mathcal{O}(\Delta r^2)$, the interpolation error in the derivatives is also accurate in $\mathcal{O}(\Delta r^2)$. Cubic spline (Quarteroni *et al.*, 2007, section 8.7) and piecewise cubic Hermite interpolating polynomial (Fritsch & Butland, 1984) interpolation, which are $\mathcal{O}(\Delta r^4)$ accurate, have also been tried in the observation operator for the grid-point case (not shown here). Use of the cubic spline interpolator did improve RMSE ratios, but it was still outperformed by lower order DG models (DG01, DG02). RMSE ratios for the piecewise cubic Hermite interpolating polynomial interpolator at high observational density were worse than those obtained using the spline interpolator. This is due to the slope limiter removing the small-scale signal in the model fields.

4 | ASSIMILATION IN THE DG SPACE: CAN WE ASSIMILATE DENSER DATA?

The current generation of geophysical DA systems are restricted in the density of observations that they can effectively assimilate. It has been shown that in standard DA systems, with uncorrelated observational errors, assimilating more observations produces more accurate analyses. However, the improvement becomes marginal when the observational density exceeds ∼ 1 observation per grid cell (Liu & Rabier, 2002). On the other hand, spatial correlations between observations errors are a common occurrence, particularly for satellite data. Such correlations will evidently be present if instrument errors are correlated. But even if they are not, preprocessing and representativeness errors stemming from unresolved processes in the model can introduce spatial correlations between observation errors (Evensen *et al.*, 2022; Janjić *et al.*, 2018). If these latter correlations are not accounted for, increasing the observational density is not only ineffective but can even deteriorate DA performance (Liu & Rabier, 2002). Ideally, any off-diagonal error covariances can be explicitly accommodated in the DA algorithm. The use of diagonal **R** is common, as the off-diagonal error covariances are often not a priori known and additional steps and assumptions are necessary to estimate them (Campbell *et al.*, 2017; Evensen, 2021; Fowler *et al.*, 2018; Michel, 2018; Miyoshi *et al.*, 2013; Rainwater *et al.*, 2015; Stewart *et al.*, 2013). Because of this, and the fact that it allows for the use of more efficient algorithms to minimise the cost function in Equation (6), it is in practice common to work with diagonal observation error covariances. This is traditionally accomplished by discarding data, an operation known as "thinning" (i.e., removing observations) and/or by spatially averaging observations (i.e., "superobbing"). In both cases, potentially useful observations are discarded or their information is not properly used. To better utilise the observation information, sophisticated schemes have been developed to optimally select the observations to be retained during the observation thinning process (Gratton *et al.*, 2015; Lazarus *et al.*, 2010; Mulia *et al.*, 2017; Ochotta *et al.*, 2007), as well as schemes in which "synthetic" representativeness errors are created by using slightly different observation locations for each ensemble member (Lea *et al.*, 2022; Zuo *et al.*, 2017, 2018). Here, we will investigate whether assimilation in DG models with representativeness errors can take advantage from higher observation densities without having to resort to non-diagonal observation error covariances or adaptive thinning schemes. In particular, we will study to what extent observation thinning is still necessary in these models. We will ignore superobbing at this stage and leave it for future work.

4.1 | Experimental set-up

The experimental set-up is chosen to resemble that in Liu and Rabier (2002); each realisation of the experiment is constructed as follows and is implemented in the DAPPER DA framework (Raanes *et al.*, 2023):

1. First, a background field is generated by drawing from a probability distribution. The pointwise expectation value of this background distribution is zero and its spatial autocovariance is

$$\mathrm{cov}(\Delta r) = 100 \times \left[\cos\left(\frac{8\pi\Delta r}{\ell}\right) + \frac{4}{3}\sin\left(\frac{8\pi\Delta r}{\ell}\right)\right] \times \mathrm{e}^{-6\pi\Delta r/\ell},$$

   with Δr the distance between points and $\ell = 8000$ km the width of the periodic domain. That is, its power spectrum can be written as $\mathcal{S}(\mathbb{E}\left[x^{\mathrm{truth}}\right])(\kappa) = |\mathcal{F}(\mathrm{cov})|^2(\kappa)$. The generation of a realisation of such a background field from this spectrum is described in more detail in Appendix A.
2. Seventeen anomalies are generated, following the procedure in Appendix A, by sampling from noise with power spectrum $\mathcal{S}(\kappa) \sim \kappa^{-\alpha}$, $\mathcal{S}(0) = 0$ (zero mean), $\int_0^\infty \mathcal{S}(\kappa)\mathrm{d}\kappa = 1$ (unit variance), with α depending on the experiment. These anomalies are added to the realisation of the background field created in the previous step and represent background errors. Keep one of the 17 members thus created aside as artificial "truth" x^{truth}. The other $N = 16$ constitute the background ensemble. We justify the use of an ensemble of this size based on the fact that its dimension-to-ensemble-size ratio is $\mathcal{O}(10^4)$ as large as the ratio typically found in operational DA systems, and that tests showed relatively little reduction in analysis error beyond $N = 16$ (not shown). That is, each realisation of the experiments starts off with a different ensemble and different ensemble mean.
3. Sample N_{obs} equally spaced observations from the "artificial truth" and add observational noise sampled from a standard normal distribution. This added observational noise represents solely the instrument error and is therefore uncorrelated by construction. The value of N_{obs} will vary between experiments.
4. Project the $N = 16$ members on both the DG space and the grid-point space with $M = 79$ grid cells. For the grid-point space this is done by evaluating the members expressed as Fourier series at the grid cell vertices. Interpolation is not yet required at this stage. The DG coefficients are obtained by integration as outlined in Equation (A.1). This projection approach was chosen, instead of the faster approach of obtaining the DG coefficients by interpolating the field to Gauss–Legendre points, to avoid aliasing of higher order DG modes into the coefficients for the lower order modes (Hesthaven & Warburton, 2007, section 5.3). This ensures that these small-scale processes will not be represented in the numerical projection as they are perpendicular to the projection space. The lost processes mimic representativeness errors in realistic, dynamic models stemming from the lack of capability to resolve dynamics at all scales. Contrary to the observational noise, these representativeness errors are spatially correlated.
5. Sample the members at the observation locations using $\mathbf{H}^{\mathrm{GP}}$ for the projection on the grid-point space and $\mathbf{H}^{\mathrm{DG}}$ for the projection on the DG spaces (see Section 3).
6. Compute the analysis using Equation (8).
7. Calculate the analysis RMSE for the field and some of its derivatives. The RMSE in the pth-order derivative is

calculated as

$$\sqrt{\sum_{m=1}^{M}\int_{D_m}\left(\mathcal{I}_p(r)-\frac{d^p x^{\mathrm{truth}}}{d^p r}(r)\right)^2 dr} \tag{11}$$

with $\mathcal{I}_p$ the interpolator for the pth-order derivative in Table 1 corresponding to either the grid-space or DG model. Notice that this RMSE contains both the background/analysis error and the representativeness error.

4.2 | Results observation density

Following the procedure in Section 4.1 using N_{obs} observations, we compute the domain-averaged forecast (prior to DA) and analysis RMSE; that is, the L_2-norm of the error between the “truth” and the ensemble mean, for the fields and their first- and second-order derivatives. This is then repeated another 49 times, each time using a different 16-member ensemble. After completion, the same 50 realisations of a 16-member ensemble are used to carry out the same experiment but now using different numbers of observations N_{obs}. As a result of the procedure, the error in the model fields consists of two parts. One part is the anomaly explicitly introduced in the second step of Section 4.1. This part corresponds to the background and model error in a model with full dynamics. The second part is an interpolation/projection error and would correspond to the representativeness error in a full model. In the DG models based on Legendre polynomials, the (derivative) of the latter error e is bounded by

$$\left\|\frac{d^d e}{dr^d}\right\| \leq \frac{h^{L-d}}{\sqrt{(L+L_{\max}-4d)!}}\left\|\frac{d^{L_{\max}} x^{\mathrm{truth}}}{dr^{L_{\max}}}\right\| \tag{12}$$

with $||\cdot||$ the L_2-norm and $L_{\max}=11$ the number of degrees of freedom per grid cell for the highest order DG scheme considered in this work (Hesthaven & Warburton, 2007, section 4.3). This shows that, even prior to any DA, the error in the model field will be smaller when high-order DG schemes (i.e., $L \gg 1$) are used and that the interpolation errors for the derivatives ($d > 0$) are higher than for the field itself ($d = 0$). As our interests lie with the performance of the DA and not the model, we try to filter out this effect by displaying RMSEs in this work as post/prior DA ratios.

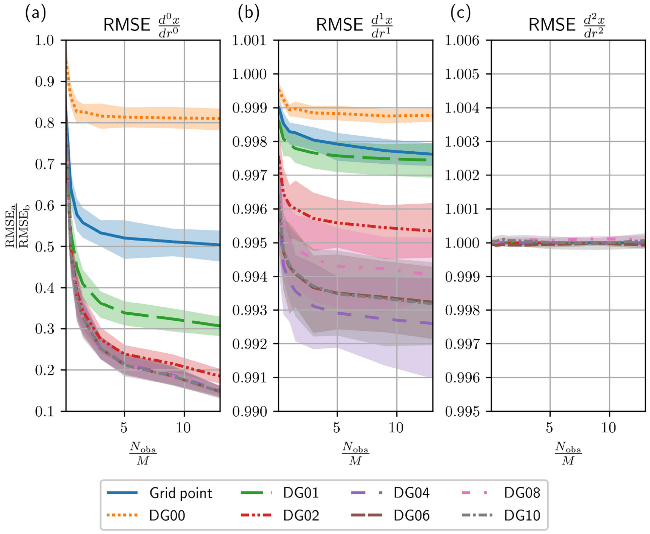


FIGURE 1 Domain-averaged root-mean-squared (RMS) analysis error ($\mathrm{RMSE_a}$)/RMS background error ($\mathrm{RMSE_{bg}}$) ratio in the (a) field, (b) first-order derivative of the field, and (c) second-order derivative of the field as a function of the average number of observations per grid cell for the grid-point model with $M = 79$ grid cells as well as discontinuous Galerkin (DG) models of orders $L-1=0$, $L-1=1$, $L-1=2$, $L-1=4$, $L-1=6$, $L-1=8$, and $L-1=10$, $M=79$ grid cells, and background error power spectrum S that scales as $S \sim \kappa^{-4}$.

The RMSE ratios from these 50 experiments and their 90% confidence intervals calculated using percentile bootstrap (Efron & Tibshirani, 1994) are shown in Figure 1a as a function of the average number of observations per grid cell. For the grid-point model, the decrease of the ratio with increasing observation densities starts to level off at $N_{\mathrm{obs}}/M \approx 1.5$ with $M = 79$ the number of grid cells. For higher order DG models (i.e., DG models using polynomials of order ≥ 1, the RMSE ratio decreases faster with increasing observation density than for the grid-point case. The difference in slope is most pronounced when $N_{\mathrm{obs}}/M \leq 5$, but the RMSE in the DG models continues to decrease even after this point. As a result, the DA in the higher order DG models outperforms the DA in the grid-point model significantly at the 90% level. This behaviour—that is, the capability to assimilate and benefit from denser data—is a consequence of the DG models having more degrees of freedom than the grid-point model. Therefore, it can provide a better fit to the observations, as will be shown at the end of this section. For the zeroth-order derivative (i.e., the model field itself) there is no DA benefit beyond a DG order of 4. This can be explained by the fact that the background error spectrum in the experiment is dominated by $S(\kappa) \sim \kappa^{-4}$. That is, it is dominated by the large scales (small κ). Higher order polynomials represent smaller scales, and consequently the background variance explained by the higher order DG coefficients is so small that effectively no DA updates are made to these coefficients. Contrary to Liu and Rabier (2002), no increase of RMSE with increasing observational density was detected. This is because the spatial observational error correlations in the work of Liu and Rabier (2002) are larger than in this experiment as they have added long spatial correlations, representing correlated instrumental errors, to the representativeness errors already present Liu and Rabier (2002). In this experiment, however, instrumental errors are assumed to be uncorrelated. The RMSE ratios for the first derivative decrease (i.e., improve) with increasing DG order, as shown in Figure 1b. The improvement over the grid-point model is significant at the 90% confidence level for higher order DG models. Furthermore, the relative improvement increases consistently with the DG order. However, the actual size of this improvement is rather marginal (<0.7%; cf. the *y*-axis ranges in Figure 1a,b). Finally, Figure 1c shows that DA has no significant impact on the errors in the second derivative.

The absence of improvement when the DG order is increased beyond 4 suggests that the potential of DG-based DA to improve over grid-point DA is related to (effectively bounded by) the scales actually present in the background error. To investigate this behaviour, we perform the same type of experiments but this time with background errors sampled from a pink-noise spectrum (i.e., a spectral slope $\alpha = -1$) that, by construction, includes smaller spatial scales. The results are shown in Figure 2. In particular, Figure 2a confirms the finding in Figure 1 that the higher order DG models outperform the grid-point model. Nevertheless, in contrast to Figure 1a, there is now a consistent reduction of the RMSE ratios with increasing DG order. It is worth noting, however, that the RMSE ratios are higher than those obtained with a spectral slope of $\alpha = -4$ (see Figure 1a). The reason for this will be discussed in more detail at the end of this section. For the first and second derivatives (Figure 2a,b) the situation is substantially different from that in Figure 1 with the RMSE ratios increasing as the DG order increases: DA deteriorates the representation of the gradient and the Laplacian of the solution.

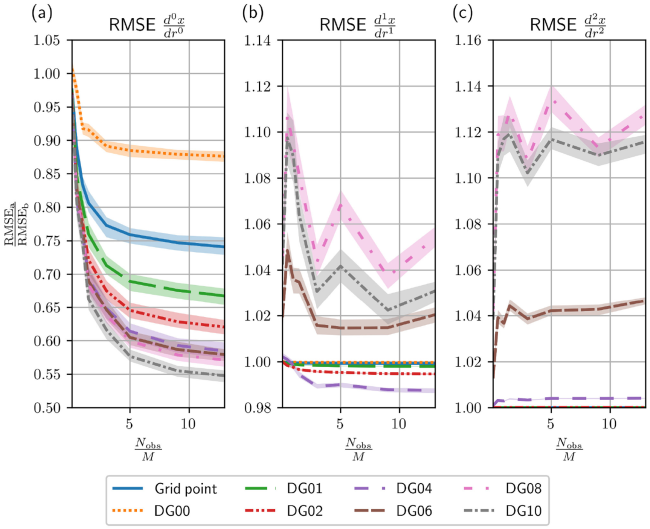
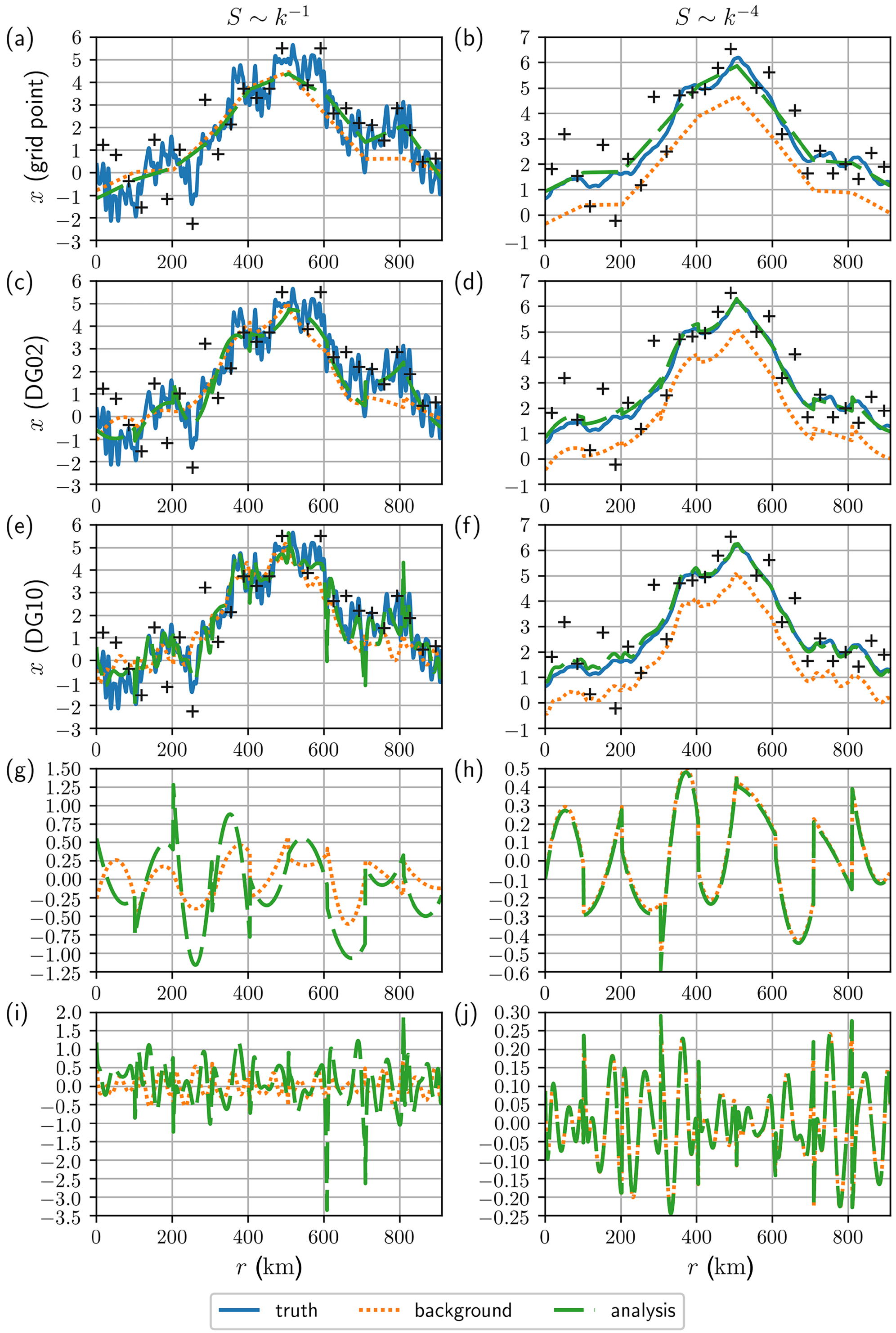


FIGURE 2 As Figure 1, but now with a background error power spectrum S that scales as $S \sim \kappa^{-1}$.

To understand and illustrate why this occurs, we plot in Figure 3 the observations, the truth, the background, and the analysis for DG orders 2 and 10, the grid-point solution, and the background error for the two noise spectra under consideration. In the case of a pink spectrum ($\alpha = -1$, Figure 3a,c,e), the spectrum of the background error in the first (second) derivative has a slope of $\alpha = 1$ ($\alpha = 3$), which follows a blue spectrum. This implies that most of the error in the first and second derivatives is in the smaller scales. As the dashed orange line (the background estimate) in Figure 3a,c shows, this small-scale portion of the spectrum is filtered out by projecting on the grid point or on the DG02 space. Therefore, there is not much error left to be removed by the DA and the RMSEs ratios are of ~ 1 (cf. Figure 2b,c). The higher order polynomials in the DG10 model contain smaller scales. The freedom contained in these scales is used to fit the observations (see Figure 3e). However, to fit the observations, the analysis field must make sharp turns (e.g., see Figure 3e near position 100 and 700 km), leading to large oscillations in the analysis, especially near the cell boundaries. These oscillations really stand out when we look at the differences in the analysis fields between DG10 and DG02 shown in Figure 3i. They are several times larger than the differences in the background and the differences between DG02 and the grid-point model shown in Figure 3g. This behaviour is similar to the Runge phenomenon observed for the interpolation polynomial: as the DG order increases, the analysis starts to fit the observations. However, increases in order do not convergence uniformly to some limit. Instead, near the boundaries, oscillations develop of which the amplitude increases with increasing order. These oscillations can result in pointwise divergence (e.g., see Figure 3e near 700 km). As the scale of these oscillations is small, their contribution to the RMSE is more noticeable for the derivatives than for the field itself. Both slope limiters and spectral filtering are popular methods to remove small-scale oscillations in finite-element methods. Tests in which a slope limiter was applied to the ensemble members (not shown here) indicate that they are of little benefit for the set-up used here. Our truth does contain small-scale oscillations, and removal of these oscillations from the ensemble by the limiter is detrimental to DA's ability to fit these oscillations. A similar issue arises when spectral filters are used to suppress higher order polynomials in the analysis. In summary, for high-order DG models the reduction of errors in the field introduces errors in the small scales that are magnified by differentiation.

FIGURE 3 Example of the the true field (blue), the background estimate (orange), and analysis (green) for (a, b) the grid-point model, (c, d) the discontinuous Galerkin (DG) order 2 model (DG02), (e, f) the DG order 10 model (DG10) with background error sampled from a spectrum with (a, c, e, g, i) $\alpha = -1$ and (b, d, f, h, j) $\alpha = -4$. On average, three observations are assimilated per grid cell. The observed values are depicted as black + signs. Only a limited part of the domain (nine grid cells) is shown. Also shown are the differences between the various models: (g) difference between panels (c) and (a); (h) difference between panels (d) and (b); (i) difference between panels (e) and (c); (j) difference between panels (f) and (d).



With a red error spectrum ($\alpha = -4$) the spatial correlations in the background errors are longer. For portions of the domain smaller than the correlation scale this gives the impression that the background is the truth plus a fixed offset (see Figure 3b,d,f) and errors in its derivatives are small. Since most of the error is located in the large scales, background estimates for the derivative are already good and little correction from DA can be achieved. Consequently, the differences in the analysis between DG02 and the grid-point model and between DG10 and DG02, shown in Figure 3h,j respectively, are a full order smaller than those observed using a pink spectrum (Figure 3g,i).

The foregoing two experiments were repeated, but now using observations randomly distributed throughout the domain with the total number of observations ranging from 4 to 1027. No significant difference with the results in Figures 1 and 2 were found (not shown). The experiments were also repeated (not shown) using observational error standard deviations of $\sigma_o = 0.5$ ($\sigma_o = 1.5$). In these cases the analysis RMSE ratio is lower (higher) than those in Figures 1 and 2, but there is no qualitative difference.

The dependence of the DA performance on the background error spectrum and its spatial scales is further explored in Figure 4. In Figure 4 the RMSE ratio for the model field and its derivatives is shown as function of the slope of the power spectrum α used to generate the background errors for three different observational densities. The left column of Figure 4 shows that for the model field itself DA performance improves in all DG orders as the spectrum from which the background errors are sampled becomes redder (lower α): the smaller the dominant scales in the error are, the smaller the improvement obtained using DG-based DA. However, the background errors benefit the most from increasing the DG order in models in which the smaller scales dominate. The centre and right column of Figure 4 show that the negative impact of DA on the derivatives in the higher order DG models starts to disappear when the background error spectrum slope drops below $\alpha < -2$.

The experiments in the preceding part of this section were all carried out under the assumption that the number of grid cells remains the same. This is because we are primarily interested in the optimal number of observations to assimilate per grid cell given a specific model set-up. This set-up resulted in lower RMSE ratios for the higher order DG models as they possess a higher number of degrees of freedom than the grid-point model. For completeness we will also compare the RMSE ratios of the DG models with the ratios in grid-point models having an equal number of degrees of freedom. Here, the number of degree of freedom in the grid-point experiment is varied by varying the number of grid cells, whereas in the DG experiments it is varied by varying the order of the DG scheme

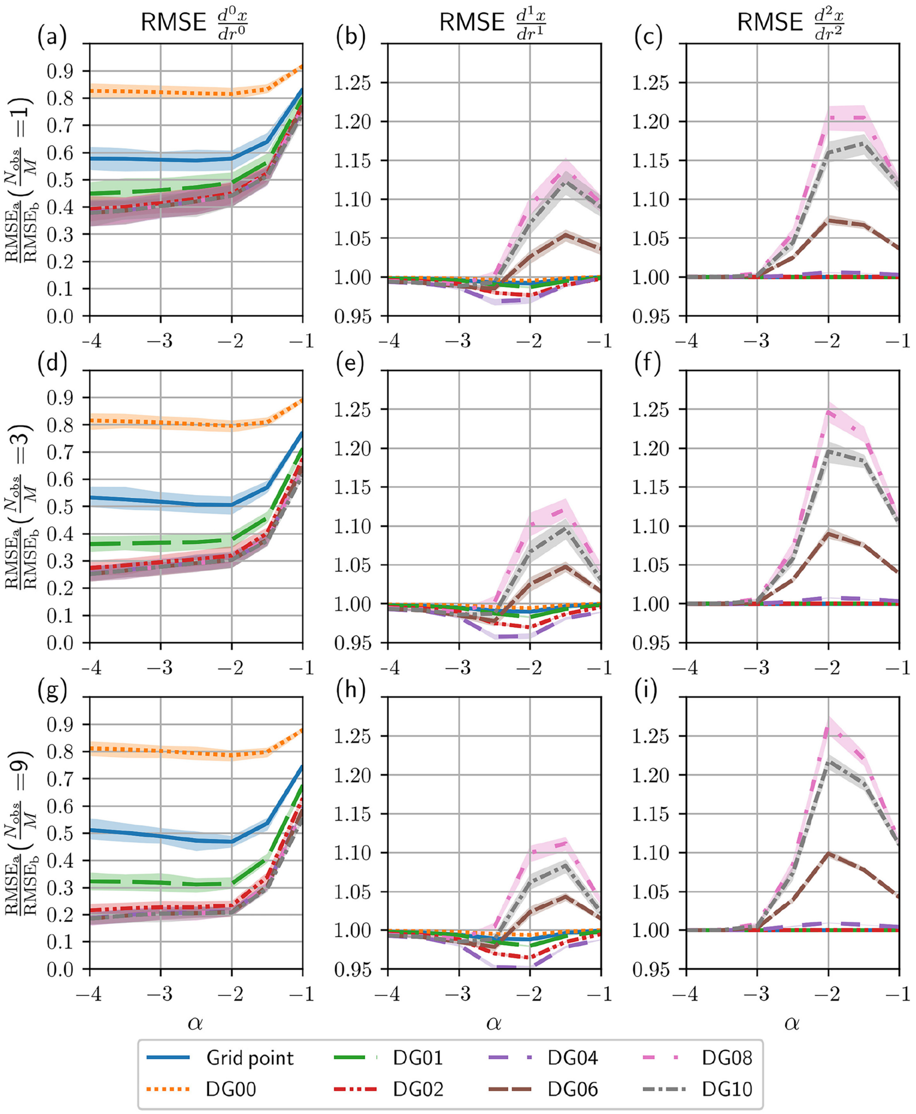


FIGURE 4 Investigation into the dependence of the root-mean-squared error (RMSE) on the scales in the background error. Shown are the analysis/background RMSE ratios of the (a, d, g) zeroth, (b, e, h) first, and (c, f, i) second derivatives as a function of the slope α of the power spectrum for experiments assimilating on average (a–c) $N_{\mathrm{obs}}/M = 1$, (d–f) $N_{\mathrm{obs}}/M = 3$, and (g–i) $N_{\mathrm{obs}}/M = 9$ observations per grid cell, with N_{obs} the total number of observations and $M = 79$ the number of grid cells.

whilst keeping the number of grid cells equal to $M = 79$. The RMSE ratios as a function of number of degrees of freedom are shown in Figure 5. The ratios are shown for different numbers of assimilated observations with qualitatively similar results. For errors generated from a red spectrum ($\alpha = -4$) the DA performance of the grid-point and DG models at an equal number of degrees of freedom is not significantly different at the 90% -confidence level. An exception to this behaviour is the DG0 model with 79 degrees of freedom, for which DA performs significantly worse than the grid-point model. This is because for a single grid cell DG0 only uses one degree of freedom to approximate the model, whereas in the grid-point model the approximation involves two points (one at each cell edge). For a pink-noise error spectrum ($\alpha = -1$), DA works better in the grid-point model than in the DG models with equal number of degrees of freedom, as the former is not affected by the Runge phenomenon. Increasing the value of $\mathbf{R}$ to weaken the impact of the observations within DA did not resolve this problem (not shown). Even though DA for the grid-point model performs, at a given number of degrees of freedom, as good as or better than the DG model, it should be noted that in a realistic setting the numerical scheme will not be determined by the requirements of DA. Instead, it will be determined by considerations related to the practical running of the model; for example, hardware architecture and parallelisation. Even with the same number of degrees of freedom, it is possible that wall times for the grid-point and DG model computations might differ.

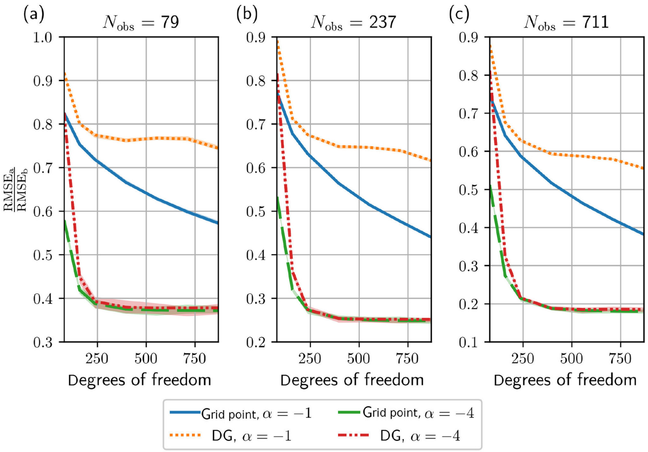


FIGURE 5 Analysis/background root-mean-squared error (RMSE) ratios as function of the number of degrees of freedom in the model when assimilating (a) $N_{\text{obs}} = 79$, (b) $N_{\text{obs}} = 237$, and (c) $N_{\text{obs}} = 711$ observations for the grid-point model with different number of grid cells and discontinuous Galerkin (DG) models with different orders but fixed number of grid cells. RMSE ratios are shown for background error spectrum $S(\kappa) = \kappa^{\alpha}$ with $\alpha = -1$ and $\alpha = -4$.

5 | DG-BASED SCALE-DEPENDENT LOCALISATION

In this section we shall consider exclusively DG models using Legendre polynomials as the basis function. We will look at their filtering properties and will show how these properties can be used to introduce a scale-dependent localisation for ensemble-based DA methods.

5.1 | Filtering properties of Legendre polynomials

The Legendre polynomials form a family of polynomials of increasing order l on domain $\tilde{D} = [-1, 1]$ given by the recursion relation (Legendre, 1785, p. 431)

$$\tilde{\phi}_l(r) = \begin{cases} 1 & \text{if } l = 0 \\ r & \text{if } l = 1 \\ \frac{2l-1}{l} r\tilde{\phi}_{l-1}(r) - \frac{l-1}{l}\tilde{\phi}_{l-2}(r) & \text{if } l > 1. \end{cases}$$

The Legendre polynomials are orthogonal; that is, $\int_{\tilde{D}} \tilde{\phi}_l(r)\tilde{\phi}_{l'}(r)\, dr = [2/(2l+1)]\delta_{ll'}$. Examples of the basis functions ϕ_{lm} based on the Legendre polynomials are shown in Figure 6a.

The power spectra in Figure 6b show that, as the polynomial order increases, a relatively larger part of the spectral power is contained in the smaller scales ($\kappa > 1$) whereas less and less ends up in the larger scales ($\kappa \sim 0$). This suggests that the Legendre polynomials can be used to construct a band-pass filter. In the following we will describe how such a filter can be constructed. For simplicity, we will assume that the model grid D consists of

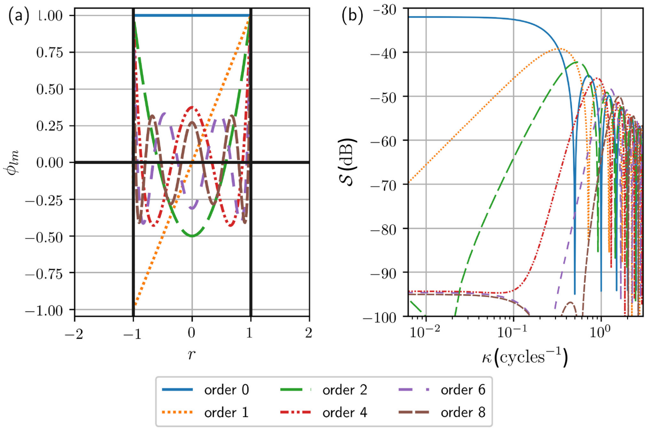


FIGURE 6 (a) Examples of Legendre basis polynomials ϕ_{lm} of different polynomial orders l on a domain consisting of 79 cells of size 2. The examples shown all have their support in the grid cell $[-1, 1]$. (b) Power spectrum of the polynomials shown in (a).

grid cells of width 2. Such a situation can always be created by applying a suitable coordinate transformation to the model's differential equations before integrating the model. After such a transformation the transform function $\Psi_m : \widetilde{D} \rightarrow D_m$ simplifies to $\Psi_m(r) = r_m - r$, where r_m is the centre of the mth grid cell. This transformation can immediately be generalised to a continuum as $\Psi_r(r') = r - r'$. Furthermore, we extend the definition of the Legendre polynomials $\tilde{\phi}_l$ from $\widetilde{D}$ to $\mathbb{R}$ as

$$\underline{\tilde{\phi}}_l(r') = \begin{cases} \tilde{\phi}_l(r') & \text{if } r' \in \widetilde{D} \\ 0 & \text{otherwise} \end{cases}$$

and use this to define a filter $\mathcal{B}_l$,

$$\begin{aligned} \mathcal{B}_l(x)(r,t) &\overset{\text{def}}{=} (x * \underline{\tilde{\phi}}_l)(r) \\ &\overset{\text{def}}{=} \int_{\mathbb{R}} \underline{\tilde{\phi}}_l(r - r')x(r',t)\, dr' \\ &= \int_{-1}^{1} \tilde{\phi}_l(r')x(\Psi_r(r'),t)\, dr', \end{aligned}$$

where $*$ denotes convolution. From the convolution theorem it now follows that

$$\begin{aligned} |\mathcal{F}\mathcal{B}_l(x)|^2(\kappa) &\overset{\text{def}}{=} |\mathcal{F}(x * \tilde{\phi}_l)|^2(\kappa) \\ &= |\mathcal{F}(x \circ \Psi_r)|^2(\kappa)|\mathcal{F}(\tilde{\phi}_l)|^2(\kappa) \\ &= |\mathcal{F}(x)|^2(\kappa)|\mathcal{F}(\tilde{\phi}_l)|^2(\kappa), \end{aligned}$$

where $\mathcal{F}$ is the Fourier transform, κ is the wave number, and by definition the power spectrum reads $\mathcal{S}(\tilde{\phi}_l) = |\mathcal{F}(\tilde{\phi}_l)|^2$. That is, the filter $\mathcal{B}_l$ modifies the spectral power contained in x at wave number κ with a factor $|\mathcal{F}(\tilde{\phi}_l)|^2(\kappa)$. The power spectra $|\mathcal{F}(\tilde{\phi}_l)|^2(\kappa)$ in Figure 6b show that the peak of this spectrum shifts to higher wave numbers as l increases. Consequently, $\mathcal{B}_l$ removes spectral power at low wave numbers (long scales) and the range of wave numbers (scales) over which this removal takes place increases with increasing l. Furthermore, using the orthogonality of the Legendre polynomials and $\Psi_{r_m} = \Psi_m$ we find that $\mathcal{B}_l(x)(r_m, t) = [2/(2l+1)]x_{lm}(t)$. Therefore, apart from a scaling constant, the DG coefficients associated with the lth-order Legendre polynomial are precisely the values of the band-pass filtered field at the grid cell centres.

The filtering properties of the Legendre polynomials are used extensively in DG models to stabilise solutions and recover high-order accuracy in models containing discontinuities (Kanevsky *et al.*, 2006). More details regarding the implementation of spectral filtering using (Legendre) polynomials in DG models and their impact on the solution can be found in Vandeven (1991); Fischer and Mullen (2001); Hesthaven and Kirby (2008). Examples of the applications of such methods in fluid dynamics are present in Blackburn and Schmidt (2003); Zhang and Stanescu (2010); Meister *et al.* (2012); Talay Akyildiz and Alshammari (2022). In the next section we will show that the use of Legendre polynomials is not limited to spectral filtering, but that they can also be used to improve ensemble estimates of the background error covariance matrix.

5.2 | Localisation matrix

In practical applications, the size of the ensemble used to estimate $\mathbf{B}$ in Equations (8) and (9) is orders of magnitude smaller than the dimension of the model space. This introduces a sampling error in the estimate $\mathbf{B}$. For correlations close to zero this sample error can overshadow the true correlation. These spurious correlations are suppressed using localisation. As discussed in Section 1, covariance localisation is to be preferred over domain localisation when dealing with dense (satellite) observations. Covariance localisation replaces $\mathbf{B}$ with $\mathbf{B} \circ \mathcal{L}$, the Schur, or entry-by-entry, product of $\mathbf{B}$ with a positive-definite, symmetric matrix $\mathcal{L}$ with both $\mathbf{B}$ and $\mathcal{L}$ acting on the full model state containing the coefficients for all the grid cells. Two types of methods are in common use to construct $\mathcal{L}$. The first is parametrised schemes in which the entries of $\mathcal{L}$ are assumed to depend on a very limited number of parameters. Parameter values are obtained from calibration experiments by minimising metrics like RMSE. The best-known scheme within this category is the Gaspari–Cohn scheme (Gaspari & Cohn, 1999), in which the entries of $\mathcal{L}$ depend on a single parameter, the localisation length scale. Other examples belonging to this category can be found in Anderson (2007); Moosavi *et al.* (2018). The second is non-parametrised schemes in which localisation factors are obtained as the expectation values of unknown distributions (Anderson, 2007, 2012; Flowerdew, 2015; Ménétrier *et al.*, 2015). These expectation values are approximated from the model under ergodicity assumptions.

Regardless of the method used, the optimal parameter values/the expectation values will depend on the signal-to-noise ratio in the sample covariance. In particular, it may be expected that if the length scales in the background errors change so does the distance at which noise starts to exceed the signal: the length scale in $\mathcal{L}$. Traditionally, localisation factors $\mathcal{L}$ have been estimated assuming a single length scale without considering the multiscale feature of the model fields. If multiple scales are present in the field then this can result in localisation factors that suppress true correlations at one scale but fail to suppress the noise at other scales. In Section 5.1 we showed that the Legendre polynomials act as a band-pass filter separating the the model field by length scale. This allows us to circumvent the aforementioned problem by applying different amounts of localisation to different scales in the signal. More formally, define

$$\mathcal{L}^{\mathrm{DG}} = \sum_{l=1}^{L-1}\sum_{l'=1}^{L-1} (\mathcal{P}^{\mathrm{DG},l})^{\mathrm{T}} \mathbf{L}^{ll'} \mathcal{P}^{\mathrm{DG},l'}, \tag{13}$$

where

$$\mathcal{P}^{\mathrm{DG},l} = \sum_{m=1}^{M} \hat{\mathbf{e}}_m \otimes \left(\otimes_{l'=0}^{L-1} \delta_{ll'} \hat{\mathbf{e}}_m^{\mathrm{T}}\right) \tag{14}$$

is the projection operator that selects only those entries from $\mathbf{x}$ that are associated with the lth basis function. Here, $\mathbf{L}^{ll'}$ only contains the localisation factors for the covariance between DG coefficients associated with Legendre polynomials l and l'. In other words, $\mathbf{L}^{ll'}$ are the $L \times L$ subtensors that act on a subspace and together make up the global localisation tensor $\mathcal{L}^{\mathrm{DG}}$ that acts on the full state consisting of all DG coefficients.

As $\mathcal{L}^{\mathrm{DG}}$ in Equation (13) requires the construction of multiple localisation matrices $\mathbf{L}^{ll'}$, the amount of calibration required makes application of the parametrised localisation scheme impractical. Instead, we construct it as $\mathbf{L}^{ll'}_{mm'} = \mathbf{L}^{ll'}[\sin(\pi|r_m - r_{m'}|/\ell)]$, where the choice of the sine function stems from the use of a periodic domain. The "kernel" matrix $\mathbf{L}^{ll'}$ is obtained using the optimal Schur filter (Ménétrier *et al.*, 2015).

For the sake of completeness we will summarise the optimal Schur filter here. We start from Ménétrier *et al.* (2015, eq. 64):

$$\mathbf{L}^{ll'}_{mm'} = \frac{N-1}{(N-2)(N+1)} \times \left(N - 1 - \frac{\mathbb{E}\left[\sum_{n=1}^{N} \mathbf{a}^{(n)}_{lm}\mathbf{a}^{(n)}_{lm} \sum_{n'=1}^{N} \mathbf{a}^{(n')}_{l'm'}\mathbf{a}^{(n')}_{l'm'}\right]}{\mathbb{E}\left[\sum_{n=1}^{N} \mathbf{a}^{(n)}_{lm}\mathbf{a}^{(n)}_{l'm'} \sum_{n'=1}^{N} \mathbf{a}^{(n')}_{lm}\mathbf{a}^{(n')}_{l'm'}\right]} \right). \tag{15}$$

Here, $\mathbb{E}[\cdot]$ is the expectation value of $\cdot$ and $\mathbf{a}$ the ensemble perturbation. An ergodicity assumption is made and the expectation values are approximated using their spatial averages

$$\begin{aligned}
\mathbb{E}\left[\sum_{n=1}^{N} \mathbf{a}^{(n)}_{lm}\mathbf{a}^{(n)}_{lm} \sum_{n'=1}^{N} \mathbf{a}^{(n')}_{l'm'}\mathbf{a}^{(n')}_{l'm'}\right] &\approx \mathbf{v}_{ll'}\left[\sin\left(\frac{\pi|r_m - r_{m'}|}{\ell}\right)\right] \\
&\overset{\mathrm{def}}{=} \frac{1}{2M}\sum_{i=1}^{M}\left(\sum_{n=1}^{N}\mathbf{a}^{(n)}_{li}\mathbf{a}^{(n)}_{li}\right)\left[\sum_{n'=1}^{N}(\mathcal{T}_{m-m'}\mathbf{a}^{(n')}_{l'})_i(\mathcal{T}_{m-m'}\mathbf{a}^{(n')}_{l'})_i \right. \\
&\quad \left. + \sum_{n'=1}^{N}(\mathcal{T}_{m'-m}\mathbf{a}^{(n')}_{l'})_i(\mathcal{T}_{m'-m}\mathbf{a}^{(n')}_{l'})_i\right] \\
&= \frac{1}{2}\left[\left(\sum_{n=1}^{N}\mathbf{a}^{(n)}_{l}\circ\mathbf{a}^{(n)}_{l}\right) * \left(\sum_{n'=1}^{N}\mathcal{S}\mathbf{a}^{(n')}_{l'}\circ\mathcal{S}\mathbf{a}^{(n')}_{l'}\right)\right]_{m-m'} \\
&\quad + \frac{1}{2}\left[\left(\sum_{n=1}^{N}\mathbf{a}^{(n)}_{l}\circ\mathbf{a}^{(n)}_{l}\right) * \left(\sum_{n'=1}^{N}\mathcal{S}\mathbf{a}^{(n')}_{l}\circ\mathcal{S}\mathbf{a}^{(n')}_{l'}\right)\right]_{m'-m}
\end{aligned} \tag{16}$$

and

$$\begin{aligned}
\mathbb{E}\left[\sum_{n=1}^{N}\sum_{n'=1}^{N} \mathbf{a}^{(n)}_{lm}\mathbf{a}^{(n)}_{l'm'}\mathbf{a}^{(n')}_{lm}\mathbf{a}^{(n')}_{k'l'm'}\right] &\approx \mathbf{c}_{ll'}\left[\sin\left(\frac{\pi|r_m - r_{m'}|}{\ell}\right)\right] \\
&\overset{\mathrm{def}}{=} \frac{1}{2M}\sum_{n'=1}^{N}\sum_{n=1}^{N}\sum_{i=1}^{M}\mathbf{a}^{(n)}_{li}\left(\mathcal{T}_{m-m'}\mathbf{a}^{(n)}_{l'} + \mathcal{T}_{m'-m}\mathbf{a}^{(n)}_{l'}\right)_i\mathbf{a}^{(n')}_{li} \\
&\quad \times\left(\mathcal{T}_{m-m'}\mathbf{a}^{(n')}_{l'} + \mathcal{T}_{m'-m}\mathbf{a}^{(n')}_{l'}\right)_i \\
&= \frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\left(\mathbf{a}^{(n)}_{l} * \mathcal{S}\mathbf{a}^{(n)}_{l'}\right)_{m-m'}\left(\mathbf{a}^{(n')}_{l} * \mathcal{S}\mathbf{a}^{(n')}_{l'}\right)_{m-m'} \\
&\quad + \frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\left(\mathbf{a}^{(n)}_{l} * \mathcal{S}\mathbf{a}^{(n)}_{l'}\right)_{m'-m}\left(\mathbf{a}^{(n')}_{l} * \mathcal{S}\mathbf{a}^{(n')}_{l'}\right)_{m'-m},
\end{aligned} \tag{17}$$

where $\mathcal{T}_j$ is a shift of the state by j grid points on a periodic grid, $\mathcal{S}$ is the mirror operator on a periodic domain, and $*$ is given by $(\mathbf{a} * \mathbf{b})_i = (1/M)\sum_{m=1}^{M}\mathbf{a}_m(\mathcal{T}_i\mathcal{S}\mathbf{b})_m$, the discrete convolution operator for a periodic domain. For non-periodic domains, $\mathcal{T}_j$ can be defined by mirroring model fields at their boundaries. The resulting expression for the localisation coefficients is

$$\begin{aligned}
\mathbf{L}^{ll'}_{mm'} &\approx \mathbf{L}^{ll'}\left[\sin\left(\frac{\pi|r_m - r_{m'}|}{\ell}\right)\right] \\
&= \frac{N-1}{(N-2)(N+1)} \\
&\quad \times\left\{N - 1 - \frac{\mathbf{v}_{ll'}\left[\sin\left(\frac{\pi|r_m - r_{m'}|}{\ell}\right)\right]}{\mathbf{c}_{ll'}\left[\sin\left(\frac{\pi|r_m - r_{m'}|}{\ell}\right)\right]}\right\}.
\end{aligned}$$

The expression for the grid-point case can be found by setting $L = 1$ and omitting any indices l, l'. A comparison of the computational cost of the scale-dependent method versus the localisation operator for the grid-point case can be found in Appendix C.

5.3 | Set-up of localisation experiments

In order to compare the differences between **B** created using scale-dependent localisation, standard scale-independent spatial localisation, and no localisation, we generate a large ensemble of 10,000 univariate ensemble members as Fourier series following the procedure outlined in Appendix A using $J = 395$ degrees of freedom in the Fourier coefficients, a pink-noise spectrum $\mathcal{S}(k) \sim \kappa^{\alpha}$, $\alpha = -1$, and a standard deviation of 1. The covariance estimated from this 10,000-member ensemble is assumed sufficiently large to have negligible sampling error.

The ensemble members generated are processed in two different ways.

1. Continuing the procedure outlined in Appendix A, each ensemble member is projected onto a fourth-order Legendre basis $(L - 1 = 4)$ defined on a periodic domain with length $\ell = 8000$ km and $M^{\mathrm{DG}} = 79$ equispaced grid cells.
2. In the grid-point approach, each ensemble member is interpolated onto an equispaced grid with J points. The grid-point state on this grid is obtained by interpolating the ensemble member expressed as Fourier series onto the grid nodes.

Subensembles of size N with sampling error are created by randomly selecting N ensemble members from the 10,000-member ensemble and storing their representation as a DG state and grid-point state. A schematic overview of the whole process is given in Figure 7.

5.4 | Resulting localisation factors

First, we verify that localisation distances are indeed scale dependent. To this end, $N = 96$ DG states are randomly selected from the 10,000-member ensemble. The localisation tensor $\mathcal{L}^{\mathrm{DG}}$ is calculated from these members using Equation (15). As a result of the ergodicity assumption in Equations (16) and (17), the localisation factors $\mathbf{L}^{ll'}_{mm'}$ are solely a function of the orders l and l' of the associated polynomials and of the difference in grid cell indices $|m - m'|$. For equispaced grids, as the one used in our experiments, the latter can be related to distance between points via $(\ell/M)|m - m'|$. For the non-scale-dependent localisation

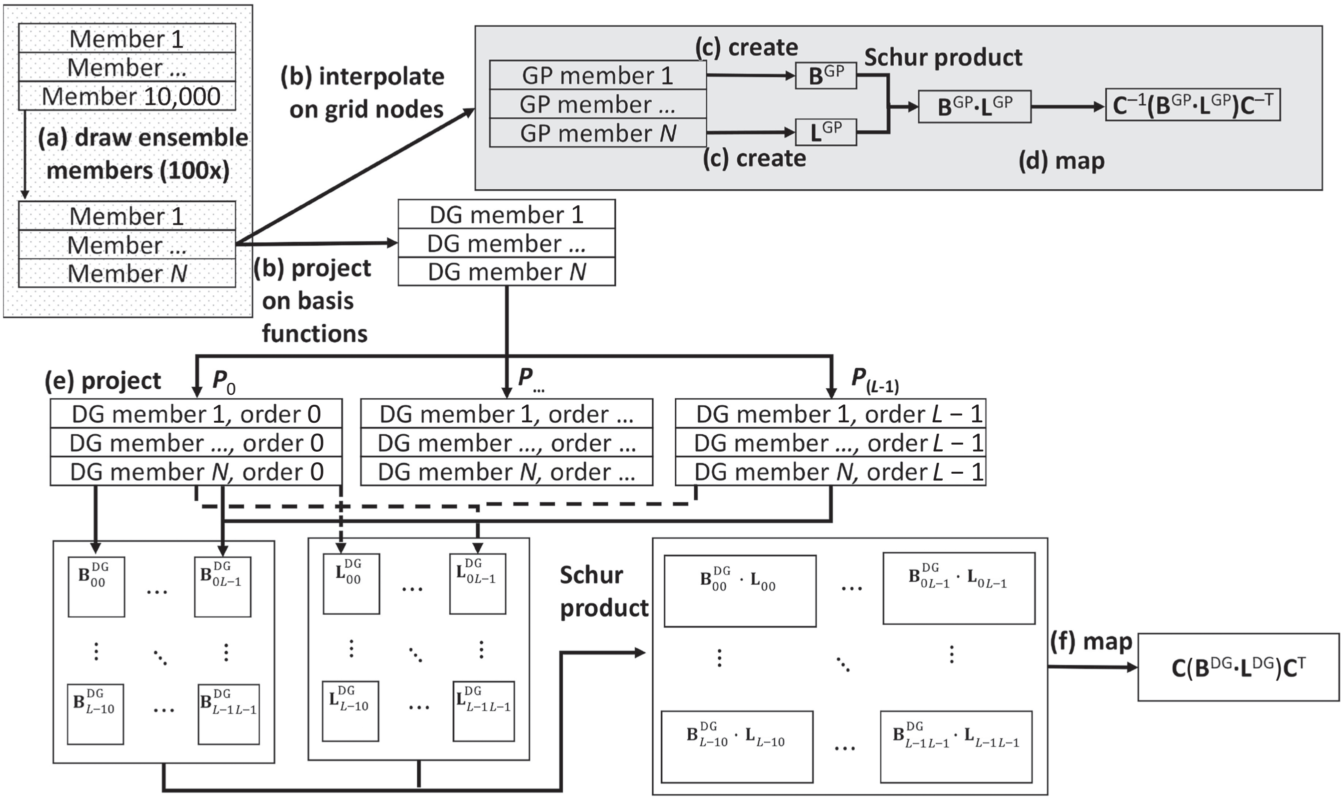


FIGURE 7 Schematic overview of the experimental set-up to create a localised covariance in grid-point space (dotted grey and solid grey background) and in discontinuous Galerkin (DG) space (dotted grey and white background). A large 10,000-member ensemble is created; (a) from this ensemble, members are randomly selected to create smaller subensembles; (b) the subensemble members are interpolated/projected to create grid-point/DG states; (c) from the grid-point states, covariance and localisation operators are created, and (d) the localised covariance operator is mapped to DG space. Simultaneously, (e) the DG states are split by the order of their coefficients, and for each pair of orders the covariances and localisation operators are calculated. (f) The resulting localised covariances are mapped from DG- to grid-point space or vice versa using the linear transform $\mathbf{C}$ defined in Equation (18). Such a transformation is only possible in this section as the number of grid points is equal to the dimension of the DG space.

factors $\mathcal{L}^{\mathrm{GP}}$, we use again Equation (15) on the grid-point states created from the same 96 ensemble members. The calculation of both $\mathcal{L}^{\mathrm{DG}}$ and $\mathcal{L}^{\mathrm{GP}}$ is repeated 100 times. Each time a different 96-member ensemble is used, rendering 100 realisations of $\mathcal{L}^{\mathrm{DG}}$ and $\mathcal{L}^{\mathrm{GP}}$. The mean of the entries in $\mathcal{L}^{\mathrm{DG}}$ and $\mathcal{L}^{\mathrm{GP}}$ over these 100 repetitions together with the 5% and 95% percentiles are shown in Figure 8.

The non-scale lines in Figure 8 show the localisation factor for the grid-point state as a function of distance between grid positions. As expected, the maximum localisation factors can be found at zero distance. Nevertheless, the inspection of the non-scale case reveals a few aspects deserving clarification. First, contrary to many other localisation schemes, the maximum localisation factor is smaller than 1 because the optimal Schur localisation scheme not only corrects for the sampling error in correlation but also for that in the sample variance (Ménétrier *et al.*, 2015). As the spectrum of the ensemble perturbations scales as $\kappa^{-\alpha}$, the first Fourier mode dominates the perturbations. As this mode is a sinusoid with wavelength ℓ, the autocorrelation as a function of distance between points first decreases towards zero and then becomes negative, whilst its absolute value increases, as distance approaches $\frac{1}{2}\ell$. As the localisation coefficient is a function of the absolute value of the autocorrelation, it also initially decreases towards zero at 1.5×10^6 m, but it starts to increase again as the distance approaches $\frac{1}{2}\ell$.

The localisation factor $\mathbf{L}^{00}_{mm'}$ closely resembles the non-scale-dependent localisation but with higher values, thus suppressing the ensemble covariance less than its non-scale-dependent counterpart. The behaviour for the higher order DG localisation factors is significantly different, as is visible in Figure 8a. Within one grid cell length the localisation factor drops from its maximum value of 0.97 to its lower bounds, which varies around 0.0. Cross-order DG localisation factors in Figure 8b also a show rapid decrease to values around 0.0, with cross-order localisation factors between zeroth-order coefficients and coefficients of order 3 and higher exhibiting no significant deviation from zero. Overall, Figure 8 shows that the localisation length scale (i.e., the length scale over which $\mathbf{L}^{ll'}_{mm'}$ goes to zero) strongly depends on the order of the coefficients l and l' being localised. In particular, the length scale for $\mathbf{L}^{00}_{mm'}$ involving solely zeroth-order DG coefficients is noticeably longer than that for any higher order DG coefficients.

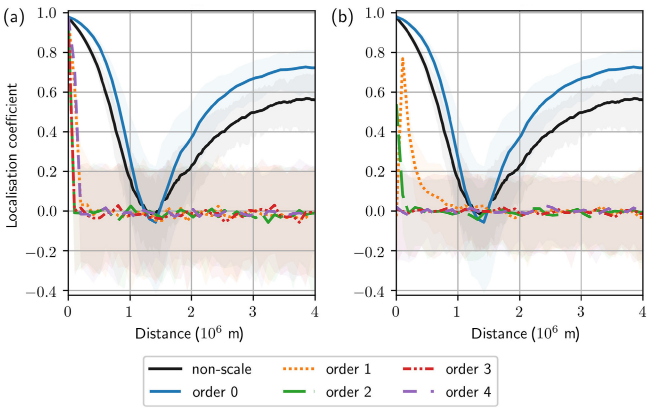


FIGURE 8 (a) Localisation coefficients $\mathbf{L}^{ll}$ (order 0-4) for different orders l as well as $\mathcal{L}^{\mathrm{GP}}$ (nonscale) as a function of distance between grid cells based on an $N = 96$-member ensemble. (b) As (a) but the colours now show cross-order coefficients $\mathbf{L}^{0l}$ between the zeroth-order coefficient and coefficients of higher orders; the black line is the same in (a) and (b).

The same localisation factors but now for an ensemble of $N = 16$ members are shown in Figure 9. Qualitatively, the dependence of the localisation factors on distance is comparable to that in Figure 8, but with lower values. Furthermore, both Figure 9 and especially Figure 8 show that the assumption that the covariance is (block-)diagonal in wavelet space—as is made in, for example, Deckmyn and Berre (2005); Pannekoucke *et al.* (2007); Chabot *et al.* (2017)—cannot be generalised to covariances in DG space: Figures 8b and 9b testify that, over short distances, cross-order localisation factors can take on values that are significantly different from zero.

5.5 | Resulting covariances

In Section 5.4 it was shown that scale-dependent localisation renders localisation factors that behave differently than those obtained by non-scale-dependent localisation. In this section, we investigate whether this results in visibly different ensemble covariances and if these covariances are a better approximations of the true covariance than those obtained with non-scale-dependent localisation.

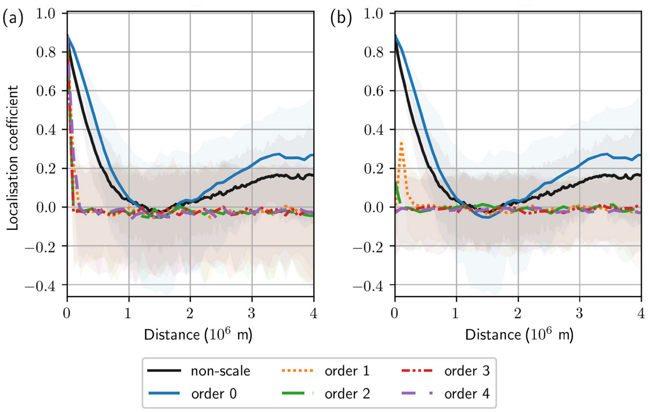


FIGURE 9 As Figure 8, but now using ensembles of $N = 16$ members.

The 10,000 ensemble members are converted to 10,000 grid-point states and 10,000 DG states. The covariance of the 10,000 grid-point states is shown in Figure 10a. A positive covariance up to 1.03 can be found in a band near the diagonal. Because of periodicity, this band is repeated near $(0, 8\times10^6)$ m and $(8\times10^6, 0)$ m. Between these bands, covariances are negative, with a low of −0.15. The equivalent covariance of DG states is shown in Figure 10b. Each box in Figure 10b is a spatial covariance, similar to the one in Figure 10a, but only involving DG coefficients of specific orders. For example, the second box (from left) on the bottom row shows the covariance $(1/N-1)\sum_{n=1}^{N}\mathbf{x}_{0m}^{(n)}\mathbf{x}_{1m'}^{(n)}$, which is the cross-covariance between the coefficients of polynomials of orders 0 and 1 at the mth and m'th grid cells respectively. We will refer to the covariance as shown in this block as the (0, 1)-DG covariance. The (0, 0)-DG covariance (bottom, left box in Figure 10b) mimics the grid-point covariance. For the (i, j)-DG covariances with $i > 0$ and/or $j > 0$, non-zero covariances are located in a small band near the diagonal, with some blocks exhibiting no covariance at all. So, when a Legendre basis is used, DG covariances exhibit distinct dependence on the order of the DG coefficient.

Not only the covariance itself, but the sampling error also depends on the order of the DG coefficients. The grid-point covariance for the 16-member ensemble is

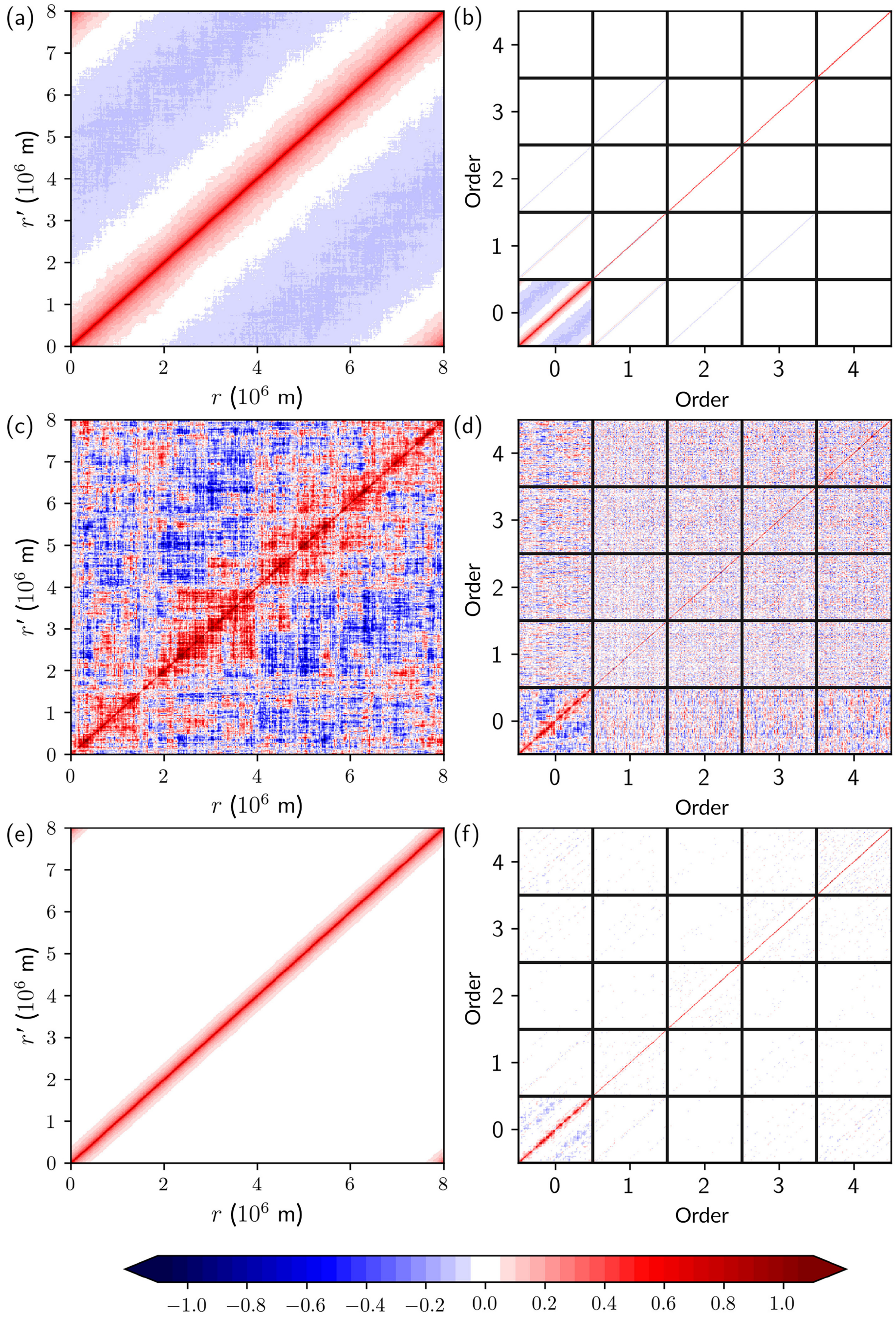


FIGURE 10 The (a, c, e) grid-point ensemble covariances and (b, d, f) discontinuous Galerkin (DG) ensemble covariances from (a, b) a 10,000-member ensemble, (c, d) a 16-member ensemble without localisation, and (e, f) a 16-member ensemble with localisation. Panels (a), (c), and (e) show the spatial covariance between grid points, whereas each tile in panels (b), (d), and (f) shows the spatial covariance between DG coefficients of the DG orders indicated on the axes.

shown in Figure 10c. As a consequence of the sampling errors, the range in values is larger ([−1.23, 2.09]) and the band structure is less discernible: we see how spurious (unrealistic) correlations appear. From the 16-member DG case in Figure 10d we see again that the (0, 0)-DG covariance mimics the grid-point covariance, but the sampling noise in the other blocks of Figure 10d exhibits less spatial structure than the error in the (0, 0)-DG covariance: the higher order DG covariances take the form of unstructured noise or unstructured noise plus positive covariance along the diagonal. That is, off-diagonal, the signal-to-noise ratio is smaller than one in these covariances. This rapid drop in signal to noise with spatial distance between points explains why the localisation factors for the higher orders in Figure 9b and Figure 8b drop rapidly with distance. The noise in Figure 10d,e is efficiently removed by the localisation (see Figure 10e,f). However, comparison of Figure 10a,b with Figure 10e,f respectively shows that it also removes the genuine negative covariances and reduces the width of the diagonal bands. This overzealous removal is not as much of a problem for the (i, j)-DG covariances with $i + j > 0$ as these covariances are truly near-diagonal.

In the aforementioned, we have established that both the covariance and the sampling errors in the covariance differ depending on the order of the DG coefficients. Now we want to determine whether accounting for those differences in the localisation procedure yields better estimates of the true covariance. To enable side-by-side comparison, either the grid-point covariance must be converted into its representation in DG space or the DG covariance must be converted into a covariance in grid-point space. A linear transformation that converts a DG state into a grid-point state can be formulated as

$$\mathbf{C} = \sum_{j=0}^{J-1}\sum_{m=1}^{M} \hat{\mathbf{e}}_j \left(\otimes_{l'=0}^{L-1} \phi_{l'm}(r_j)\delta_{ll'}\hat{\mathbf{e}}_m^{\mathrm{T}}\right), \tag{18}$$

with $r_j = (\ell/J)(j-1)$ the position of the jth grid-point point. As it is a linear transformation, the grid-point representation of the localised DG covariance is given as

$$\mathbf{C}(\mathbf{B}^{\mathrm{DG}} \circ \mathcal{L}^{\mathrm{DG}})\mathbf{C}^{\mathrm{T}}. \tag{19}$$

Since in our experiments the dimension of the grid-point space is equal to that of the DG space, Equation (18) can be inverted to get the localised grid-point covariance in the DG space as

$$\mathbf{C}^{-1}(\mathbf{B}^{\mathrm{GP}} \circ \mathcal{L}^{\mathrm{GP}})\mathbf{C}^{-\mathrm{T}}. \tag{20}$$

Similar to Figure 10e, the ensemble covariance of 16 grid-point states after non-scale-dependent localisation is shown in Figure 11a. Next to it in Figure 11b is its representation in DG space obtained by applying Equation (20) to Figure 11a. If we compare the figure with the true DG covariance in Figure 10b, we see for the (i, j)-DG covariances with $i > 0$ and/or $j > 0$ that spurious localisation appears as bands parallel to the diagonal. This indicates that a strategy in which DG states are mapped to and from grid-point states and in which localisation and DA is carried out in grid-point space is troublesome: for the high-order DG coefficients the effective background error covariances obtained in this way do not match the covariance from the ensemble of DG states (Figure 10b). On the other hand, if we start of with the DG states, localise in DG space (Figure 11d), and map this to grid-point space using Equation (19) we get a covariance (Figure 11c) that, although not flawless, better captures the width of the true covariance (Figure 10a) around the diagonal as well as the occurrence of negative covariances than the covariance obtained using non-scale-dependent localisation (Figure 10e). This improvement in the off-diagonal covariances suggests that the benefits of the scale-dependent localisation scheme are not limited to subgrid scales, but reach into the larger scales as well. The reason for this can be inferred from Figures 8 and 9: as the subgrid scales with small localisation radii are treated separately from the grid-scale processes that are represented by the zeroth-order coefficients, the localisation coefficients for the latter (blue lines in the figures) are larger than those when non-scale-dependent localisation is used (black lines). This results in the truthful long-distance covariances suffering less suppression during localisation.

To quantify the differences visible in Figure 11, the relative error between the true grid-point covariance and the covariance of grid-point states without localisation and grid-point states with non-scale-dependent localisation is shown in Figure 12 for different ensemble sizes. Also shown in the figure is the covariance created from DG states using scale-dependent localisation after mapping it to the grid-point space. The error norm in the grid-point space is calculated using two different methods: the Frobenius norm

$$\sqrt{\sum_{j=0}^{J-1}\sum_{j'=0}^{J-1}(\Delta \mathbf{B}_{ll'})^2}$$

in Figure 12a, and the operator/spectral norm, which is the largest absolute eigenvalue of the error $\Delta\mathbf{B}$, in Figure 12b. The errors are normalised by the Frobenius or spectral norm of the true covariance respectively to obtain the relative error. The computation of the norms is repeated 100 times with different ensembles and the mean of the relative norms with its 90% confidence interval is shown in Figure 12. Both non-scale- and scale-dependent

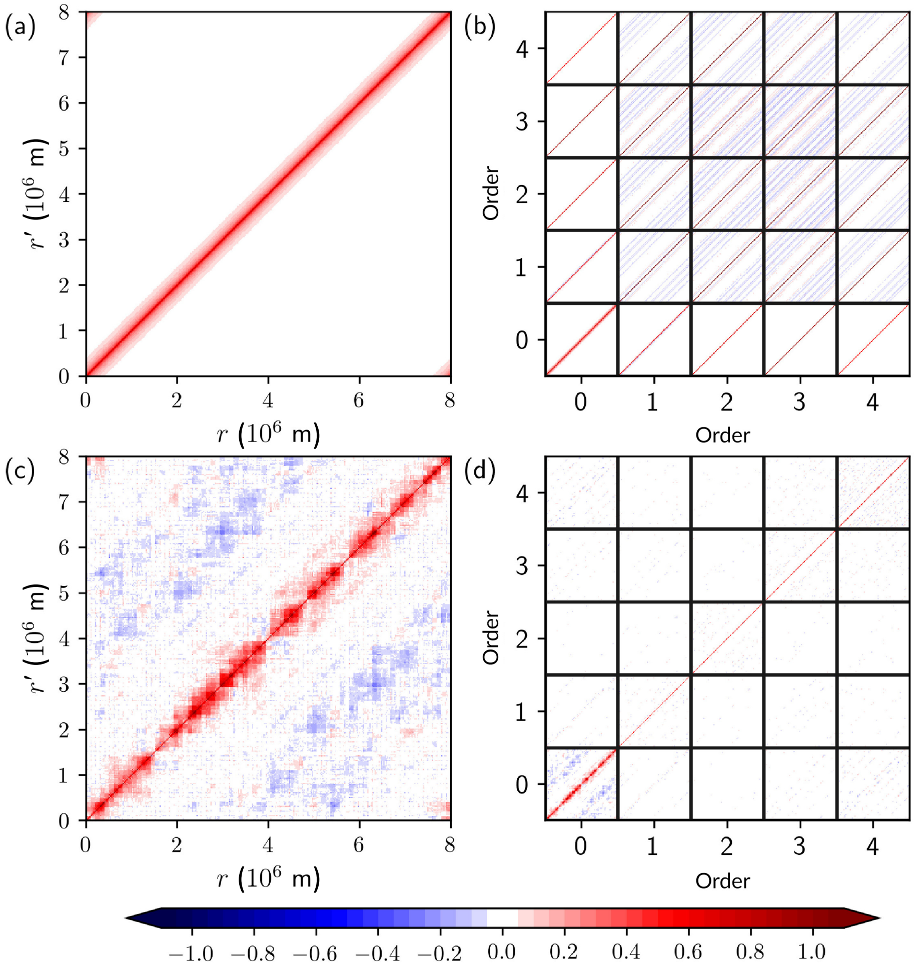


FIGURE 11 (a) The 16-member ensemble covariance localised in grid-point space and (b) its transformation into DG space. (d) The same 16-member ensemble, but now localised in DG space using scale-dependent localisation together with (c) its transformation to grid-point space.

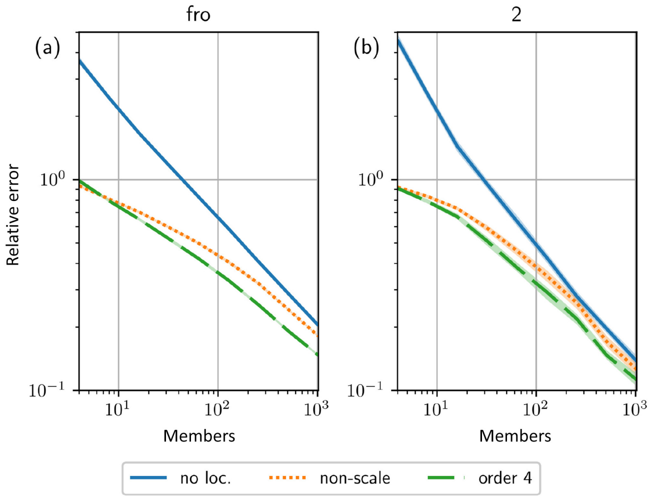


FIGURE 12 (a) Relative Frobenius and (b) operator/spectral error between the true ensemble covariance in grid-point space as a function of the number of ensemble members N for covariance without localisation, with non-scale-dependent localisation, and with scale-dependent localisation. The 90% confidence intervals in the relative error are shown as opaque shading.

localisation reduce the relative error strongly compared with no localisation, with the reduction being more profound for the smaller ensemble sizes. The scale-dependent localisation produces relative errors that for $N \geq 8$ are slightly, but significantly, better at the 90% confidence level than the covariances using non-scale-dependent localisations are. In summary, non-scale-dependent localisation applies too much localisation to the largest scales in the errors, resulting in the removal of the negative, off-diagonal covariances and too little to the smallest scales, leaving spurious off-diagonal covariances for the higher order DG coefficients. Scale-dependent localisation addresses these issues and renders an overall more accurate representation of the background error covariance matrix.

6 | DISCUSSION AND CONCLUSIONS

DG models resolve the solution of a system of partial differential equations as a linear combination of basis functions. We have shown in this study that, if Legendre polynomials are used as basis functions, these functions act as a filter bank: each polynomial in each grid cell represents portions of the solution in specific portions of the physical space and of the spectral domain. Furthermore, the polynomials resolve the solution within grid cells, not solely on a finite number of points as in the case in finite-difference or finite-volume methods. Both these characteristics can, in principle, be exploited by DA. This study is a first exploration along this line. Here, we have used a highly idealised DA set-up to investigate if, and under what conditions, these theoretical benefits can be realised. We tackled three aspects: the dependence of the DA performance on the observation density, the ability of DA–DG combination to improve estimates of the derivatives, and the DG-based scale-dependent localisation.

Dependency of RMSE ratios, ratios of norm of difference between ensemble means and the “truth,” on the observation density per grid cell shows that DA in DG models benefits from increasing observation density and, consequently, that less data thinning should be applied when deploying DA in DG models. In agreement with the data thinning finding in Liu and Rabier (2002), we see that the RMSE reduction rate levels off in all models, but in the DG models this happens at higher densities than in the grid-point model. In the set-up presented in this article, little improvement in the RMSE ratio was obtained when assimilating more than five observations per grid cell both for red and pink noise. As the ability to correct the error will depend on the extent to which the latter is resolved by the basis functions, we expect the optimal number of observations beyond which no substantial amelioration of the analysis is obtained to be different when a different family of basis functions or family of error spectra is used. Furthermore, in this study, correlations stemmed solely from representativeness errors. In reality, instruments might also introduce spatial correlations, albeit their estimation and treatment in DA is more straightforward. More details on how to deal with such inherent correlations can be found in Stewart *et al.* (2013); Campbell *et al.* (2017). When keeping the number of grid cells constant, DA performance increases with increasing order of the DG scheme. The magnitude of the latter improvement, however, depends on the scales contained in the background errors. If small-scale errors dominate, increasing DG order renders a monotonic reduction of RMSE. On the other hand, if large-scale errors dominate, increasing DG order beyond 2 yields no additional benefit. In both cases, most of the benefit is attained already for orders as low as 2, an order that is achievable in operational models and will be used in neXtSIM$_{\mathrm{DG}}$. When large scales dominate the background error, the RMSE ratios for a given DG-order are smaller (better DA performance) than when small scales dominate. This is in agreement with the findings of Fowler *et al.* (2018), who showed that DA performs best when observations and background are accurate at different scales. As we have used white, uncorrelated, noise (thus small scale) to simulate observational errors, we get the greatest reduction in RMSE ratio if the large scales dominate the error background errors.

This study could not find any indication that DG models can help DA with reducing errors in the derivatives of the field. When small-scale errors are dominant in the background errors of higher order DG models, the improvements in DA performance actually come at the cost of deteriorating the estimates of the higher order derivatives. This suggests that if accurate estimates for the derivatives are required then methods that include such derivatives in the cost function are needed. One way to achieve this would be to assimilate the derivatives or their approximations as observations. A successful example of such an approach can be found in Bédard and Buehner (2020), in which not only a thinned set of observations is assimilated, but also the differences between observation. The latter is found to reduce analysis errors for intermediate wavelengths. An alternative approach that could help with improving the higher order derivatives is adding penalty terms for discontinuities in the field and its derivatives at the cell boundaries. This would suppress the detrimental Runge effect. In any case, our findings have pointed out that the straightforward DG–DA approach followed in this work is insufficient to correct the derivatives of the field.

When the Legendre basis is used, it was found that the functional dependence of the optimal localisation coefficients depends strongly on the order of the DG coefficient and that consequently tuning localisation on the order of the DG coefficients is appropriate. When this is done, it produces covariances that are closer to the true background error covariance and are apter in retaining weak long-distance covariances. Wavelet-based localisation algorithms, like the ones in Deckmyn and Berre (2005); Pannekoucke *et al.* (2007); Varella *et al.* (2011), have the advantage that they can vary their localisation scales in space. This aspect has not been covered in this study as we assumed our background error statistics to be homogeneous in space. Because of the similarities with the schemes used in aforementioned studies, the DG-based localisation scheme in this study should be able to deal with varying correlation scales too. However, it is important to note when Legendre polynomials are used as basis functions that all scales larger than the grid cell width are grouped together. That is, as the grid cell sizes decreases, DG-based localisation is expected to tend to the same result as localisation in the grid-point space. Another well-known issue with the optimal localisation scheme, and one that has not been addressed in this study, is that localisation coefficient matrices produced by optimal localisation are not bounded to be positive definite. One proposed way to overcome this problem is to fit positive-definite kernels to $\mathcal{L}$ (Michel *et al.*, 2016). Such an approach could also directly be applied to the DG-based scale-dependent localisation. Alternatively, positive-definiteness can be achieved by actively forcing the Fourier coefficients of the distance-dependent localisation function to be positive (Bochner, 1933). Finally, the DA increments, regardless of whether scale-dependent localisation is used or not, might induce non-physical discontinuities in the solution at the grid cell boundaries. In sequential DA in an operational setting, the analysis will serve as an initial condition for a forward model and these discontinuities could give rise to numerical instability and model imbalances. The issue is not unique to DG systems, as finite-volume methods also contain discontinuities. However, the issue might be more prominent in DG systems as the presence of higher order modes might allow these discontinuities to disperse in the solutions as small-scale signal. Future work using a cycling DA system should be carried out to assess whether this leads to problems in practice and what countermeasures can be put in place.

Once these problems are solved, the scale-dependent localisation approach in this work could also be applied to schemes using scale-dependent basis functions other than Legendre polynomials. Currently, DA schemes for the third-generation wave models Wavewatch III, SWAN, and WAM only directly correct significant wave height (Almeida *et al.*, 2016; Houghton *et al.*, 2022, 2023; Moazzami *et al.*, 2017; Qi & Cao, 2015; Waters *et al.*, 2013) or do not use localisation (Caires *et al.*, 2018). In these models, the wave-number–wave direction combinations are basis functions and scale-dependent localisation can be achieved by applying different localisation coefficients to coefficients associated with different wave numbers.

The work in this study was conceived as an initial exploration of the changes in best DA practices when DA is applied to DG models and consequently was intentionally highly idealised. Nevertheless, its findings should be readily extendable to two- and three-dimensional models provided that they are defined on tensor product grids (e.g., Cartesian grids). In that case, tensor products of the Legendre polynomials can be used as basis functions and the results in this study are applicable in each direction separately. When using different grid topologies or basis functions, DA in higher order DG models is still expected to produce greater error reduction, as these models resolve the solution on a subgrid level. However, scale-dependent localisation might not be feasible, as other basis functions may not possess the same filtering properties as the Legendre polynomials.

ACKNOWLEDGEMENTS

We acknowledge the support of the Scale-Aware Sea Ice Project (SASIP) funded by Schmidt Sciences, a philanthropic initiative that seeks to improve societal outcomes through the development of emerging science and technologies. We would like to thank Piotr Minakowski and Jan Hesthaven for the their aid with understanding DG methods, Vèronique Dansereau for her guidance on sea-ice modelling, and two anonymous reviewers for their suggested improvements to the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

Code to reproduce model output and figures is available at https://doi.org/10.5281/zenodo.10647102.

ENDNOTE

[1] Flop will be treated as a *zero plural* to avoid confusion with flops as abbreviation for floating-point operations per second.

ORCID

Ivo Pasmans https://orcid.org/0000-0001-5076-5421
Yumeng Chen https://orcid.org/0000-0002-2319-6937
Alberto Carrassi https://orcid.org/0000-0003-0722-5600

REFERENCES

Adcroft, A., Campin, J.-M., Doddridge, S.D., Evangelinos, C., Ferreira, D., Fol- lows, M. et al. (2022) *MITgcm documentation*. MIT. Technical report checkpoint68l-2-g20dee6164.

Almeida, S., Rusu, L. & Soares, C.G. (2016) Data assimilation with the ensemble Kalman filter in a high-resolution wave forecasting model for coastal areas. *Journal of Operational Oceanography*, 9(2), 103–114. Available from: https://doi.org/10.1080/1755876X.2016.1244232

Anderson, J.L. (2007) Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, 230(1–2), 99–111. Available from: https://doi.org/10.1016/j.physd.2006.02.011

Anderson, J.L. (2012) Localization and sampling error correction in ensemble Kalman filter data assimilation. *Monthly Weather Review*, 140(7), 2359–2371. Available from: https://doi.org/10.1175/MWR-D-11-00013.1

Auligné, T., Ménétrier, B., Lorenc, A.C. & Buehner, M. (2016) Ensemble–variational integrated localized data assimilation. *Monthly Weather Review*, 144(10), 3677–3696. Available from: https://doi.org/10.1175/MWR-D-15-0252.1

Bannister, R.N. (2017) A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703), 607–633. Available from: https://doi.org/10.1002/qj.2982

Bédard, J. & Buehner, M. (2020) A practical assimilation approach to extract smaller-scale information from observations with spatially correlated errors: an idealized study. *Quarterly Journal of the Royal Meteorological Society*, 146(726), 468–482. Available from: https://doi.org/10.1002/qj.3687

Bellman, R. (1970) Chapter 5: The Bubnov-Galerkin method. In: *Mathematics in science and engineering*. *Methods of nonlinear analysis*, Vol. 61. Amsterdam: Elsevier, pp. 187–224. Available from: https://doi.org/10.1016/S0076-5392(08)60463-4

Berre, L., Varella, H. & Desroziers, G. (2015) Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-var at Météo-France. *Quarterly Journal of the Royal Meteorological Society*, 141(692), 2803–2812. Available from: https://doi.org/10.1002/qj.2565

Blackburn, H.M. & Schmidt, S. (2003) Spectral element filtering techniques for large eddy simulation with dynamic estimation. *Journal of Computational Physics*, 186(2), 610–629. Available from: https://doi.org/10.1016/S0021-9991(03)00088-3

Bochner, S. (1933) Monotone funktionen, stieltjessche integrale und harmonische analyse. *Mathematische Annalen*, 108(1), 378–410.

Bonavita, M., Hólm, E., Isaksen, L. & Fisher, M. (2016) The evolution of the ECMWF hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 287–303. Available from: https://doi.org/10.1002/qj.2652

Bouillon, S. & Rampal, P. (2015) Presentation of the dynamical core of neXtSIM, a new sea ice model. *Ocean Modelling*, 91, 23–37. Available from: https://doi.org/10.1016/j.ocemod.2015.04.005

Buehner, M., Caya, A., Pogson, L., Carrieres, T. & Pestieau, P. (2013) A new Environment Canada regional ice analysis system. *Atmosphere-Ocean*, 51(1), 18–34. Available from: https://doi.org/10.1080/07055900.2012.747171

Buehner, M. & Charron, M. (2007) Spectral and spatial localization of background-error correlations for data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 133(624), 615–630. Available from: https://doi.org/10.1002/qj.50

Caires, S., Kim, J. & Groeneweg, J. (2018) Korean East Coast wave predictions by means of ensemble Kalman filter data assimilation. *Ocean Dynamics*, 68(11), 1571–1592. Available from: https://doi.org/10.1007/s10236-018-1214-0

Campbell, W.F., Satterfield, E.A., Ruston, B. & Baker, N.L. (2017) Accounting for correlated observation error in a dual-formulation 4D variational data assimilation system. *Monthly Weather Review*, 145(3), 1019–1032. Available from: https://doi.org/10.1175/MWR-D-16-0240.1

Carrassi, A., Bocquet, M., Bertino, L. & Evensen, G. (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *WIREs Climate Change*, 9(5), e535. Available from: https://doi.org/10.1002/wcc.535

Chabot, V., Berre, L. & Desroziers, G. (2017) Diagnosis and normalization of gridpoint background-error variances induced by a block-diagonal wavelet covariance matrix. *Quarterly Journal of the Royal Meteorological Society*, 143(704), 1268–1279. Available from: https://doi.org/10.1002/qj.3003

Chen, C., Beardsley, R.C. & Cowles, G. (2006) An unstructured grid, finite-volume coastal ocean model (FVCOM) system. *Oceanography*, 19(1), 78–89. Available from: https://doi.org/10.5670/oceanog.2006.92

Cheng, S., Chen, Y., Aydoğdu, A., Bertino, L., Carrassi, A., Rampal, P. et al. (2023) Arctic sea ice data assimilation combining an ensemble Kalman filter with a novel Lagrangian sea ice model for the winter 2019–2020. *The Cryosphere*, 17(4), 1735–1754. Available from: https://doi.org/10.5194/tc-17-1735-2023

Cohn, S.E. (1997) An introduction to estimation theory. *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B), 257–288. Available from: https://doi.org/10.2151/jmsj1965.75.1B_257

Cooley, J.W., Lewis, P.A.W. & Welch, P.D. (1967) Historical notes on the fast Fourier transform. *Proceedings of the IEEE*, 55(10), 1675–1677. Available from: https://doi.org/10.1109/PROC.1967.5959

Deckmyn, A. & Berre, L. (2005) A wavelet approach to representing background error covariances in a limited-area model. *Monthly Weather Review*, 133(5), 1279–1294. Available from: https://doi.org/10.1175/MWR2929.1

Drüe, C. & Heinemann, G. (2004) High-resolution maps of the sea-ice concentration from MODIS satellite data. *Geophysical Research Letters*, 31(20), L20403. Available from: https://doi.org/10.1029/2004GL020808

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F. et al. (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36. Available from: https://doi.org/10.1016/j.rse.2011.11.026

Efron, B. & Tibshirani, R.J. (1994) *An introduction to the bootstrap*. *Number 57 in monographs on statistics and applied probability*. Boca Raton, FL: CRC Press.

Ehrendorfer, M. (2007) A review of issues in ensemble-based Kalman filtering. *Meteorologische Zeitschrift*, 16, 795–818. Available from: https://doi.org/10.1127/0941-2948/2007/0256

Emery, W.J., Fowler, C. & Maslanik, J. (1994) Arctic sea ice concentrations from special sensor microwave imager and advanced very high resolution radiometer satellite data. *Journal of Geophysical Research: Oceans*, 99(C9), 18329–18342. Available from: https://doi.org/10.1029/94JC01413

Evensen, G. (2021) Formulating the history matching problem with consistent error statistics. *Computational Geosciences*, 25(3), 945–970. Available from: https://doi.org/10.1007/s10596-021-10032-7

Evensen, G., Vossepoel, F.C. & van Leeuwen, P.J. (2022) *Data assimilation fundamentals: a unified formulation of the state and parameter estimation problem*. Cham: Springer Nature. Available from: https://doi.org/10.1007/978-3-030-96709-3

Fischer, P. & Mullen, J. (2001) Filter-based stabilization of spectral element methods. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 332(3), 265–270. Available from: https://doi.org/10.1016/S0764-4442(00)01763-8

Flowerdew, J. (2015) Towards a theory of optimal localisation. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1), 25257. Available from: https://doi.org/10.3402/tellusa.v67.25257

Fowler, A.M., Dance, S.L. & Waller, J.A. (2018) On the interaction of observation and prior error correlations in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(710), 48–62. Available from: https://doi.org/10.1002/qj.3183

Fritsch, F.N. & Butland, J. (1984) A method for constructing local monotone piecewise cubic interpolants. *SIAM Journal on Scientific and Statistical Computing*, 5(2), 300–304. Available from: https://doi.org/10.1137/0905021

Gaspari, G. & Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554), 723–757. Available from: https://doi.org/10.1002/qj.49712555417

Giraldo, F. (2011) *The development of the non-hydrostatic unified model of the atmosphere (NUMA)*. U.S. Naval Postgraduate School. Technical report: ADA560004.

Gong, P., Wang, J., Le, Y., Zhao, Y., Yuanyuan Zhao, L., Liang, Z.N. et al. (2013) Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34(7), 2607–2654. Available from: https://doi.org/10.1080/01431161.2012.748992

Gourmelen, N., Escorihuela, M.J., Shepherd, A., Foresta, L., Muir, A., Garcia-Mondéjar, A. et al. (2018) CryoSat-2 swath interferometric altimetry for mapping ice elevation and elevation change. *Advances in Space Research*, 62(6), 1226–1242. Available from: https://doi.org/10.1016/j.asr.2017.11.014

Gratton, S., Rincon-Camacho, M., Simon, E. & Toint, P.L. (2015) Observation thinning in data assimilation computations. *EURO Journal on Computational Optimization*, 3(1), 31–51. Available from: https://doi.org/10.1007/s13675-014-0025-4

Greybush, S.J., Kalnay, E., Miyoshi, T., Ide, K. & Hunt, B.R. (2011) Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, 139(2), 511–522. Available from: https://doi.org/10.1175/2010MWR3328.1

Gürol, S., Weaver, A.T., Moore, A.M., Piacentini, A., Arango, H.G. & Gratton, S. (2014) B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Quarterly Journal of the Royal Meteorological Society*, 140(679), 539–556. Available from: https://doi.org/10.1002/qj.2150

Hamill, T.M., Whitaker, J.S. & Snyder, C. (2001) Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11), 2776–2790. Available from: https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Harris, L., Chen, X., Putman, W., Zhou, L. & Chen, J.-H. (2021) *A scientific description of the GFDL finite-volume cubed-sphere dynamical core*. NOAA. Technical memorandum. Available from: https://doi.org/10.25923/6nh-s5897

Hebert, D.A., Allard, R.A., Joseph Metzger, E., Posey, P.G., Preller, R.H., Wallcraft, A.J. et al. (2015) Short-term sea ice forecasting: an assessment of ice concentration and ice drift forecasts using the U.S. Navy's arctic cap nowcast/forecast system. *Journal of Geophysical Research: Oceans*, 120(12), 8327–8345. Available from: https://doi.org/10.1002/2015JC011283

Hesthaven, J.S. & Kirby, R.M. (2008) Filtering in legendre spectral methods. *Mathematics of Computation*, 77(263), 1425–1452.

Hesthaven, J.S. & Warburton, T. (2007) *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*. New York: Springer Science & Business Media.

Houghton, I.A., Hegermiller, C., Teicheira, C. & Smit, P.B. (2022) Operational assimilation of spectral wave data from the Sofar spotter network. *Geophysical Research Letters*, 49(15), e2022GL098973. Available from: https://doi.org/10.1029/2022GL098973

Houghton, I.A., Penny, S.G., Hegermiller, C., Cesaretti, M., Teicheira, C. & Smit, P.B. (2023) Ensemble-based data assimilation of significant wave height from Sofar spotters and satellite altimeters with a global operational wave model. *Ocean Modelling*, 183, 102200. Available from: https://doi.org/10.1016/j.ocemod.2023.102200

Houtekamer, P.L. & Mitchell, H.L. (1998) Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3), 796–811. Available from: https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L. et al. (2018) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1257–1278. Available from: https://doi.org/10.1002/qj.3130

Janjić, T. & Cohn, S.E. (2006) Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, 134(10), 2900–2915. Available from: https://doi.org/10.1175/MWR3229.1

Kalnay, E. (2002) *Atmospheric modeling, data assimilation and predictability*. Cambridge, UK: Cambridge University Press.

Kanevsky, A., Carpenter, M.H. & Hesthaven, J.S. (2006) Idempotent filtering in spectral and spectral element methods. *Journal of Computational Physics*, 220(1), 41–58. Available from: https://doi.org/10.1016/j.jcp.2006.05.014

Kärnä, T., Kramer, S.C., Mitchell, L., Ham, D.A., Piggott, M.D. & Baptista, A.M. (2018) Thetis coastal ocean model: discontinuous Galerkin discretization for the three-dimensional hydrostatic equations. *Geoscientific Model Development*, 11(11), 4359–4382. Available from: https://doi.org/10.5194/gmd-11-4359-2018

Kühnlein, C., Deconinck, W., Klein, R., Malardel, S., Piotrowski, Z.P., Smolarkiewicz, P.K. et al. (2019) FVM 1.0: a nonhydrostatic finite-volume dynamical core for the IFS. *Geoscientific Model Development*, 12(2), 651–676. Available from: https://doi.org/10.5194/gmd-12-651-2019

Lazarus, S.M., Splitt, M.E., Lueken, M.D., Ramachandran, R., Li, X., Movva, S. et al. (2010) Evaluation of data reduction algorithms for real-time analysis. *Weather and Forecasting*, 25(3), 837–851. Available from: https://doi.org/10.1175/2010WAF2222296.1

Lea, D.J., While, J., Martin, M.J., Weaver, A., Storto, A. & Chrust, M. (2022) A new global ocean ensemble system at the met office: assessing the impact of hybrid data assimilation and inflation settings. *Quarterly Journal of the Royal Meteorological Society*, 148(745), 1996–2030. Available from: https://doi.org/10.1002/qj.4292

Legendre, A.M. (1785) *Recherches sur l'attraction des spheroides homogenes. Mémoires de mathématique et de physique: prés. à l'Académie Royale des Sciences, par divers savans, et lûs dans ses assemblées*, Vol. 1785. Frankfurt, Germany: Universitätsbibliothek Johann Christian Senckenberg, pp. 411–434.

Liu, Z.-Q. & Rabier, F. (2002) The interaction between model resolution, observation resolution and observation density in data assimilation: a one-dimensional study. *Quarterly Journal of the Royal Meteorological Society*, 128(582), 1367–1386. Available from: https://doi.org/10.1256/003590002320373337

Marras, S., Kelly, J.F., Moragues, M., Müller, A., Kopera, M.A., Vázquez, M. et al. (2016) A review of element-based Galerkin methods for numerical weather prediction: finite elements, spectral elements, and discontinuous Galerkin. *Archives of Computational Methods in Engineering*, 23(4), 673–722. Available from: https://doi.org/10.1007/s11831-015-9152-1

Massonnet, F., Mathiot, P., Fichefet, T., Goosse, H., Beatty, C.K., Vancoppenolle, M. et al. (2013) A model reconstruction of the Antarctic sea ice thickness and volume changes over 1980–2008 using data assimilation. *Ocean Modelling*, 64, 67–75. Available from: https://doi.org/10.1016/j.ocemod.2013.01.003

Meister, A., Ortleb, S. & Sonar, T. (2012) Application of spectral filtering to discontinuous Galerkin methods on triangulations.

Numerical Methods for Partial Differential Equations, 28(6), 1840–1868. Available from: https://doi.org/10.1002/num.20705

Ménétrier, B., Montmerle, T., Michel, Y. & Berre, L. (2015) Linear filtering of sample covariances for ensemble-based data assimilation. Part I: optimality criteria and application to variance filtering and covariance localization. *Monthly Weather Review*, 143(5), 1622–1643. Available from: https://doi.org/10.1175/MWR-D-14-00157.1

Mercier, F., Michel, Y., Montmerle, T., Jolivet, P. & Gürol, S. (2019) Speeding up the ensemble data assimilation system of the limited-area model of Météo-France using a block Krylov algorithm. *Quarterly Journal of the Royal Meteorological Society*, 145(720), 910–929. Available from: https://doi.org/10.1002/qj.3428

Michel, Y. (2018) Revisiting Fisher's approach to the handling of horizontal spatial correlations of observation errors in a variational framework. *Quarterly Journal of the Royal Meteorological Society*, 144(716), 2011–2025. Available from: https://doi.org/10.1002/qj.3249

Michel, Y., Ménétrier, B. & Montmerle, T. (2016) Objective filtering of the local correlation tensor. *Quarterly Journal of the Royal Meteorological Society*, 142(699), 2314–2323. Available from: https://doi.org/10.1002/qj.2824

Miyoshi, T., Kalnay, E. & Li, H. (2013) Estimating and including observation-error correlations in data assimilation. *Inverse Problems in Science and Engineering*, 21(3), 387–398. Available from: https://doi.org/10.1080/17415977.2012.712527

Moazzami, H., Siadatmousavi, S.M. & Mazaheri, S. (2017) Data assimilation for wave data in Persian Gulf using WAVEWATCH-III spectral model. *Journal of Marine Engineering*, 12(24), 115–125.

Moosavi, A., Attia, A. & Sandu, A. (2018) A machine learning approach to adaptive covariance localization. Available from: https://doi.org/10.48550/arXiv.1801.00548

Mu, L., Losch, M., Yang, Q., Ricker, R., Losa, S.N. & Nerger, L. (2018) Arctic-wide sea ice thickness estimates from combining satellite remote sensing data and a dynamic ice-ocean model with data assimilation during the CryoSat-2 period. *Journal of Geophysical Research: Oceans*, 123(11), 7763–7780. Available from: https://doi.org/10.1029/2018JC014316

Mulia, I.E., Inazu, D., Waseda, T. & Gusman, A.R. (2017) Preparing for the future Nankai trough tsunami: a data assimilation and inversion analysis from various observational systems. *Journal of Geophysical Research: Oceans*, 122(10), 7924–7937. Available from: https://doi.org/10.1002/2017JC012695

Navon, I.M. (2009) Data assimilation for numerical weather prediction: a review. In: Park, S.K. & Liang, X. (Eds.) *Data assimilation for atmospheric, oceanic and hydrologic applications*. Berlin: Springer, pp. 21–65. Available from: https://doi.org/10.1007/978-3-540-71056-1_2

Ngodock, H., Souopgui, I., Carrier, M., Smith, S., Osborne, J. & D'Addezio, J. (2020) An ensemble of perturbed analyses to approximate the analysis error covariance in 4dvar. *Tellus A: Dynamic Meteorology and Oceanography*, 72(1), 1–12. Available from: https://doi.org/10.1080/16000870.2020.1771069

Ochotta, T., Gebhardt, C., Bondatenko, V., Saupe, D. & Wergen, W. (2007) On thinning methods for data assimilation of satellite observations. In: *International Conference on IIPS for Meteorology, Oceanography and Hydrology, San Antonio, Texas*.

Oke, P.R., Brassington, G.B., Griffin, D.A. & Schiller, A. (2008) The bluelink ocean data assimilation system (BODAS). *Ocean Modelling*, 21(1), 46–70. Available from: https://doi.org/10.1016/j.ocemod.2007.11.002

Pannekoucke, O., Berre, L. & Desroziers, G. (2007) Filtering properties of wavelets for local background-error correlations. *Quarterly Journal of the Royal Meteorological Society*, 133(623), 363–379. Available from: https://doi.org/10.1002/qj.33

Pasmans, I. & Kurapov, A.L. (2019) Ensemble of 4DVARs (En4DVar) data assimilation in a coastal ocean circulation model, part I: methodology and ensemble statistics. *Ocean Modelling*, 144, 101493. Available from: https://doi.org/10.1016/j.ocemod.2019.101493

Pasmans, I., Kurapov, A.L., Barth, J.A., Michael Kosro, P. & Kipp Shearman, R. (2020) Ensemble 4DVAR (En4DVar) data assimilation in a coastal ocean circulation model. Part II: implementation offshore Oregon–Washington, USA. *Ocean Modelling*, 154, 101681. Available from: https://doi.org/10.1016/j.ocemod.2020.101681

Qi, P. & Cao, L. (2015) Establishment and tests of EnOI assimilation module for WAVEWATCH III. *Chinese Journal of Oceanology and Limnology*, 33(5), 1295–1308. Available from: https://doi.org/10.1007/s00343-015-4282-8

Quarteroni, A., Sacco, R. & Saleri, F. (2007) *Numerical mathematics. Text in applied mathematics*, Vol. 37. Berlin: Springer Science & Business Media.

Raanes, P.N., Chen, Y., Grudzien, C., Tondeur, M. & Dubois, R. (2023) *DAPPER*. Bergen, Norway: Nansen Environmental and Remote Sensing Center. Available from: https://github.com/nansencenter/DAPPER [Accessed 18th March 22].

Rainwater, S., Bishop, C.H. & Campbell, W.F. (2015) The benefits of correlated observation errors for small scales. *Quarterly Journal of the Royal Meteorological Society*, 141(693), 3439–3445. Available from: https://doi.org/10.1002/qj.2582

Rampal, P., Bouillon, S., Ólason, E. & Morlighem, M. (2016) neXtSIM: a new Lagrangian sea ice model. *The Cryosphere*, 10(3), 1055–1073. Available from: https://doi.org/10.5194/tc-10-1055-2016

Richter, T., Dansereau, V., Lessig, C. & Minakowski, P. (2023) A dynamical core based on a discontinuous Galerkin method for higher-order finite element sea ice modeling. *Geoscientific Model Development*, 16, 3907-3926. Available from: https://doi.org/10.5194/gmd-16-3907-2023

Riedel, L., De Ríos, S.O.L., Häfner, D. & Klein, O. (2020) DORiE: a discontinuous Galerkin solver for soil water flow and passive solute transport based on DUNE. *Journal of Open Source Software*, 5(52), 2313. Available from: https://doi.org/10.21105/joss.02313

Sakov, P., Counillon, F., Bertino, L., Lisæter, K.A., Oke, P.R. & Korablev, A. (2012) TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science*, 8(4), 633–656. Available from: https://doi.org/10.5194/os-8-633-2012

Sakov, P. & Oke, P.R. (2008) A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A: Dynamic Meteorology and Oceanography*, 60(2), 361–371. Available from: https://doi.org/10.1111/j.1600-0870.2007.00299.x

Samaké, A., Rampal, P., Bouillon, S. & Ólason, E. (2017) Parallel implementation of a Lagrangian-based model on an adaptive mesh in C++: application to sea-ice. *Journal of Computational*

Physics, 350, 84–96. Available from: https://doi.org/10.1016/j.jcp.2017.08.055

Sampson, C., Carrassi, A., Aydoğdu, A. & Jones, C.K.R.T. (2021) Ensemble Kalman filter for nonconservative moving mesh solvers with a joint physics and mesh location update. *Quarterly Journal of the Royal Meteorological Society*, 147(736), 1539–1561. Available from: https://doi.org/10.1002/qj.3980

Shchepetkin, A.F. & McWilliams, J.C. (2005) The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, 9(4), 347–404. Available from: https://doi.org/10.1016/j.ocemod.2004.08.002

Stewart, L.M., Dance, S.L. & Nichols, N.K. (2013) Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1), 19546. Available from: https://doi.org/10.3402/tellusa.v65i0.19546

Talay Akyildiz, F. & Alshammari, F. (2022) A new analysis of Galerkin legendre spectral methods for coupled hyperbolic/-parabolic system arising in unsteady MHD flow of Maxwell fluid and numerical simulation. *Applied Numerical Mathematics*, 176, 83–103. Available from: https://doi.org/10.1016/j.apnum.2022.02.009

Tangborn, A. (2004) Wavelet approximation of error covariance propagation in data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, 56(1), 16–28. Available from: https://doi.org/10.3402/tellusa.v56i1.14388

Toyoda, T., Fujii, Y., Yasuda, T., Usui, N., Ogawa, K., Kuragano, T. et al. (2016) Data assimilation of sea ice concentration into a global ocean–sea ice model with corrections for atmospheric forcing and ocean temperature fields. *Journal of Oceanography*, 72(2), 235–262. Available from: https://doi.org/10.1007/s10872-015-0326-0

Vandeven, H. (1991) Family of spectral filters for discontinuous problems. *Journal of Scientific Computing*, 6(2), 159–192. Available from: https://doi.org/10.1007/BF01062118

Varella, H., Berre, L. & Desroziers, G. (2011) Diagnostic and impact studies of a wavelet formulation of background-error correlations in a global model. *Quarterly Journal of the Royal Meteorological Society*, 137(658), 1369–1379. Available from: https://doi.org/10.1002/qj.845

Vos, P.E.J., Sherwin, S.J. & Kirby, R.M. (2010) From h to p efficiently: implementing finite and spectral/hp element methods to achieve optimal performance for low- and high-order discretisations. *Journal of Computational Physics*, 229(13), 5161–5181. Available from: https://doi.org/10.1016/j.jcp.2010.03.031

Waters, J., Wyatt, L.R., Wolf, J. & Hines, A. (2013) Data assimilation of partitioned HF radar wave data into Wavewatch III. *Ocean Modelling*, 72, 17–31. Available from: https://doi.org/10.1016/j.ocemod.2013.07.003

Xie, J., Counillon, F. & Bertino, L. (2018) Impact of assimilating a merged sea-ice thickness from CryoSat-2 and SMOS in the Arctic reanalysis. *The Cryosphere*, 12(11), 3671–3691. Available from: https://doi.org/10.5194/tc-12-3671-2018

Zhang, X. & Stanescu, D. (2010) Large eddy simulations of round jet with spectral element method. *Computers & Fluids*, 39(2), 251–259. Available from: https://doi.org/10.1016/j.compfluid.2009.09.002

Zhu, S., Wang, B., Zhang, L., Liu, J., Liu, Y., Gong, J. et al. (2022) A 4DEnVar-based ensemble four-dimensional variational (En4DVar) hybrid data assimilation system for global NWP: system description and primary tests. *Journal of Advances in Modeling Earth Systems*, 14(8), e2022MS003023. Available from: https://doi.org/10.1029/2022MS003023

Zuo, H., Alonso-Balmaseda, M., de Boisseson, E., Chrust, S.H.M. & de Rosnay, P. (2017) *A generic ensemble generation scheme for data assimilation and ocean analysis*. ECMWF. Technical memorandum 795. Available from: https://doi.org/10.21957/cub7mq0i4

Zuo, H., Alonso-Balmaseda, M., Mogensen, K. & Tietsche, S. (2018) *OCEAN5: the ECMWF OCEAN reanalysis system and its real-time analysis component*. ECMWF. Technical memorandum 823. Available from: https://doi.org/10.21957/la2v0442

How to cite this article: Pasmans, I., Chen, Y., Carrassi, A. & Jones, C.K.R.T. (2024) Tailoring data assimilation to discontinuous Galerkin models. *Quarterly Journal of the Royal Meteorological Society*, 1–28. Available from: https://doi.org/10.1002/qj.4737

APPENDIX A. FIELD GENERATION

Assume that the power spectrum density $\hat{S}(x)$ of the field x is given. Then, a random field on a periodic domain of length ℓ is created as

$$x(r) = \Re\left[\sum_{j=0}^{J}(\epsilon_{1,j} + i\epsilon_{2,j})\sqrt{\hat{S}(|\kappa_j|)}\,\mathrm{e}^{\mathrm{i}\kappa_j r},\right]$$

with $\Re[\cdot]$ the real part of $\cdot$, $\epsilon_{1,j}$ and $\epsilon_{2,j}$ drawn from a normal distribution, and $\kappa_j = j(2\pi/\ell)$ the wave number, $J = 829$ for the experiments in Section 4 and $J = 395$ in Section 5. Then,

$$\mathcal{F}(x)(\kappa) = \frac{1}{2}\sum_{j=-J}^{J}(1+\delta_{j0})(\epsilon_{1,j} + \mathrm{i}\ \mathrm{sign}(\kappa_j)\epsilon_{2,j}) \times \sqrt{\hat{S}(|\kappa_j|)}\delta(\kappa - \kappa_j)$$

and

$$\begin{aligned}\mathcal{S}(x)(\kappa) &= \mathbb{E}\left[\mathcal{F}(x)(\kappa)^{*}\mathcal{F}(x)(\kappa)\right] \\ &= \frac{1}{2}\sum_{j=-\lfloor\frac{J}{2}\rfloor}^{\lfloor\frac{J}{2}\rfloor}(1+\delta_{j0})\hat{S}(|\kappa_j|)\delta(\kappa-\kappa_j),\end{aligned}$$

with $\cdot^{*}$ denoting the complex conjugate. Then, in its one-sided form, $\mathcal{S}(x)(|\kappa_j|) = \hat{S}(|\kappa_j|)$ as desired.

After generation, the signal can be converted into Legendre coefficients by exploiting the orthogonality condition of the Legendre polynomials $\tilde{\phi}_l$:

$$
\begin{aligned}
\mathbf{x}_{lm} &= \frac{2l+1}{2}\int_{D_m}\phi_{lm}(r)x(r)\,dr\\
&= \frac{2l+1}{2}\int_{m\Delta r}^{m\Delta r+\Delta r}\tilde{\phi}_l\left(\frac{2r-2m\Delta r}{\Delta r}-1\right)x(r)\,dr\\
&= \frac{2l+1}{2}\Re\left[\sum_{j=1}^{J}(\epsilon_{1,j}+\mathrm{i}\epsilon_{2,j})\sqrt{\widehat{S}(|\kappa_j|)}\right.\\
&\quad\left.\int_{m\Delta r}^{m\Delta r+\Delta r}\tilde{\phi}_l\left(\frac{2r-2m\Delta r}{\Delta r}-1\right)\,\mathrm{e}^{\mathrm{i}\kappa_j r}\,dr\right]\\
&\quad+\frac{1}{2}\sqrt{\widehat{S}(|\kappa_0|)}\epsilon_{1,0}\\
&\quad\times\int_{m\Delta r}^{m\Delta r+\Delta r}\tilde{\phi}_l\left(\frac{2r-2m\Delta r}{\Delta r}-1\right)\,dr
\end{aligned}
\tag{A.1}
$$

with $\Delta r = \ell/N$ the grid cell width and $2/(2l+1)$ a normalisation factor. As the functions $\tilde{\phi}_l$ are polynomials, the integrals in Equation (A.1) can be carried out using integration by parts.

APPENDIX B. PROJECTION OPERATOR

The "truth" x^{truth} is generated as a Fourier series. The projection on the basis function ϕ_{lm} can be found by integration, exploiting the fact that the basis is orthonormal and that ϕ_{lm} has its support in grid cell m:

$$
\begin{aligned}
x^{\mathrm{truth}}(r) &= \Re\left[\int_0^{\ell}\phi_{lm}(r)\sum_{j=0}^{J}A_j\,\mathrm{e}^{\mathrm{i}\kappa_j r}\right]\\
&= \Re\left[\int_{r_m-\frac{1}{2}\Delta r}^{r_m+\frac{1}{2}\Delta r}\phi_{lm}(r)\sum_{j=0}^{J}A_j\,\mathrm{e}^{\mathrm{i}\kappa_j r}\,dr\right]\\
&= \Re\left[\int_{-1}^{1}P(r')\sum_{j=0}^{J}A_j\,\mathrm{e}^{\mathrm{i}\kappa r_m}\,\mathrm{e}^{\mathrm{i}\Delta r\kappa r'/2}\,dr'\right]\\
&= \Re\left[\sum_{j=0}^{J'}A'_j\int_{-1}^{1}P_l(r')\,\mathrm{e}^{\mathrm{i}\kappa'_j r'}\,dr'\right]\\
&\quad+\Re\left[\sum_{j=J'+1}^{J}A'_j\int_{-1}^{1}P_l(r')\,\mathrm{e}^{\mathrm{i}\kappa'_j r'}\,dr'\right],
\end{aligned}
$$

with $\Delta r = \ell/M$ the grid cell width, $\kappa_j = 2\pi j/\ell$ the wave number, $\kappa'_j = \pi\Delta r j/\ell$ the scaled wave number, and $A'_j = \frac{1}{2}\Delta r A_j$ the scaled Fourier amplitude.

Let J' be the largest value of j for which $\kappa'_j < 0.2$. Then, for $j > J'$, the integrals by applying integration by parts repeatedly give

$$
\int_{-1}^{1}P_l(r')\,\mathrm{e}^{\mathrm{i}\kappa'_j r'}\,dr' = \sum_{l'=0}^{l}(-1)^{l'}(\mathrm{i}\kappa'_j)_j^{-l'-1}\frac{d^{l'}P_l}{dr'^{l'}}(r')\,\mathrm{e}^{\mathrm{i}\kappa'_j r'}\Bigg|_{-1}^{1}
$$

For $\kappa'_j < 0.2$ the terms $\kappa^{-l\prime-1}$ become so large that the summation fails as a result of numerical errors. Therefore, integration for $0 < j \leq J'$ is carried out using Gaussian quadrature:

$$
\mathrm{e}^{\mathrm{i}\kappa'_j r'} \approx Q(r') \overset{\mathrm{def}}{=} \sum_{q=1}^{8}\mathrm{e}^{\mathrm{i}\kappa\prime_j r\prime_q}\prod_{s=1,s\neq q}^{8}\frac{r'-r'_s}{r'_q-r'_s},
$$

with r'_q the Gauss–Lobatto points on the interval $[-1,1]$, and

$$
\int_{-1}^{1}P_l(r')\,\mathrm{e}^{\mathrm{i}\kappa'_j r'}\Delta r' \approx \int_{-1}^{1}P_l(r')Q(r')\,dr'.
$$

APPENDIX C. COMPUTATIONAL COST SCALE-DEPENDENT LOCALISATION

To apply covariance localisation in the D-E3DVar algorithm, $\mathbf{B}$ in Equation (10) needs to be replaced with $\mathbf{B}\circ\mathcal{L}$. As D-E3DVar is an iterative solver, it requires repeated application of $\mathbf{B}\circ\mathcal{L}$ to a vector of the form $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1/2}\chi$ with $\chi \in \mathbb{R}^{N_{\mathrm{obs}}}$. Here, we provide an estimate of the algorithm complexity as measured by the number of floating-point operations, "flop"[1] for short, necessary for the application of $\mathbf{B}\circ\mathcal{L}$ to a vector $\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1/2}\chi$. Flop necessary to calculate the latter vector are not included as they are necessary regardless of whether localisation is used. The number of flop will be listed in big-O notation: $\mathcal{O}(N)$ means there is $c \in \mathbb{R}$: flop $< cN$.

For given l and l', $\mathbf{L}^{ll'}_{mm'}$ can be calculated using the convolutions in Equations (16) and (17). Calculation of each term in the convolution in Equation (16) requires $\mathcal{O}(MN)$ flop. The convolution can be carried out most efficiently using a fast-Fourier transform at the expense of $\mathcal{O}(M\log M)$ flop (Cooley *et al.*, 1967). For Equation (17), N convolutions need to be carried out at a total cost of $NM\log M$ flop and the results need to be multiplied and summed, costing another $\mathcal{O}(MN)$ flop. The division, summation, and scalar multiplication in Equation (15) require an additional $\mathcal{O}(M)$ flop, bringing the total cost to calculate $\mathbf{L}^{ll'}_{mm'}$ for all m and m' combined to $\mathcal{O}(NM\log M)$.

Each entry of $(\mathbf{B}^{\mathrm{DG}}\circ\mathcal{L}^{\mathrm{DG}})\chi$ can be written as

$$
\left((\mathbf{B}^{\mathrm{DG}}\circ\mathcal{L}^{\mathrm{DG}})\chi\right)_{lm} = \sum_{l'=0}^{L_k-1}\sum_{m'=1}^{M}\frac{1}{N-1}\times\sum_{n=1}^{N}\mathbf{a}^{(n)}_{lm}\mathbf{a}^{(n)}_{l'm'}\mathbf{L}^{ll'}_{mm'}\chi_{l'm'}.
\tag{C.1}
$$

Based on the foregoing, for each l, $\mathcal{O}(LNM\log M)$ flop are necessary for $\mathbf{L}^{ll'}_{mm'}$. For each l and m, $\mathcal{O}(LMN)$ multiplications/additions appear in Equation (C.1). Thus, in total, the calculation of $(\mathbf{B}^{\mathrm{DG}}\circ\mathcal{L}^{\mathrm{DG}})\chi$ requires

$\sum_{l=0}^{L-1}\mathcal{O}(LNM\log M)+M\mathcal{O}(LMN)=\mathcal{O}(L^2NM\log M)+\mathcal{O}(L^2NM^2)$ flop.

The number of flop for non-scale-dependent localisation on a grid with M^{GP} nodes can be found by setting $L=1$ in the foregoing and is $\mathcal{O}(NM^{\mathrm{GP}}\log M^{\mathrm{GP}})+\mathcal{O}(N(M^{\mathrm{GP}})^2)$. So, for equal number of grid cells, $M^{\mathrm{GP}}=M$, the scale-dependent localisation is L^2 times more expensive than the non-scale-dependent localisation. If the dimension of the state is equal in both cases (i.e., $M^{\mathrm{GP}}=LM$), then the scale-dependent approach is more expensive by

$$\frac{1+(M)^{-1}\log M}{1+(LM)^{-1}\log(LM)}<1+(M)^{-1}\log M\leq 1+\mathrm{e}^{-1}.$$

So, if DG and grid-point space have similar dimensions, scale-dependent localisation can be up to 37% more expensive than non-scale-dependent localisation.