Modelling spatially correlated observation errors in variational data assimilation using a diffusion operator on an unstructured mesh

O. Guillet[1,2*], A. T. Weaver[1], X. Vasseur[4], Y. Michel[2], S. Gratton[3] and S. Gürol[1]

[1] *CERFACS / CECI CNRS UMR 5318, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France*

[2] *CNRM UMR 3589, Météo-France and CNRS, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France*

[3] *INPT-IRIT, University of Toulouse and ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse Cedex 7, France*

[4] *ISAE-SUPAERO, University of Toulouse, 10 avenue Edouard Belin, BP 54032, 31055 Toulouse Cedex 4, France*

*Correspondence to: Oliver Guillet, Météo-France, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France.

E-mail: oliver.guillet@meteo.fr

We propose a method for representing spatially correlated observation errors in variational data assimilation. The method is based on the numerical solution of a diffusion equation, a technique commonly used for representing spatially correlated background errors. The discretization of the pseudo-time derivative of the diffusion equation is done implicitly using a backward Euler scheme. The solution of the resulting elliptic equation can be interpreted as a correlation operator whose kernel is a correlation function from the Matérn family.

In order to account for the possibly heterogeneous distribution of observations, a spatial discretization technique based on the finite element method (FEM) is chosen where the observation locations are used to define the nodes of an *unstructured* mesh on which the diffusion equation is solved. By construction, the method leads to a convenient operator for the *inverse* of the observation error correlation matrix, which is an important requirement when applying it with standard minimization algorithms in variational data assimilation. Previous studies have shown that spatially correlated observation errors can also be accounted for by assimilating the observations together with their directional derivatives up to arbitrary order. In the continuous framework, we show that the two approaches are formally equivalent for certain parameter specifications. The FEM provides an appropriate framework for evaluating the derivatives numerically, especially when the observations are heterogeneously distributed.

Numerical experiments are performed using a realistic data distribution from the Spinning Enhanced Visible and InfraRed Imager (SEVIRI). Correlations obtained with the FEM-discretized diffusion operator are compared with those obtained using the analytical Matérn correlation model. The method is shown to produce an accurate representation of the target Matérn function in regions where the data are densely distributed. The presence of large gaps in the data distribution degrades the quality of the mesh and leads to numerical errors in the representation of the Matérn function. Strategies to improve the accuracy of the method in the presence of such gaps are discussed.

Keywords : observation errors; correlation functions; diffusion operator; variational assimilation; unstructured mesh; finite element method

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.3237

1. Introduction

Specifying background and observation error covariance matrices ($\boldsymbol{B}$ and $\boldsymbol{R}$, respectively) that are accurate approximations of the true error covariance matrices is a challenging problem in operational data assimilation for the atmosphere and ocean. Over the past two decades, there has been considerable research devoted to the estimation of background error covariances, notably through the use of ensemble methods, and to the development of covariance models for representing them efficiently in $\boldsymbol{B}$ (*e.g.*, see the review articles by Bannister (2008a,b, 2017)). Comparatively fewer studies have addressed the estimation and modelling of observation error covariances, especially correlations. One key aspect of the problem for variational data assimilation is that standard minimization algorithms require an operator for the precision matrix $\boldsymbol{R}^{-1}$, either for the computation of the gradient of the cost function or for preconditioning (Michel 2018). Thus, even if we have an accurate $\boldsymbol{R}$ operator at our disposal, we still need to specify an efficient $\boldsymbol{R}^{-1}$ operator for computational purposes. Designing such an operator for large problems can be difficult.

In practice, certain assumptions are invoked that greatly simplify the structure of $\boldsymbol{R}$ that is specified in operational data assimilation systems. In particular, observation errors from one observing system are assumed to be uncorrelated with those from another observing system. In a multi-instrument observing system, this assumption is usually extended to the individual instruments themselves. As a result, $\boldsymbol{R}$ is defined as a block-diagonal matrix where the specification of the observation error covariances associated with each block can be treated independently for each observing system or instrument. Despite this simplification, each block typically corresponds to a large number of observations (*e.g.,* several millions for certain satellite observations).

Satellite radiance observations are well known to have correlated errors. For example, significant horizontal error correlations in radiances have been diagnosed by Bormann *et al.* (2010) for the Infrared Atmospheric Sounding Interferometer (IASI) and by Waller *et al.* (2016a) for the Spinning Enhanced Visible and InfraRed Imager (SEVIRI). Furthermore, highly correlated observation errors are expected from future satellite instruments, such as infrared sounders of the Meteosat Third Generation, which will provide high-resolution information about water vapour and temperature structures of the atmosphere (Stuhlmann *et al.* 2005).

For radiances, it is customary to separate the vertical (or inter-channel) correlations from the horizontal spatial correlations. In recent years, substantial progress has been made in representing inter-channel error correlations (Bormann *et al.* 2010; Bormann and Bauer 2010; Stewart *et al.* 2014; Weston *et al.* 2014; Waller *et al.* 2016a; Campbell *et al.* 2017). The size of the matrices required to represent these correlations is rather small (less than 10^3 rows or columns), which makes them straightforward to handle computationally using, for example, Cholesky decomposition. A similar technique has been used by Järvinen *et al.* (1999) to model temporal correlations in surface pressure observations.

Matrices associated with horizontal correlations are much larger than those associated with inter-channel or temporal correlations. Furthermore, due to the irregular nature of the spatial distribution of observations, they tend to have more complicated structure than those associated with correlated background error. These two features make horizontally correlated observation error difficult to handle computationally. For this reason, horizontal correlations are often neglected altogether, although this has to be done with caution, especially when considering high-density observations.

Rather than explicitly accounting for horizontal correlations in $\boldsymbol{R}$, mitigating strategies such as variance inflation, thinning and superobbing are typically employed (Rabier 2006). Inflating the observation error variances has the effect of downweighting the influence of the observations, as is the case when correlations are explicitly accounted for. Thinning is used to reduce the spatial and spectral resolution of the observations (and hence their error correlations) by selecting a reduced set of locations and channels. Superobbing combines locations or channels at different positions and can help reduce the observation error variances as well as their correlations. However, these procedures are ultimately suboptimal as they involve discarding potentially

valuable observational information (Liu and Rabier 2002; Dando *et al.* 2007; Stewart *et al.* 2008).

Brankart *et al.* (2009) proposed a method to account for spatially correlated observation errors by focusing on modelling $\boldsymbol{R}^{-1}$ rather than $\boldsymbol{R}$. In particular, assuming that $\boldsymbol{R}$ is constructed from an exponential function, then $\boldsymbol{R}^{-1}$ is very sparse and can be accounted for indirectly by assimilating the observations together with their spatial derivatives, where the weights given to the spatial derivatives are related to the length-scale of the (exponential) correlation function. Chabot *et al.* (2015) discuss a related technique to account for spatially correlated errors in image observations. The method is appealing especially when the observations are sufficiently structured to simplify the computation of the spatial derivatives. For example, Ruggiero *et al.* (2016) used the Brankart *et al.* (2009) method to account for spatially correlated observation errors in simulated altimeter products from the future Surface Water and Ocean Topography (SWOT) satellite mission.

Following earlier work by M. Fisher at the European Centre for Medium-Range Weather Forecasts, Michel (2018) has shown that it is possible to carry out the main correlation operator computations on an auxiliary grid with simplified structure. The correlation operator in the space of the (possibly unstructured) observations is then obtained using an interpolation operator and its adjoint. While the method provides an efficient model for $\boldsymbol{R}$, it does not lead to a convenient and inexpensive expression for $\boldsymbol{R}^{-1}$, as required for variational data assimilation. Michel (2018) used a sequential Lanczos algorithm to build a low-rank approximation of $\boldsymbol{R}^{-1}$ in terms of its dominant eigenpairs. However, the method can be costly, as many eigenpairs may be required to obtain an adequate approximation of $\boldsymbol{R}^{-1}$.

In this article, we present an alternative method for modelling spatially correlated observation errors in variational data assimilation. Our starting point for modelling correlations in $\boldsymbol{R}$ is the framework for modelling correlations in $\boldsymbol{B}$ for which an extensive body of research exists. However, many of the standard methods used for modelling background error correlations, such as those based on spectral or (first-generation) wavelet transforms, require structured grids and thus are not appropriate for modelling $\boldsymbol{R}$. Diffusion operators can be used to model a class of correlation functions from the Matérn family (Guttorp and Gneiting 2006) and are popular for modelling $\boldsymbol{B}$ in ocean applications of variational data assimilation (Weaver and Courtier 2001; Carrier and Ngodock 2010). For numerical applications, the diffusion method provides useful flexibility regarding the choice of spatial and temporal discretization schemes. In particular, spatial discretization schemes based on the Finite Element Method (FEM) or Finite Volume Method (FVM) can be used to adapt the diffusion operator to an unstructured mesh, as desired for modelling $\boldsymbol{R}$. Furthermore, temporal discretization schemes based on backward Euler implicit methods provide immediate access to an inverse correlation operator, which greatly simplifies the specification of $\boldsymbol{R}^{-1}$.

A similar method for modelling spatial correlations on an unstructured mesh was developed by Lindgren *et al.* (2011) for spatial interpolation (kriging) applications in geostatistics and by Bui-Thanh *et al.* (2013) for modelling prior (background) error covariances in a seismic inverse problem. Lindgren *et al.* (2011) (see also Simpson *et al.* (2012) and Bolin and Lindgren (2013)) use the fact that Gaussian fields with a specific covariance function are solutions to a linear stochastic partial differential equation (SPDE). Solving the SPDE is a convenient way of imposing this specific covariance structure on a random field. In fact, the SPDE can be interpreted as a stochastic diffusion equation and is related to the "square-root" of a diffusion-based covariance operator. In Lindgren *et al.* (2011), the SPDE is discretized on a triangular two-dimensional (2D) mesh, where the nodes of the mesh include the observation locations as well as other locations where interpolated values are desired. In our approach, the diffusion equation is also discretized on a triangular 2D mesh, which is built exclusively from the observation locations (*i.e.*, there are no additional nodes as in Lindgren *et al.* (2011)). This approach allows spatial correlations to be modelled directly between observation locations, as required for $\boldsymbol{R}$.

The structure of the article is as follows. Section 2 introduces the theoretical framework for correlation modelling with the diffusion equation. In particular, this section discusses the relationship between the diffusion equation and correlation functions from the Matérn family. Generalizations of the method are then introduced and discussed within the context of modelling

$\boldsymbol{R}$ for certain observation types. Section 3 addresses the issue of discretizing the diffusion equation on unstructured grids using the FEM. Here, we derive explicit expressions for the matrix operators $\boldsymbol{R}$ and $\boldsymbol{R}^{-1}$ in terms of the components of the FEM-discretized diffusion operator. In Section 4, we establish a formal link between the diffusion-based approach and the method of Brankart *et al.* (2009) that involves assimilating successive derivatives of the observed field, up to a certain order. In Section 5, the diffusion method is applied to a realistic distribution of observations from SEVIRI and the accuracy of the method is assessed by comparing results with the analytical Matérn correlation model. Section 6 provides a summary and discusses future research directions for improving the accuracy of the method.

2. Correlation modelling with a diffusion operator

In this section, we introduce key aspects of the theory of diffusion-based correlation operators, as required for our study. We focus on correlation operators defined on the Euclidean space $\mathbb{R}^2$ and subdomains of $\mathbb{R}^2$ since the application considered in Section 5 concerns the modelling of 2D spatial observation error correlations on a plane. The reader can find a more general presentation in Weaver and Mirouze (2013), and references therein, where diffusion-based correlation operators are formulated on Euclidean spaces other than $\mathbb{R}^2$ and on the sphere $\mathbb{S}^2$.

In what follows, we will adopt the notation where continuous functions and operators are in italics, while vectors and matrices are in boldface.

2.1. Correlation and covariance operators

We consider correlation operators on the spatial domain Ω contained in $\mathbb{R}^2$. Let $f : \boldsymbol{z} \mapsto f(\boldsymbol{z})$ be a square-integrable function ($f \in L^2(\Omega)$) of the spatial coordinates $\boldsymbol{z} = (z_1, z_2)^{\mathrm{T}} \in \Omega$. A *correlation operator* $\mathcal{C} : f \mapsto \mathcal{C}[f]$ is an integral operator of the form

$$\mathcal{C}[f](\boldsymbol{z}) = \int_\Omega c(\boldsymbol{z}, \boldsymbol{z}') f(\boldsymbol{z}') \mathrm{d}\boldsymbol{z}', \tag{1}$$

where $\mathrm{d}\boldsymbol{z} = \mathrm{d}z_1 \mathrm{d}z_2$ is the Lebesgue measure on $\mathbb{R}^2$, and $c : (\boldsymbol{z}, \boldsymbol{z}') \mapsto c(\boldsymbol{z}, \boldsymbol{z}')$ is a correlation function where $(\boldsymbol{z}, \boldsymbol{z}') \in \Omega \times \Omega$. The correlation operator is symmetric and positive definite in the sense of the $L^2(\Omega)$-inner product:

$$\begin{aligned} \int_\Omega \mathcal{C}[f_1](\boldsymbol{z})\, f_2(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} &= \int_\Omega f_1(\boldsymbol{z})\, \mathcal{C}[f_2](\boldsymbol{z})\, \mathrm{d}\boldsymbol{z}, \\ &\quad \forall f_1, f_2 \in L^2(\Omega), \\ \int_\Omega \mathcal{C}[f_1](\boldsymbol{z})\, f_1(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} &> 0, \quad \forall f_1 \neq 0 \in L^2(\Omega). \end{aligned} \tag{2}$$

Notably, the first of these properties implies that the correlation function is symmetric: $c(\boldsymbol{z}, \boldsymbol{z}') = c(\boldsymbol{z}', \boldsymbol{z})$ for any pair $(\boldsymbol{z}, \boldsymbol{z}') \in \Omega \times \Omega$. A correlation function also has unit amplitude ($c(\boldsymbol{z}, \boldsymbol{z}) = 1$).

For data assimilation, we need to define *covariance operators*. In particular, if $\bar{f} \in L^2(\Omega)$ then $\mathcal{R} : \bar{f} \mapsto \mathcal{R}[\bar{f}]$ is the observation error covariance operator defined as

$$\mathcal{R}[\bar{f}](\boldsymbol{z}) = \int_\Omega \bar{c}(\boldsymbol{z}, \boldsymbol{z}') \bar{f}(\boldsymbol{z}') \mathrm{d}\boldsymbol{z}', \tag{3}$$

where $\bar{c} : \boldsymbol{z}, \boldsymbol{z}' \mapsto \bar{c}(\boldsymbol{z}, \boldsymbol{z}') = \sigma(\boldsymbol{z})\sigma(\boldsymbol{z}')c(\boldsymbol{z}, \boldsymbol{z}')$ is the covariance function, and $\sigma(\boldsymbol{z}) = \sqrt{\bar{c}(\boldsymbol{z}, \boldsymbol{z})}$ is the standard deviation at the location $\boldsymbol{z}$, which we assume is non-zero so that $\mathcal{R}$ is strictly positive definite. Combining Eqs (1) and (3) yields the standard relationship

$$\left.\begin{aligned} f(\boldsymbol{z}') &= \sigma(\boldsymbol{z}')\bar{f}(\boldsymbol{z}') \\ \mathcal{R}[\bar{f}](\boldsymbol{z}) &= \sigma(\boldsymbol{z})\, \mathcal{C}[f](\boldsymbol{z}) \end{aligned}\right\}, \tag{4}$$

which allows us to separate the specification of $\sigma(\boldsymbol{z})$ and $\mathcal{C}$. In this study, we focus on computational aspects of specifying $\mathcal{C}$.

2.2. Inverse correlation and inverse covariance operators

The inverse of the correlation operator $\mathcal{C}$ is defined as the operator $\mathcal{C}^{-1} : g \mapsto \mathcal{C}^{-1}[g] = f$ where $g = \mathcal{C}[f]$. The inverse correlation operator is also symmetric and positive definite in the sense of the $L^2(\Omega)$-inner product:

$$\begin{aligned} \int_\Omega \mathcal{C}^{-1}[g_1](\boldsymbol{z})\, g_2(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} &= \int_\Omega g_1(\boldsymbol{z})\, \mathcal{C}^{-1}[g_2](\boldsymbol{z})\, \mathrm{d}\boldsymbol{z}, \\ &\quad \forall g_1, g_2 \in L^2(\Omega), \\ \int_\Omega \mathcal{C}^{-1}[g_1](\boldsymbol{z})\, g_1(\boldsymbol{z})\, \mathrm{d}\boldsymbol{z} &> 0, \quad \forall g_1 \neq 0 \in L^2(\Omega). \end{aligned}$$

In general, $\mathcal{C}^{-1}$ is a differential operator, which cannot be expressed as an integral operator with an ordinary function as

its kernel (as in Eq. (1)). However, it is possible to express $\mathcal{C}^{-1}$ as an integral operator if the kernel is considered to be a sum of generalized functions (Jones 1982).

If $\bar{g} = \mathcal{R}[\bar{f}]$ then the inverse of the covariance operator $\mathcal{R}$ is the operator $\mathcal{R}^{-1} : \bar{g} \mapsto \mathcal{R}^{-1}[\bar{g}]$ where

$$\left.\begin{aligned} g(\boldsymbol{z}') &= \frac{1}{\sigma(\boldsymbol{z}')}\,\bar{g}(\boldsymbol{z}') \\ \mathcal{R}^{-1}[\bar{g}](\boldsymbol{z}) &= \frac{1}{\sigma(\boldsymbol{z})}\,\mathcal{C}^{-1}[g](\boldsymbol{z}) \end{aligned}\right\}. \tag{5}$$

2.3. *Matérn correlation functions*

A well-known class of isotropic and homogeneous correlation functions is the Matérn class (Stein 1999; Guttorp and Gneiting 2006). Here, we are interested in a subclass of Matérn functions that have the specific form

$$c_{m,\ell}(r) = \frac{2^{2-m}}{(m-2)!}\left(\frac{r}{\ell}\right)^{m-1} K_{m-1}\left(\frac{r}{\ell}\right), \tag{6}$$

where $m > 1$ is an integer, $K_m(\cdot)$ is the modified Bessel function of the second kind of order m, $r = \|\boldsymbol{z} - \boldsymbol{z}'\|_2$ is the Euclidean distance between $\boldsymbol{z}$ and $\boldsymbol{z}'$, and ℓ is a scale parameter. The parameter m controls the scale-dependent smoothness properties of $c_{m,\ell}$, with larger values of m providing more selective damping at small scales. The parameter ℓ controls the spatial extent of the smoothing.

Matérn functions with $m > 2$ are differentiable at the origin ($r = 0$). For these functions, it is customary to define the length-scale D of $c_{m,\ell}$ in terms of the local curvature of the correlation function near the origin (the Daley length-scale). It is a quantity of practical interest since it can be estimated locally from derivatives of an ensemble of simulated errors (Belo Pereira and Berre 2006). In terms of ℓ and m, the Daley length-scale of Eq. (6) is given by (Weaver and Mirouze 2013)

$$D = \sqrt{2m-4}\,\ell. \tag{7}$$

The Daley length-scale is the natural scale parameter of the Gaussian function defined by $c_g(r) = \exp(-r^2/2D^2)$. The Gaussian function can be derived as a limiting case of Eq. (6) as $m \to +\infty$ with ℓ simultaneously decreased to keep D fixed (Weaver and Mirouze 2013). The correlation functions with small values of m have fatter tails than those with larger values of m (for the same value of D), as illustrated in Figure 1.

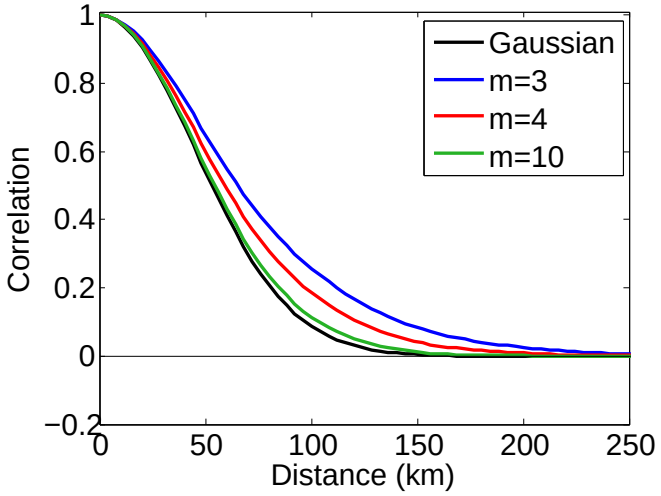


Figure 1. Cross section of a two-dimensional correlation function generated with Eq. (6) for different values of the parameter m and a fixed value of $D = 45$ km.

When $m = 2$, properties (2) still hold, but the correlation functions are no longer differentiable at the origin. For these functions, we can define the correlation length-scale as the scale parameter ℓ itself or some other characteristic measure. Correlation functions with $m = 2$ are displayed in Figure 2 for different values of ℓ. These functions have fat tails, and sharper correlations near the origin than those of the differentiable Matérn functions. We will come back to this point in Section 5 when considering the application to SEVIRI observations.

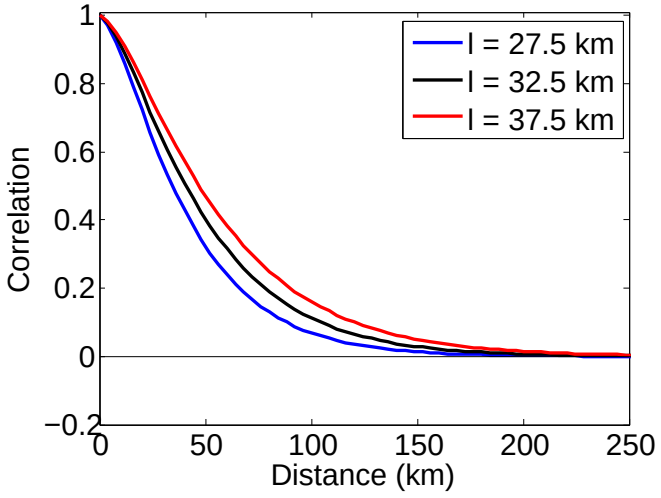


Figure 2. Cross section of a two-dimensional correlation function generated with Eq. (6) for different values of the parameter ℓ (see legend) and a fixed value of $m = 2$.

2.4. The inverse correlation operator

Let $\mathcal{C} : f \mapsto \mathcal{C}[f]$ be the correlation operator that has $c(\boldsymbol{z}, \boldsymbol{z}') = c_{m,\ell}(r)$ given by Eq. (6) as its kernel. Furthermore, suppose that Ω extends to include the whole of $\mathbb{R}^2$ and that $g = \mathcal{C}[f]$ and its derivatives vanish as $r \to \infty$. Since $c_{m,\ell}$ is homogeneous, $\mathcal{C}$ is a convolution operator,

$$g(\boldsymbol{z}) = \int_{\mathbb{R}^2} c_{m,\ell}(\boldsymbol{z} - \boldsymbol{z}') f(\boldsymbol{z}') \mathrm{d}\boldsymbol{z}', \tag{8}$$

and we can use the Fourier Transform (FT) to derive an expression for $\mathcal{C}^{-1}$ (*e.g.*, see Jones (1982)).

Let $\hat{f} : \hat{\boldsymbol{z}} \mapsto \hat{f}(\hat{\boldsymbol{z}})$, $\hat{g} : \hat{\boldsymbol{z}} \mapsto \hat{g}(\hat{\boldsymbol{z}})$ and $\hat{c}_{m,\ell} : \hat{\boldsymbol{z}} \mapsto \hat{c}_{m,\ell}(\hat{\boldsymbol{z}})$ denote the FT of f, g and $c_{m,\ell}$, respectively, where $\hat{\boldsymbol{z}}$ is the vector of spectral wave-numbers. Taking the FT of Eq. (8) yields

$$\hat{g}(\hat{\boldsymbol{z}}) = \hat{c}_{m,\ell}(\hat{\boldsymbol{z}})\, \hat{f}(\hat{\boldsymbol{z}}) \tag{9}$$

where (Whittle 1963)

$$\hat{c}_{m,\ell}(\hat{\boldsymbol{z}}) = \frac{\gamma^2}{\left(1 + \ell^2\, \hat{\boldsymbol{z}}^2\right)^m} \tag{10}$$

and

$$\gamma^2 = 4\pi(m-1)\, \ell^2. \tag{11}$$

A necessary and sufficient condition for a homogeneous and isotropic function to yield a positive definite operator, in the sense of the second property in (2), is that its FT is non-negative (see Theorem 2.10 in Gaspari and Cohn (1999)). This is clearly satisfied by Eq. (10).

Let $\mathcal{C}^{-1} : g \mapsto \mathcal{C}^{-1}[g]$ be the inverse correlation operator and let $f = \mathcal{C}^{-1}[g]$. From Eq. (9), we have

$$\hat{f}(\hat{\boldsymbol{z}}) = \frac{1}{\hat{c}_{m,\ell}(\hat{\boldsymbol{z}})}\, \hat{g}(\hat{\boldsymbol{z}}). \tag{12}$$

Taking the inverse FT of Eq. (12) leads to the elliptic equation (Whittle 1963)

$$\frac{1}{\gamma^2}\left(1 - \ell^2 \nabla^2\right)^m g(\boldsymbol{z}) = f(\boldsymbol{z}), \tag{13}$$

where I is the identity operator and $\nabla^2 \equiv \partial^2/\partial z_1^2 + \partial^2/\partial z_2^2$ is the 2D Laplacian operator. We can then identify $\mathcal{C}^{-1}$ as the elliptic operator in Eq. (13):

$$\mathcal{C}^{-1} = \frac{1}{\gamma^2}\left(I - \ell^2 \nabla^2\right)^m. \tag{14}$$

The constant γ^2 ensures that the correlation functions are properly normalized to have unit amplitude. Notice that γ^2 has physical units of length squared and can be interpreted as the natural "variance" of the Matérn covariance function associated with the (unnormalized) elliptic equation.

On a finite domain Ω, Eq. (13) must be supplied with appropriate boundary conditions. In this study, we use Neumann conditions on the boundaries of Ω. Because of the boundary conditions, the correlation functions near the boundaries are not of precise Matérn form (Mirouze and Weaver 2010). This has implications on the normalization factor, which is no longer adequately described by a constant (Eq. (11)) near the boundaries. This point will be discussed further in Section 5.

2.5. Computational aspects

Consider a triangular mesh represented by a set of nodes $(\boldsymbol{z}_i)_{i\in[1,p]}$. In this study, $(\boldsymbol{z}_i)_{i\in[1,p]}$ are taken to be the observation locations at which spatial correlations need to be defined. Applying the inverse correlation operator $\mathcal{C}^{-1}$ numerically on the mesh requires m successive applications of a discretized representation of the operator $I - \ell^2\nabla^2$ (Eq. (13)). With appropriately chosen basis functions, this is a straightforward and computationally inexpensive operation since it involves multiplication by sparse matrices after discretization. In contrast, to apply the correlation operator $\mathcal{C}$ in the integral form (1) using the expression (6), one has to compute $c_{m,\ell}(\|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2)$ for every pair $(\boldsymbol{z}_i, \boldsymbol{z}_j)$, which becomes unaffordable when the number of nodes (p) is large. For this reason, it is preferable to apply $\mathcal{C}$ by seeking a numerical solution of the elliptic equation (13) rather than attempting to integrate Eq. (1) numerically. Solving the elliptic equation requires solving m symmetric positive definite (SPD) linear systems in sequence, for which efficient methods are available (*e.g.*, see Saad (2003) for a general review). The

numerical aspects of the solution algorithm will be discussed in Section 3.

2.6. *Interpretation as an implicit diffusion operator*

Equation (13) can be interpreted as a semi-discretized representation of a standard diffusion equation using a backward Euler temporal scheme. In particular, consider the 2D diffusion equation

$$\frac{\partial g}{\partial s} - \kappa \nabla^2 g = 0, \tag{15}$$

subject to the initial condition $g(\boldsymbol{z})\big|_{s=0} = \gamma^2 f(\boldsymbol{z})$, and to homogeneous Neumann boundary conditions $\nabla g\big|_{\partial\Omega} \cdot \hat{\boldsymbol{n}} = 0$, where $\hat{\boldsymbol{n}}$ is the unit outward normal vector to the boundary $\partial\Omega$, ∇ is the 2D gradient operator, and $\cdot$ denotes the dot product. Here, s represents a non-dimensional pseudo-time coordinate, and κ is a constant pseudo-diffusion coefficient. Discretizing Eq. (15) using a backward Euler scheme with a pseudo-time step of unit size ($\Delta s = 1$) leads to the semi-discrete elliptic equation

$$\big(I - \kappa \nabla^2\big) g_{n+1}(\boldsymbol{z}) = g_n(\boldsymbol{z}) \tag{16}$$

where n is the pseudo-time discretization index, and $g_0(\boldsymbol{z}) = \gamma^2 f(\boldsymbol{z})$ is the "initial" condition. Considering the diffusion problem on the pseudo-time interval $n = [0, m-1]$ allows us to write Eq. (16) in the form of Eq. (13) with $\kappa = \ell^2$ and $g_m(\boldsymbol{z}) = g(\boldsymbol{z})$. We can thus interpret the self-adjoint operator

$$\mathcal{L}^{-1} = \big(I - \kappa \nabla^2\big)^m \tag{17}$$

as an inverse diffusion operator acting backwards in pseudo-time over m-steps.

2.7. *More general functional shapes*

Numerically "time"-stepping an implicitly formulated diffusion equation is an efficient way to apply a correlation operator with Matérn kernel of the specific form (6). More general correlation functions than those from the Matérn family can be modelled by constructing $\mathcal{C}^{-1}$ as a linear combination of powers of the Laplacian operator (Weaver and Courtier 2001; Yaremchuk and Smith 2011; Weaver and Mirouze 2013):

$$\mathcal{C}^{-1} = \frac{1}{\widetilde{\gamma}^2}\big(I - a_1 \ell^2 \nabla^2 + a_2 \ell^4 \nabla^4 + \ldots + (-1)^p a_q \ell^{2q} \nabla^{2q}\big) \tag{18}$$

where p is a positive integer, $(a_k)_{k\in[1,q]}$ are constant coefficients, and $\widetilde{\gamma}$ is a normalization constant. The operator (14) is a special case of Eq. (18) with $\widetilde{\gamma} = \gamma$, $q = m$, and $(a_k)_{k\in[1,m]}$ defined by the binomial coefficient:

$$a_k \equiv b_k = \frac{m!}{k!(m-k)!}. \tag{19}$$

Following the procedure outlined in Section 2.4, we can easily derive, from Eq. (18), the FT of the kernel associated with $\mathcal{C}$:

$$\hat{c}(\hat{\boldsymbol{z}}) = \frac{\widetilde{\gamma}^2}{1 + a_1 \ell^2 \hat{\boldsymbol{z}}^2 + a_2 \ell^4 \hat{\boldsymbol{z}}^4 + \ldots + a_q \ell^{2q} \hat{\boldsymbol{z}}^{2q}}. \tag{20}$$

Positiveness of $\hat{c}(\hat{\boldsymbol{z}})$ ensures that the kernel (the inverse FT of $\hat{c}$) is a valid correlation function. This is clearly guaranteed when the coefficients $(a_k)_{k\in[1,q]}$ are all positive. Negative coefficients can be used to generate functions with damped oscillatory behaviour about the zero-correlation axis, but special care is required to ensure that the resulting formulation leads to positive $\hat{c}(\hat{\boldsymbol{z}})$ (Weaver and Mirouze 2013; Barth *et al.* 2014). Negative correlations have been observed, for example, in the simulation of roll errors with wide-swath satellite altimeter measurements (Ruggiero *et al.* 2016).

2.8. *Anisotropy and inhomogeneity*

A further generalization is to replace $\ell^{2k}\nabla^{2k}$ in Eq. (18) with $\big(\nabla \cdot \boldsymbol{\kappa} \nabla\big)^k$, where $\nabla\cdot$ is the 2D divergence operator, and $\boldsymbol{\kappa}$ is a constant (anisotropic) diffusion tensor; *i.e.*, an SPD 2×2 matrix that allows the principal axes of the correlation functions to be stretched and rotated relative to the axes of the computational coordinates. This flexibility is desirable for representing spatial observation error correlations from polar orbiting satellites, whose principal axes may be preferentially aligned with the along- and across-track directions of the satellite path (Ruggiero *et al.* 2016).

For the Matérn family, the correlation functions associated with a constant diffusion tensor are still given by Eq. (6), but with the normalized distance measure r/ℓ replaced with

$\sqrt{(\boldsymbol{z}-\boldsymbol{z}')^{\mathrm{T}}\boldsymbol{\kappa}^{-1}(\boldsymbol{z}-\boldsymbol{z}')}$. Furthermore, the parameter ℓ^2 in the normalization constant (11) for $\mathcal{C}^{-1}$ must be replaced with $\sqrt{\det(\boldsymbol{\kappa})}$ where $\det$ is the determinant.

Spatially constant correlation functions can be overly restrictive. This is particularly true when representing correlations of background error, which generally exhibit significant spatial variations due to the heterogeneous nature of atmospheric/ocean dynamics and of the observational network. Spatial variations can also be present in observation error correlations. For example, Waller *et al.* (2016c) showed that Doppler radar radial winds have error correlations that depend on both the height of the observation and on the distance of the observation away from the radar.

It is straightforward within the diffusion framework to account for inhomogeneous error correlations by making the diffusion tensor $\boldsymbol{\kappa}(\boldsymbol{z})$ spatially dependent. With this extension, the exact analytical form of the underlying correlation function is generally not known. However, when the spatial variation of $\boldsymbol{\kappa}(\boldsymbol{z})$ is sufficiently slow, the kernel of the integral solution of Eq. (13), with $\nabla\cdot\boldsymbol{\kappa}(\boldsymbol{z})\nabla$ used instead of $\ell^2\nabla^2$, can be expected to be approximately given by Eq. (6) in the vicinity of a given point (Mirouze and Weaver 2010; Weaver and Mirouze 2013; Yaremchuk and Nechaev 2013).

The exact normalization factors are no longer constant when the diffusion tensor is spatially dependent. For slowly varying $\boldsymbol{\kappa}(\boldsymbol{z})$, they can be approximated by (cf. Eq. (11))

$$\gamma(\boldsymbol{z}) \approx \sqrt{4\pi(m-1)\sqrt{\det(\boldsymbol{\kappa}(\boldsymbol{z}))}} \tag{21}$$

or a suitably smoothed version of Eq. (21) (Purser *et al.* 2003; Yaremchuk and Carrier 2012). Furthermore, to maintain symmetry of the correlation functions, they must be introduced symmetrically in the elliptic equation:

$$\frac{1}{\gamma(\boldsymbol{z})}\left(I-\nabla\cdot\boldsymbol{\kappa}(\boldsymbol{z})\nabla\right)^m\frac{1}{\gamma(\boldsymbol{z})}\,g(\boldsymbol{z}) = f(\boldsymbol{z}). \tag{22}$$

2.9. Estimating parameters of the correlation model

The Matérn correlation model requires specifying the smoothness parameter m and scale parameter ℓ. In the generalized correlation models described in Sections 2.7 and 2.8, the parameters to set are the coefficients $(a_k)_{k\in[1,q]}$ of the Laplacian operators up to order q (instead of the single parameter m), and the spatially dependent diffusion tensors $\boldsymbol{\kappa}(\boldsymbol{z})$ (instead of the single parameter ℓ).

The correlation model parameters need to be estimated from knowledge of the actual observation error statistics. For this purpose, the *a posteriori* diagnostic from Desroziers *et al.* (2005) is frequently used. This diagnostic provides an estimate of the total observation error covariances (*i.e.*, the combined components of measurement and representativeness error) from the cross-covariances between the analysis and the background residuals in observation space. The effectiveness of the technique for estimating observation error correlations is discussed by Waller *et al.* (2016b). The statistics are usually averaged in space and in time in order to increase the sample size and thus improve the robustness of the estimated covariances (Bormann *et al.* 2010; Bormann and Bauer 2010; Waller *et al.* 2016a; Michel 2018). Together with the fact that the method itself is based on some questionable assumptions, this suggests that this diagnostic should be used to provide only *coarse* estimates of the covariances. In this respect, the basic two-parameter Matérn correlation function (6) may be adequate for representing the statistics.

Some observation types may come equipped with an instrument error simulator. In particular, this is the case for the SWOT altimeter mission (Ubelmann *et al.* 2016). Assuming that the sources of measurement error are accurately modelled by the simulator, it can be used to provide more detailed sample estimates of the measurement component of the observation error covariances. Ruggiero *et al.* (2016) used the SWOT simulator to estimate parameters of the Brankart *et al.* (2009) covariance model. Complementary techniques for estimating the representativeness component of the observation error covariances are discussed in the recent review article by Janjić *et al.* (2018).

When reliable, comprehensive estimates of the observation error covariances are available, the multi-parameter formulations of the diffusion-based correlation model are appropriate. The approach considered in Sections 3 and 5 will focus on the two-parameter model, but can be adapted to the more general cases if necessary.

3. Finite element discretization

3.1. Motivation

We now investigate strategies to discretize the diffusion equation (13) in space. Given a set of observations at locations $(\boldsymbol{z}_i)_{i\in[1,p]}$, we wish to compute the solution of the diffusion equation at these same locations. Hence our choice for discretizing Eq. (13) is to build a computational mesh with nodes at observation locations, so that the solution can be computed directly at the nodes of this mesh. The FEM is one popular class of discretization strategies well known for handling such unstructured data distributions, and is the focus of this article.

Efficient solution techniques for partial differential equations (PDEs) convert a continuous operator problem to a discrete problem by a suitable projection onto a finite-dimensional subspace. Let Eq. (13) be written as

$$\left.\begin{aligned} g_0(\boldsymbol{z}) \;&=\; \gamma(\boldsymbol{z})\, f(\boldsymbol{z}) \\ \big(I-\ell^2\nabla^2\big)\, g_{n+1}(\boldsymbol{z}) \;&=\; g_n(\boldsymbol{z}) \quad ; \quad n=[0,m-1] \\ g(\boldsymbol{z}) \;&=\; \gamma(\boldsymbol{z})\, g_m(\boldsymbol{z}) \end{aligned}\right\}, \tag{23}$$

where the normalization factors have been introduced symmetrically as in Eq. (22). This is necessary for numerical applications, even with constant ℓ, since the exact normalization factors depend on the local accuracy of the numerical solution (which depends on the local quality of the mesh) and the boundary conditions. Here, we use Neumann boundary conditions on the spatial domain of interest Ω.

The FEM is a standard technique for solving PDEs numerically (Ciarlet 2002; Brenner and Scott 2013). The basic procedure involves defining a variational formulation of the infinite-dimensional continuous problem. This variational formulation is then solved by approximating the solution in a carefully chosen finite-dimensional subspace. Applying this procedure to Eq. (23) leads to a matrix formulation of the diffusion equation in the space of the observations. We will outline the procedure below and show how the resulting expressions can be used in formulations of the observation error covariance matrix and its inverse.

3.2. Galerkin approximation

Let $(\varphi_j)_{j\in\mathcal{I}}$ be an independent set of test functions used to discretize the diffusion equation. The $(\varphi_j)_{j\in\mathcal{I}}$ are called "degrees of freedom" and $\mathcal{I}$ is a set of indices of finite cardinality. Multiplying both sides of the elliptic equation in Eq. (23) by φ_i and integrating over Ω leads to the weak formulation of the PDE:

$$\int_\Omega \big(I-\ell^2\nabla^2\big) g_{n+1}(\boldsymbol{z})\, \varphi_j(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \;=\; \int_\Omega g_n(\boldsymbol{z})\, \varphi_j(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \tag{24}$$

for $j\in\mathcal{I}$ and $n=[0,m-1]$.

Now we introduce the Galerkin approximation in which $g_n(\boldsymbol{z})$ and $g_{n+1}(\boldsymbol{z})$ are represented by finite expansions in terms of $(\varphi_i)_{i\in\mathcal{I}}$:

$$\left.\begin{aligned} g_n(\boldsymbol{z}) \;&=\; \sum_{i\in\mathcal{I}} \alpha_n^{(i)}\, \varphi_i(\boldsymbol{z}) \\ \text{and}\quad g_{n+1}(\boldsymbol{z}) \;&=\; \sum_{i\in\mathcal{I}} \alpha_{n+1}^{(i)}\, \varphi_i(\boldsymbol{z}) \end{aligned}\right\}. \tag{25}$$

Substituting these expressions into Eq. (24) and using Green's first identity together with Neumann boundary conditions yields the matrix equation

$$(\boldsymbol{M}+\boldsymbol{K})\boldsymbol{\alpha}_{n+1} \;=\; \boldsymbol{M}\boldsymbol{\alpha}_n, \tag{26}$$

where $\boldsymbol{\alpha}_n$ is a vector containing the coordinates $(\alpha_n^{(i)})_{i\in\mathcal{I}}$, $\boldsymbol{M}$ is the Gram mass matrix, and $\boldsymbol{K}$ is the stiffness matrix, with elements given by

$$\boldsymbol{M}_{ij} \;=\; \int_\Omega \varphi_i(\boldsymbol{z})\varphi_j(\boldsymbol{z})\mathrm{d}\boldsymbol{z}, \tag{27}$$

$$\text{and}\quad \boldsymbol{K}_{ij} \;=\; \ell^2 \int_\Omega \nabla\varphi_i(\boldsymbol{z})\cdot\nabla\varphi_j(\boldsymbol{z})\mathrm{d}\boldsymbol{z}. \tag{28}$$

The stiffness matrix $\boldsymbol{K}$ is symmetric and positive semi-definite. The Gram mass matrix $\boldsymbol{M}$ is symmetric and positive definite since $(\varphi_i)_{i\in\mathcal{I}}$ form an independent set of functions. It defines the weighting matrix for the $L^2(\Omega)$-inner product measured with respect to vectors $\boldsymbol{\alpha}_k$ and $\boldsymbol{\alpha}_l$ of basis coefficients; *i.e.*,

$$\int_\Omega g_k(\boldsymbol{z})\, g_l(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \;=\; \boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\alpha}_l,$$

which using standard inner-product notation reads

$$\langle\, g_k,\, g_l \,\rangle \;=\; \langle\, \boldsymbol{\alpha}_k, \boldsymbol{\alpha}_l \,\rangle_{\boldsymbol{M}}. \tag{29}$$

Applying Eq. (26) on $n = [0, m-1]$ leads to a sequence of linear systems

$$\left.\begin{aligned} \big(\boldsymbol{M}+\boldsymbol{K}\big)\boldsymbol{\alpha}_1 &= \boldsymbol{M}\,\boldsymbol{\alpha}_0 \\ \big(\boldsymbol{M}+\boldsymbol{K}\big)\boldsymbol{\alpha}_2 &= \boldsymbol{M}\,\boldsymbol{\alpha}_1 \\ &\vdots \\ \big(\boldsymbol{M}+\boldsymbol{K}\big)\boldsymbol{\alpha}_m &= \boldsymbol{M}\,\boldsymbol{\alpha}_{m-1} \end{aligned}\right\}. \tag{30}$$

After multiplying both sides of the equations in (30) by $\boldsymbol{M}^{-1}$, we can combine the resulting equations into a single equation

$$\big[\boldsymbol{M}^{-1}\big(\boldsymbol{M}+\boldsymbol{K}\big)\big]^m\boldsymbol{\alpha}_m \;=\; \boldsymbol{\alpha}_0,$$

which can be identified as the discretized weak formulation of $\mathcal{L}^{-1}$ in Eq. (17), defined for the vector $\boldsymbol{\alpha}$ of basis coefficients. We denote this matrix operator by

$$\boldsymbol{L}_{\boldsymbol{M}}^{-1} \;=\; \big[\boldsymbol{M}^{-1}\big(\boldsymbol{M}+\boldsymbol{K}\big)\big]^m \tag{31}$$

where the notation $\boldsymbol{L}_{\boldsymbol{M}}^{-1}$ indicates that this matrix is self-adjoint with respect to the $\boldsymbol{M}$-inner product (Eq. (29)); *i.e.*,

$$\boldsymbol{L}_{\boldsymbol{M}}^{-1} \;=\; \boldsymbol{M}^{-1}\big(\boldsymbol{L}_{\boldsymbol{M}}^{-1}\big)^{\mathrm{T}}\boldsymbol{M}. \tag{32}$$

The self-adjointness of $\boldsymbol{L}_{\boldsymbol{M}}^{-1}$ with respect to the $\boldsymbol{M}$-inner product corresponds to the self-adjointness of $\mathcal{L}^{-1}$ with respect to the $L^2(\Omega)$-inner product. The matrix $\boldsymbol{M}\boldsymbol{L}_{\boldsymbol{M}}^{-1}$ is symmetric in the usual sense.

3.3. Discrete diffusion operator

In this section, we drop the pseudo-time index n for clarity of notation. Let $\boldsymbol{g}$ be a vector of dimension card($\mathcal{I}$), which contains the values of $\boldsymbol{g}$ at observation locations $(\boldsymbol{z}_i)$, $i = [1, p]$. Equation (25) describes the relation between the values $\boldsymbol{g}(\boldsymbol{z}_i)$ at observation locations and the coordinates $(\alpha^{(i)})_{i\in\mathcal{I}}$ of the basis functions $(\varphi_i)_{i\in\mathcal{I}}$. It can be written in matrix form as

$$\boldsymbol{g} \;=\; \boldsymbol{G}\,\boldsymbol{\alpha} \tag{33}$$

where the elements of $\boldsymbol{G}$ are defined through the relation

$$\boldsymbol{G}_{ij} \;=\; \varphi_j(\boldsymbol{z}_i) \tag{34}$$

with $j \in \mathcal{I}$ and $i = [1, p]$. In the following, we will only consider the standard $\mathbb{P}_1$-FEM approximation for which $\boldsymbol{G}$ is the identity matrix (card($\mathcal{I}$) $= p$). Therefore, we will later omit $\boldsymbol{G}$. Nevertheless, we note that other approximations lead to more complex expressions for $\boldsymbol{G}$ (*e.g.*, when the $(\varphi_i)_{i\in\mathcal{I}}$ are harmonic functions and $\boldsymbol{G}$ is the corresponding spectral transform).

From now on, let us assume that a triangular mesh supporting the (observation) nodes is available, and that each node i in this triangulation corresponds to the point $\boldsymbol{z}_i$. Here, we choose the basis functions $(\varphi_i)_{i\in\mathcal{I}}$ to be continuous and linear inside each triangle, with the property (Ern and Guermond 2010, Chapter 8)

$$\varphi_i(\boldsymbol{z}_j) = \delta_{ij}, \tag{35}$$

where δ_{ij} is the Kronecker delta. An immediate consequence of Eq. (35) is that $\boldsymbol{G}$ is the $p \times p$ identity matrix ($\boldsymbol{I}$). As already mentioned, this simple choice of basis functions corresponds to the $\mathbb{P}_1$-FEM and guarantees that every function has a local compact support (see Figure 3). In Eqs (27) and (28), the integrals can be computed exactly over the triangular nodes using standard integration techniques, since the integrands are polynomials of at most second order (Canuto *et al.* 1987). (Higher order finite elements may require using quadrature formula to evaluate the integrals.) Hence, the non-diagonal entries $\boldsymbol{M}_{ij}$ and $\boldsymbol{K}_{ij}$ equal zero as soon as $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ do not belong to the same triangle. Therefore, the choice of the $\mathbb{P}_1$ element is responsible for the sparsity of the matrices $\boldsymbol{M}$ and $\boldsymbol{K}$. The profile of $\boldsymbol{M}$ for the unstructured distribution of satellite observations considered in Section 5 is depicted in Figure 4.

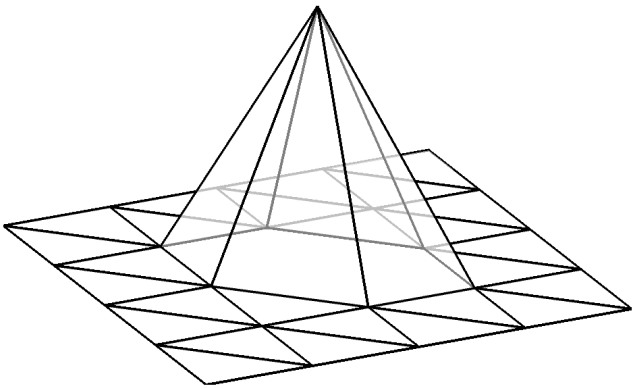


Figure 3. Representation of one $\mathbb{P}_1$-FEM basis function and its compact support. The function has a value equal to 1 at node $\boldsymbol{z}_i$ (the point (3,3) in the figure) and a value of 0 at other nodes.

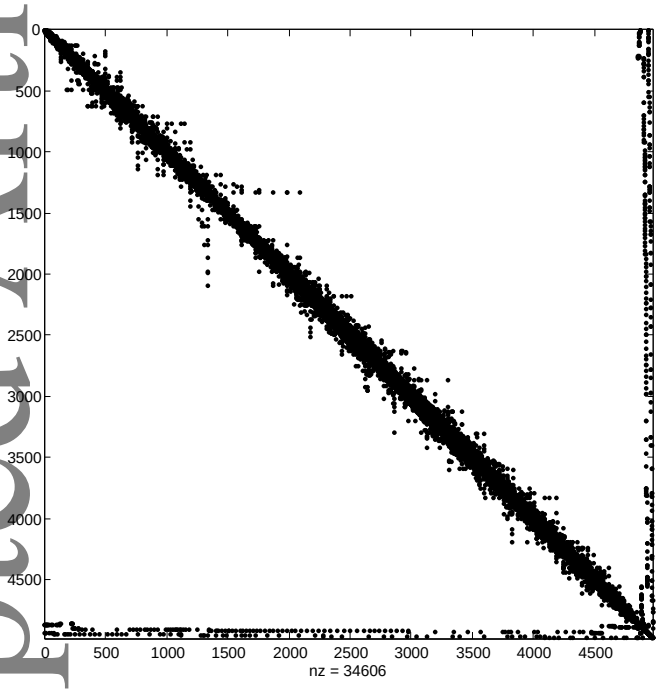


Figure 4. Profile of the mass matrix $\boldsymbol{M}$ for the unstructured satellite (SEVIRI) observations considered in Section 5. The profile of the stiffness matrix $\boldsymbol{K}$ is similar. The labelling on the horizontal and vertical axes corresponds to the column and row indices. The total number of non-zero ('nz') entries of the matrix is indicated below the horizontal axis. The non-zero entries that appear far from the diagonal correspond to boundary elements in the mesh (see section 3.6 for a discussion).

We can write the discrete form of the $L^2(\Omega)$-inner product on the left-hand side of property (29) as

$$\langle \boldsymbol{g}, \boldsymbol{g} \rangle_{\boldsymbol{W}} \;=\; \boldsymbol{g}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{g} \tag{36}$$

where $\boldsymbol{W}$ is a symmetric and positive definite matrix of grid-dependent weighting factors. Since $\boldsymbol{G} = \boldsymbol{I}$ from our choice of basis functions, Eqs (29), (33) and (36) imply the simple relation

$$\boldsymbol{W} = \boldsymbol{M}. \tag{37}$$

3.4. *Formulation of $\boldsymbol{R}$ and $\boldsymbol{R}^{-1}$ in variational assimilation*

Consider the discrete form of the cost function in variational data assimilation, focusing on the observation term J_{o}. Let $\boldsymbol{d}$ be a p-dimensional vector where p is the number of assimilated observations. The ith component of $\boldsymbol{d}$ corresponds to the difference between the ith observation and its model counterpart. The observation term is formulated as the $\boldsymbol{R}^{-1}$-norm of the difference vector:

$$J_{\mathrm{o}} \;=\; \frac{1}{2} \boldsymbol{d}^{\mathrm{T}} \, \boldsymbol{R}^{-1} \, \boldsymbol{d}$$

where $\boldsymbol{R}$ is the (symmetric positive definite) observation error covariance matrix.

We can express $\boldsymbol{R}^{-1}$ in the standard factored form

$$\boldsymbol{R}^{-1} \;=\; \boldsymbol{\Sigma}^{-1} \, \boldsymbol{C}^{-1} \, \boldsymbol{\Sigma}^{-1}, \tag{38}$$

where $\boldsymbol{C}^{-1}$ is the inverse of the observation error correlation matrix, and $\boldsymbol{\Sigma}$ is a diagonal matrix containing the observation error standard deviations. Assuming that the errors are correlated and modelled by a discretized diffusion operator, then we can express $\boldsymbol{C}^{-1}$ as

$$\boldsymbol{C}^{-1} \;=\; \boldsymbol{\Gamma}^{-1} \boldsymbol{W} \, \boldsymbol{L}_{\boldsymbol{W}}^{-1} \, \boldsymbol{\Gamma}^{-1}, \tag{39}$$

where $\boldsymbol{\Gamma}^2$ is a diagonal matrix of normalization factors, and $\boldsymbol{L}_{\boldsymbol{W}}^{-1}$ is the inverse diffusion operator, which is self-adjoint with respect to the $\boldsymbol{W}$-inner product (Lorenc 1997; Weaver and Courtier 2001). The appearance of $\boldsymbol{W}$ in Eq. (39) comes from the discrete representation of spatial integration implied by the $L^2(\Omega)$-inner product. Here, we have implicitly assumed that $\boldsymbol{d}$ is the discrete representation of a square-integrable function.

When the diffusion operator is discretized using the FEM then, from Eq. (37), Eq. (39) becomes

$$\boldsymbol{C}^{-1} \;=\; \boldsymbol{\Gamma}^{-1} \boldsymbol{M} \, \boldsymbol{L}_{\boldsymbol{M}}^{-1} \, \boldsymbol{\Gamma}^{-1}. \tag{40}$$

Expressions for the covariance and correlation matrices follow directly from Eqs (31), (38) and (40):

$$\boldsymbol{R} \;=\; \boldsymbol{\Sigma} \, \boldsymbol{C} \, \boldsymbol{\Sigma} \tag{41}$$

where

$$\boldsymbol{C} \;=\; \boldsymbol{\Gamma}\,\boldsymbol{L}_{\boldsymbol{M}}\,\boldsymbol{M}^{-1}\,\boldsymbol{\Gamma} \tag{42}$$

and

$$\boldsymbol{L}_{\boldsymbol{M}} \;=\; \big[\big(\boldsymbol{M}+\boldsymbol{K}\big)^{-1}\boldsymbol{M}\big]^{m}. \tag{43}$$

3.4.1. Square-root formulation

It is convenient to construct $\boldsymbol{R}$ and $\boldsymbol{R}^{-1}$ as factored products

$$\left.\begin{aligned} \boldsymbol{R} \;&=\; \boldsymbol{V}\,\boldsymbol{V}^{\mathrm{T}} \\ \text{and}\quad \boldsymbol{R}^{-1} \;&=\; \big(\boldsymbol{V}^{-1}\big)^{\mathrm{T}}\,\boldsymbol{V}^{-1} \end{aligned}\right\}.$$

This factorization ensures that $\boldsymbol{R}$ and $\boldsymbol{R}^{-1}$ remain symmetric and positive definite in numerical applications. Furthermore, it gives us access to a "square-root" operator $\boldsymbol{V}$, which is a valuable tool for randomization applications; *i.e.*, for generating a spatially correlated random vector with covariance matrix equal to $\boldsymbol{R}$ given white noise as an input vector.

By restricting the number of diffusion steps m to be even, it follows from Eqs (41), (42) and (43), and Eqs (31), (38) and (39) that

$$\left.\begin{aligned} \boldsymbol{V} \;&=\; \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}\,\boldsymbol{L}_{\boldsymbol{M}}^{1/2}\big(\boldsymbol{M}^{-1/2}\big)^{\mathrm{T}} \\ \text{and}\quad \boldsymbol{V}^{-1} \;&=\; \big(\boldsymbol{M}^{1/2}\big)^{\mathrm{T}}\,\boldsymbol{L}_{\boldsymbol{M}}^{-1/2}\,\boldsymbol{\Gamma}^{-1}\,\boldsymbol{\Sigma}^{-1} \end{aligned}\right\},$$

where

$$\left.\begin{aligned} \boldsymbol{L}_{\boldsymbol{M}}^{1/2} \;&=\; \big[\big(\boldsymbol{M}+\boldsymbol{K}\big)^{-1}\boldsymbol{M}\big]^{m/2} \\ \text{and}\quad \boldsymbol{L}_{\boldsymbol{M}}^{-1/2} \;&=\; \big[\boldsymbol{M}^{-1}\big(\boldsymbol{M}+\boldsymbol{K}\big)\big]^{m/2} \end{aligned}\right\}.$$

In deriving these expressions, we have used the relation

$$\boldsymbol{L}_{\boldsymbol{M}}\boldsymbol{M}^{-1} \;=\; \boldsymbol{L}_{\boldsymbol{M}}^{1/2}\big(\boldsymbol{M}^{-1/2}\big)^{\mathrm{T}}\boldsymbol{M}^{-1/2}\big(\boldsymbol{L}_{\boldsymbol{M}}^{1/2}\big)^{\mathrm{T}},$$

which follows from the identity (32) and the standard factorization $\boldsymbol{M} \;=\; \boldsymbol{M}^{1/2}\big(\boldsymbol{M}^{1/2}\big)^{\mathrm{T}}$ for an SPD matrix.

3.5. Mass lumping

The expressions for $\boldsymbol{C}$ (Eqs (42) and (31)) and $\boldsymbol{C}^{-1}$ (Eqs (40) and (43)) involve the inverse of the mass matrix, $\boldsymbol{M}^{-1}$. The choice of our basis functions renders $\boldsymbol{M}$ sparse and hence amenable to the use of a sparse direct SPD solver based on Cholesky decomposition. Nevertheless, it can be convenient to simplify computations further by using a technique known as *mass lumping*, which involves approximating the *consistent* mass matrix $\boldsymbol{M}$ in Eq. (31) with a diagonal matrix $\widetilde{\boldsymbol{M}}$ called the lumped mass matrix. For the case of the $\mathbb{P}_1$-FEM considered here, it is simply obtained by summing the coefficients of $\boldsymbol{M}$ on each row or column. This process is equivalent to computing the integral (27) using a low-order quadrature formula or replacing the basis functions $(\varphi_i)_{i\in\mathcal{I}}$ with piecewise constant functions $(\widetilde{\varphi}_i)_{i\in\mathcal{I}}$ on each element (Canuto *et al.* 1987). Hence the formula for the coefficients of $\widetilde{\boldsymbol{M}}$ becomes

$$\widetilde{\boldsymbol{M}}_{ii} \;=\; \sum_j \boldsymbol{M}_{ij} \;=\; \sum_j \int_\Omega \varphi_i(\boldsymbol{z})\varphi_j(\boldsymbol{z})\mathrm{d}\boldsymbol{z}.$$

In Section 5, we evaluate this approximation in terms of its effect on the representation of Matérn correlations.

3.6. Boundary nodes and the inverse correlation operator

The nodes added at the artificial boundaries of the domain are required to make the elliptic problem well posed. However, they do not correspond to actual observation locations, so must be discarded once the correlation operator has been applied. This specific feature of the correlation operator has not been made explicit in the formulation of $\boldsymbol{R}$ presented in Section 3.4, but has implications for the specification of the inverse correlation operator as discussed in this section.

Let $\boldsymbol{C}_{\mathrm{b}}$ denote the correlation operator that includes the extra boundary nodes. It is formulated as an FEM-discretized diffusion operator according to Eq. (42). A correlation operator $\boldsymbol{C}$ associated with the actual observations can be obtained from $\boldsymbol{C}_{\mathrm{b}}$ using the formulation

$$\boldsymbol{C} \;=\; \boldsymbol{S}\,\boldsymbol{C}_{\mathrm{b}}\,\boldsymbol{S}^{\mathrm{T}} \tag{44}$$

where $\boldsymbol{S}$ is a selection matrix (a rectangular matrix of 0s and 1s) that, together with $\boldsymbol{S}^{\mathrm{T}}$, picks out the submatrix of $\boldsymbol{C}_{\mathrm{b}}$ whose elements correspond to the correlations at the actual observation locations. If p_{b} denotes the number of boundary nodes, then $\boldsymbol{C}_{\mathrm{b}}$ is a full-rank $(p+p_{\mathrm{b}})\times(p+p_{\mathrm{b}})$ matrix, while $\boldsymbol{C}$ is a full-rank $p\times p$ matrix.

The inverse correlation matrix associated with Eq. (44) is

$$\boldsymbol{C}^{-1} \;=\; \big(\boldsymbol{S}\,\boldsymbol{C}_{\mathrm{b}}\,\boldsymbol{S}^{\mathrm{T}}\big)^{-1}, \tag{45}$$

which is not an explicit operator. To apply it requires solving a linear system. Rather than using Eq. (45), we can approximate the inverse as

$$\boldsymbol{C}^{-1} \;\approx\; \widetilde{\boldsymbol{C}}^{-1} \;=\; \boldsymbol{S}\,\boldsymbol{C}_{\mathrm{b}}^{-1}\,\boldsymbol{S}^{\mathrm{T}}, \tag{46}$$

which is straightforward to apply using the expressions presented in Section 3.4. However, the extent to which Eq. (46) is a good approximation of Eq. (45) is not obvious. As a trivial example, consider $\boldsymbol{C}_{\mathrm{b}}$ to be a 2×2 correlation matrix (*i.e.*, 2 nodes) with correlation coefficient (off-diagonal element) equal to ρ. If $\boldsymbol{S}$ and $\boldsymbol{S}^{\mathrm{T}}$ act to select one of the nodes then $\boldsymbol{C}$ and $\boldsymbol{C}^{-1}$ have a single element equal to one. For comparison, $\widetilde{\boldsymbol{C}}^{-1}$ has a single element equal to $1/(1-\rho)$, which shows that it is a good approximation of $\boldsymbol{C}^{-1}$ if the two points are weakly correlated ($\rho \approx 0$).

For the mesh used in our experiments, $\widetilde{\boldsymbol{C}}^{-1}$ and $\boldsymbol{C}^{-1}$ are practically equivalent: for a random vector $\boldsymbol{v}$, $\|\widetilde{\boldsymbol{C}}^{-1}\boldsymbol{C}\boldsymbol{v} - \boldsymbol{v}\|_\infty < 10^{-14}$. This is perhaps not surprising in view of the simple analysis above, since the distance between the artificial boundary nodes and the interior nodes is typically much larger than the correlation length-scale that is used in our experiments (see Section 5).

3.7. Computational aspects

The discrete inverse covariance matrix $\boldsymbol{R}^{-1}$ (taking into account the approximation (46)) is built from a combination of diagonal matrices $\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{\Gamma}^{-1}$ and $\widetilde{\boldsymbol{M}}$ (assuming a lumped mass matrix), and a product of m matrices involving the left-scaled, shifted stiffness matrix, $\widetilde{\boldsymbol{M}}^{-1}\big(\widetilde{\boldsymbol{M}} + \boldsymbol{K}\big)$. The resulting operator is well suited for a parallelization strategy based on domain decomposition in a distributed memory environment, where the observations are split between processors according to their spatial location, and Message Passing Interface communications are performed at the domain boundaries before each application of the stiffness matrix. As an operator, $\boldsymbol{R}^{-1}$ can be applied cheaply and is therefore ideal in variational data assimilation for minimization algorithms that require $\boldsymbol{R}^{-1}$, but not $\boldsymbol{R}$.

Applying $\boldsymbol{R}$ is computationally more demanding than applying $\boldsymbol{R}^{-1}$. The discrete covariance matrix $\boldsymbol{R}$ is built from a combination of diagonal matrices $\boldsymbol{\Sigma}$, $\boldsymbol{\Gamma}$ and $\widetilde{\boldsymbol{M}}^{-1}$, and a product of m matrices involving the right-scaled, *inverse* of the shifted stiffness matrix, $\big(\widetilde{\boldsymbol{M}} + \boldsymbol{K}\big)^{-1}\widetilde{\boldsymbol{M}}$. An efficient way to apply the latter is to solve, in sequence, each of the linear systems in (30) involving the sparse symmetric positive definite matrix $\widetilde{\boldsymbol{M}} + \boldsymbol{K}$.

The sparsity of $\widetilde{\boldsymbol{M}} + \boldsymbol{K}$ depends on the orthogonality of the basis functions $(\varphi_i)_{i\in\mathcal{I}}$ with respect to the $L^2(\Omega)$-inner product. We have chosen here to use compactly-supported piecewise polynomial functions, which results in a large number of zero entries in $\boldsymbol{K}$ (and $\boldsymbol{M}$). Choosing different types of functions would result in alternative covariance operators that would generally be more costly to apply.

For most applications related to 2D computational domains, the linear systems in (30) can be solved up to machine precision using a direct method based on Cholesky decomposition (Duff *et al.* 1989; Davis 2006). Iterative methods can be used to solve the linear system approximately when the size of the matrix is very large. Weaver *et al.* (2016) highlight the importance of using a *linear* iterative solver, together with the adjoint of the solver, in a square-root formulation of the correlation matrix in order to preserve numerical symmetry of the correlation matrix when using a modest convergence criterion. Linear iterative solvers based on multi-grid (Gratton *et al.* 2011) or the Chebyshev Iteration (Weaver *et al.* 2016, 2018) are particularly well suited for this problem.

For the experiments described in Section 5, a direct method has been used to solve the linear systems in (30) to an accuracy largely below the discretization error of the FEM.

4. Link between a diffusion-based covariance model and assimilating derivatives of observations

The method of Brankart *et al.* (2009) (referred to as the Brankart method hereafter) for accounting for correlated observation errors has gained popularity in recent years, particularly in oceanography for the assimilation of high-resolution altimeter data from SWOT (Ruggiero *et al.* 2016). The Brankart method involves assimilating the observations together with successive derivatives of the observations. This method can be viewed, under

certain assumptions, as a diffusion-based approach for modelling correlated error. The purpose of this section is to establish a formal mathematical link between the two methods in order to help improve our understanding of the advantages and disadvantages of each method.

4.1. *Continuous formulation*

The approach presented in Brankart *et al.* (2009) involves linearly transforming the observations into an augmented set of observations. In this subsection, we consider the approach in a continuous framework before treating the discrete problem in the next subsection. Let the observations be denoted by a continuous function $y : \boldsymbol{z} \mapsto y(\boldsymbol{z})$ where $y \in L^2(\Omega)$. We introduce the linear transform operator $\mathcal{T}[y]$ such that the resulting function contains both y and successive derivatives of y. Brankart *et al.* (2009) focus mainly on assimilating the first-order derivatives of y, while Ruggiero *et al.* (2016) consider both first- and second-order derivatives of y. While it is possible to assimilate higher order derivatives, for reasons of clarity, we choose to consider only the first- and second-order derivative information, as in Ruggiero *et al.* (2016). This will be sufficient to illustrate the link with the diffusion approach. We adopt similar notation to that of Brankart *et al.* (2009) and, as in the previous section, we focus on the domain Ω contained in $\mathbb{R}^2$.

The Brankart method involves formulating the inverse observation error correlation operator as

$$\mathcal{R}_{\mathrm{B}}^{-1} = \mathcal{T}^{\mathrm{T}} \left(\mathcal{R}^{+}\right)^{-1} \mathcal{T} \tag{47}$$

where

$$\left(\mathcal{R}^{+}\right)^{-1} = \mathrm{diag}\left(a_0 I,\, a_1 I,\, a_2 I,\, a_3 I,\, a_4 I,\, a_5 I\right),$$

and the operator $\mathcal{T}$ and its transpose are defined as

$$\mathcal{T} = \begin{pmatrix} I \\ \partial/\partial z_1 \\ \partial/\partial z_2 \\ \partial^2/\partial z_1^2 \\ \partial^2/\partial z_2^2 \\ \partial^2/\partial z_1 \partial z_2 \end{pmatrix} \tag{48}$$

and

$$\mathcal{T}^{\mathrm{T}} = \begin{pmatrix} I & -\dfrac{\partial}{\partial z_1} & -\dfrac{\partial}{\partial z_2} & \dfrac{\partial^2}{\partial z_1^2} & \dfrac{\partial^2}{\partial z_2^2} & \dfrac{\partial^2}{\partial z_1 \partial z_2}, \end{pmatrix}.$$

The elements $(a_i)_{i=[0,5]}$ can be functions of $\boldsymbol{z}$ but here we consider them to be constant. The last component of $\mathcal{T}$ involving cross-derivatives is not considered by Brankart *et al.* (2009) or Ruggiero *et al.* (2016) but is required here to compare with the 2D diffusion-based formulation since the latter involves powers of the Laplacian operator in a general coordinate system $\boldsymbol{z} = (z_1, z_2)^{\mathrm{T}}$ where z_1 and z_2 are not necessarily aligned with the principal axes of the 2D correlation functions. The operator $\left(\mathcal{R}^{+}\right)^{-1}$ is to be interpreted as the inverse error covariance operator of the augmented set of observations y, $\partial y/\partial z_1$, $\partial y/\partial z_2$, $\partial^2 y/\partial z_1^2$, $\partial^2 y/\partial z_2^2$, $\partial^2 y/\partial z_1 \partial z_2 \in L^2(\Omega)$. The operator $\mathcal{R}_{\mathrm{B}}^{-1}$ is symmetric with respect to the $L^2(\Omega)$-inner product. Note that the components of the second derivatives are symmetric, while those of the first derivatives are anti-symmetric (Tarantola 2005, pp. 130-131).

Expanding Eq. (47) allows us to write

$$\begin{aligned} \mathcal{R}_{\mathrm{B}}^{-1} \;=\; & a_0 \,-\, a_1 \frac{\partial^2}{\partial z_1^2} \,-\, a_2 \frac{\partial^2}{\partial z_2^2} \\ & +\, a_3 \frac{\partial^4}{\partial z_1^4} \,+\, a_4 \frac{\partial^4}{\partial z_2^4} \,+\, a_5 \frac{\partial^4}{\partial z_1^2 \partial z_2^2}. \end{aligned} \tag{49}$$

Now consider the inverse of an implicit diffusion-based covariance operator assuming that the variance σ^2 is constant:

$$\mathcal{R}^{-1} = \frac{1}{\sigma^2 \gamma^2} \left(I - \ell^2 \nabla^2\right)^m \tag{50}$$

(see Eq. (14) for the corresponding inverse correlation operator). By comparing Eqs (49) and (50), it is easy to see that they are equivalent when $m = 2$ and when the elements of $\left(\mathcal{R}^{+}\right)^{-1}$ are

chosen to be

$$\begin{aligned} a_0 &= \frac{1}{\sigma^2\gamma^2}, \\ a_1 &= a_2 = \frac{2\,\ell^2}{\sigma^2\gamma^2}, \\ a_3 &= a_4 = \frac{\ell^4}{\sigma^2\gamma^2}, \\ \text{and} \quad a_5 &= \frac{2\,\ell^4}{\sigma^2\gamma^2}. \end{aligned}$$

The equivalence of the two methods is easily generalized to account for an arbitrary value of m by augmenting Eq. (48) to include derivatives and cross-derivatives of y up to order m, and by extending $\left(\mathcal{R}^+\right)^{-1}$ to include additional coefficients defined appropriately in terms of the binomial coefficients (Eq. (19)).

Unlike the diffusion-based approach, the Brankart method does not distinguish the inverse of the correlation operator from the inverse of the covariance operator. The parameters σ^2, γ^2 and ℓ are defined jointly via the coefficients a_k. Procedures for estimating these coefficients as spatially-dependent quantities are described by Ruggiero *et al.* (2016) and Yaremchuk *et al.* (2018). In the diffusion-based approach, the parameters σ^2 and ℓ (or, in general, the diffusion tensor $\boldsymbol{\kappa}$), and spatially-dependent generalizations of these parameters, can be estimated separately based on knowledge of the underlying (Matérn) covariance function to which sample estimates of the covariances can be fit. The relationship between the two methods becomes more difficult to quantify as soon as the parameters are made spatially dependent.

4.2. *Discrete formulation*

Consider the expression for the inverse of the covariance matrix associated with an FEM diffusion-based formulation for $m = 2$. From Eqs. (31), (38) and (40)

$$\begin{aligned} \boldsymbol{R}^{-1} &= \boldsymbol{\Sigma}^{-1}\,\boldsymbol{\Gamma}^{-1}\big(\boldsymbol{M}+\boldsymbol{K}\big)\,\boldsymbol{M}^{-1}\big(\boldsymbol{M}+\boldsymbol{K}\big)\,\boldsymbol{\Gamma}^{-1}\,\boldsymbol{\Sigma}^{-1} \\ &= \boldsymbol{\Sigma}^{-1}\,\boldsymbol{\Gamma}^{-1}\big(\boldsymbol{M}+2\boldsymbol{K}+\boldsymbol{K}\boldsymbol{M}^{-1}\boldsymbol{K}\big)\,\boldsymbol{\Gamma}^{-1}\,\boldsymbol{\Sigma}^{-1}, \end{aligned}$$

which can be written as

$$\boldsymbol{R}^{-1} = \boldsymbol{\Sigma}^{-1}\,\boldsymbol{\Gamma}^{-1}\,\boldsymbol{T}\left(\boldsymbol{R}^+\right)^{-1}\boldsymbol{T}^{\mathrm{T}}\,\boldsymbol{\Gamma}^{-1}\,\boldsymbol{\Sigma}^{-1} \tag{51}$$

where

$$\widehat{\boldsymbol{T}} = \begin{pmatrix} \left(\boldsymbol{M}^{1/2}\right)^{\mathrm{T}} \\ \left(\boldsymbol{K}^{1/2}\right)^{\mathrm{T}} \\ \boldsymbol{M}^{-1/2}\,\boldsymbol{K} \end{pmatrix}, \qquad \left(\widehat{\boldsymbol{R}}^+\right)^{-1} = \begin{pmatrix} \boldsymbol{I} & & \\ & 2\boldsymbol{I} & \\ & & \boldsymbol{I} \end{pmatrix},$$

$\boldsymbol{K} = \boldsymbol{K}^{1/2}\left(\boldsymbol{K}^{1/2}\right)^{\mathrm{T}}$ and $\boldsymbol{M} = \boldsymbol{M}^{1/2}\left(\boldsymbol{M}^{1/2}\right)^{\mathrm{T}}$.

The length-scales (diffusion tensor) are hidden in the definition of $\boldsymbol{K}$. If we assume a constant length-scale ℓ then we can make it explicit in the expressions above by writing $\boldsymbol{K} = \ell^2\widehat{\boldsymbol{K}}$. If we assume further that $\Sigma = \sigma\boldsymbol{I}$ and $\Gamma = \gamma\boldsymbol{I}$ then Eq. (51) can be written in the Brankart form

$$\boldsymbol{R}^{-1} = \boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{R}^+\right)^{-1}\boldsymbol{T}$$

where

$$\boldsymbol{T} = \begin{pmatrix} \left(\boldsymbol{M}^{1/2}\right)^{\mathrm{T}} \\ \left(\widehat{\boldsymbol{K}}^{1/2}\right)^{\mathrm{T}} \\ \boldsymbol{M}^{-1/2}\,\widehat{\boldsymbol{K}} \end{pmatrix} \quad \text{and} \quad \left(\boldsymbol{R}^+\right)^{-1} = \frac{1}{\sigma^2\gamma^2}\begin{pmatrix} \boldsymbol{I} & & \\ & 2\ell^2\boldsymbol{I} & \\ & & \ell^4\boldsymbol{I} \end{pmatrix}.$$

Note that $\boldsymbol{M}$ does not contain any information about derivatives, while $\boldsymbol{K}$ contains products of gradients (see Eqs (27) and (28)). Therefore, multiplying by $\boldsymbol{M}^{1/2}$, $\left(\widehat{\boldsymbol{K}}^{1/2}\right)^{\mathrm{T}}$ and $\boldsymbol{M}^{-1/2}\widehat{\boldsymbol{K}}$ corresponds to differentiation to the zeroth, first and second order, respectively (*cf.* Eq. (48)).

The FEM diffusion-based approach has distinct advantages over the Brankart method for unstructured meshes resulting from sparse or heterogeneously-distributed observations. The discretization of operator $\mathcal{T}$ in Eq. (48) relies directly on the ability to estimate first- and second-order derivatives on the mesh supporting the observations. While this is straightforward when considering structured data on regular grids, it becomes difficult when gaps appear in the spatial distribution of the observations. The FEM discretization described in this study offers a natural framework for handling such difficulties. As the computations rely on the triangulation supporting the observations, the derivatives are estimated at each point using all the information in its neighbourhood. This approximation involves all neighbouring points, even those that are close but do not share exactly the same latitude or longitude.

5. Application to unstructured satellite observations

In this section, we consider a realistic distribution of satellite observations from SEVIRI to illustrate how the FEM-discretized diffusion operator can be used to represent spatially correlated errors. In doing so, we discuss the accuracy of the method by comparing results with those obtained using the theoretical reference (Matérn) correlation function that the diffusion model is intended to represent.

5.1. SEVIRI observations

SEVIRI is a radiometer on board the Meteosat Second Generation satellite, which measures radiances at the top of the atmosphere from 12 different spectral channels (Schmetz *et al.* 2002). SEVIRI radiances provide useful information about temperature and humidity in the troposphere and lower stratosphere. In global numerical weather prediction, SEVIRI radiances are usually assimilated through the clear-sky radiance product, which undergoes cloud-clearing as well as superobbing to 16 pixel by 16 pixel squares (Szyndel *et al.* 2005). In the operational limited-area model AROME* at Météo-France, the raw SEVIRI radiances are assimilated as described by Montmerle *et al.* (2007) with some recent adjustments such as the use of a variational bias correction (Auligné *et al.* 2007). The infrared channels are assimilated in clear-sky conditions and above low clouds (Table 1 of Michel (2018)). In this study, we focus on radiances from Channel 5 (wavelength 6.2 μm), which provides information about humidity in the upper troposphere.

The SEVIRI measurements are known for having spatially correlated observation errors (Waller *et al.* 2016a; Michel 2018). Therefore, they are thinned at a spatial resolution of 70km before assimilation in AROME. This thinning, as well as the screening step to remove cloud-contaminated data, result in a large amount of observations that is discarded. It also causes gaps in the spatial distribution of the observations that depend on the meteorological situation. Those gaps can be responsible for the presence of ill-shaped triangular elements in the mesh supporting the observations.

*"Applications de la Recherche à l'Opérationnel à Méso-Echelle" (see Seity *et al.* (2011)).

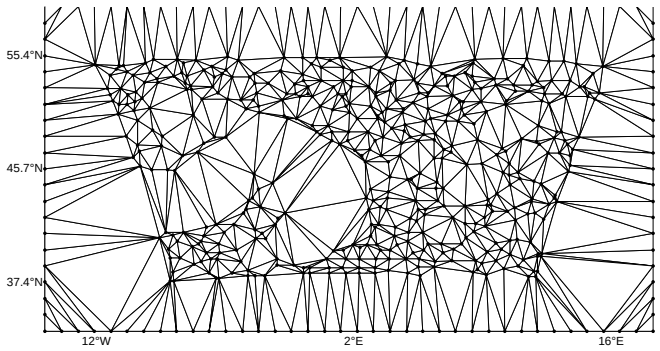
5.2. Mesh generation

The spatial domain is that of AROME. It covers France over an extended region between 12°W to 16°E and 37°N to 55°N. We define a rectangular domain containing the observations, with outer boundaries chosen far from the observation locations, relative to the correlation length-scale (see later), to minimize boundary effects on the solution of the diffusion equation in the interior of the domain. We impose Neumann boundary conditions as they have been naturally accounted for in Eq. (26) through the elimination of the boundary terms after integrating Eq. (24) by parts. The mesh is then built using a constrained Delaunay triangulation algorithm (Edelsbrunner *et al.* 1992), in such a way that the triangular nodes are (exactly) located at the observation locations. Figures 5(a) and 5(b) show examples of the mesh generated from SEVIRI observation locations resulting from two different levels of observation thinning. Figure 5(a) corresponds to the mesh resulting from the thinning algorithm used in the operational AROME model. Figure 5(b) corresponds to the mesh used for the experiments in this study. For this mesh, the total number of nodes is 4980, of which 124 are additional nodes at the artificial domain boundaries.

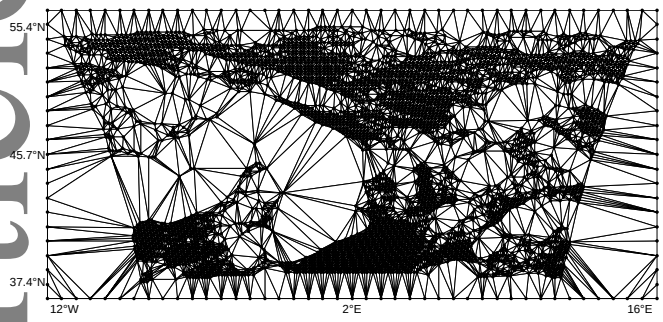
5.3. Impulse response of the spatial correlation operator

In this section, we evaluate the quality of the spatial correlations produced using the FEM-discretized diffusion operator by comparing them to the correlations produced using the analytical Matérn function. We choose constant values for m and ℓ to ensure consistency between the diffusion model and analytical Matérn model. With constant parameters, these models are expected, from theory, to give identical results. The procedure for estimating the values of the parameters m and ℓ is discussed below.

Actual estimates of observation error correlations for Channel 5 SEVIRI radiances have been computed by Waller *et al.* (2016a) and Michel (2018) using Desroziers diagnostics (Desroziers *et al.* 2005). A distinguishing feature of these estimates is the sharpness of the correlations near the origin and the rather slow decay of the correlations at large distances from the origin. As discussed in Section 2.3, this suggests that a Matérn function with a value of $m = 2$ is more appropriate than a Matérn function with a larger



(a) Mesh built from substantially thinned data



(b) Mesh built from mildly thinned data

Figure 5. Triangular mesh constructed from the locations of SEVIRI measurements that have undergone two different levels of thinning. The average distance between observations after thinning is approximately 70 km in (a) and 12 km in (b). The thinning used to generate the mesh in (a) is based on that used in the operational AROME configuration at Météo-France. The experiments in this article employ the mesh in (b). Thin or flat triangles are called "ill-shaped" because their presence is likely to induce numerical errors.

value of m. Therefore, we choose to use this value of m for the diffusion model. Furthermore, we use the Channel 5 correlation estimates from Figure 5a of Waller *et al.* (2016a) as a guideline for choosing a value of ℓ. In particular, we use their result that the distance at which the Channel 5 correlations drop to 0.2 is about 80 km. Correlations beyond 0.2 can be considered insignificant (Liu and Rabier 2003). Assuming that the correlation function is of Matérn type with $m = 2$, then we can invert Eq. (6) to determine ℓ such that $c_{m,\ell}(80\ \mathrm{km}) = 0.2$. This gives $\ell = 32.5$ km, which is the value of ℓ used in the following experiments.

The spatial correlation function at a given point $\boldsymbol{z}_i$ corresponds to the ith column of the correlation matrix $\boldsymbol{C}$. It can be visualized by plotting the result of applying $\boldsymbol{C}$ to a vector that has a value of one at $\boldsymbol{z}_i$ and a value of zero at all other points. Figure 6(a) displays the result of applying the diffusion-based correlation operator (Eq. (42) *without* mass lumping) at six different points in the rectangular domain. The points have been selected to be sufficiently far apart so that the correlation functions do not intersect in any significant way. The points have also been chosen to sample different characteristics of the observation distribution. They include regions where the distribution is dense, sparse, near large gaps, and next to the artificial boundary nodes. A first, qualitative remark to make is that, for all points, the diffusion operator produces sensible, localized structures with spatial extent roughly consistent with the prescribed length-scale and with maximum amplitude close to one.

We can quantify the accuracy of the diffusion-modelled correlation functions by computing their difference with the corresponding Matérn function $c_{m,\ell}$. Denoting the difference field by $\varepsilon_i(\boldsymbol{z}_j)$, $j \in [1, p]$, then for each of the six points $\boldsymbol{z}_i$, $i = [1, 6]$, we have

$$\varepsilon_i(\boldsymbol{z}_j) \;=\; C_{ij} - \big(C_{m,\ell}\big)_{ij}$$

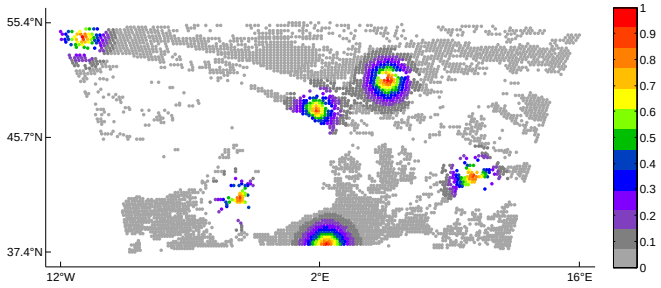
where $C_{ij} = c(\boldsymbol{z}_j, \boldsymbol{z}_i)$ and $\big(C_{m,\ell}\big)_{ij} = c_{m,\ell}(r_{ij})$, with $r_{ij} = \|\boldsymbol{z}_j - \boldsymbol{z}_i\|_2$, are the elements of the diffusion-modelled and Matérn correlation matrices, respectively. The difference field is displayed in Figure 6(b). The errors are small in magnitude (less than 5%) for the points in the densely observed regions, but are up to 20% for the points in sparsely observed regions and near large data gaps. The errors manifest themselves as inaccuracies in the diagonal and off-diagonal elements of the correlation matrix. They are mainly associated with ill-shaped elements in the mesh and the boundary conditions. In the following subsections, we present diagnostics to investigate these errors in more detail.

5.4. Accuracy of the diagonal elements of $\boldsymbol{C}$

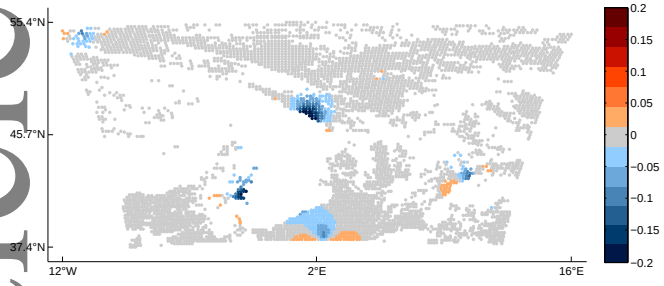
The diagonal elements of the diffusion-based correlation matrix correspond to the amplitude (variance) of the correlation function at each node and should be equal to one. To quantify the amplitude errors of the actual estimates C_{ii}, we compute at each node i the difference

$$\varepsilon_i^{\mathrm{amp}} \;=\; C_{ii} - 1. \tag{52}$$

The amplitude errors are shown in Figure 7. They appear to be minimal far from the boundaries and away from large data gaps. The errors associated with the latter are related to the quality of the mesh in these regions. This will be discussed further in Section 5.6.



(a) Impulse response



(b) Difference with respect to Matérn

Figure 6. (a) Impulse response of the FEM diffusion-based correlation operator at six points in the domain. (b) Difference between the numerical response in (a) and the values of the corresponding Matérn function centred on the same points.

The amplitude errors can be eliminated entirely by re-normalizing the diffusion operator at each point using the actual numerical values of the amplitude at each point. The square-root of the normalization factors are stored in the diagonal matrix $\mathbf{\Gamma}$ of Eq. (40). To diagnose the exact normalization factors requires as many applications of the square-root of the diffusion operator as the number of nodes on the mesh. These computations are expensive, so approximate methods are usually used instead (Weaver and Courtier 2001; Yaremchuk and Carrier 2012). Randomization is one such method, but requires a large number of random vectors to reduce the amplitude error to a satisfactory level (*e.g.*, 1000 vectors are required to reduce the errors to about 4%). Randomization is typically of interest for much larger problems than the one considered in this study.

Near the artificial boundary nodes, Neumann boundary conditions prevent flux exchanges across the boundary, causing the amplitude to double directly at points along straight boundaries (Mirouze and Weaver 2010, Appendix B) and to increase even more in the corners of the domain. The opposite occurs with Dirichlet boundary conditions (*i.e.*, the amplitude is diminished near the boundaries). Procedures to correct the amplitude near boundaries have been proposed in the literature. For example, Mirouze and Weaver (2010) show that the correct amplitude can be obtained by redefining the correlation operator as an average of two diffusion operators, one that employs Neumann boundary conditions and the other that employs Dirichlet boundary conditions. Their analysis was based on solutions of the continuous, one-dimensional (1D) diffusion equation in the presence of an isolated, straight boundary. The method has limitations, however, when applied in higher dimensions and in the presence of complex geometry. Furthermore, since the method involves a sum of diffusion operators, it results in a complicated expression for the inverse of the correlation operator. For this latter reason in particular, it is not considered appropriate for the problem at hand.

Building on the continuous, 1D theoretical analysis of Mirouze and Weaver (2010), Mirouze and Storto (2016) proposed a simple analytical correction to the normalization factor near the boundary as an alternative to the less practical "double-diffusion" approach of Mirouze and Weaver (2010). With Neumann boundary conditions, their analysis suggests that the normalization coefficient should be corrected by a factor $\xi = 1/(1 + c_{m,\ell}(r_\mathrm{b}))$ where r_b is the Euclidean distance to the closest boundary point. For example, directly at the boundary, ξ equals $1/2$ to compensate for the doubling of the amplitude there with Neumann boundary conditions. The expression for the correction factor also suggests that nodes located at distances beyond the correlation length-scale (*i.e.*, such that $c_{m,\ell}(r_\mathrm{b})$ is small) will be largely unaffected by the artificial boundaries. This point has been analyzed in mathematical detail in a recent article by Khristenko *et al.* (2018).

From Figure 7, it is interesting to notice that, apart from a few isolated points in the interior of the domain, the amplitude errors are negative; *i.e.*, the amplitude is mostly underestimated. This suggests that, for the points near the boundary nodes, the amplitude errors are dominated by the effects of large or ill-shaped triangular elements in the mesh, not the (Neumann) boundary conditions. Moving the boundary nodes closer to the interior nodes may reduce the mesh-related errors but at the expense of increasing the boundary condition-related errors. In such a case, corrections like those proposed by Mirouze and Storto (2016) would be needed.

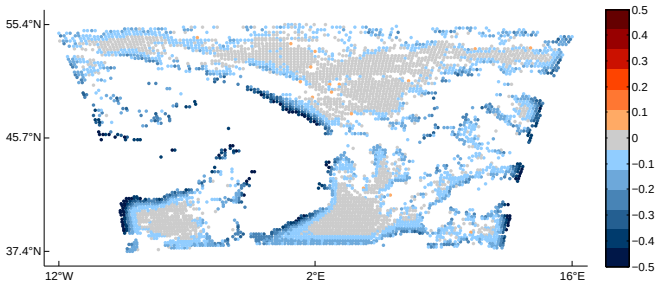


Figure 7. Amplitude error (Eq. (52)) at each node in the domain.

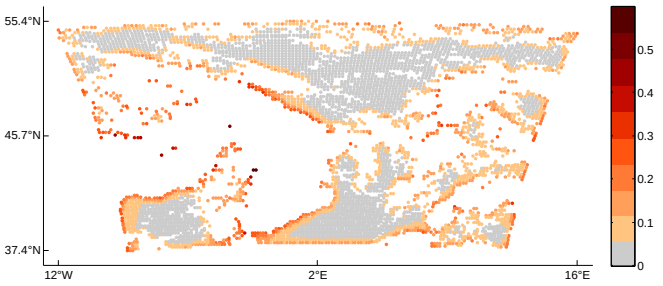


Figure 8. Normalized RMSE of the shape of the correlation function centered on each node in the domain (Eq. (53)).

5.5. *Accuracy of the re-normalized off-diagonal elements of* $\boldsymbol{C}$

The second kind of error concerns the overall shape of the correlation function, which is associated with the accuracy of the off-diagonal elements C_{ij} for $i \neq j$. Even if the amplitude at a particular node is correctly specified (*e.g.*, through re-normalization), the correlation of that node with other nodes might be underestimated or overestimated. We refer to these inaccuracies as shape errors to distinguish them from the amplitude errors discussed in the previous section.

To quantify the shape errors, we can compute the normalized root mean square error (RMSE) between the diffusion-modelled and analytical estimates of the off-diagonal elements:

$$\varepsilon_i^{\text{shape}} = \frac{\left(\sum_j \left|\widehat{C}_{ij} - \left(C_{m,l}\right)_{ij}\right|^2\right)^{1/2}}{\left(\sum_j \left|\left(C_{m,l}\right)_{ij}\right|^2\right)^{1/2}} \tag{53}$$

where

$$\widehat{C}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}}\sqrt{C_{jj}}}$$

is the *exactly* normalized ijth element of matrix $\boldsymbol{C}$.

The shape errors are shown in Figure 8. They exhibit the same basic structure as the amplitude errors in Figure 7, with smallest errors at points where there is a high density of observations and largest errors at points near data gaps and the boundary regions. For the latter, the magnitude of the errors is generally between 10% and 30%, but reaches up to 50% at a few points. Errors within this range ($< 30\%$) can still provide a better approximation to $\boldsymbol{R}$ than assuming strictly uncorrelated errors (Stewart *et al.* 2013).

5.6. *Link between the accuracy of* $\boldsymbol{C}$ *and the quality of the mesh*

The accuracy of the diagonal and off-diagonal elements of the correlation matrix generated by the FEM-discretized diffusion operator is closely linked to the quality of the mesh. In this section, we provide an additional diagnostic to explore this link further.

The aspect ratio ($a(\tau)$) of a triangular element τ is used to compute error bounds in standard applications of the FEM (Ern and Guermond 2010, Section 1.5.1):

$$a(\tau) = \frac{h(\tau)}{\rho(\tau)} \tag{54}$$

where $h(\tau)$ is the size of the largest side of τ and $\rho(\tau)$ is the radius of its inscribed circle (see Figure 9). A large value of the aspect ratio indicates the presence of "flat" elements in the mesh (depicted schematically by the triangular element in the middle in Figure 9), which are typically responsible for causing interpolation errors. The aspect ratio has the property of being scale-invariant; *i.e.*, it only depends on the measure of the angles, but is not affected by the actual size of the edges. Therefore, in using the aspect ratio as a criterion for mesh quality, there is an implicit assumption that the mesh size is locally homogeneous; *i.e.*, any two elements found in the same region of the mesh are assumed to be approximately the same size, so that their "quality" only differs in their aspect ratio. This is generally ensured by mesh generators in standard applications of the FEM in numerical modelling.

In our application, however, the mesh is constrained by the observation locations, which can result in contiguous elements of very different size. As a consequence, the mesh generated from the observations does not satisfy the local homogeneous assumption required for the aspect ratio (54) to be a reliable indicator of mesh

quality. Therefore, we seek an alternative indicator that detects the presence of overly-large elements as well as ill-shaped elements (depicted schematically by the triangular elements on the left and in the middle in Figure 9). Here, we propose the value of the circumradius $r(\tau)$ as one such indicator. It has the advantage of being high both when the triangles contain large angles (large aspect ratio) and when their size is significantly larger than others in the mesh.

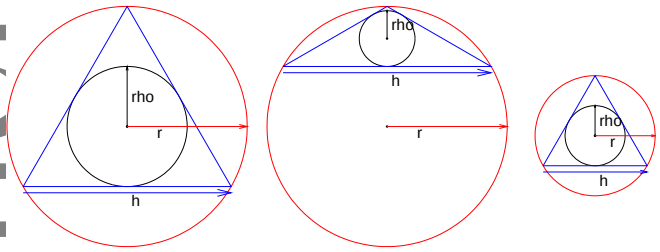


Figure 9. Inscribed circle (black) with radius "rho" (ρ in Eq. (54)) and circumcircle (red) with radius r for three different types of triangular elements whose largest side is denoted by h. The large and "flat" triangular elements correspond to elements of poor quality; they have a large circumradius. The small triangular element corresponds to an element of good quality; it has a small circumradius.

There is no guarantee that given a mesh constructed from an arbitrary distribution of observations, the FEM will lead to small errors in both shape and amplitude. On the contrary, heterogeneously distributed observation locations may cause the elements in the mesh to become ill-shaped, thus leading to increased errors in the FEM discretization. Figure 10 measures the mesh quality for the SEVIRI observation locations in terms of the circumradius of each triangle. Comparing this figure with the error maps (Figures 7 and 8) shows that the locations of the largest errors in both shape and amplitude are highly correlated with the presence of triangles with a large circumcircle radius.

One possible way to improve the quality of the mesh is to eliminate those observations that lead to ill-shaped elements

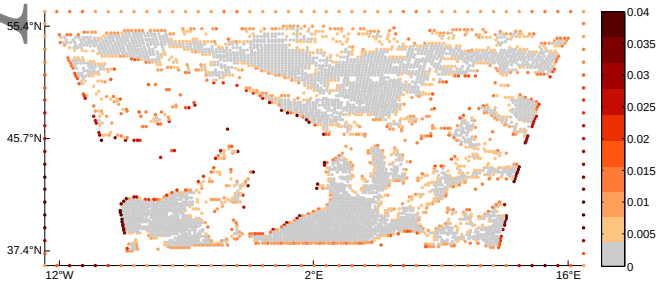
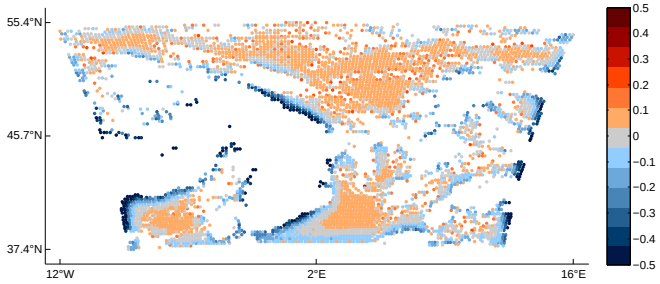


Figure 10. The value of the circumradius of each triangle at each node in the domain. For each node in the triangulation, the largest circumradius of the neighbouring triangles is taken. The map displays the values of these maximum circumradii at all nodes, including those on the artificial boundary.

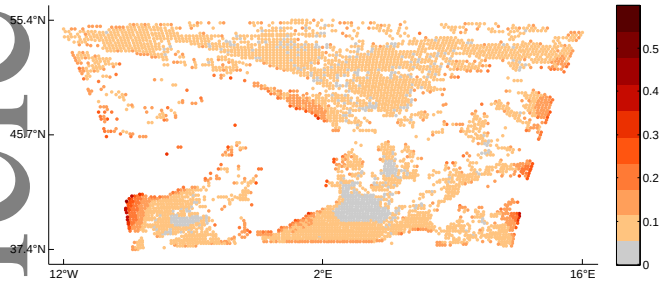
in the FEM. Since the number of observations assimilated in current operational weather prediction systems is very small compared to the number of observations that is actually available (in some cases the number is smaller than 1% of the original set), performing an additional selection of observations based on a mesh-generation criterion is unlikely to deteriorate this ratio significantly. In such a case, removing a few observations would seem to be a reasonable compromise for building a suitable mesh. However, finding an objective criterion to do this correctly is non-trivial. Furthermore, in an operational environment, it would need to be automated and designed to reject as few observations as possible. While an interesting possibility, it is left as a future research direction to develop this idea further.

Another possibility to improve the accuracy of the method is to introduce artificial nodes in order to provide a mesh of better quality. This procedure could be automated using standard mesh-refinement techniques (Ern and Guermond 2010, Chapter 10), and thus seems particularly appealing. However, it leads to complicating issues similar to those encountered by Michel (2018) (and discussed in Section 3.6 for the particular case of the boundary nodes) concerning the representation of $\boldsymbol{C}^{-1}$.

Even in areas where the mesh-related errors are largest, the FEM-based diffusion operator produces a reasonable representation of the spatial correlations, which may be adequate for practical applications, especially in view of our typically inaccurate knowledge of the true observation error covariances. As pointed out by Stewart *et al.* (2008, 2013), it is generally better to have a slightly approximate model for the correlations in $\boldsymbol{R}$ than to neglect them altogether. Using simple analytical models, Fisher (2007) (see also Section 4.9 in Daley (1991)) examined the effects of mis-specifying *background* error covariance parameters on analysis error. His simple scalar example, which illustrates the effects of mis-specifying the background error variance, is equally applicable to the problem of mis-specifying the observation error variance. Specifically, his Figure 3 shows that the analysis error standard deviation is degraded by less than 5% when the background or observation error variance is mis-specified by a factor between 0.5 and 2. This is within the amplitude mis-specification bounds in our experiment, which are roughly



(a) Amplitude error



(b) Shape error

Figure 11. (a) Amplitude error at each node in the domain (Eq. (52)), and (b) normalized RMSE of the shape of the correlation function centered on each node in the domain (Eq. (53)), when the mass lumping approximation is used (cf. Figures 7 and 8).

between 0.5 and 1.2 when using an analytical estimate of the normalization factors (see Figure 7).

5.7. Effect of mass lumping

We recall from Section 3.5 that mass lumping results in a diagonal approximation of the mass matrix. In this section, we examine the effect of mass lumping on the representation of the Matérn correlation functions. The amplitude and shape errors that result from using a mass-lumped matrix are shown in Figures 11(a) and 11(b), respectively. These figures should be compared with Figures 7 and 8, which are the corresponding errors resulting from using the consistent (unapproximated) mass matrix.

The amplitude and shape errors associated with the mass-lumped correlation matrix have similar structures to those associated with the consistent mass matrix. This implies that any mesh-dependent criterion used to predict the errors in the consistent mass formulation will also be relevant for the mass-lumped formulation. However, the magnitude of the errors is significantly larger at some points near data gaps and about 10% larger in areas where there is a high density of observations. Furthermore, mass lumping has a tendency to overestimate the amplitude, as evident by the large patches of positive error in Figure 11(a).

6. Summary and discussion

In this article, we addressed modelling and computational issues that arise when accounting for spatially correlated observation errors in variational data assimilation. Key requirements include the need to handle large covariance matrices, built from heterogeneously distributed observations, and the need to provide an efficient operator for the inverse of the observation error covariance matrix ($\boldsymbol{R}^{-1}$) as well as the observation error covariance matrix ($\boldsymbol{R}$) itself.

We showed how to construct a spatial correlation operator for observation error using a diffusion operator that is discretized with a finite element method (FEM) on a triangular 2D mesh whose nodes are defined by the observation locations. The basic technique has many similarities to the stochastic PDE approach developed by Lindgren *et al.* (2011) for spatial interpolation in geostatistical applications.

The theoretical basis of the diffusion operator approach to correlation modelling is well documented. Here, we considered a diffusion operator that results from integrating a diffusion equation over a finite number of steps with a backward Euler (implicit) scheme. The re-scaled solution of the resulting elliptic equation can be interpreted as a correlation operator whose kernel is a correlation function from the Matérn family. Crucially, the elliptic operator itself provides the corresponding inverse correlation operator, which can be used for defining $\boldsymbol{R}^{-1}$.

In the continuous framework of spatial correlation modelling, we established a formal link between the diffusion operator approach and the method of assimilating directional derivatives of the observations, up to arbitrary order, as proposed by Brankart *et al.* (2009). While the two methods are closely related, the diffusion framework offers better controllable flexibility and a clearer connection with theoretical correlation models. In the discrete framework, we showed how Brankart *et al.* (2009)'s method can be adapted to an unstructured mesh, by formulating the numerical representation of the derivative operators in terms of the mass and stiffness matrices of the FEM-discretized diffusion operator.

The correlation model based on a FEM-discretized diffusion operator was evaluated on an operational data-set from SEVIRI.

To assess the accuracy of the method, results were compared to those produced using the analytical Matérn function, which should be identical for the constant correlation parameter settings considered. A qualitative assessment of the spatially correlated fields centred at various points in the domain showed that the diffusion operator produced sensible structures, with amplitude close to one and spatial scale roughly consistent with the length-scale (square-root of the diffusion coefficient) that was specified. This was the case even in areas where the distribution of observations was extremely irregular.

To quantify the numerical errors, we evaluated separately the diagonal and off-diagonal elements of the diffusion-modelled correlation matrix. The diagonal elements are associated with the amplitude of the modelled correlation functions centered at each point. They were compared with their expected value of one. The errors were shown to be small ($< 5\%$) for points in densely observed regions, but reached 20% for points in sparsely observed regions and near large data gaps. In practice, amplitude errors can be corrected using a re-normalization procedure, although to do so accurately can be costly for very large data-sets.

The off-diagonal elements from the perfectly re-normalized diffusion-modelled matrix were compared with those from the correlation matrix built from the Matérn function. Discrepancies between the two are associated with inaccuracies in the overall shape of the modelled correlation function centred at each point. The shape errors were shown to have a similar spatial structure to the amplitude errors, with errors generally between 10% and 30%. Using the mass lumping approximation for the mass matrix led to a further increase in the magnitude of the errors of about 10%, but did not change the spatial structure of the errors.

Further analysis showed that the largest errors occurred predominantly in areas where the triangular elements in the mesh were ill-shaped; *i.e.*, either too "flat" or too "large". A diagnostic based on the radius of the circumcircle of an element was shown to be a reliable indicator of the quality of the mesh, with large (small) radii being well correlated with large (small) shape and amplitude errors. Although not explored in this study, one interesting possibility to improve the accuracy of the method is to reduce the number of ill-shaped elements in the mesh by using the circumcircle diagnostic in an objective data-thinning criterion prior to the data assimilation step.

Rather than thinning the data to improve the accuracy of the method, another possibility is to refine the mesh by adding extra nodes in areas where observations are missing, as in the Lindgren *et al.* (2011) approach. This procedure results in an auxiliary (higher resolution) mesh, different from the one supporting only the observations, on which the main computations are carried out. An interpolation operator and its adjoint are required to transfer fields between the original mesh and the auxiliary mesh, which results in an $\boldsymbol{R}$ operator with rectangular matrix components, as in the method described by Michel (2018). Therefore, although this approach provides a convenient and more accurate model for $\boldsymbol{R}$, it leads to difficulties in defining $\boldsymbol{R}^{-1}$, which can no longer be represented as an explicit operator.

The method described in this article is generic and could be adapted to other observation types, such as Doppler radar observations, and satellite-derived sea-surface temperature and altimeter observations, which are known to have spatially correlated errors.

Acknowledgement

The authors gratefully acknowledge the support of the French national programme LEFE/INSU. O. Guillet benefitted from an FCPLR (Formation Complémentaire Par la Recherche) grant from Météo-France. The authors thank two anonymous reviewers whose comments helped to improve an earlier version of the paper.

References

Auligné T, McNally AP, Dee DP. 2007. Adaptive bias correction for satellite data in a numerical weather prediction system. *Q. J. R. Meteorol. Soc.* **133**: 631–642, doi:10.1002/qj.56.

Bannister RN. 2008a. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Q. J. R. Meteorol. Soc.* **134**: 1951–1970.

Bannister RN. 2008b. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Q. J. R. Meteorol. Soc.* **134**: 1971–1996.

Bannister RN. 2017. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **143**: 607–633.

Barth A, Beckers JM, Troupin C, Alvera-Azeárate A, Vandenbulcke L. 2014. divand-1.0: n-dimensional variational data analysis for ocean observations. *Geosci. Model Dev.* **7**: 225–241.

Belo Pereira M, Berre L. 2006. The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Weather Rev.* **134**: 2466–2489.

Bolin D, Lindgren F. 2013. A comparison between Markov approximations and other methods for large spatial data sets. *Comp. Statist. Data Anal.* **61**: 7–32.

Bormann N, Bauer P. 2010. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Q. J. R. Meteorol. Soc.* **136**: 1036–1050.

Bormann N, Collard A, Bauer P. 2010. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data. *Q. J. R. Meteorol. Soc.* **136**: 1051–1063.

Brankart JM, Ubelmann C, Testut CE, Cosme E, Brasseur P, Verron J. 2009. Efficient parameterization of the observation error covariance matrix for square root or Ensemble Kalman Filters: Application to ocean altimetry. *Mon. Weather Rev.* **137**: 1908–1927.

Brenner S, Scott LR. 2013. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics, Springer: New York, NY.

Bui-Thanh T, Ghattas O, Martin J, Stadler G. 2013. A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with application to global seismic inversion. *SIAM J. Sci. Comput.* **35**: A2494–A2523.

Campbell WF, Satterfield E, Ruston B, Baker N. 2017. Accounting for correlated observation error in a dual formulation 4d-variational data assimilation system. *Mon. Weather Rev.* **145**: 1019–1032.

Canuto C, Hussaini MY, Quarteroni A, Zang T. 1987. *Spectral Methods in Fluid Dynamics*. Springer: New York, NY.

Carrier MJ, Ngodock H. 2010. Background-error correlation model based on the implicit solution of a diffusion equation. *Ocean Model.* **35**: 45–53.

Chabot V, Nodet M, Papadakis N, Vidard A. 2015. Accounting for observation errors in image data assimilation. *Tellus A* **67**: 23 629, doi:10.3402/tellusa.v67.23629.

Ciarlet P. 2002. *The Finite Element Method for Elliptic Problems*. Classics in Applied Mathematics, SIAM: Philadelphia, PA.

Daley R. 1991. *Atmospheric Data Analysis*. Cambridge Atmospheric and Space Sciences Series, Cambridge University Press: Cambridge, UK.

Dando ML, Thorpe AJ, Eyre JR. 2007. The optimal density of atmospheric sounder observations in the Met Office NWP system. *Q. J. R. Meteorol. Soc.* **133**: 1933–1943.

Davis TS. 2006. *Direct Methods for Sparse Linear Systems*. SIAM: Philadelphia, PA.

Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**: 3385–3396.

Duff IS, Erisman AM, Reid JK. 1989. *Direct Methods for Sparse Matrices*. Oxford University Press: Oxford, UK.

Edelsbrunner H, Tan T, Waupotitsch R. 1992. An $o(n^2 \log n)$ time algorithm for the minmax angle triangulation. *SIAM J. Sci. Comput.* **13**: 994–1008.

Ern A, Guermond JL. 2010. *Theory and Practice of Finite Elements*, *Applied Mathematical Series*, vol. 159. Springer: New York, NY.

Fisher M. 2007. The sensitivity of analysis errors to the specification of background error covariances. In: *Workshop on Flow-dependent Aspects of Data Assimilation*. ECMWF, Reading, UK, pp. 27–36.

Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**: 723–757.

Gratton S, Toint P, Tshimanga J. 2011. A comparison between conjugate gradients and multigrid solvers for covariance modelling in data assimilation. *Q. J. R. Meteorol. Soc.* **139**: 1481–1487.

Guttorp P, Gneiting T. 2006. Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika* **93**: 989–995.

Janjić T, Bormann N, Bocquet M, Carton JA, Cohn SE, Dance SL, Losa SN, Nichols NK, Potthast R, Waller JA, Weston P. 2018. On the representation error in data assimilation. *Q. J. R. Meteorol. Soc.* **144**: 1257–1278, doi: 10.1002/qj.3130.

Järvinen H, Andersson E, Bouttier F. 1999. Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus A* **51**: 469–488.

Jones GH. 1982. *The Theory of Generalised Functions*. Cambridge University Press: Cambridge, UK, 3nd edn.

Khristenko U, Scarabosio L, Swierczynski P, Ullmann E, Wohlmuth B. 2018. Analysis of boundary effects on PDE-based sampling of Whittle-Matérn random fields. *ArXiv e-prints* 1809.07570.

Lindgren F, Rue H, Lindström J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. Roy. Stat. Soc.: Series B Stat. Method.* **73**: 423–498.

Liu ZQ, Rabier F. 2002. The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q. J. R. Meteorol. Soc.* **128**: 1367–1386.

Liu ZQ, Rabier F. 2003. The potential of high-density observations for numerical weather prediction: a study with simulated observations. *Q. J. R. Meteorol. Soc.* **129**: 3013–3035.

Lorenc AC. 1997. Development of an operational variational assimilation scheme. *J. Meteorol. Soc. Jpn.* **75**: 339–346.

Michel Y. 2018. Revisiting Fisher's approach to the handling of horizontal spatial correlations of the observation errors in a variational framework. *Q. J. R. Meteorol. Soc.* **144**: 2011–2025, doi:10.1002/qj.3249.

Mirouze I, Storto A. 2016. Handling boundaries with the one-dimensional first-order recursive filter. *Q. J. R. Meteorol. Soc.* **142**: 2478–2487.

Mirouze I, Weaver AT. 2010. Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Q. J. R. Meteorol. Soc.* **136**: 1421–1443.

Montmerle T, Rabier F, Fischer C. 2007. Relative impact of polar-orbiting and geostationary satellite radiances in the ALADIN/France numerical weather prediction system. *Q. J. R. Meteorol. Soc.* **133**: 655–671, doi:10.1002/qj.34.

Purser RJ, Wu WS, Parrish DF, Roberts NM. 2003. Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: spatially inhomogeneous and anisotropic general covariances. *Mon. Weather Rev.* **131**: 1536–1548.

Rabier F. 2006. Importance of data: A meteorological perspective. In: *Ocean Weather Forecasting: An Integrated View of Oceanography*, Chassignet EP, Verron J (eds), Springer: Dordrecht, NL, pp. 343–360.

Ruggiero GA, Cosme E, Brankart JM, Le Sommer J. 2016. An efficient way to account for observation error correlations in the assimilation of date from the future swot high-resolution altimeter mission. *J. Atmospheric Ocean. Technol.* **33**: 2755–2768.

Saad Y. 2003. *Iterative Methods for Sparse Linear Systems*. SIAM: Philadelphia, PA, 2nd edn.

Schmetz J, Pili P, Tjemkes S, Just D, Kerkmann J, Rota S, Ratier A. 2002. An introduction to Meteosat Second Generation (MSG). *Bull. Am. Meteorol. Soc.* **83**: 977–992.

Seity Y, Brousseau P, Malardel S, Hello G, Bénard P, Bouttier F, Lac C, Masson V. 2011. The AROME-France convective-scale operational model. *Mon. Weather Rev.* **139**: 976–991, doi:10.1175/2010MWR3425.1.

Simpson D, Lindgren F, Rue H. 2012. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics* **23**: 65–74.

Stein ML. 1999. *Interpolation of Spatial Data. Some Theory for Kriging*. Springer: New York, NY.

Stewart LM, Dance SL, Nichols. 2013. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. *Tellus A.* **65**: 1, doi:10.3402/tellusa.v65i0.19546.

Stewart LM, Dance SL, Nichols NK. 2008. Correlated observation errors in data assimilation. *Int. J. Numer. Methods Fluids.* **56**: 1521–1527.

Stewart LM, Dance SL, Nichols NK, Eyre JR, Cameron J. 2014. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. *Q. J. R. Meteorol. Soc.* **140**: 1236–1244.

Stuhlmann R, Rodriguez A, Tjemkes S, Grandell J, Arriaga A, Bézy JL, Aminou D, Bensi P. 2005. Plans for EUMETSAT's Third Generation Meteosat geostationary satellite programme. *Adv. Space Res.* **36**(5): 975–981, doi:https://doi.org/10.1016/j.asr.2005.03.091.

Szyndel M, Kelly G, Thépaut JJ. 2005. Evaluation of potential benefit of assimilation of SEVIRI water vapour radiance data from Meteosat-8 into global numerical weather prediction analyses. *Atmos. Sci. Letters* **6**: 105–111.

Tarantola A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM: Philadelphia, PA.

Ubelmann C, Gaultier L, Fu LL. 2016. SWOT simulator for ocean science. Online documentation. Available at https://github.com/SWOTsimulator/swotsimulator/commits/master/doc/source/science.rst.

Waller JA, Ballard S, Dance SL, Kelly G, Nichols NK, Simonin D. 2016a. Diagnosing horizontal and inter-channel observation error correlations for SEVIRI observations using observation-minus-background and observation-minus-analysis statistics. *Remote Sens.* **8**: 581, doi:10.3390/rs8070581.

Waller JA, Dance SL, Nichols NK. 2016b. Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis residuals. *Q. J. R. Meteorol. Soc.* **142**: 418–431.

Waller JA, Simonin D, Dance SL, Nichols NK, Ballard S. 2016c. Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Mon. Weather Rev.* **144**: 3533–3551.

Weaver AT, Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* **127**: 1815–1846.

Weaver AT, Gürol S, Tshimanga J, Chrust M, Piacentini A. 2018. "Time"-parallel diffusion-based correlation operators. *Q. J. R. Meteorol. Soc.* **144**: 2067–2088, doi:10.1002/qj.3302.

Weaver AT, Mirouze I. 2013. On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Q. J. R. Meteorol. Soc.* **139**: 242–260.

Weaver AT, Tshimanga J, Piacentini A. 2016. Correlation operators based on an implicitly formulated diffusion equation solved with the Chebyshev iteration. *Q. J. R. Meteorol. Soc.* **142**: 455–471.

Weston PP, Bell W, Eyre JR. 2014. Accounting for correlated error in the assimilation of high-resolution sounder data. *Q. J. R. Meteorol. Soc.* **140**: 2420–2429.

Whittle P. 1963. Stochastic processes in several dimensions. *Bull. Inst. Internat. Statist.* **40**: 974–994.

Yaremchuk M, Carrier M. 2012. On the renormalization of the covariance operators. *Mon. Weather Rev.* **140**: 637–649.

Yaremchuk M, D'Addezio JM, Panteleev G, Jacobs G. 2018. On the approximation of the inverse error covariances of high resolution satellite altimetry data. *Q. J. R. Meteorol. Soc.* **144**: 1927–1932, doi:10.1002/qj.3336.

Yaremchuk M, Nechaev D. 2013. Covariance localization with the diffusion-based correlations models. *Mon. Weather Rev.* **141**: 848–860.

Yaremchuk M, Smith S. 2011. On the correlation functions associated with polynomials of the diffusion operator. *Q. J. R. Meteorol. Soc.* **137**: 1927–1932.