Covariance kernels investigation from diffusive wave equations for data assimilation in hydrology

Thibault Malou, Jerome Monnier

Covariance kernels investigation from diffusive wave equations for data assimilation in hydrology

T. Malou[1,2,3], J. Monnier[1,2]
[1] INSA Toulouse, France
[2] Institut de Mathématiques de Toulouse, France
[3] Collecte Localisation Satellite (CLS), Toulouse, France

Abstract

In data assimilation, the estimation of the background error covariance operator is a classical and still open topic. However, this operator is often modeled using empirical information. In order to exploit at best the potential of the knowledge of the physics, the present study proposes a method to derive covariance operators from the underlying equations. In addition, Green's kernels can be used to model covariance operators and are naturally linked to them. Therefore, Green's kernels of equations representing physics can provide physically-derived estimates of the background error covariance operator, and also physically-consistent parameters. In this context, the present covariance operators are used in a Variational Data Assimilation (VDA) process of altimetric data to infer bathymetry in the Saint-Venant equations. In order to investigate these new physically-derived covariance operators, the associated VDA results are compared to the VDA results using classical operators with physically-consistent and arbitrary parameters. The physically-derived operators and physically-consistent exponential operator provide better accuracy and faster convergence than empirical operators, especially during the first iterations of the VDA optimization process.

Keywords. Variational data assimilation; background error; covariance modeling; Green's kernel; diffusive wave equations; river hydraulics.

1 Introduction

Data Assimilation (DA) is a class of inverse problems that aims to improve the background value of a control by combining a physical model with observations of the system state. For example, in meteorology, DA aims to improve the initial state of the atmosphere using a weather prediction model and observations of the atmosphere on an assimilation time window, see e.g. Bouttier and Courtier [2002]. In hydrology, DA also aims to improve bathymetry, roughness coefficient and boundary conditions from observations of the water surface elevation, see e.g. Honnorat et al. [2009]. In the Variational DA methods (VDA), the assimilation is based on the minimization of a cost function which aims at fitting the model outputs to the observations, see e.g. Bouttier and Courtier [2002], Le Dimet and Talagrand [1986], Sasaki [1958].

The quality of the VDA depends strongly on the estimation of the covariance of the background error (i.e. the error between the true and the background value of the control). This topic is therefore crucial but also complex, e.g. see a review on the topic in Bannister [a,b].

The covariance can be estimated statistically by means of sample covariance using methods such as observation-based methods (see e.g. Hollingsworth and Lönnberg [1986]), the NMC method (see e.g. Parrish and Derber [1992])

or ensemble methods (see e.g. Carrassi et al. [2018]). Another possibility is to model the covariance operator. In the present study, the modeling is performed by specifying the covariance kernel. However, the choice of the covariance kernel and the parameters are so far based on empirical information and useful mathematical properties of the kernels. For example, Egbert et al. [1994], Weaver and Courtier [2001] and Mirouze and Weaver [2010] use the Gaussian and Matérn kernels, respectively, for their connection with the diffusion equation. The Gaussian kernel is also studied in Haben [2011], used in Pannekoucke et al. [2016, 2018] and mentioned in Daley [1993] p. 117. The second order auto-regressive kernel (SOAR, a special case of the Matérn kernel) is also widely used, e.g. in Ingleby [2001] for the Met Office 3DVar system, but see also in Daley [1993] p. 117 and Haben et al. [2011], Pannekoucke et al. [2016]. In Monnier and Zhu [2019], the decreasing exponential kernel (another specific case of the Matérn kernel) is used because of the known expression of the associated weighted L_2 norm. In atmospheric science, the turbulent microscale is used to estimate the correlation length as mentioned in Daley [1993] p. 110. However, alternative estimates are often discussed as mentioned in Mirouze and Weaver [2010] and references therein. In Larnier and Monnier [2020], which treats an inverse problem including the present ones, the correlation lengths are taken empirically.

Several studies have provided methods to improve the covariance estimate based on prior knowledge of physics or topography.
Inflation is one method used to improve the statistical estimate of the covariance, see e.g. Carrassi et al. [2018] Sec. 4.4. In Wang et al. [2008], the statistical estimate is improved by using a static covariance that includes more long term knowledge/physics, for example climatological information in a meteorological context. In Lopez-Restrepo et al. [2021], prior topographic knowledge is used. This study also highlights that prior knowledge of the physics, parameters, any environmental information or expertise on the covariance structure could be used to improve the statistical estimate.
In order to reduce the number of stored parameters, the Lagrangian Kalman filter (see Lyster et al. [2004]) takes advantage of a transport model. In this method, the covariance is physically transported along the characteristics. In Pannekoucke et al. [2016], the parametric Kalman filter aims not only to physically transport but also to diffuse an initially known covariance matrix. In Pannekoucke et al. [2018], the previous method is extended to nonlinear behaviors with the example of the Burgers equation. These studies use topographical or physical information. However, this information is only used to improve an already known estimate of the covariance operator.

The present study investigates a method for modeling the covariance kernel based on knowledge of the physics and the underlying equations. It allows (1) to provide covariance kernels consistent with the physical impact of the error at a point on the rest of the domain and (2) to provide an estimate of the correlation length consistent with physics.
The presented method is based on Green's kernels and an extension, called here, the Green-like kernels. These kernels represent the response of the system to a Dirac distribution, i.e. the impact of a perturbation at a given location on the rest of the domain. The Green's kernel can naturally be related to the covariance kernel, see e.g. the application to machine learning in Nagai [2020] and references therein.
An alternative approach to the DA problem is to assimilate noisy data into a stochastic partial differentiable equation in order to infer, using a Bayesian estimator such as the maximum a posteriori, the low-frequency part of a parameter and to estimate the uncertainty due to the high-frequency part. A strategy for this inference problem is proposed by Nolen and Papanicolaou [2009] and is applied to infer the diffusion coefficient in a steady diffusion equation with random potential. This approach also makes a link between Green's kernels and covariance operators.
The proposed method is investigated through the inference of the bathymetry in the Saint-Venant equations. These equations can be simplified and provide the double-scale diffusive wave equations, see Malou and Monnier [2021].

These last equations are adapted to river altimetry observations that we assimilate. Moreover, they allow to derive Green-like kernels.

The outline of the article is as follow. In a first time, the VDA is adapted to infer bathymetry in the Saint-Venant equations form altimetry data. The inference is first performed using so-called classical covariance operators, i.e. based on empirical information. Then, the method to derive covariance kernels from the equations is presented. Finally, the results of the inference using the physically derived covariance operators are compared to those using the classical covariance operators.

2 Variational Data Assimilation based on classical covariance operators

In this 1^{st} section, the formulation of the Variational Data Assimilation (VDA) for a river flow model, the 1D Saint-Venant equations, is presented. Then, covariance operators, which constitute the central part of this study, are recalled, as well as their link with VDA.
Numerical experiments to compare these covariance operators are then defined. A real-like dataset will be used in these experiments and will be presented.
The classical covariance operators are investigated in this section. This will also allow us to determine a reference classical covariance.

2.1 Inference of the bathymetry in the Saint-Venant equations

First, the direct model is presented. Then, the VDA formulation to infer the bathymetry in these equations is presented.

2.1.1 Direct model: the Saint-Venant equations

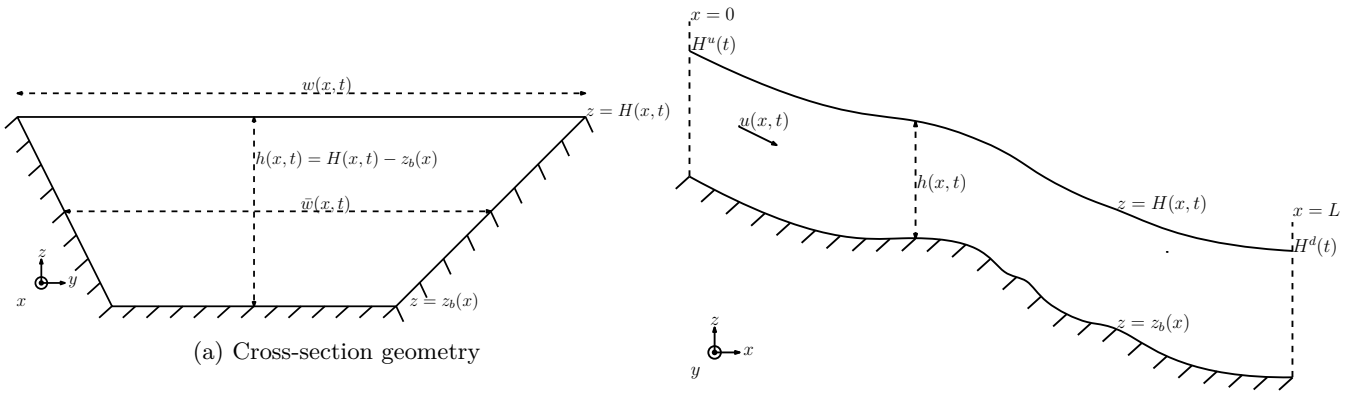
The most classical dynamic river flow model is the one-dimension Saint-Venant equations, see e.g. Chow [1964]. In these equations, the natural variables are the wet area $A(x,t)$ and the discharge $Q(x,t)$. The equations are the following:

$$\begin{cases} \frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \\ \frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}(\frac{Q^2}{A}) + gA\frac{\partial H}{\partial x} = -gAS_f \end{cases} \tag{1}$$

with g the gravity magnitude and $H(x,t)$ the free surface height. The friction term $S_f(x,t)$ is defined by the classical Manning-Strickler relation $S_f = \frac{Q^2}{K_s^2 A^2 h^{\frac{4}{3}}}$ (assuming Q is positive and that for large river, the hydraulics radius is $R_h \approx h$) with K_s the Strickler roughness coefficient. The bathymetry is denoted $z_b(x)$, the average cross-section velocity $u(x,t) = \frac{Q(x,t)}{A(x,t)}$, the water depth $h(x,t) = H(x,t) - z_b(x)$ and the width $w(x,t)$, see Fig. 1. The average width $\bar{w}$ is the river width averaged over a cross-section at a given time. The abscissa x is the curvilinear abscissa of the centerline of the river and $x \in \Omega = [0; L]$. The modeling time window is $t \in [0; T]$. The Saint-Venant equations are solved using the Preissmann scheme [Cunge, 1980] implemented in the DassFlow-1D software [Monnier and Larnier, 2018].

2.1.2 VDA for bathymetry $z_b(x)$ inference

The goal of VDA is to determine the control (denoted k, in a first time $k(x) = z_b(x)$ and in a second time $k(x) = (z_b(x), K_s(x))$) that makes the physical model (called direct, here the Saint-Venant equations (1)) fits with



(a) Cross-section geometry

(b) Longitudinal representation geometry

Figure 1: Assumed geometry of the river and summary of the notations.

some observations. The identification of the optimal control is done by minimizing a cost function j, see e.g. Bouttier and Courtier [2002]:

$$\begin{cases} \text{find } k^*(x) \text{ such that:} \\ k^*(x) = \arg\min_{k(x)} j(k) \end{cases} \tag{2}$$

The cost function is commonly defined as:

$$j(k) = j_{obs}(k) + \alpha_{reg} j_{reg}(k) \tag{3}$$

with α_{reg} the regularization weight coefficient. The observation term j_{obs} aims to minimize the distance between the observations H^{obs} and the output of the direct model (1) $H(z_b)$, see e.g. Bouttier and Courtier [2002], Le Dimet and Talagrand [1986], Sasaki [1958]:

$$j_{obs}(k) = \int_0^T \|H(k;t) - H^{obs}(t)\|^2_{L^2(\Omega^{obs})} dt \tag{4}$$

where Ω^{obs} is the domain of observations.

A regularization term j_{reg} is added to deal with the non-uniqueness of the solution of the problem (2) (Thikhonov regularization term, see e.g. Kaltenbacher et al. [2008]).

For a sake of simplicity, the following is detailed for a single control variable, namely $k = z_b$.

The commonly used expression for j_{reg} is $j_{reg}(z_b) = \|z_b - z_b^b\|^2_{\mathbf{C}^{-1}} = \|r_b\|^2_{\mathbf{C}^{-1}} = \langle r_b, \mathbf{C}^{-1}(r_b)\rangle_{L^2(\Omega)}$. This term tends to minimize the background residual $r_b = z_b - z_b^b$ with z_b^b the background value of z_b. This minimization is done with respect to a norm defined by an positive-definite and symmetric operator $\mathbf{C}^{-1}$.
In the Bayesian framework and under the assumption of Gaussian errors, this expression of j_{reg} follows naturally from the maximum log-likehood of the conditional probability density function of the analysis error $z_b - z_b^t$, with z_b^t the true value of the control. In the same framework, the operator $\mathbf{C}$ is the covariance operator of the background error $\varepsilon_b = z_b^b - z_b^t$ (sometimes denoted $\varepsilon_b^{z_b}$, see Sec. 2.2 and e.g. Bouttier and Courtier [2002]).

Following Parrish and Derber [1992], one introduces the change of variable:

$$z_b(x) = \mathbf{L}\tilde{z}_b(x) + z_b^b(x) \tag{5}$$

This change of variable implies that the operator $\mathbf{L}$, that satisfies $\mathbf{C} = \mathbf{L}\mathbf{L}^T$, must be estimated instead of $\mathbf{C}^{-1}$. This change of variable acts as a preconditioning of the Hessian of j, see Haben et al. [2011]. The change of variable can be understood as an implicit regularization, while the addition of a j_{reg} is an explicit regularization.
From now, j denotes the cost function computed without distinction from z_b or from $\tilde{z}_b$ defined by (5).

The regularization term $j_{reg}(z_b) = \|r_b\|^2_{\mathbf{C}^{-1}}$ does not prevent high frequency oscillations of z_b. The following regularization term is here more adequate:

$$j_{reg}(\tilde{z}_b) = \|\partial_x z_b\|^2_{L^2(\Omega)} = \|\partial_x \left(\mathbf{L}\tilde{z}_b + z_b^b\right)\|^2_{L^2(\Omega)} \tag{6}$$

This regularization term aims to minimize the variations of z_b with respect to the background value z_b^b.
It is also common to infer the Strickler coefficient (as in the sequel experiences, see Sec. 2.3.1) and/or the inflow discharge in addition to the bathymetry in order to provide an effective model (see e.g. Larnier and Monnier [2020] for the inversion of the triplet and Honnorat et al. [2009]).

The optimization algorithm used to solve the problem (2) is the L-BFGS-B algorithm, see Gilbert and Lemaréchal [1989], implemented in the minimize function of the python package scipy.optimize[1]. The gradient of j is estimated by DassFlow-1D solver, see Monnier and Larnier [2018], using the automatic differentiation algorithm TAPENADE, see Hascoet and Pascual [2013].

2.2 Covariance operators in VDA

In the previous section, the change of variable (5) introduces the ε_b covariance operator. $\mathbf{C}$ is a linear integral operator defined by a covariance kernel c:

$$\mathbf{C} \ : \ z \mapsto \left(x \mapsto \int_\Omega c(x;x')z(x') \ dx' \right)$$

Assuming that the distribution of ε_b is known, c is defined by: $c(x;x') = \mathbb{E}\left[(\varepsilon_b(x) - \mathbb{E}[\varepsilon_b(x)])(\varepsilon_b(x') - \mathbb{E}[\varepsilon_b(x')])\right]$.
As mentioned previously, this expression can be estimated using the sample covariance if there are enough realizations or estimations of $\varepsilon_b(x)$ and $\varepsilon_b(x')$.
This expression directly implies that c (and the associated covariance matrix, also denoted $\mathbf{C}$) is symmetric. Moreover, one can show that c is also positive semi-definite. However, since it defines a norm and must be invertible, c is assumed to be positive-definite.
The modeling of the covariance consists in assuming the expression of c. We first consider homogeneous covariance kernels, i.e. which can be written : $c(x;x') = \mathbf{c}(x - x') \Rightarrow \mathbf{C} \ : z \mapsto \mathbf{c} * z$.
Only very few covariance operators are used in the literature on inverse problem and VDA. Let us mention them.

- $\mathbf{C}^{Id}$ is defined by the identity kernel $c^{Id} = Id$ for uncorrelated variables.

[1] see the documentation of the minimize function of the scipy.optimize package and its implementation of the L-BFGS-B algorithm in https://docs.scipy.org/doc/scipy/reference/optimize.minimize-lbfgsb.html

- $\mathbf{C}^G_{L_c}$ is defined by the Gaussian kernel (e.g. in meteorology Egbert et al. [1994]):

$$c^G_{L_c}(x,x') = \frac{1}{\sqrt{2\pi L_c^2}} \exp\left(-\frac{1}{2}\frac{(x-x')^2}{L_c^2}\right) \tag{7}$$

- $\mathbf{C}^e_{L_c}$ is defined by the decreasing exponential kernel (also called First Order Auto-Regressive, FOAR, e.g. used in glaciology Monnier and Zhu [2019]):

$$c^e_{L_c}(x,x') = \exp\left(-\frac{|x-x'|}{L_c}\right) \tag{8}$$

with L_c the correlation length.
Note that the Gaussian and decreasing exponential kernels are specific cases of the Matérn covariance kernel, e.g. used in Mirouze and Weaver [2010]. The Second Order Auto-Regressive kernel (SOAR) is another specific case of the Matérn kernel often used, e.g. in Haben et al. [2011].
In Tarantola [2005], the expression of the $\|\cdot\|_{\mathbf{C}^{-1}}$ norm associated to $\mathbf{C}^e_{L_c}$, see (8), is calculated. One has:

$$\|r_b\|^2_{\mathbf{C}^{-1}} = \frac{1}{L_c}\|r_b\|_2^2 + L_c\|\partial_x r_b\|_2^2 \tag{9}$$

In this first part of the study, the operator $\mathbf{C}^e_{L_c}$ is mainly used to model $\mathbf{C}$.

In the context of inferring z_b in the Saint-Venant equations, no prior information provides a value of L_c. Recall that this correlation length is often estimated from empirical information. For example, in atmospheric data assimilation, the turbulent microscale is used to estimate L_c, see Daley [1993] p.110, Pannekoucke et al. [2008], Mirouze and Weaver [2010] and references therein for related discussions.
In the sequel, the *classical covariance operator* is $\mathbf{C}^e_{L_c}$ defined by (8) with the arbitrary correlation length values L_c equal to $3\Delta x$, $30\Delta x$ and $300\Delta x$, Δx the average space step of the numerical grid.
By abuse of notation, the matrix associated to a covariance operator discretized on the numerical grid, called covariance matrix, is noted by the same symbol.
Since the kernels presented above satisfy $c(x,x) = 1$, they are more specifically correlation kernels. The covariance kernel should be $\sigma(x)\sigma(x')c(x,x')$ with σ the variances.
If σ is considered to be non-constant, the choice of L_c may be influenced by issues related to the positive-definite property, see e.g. Monnier and Zhu [2019]. Here σ is assumed to be constant.
In the case z_b is the only control variable, the constant σ acts as a rescaling factor. Therefore, σ has no impact on the results. In order to have equivalent termination criteria for the optimization algorithm equivalent for each operator, the kernels will be rescaled so that the maximum is 1. In this case, by abuse of language, the correlation kernels are still called covariance kernels.
If two (or more) controls are inferred, the rescaled kernels is then weighted by specified constant variances.

2.3 Numerical experiments description

The numerical experiments, which are analysed in next sections, are here presented. They aim at comparing the classical covariance operator, see Sec. 2.2 and later Sec. 2.4, and the original ones derived from physics, see later Sec. 3.3.3 and Sec. 4. These VDA experiments are performed over a real-like case presented in the last part of this section.

2.3.1 Fully- and partially-observed experiments

The fully-observed experiment

The purpose of the experiments is to investigate the inference of z_b using different covariance operators. To do this, the experiments are based on a target bathymetry z_b^t.

From this target, the observations are $H^{obs} = H(z_b^t)\ \forall x \in \Omega$, see upper Fig. 2a. Moreover, the observations are made over almost the entire time window (every 6 hours on a 10-days time window). This experiment is called *fully-observed*.

To fit to a real-like inference problem, the background z_b^b is assumed to be the linear interpolation of two points (the upstream and downstream points of the target, see upper Fig. 2a). For the same reason, z_b^b is the first guess value of the iterative VDA algorithm.

The partially-observed experiment

In order to consider an experiment closer to real world problems, H^{obs} is in this experiment computed at the km scale by averaging $H(z_b^t) + \varepsilon_{obs}$ between the observations points, with $\varepsilon_{obs} \sim \mathcal{N}(0, 0.1^2)$. Moreover, H^{obs} is taken at only two times ($t = 6$hour and $t = 234$hour) over the (10-days) time window. This is the so-called *partially-observed*.

Moreover, the partially-observed experiment is a multivariate inverse problem that aims at inferring the pair (z_b, K_s). The two control variables are assumed to be uncorrelated: the cross-covariance is $Cov(\varepsilon_b^K(x), \varepsilon_b^{z_b}(x')) = 0\ \forall x, x'$. Hence, the multivariate covariance matrix (containing both the univariate covariance matrices $\mathbf{C}_{K_s}$ and $\mathbf{C}_{z_b}$, and cross-covariance matrix) is block diagonal. This assumption may not be accurate. However, this assumption is necessary without further knowledge of the cross-covariance and hence, is classically made.

The univariate covariance of K_s is assumed to be modeled by $\mathbf{C}_{K_s} = \mathbf{C}^e_{\Delta x}$. As mentioned in Sec. 2.3.2, the assumed value of the Strickler is $K_s^t = 30$. We assume that the background and initial value is $K_s^b = 45$.

In the present study, two pairs of constant are used.

- The 1^{st} pair $(\sigma_{z_b}, \sigma_{K_s}) = (0.5\overline{h}, 10)$ is consistent with the estimate suggested by Larnier and Monnier [2020]. It denotes a realistic estimate of the accuracy of $k^b = (z_b^b, K_s^b)$.
- The 2^{nd} pair is $(\sigma_{z_b}, \sigma_{K_s}) = (0.1\overline{h}, 50)$. This means that the accuracy of z_b^b and K_s^b are respectively overestimated and underestimated. The purpose of this pair is to investigate the inference of z_b in a case in which K_s is mainly inferred .

In the present case, the mean water depth value $\overline{h} \approx 4m$.

Performance criteria

The performance of the optimization with a given c is evaluated by the evolution through the optimization iterations of the Root Mean Square Errors ($RMSE\ X = \sqrt{\overline{(X - X')^2}}$, in m). The $RMSE$ between z_b and z_b^t is denoted $RMSE\ z_b$; $RMSE\ H$ denotes the $RMSE$ between $H(z_b)$ and H^{obs}.

According to Morozov's discrepancy principle, one should optimize until the distance between the observations and the system state reaches the accuracy of the observations, see e.g. Kaltenbacher et al. [2008]. In real-like experiments, the error of altimetric H^{obs} is of the order of $10cm$. Therefore, the evolution of $RMSE$ and z_b until the accuracy $RMSE\ H \approx 10cm$ is reached, is also studied as a criterion of the inference performance.

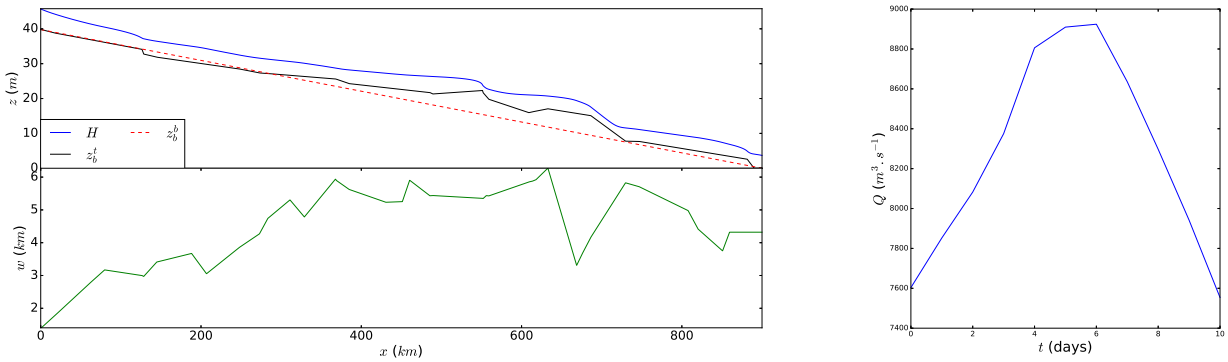
The performance criteria and their evolution through the optimization iterations are also studied until the optimization algorithm terminates. The termination criteria of the L-BFGS-B algorithm used[2] are based on the evolution of j between two iterations and on the maximum of the gradient of j. The values of the tolerances of these criteria

[2] see more details in the documentation of the L-BFGS-B algorithm in https://docs.scipy.org/doc/scipy/reference/optimize.minimize-lbfgsb.html

(resp. 10^{-4} and 3×10^{-2}) are taken so that the first fully-observed experiment (presented in Sec. 2.4.1 and on Fig. 3) has, by expertise, converged sufficiently well. In the following, this stage of the optimization process is called "complete convergence".
The number of iterations to reach these two stages of the optimization process, the stage $RMSE\ H \approx 10cm$ and the complete convergence, is also studied to highlight the number of iteration needed to reach a given accuracy on z_b or on $H(z_b)$ and as a measure of the convexity of the cost j around a given z_b.

2.3.2 The real-like Rio Negro dataset

The VDA experiments are performed over a dataset derived from a Rio Negro dataset (Amazon basin). On Fig. 2a, the width w is measured using the Peckel water mask, as in Pujol et al. [2020], at the intersection of the river with Sentinel-3A/B and Jason3 satellite tracks called virtual stations. Since only the value of w at the free surface is provided, the cross-sections are assumed to be rectangular.
The target z_b^t is estimated from satellite measurements of H and Q estimated by the large-scale hydraulic model MGB (see Collischonn et al. [2007]) at the same virtual stations, see Fig. 2a. The inflow Q, see Fig. 2b, is also estimated by the daily large-scale hydraulic model MGB over a 10-days time interval. More details on the test case construction can be found in Malou et al. [2021]. For simplicity, the Strickler coefficient is considered as constant: $K_s = 30$. The numerical mesh has a mean space step $\Delta x \sim 250m$. The time step of the solver is fixed to $\Delta t = 10min$.



(a) Geometry z_b^t and w, resulting $H(z_b^t)$ and background z_b^b

(b) Inflow discharge

Figure 2: Given geometry and inflow discharge, and target outputs at $t = 10$days of the Rio Negro-like channel test case.

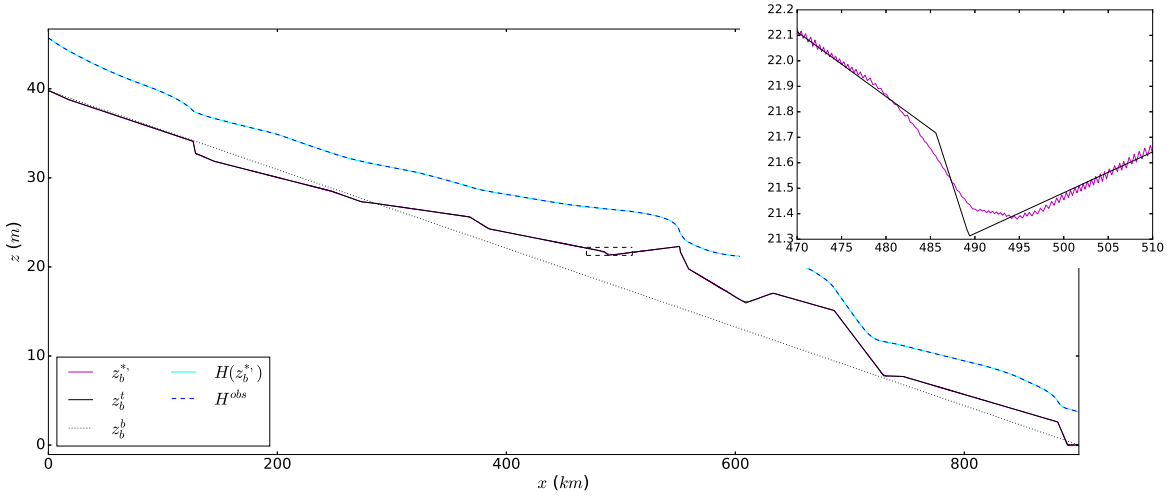
2.4 Numerical results of the fully-observed experiment using classical covariance operators

In this section, the inference of z_b is performed using the different classical covariance operators: the decreasing exponential $\mathbf{C}^e_{L_c}$ defined by (8) with different arbitrary values of L_c, and the identity $\mathbf{C}^{Id}$, see Sec. 2.2.

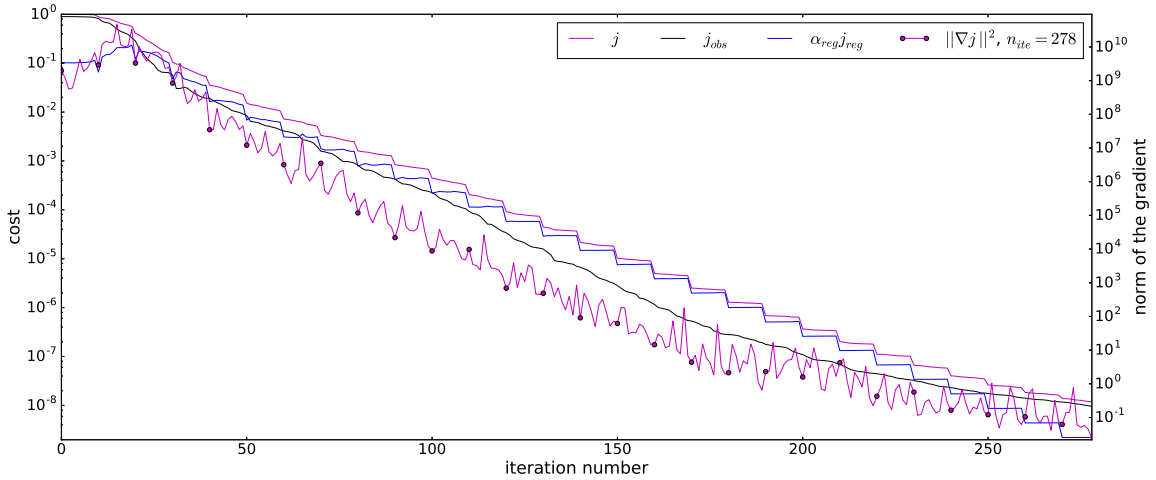
2.4.1 With the identity covariance operator

By expertise, the change of variable (5) with $\mathbf{C}^{Id}$ may not regularize the optimization problem (2) enough: the regularization term of the cost function j_{reg}, defined here by (6), may be needed to converge. In the present experiment, the regularization weight coefficient α_{reg}, see the cost defined by (3), is taken such that $j_{obs}(z_b^b) = 10\% j(z_b^b)$. Moreover, an adaptive regularization strategy is adopted, see e.g. Kaltenbacher et al. [2008]: α_{reg} is divided by 2 every 10 iterations, see on Fig. 3b.

The resulting z_b^* is then very close to z_b^t (nearly indiscernible on Fig. 3a). The main difference comes from



(a) Bathymetry $z_b^{\square}(x)$ (target ^{t}, background ^{b} and optimal *) and the resulting H at $t = 10$days.



(b) Cost: total j, observation j_{obs} and regularization $\alpha_{reg} j_{reg}$ (normalized by the initial total cost $j(z_b^b)$, log scale), and the norm of the gradient vs optimization iterations.

Figure 3: Inference of $z_b(x)$ (fully-observed experiment) with $\mathbf{C}^{Id}$ and the regularization term (6) (α_{reg} initially such that $\alpha_{reg} j_{reg}(z_b^b) \sim 10^{-1} j(z_b^b)$ and divided by 2 every 10 iterations).

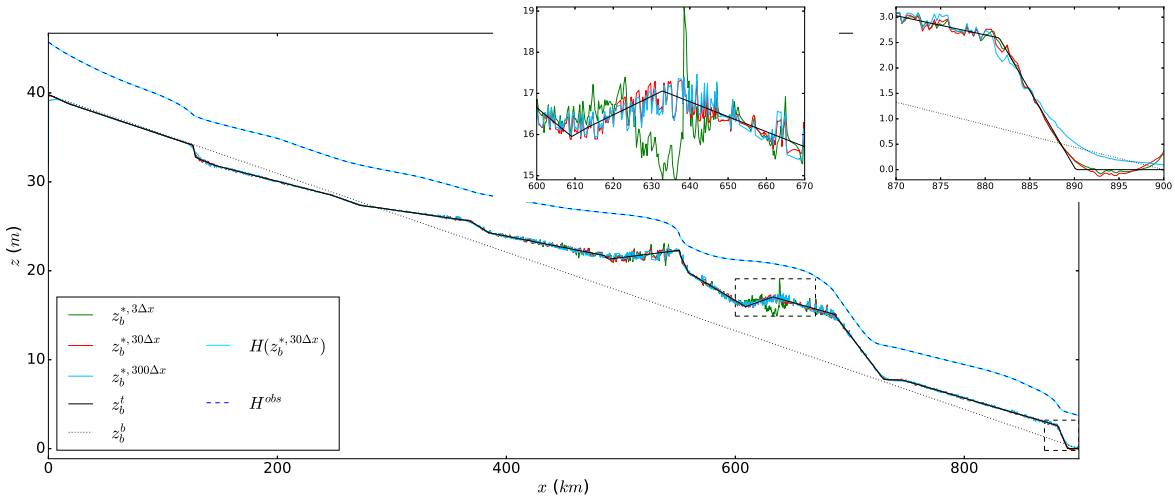
the corners of the z_b^t curve which are smoothed on the z_b^* curve, see the zoom on $x \in [470; 510] km$. This is the

consequence of j_{reg} (6) that aims to provide smoother solutions (but that is needed to converge).

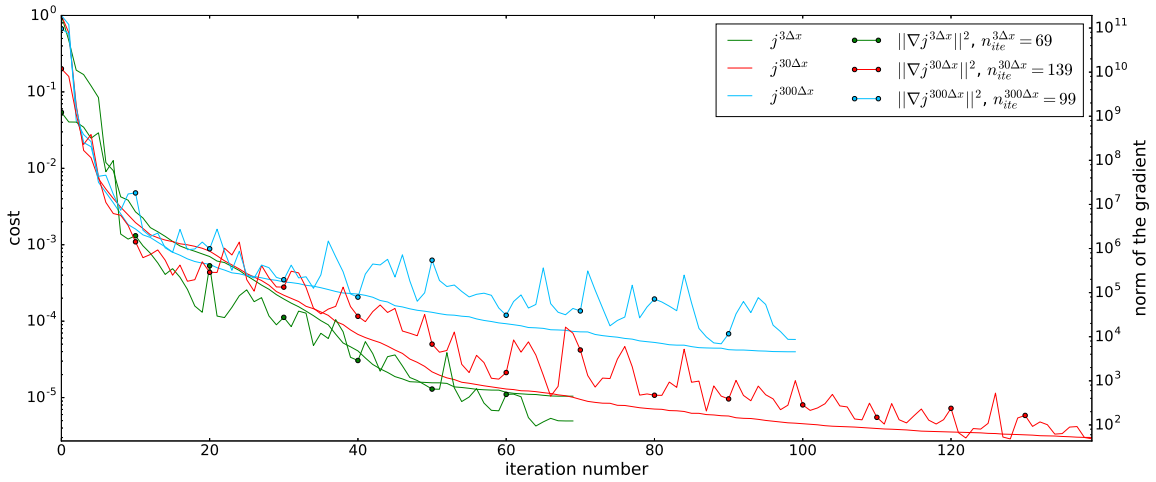
2.4.2 Exponential covariance operator

Following (9), high values of L_c tend to provide smooth solutions z_b^*. On the other hand, low values of L_c tend to provide z_b^* closer to z_b^b (in this case $\mathbf{C}^e_{L_c}$ tends to behave like $\mathbf{C}^{Id}$). By expertise, $\mathbf{C}^e_{L_c}$ regularizes (implicitly) well the optimization problem (2). Therefore, no additional/explicit regularization terms ($\alpha_{reg} = 0$) are needed to converge to a convincing z_b^*.

On Fig. 4a, z_b^* obtained using $L_c = 3\Delta x$ has oscillations of larger amplitudes (e.g. over $x \in [600; 670]km$) than



(a) Bathymetry $z_b^{\square}(x)$ (target ^{t}, background ^{b} and optimal with $\mathbf{C}^e_{L_c}{}^{*,L_c}$) and the resulting H at $t = 10$days.



(b) Total cost j^{L_c} (with $\mathbf{C}^e_{L_c}$, normalized by the initial total cost $j^{L_c}(z_b^b)$, log scale) and the norm of the gradient vs optimization iterations.

Figure 4: Inference of $z_b(x)$ (fully-observed experiment) with $\mathbf{C}^e_{L_c}$ and arbitrary values of L_c ($L_c = 3\Delta x$ in green, $L_c = 30\Delta x$ in red and $L_c = 300\Delta x$ in blue) and no regularization term ($\alpha_{reg} = 0$). The total number of iteration is denoted n_{ite}.

when $L_c = 30\Delta x$ or $L_c = 300\Delta x$ are used. Also, the use of $L_c = 300\Delta x$ seems to provide smoother z_b^* than smaller L_c values, especially at the location of the corners of the z_b^t curve, e.g. on $x \in [870; 900]km$.

2.4.3 Comparisons and choice of the reference method

Covariance operators		$\mathbf{C}^{Id}$	$\mathbf{C}^e_{3\Delta x}$	$\mathbf{C}^e_{30\Delta x}$	$\mathbf{C}^e_{300\Delta x}$
till $RMSE\ H \approx 10cm$	n_{ite}	69	11	10	9
	$RMSE\ z_b\ (m)$	0.44	0.6	0.78	0.31
till complete convergence	n_{ite}	278	69	139	99
	$RMSE\ z_b\ (m)$	0.02	0.23	0.14	0.22
	$RMSE\ H\ (m)$	0.21×10^{-3}	6.61×10^{-3}	3.55×10^{-3}	1.29×10^{-2}

Table 1: Performance (number of iteration n_{ite} and $RMSE\ z_b$ and H) of the inference of z_b using classical covariance operator and $\mathbf{C}^{Id}$ at different stage of the optimization process (till $RMSE\ H \approx 10cm$ and till complete convergence as described in Sec. 2.3.1). The reference classical covariance operator is indicated in blue.

The optimization algorithm with $\mathbf{C}^{Id}$ converges very slowly (278 iterations to reach z_b^* and 69 iterations to reach the $RMSE\ H \approx 10cm$ stage, see Tab. 1). On the other hand, optimizations with $\mathbf{C}^e_{L_c}$ reach the $RMSE\ H \approx 10cm$ stage and z_b^* faster (less than 11 and 139 iterations respectively) than with $\mathbf{C}^{Id}$.
Therefore, $\mathbf{C}^e_{L_c}$ greatly improves the convergence speed of the optimization algorithm, especially in the first iterations.
Moreover, $\|\nabla j\|_2^2$ with $\mathbf{C}^{Id}$ increases a lot during the 15 first iterations (by a factor 10^2). With $\mathbf{C}^e_{L_c}$, $\|\nabla j\|_2^2$ decreases very quickly during the first iterations. This can be explained by the fact that $\mathbf{C}^e_{L_c}$ provides "more convex" cost functions around the reached (local) minimum. This difference in the behavior of ∇j is in agreement with the difference in convergence speed.
One can underline that $\mathbf{C}^{Id}$ provides a higher accuracy. This accuracy comes from the regularization term (6) with a well chosen α_{reg}. Recall that $\alpha_{reg} \neq 0$ is necessary with $\mathbf{C}^{Id}$, unlike with $\mathbf{C}^e_{L_c}$. However, this accuracy is unnecessary in practice and all estimates of $\mathbf{C}$ provide sufficiently accurate z_b^*, see Sec. 2.3.1.

With the operator $\mathbf{C}^e_{L_c}$, using a large value of L_c (e.g. $L_c = 300\Delta x$) implies that the corners of the z_b^t curve are smoothed on the z_b^* curve. On the other hand, using a low value of L_c (e.g. $L_c = 3\Delta x$) implies some oscillations on z_b^*. This explains why a low value of L_c implies at complete convergence a similar $RMSE\ z_b$ but a lower $RMSE\ H$ than a large value of L_c.
Using an intermediate value of L_c (e.g. $L_c = 30\Delta x$) implies that z_b^* has less oscillations than when a low value of L_c is used, and the corners of the z_b^t curve are better approximated than when a high value of L_c is used. This implies that the $RMSE\ H$ and the $RMSE\ z_b$ are (slightly) better, see Fig. 9.

The operator $\mathbf{C}^e_{L_c}$ has the advantage of providing a self-sufficient implicit regularization through the change of variable (5). Recall that no explicit regularization term is needed, i.e. $\alpha_{reg} = 0$. This implicit regularization improves the convergence speed of the optimization algorithm. The algorithm also convergences to a solution with more than sufficient accuracy. Moreover, $\mathbf{C}^e_{L_c}$ with $L_c = 30\Delta x$ is a good compromise since it benefits from the advantage of a high value of L_c (less oscillations than a low value of L_c) but mitigate its disadvantage (the corners of the z_b^t curve are better approximated).
In the sequel, $\mathbf{C}^e_{30\Delta x}$ is then selected as the so-called "classical covariance operator of reference".

3 Covariance operators from Green-like kernels of the double scale diffusive wave equations

The purpose of this section is to derive physically-consistent estimates of $\mathbf{C}$, see Sec. 2.1.2. These physically-consistent covariance operators are derived using Green-like kernels.
In a first part, the definitions of Green's kernel and Green-like kernel are given. Then, a link between a Green's kernel and a covariance operator is established. The inference is made using the Saint-Venant equations (1). However, these equations do not allow to derive Green's kernels (or Green-like). Therefore, these kernels will be derived from the double-scale diffusive wave equations (see Malou and Monnier [2021]), which are first recalled.
These Green-like kernels also provide a tool to quantify the sensitivity of the equations with respect to the control, here z_b. This tool is called the "sensitivity map".
In this context, Green-like kernels do not directly provide covariance kernels. The last part of this section is about obtaining a covariance operator from a Green-like kernel.

3.1 Green-like kernels and their link to covariance operators

3.1.1 Green-like kernels: definition

Suppose a time-dependent 1D linear PDE in the unknown H: $\partial_t H(x,t) + \mathcal{A}H(x,t) = \mathcal{B}k(x,t)$. The linear differential operator $\mathcal{A}$ is assumed to be with constant coefficients, the right hand side operator $\mathcal{B}$ is linear in k with constant coefficient too.

The Green's kernel associated to this PDE is the function $G^{\mathcal{B}}$ which satisfies the equation: $\partial_t G^{\mathcal{B}} + \mathcal{A}G^{\mathcal{B}}(x,t) = \delta(x,t)$, where δ is the Dirac distribution, see Green [1828] and e.g. Evans [1998] for modern use with the similar concept of fundamental solutions.
Since the operator $\mathcal{A}$ is assumed to be linear with constant coefficients, the solution of the PDE can be written as $H(x,t) = \left[G^{\mathcal{B}} * \mathcal{B}k\right](x,t)$. The Malgrange-Ehrenpreis theorem states that a Green kernel exists, see e.g. Brezis [2010].

One can be interested in the response of the system to an impulse in k. Hence, the kernel G, called here *Green-like kernel*, satisfies the equation:

$$\boxed{\partial_t G + \mathcal{A}G(x,t) = \mathcal{B}\delta(x,t)} \tag{10}$$

Since the operator $\mathcal{B}$ is a linear operator with constant coefficients, the solution of the PDE can be written as:

$$H(x,t) = \left[G * k\right](x,t) \tag{11}$$

with G solution of (10).
If the Green kernel exists and is unique, and if the operator $\mathcal{B}$ is linear continuous, then the Riesz representation theorem (see e.g. Brezis [2010]) ensures the existence and uniqueness of the Green-like kernel.
Solving the equation (10) with non-linear operators $\mathcal{A}$ and $\mathcal{B}$, or operators with non-constant coefficients, is generally not possible. Moreover, in such cases, the existence and uniqueness of such kernels are not ensured.

3.1.2 When the Green kernel is naturally a covariance kernel

Let us consider the 1D heat equation with source term:

$$\partial_t H(x,t) - \frac{D}{2}\partial^2_{xx} H(x,t) = k(x,t) \tag{12}$$

with the initial condition $H(x,t=0)=0$, D the diffusion coefficient. The associated Green's kernel reads :

$$G^{\mathcal{B}}(x,t) = \frac{1}{\sqrt{2\pi Dt}} \exp\left(-\frac{1}{2}\frac{x^2}{Dt}\right) \tag{13}$$

This Green's kernel is the Gaussian covariance kernel (7) with the correlation length $L_c = \sqrt{Dt}$: $G^{\mathcal{B}}(x-x',t) = c^G_{\sqrt{Dt}}(x,x')$. Therefore, if this diffusion equation (12) is used to model a given physical phenomenon, then the value of the diffusion coefficient is naturally obtained. Thus, the correlation length can be estimated from D and a characteristic time. Recall from Sec. 2.2 that the correlation lengths for more complex phenomena are classically taken empirically.

Note that here the Green's kernel is naturally symmetric and positive-definite.

The solution of Eqn (12) can then be written as (11): $H = G^{\mathcal{B}} * k = c^G_{\sqrt{Dt}} * k$.

Observe that the same analysis can be done for the 1D homogeneous heat equation $\partial_t H(x,t) - \frac{D}{2}\partial^2_{xx} H(x,t) = 0$ with the initial condition $H(x,t=0) = k(x)$. As suggested by Egbert et al. [1994], the solution of this equation can then be used to estimate $\mathbf{C}^G_{\sqrt{Dt}}(k) = c^G_{\sqrt{Dt}} * k$ without storing the matrix of $\mathbf{C}^G_{\sqrt{Dt}}$. Mirouze and Weaver [2010] links the implicit time discretization of this equation and the Matérn kernel.

If the PDE contains more terms than a simple diffusion term, it is likely that the associated Green's kernel (or Green-like kernel, as in the sequel) provides a covariance kernel containing more dynamics information. The resulting kernel may provide physically-consistent parameters. However, this is valid as long as the resulting Green's kernel is symmetric positive-definite. If this is not the case, one may build up a symmetric positive-definite approximation of the Green's kernel.

3.2 Green-like kernels derived from the double scale diffusive wave

3.2.1 The double scale diffusive wave model

We recall here the double scale-diffusive wave equations which are dedicated to altimetric observations, see Malou and Monnier [2021]. These equations are derived from the Saint-Venant equations (1) under the low Froude assumption ($Fr^2 \ll 1$) and a double scale assumption. The double-scale distinguishes the physical scale and the observations scale. The physical scale variables are denoted by (x,t), the observations scale variables by (y,τ). These two sets of space-time variables are linked by $(y,\tau) = \epsilon_s\,(x,t)$ with ϵ_s the scaling factor [Malou and Monnier, 2021]. The double scale assumption aims at taking into account width variations that are neglected in the classical diffusive wave equation (the "physical scale" equation) but which are not negligible at the observations scale. These are the following two scalar equations:

$$\begin{aligned}
\left(\ \partial_t H - \mu(H)\partial^2_{xx} H + v(H)\partial_x H \qquad\quad \right)(x,t) &= \left(\ v(H)\partial_x z_b \qquad\quad \right)(x,t) \\
\left(\ \partial_\tau H - \mu(H)\partial^2_{yy} H + v(H)\partial_y H + \rho(H) H \ \right)(y,\tau) &= \left(\ v(H)\partial_y z_b + \rho(H) z_b \ \right)(y,\tau)
\end{aligned} \tag{14}$$

with the initial condition $H(\mathcal{X},\mathcal{T}=0)=H_0(\mathcal{X})$. The effective diffusion coefficient μ, wave velocity v, width variation coefficients ρ and the fluid velocity u are defined by:

$$\mu(H;\mathcal{X},\mathcal{T})=\frac{1}{2}u(H;\mathcal{X},\mathcal{T})\frac{(H(\mathcal{X},\mathcal{T})-z_b(\mathcal{X}))}{|\partial_{\mathcal{X}}H(\mathcal{X},\mathcal{T})|} \qquad v(H;\mathcal{X},\mathcal{T})=\gamma u(H;\mathcal{X},\mathcal{T}) \tag{15}$$

$$\rho(H;y,\tau)=\frac{1}{\bar{w}(y,\tau)}\left(\partial_\tau\bar{w}(y,\tau)+\gamma u(H;y,\tau)\partial_y\bar{w}(y,\tau)\right) \quad u(H;\mathcal{X},\mathcal{T})=K_s\left(H(\mathcal{X},\mathcal{T})-z_b(\mathcal{X})\right)^{\gamma-1}\left|\partial_{\mathcal{X}}H(\mathcal{X},\mathcal{T})\right|^{\frac{1}{2}}$$

The expression of u comes from the Manning-Strickler parametrization (with $\gamma=\frac{5}{3}$), see Sec. 2.1.1.
The generic variables $(\mathcal{X},\mathcal{T})$ denote either (x,t) for the physical scale or (y,τ) for the observations scale.

The non-linear equations (14) do not provide a framework for deriving Green's kernels and their existence is not even ensured, see Sec. 3.1.1. Therefore, the double-scale diffusive wave equations (14) are linearized around a reference state that gives constant coefficients.
The linearized diffusive wave equations (14) are non-homogeneous linear reaction-advection-diffusion equations: $\forall(x,t)\in[0,L]\times[0;T]$,

$$\left(\partial_t H-\mu_r^x\partial^2_{xx}H+v_r\partial_x H\right)(x,t)=\left(v_r\partial_x z_b-\eta_r^x z_b\right)(x,t) \tag{16a}$$

$$\left(\partial_\tau H-\mu_r^y\partial^2_{yy}H+v_r\partial_y H+\rho_r H\right)(y,\tau)=\left(v_r\partial_y z_b+\left(\rho_r-\eta_r^y\right)z_b\right)(y,\tau) \tag{16b}$$

with the reference coefficients[3] :

$$\mu_r^{\mathcal{X}}=\frac{1}{2}u_r\frac{H_r}{|\partial_{\mathcal{X}}H_r|},\ v_r=\gamma u_r\ ,\ \eta_r^{\mathcal{X}}=\frac{1}{2}u_r\frac{\partial^2_{\mathcal{X}\mathcal{X}}H_r}{|\partial_{\mathcal{X}}H_r|} \tag{17}$$

$$\rho_r=\frac{1}{\bar{w}_r}\left(\partial_\tau\bar{w}_r+\gamma u_r\partial_y\bar{w}_r\right) \tag{18}$$

The reference state is given by the reference water surface height H_r, the reference velocity u_r and the reference mean width $\bar{w}_r$. We must make sure that μ_r, defined in (17), is positive; if not, a change of reference height solves the issue.
Note that v_r does not depend on the scale but that $\mu_r^{\mathcal{X}}$ and $\eta_r^{\mathcal{X}}$ do (see Tab. 2).

scale	$\mu_r^{\mathcal{X}}$	v_r	$\eta_r^{\mathcal{X}}$	ρ_r
physical scale	10^5	1	10^{-6}	
observation scale	10^2	1	10^{-5}	10^{-4}

Table 2: Magnitudes of the reference coefficients, defined by (17) and (18), at both physical and observations scale with a scaling factor $\epsilon_s=1/250$

From now, the calculations are performed at the observations scale. Indeed and if not specified, the calculations also hold for the physical scale[4] assuming that $\rho_r=0$.
We set: $\mathcal{A}=-\mu_r^{\mathcal{X}}\partial^2_{\mathcal{X}\mathcal{X}}+v_r\partial_{\mathcal{X}}+\rho_r$ and $\mathcal{B}=v_r\partial_{\mathcal{X}}+\left(\rho_r-\eta_r^{\mathcal{X}}\right)$ respectively the right and left hand side operators of (16b).
The operator $\mathcal{B}$ is linear and continuous and the existence of a unique Green's kernel is ensured for the linearized diffusive wave equations (16). Therefore, the existence of a unique Green-like kernel is ensured (by vertue of the Riesz representation theorem, see Sec. 3.1.1).

[3] the subscript $_r$ relates to the reference state and the resulting coefficients
[4] If needed, the observations and physical scales are distinguished by the subscripts $_{ps}$ and $_{os}$.

3.2.2 Derivation of Green-like kernels

First, the classical Green's kernel of the linear diffusive wave equations (16b), denoted by $G^{\mathcal{B}}$, is derived[4]. The Green's kernel is calculated using the Fourier transform in space and the Laplace transform in time (following e.g. Evans [1998]). Moreover, it is assumed that $G^{\mathcal{B}}(\mathcal{X}, \mathcal{T}=0)=0$. To perform the Fourier transform, the equations (16) are defined $\forall \mathcal{X} \in \mathbb{R}$.

We are interested in the response of the system to a perturbation, and not in finding the solution of the initial and boundary values problem. Therefore, the Green's kernel (and later Green-like kernels) does not seek to satisfy either the boundary conditions or the initial condition.

The obtained Green's kernel expression is the following[5] :

$$G^{\mathcal{B}}(\mathcal{X},\mathcal{T}) = \frac{1}{\sqrt{2\pi\left(2\mu_r^{\mathcal{X}}\mathcal{T}\right)}} \exp\left(-\frac{1}{2}\frac{\left(\mathcal{X}-v_r\mathcal{T}\right)^2}{2\mu_r^{\mathcal{X}}\mathcal{T}}\right) \exp\left(-\rho_r\mathcal{T}\right) \mathbf{1}_{]0;+\infty[}(\mathcal{T}) \tag{19}$$

The advection and diffusion terms of the equations (16) involve the Gaussian term (the classical Green's kernel) in $G^{\mathcal{B}}$, see Sec. 3.1.2 or e.g. Evans [1998]. This Gaussian is centered in $v_r\mathcal{T}$ with the correlation length $L_c = \sqrt{2\mu_r^{\mathcal{X}}\mathcal{T}}$ ($\mu_r^{\mathcal{X}} > 0$, v_r is defined by (17)). Therefore, the Green's kernel (19) provides a physically meaningful estimate of the characteristic length L_c, see later (29). Moreover, because of the Gaussian term in the expression of $G^{\mathcal{B}}$, the Gaussian operator, defined by (7), seems to be a natural physically-consistent estimate of $\mathbf{C}$ in this context.

Since the equation (16b) is linear with constant coefficients, the solution $\forall(\mathcal{X},\mathcal{T}) \in \mathbb{R}\times\mathbb{R}^{+*}$ can be written as: $H(\mathcal{X},\mathcal{T}) = \left[\, G^{\mathcal{B}} * \mathcal{B}z_b \,\right](\mathcal{X},\mathcal{T})$. On the other hand, the Green-like kernel, denoted G^{ts}, implies that the solution can also be written as the relation (11). Therefore, the following equality holds: $\left[\, G^{\mathcal{B}} * \mathcal{B}z_b \,\right](\mathcal{X},\mathcal{T}) = \left[G^{ts} * z_b\right](\mathcal{X},\mathcal{T})$.

Using the convolution product differentiation property, the Green-like kernel reads as follows[6]:

$$G^{ts}(\mathcal{X},\mathcal{T}) = \left(-v_r\frac{\left(\mathcal{X}-v_r\mathcal{T}\right)}{2\mu_r^{\mathcal{X}}\mathcal{T}} + \rho_r - \eta_r^{\mathcal{X}}\right) G^{\mathcal{B}}(\mathcal{X},\mathcal{T})\mathbf{1}_{]0;+\infty[}(\mathcal{T}) \tag{20}$$

with $G^{\mathcal{B}}$ defined by (19).

However, z_b is constant in time. Therefore, the appropriate perturbation to consider is one that is local in space and constant in time (not local in time). To do this, we must first consider a Dirac comb of $n+1$ equidistant Dirac distributions on the time interval $[0;T]$: $\text{Ш}(\mathcal{T}) = \sum_{i=0}^{n} \delta(\mathcal{T} - \frac{T}{n}i)$.

We denote by G^{dc}[7] the kernel that formally satisfies: $(\partial_{\mathcal{T}} G^{dc} + \mathcal{A}G^{dc})(\mathcal{X},\mathcal{T}) = \mathcal{B}\text{Ш}(\mathcal{T})\delta(\mathcal{X})$. This kernel is easily calculated from the kernels (20) using the linearity and time shifting property of the Laplace transform:

$$G^{dc}(\mathcal{X},\mathcal{T}) = \sum_{i=0}^{n_t} G^{ts}(\mathcal{X},\mathcal{T}-\frac{T}{n}i) \tag{21}$$

with G^{ts} defined by (20) and the index of the current time $n_t = \lfloor \frac{n}{T}\mathcal{T} \rfloor$.

[5] the superscript $^{\mathcal{B}}$ relates to the result of a Dirac distribution as right hand side $\mathcal{B}z_b$

[6] the superscript ^{ts} refers to the result of a Dirac distribution local in time and space as z_b

[7] the superscript ^{dc} refers to the result of a Dirac comb as z_b

The next step is to sum an infinite number of Dirac distributions, i.e. $n \to +\infty$. Hence, formally, $\text{Ш}(\mathcal{T}) \xrightarrow[n\to+\infty]{} \mathbf{1}_{[0;T]}(\mathcal{T})$. Then, G^{dc} tends to solve the equation: $\forall \mathcal{T} \in [0;T] \;\; (\partial_{\mathcal{T}} G + \mathcal{A}G)(\mathcal{X},\mathcal{T}) = \mathcal{B}\delta(\mathcal{X})$.
Since, formally, Dirac combs are related to the rectangle rule:

$$G^{dc}(\mathcal{X},\mathcal{T}) \xrightarrow[n\to+\infty]{} \int_0^{\mathcal{T}} G^{ts}(\mathcal{X},\mathcal{T}')d\mathcal{T}' \tag{22}$$

This limit (22) provides the following new Green-like kernels :

$$\boxed{\begin{aligned} G_{ps}(x,t) &= \int_0^t \left(-v_r \frac{(x-v_r t')}{2\mu_r^x t'} - \eta_r^x\right) \frac{1}{\sqrt{2\pi\,(2\mu_r^x t')}} \exp\left(-\frac{1}{2}\frac{(x-v_r t')^2}{2\mu_r^x t'}\right) dt' \\ G_{os}(y,\tau) &= \int_0^\tau \left(-v_r \frac{(y-v_r \tau')}{2\mu_r^y \tau'} + \rho_r - \eta_r^y\right) \frac{1}{\sqrt{2\pi(2\mu_r^y \tau')}} \exp\left(-\frac{1}{2}\frac{(y-v_r\tau')^2}{2\mu_r^y\tau'}\right) \exp\left(-\rho_r \tau'\right) d\tau' \end{aligned}} \tag{23}$$

with the coefficients $\mu_r^{\mathcal{X}}$, v_r, ρ_r and $\eta_r^{\mathcal{X}}$ defined by (17).

Note that since $G^{\mathcal{B}}(\mathcal{X},0) = 0 \;\; \forall \mathcal{X} \in \mathbb{R}$ and formally $G^{\mathcal{B}}(\mathcal{X},\mathcal{T}) \xrightarrow[\mathcal{T}\to 0]{} \delta(\mathcal{X}) \;\; \forall \mathcal{X} \in \mathbb{R}$, see e.g. Evans [1998], the kernel $G^{\mathcal{B}}$ is discontinuous in $(0,0)$.
The kernel G^{ts} (20) also satisfies $G^{ts}(\mathcal{X},\mathcal{T}) \xrightarrow[\mathcal{T}\to 0]{} 0 \;\; \forall \mathcal{X} \in \mathbb{R}^*$. Since $G^{ts}(\mathcal{X},0) = 0 \;\; \forall \mathcal{X} \in \mathbb{R}$, the kernel defined by (20) is continuous for $\forall (\mathcal{X},\mathcal{T}) \in \mathbb{R} \times \mathbb{R}^+ \setminus (0,0)$.
This continuity property of the kernel G^{ts} implies that the sum G^{dc}, defined by (21), and the integrals (22) are well-defined $\forall \mathcal{X} \in \mathbb{R}^*$. Hence, the kernels G_{ps} and G_{os} obtained above are well-defined $\forall (\mathcal{X},\mathcal{T}) \in \mathbb{R}^* \times \mathbb{R}^+$.
There is, nevertheless, no proof that the integral (22) is defined and finite $\forall (\mathcal{X},\mathcal{T}) \in \{0\} \times \mathbb{R}^+$. However, we are interested in the response the impulsion, not the impulsion itself. Hence, the impulsion (i.e. when $\mathcal{T}' = 0$) is not taken into account by the numerical integration.

Note that $\mathcal{X} \mapsto G^{\mathcal{B}}(\mathcal{X},\mathcal{T}) \;\; \forall \mathcal{T} \in \mathbb{R}^+$, defined by (19), is not an even function because of the advection term in the equations (16) ($v_r \neq 0$). Similarly, the kernels (20) and (23) are not even with respect to $\mathcal{X}$. Therefore, $(\mathcal{X},\mathcal{X}') \mapsto G(\mathcal{X}-\mathcal{X}',\mathcal{T})$ is not symmetric. Moreover, there is no proof that this kernel are positive-definite.
Therefore, the kernels G_{ps} and G_{os} do not directly provide a covariance kernel.

3.2.3 A by-product: sensitivity map

The purpose of the following section is to highlight the sensitivity of the double-scale diffusive wave equations (14) with respect to k (here z_b) resulting from the Green-like kernels previously derived. This investigation also stands for the Saint-Venant equations (1) as long as the low Froude and double-scale assumptions stand.
Since v_r is constant, $v_r\mathcal{T}$ is the distance traveled by the wave till the time $\mathcal{T}$. Since v, defined in (15), is a function of $(\mathcal{X},\mathcal{T})$, it can be integrated along the characteristic curve to provide the traveled length l.
The same way, α and ω are respectively the diffusion and the width variation coefficients integrated along the characteristic curve.
From the sum G^{dc} defined by (21), we know that the integrands of the kernels (23) at a given $\mathcal{T}'$ are actually the propagation of this integrand from the time $\mathcal{T} - \mathcal{T}'$ till the time $\mathcal{T}$.
The characteristic curve $X(\mathcal{X},\mathcal{T}_1,\mathcal{T}_2)$ is the position at time $\mathcal{T}_2$ of the particle that was located at the position $\mathcal{X}$

at time $\mathcal{T}_1$: $X(\mathcal{X}, \mathcal{T}-\mathcal{T}', \mathcal{T}) = \mathcal{X} + l(\mathcal{X}, \mathcal{T}, \mathcal{T}')$.

Hence, the traveled length, diffusion coefficient and width variation coefficient integrated along the characteristic curve read:

$$\begin{aligned} l(\mathcal{X}, \mathcal{T}, \mathcal{T}') &= \int_{\mathcal{T}-\mathcal{T}'}^{\mathcal{T}} v(X(\mathcal{X}, \mathcal{T}-\mathcal{T}', s), s) ds \\ \alpha(\mathcal{X}, \mathcal{T}, \mathcal{T}') &= \int_{\mathcal{T}-\mathcal{T}'}^{\mathcal{T}} 2\mu(X(\mathcal{X}, \mathcal{T}-\mathcal{T}', s), s) ds \\ \omega(\mathcal{X}, \mathcal{T}, \mathcal{T}') &= \int_{\mathcal{T}-\mathcal{T}'}^{\mathcal{T}} \rho(X(\mathcal{X}, \mathcal{T}-\mathcal{T}', s), s) ds \end{aligned} \tag{24}$$

with the coefficients v, ρ and u defined in (15). The coefficients μ and η are similarly defined as in (17):

$$\mu(\mathcal{X}, \mathcal{T}) = \frac{1}{2} u(\mathcal{X}, \mathcal{T}) \frac{H(\mathcal{X}, \mathcal{T})}{|\partial_{\mathcal{X}} H(\mathcal{X}, \mathcal{T})|}, \quad \eta(\mathcal{X}, \mathcal{T}) = \frac{1}{2} u(\mathcal{X}, \mathcal{T}) \frac{\partial^2_{\mathcal{X}\mathcal{X}} H(\mathcal{X}, \mathcal{T})}{|\partial_{\mathcal{X}} H(\mathcal{X}, \mathcal{T})|} \tag{25}$$

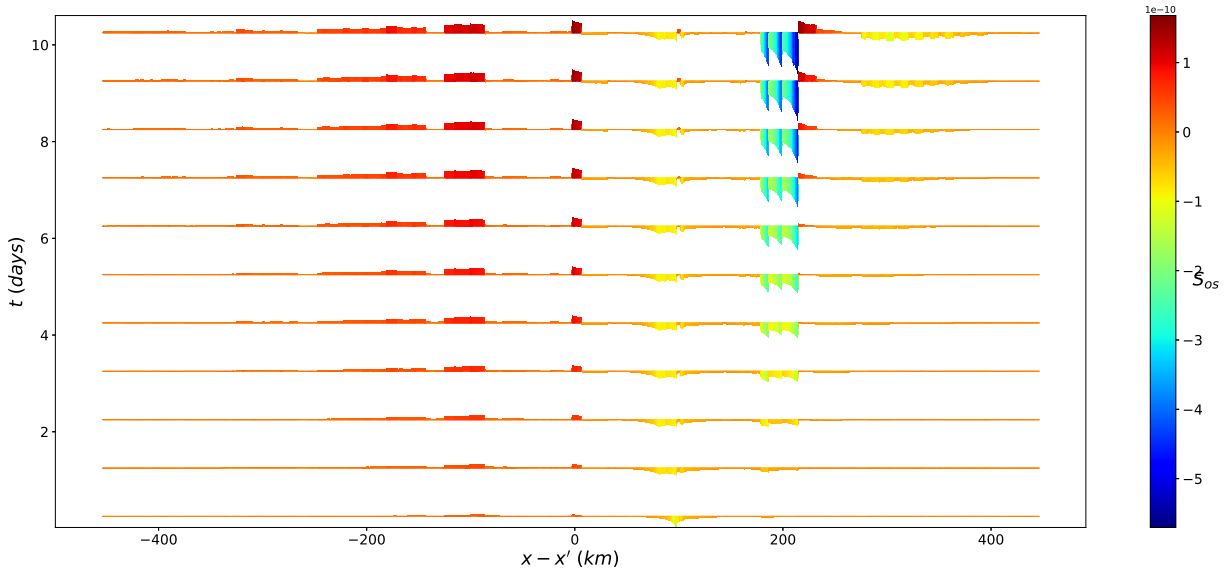
Then an expression of the sensitivities derives from the Green kernels (23). For a perturbation at $\mathcal{X}'$, i.e. for a Dirac distribution $\delta(\mathcal{X} - \mathcal{X}')$, the expressions of the sensitivities S are the following :

$$\begin{aligned} S(\mathcal{X}, \mathcal{X}', \mathcal{T}) = \int_0^{\mathcal{T}} \left(-v(\mathcal{X}, \mathcal{T}-\mathcal{T}') \frac{(\mathcal{X} - \mathcal{X}' - l(\mathcal{X}, \mathcal{T}, \mathcal{T}'))}{\alpha(\mathcal{X}, \mathcal{T}, \mathcal{T}')} + \rho(\mathcal{X}, \mathcal{T}-\mathcal{T}') - \eta(\mathcal{X}, \mathcal{T}-\mathcal{T}') \right) \\ \times \frac{1}{\sqrt{2\pi\alpha(\mathcal{X}, \mathcal{T}, \mathcal{T}')}} \exp\left(-\frac{1}{2} \frac{(\mathcal{X} - \mathcal{X}' - l(\mathcal{X}, \mathcal{T}, \mathcal{T}'))^2}{\alpha(\mathcal{X}, \mathcal{T}, \mathcal{T}')} \right) \exp\left(-\omega(\mathcal{X}, \mathcal{T}, \mathcal{T}')\right) d\mathcal{T}' \end{aligned} \tag{26}$$

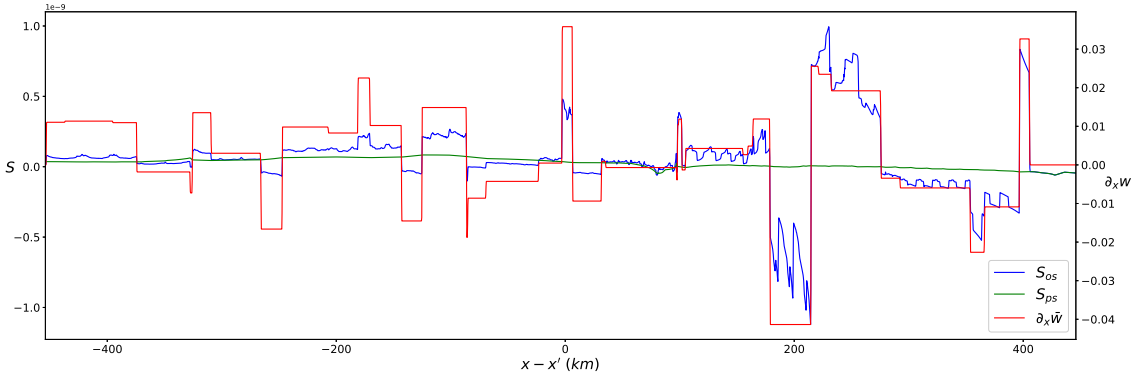
The so-called *sensitivity map* represents the sensitivity, at a given scale, of the control at one point (located at a given $\mathcal{X}'$) on the rest of the domain through time: $S^{\mathcal{X}'} : (\mathcal{X}, \mathcal{T}) \mapsto S(\mathcal{X}, \mathcal{X}', \mathcal{T})$.
Here the sensitivity maps are computed on the whole domain for a perturbation in the middle of the domain $\mathcal{X}' = 0.5L$. The integrals are computed using a rectangle method on the right with a time step of 1h.

We notice that the sensitivity at observations scale S_{os} is discontinuous with respect to $\mathcal{X}$, see Fig. 5. Moreover, on Fig. 5b, the discontinuities are located at the same positions as the discontinuities in the space derivative of the width, see Sec. 2.3.2 for the construction of the width.
The sensitivities are mostly positive upstream and negative downstream of $\mathcal{X}'$.
In absolute value, S_{os} seems to increase in time, see Fig. 5a. Also, the sensitivity map highlights that more information are located far from $\mathcal{X}'$ when the time increases. These phenomena come from the increasing integration domain in time of the integral and from the advection and diffusion phenomena in the integrand of the sensitivities (26).

Finally, one can deduce that the sensitivity of the double scale diffusive wave equations with respect to the z_b is related to the w variations. Moreover, introducing an error in z_b at a point has a positive impact upstream and negative impact downstream on H, Fig. 5a.
This sensitivity map represents the spatio-temporal propagation of the local sensitivity. This sensitivity analysis also stands for the Saint-Venant equations under the low Froude and double gradually varied assumptions.



(a) Sensitivity map at observations scale



(b) Sensitivity maps ($S_{ps}^{\mathcal{X}'}$ at the physical scale and $S_{os}^{\mathcal{X}'}$ at the observations scale) at $t = 10days$ and the river width derivative.

Figure 5: Sensitivity maps (26) for perturbations in the middle of the domain ($\mathcal{X}' = 0.5L$) with the coefficients (15), (24) and (25) computed with the output of the Saint-Venant equations for Rio Negro-like case, see Fig. 2.

3.3 Building physically-derived covariance operators from Green-like kernels

Covariance operators and their associated matrices are by definition symmetric and positive-definite. The kernels $(\mathcal{X}, \mathcal{X}') \mapsto G(\mathcal{X} - \mathcal{X}', \mathcal{T})$ defined by (23) are not symmetric. Moreover, there is no proof that they are positive-definite.

Note that the same issues arise when $\mathbf{C}$ is estimated statistically from the sample covariance, see e.g. Bannister [a]. In this case, the number of realizations or estimates of ε_b may not be sufficient to correctly estimate the covariance. Consequently, spurious correlations may appear and the resulting matrix may not be symmetric positive-definite.

If the inverse operator of the Green-like kernel (the operator G^{-1} such that $z_b = G^{-1}H$) and if a covariance oper-

ator of H is known, an estimate of $\mathbf{C}$ can be computed using the bilinear property of the covariance operator, see e.g. Rasmussen and Williams [2006] chap. 4. However, it is based on a prior knowledge of the covariance of H. Therefore, this approach is not used here.
Thus, the purpose of this section is to construct covariance operators from the Green-like kernels G previously defined.

3.3.1 On the symmetry of the covariance kernel

The kernels $(\mathcal{X}, \mathcal{X}') \mapsto G(\mathcal{X} - \mathcal{X}', \mathcal{T})$, see (23), are not symmetric because of the advection of h, $h = H - z_b$, in (14). This advection implies :

- the Gaussian term of $(\mathcal{X}, \mathcal{X}') \mapsto G^{\mathcal{B}}(\mathcal{X} - \mathcal{X}', \mathcal{T}')$, defined by (19), is not centered in $\mathcal{X}'$ but in $(\mathcal{X}' + v_r\mathcal{T}')$ (the wave travels from upstream to downstream),
- the term $v_r \partial_{\mathcal{X}} z_b$ in the right hand side of (16) implies that the term $(-v_r \frac{\mathcal{X}-\mathcal{X}'-v_r\mathcal{T}'}{2\mu_r^{\mathcal{X}}\mathcal{T}'})$ in $G^{ts}(\mathcal{X} - \mathcal{X}', \mathcal{T}')$, see (20) ($\mathcal{X} \mapsto G^{ts}(\mathcal{X} - \mathcal{X}', \mathcal{T}')$ is positive upstream of $\mathcal{X}' + v_r\mathcal{T}'$ and negative downstream).

In Cheng et al. [2020], the symmetry is forced by considering the kernel $(\mathcal{X}, \mathcal{X}') \mapsto \frac{1}{2}\left(G\left(\mathcal{X} - \mathcal{X}', \mathcal{T}\right) + G\left(\mathcal{X}' - \mathcal{X}, \mathcal{T}\right)\right)$.
Another possibility to force the symmetry is to neglect the advection i.e. to assume that $v_r \approx 0$.
However, this two approximations are inefficient in our case.
Assuming that $v_r \approx 0$ implies that the integrands of Green-like kernels (23) are centered Gaussian. Moreover, the numerical experiments performed below show that the Gaussian kernel is a poor estimate of $\mathbf{C}$ in this context, see Sec. 4 and Sec. 5.
In the present context, forcing symmetry by considering the mean kernel $(\mathcal{X}, \mathcal{X}') \mapsto \frac{1}{2}\left(G\left(\mathcal{X} - \mathcal{X}', \mathcal{T}\right) + G\left(\mathcal{X}' - \mathcal{X}, \mathcal{T}\right)\right)$ provides a kernel that resembles a Gaussian kernel. The results of the optimization algorithm obtained using this mean symmetric kernel are similar to those obtained with the Gaussian kernel. Hence, forcing symmetry in this way provides a poor estimate of $\mathbf{C}$.
Here in order to derive a locally physically-consistent symmetric kernel from G, one consider the kernel $(\mathcal{X}, \mathcal{X}') \mapsto G(|\mathcal{X} - \mathcal{X}'|, \mathcal{T})$.

3.3.2 On the positive-definite property of the covariance kernel

Given the symmetric kernels $(\mathcal{X}, \mathcal{X}') \mapsto G(|\mathcal{X} - \mathcal{X}'|, \mathcal{T})$ with the Green-like kernels (23), we now ensure that the resulting operators are positive-definite.
As mentioned earlier, when $\mathbf{C}$ is estimated by the sample covariance, the estimate may not be positive-definite. Several methods have been developed in the Data Assimilation community to solve this issue, see a review in Carrassi et al. [2018] Sec. 4.4. A first method, called localization, consists in cutting off spurious correlations when $|\mathcal{X} - \mathcal{X}'|$ is empirically too large. However, the cut-off distance which is arbitrary chosen implies a loss of information.
Another method, called inflation, consists in applying a convex combination of the covariance matrix estimate with a target matrix. Usually, this target matrix is either a diagonal or contains some prior/empirical information from static/long-term knowledge (see e.g. Wang et al. [2008]), topography knowledge (see e.g. Lopez-Restrepo et al. [2021]) or expertise on the covariance structure.

In the present study, we try to respect the physics as much as possible. Thus, the previous empirical approaches are not adopted. First, several symmetric kernels with more or less physical terms are presented. Next, given the symmetric matrices associated with these kernels, we enforce the positive-definite property. This results to covariance matrices, defined on the numerical grid.

In (23), the constant coefficients $\mu_r^{\mathcal{X}}$, v_r, ρ_r and $\eta_r^{\mathcal{X}}$ are obtained by averaging their expressions over the space, see (15) and (25), using the output of the direct model (1) at initial time.

Building up symmetric kernels

Using the rectangle method and setting $\mathcal{T} = \Delta\mathcal{T}$ with $\Delta\mathcal{T}$ the integration time step, the kernel $(\mathcal{X}, \mathcal{X}') \mapsto G(|\mathcal{X} - \mathcal{X}'|, \mathcal{T})$ becomes:

$$K^{ts}(\mathcal{X}, \mathcal{X}') = \Delta\mathcal{T}\left(-v_r\frac{(|\mathcal{X}-\mathcal{X}'| - v_r\Delta\mathcal{T})}{2\mu_r^{\mathcal{X}}\Delta\mathcal{T}} + \rho_r - \eta_r^{\mathcal{X}}\right)\frac{1}{\sqrt{2\pi(2\mu_r^{\mathcal{X}}\Delta\mathcal{T})}}\exp\left(-\frac{1}{2}\frac{(|\mathcal{X}-\mathcal{X}'| - v_r\Delta\mathcal{T})^2}{2\mu_r^{\mathcal{X}}\Delta\mathcal{T}}\right)\exp\left(-\rho_r\Delta\mathcal{T}\right) \tag{27}$$

This approximated kernel, denoted by K^{ts}, can be linked to the symmetrization of the Green-like kernel (23), which is the response to a Dirac distribution local in time and space.
This kernel is the product of the Gaussian symmetrized with an exponential and the term derived from the right hand side of the diffusive wave equations (16). The Gaussian results from the physical diffusion and the advection phenomena. The exponential results from the reaction term of the observations-scale diffusive wave equation (16b) and is constant.

Next if we assume that the term derived from the right hand side of the diffusive wave equations $(-v_r\left(\frac{|\mathcal{X}-\mathcal{X}'|-v_r\Delta\mathcal{T}}{2\mu_r^{\mathcal{X}}\Delta\mathcal{T}}\right)+\rho_r-\eta_r^{\mathcal{X}})$ is constant and if omitting the constant multiplicative coefficient for scaling reasons, see Sec. 2.2, the expression (27) becomes:

$$K^{\mathcal{B}}(\mathcal{X}, \mathcal{X}') = \frac{1}{\sqrt{2\pi(2\mu_r^{\mathcal{X}}\Delta\mathcal{T})}}\exp\left(-\frac{1}{2}\frac{(|\mathcal{X}-\mathcal{X}'| - v_r\Delta\mathcal{T})^2}{2\mu_r^{\mathcal{X}}\Delta\mathcal{T}}\right) \tag{28}$$

This symmetrized Gaussian kernel can be seen as the symmetrization of the Green kernel (19).

The derivation of the Green's kernel (19) provide a physically meaningful estimate of L_c depending on a characteristic time. By taking the integration time step as characteristic time, the *physically-derived correlation length* reads :

$$\boxed{L_{pc} = \sqrt{2\mu_r^x\Delta t}} \tag{29}$$

In the forthcoming test cases, one has $\Delta t = 1h$ and $L_{pc} \sim 140\Delta x \sim 35km$.

Also, one can couple the classical covariance kernels $c_{L_c}^G$ defined by (7), which is equivalent to (28) under the assumption that the advection is negligible, and $c_{L_c}^e$ defined by (8) with the physically-derived correlation length L_{pc}.

Symmetric positive-definite approximation of a symmetric matrix

As already mentioned, the operators $\mathbf{K}^{ts}$ and $\mathbf{K}^{\mathcal{B}}$, defined by (27) and (28) respectively, are not guaranteed to be positive definite. We here apply an approximation that imposes this property, and evaluate the consequence of the approximation.
The presented approximation is based on the work of Higham [2002] in a mathematical finance context. The method presented in Higham [2002] aims at computing the closest correlation matrix to a symmetric matrix for a distance

based on a weighted Frobenius norm. Since the set of symmetric positive-semi definite matrices is a closed set, the projection of a symmetric matrix onto this set is defined for this Frobenius norm [Higham, 2002].

The spectral decomposition of a symmetric matrix $M \in \mathcal{M}^{n\times n}(\mathbb{R})$ reads: $M = V\Lambda V^T$. The matrix V is the orthogonal matrix containing the eigenvectors of M as columns and Λ is the diagonal matrix containing the eigenvalues $\Lambda_{i,i} = \lambda_i \ \forall i \in [\![1, n]\!]$. The Frobenius norm is: $\|M\|_{Fr} = \sqrt{tr(MM^T)} = \sqrt{\sum_{i,j=1}^{n} M_{i,j}^2} = \sqrt{\sum_{i=1}^{n} \lambda_i^2}$.

Higham [2002] demonstrates that the projection p of a symmetric matrix on the set of the symmetric positive-semidefinite matrices for $\|\cdot\|_{Fr}$ is: $p(M) = V\Lambda^+V^T$ with Λ^+ is the diagonal matrix such that $\Lambda^+_{i,i} = max(\lambda_i, 0) \ \forall i \in [\![1, n]\!]$.

As recalled before, we are looking for strictly positive-definite not positive-semidefinite matrices. Therefore, the same projection p is not suitable. Also, the set of positive-definite matrices is not a closed set. Hence, the projection of symmetric matrices on this set is not defined. To overcome this issue, the following natural positive-definite approximation is used:

$$p^\epsilon(M) = V\Lambda^\epsilon V^T \tag{30}$$

with Λ^ϵ the approximation of Λ^+ such that $\Lambda^\epsilon_{i,i} = max(\lambda_i, \epsilon) \ \forall i \in [\![0, n]\!]$. Ideally, the threshold should satisfy $\epsilon < \min_{\lambda\in\{\lambda_i | i\in[\![1,n]\!] \text{ and } \lambda_i>0\}} \lambda$.

Symmetric positive-definite approximation of the kernels' matrix

The positive-definite approximation (30) is now applied on the matrix of kernels (27) at observations scale and (28), respectively denoted $\mathbf{K}^{ts}$ and $\mathbf{K}^{\mathcal{B}}$.
The approximation is performed with the threshold $\epsilon = 10^{-14}$. The choice of the value of ϵ is discussed in the last paragraph of this section.
The eigenvalues are computed using the eigh routine of the numpy.linalg python package[8].
We denote: n the number of nodes in the discretization of the domain Ω, $\lambda = \{\lambda_i | i \in [\![1, n]\!]\}$, $\lambda_+ = \{\lambda_i \in \lambda | \lambda_i > 0\}$ and $n_+ = \#\lambda_+$. We have: $n = 3548$.
In order to quantify the accuracy of the approximations, for a matrix $\mathbf{K}$ and two approximations $\mathbf{K}_1$ and $\mathbf{K}_2$, we compute the following criteria:

- the averaged relative difference $\bar{\varepsilon}_{Fr}(\mathbf{K}_1, \mathbf{K}_2) = \frac{1}{|\overline{\mathbf{K}}|n^2}\|\mathbf{K}_1 - \mathbf{K}_2\|_{Fr}$
- the maximal relative difference $\bar{\varepsilon}_{\infty}(\mathbf{K}_1, \mathbf{K}_2) = \frac{1}{|\overline{\mathbf{K}}|} \max_{i,j} \left(|\, (\mathbf{K}_1 - \mathbf{K}_2)_{i,j} \,| \right)$.

$\mathbf{K}$	n_+/n	$\min \lambda$	$\min \lambda_+$	$\overline{\lambda}$	$\max \lambda$
$\mathbf{K}^{ts}$	3538/3548	-2.325×10^{-4}	2.892×10^{-9}	1.255×10^{-7}	8.733×10^{-5}
$\mathbf{K}^{\mathcal{B}}$	26/3548	-2.211×10^{-5}	2.235×10^{-6}	1.140×10^{-5}	4.258×10^{-3}
$\mathbf{C}^G_{L_{pc}}$	1725/3548	-2.294×10^{-18}	7.365×10^{-24}	1.144×10^{-5}	4.006×10^{-3}

Table 3: Spectral properties (the number of positive eigenvalues, the minimum, mean and maximum of the eigenvalues, the minimum of the positive eigenvalues estimated numerically) of the different kernels.

[8] See the documentation of the eigh function in https://numpy.org/doc/stable/reference/generated/numpy.linalg.eigh.html. This function compute the eigenvalues and eigenvectors for a real symmetric matrix. The eigenvalues are computed using a QR algorithm

$\mathbf{K}$	$\overline{\varepsilon}_{Fr}(\mathbf{K}, p^\epsilon(\mathbf{K}))$	$\overline{\varepsilon}_{\infty}(\mathbf{K}, p^\epsilon(\mathbf{K}))$	$\overline{\varepsilon}_{Fr}(p^\epsilon(\mathbf{K}), p(\mathbf{K}))$	$\overline{\varepsilon}_{\infty}(p^\epsilon(\mathbf{K}), p(\mathbf{K}))$
$\mathbf{K}^{ts}$	6.089×10^{-4}	6.306	4.029×10^{-15}	1.708×10^{-10}
$\mathbf{K}^{\mathcal{B}}$	5.442×10^{-6}	1.893×10^{-1}	4.102×10^{-15}	8.645×10^{-10}
$\mathbf{C}^G_{L_{pc}}$	4.333×10^{-15}	9.098×10^{-10}	4.333×10^{-15}	9.098×10^{-10}

Table 4: Performance criteria (average and maximal relative error) of the approximation of the matrices (between the matrix and its positive-definite approximation and between the semidefinite-positive projection and its definite positive approximation).

The approximation $p^\epsilon(\mathbf{K}^{ts})$ is accurate on averaged: $\overline{\varepsilon}_{Fr}(\mathbf{K}^{ts}, p^\epsilon(\mathbf{K}^{ts})) \sim 10^{-4}$, see Tab. 4. It is less good in terms of maximum relative difference: $\overline{\varepsilon}_{\infty}(\mathbf{K}^{ts}, p^\epsilon(\mathbf{K}^{ts})) \approx 6.306$. This can be explained on Fig. 6 by the fact that the values close to zero are fairly well approximated but the higher values are less so.

The kernel $\mathbf{K}^{\mathcal{B}}$ is closer to its positive-definite approximation than $\mathbf{K}^{ts}$: $\overline{\varepsilon}_{Fr}(\mathbf{K}^{\mathcal{B}}, p^\epsilon(\mathbf{K}^{\mathcal{B}})) \sim 10^{-6}$ and $\overline{\varepsilon}_{\infty}(\mathbf{K}^{\mathcal{B}}, p^\epsilon(\mathbf{K}^{\mathcal{B}})) \sim 10^{-1}$. Note that, since $\mathbf{C}^G_{L_c}$ (7) is positive-definite, $\mathbf{K}^{\mathcal{B}}$ is not positive-definite because of the advection phenomena.

The Gaussian operator $\mathbf{C}^G_{L_{pc}}$ defined by (7), with (29), is a well-known covariance operator. Hence, the associated matrix is positive-definite. However, the numerical estimation of the eigenvalues shows the opposite for the matrix $\mathbf{C}^G_{L_{pc}}$, see the negative eigenvalues in Tab. 3. This issue may be the result of positive eigenvalues that are smaller than the accuracy of the algorithm used to estimate them and that are estimated by negative values. The approximation p^ϵ will be applied even if $\mathbf{C}^G_{L_{pc}}$ is positive-definite in theory.
The very small eigenvalues imply that the matrix is ill-conditioned. The ill-conditioned issue of $\mathbf{C}^G_{L_c}$ has already been studied in Koivunen and Kostinski [1999] and in Haben [2011] with respect to L_c.
The approximation $p^\epsilon(\mathbf{C}^G_{L_{pc}})$ in Tab. 4 is still very close to $\mathbf{C}^G_{L_{pc}}$: $\overline{\varepsilon}_{Fr}(\mathbf{C}^G_{L_{pc}}, p^\epsilon(\mathbf{C}^G_{L_{pc}})) \sim 10^{-15}$ and $\overline{\varepsilon}_{\infty}(\mathbf{C}^G_{L_{pc}}, p^\epsilon(\mathbf{C}^G_{L_{pc}})) \sim 10^{-10}$.

The differences $\overline{\varepsilon}_{Fr}(p^\epsilon(\mathbf{K}), p(\mathbf{K}))$ and $\overline{\varepsilon}_{\infty}(p^\epsilon(\mathbf{K}), p(\mathbf{K}))$ for all three kernels imply that p^ϵ is almost as accurate as the projection p (resp. $\sim 10^{-15}$ and $\sim 10^{-10}$ in Tab. 4).

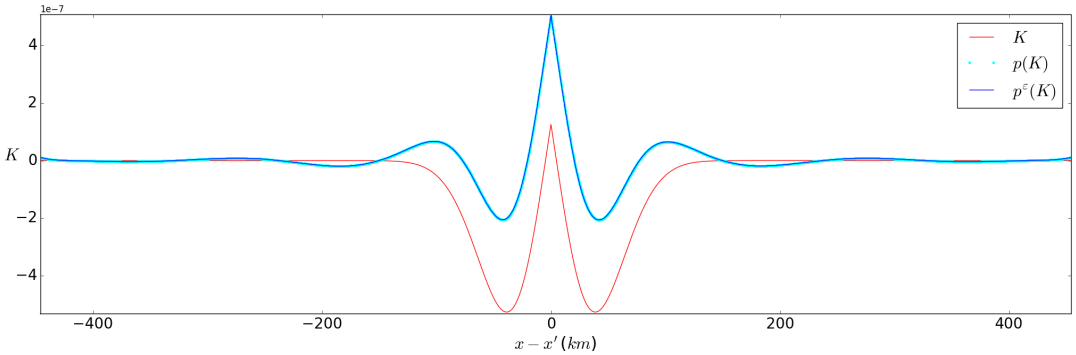


Figure 6: Kernel K^{ts}_{os} (27) (in red), its semidefinite-positive projection p (in cyan) and its positive-definite approximation p^ϵ (30) (in blue) for $x' = 0.5L$.

On the choice of the threshold ϵ

As mentioned previously, the threshold ϵ should ideally satisfy $\epsilon < \min_{\lambda \in \lambda_+} \lambda$. The threshold is here set to $\epsilon = 10^{-14}$.

The approximation p^ϵ has also been tested on the matrix $\mathbf{K}^{ts}$ with $\epsilon = 10^{-9}$, which is still less than the lowest positive eigenvalue, see Tab. 3.
The results of the VDA associated to $\epsilon = 10^{-9}$ are much less accurate than those obtained with $\epsilon = 10^{-14}$ presented below. For example, at complete convergence (see Sec. 2.3.1), the accuracy reached on z_b implies $RMSE\ z_b = 87cm$ when $\epsilon = 10^{-9}$ and $RMSE\ z_b = 33cm$ when $\epsilon = 10^{-14}$, the latter being used in the following and see the performance in Tab. 5.
Therefore, these small eigenvalues may have an impact on the results of the VDA. Let us note that, with $\epsilon = 10^{-9}$, the positive definite approximation p^ϵ is still close to the semidefinite-positive projection p, $\bar{\varepsilon}_{Fr}(p^\epsilon(\mathbf{K}^{ts}), p(\mathbf{K}^{ts})) \approx 10^{-9}$ and $\bar{\varepsilon}_{\infty}(p^\epsilon(\mathbf{K}^{ts}), p(\mathbf{K}^{ts})) \approx 10^{-4}$.

On the other hand, a too low value of ϵ implies that the matrix obtained using the positive-definite approximation is ill-conditioned. This may lead to numerical issues: the Cholesky decomposition algorithm, that is used to compute the change of variable (5), may fail computing the Cholesky decomposition.
This may imply that, as in the case of the Gaussian operator $\mathbf{C}^G_{L_{pc}}$ mentioned above, some very small but positive eigenvalues are lower than the threshold ϵ.

Hence, the value of ϵ is chosen as small as possible in order not to lose information and to have an approximation p^ϵ as close as possible to the projection p (since p^ϵ tends towards p when ϵ goes towards 0).
However, it should not be too small to avoid numerical issues with the Cholesky decomposition algorithm.

3.3.3 Summary of the investigated physically-derived covariance operators

The pseudo-physical covariance operator

In order to get a physically-derived covariance matrix, the approximation p^ϵ is applied to the symmetric matrix $\mathbf{K}^{ts}$. The symmetric matrix $\mathbf{K}^{ts}$ is obtained by forcing the symmetry of the Green-like kernels G given by (23). In the present study, the way considered to force symmetry is $(\mathcal{X}, \mathcal{X}') \mapsto G(|\mathcal{X} - \mathcal{X}'|, \mathcal{T})$. Other ways to force symmetry have been mentioned in Sec. 3.3.1, but they provide poor estimates of $\mathbf{C}$. There are therefore several ways to force symmetry and, so far, the choice of which one to use is based on the experiments.
As in the previous section, the approximation p^ϵ (30) is applied with a threshold $\epsilon = 10^{-14}$. The numerical study of this approximation highlights that one may seek to obtain the smallest ϵ possible to avoid losing some information, but not too small to avoid the matrix being too ill-conditioned and numerical issues.
In the following, the approximation p^ϵ of $\mathbf{K}^{ts}$ is referred as the *pseudo-physical covariance operator*[9]:

$$\boxed{\mathbf{C}^{pp} = p^\epsilon(\mathbf{K}^{ts})} \tag{31}$$

The operator (31) is here used as physically-consistent estimates of the background error covariance operator $\mathbf{C}$.

The physically-derived Gaussian and physically-derived exponential covariance operators

The covariance operators $\mathbf{C}^G_{L_{pc}}$ and $\mathbf{C}^e_{L_{pc}}$, resp. defined by (7) and (8) with L_{pc} the physically-derived correlation length (29), are called *physically-derived Gaussian and physically-derived exponential covariance operators*.
The physically-derived correlation length L_{ps} is directly obtained from the Green's kernel (19). Let us note that no approximation is needed to derive this correlation length.
So far there is here no physical argument for which of these classical operators should be considered. However, they are the most employed covariance operators in the literature for complex highly non linear problems.

[9] the superscript ^{pp} refers to the pseudo-physical covariance

Comparison of the physically-derived covariance operators

From Fig. 7, the curves of c^{pp}_{ps} and c^{pp}_{os} (the kernel c^{pp}, associated to the operator (31), resp. physical and observations scale) are rather close to the curve of $c^{e}_{30\Delta x}$. The curves of the kernels c^{pp}_{ps} and c^{pp}_{os} (and especially their decrease near $\mathcal{X} = \mathcal{X}'$) are even closer to the curve of $c^{e}_{30\Delta x}$ than to the curve of $c^{G}_{L_c}$ with any L_c. This is contrary to the idea that $\mathbf{C}^{G}_{L_c}$ is a natural covariance operator as suggested by the kernels $G^{\mathcal{B}}$ (19) and $K^{\mathcal{B}}$ (28).
The sensitivity map, see Fig. 5, highlights that the variations of w can have a major impact on the sensitivity of the double-scale diffusive wave equations (14) with respect to z_b, see Sec. 3.2.3. However, the curves of c^{pp}_{ps} and c^{pp}_{os} are quite similar on Fig. 7. Therefore, the coefficient ρ_r in the kernel (27), specific to the observations scale and defined by (18), does not have a great impact on the covariance operator. This could be the result of averaging over the space to obtain constant coefficients.
In all the sequel and if not specified, $\mathbf{C}^{pp}$ is considered at observations scale.

Note that one can try to take spatially-distributed coefficients in the expression of K^{ts} (27), as done in Sec. 3.2.3.

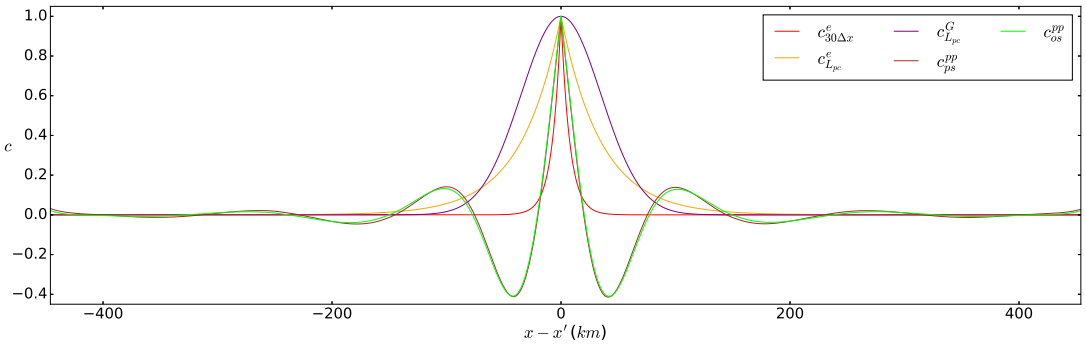


Figure 7: Pseudo-physical covariance kernels c^{pp}_{ps} and c^{pp}_{os} (31) (resp. in brown and in lime green), Gaussian kernel $c^{G}_{L_{pc}}$ (in purple) and exponential kernels $c^{e}_{L_{pc}}$ (in orange) and $c^{e}_{30\Delta x}$ (in red) for $x' = 0.5L$.

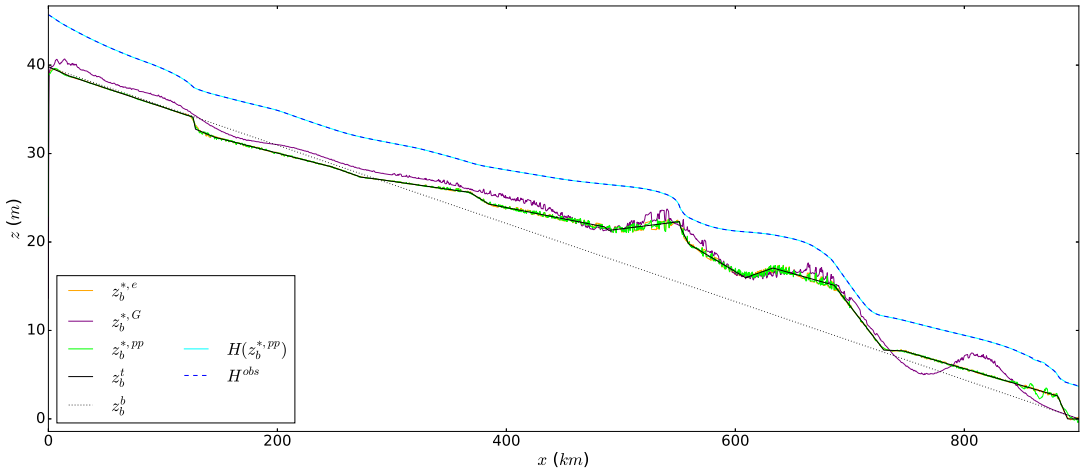
However, the associated fully-observed experiment has shown that these spatially-distributed coefficients imply the kernels to vary in space by several order of magnitude. As a result, the optimization algorithm does not converge correctly.

4 Inference of the bathymetry $z_b(x)$ using the physically-derived covariance operators

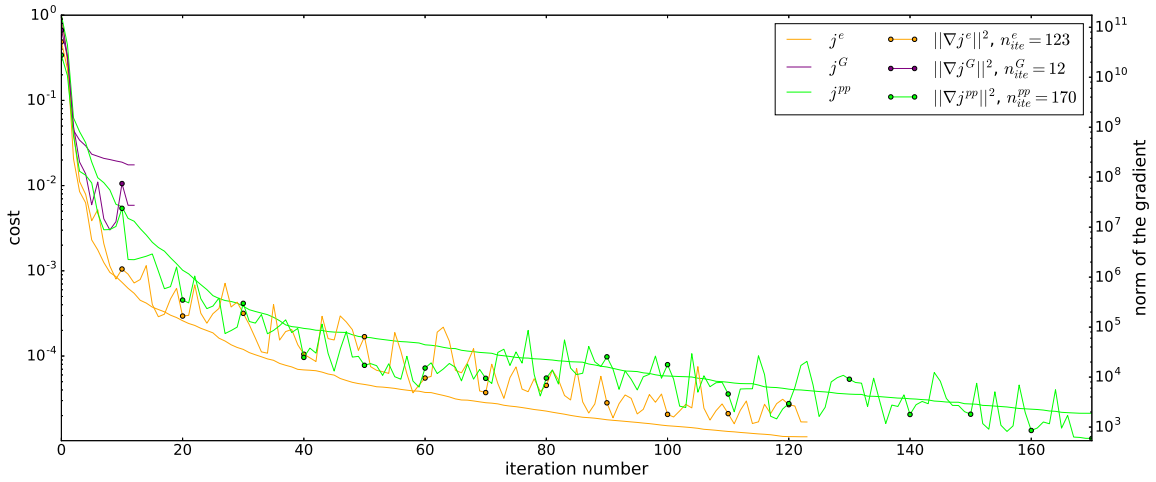
In this section, we infer the river bathymetry z_b following the fully-observed experiment by using the different covariance operators previously defined. The performance of the physically-derived covariance operators, Gaussian $\mathbf{C}^{G}_{L_{pc}}$, exponential $\mathbf{C}^{e}_{L_{pc}}$ and pseudo-physical $\mathbf{C}^{pp}_{os}$, derived in Sec. 3.3, are compared with the performance of the reference covariance operator $\mathbf{C}^{e}_{30\Delta x}$, see Sec. 2.4.

Results and comparison

Overall in Tab. 5 and on Fig. 9, with the exception of $\mathbf{C}^{G}_{L_{pc}}$, the fully-observed experiment with the physically-derived covariance operators and the reference covariance operator provide similar z^{*}_{b}.
In terms of convergence speed, $\mathbf{C}^{e}_{L_{pc}}$ is better especially in the first iterations. Still, $\mathbf{C}^{e}_{30\Delta x}$ and $\mathbf{C}^{pp}_{os}$ have similar convergence speed.



(a) Bathymetry $z_b^{\square}(x)$ (target ^{t}, background ^{b} and optimal with the covariance kernel i ^{*,i}) and the resulting H at $t = 10$days.



(b) Total cost j^i (with the covariance kernel i, normalized by the initial total cost $j^i(z_b^b)$) and the norm of the gradient vs optimization iterations, log scale.

Figure 8: Inference of the bathymetry (fully-observed experiment) with the pseudo-physical at observations scale covariance operator (31) ($i = pp$, in lime green), physically-derived Gaussian (7) ($i = G$, in purple) and exponential (8) ($i = e$, in orange) covariance operators, with $L_c = L_{pc}$ (29)) and no regularization term ($\alpha_{reg} = 0$).

The operator $\mathbf{C}^G_{L_{pc}}$ does not enable the VDA algorithm to converge properly: the minimization stops after 12 iterations without reaching the $RMSE\ H \approx 10cm$ stage, see Figs. 8 and 9. As mentioned in Sec. 3.3.2, the matrix $\mathbf{C}^G_{L_{pc}}$ has many very low eigenvalues. This fact may explain the bad convergence.
Recall from Sec. 3.3.3 that there is no physical argument for using a $\mathbf{C}^G_{L_c}$ as estimate of $\mathbf{C}$.

On Fig. 9, $RMSE\ H$ decreases rather rapidly during the first iterations for $\mathbf{C}^{pp}_{os}$, $\mathbf{C}^e_{L_{pc}}$ and $\mathbf{C}^e_{30\Delta x}$.
Moreover, during the first iterations, $RMSE\ z_b$ decreases faster with $\mathbf{C}^{pp}_{os}$ and $\mathbf{C}^e_{L_{pc}}$ (respectively $0.32m$ and $0.49m$ at $RMSE\ H \approx 10cm$) than with $\mathbf{C}^e_{30\Delta x}$ ($0.78m$).

Covariance operators		Reference a priori	Physically-derived	
		$\mathbf{C}^e_{30\Delta x}$	$\mathbf{C}^e_{L_{pc}}$	$\mathbf{C}^{pp}_{os}$
till $RMSE\ H \approx 10cm$	n_{ite}	10	5	15
	$RMSE\ z_b\ (m)$	0.78	0.33	0.49
till complete convergence	n_{ite}	139	123	170
	$RMSE\ z_b\ (m)$	0.15	0.18	0.33
kernels' expressions		a priori L_c	$L_c = L_{pc}$, see (29)	c^{pp}, see (31)
		$c^e_{L_c}$, see (8)		

Table 5: Performance (number of iteration n_{ite} and $RMSE\ z_b$) of the inference of z_b using different covariance operators (classical/reference and physically-derived) at different stage of the optimization process (till $RMSE\ H \approx 10cm$ and till complete convergence as described in Sec. 2.3.1).

These physically-derived covariance operators seem to improve the convexity feature of the cost function. Indeed, better descent directions are promoted during the first iterations.

In Sec. 2.4.3, the classical operator $\mathbf{C}^e_{30\Delta x}$ is chosen as reference. The choice is motivated by a good performance compromise among the decreasing exponential operator $\mathbf{C}^e_{L_c}$ with arbitrary L_c values.
During the first few iterations, the optimization using $\mathbf{C}^e_{L_{pc}}$ converges faster to a more accurate z_b than the optimization using any of the classical $\mathbf{C}^e_{L_c}$, see the dots on Fig. 9. Also, z_b^* computed with $\mathbf{C}^e_{L_{pc}}$ has very similar accuracy to the one obtained with $\mathbf{C}^e_{30\Delta x}$ but is reached more quickly, see Tab. 5. In this regard, the physically-derived $\mathbf{C}^e_{L_{pc}}$ offers a better performance trade-off than the reference $\mathbf{C}^e_{30\Delta x}$.

The performance (accuracy and speed) of $\mathbf{C}^{pp}_{os}$ is similar to all other $\mathbf{C}^e_{L_c}$ used on Fig. 9. This highlights that the derivation of a covariance operator from a Green-like kernel is promising and provides a credible covariance operator.

Summary

Finally, this investigation of the physically-derived covariance operators through the fully-observed experiment can be summarized as follows.

- The proposed method for estimating the background error covariance operator $\mathbf{C}$ from physics provides good estimates, especially the pseudo-physical covariance operator $\mathbf{C}^{pp}_{os}$, although some approximations have been made in Sec. 3.3.2 to provide symmetric positive-definite matrix. These pseudo-physical operators rely on approximate physics only. However, they rely on the most physics to date.
- The exponential operator $\mathbf{C}^e_{L_c}$ is a good approximation of $\mathbf{C}$. The curve of its kernel (on Fig. 7) is consistent with the curve of c^{pp} (unlike $c^G_{L_c}$).
- In the present context, the widely used kernel $\mathbf{C}^G_{L_c}$ is not a good estimate of $\mathbf{C}$.
- The derivation of the Green-like kernels, see Sec. 3.2.2, provides an estimate of the correlation length consistent with the physics: $L_c = L_{pc}$ defined by (29). The operator $\mathbf{C}^e_{L_c}$ coupled with L_{pc} provides the best results.

Experiments have been performed with lower quality data as in the partially-observed experiment, see Sec. 2.3.1, but inferring only z_b. The results are not shown here because the conclusion is the same than the present one.

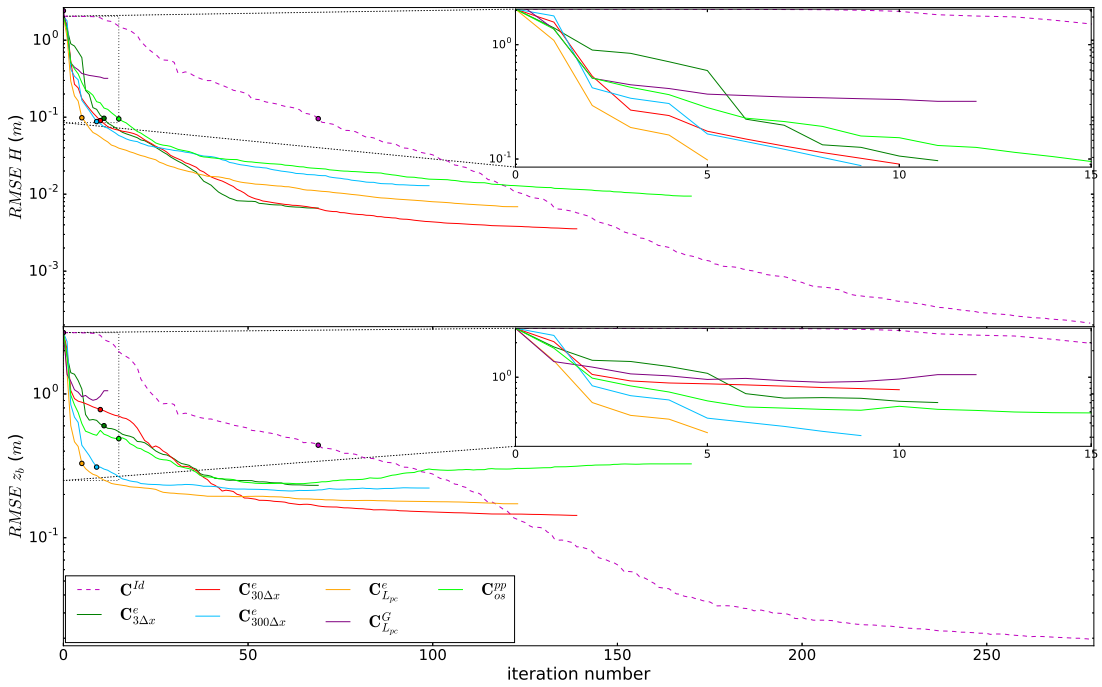


Figure 9: $RMSE$ z_b (upper) and $RMSE$ H (lower, log scale) vs optimization iteration for different covariance operators: $\mathbf{C}^{Id}$ and $\alpha_{reg} \neq 0$ (in purple dashed line), see Fig. 3, $\mathbf{C}^{e}_{L_c}$ (8) with several L_c, see Fig. 4 including L_{pc} (with the empirical values of L_c in green, red and blue, with $L_c = L_{pc}$ in orange), $\mathbf{C}^{G}_{L_{pc}}$ (7) (in purple) and the pseudo-physical covariance kernel $\mathbf{C}^{pp}_{os}$ (31) (in lime green), see Fig. 8, L_{pc} is the physical correlation length (29). The zooms and the dots correspond to the first iterations until $RMSE$ $H \approx 10cm$.

5 Inference of the pair $(z_b, K_s)(x)$ from lower quality data

In this section, the VDA based on the same various estimates of $\mathbf{C}$ is analysed through the partially-observed experiment, that is the inference of the composite control variable (z_b, K_s) from perturbed data (and not perfect ones), see Sec. 2.3.1.

5.1 Results

Recall from Sec. 2.3.1 that, for this multivariate inverse problem, the covariance matrix is assumed to be block diagonal with the block related to the friction $\mathbf{C}_{K_s} = \mathbf{C}^{e}_{\Delta x}$. We consider two pairs of constant variances: $(\sigma_{z_b}, \sigma_{K_s}) = (0.5\overline{h}, 10)$ and $(\sigma_{z_b}, \sigma_{K_s}) = (0.1\overline{h}, 50)$, see Sec. 2.3.1.

5.1.1 Using the variances $(\sigma_{z_b}, \sigma_{K_s}) = (0.5\overline{h}, 10)$

On Fig. 10, except for the result computed with $\mathbf{C}^{G}_{L_{pc}}$, z_b^* is similar to the one obtained in the fully-observed experiment, see Fig. 8. Moreover, K_s^* varies very little from K_s^b.
One can note that z_b^* is almost a translation upward of z_b^t. This comes from the fact that essentially z_b is optimized

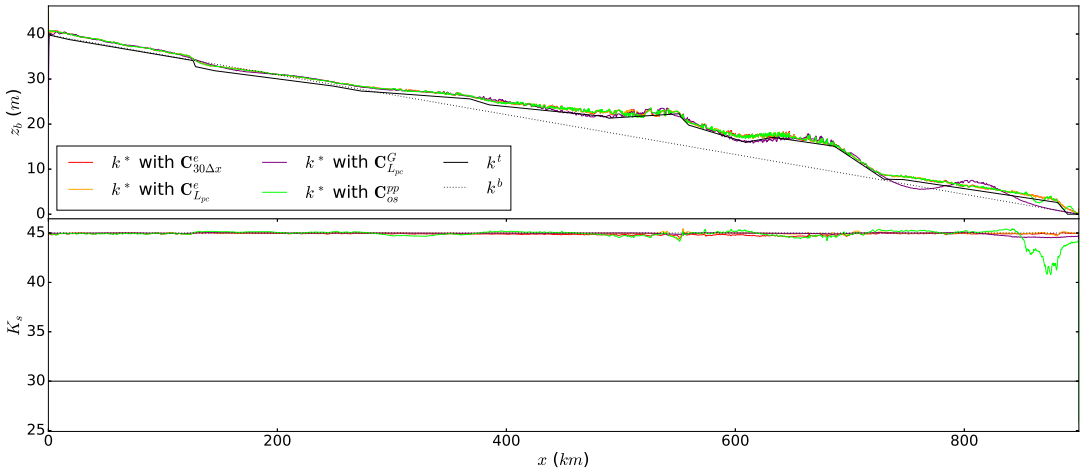


Figure 10: Control $k(x) = (z_b, K_s)(x)$ (target ^{t}, background ^{b} and optimal *) of the partially-observed experiment computed with the variances $(\sigma_{z_b}, \sigma_{K_s}) = (0.5\overline{h}, 10)$, using the reference $\mathbf{C}^e_{30\Delta x}$ ((8) with $L_c = 30\Delta x$, in red), the physically-derived $\mathbf{C}^e_{L_{pc}}$, $\mathbf{C}^G_{L_{pc}}$ (resp. (8) in orange and (7) in purple with $L_c = L_{pc}$ (29)) and the pseudo-physical $\mathbf{C}^{pp}_{os}$ (31) (in lime green) as estimates of $\mathbf{C}_{z_b}$, with no regularization term ($\alpha_{reg} = 0$).

in this case. The optimization compensates the discrepancy between K_s^* and K_s^t by this translation on z_b^*. Note that $\mathbf{C}^G_{L_{pc}}$ still does not allow the inference to reach the $RMSE$ $H \approx 10cm$ stage.

5.1.2 Using the variances $(\sigma_{z_b}, \sigma_{K_s}) = (0.1\overline{h}, 50)$

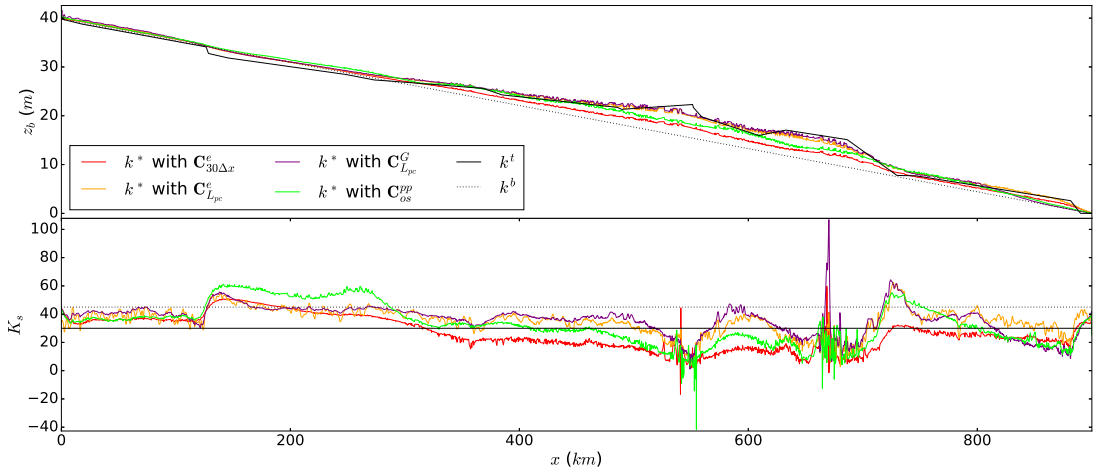


Figure 11: Control $k(x) = (z_b, K_s)(x)$ (target ^{t}, background ^{b} and optimal *) of the partially-observed experiment computed with the variances $(\sigma_{z_b}, \sigma_{K_s}) = (0.1\overline{h}, 50)$, using the reference $\mathbf{C}^e_{30\Delta x}$ ((8) with $L_c = 30\Delta x$, in red), the physically-derived $\mathbf{C}^e_{L_{pc}}$, $\mathbf{C}^G_{L_{pc}}$ (resp. (8) in orange and (7) in purple with $L_c = L_{pc}$ (29)) and the pseudo-physical $\mathbf{C}^{pp}_{os}$ (31) (in lime green) as estimates of $\mathbf{C}_{z_b}$, with no regularization term ($\alpha_{reg} = 0$).

On Fig. 11, for each covariance operator, the z_b^* curve is a combination of the z_b^t curve (e.g. on $x \in [400; 700]km$) and the constant slope z_b^b curve.
On the other hand, K_s^* varies a lot. These variations compensate the discrepancy between z_b^* and z_b^t.
In this case the operator $\mathbf{C}_{L_{pc}}^G$ provides good results.

With $\mathbf{C}_{30\Delta x}^e$ and $\mathbf{C}_{os}^{pp}$, K_s^* is locally strictly negative, despite the fact that, physically, $K_s(x)$ must be strictly positive (for all x). However these negative values are very local and can be easily avoided by adding a regularization term such as $j_{reg}(K_s) = \|\partial_x K_s\|_2^2$.
In order to study only the effect of the estimate of $\mathbf{C}$, such additional regularization term is not used here.

5.2 Comparisons

		Covariance operators	Reference a priori	Physically-derived	
			$\mathbf{C}_{30\Delta x}^e$	$\mathbf{C}_{L_{pc}}^e$	$\mathbf{C}_{os}^{pp}$
$\sigma_{z_b} = 0.5\overline{h}$ $\sigma_{K_s} = 10$	At $RMSE\ H \approx 10cm$	n_{ite}	11	7	18
		$RMSE\ z_b\ (m)$	0.97	0.9	0.88
	till complete convergence	n_{ite}	57	43	48
		$RMSE\ z_b\ (m)$	0.86	0.87	0.87
$\sigma_{z_b} = 0.1\overline{h}$ $\sigma_{K_s} = 50$	At $RMSE\ H \approx 10cm$	n_{ite}	61	25	67
		$RMSE\ z_b\ (m)$	1.67	0.82	1.32
	till complete convergence	n_{ite}	173	102	151
		$RMSE\ z_b\ (m)$	1.67	0.75	1.32

Table 6: Performance (number of iteration n_{ite} and $RMSE\ z_b$) of the inference of (z_b, K_s) using different covariance operator (classical/reference and physically-derived) at different stage of the optimization process (till $RMSE\ H \approx 10cm$ and till complete convergence as described in Sec. 2.3.1).

With the variances $(\sigma_{z_b}, \sigma_{K_s}) = (0.5\overline{h}, 10)$

Overall, in this case, the comparison of the covariance operators is similar to the comparison in the fully-observed experiment, see Sec. 4.
At the $RMSE\ H \approx 10cm$ stage, $RMSE\ z_b$ is slightly better using $\mathbf{C}_{L_{pc}}^e$ and $\mathbf{C}_{os}^{pp}$ (resp. 0.9 and $0.87m$) than with $\mathbf{C}_{30\Delta x}^e$ ($0.97m$, see Tab. 6).
In terms of convergence speed, $\mathbf{C}_{L_{pc}}^e$ is better especially during the first iterations but the result remains similar to the results with $\mathbf{C}_{30\Delta x}^e$ and with $\mathbf{C}_{os}^{pp}$.

On Fig. 12, the inference of z_b with $\mathbf{C}_{30\Delta x}^e$ follows three different steps. At the 2^{nd} iteration, the resulting z_b is close to z_b^t. However, z_b^t, and especially its variations, is not well-approximated everywhere by z_b. For example, z_b^t is well-approximated by z_b on $x \in [400; 550]km$ but not on $x \in [600; 660]km$. Then, z_b is slowly shifted upward until the 6^{th} iteration. After this 6^{th} iterations, the variations of z_b evolves slowly toward the variations of z_b^t, especially where the variations of z_b^t was not well-approximated by z_b during the first iterations.
On the other hand, with $\mathbf{C}_{os}^{pp}$ and especially with $\mathbf{C}_{L_{pc}}^e$, z_b approximates well z_b^t, and its variations, more quickly and everywhere at once. Moreover, with $\mathbf{C}_{os}^{pp}$ and $\mathbf{C}_{L_{pc}}^e$, $RMSE\ z_b$ hardly decreases between the $RMSE\ H \approx 10cm$ stage and the complete convergence, unlike with $\mathbf{C}_{30\Delta x}^e$.

Note that with $\mathbf{C}^e_{30\Delta x}$ and $\mathbf{C}^e_{L_{pc}}$, on Fig. 12, after resp. the 2^{nd} and 3^{rd} iteration, the $RMSE$ z_b increases a bit. Recall from Sec. 2.3.1 that the optimization algorithm starts from z_b^b. Thus, at some point in the optimization iterations, z_b becomes quite close to z_b^t. Then, z_b is translated upward to reach z_b^*.

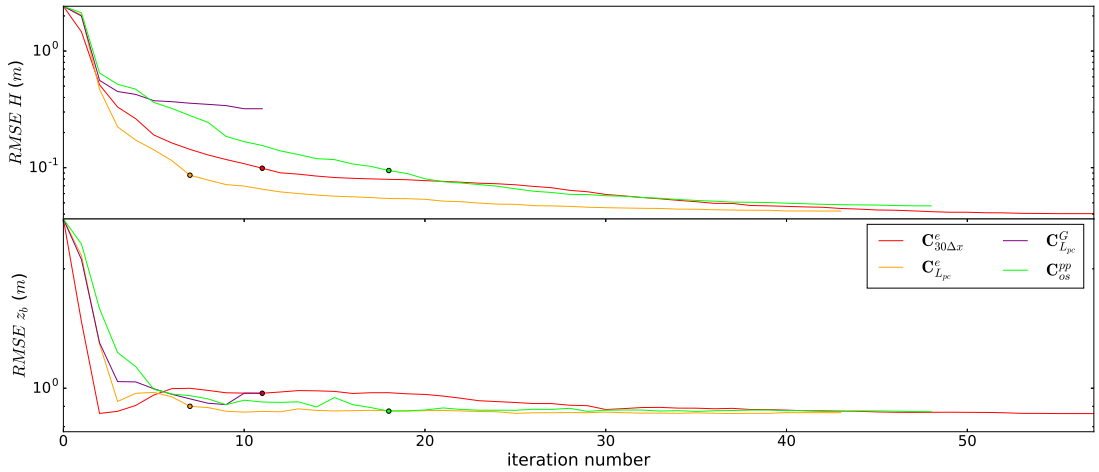


Figure 12: $RMSE$ z_b and $RMSE$ H (log scale) during the partially observed experiment with the variances $(\sigma_{z_b}, \sigma_{K_s}) = (0.5\overline{h}, 10)$ using the reference exponential covariance operator $\mathbf{C}^e_{30\Delta x}$ (8) (in red), the physically-derived exponential $\mathbf{C}^e_{L_{pc}}$(8) (in orange), Gaussian $\mathbf{C}^G_{L_{pc}}$ (7) (in purple, with L_{pc} (29)) and the pseudo-physical $\mathbf{C}^{pp}_{os}$ covariance operators (31) (in lime green) as estimates of the covariance operator $\mathbf{C}_{z_b}$ vs iteration number of the $(z_b, K)(x)$ inference. Dots corresponds to $RMSE$ $H \approx 10cm$.

With the variances $(\sigma_{z_b}, \sigma_{K_s}) = (0.1\overline{h}, 50)$

The results in this case are different from the previous cases. The inference of z_b with $\mathbf{C}^e_{L_{pc}}$ provides the most accurate z_b^* ($RMSE$ $z_b \approx 0.82m$ at complete convergence). Moreover, the inference provides more accurate results with $\mathbf{C}^{pp}_{os}$ than with $\mathbf{C}^e_{30\Delta x}$ (resp. $RMSE$ $z_b \approx 1.31m$ and $RMSE$ $z_b \approx 1.67m$).
The inference with $\mathbf{C}^e_{L_{pc}}$ converges much faster (resp. 25 et 102 iterations to reach the $RMSE$ $H \approx 10cm$ stage and the complete convergence) than with $\mathbf{C}^{pp}_{os}$ (resp. 67 and 151 iterations) or with $\mathbf{C}^e_{30\Delta x}$ (resp. 61 and 173 iterations). In this case, $\mathbf{C}^G_{L_{pc}}$ provides results close to those computed with $\mathbf{C}^e_{L_{pc}}$ (just a little faster to reach complete convergence).
Between the 6^{th} and 8^{th} iterations with $\mathbf{C}^e_{L_{pc}}$, the $RMSE$ z_b decreases a lot, from $RMSE$ $z_b \approx 1.84m$ to $RMSE$ $z_b \approx 0.91m$, see Fig. 13. Moreover, between these iterations, the norm of the gradient increases a lot (by a factor 10^2) and then decreases very rapidly (by a factor 10^4 between the 8^{th} and 15^{th} iterations).
Similarly, during the first iterations with $\mathbf{C}^{pp}_{os}$, the $RMSE$ z_b decreases rapidly (compared to with $\mathbf{C}^e_{30\Delta x}$).

This investigation of the physically-derived covariance operators through the partially-observed experiments can be summarized as follows.

- The operators $\mathbf{C}^{pp}_{os}$ and $\mathbf{C}^e_{L_{pc}}$ promote better descent directions than the reference operator $\mathbf{C}^e_{30\Delta x}$, especially during the first iterations of the optimization algorithm. With $(\sigma_{z_b}, \sigma_{K_s}) = (0.1\overline{h}, 50)$ and $\mathbf{C}^e_{L_{pc}}$, the cost function is such that the descent directions and the line search allow for a change in local minima during the optimization iterations. Thus, in this case, the optimization finds a local minima closer to z_b^t more quickly.

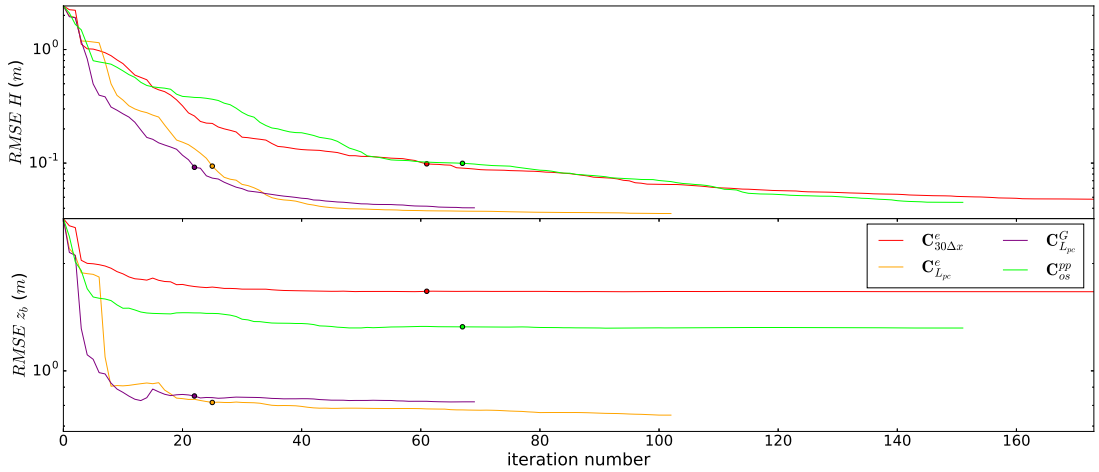


Figure 13: $RMSE$ z_b and $RMSE$ H (log scale) during the partially observed experiment with the variances $(\sigma_{z_b}, \sigma_{K_s}) = (0.1\overline{h}, 50)$ using the reference exponential covariance operator $\mathbf{C}^e_{30\Delta x}$ (8) (in red), the physically-derived exponential $\mathbf{C}^e_{L_{pc}}$ (8) (in orange), Gaussian $\mathbf{C}^G_{L_{pc}}$ (7) (in purple, with L_{pc} (29)) and the pseudo-physical $\mathbf{C}^{pp}_{os}$ covariance operators (31) (in lime green) as estimates of the covariance operator $\mathbf{C}_{z_b}$ vs iteration number of the $(z_b, K)(x)$ inference. Dots corresponds to $RMSE$ $H \approx 10cm$.

- The conclusions of the fully-observed experiment are confirmed: the pseudo-physical operator $\mathbf{C}^{pp}_{os}$ is a good estimate and the physically-derived $\mathbf{C}^e_{L_{pc}}$ provides the best results. The Gaussian operator $\mathbf{C}^G_{L_{pc}}$ remains a poor estimate of $\mathbf{C}$ even if it provides good results in one case.

- The physically-derived operators $\mathbf{C}^{pp}_{os}$ and $\mathbf{C}^e_{L_{pc}}$ seem to improve the convexity feature of the cost function. As a result, the optimization algorithm converges to a more accurate z_b^*, especially when the second control, here K_s, is assumed highly uncertain.

6 Conclusion

This paper investigates, for Data Assimilation purposes, a method for deriving estimates of the background error covariance operator $\mathbf{C}$, see Sec. 2.1.2, from governing physical equations. The physically-derived kernels are built up from Green-like kernels representing the system response to a perturbation of the control, see (10).
In the literature, $\mathbf{C}$ is classically modeled from empirical or arbitrary information. The present investigation compares these classical empirical operators with the physically-derived ones.

The method is investigated for the inference of the bathymetry z_b in the Saint-Venant equations (1) in a satellite observation framework. In this context, the resulting Green-like kernels (23) are neither symmetric (due to the advection phenomena) nor positive-definite. The present method constructs, from the Green-like kernels, discrete covariance operators partly consistent with the underlying physics. These operators are obtained by applying (i) a locally symmetric approximation and (ii) the positive-definite approximation (30).
This provides an original covariance operator $\mathbf{C}^{pp}$, see (31), called pseudo-physical covariance operator. Moreover, a physically-consistent estimate of the correlation length L_{pc}, defined by (29), is also provided. Thus, the correlation length value $L_c = L_{pc}$ is tested with covariance operators such as the Gaussian $\mathbf{C}^G_{L_c}$ and the decreasing exponential

$\mathbf{C}^e_{L_c}$ operators.
The operator $\mathbf{C}^e_{30\Delta x}$ is a priori the empirical covariance operator that provide the best numerical results and features. Thus, it is chosen as the reference operator.

During all the numerical experiments, the operator $\mathbf{C}^e_{L_{pc}}$ provides better results than the $\mathbf{C}^e_{30\Delta x}$.
The operator $\mathbf{C}^{pp}$ provides results at least comparable, especially during the fully-observed experiment, to $\mathbf{C}^e_{30\Delta x}$: the convergence speeds and the accuracies are similar.
Here, the classical operator $\mathbf{C}^G_{L_c}$ has no physical justification and the associated numerical experiments provide poor results. Hence, it is a bad estimate of $\mathbf{C}$.
From these numerical experiments and especially from the partially-observed experiment, we can observe that the physically-derived operators promote better descent directions and converge faster and with better accuracy, especially during the first iterations.
This highlights that the physically-derived covariance operators are better preconditioning of the Hessian of the cost function therefore improving convexity features of the cost function.
Therefore, the method for deriving physically-consistent covariance operators provides good estimates of $\mathbf{C}$ such as $\mathbf{C}^{pp}$. Furthermore, it provides a physically-consistent L_{pc} which is more efficient than empirical correlation lengths when combined with $\mathbf{C}^e_{L_c}$.

The presented method is available in the open-source computational software Dassflow [Monnier and Larnier, 2018]. This method has been successfully implemented to the 2D version of the Shallow-water river flow models too, see e.g. Monnier et al. [2016].

Acknowledgments

This work is part of the PhD thesis work of TM (CIFRE PhD 2018/0725, funded by ANRT).

References

R. N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970, a. doi: 10.1002/qj.339.

R. N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. ii: Modelling the forecast error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134 (637):1971–1996, b. doi: https://doi.org/10.1002/qj.340.

F. Bouttier and P. Courtier. Data assimilation concepts and methods march 1999. *Meteorological training course lecture series. ECMWF*, page 59, 2002. URL http://msi.ttu.ee/~elken/Assim_concepts.pdf.

H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer, 2010.

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

S. Cheng, J.-P. Argaud, B. Iooss, D. Lucor, and A. Ponçot. Error covariance tuning in variational data assimilation: application to an operating hydrological model. *Stochastic Environmental Research and Risk Assessment*, pages 1–20, 2020. doi: https://doi.org/10.1007/s00477-020-01933-7.

V. T. Chow. Handbook of applied hydrology. *McGraw-Hill Book Co., New-York, 1467 pages*, 1964.

W. Collischonn, D. Allasia, B. C. Da Silva, and C. E. M. Tucci. The mgb-iph model for large-scale rainfall—runoff modelling. *Hydrological Sciences Journal*, 52(5):878–895, 2007. doi: https://doi.org/10.1623/hysj.52.5.878.

J. Cunge. Practical aspects of computational river hydraulics. *Pitman Publishing Ltd. London,(17 CUN), 1980, 420*, 1980.

R. Daley. *Atmospheric data analysis*. Number 2. Cambridge university press, 1993. URL https://www.cambridge.org/fr/academic/subjects/earth-and-environmental-science/atmospheric-science-and-meteorology/atmospheric-data-analysis?format=PB&isbn=9780521458252.

G. D. Egbert, A. F. Bennett, and M. G. G. Foreman. Topex/poseidon tides estimated using a global inverse model. *Journal of Geophysical Research: Oceans*, 99(C12):24821–24852, 1994. doi: https://doi.org/10.1029/94JC01894.

L. C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998. ISBN 9780821807729.

J.-C. Gilbert and C. Lemaréchal. Some numerical experiments with variable storage quasi-newton algorithms. *Mathematical Programming*, 45:407–435, 08 1989. doi: https://doi.org/10.1007/BF01589113.

G. Green. An essay on the application of mathematical analysis to the theories of electricity and magnetism, 1828.

S. A. Haben. *Conditioning and preconditioning of the minimisation problem in variational data assimilation*. PhD thesis, University of Reading, 2011. URL http://www.reading.ac.uk/web/files/maths/HabenThesis.pdf.

S. A. Haben, A. S. Lawless, and N. K. Nichols. Conditioning and preconditioning of the variational data assimilation problem. *Computers and Fluids*, 46(1):252–256, 2011. ISSN 0045-7930. doi: https://doi.org/10.1016/j.compfluid.2010.11.025. 10th ICFD Conference Series on Numerical Methods for Fluid Dynamics (ICFD 2010).

L. Hascoet and V. Pascual. The tapenade automatic differentiation tool: Principles, model, and specification. 39 (3), 2013. ISSN 0098-3500. doi: https://doi.org/10.1145/2450153.2450158.

N. J. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002. doi: https://doi.org/10.1093/imanum/22.3.329.

A. Hollingsworth and P. Lönnberg. The statistical structure of short-range forecast errors as determined from radiosonde data. part i: The wind field. *Tellus A*, 38A(2):111–136, 1986. doi: https://doi.org/10.1111/j.1600-0870.1986.tb00460.x.

M. Honnorat, J. Monnier, and F.-X. Le Dimet. Lagrangian data assimilation for river hydraulics simulations. *Computing and visualization in science*, 12(5):235–246, 2009. doi: https://doi.org/10.1007/s00791-008-0089-x.

N. B. Ingleby. The statistical structure of forecast errors and its representation in the met. office global 3-d variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 127(571):209–231, 2001. doi: https://doi.org/10.1002/qj.49712757112.

B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6. Walter de Gruyter, 2008. doi: https://doi.org/10.1515/9783110208276.

A. C. Koivunen and A. B. Kostinski. The feasibility of data whitening to improve performance of weather radar. *Journal of Applied Meteorology*, 38(6):741–749, 1999. doi: https://doi.org/10.1175/1520-0450(1999)038<0741:TFODWT>2.0.CO;2.

K. Larnier and J. Monnier. Hybrid neural network – variational data assimilation algorithm to infer river discharges from swot-like data. *Nonlinear Processes in Geophysics Discussions*, 2020:1–30, 2020. doi: https://doi.org/10.1016/10.5194/npg-2020-32.

F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A: Dynamic Meteorology and Oceanography*, 38(2):97–110, 1986. doi: https://doi.org/10.3402/tellusa.v38i2.11706.

S. Lopez-Restrepo, E. D. Nino-Ruiz, L. G. Guzman-Reyes, A. Yarce, O. L. Quintero, N. Pinel, A. Segers, and A. W. Heemink. An efficient ensemble kalman filter implementation via shrinkage covariance matrix estimation: exploiting prior knowledge. *Computational Geosciences*, pages 1–19, 2021. doi: https://doi.org/10.1007/s10596-021-10035-4.

P. M. Lyster, S. E. Cohn, B. Zhang, L.-P. Chang, R. Ménard, K. Olson, and R. Renka. A lagrangian trajectory filter for constituent data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 130(601):2315–2334, 2004. doi: https://doi.org/10.1256/qj.02.234.

T. Malou and J. Monnier. Double scale diffusive wave equation dedicated to spatial rivers observations. 2021.

T. Malou, P.-A. Garambois, A. Paris, J. Monnier, and K. Larnier. Generation and analysis of stage-fall-discharge laws from coupled hydrological-hydraulic river network model integrating sparse multi-satellite data. *Journal of Hydrology*, 2021.

I. Mirouze and A. Weaver. Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1421–1443, 2010. doi: https://doi.org/10.1002/qj.643.

J. Monnier and K. et al. Larnier. Dassflow: Data assimilation for free surface flows. open-source computational software. Technical report, Mathematics Institute of Toulouse - INSA - CNES - CNRS, 2018. URL http://www.math.univ-toulouse.fr/DassFlow.

J. Monnier and J. Zhu. Inference of the bottom topography in anisothermal mildly-sheared shallow ice flows. *Computer Methods in Applied Mechanics and Engineering*, 348:954–977, 2019. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2019.01.003.

J. Monnier, F. Couderc, D. Dartus, K. Larnier, R. Madec, and J-P. Vila. Inverse algorithms for 2D shallow water equations in presence of wet dry fronts. application to flood plain dynamics. *Advances in Water Resources*, 97: 11–24, 2016. doi: https://doi.org/10.1016/j.advwatres.2016.07.005.

T. Nagai. The covariance matrix of green's functions and its application to machine learning, 2020. URL https://arxiv.org/abs/2004.06481.

J. Nolen and G. Papanicolaou. Fine scale uncertainty in parameter estimation for elliptic equations. *Inverse Problems*, 25(11):115021, nov 2009. doi: 10.1088/0266-5611/25/11/115021. URL https://doi.org/10.1088/0266-5611/25/11/115021.

O. Pannekoucke, L. Berre, and G. Desroziers. Background-error correlation length-scale estimates and their sampling statistics. *Quarterly Journal of the Royal Meteorological Society*, 134(631):497–508, 2008. doi: https://doi.org/10.1002/qj.212. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.212.

O. Pannekoucke, S. Ricci, S. Barthelemy, R. Ménard, and O. Thual. Parametric kalman filter for chemical transport models. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1):31547, 2016. doi: https://doi.org/10.3402/tellusa.v68.31547.

O. Pannekoucke, M. Bocquet, and R. Ménard. Parametric covariance dynamics for the nonlinear diffusive burgers equation. *Nonlinear Processes in Geophysics*, 25(3):481–495, 2018. doi: https://doi.org/10.5194/npg-25-481-2018.

D. F. Parrish and J. C. Derber. The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8):1747 – 1763, 1992. doi: https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.

L. Pujol, P.-A. Garambois, P. Finaud-Guyot, J. Monnier, K. Larnier, R. Mosé, S. Biancamaria, H. Yésou, D. Moreira, A. Paris, and S. Calmant. Estimation of multiple inflows and effective channel by assimilation of multi-satellite hydraulic signatures: The ungauged anabranching negro river. *Journal of Hydrology*, 2020. doi: https://doi.org/10.1016/j.jhydrol.2020.125331.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Y. Sasaki. An objective analysis based on the variational method. *Journal of the Meteorological Society of Japan. Ser. II*, 36(3):77–88, 1958. doi: https://doi.org/10.2151/jmsj1923.36.3_77.

A. Tarantola. *Inverse problem theory and methods for model parameter estimation*, volume 89. SIAM, 2005. URL http://www.ipgp.fr/~tarantola/Files/Professional/Books/InverseProblemTheory.pdf.

X. Wang, D. M. Barker, C. Snyder, and T. M. Hamill. A hybrid etkf–3dvar data assimilation scheme for the wrf model. part i: Observing system simulation experiment. *Monthly Weather Review*, 136(12):5116–5131, 2008. doi: https://doi.org/10.1175/2008MWR2444.1.

A. Weaver and P. Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, 127(575):1815–1846, 2001.