

Second-Order Methods in Variational Data Assimilation



François-Xavier Le Dimet and Victor Shutyaev

Abstract The chapter presents the second-order methods in variational data assimilation. The algorithms to compute the Hessian of the cost function are discussed, the second-order adjoint method among them. General sensitivity analysis for the optimality system is presented. Using the Hessian, the sensitivity of the optimal solution and its functionals is studied with respect to observations and uncertainties in model parameters. Numerical examples for joint state and parameter estimation for a sea thermodynamics model are presented.

1 Introduction

The methods of data assimilation (DA) have become an important tool for analysis of complex physical phenomena in various fields of science and technology. These methods allow us to combine mathematical models, data from observations and a priori information.

Currently, there is an increasing interest in computational technologies that combine the flows of real data and hydrodynamic forecasts using mathematical models. This is especially true for 4D technologies - the combination of the flows of observational data and forecasts in a certain spatio-temporal domain. These methods have received the greatest applications in meteorology and oceanography, where observations are assimilated into numerical models. Geophysical flows are governed by equations derived from fluid dynamics: a set of nonlinear partial differential equations of the first order with respect to time. Formally, it is a Cauchy problem, and an initial condition is necessary to integrate these equations, to carry out a prediction. The purpose of assimilation procedures is to construct or refine the initial and bound-

F.-X. Le Dimet

LJK, Université Grenoble Alpes, 700 Avenue Centrale, 38401 Domaine Universitaire de Saint-Martin-d'Hères, Grenoble, France

V. Shutyaev (✉)

Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, 119333 Gubkina 8, Moscow, Russia

e-mail: victor.shutyaev@mail.ru

ary conditions (or other model parameters) to improve the accuracy of a prediction model Le Dimet and Talagrand (1986), Asch et al. (2016), Fletcher (2017), Carrassi et al. (2018).

At present, two main approaches are well known for the assimilation of observational data in models of geophysical hydrodynamics and oceanography. The first is the Statistical approach which is based on the methods of probability theory and mathematical statistics. Historically, its rigorous justification and limits of applicability were given by Markov (1900) and Kolmogorov (1946). From a methodological point of view, this method gave rise to the methods of optimal interpolation, the Kalman filter methods and their subsequent modifications, widely used in various fields of science and technology. This approach is used to estimate unknown quantities from measurement data, taking into account the random nature of measurement errors.

The second approach is based on the methods of calculus of variations, optimal control (see, e.g., Lions (1968), Pontryagin et al. (1964)) and the theory of adjoint equations (see Marchuk (1995)). Compared to the statistical method, the variational method has greater versatility. It allows, on a unified methodological basis, to solve the problems of initializing hydrophysical fields, assessing the sensitivity of a model solution, identifying model parameters, etc. The variational approach can be applied by assimilating information of various types and measuring systems. In this case this approach is referred to as variational data assimilation (VDA) Le Dimet and Talagrand (1986), Asch et al. (2016), Fletcher (2017), Carrassi et al. (2018). The main idea of the method is to minimize some functional that describes the deviation of the model solution from the observational data, and the minimum of this functional is sought on the model trajectories, in other words, in the subspace of model solutions.

Basically, as seen as a problem of optimal control, VDA is an optimization problem and as such we need to exhibit a necessary optimality condition derived from the evaluation of the gradient of the cost function, which should be zero at the optimum. Information on the gradient of the cost function (first-order information) is used to construct the optimality system (OS). To this aim and for the numerical solution of the optimization problem, the representation of the gradient through adjoint equations (first-order adjoint problem) is often used Le Dimet and Talagrand (1986), Marchuk (1995). In the case of discontinuous processes in the physics (rain, deep convection, etc.) the cost function is no longer differentiable and the formal application of the adjoint operator will evaluate a sub-gradient.

To study the variational data assimilation problem (as an optimal control problem) and to develop efficient algorithms for its numerical solution, second-order information is needed. This is information about the Hessian of the cost function. A necessary and sufficient optimality condition is to get the Hessian positive definite at the optimum; therefore, a second-order analysis must be carried out. Often, to construct the Hessian, it is necessary to differentiate the optimality system. In this case, a second-order adjoint problem arises Le Dimet et al. (2002). The investigation of the second-order adjoint equations and the Hessian of the cost functional plays an important role in the study of the solvability of the variational assimilation problem, the construction of algorithms for its numerical solution based on the modification

of Newton type methods, the identification of model parameters, and the study of the sensitivity of the optimal solution and its functionals. These issues are the subject of this chapter.

2 Variational Data Assimilation

Variational methods were introduced in meteorology in 1958 by Sasaki (1958). These methods consider the equations governing the flow as constraints and the problem is closed by using a variational principle, e.g. the minimization of the discrepancy between the model and the observations. Using Optimal Control Techniques (Lions (1968)) was proposed by Le Dimet (1982), Le Dimet and Talagrand (1986), Talagrand and Courtier (1987), Penenko and Obraztsov (1976), Marchuk et al. (1978).

Consider the mathematical model of a physical process that is described by the nonlinear evolution problem

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi) + f, & t \in (0, T) \\ \varphi|_{t=0} = u, \end{cases} \quad (1)$$

where the initial state u is supposed to be from a Hilbert space X , the unknown function $\varphi = \varphi(t)$ belongs to $Y = L_2(0, T; X)$ with the norm $\|\varphi\|_Y = (\varphi, \varphi)_Y^{1/2} = (\int_0^T \|\varphi(t)\|_X^2 dt)^{1/2}$, F is a nonlinear operator mapping Y into Y , $f \in Y$. We suppose that for given $u \in X$, $f \in Y$ there exists a unique solution $\varphi \in Y$ to (1) with $\frac{\partial \varphi}{\partial t} \in Y$.

Often, the the initial state u is supposed to be unknown, and one would like to find it using the information from observations. Let us introduce the cost function as a functional on X in the form

$$J(u) = \frac{1}{2}(V_1(u - u_b), u - u_b)_X + \frac{1}{2}(V_2(C\varphi - \varphi_{obs}), C\varphi - \varphi_{obs})_{Y_{obs}}, \quad (2)$$

where $u_b \in X$ is a prior (background) function, $\varphi_{obs} \in Y_{obs}$ is a prescribed function (observational data), Y_{obs} is a Hilbert space (observation space), $C : Y \rightarrow Y_{obs}$ is a linear bounded operator (observation operator), $V_1 : X \rightarrow X$ and $V_2 : Y_{obs} \rightarrow Y_{obs}$ are symmetric positive definite bounded operators. Usually, V_1, V_2 are chosen as inverse covariancve operators of background and observation errors, respectively, Asch et al. (2016), Carrassi et al. (2018).

Let us consider the following data assimilation problem with the aim to find the initial value u : for given $f \in Y$, $\varphi_{obs} \in Y_{obs}$, $u_b \in X$, find $u \in X$ and $\varphi \in Y$ such that they satisfy (1), and on the set of solutions to (1), the functional $J(u)$ takes the minimum value, i.e.

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi) + f, & t \in (0, T) \\ \varphi|_{t=0} = u, \\ J(u) = \inf_{w \in X} J(w). \end{cases} \quad (3)$$

This is a so-called hind-cast (initialization) variational DA problem, a typical DA problem often considered in numerical weather prediction and oceanographic applications Le Dimet and Talagrand (1986), Asch et al. (2016), Fletcher (2017), Carrassi et al. (2018). We suppose that the solution of (3) exists. To derive the optimality system, we assume the solution φ and the operator $F(\varphi)$ in (1)–(2) are regular enough, and for $u, w \in X$ introduce the directional (Gâteaux) derivative with respect to u in the direction w (Gâteaux differential):

$$dJ(u, w) = \lim_{\tau \rightarrow 0} \frac{J(u + \tau w) - J(u)}{\tau} = \frac{d}{d\tau} J(u + \tau w) \Big|_{\tau=0}.$$

If $dJ(u, w)$ is linear with respect to w , then it may be represented as follows:

$$dJ(u, w) = J'(u)w,$$

where $J'(u)$ is the gradient of J with respect to u . From (1)–(2) we get

$$dJ(u, w) = (V_1(u - u_b), w)_X + (C^*V_2(C\varphi - \varphi_{obs}), \tilde{\phi})_Y, \quad (4)$$

where $\tilde{\phi}$ is the solution to the tangent linear problem:

$$\begin{cases} \frac{\partial \tilde{\phi}}{\partial t} = F'_\varphi(\varphi)\tilde{\phi}, & t \in (0, T), \\ \tilde{\phi}|_{t=0} = w. \end{cases} \quad (5)$$

Here $F'_\varphi(\varphi) : Y \rightarrow Y$ is the Fréchet derivative of F Marchuk et al. (1996) with respect to φ , and C^* is the adjoint operator to C defined by $(C\varphi, \psi)_{Y_{obs}} = (\varphi, C^*\psi)_Y$, $\varphi \in Y$, $\psi \in Y_{obs}$.

Let us introduce the adjoint operator $(F'_\varphi(\varphi))^* : Y \rightarrow Y$ and consider the adjoint problem:

$$\begin{cases} \frac{\partial \varphi^*}{\partial t} + (F'_\varphi(\varphi))^* \varphi^* = C^*V_3(C\varphi - \varphi_{obs}), & t \in (0, T) \\ \varphi^*|_{t=T} = 0. \end{cases} \quad (6)$$

The problem (6) is adjoint with respect to the linearized (tangent linear) problem (5), therefore, it is also linear in φ^* , however, it is still nonlinear in φ .

In the below consideration, we assume that the direct and adjoint linear problems of the form

$$\begin{cases} \frac{\partial \phi}{\partial t} - F'_\varphi(\varphi)\phi = p, & t \in (0, T) \\ \phi|_{t=0} = q, \end{cases}$$

$$\begin{cases} -\frac{\partial \phi^*}{\partial t} - (F'_\varphi(\varphi))^* \phi^* = g, & t \in (0, T) \\ \phi^*|_{t=T} = 0 \end{cases}$$

with $p, g \in Y, q \in X$ have the unique solutions $\phi, \phi^* \in Y$ and $\frac{\partial \phi}{\partial t}, \frac{\partial \phi^*}{\partial t} \in Y$. From (4)–(6) we get

$$dJ(u, w) = (V_1(u - u_b), w)_X - (\varphi^*|_{t=0}, w)_X. \tag{7}$$

The relation (7) exhibits the linear dependence of $dJ(u, w)$ with respect to w . Thus, $dJ(u, w) = J'(u)w$, and the gradient of J with respect to u is defined by

$$J'(u) = V_1(u - u_b) - \varphi^*|_{t=0}.$$

The necessary optimality condition Lions (1968) is $J'(u) = 0$. From (3)–(7) we obtain the Optimality System :

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi) + f, & t \in (0, T), \\ \varphi|_{t=0} = u, \end{cases} \tag{8}$$

$$\begin{cases} \frac{\partial \varphi^*}{\partial t} + (F'_\varphi(\varphi))^* \varphi^* = C^* V_3(C\varphi - \varphi_{obs}), & t \in (0, T) \\ \varphi^*|_{t=T} = 0, \end{cases} \tag{9}$$

$$V_1(u - u_b) - \varphi^*|_{t=0} = 0. \tag{10}$$

It worth to point out that there is no approximation in the derivation of the optimality system and the only assumption is the differentiability of the operator of the model. Some authors consider, at this level, a so-called "linear tangent approximation", it is fully unnecessary.

We suppose that the system (8)–(10) has a unique solution $\varphi, \varphi^* \in Y, u \in X$. The system (8)–(10) may be considered as a generalized model of the form $\mathcal{A}(U) = 0$ with the state variable $U = (\varphi, \varphi^*, u)$, and it contains the information on the observation data $\varphi_{obs} \in Y_{obs}$. The optimality system plays a fundamental role in studying the solvability of the original data assimilation problem, searching efficient algorithms for its solution, and studying the sensitivity of the optimal solution with respect to observations.

3 Computing the Hessian

Consider the Hessian $\mathcal{H}(u)$ of the functional (2); it depends on $u \in X$ (which may be the exact solution, the optimal solution, or some arbitrary function $u \in X$). For a fixed $u \in X$ the Hessian $\mathcal{H}(u)$ is defined by the successive solutions of the below-

formulated problems. First we find φ and φ^* by solving the direct and adjoint problems (like in the optimality system):

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi) + f, & t \in (0, T) \\ \varphi|_{t=0} = u, \end{cases} \quad (11)$$

$$\begin{cases} -\frac{\partial \varphi^*}{\partial t} - (F'(\varphi))^* \varphi^* = -C^* V_2 (C\varphi - \varphi_{obs}), & t \in (0, T) \\ \varphi^*|_{t=T} = 0. \end{cases} \quad (12)$$

Note that here u is not necessarily the optimal solution from the optimality system (8)–(10); it is just some fixed function at which we would like to compute the Hessian. (Hence, in general, the functions φ and φ^* do not satisfy the optimality system). Note also that (11)–(12) are usual two steps when we compute the gradient of the functional $J(u)$ (at the point u) using the adjoint problem. If for a fixed u the functions φ , φ^* are computed from (11)–(12), the gradient of J with respect to u is defined by

$$J'(u) = V_1(u - u_b) - \varphi^*|_{t=0}. \quad (13)$$

To find the Hessian we should differentiate (11)–(13) with respect to u . Then, the action of the Hessian $\mathcal{H}(u)$ on the function $v \in X$ is defined by the successive solutions of the following problems:

$$\begin{cases} \frac{\partial \psi}{\partial t} - F'(\varphi)\psi = 0, & t \in (0, T), \\ \psi|_{t=0} = v, \end{cases} \quad (14)$$

$$\begin{cases} -\frac{\partial \psi^*}{\partial t} - (F'(\varphi))^* \psi^* = (F''(\varphi)\psi)^* \varphi^* - C^* V_2 C \psi, & t \in (0, T) \\ \psi^*|_{t=T} = 0, \end{cases} \quad (15)$$

$$\mathcal{H}(u)v = V_1 v - \psi^*|_{t=0}. \quad (16)$$

Here φ and φ^* are involved, being taken from (11)–(12). The problem (15) is a so-called *second-order adjoint* problem Le Dimet et al. (2002). It involves the second derivative $F''(\varphi)$ of the model operator $F(\varphi)$ and depends on the solution φ^* of the first-order adjoint problem (12).

If u is the optimal solution, then φ and φ^* are exactly the functions from the optimality system (8)–(10).

Formulas (11)–(16) may be used to compute the Hessian of the original cost functional. To solve the second-order adjoint problem (15), no additional software is needed to be developed. To this aim, one can use the existing code for the first-order adjoint problem (12), taking into account the new right-hand side involving the term $(F''(\varphi)\psi)^* \varphi^*$. An alternative method to compute the Hessian \mathcal{H} is the method of finite differences described in Gill et al. (1981). However, this method is not sufficiently accurate due to truncations used in a local Taylor expansion and

is expensive for practical implementation. The sensitivity matrix method Thacker (1989) is computationally efficient if the dimension of the observation vector is much smaller than the dimension of the state vector, and so is feasible mainly for the 3D-VAR applications. It requires full storage of the resulting matrix. The above-formulated second-order adjoint method allows the action $\mathcal{H}v$ to be computed, thus does not require full storage of \mathcal{H} .

In the finite-dimensional space, $\mathcal{H}(u)$ is a matrix. To obtain the first column of this matrix, one can choose v in (14)–(16) to be the first basis vector $v = (1, 0, \dots, 0)$. To obtain the second column of this matrix, one can choose v in (14)–(16) to be the second basis vector $v = (0, 1, 0, \dots, 0)$, and so on.

In the linear case, the solution is unique if the Hessian is positive definite. In this case, the necessary optimality condition given by the optimality system is also a sufficient condition. From a general point of view the information given by the Hessian is important for theoretical, numerical and practical issues. For operational models it is impossible to compute the Hessian itself, as it is a square matrix with around 10^{18} terms, nevertheless the most important information can be extracted from the spectrum of the Hessian which can be estimated without an explicit determination of this matrix. This information is of importance for estimating the condition number of the Hessian for preparing an efficient preconditioning.

The above-obtained system with the second order adjoint is used to compute the product of the Hessian by any vector. Of course, if we consider all the vectors of the canonical base, then it will be possible to get the complete Hessian.

The determination of this product permits to access some information. So, by using Lanczos type methods and deflation, it is possible to compute the eigenvectors and eigenvalues of the Hessian. Also, to solve the variational data assimilation problem, second-order optimization methods of Newton-type are used for equations of the form:

$$J'(u) = 0.$$

The iterations are

$$u_{n+1} = u_n - \mathcal{H}^{-1}(u_n) J'(u_n),$$

where \mathcal{H} is the Hessian of J , or its approximation. At each iteration a linear system should be solved. This is done by carrying out some iterations of a conjugate gradient methods which require computing the Hessian-vector product. To construct an approximation of the inverse Hessian, the quasi-Newton BFGS algorithm may be used Polak (1997). This algorithm generates \mathcal{H}^{-1} in the course of a minimization process.

In some applications (such as sensitivity analysis) one needs to solve the system of equations in the form $\mathcal{H}(u)v = p$. In this case, computing the Hessian-vector product by (11)–(16) may be efficient for using iterative algorithms. The following directions to construct a specialized solver for the equation $\mathcal{H}(u)v = p$ could be considered: the use of a multi-grid strategy; the use of reduced order models (Proper Orthogonal Decomposition) or local approximations (splines, wavelets); decomposition of the

spatial domain by the ‘region of influence’ principle, hence decomposition of a global DA problem into a set of local open boundary DA problems.

The inverse Hessian or its approximations may be used also to estimate the optimal solution error covariances Gejadze et al. (2008, 2011, 2013), Shutyaev et al. (2012). Assuming the so-called tangent linear hypothesis (TLH), the covariance is often approximated by the inverse Hessian of the objective function. In practice, the same approximation could be valid even though the TLH is clearly violated. However, often we deal with such a highly nonlinear dynamics that the inverse Hessian approach is no longer valid. In this case a new method for computing the covariance matrix named the ‘effective inverse Hessian’ method can be used Shutyaev et al. (2012). This method yields a significant improvement of the covariance estimate as compared to the inverse Hessian. The method is potentially feasible for large-scale applications because it can be used in the multiprocessor environment and operates in terms of the Hessian-vector products. The software blocks needed for its implementation are the standard blocks of any existing 4D-Var system. The results given by the method are consistent with the assumption on a ‘close-to-normal’ nature of the optimal solution error. This should be expected taking into account the consistency and asymptotic normality of the estimator and the fact that the observation window in variational DA is usually quite large.

4 Parameter Estimation

We should mention the importance of the parameter estimation problem itself. A precise determination of the initial condition is very important in view of forecasting, however the use of variational data assimilation is not limited to operational forecasting. In many domains (e.g. hydrology) the uncertainty in the parameters is more crucial than the uncertainty in the initial condition (e.g. White et al. (2003)). In some problems the quantity of interest can be represented directly by the estimated parameters as controls. For example, in Agoshkov et al. (2015) the sea surface heat flux is estimated in order to understand its spatial and temporal variability. The problems of parameter estimation are common inverse problems considered in geophysics and in engineering applications (see Alifanov et al. (1996), Sun (1994), Zhu and Navon (1999), Storch et al. (2007)). Last years an interest is rising to the parameter estimation using 4D-Var (Bocquet (2012), Schirber et al. (2013), Smith et al. (2013), Yuepeng et al. (2018), Agoshkov and Sheloput (2017)).

We consider a dynamic formulation of variational data assimilation problem for parameter estimation in a continuous form. Of course, the initial condition function may be also considered as a parameter, however, in our dynamic formulation we have two equations for the model: one equation for describing an evolution of the model operator (involving model parameters such as right-hand sides, coefficients, boundary conditions etc.), and another equation is considered as an initial condition.

Let the model be governed by the evolution problem of the form (1):

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda) + f, & t \in (0, T) \\ \varphi|_{t=0} = u, \end{cases} \quad (17)$$

where F is a nonlinear operator mapping $Y \times Y_p$ into Y , Y_p is a Hilbert space (space of control parameters, or control space). Suppose that for given $u \in X$, $f \in Y$ and $\lambda \in Y_p$ there exists a unique solution $\varphi \in Y$ to (17) with $\frac{\partial \varphi}{\partial t} \in Y$. The function λ is an unknown model parameter.

Let us introduce the cost function

$$J(\lambda) = \frac{1}{2}(V_1(\lambda - \lambda_b), \lambda - \lambda_b)_{Y_p} + \frac{1}{2}(V_2(C\varphi - \varphi_{obs}), C\varphi - \varphi_{obs})_{Y_{obs}}, \quad (18)$$

where $\lambda_b \in Y_p$ is a prior (background) function, $\varphi_{obs} \in Y_{obs}$ is a prescribed function (observational data), Y_{obs} is a Hilbert space (observation space), $C : Y \rightarrow Y_{obs}$ is a linear bounded observation operator, $V_1 : Y_p \rightarrow Y_p$ and $V_2 : Y_{obs} \rightarrow Y_{obs}$ are symmetric positive definite bounded operators.

Let us consider the following data assimilation problem with the aim to estimate the parameter λ : for given $u \in X$, $f \in Y$, find $\lambda \in Y_p$ and $\varphi \in Y$ such that they satisfy (17), and on the set of solutions to (17), the functional $J(\lambda)$ takes the minimum value, i.e.

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda) + f, & t \in (0, T) \\ \varphi|_{t=0} = u, \\ J(\lambda) = \inf_{v \in Y_p} J(v). \end{cases} \quad (19)$$

We suppose that the solution of (19) exists. Let us note that the solvability of the parameter estimation problems (or identifiability) has been addressed, e.g., in Chavent (1983), Navon (1998). To derive the optimality system, we assume the solution φ and the operator $F(\varphi, \lambda)$ in (17)–(18) are regular enough, and for $v \in Y_p$ find the gradient of the functional J with respect to λ :

$$\begin{aligned} J'(\lambda)v &= (V_1(\lambda - \lambda_b), v)_{Y_p} + (V_2(C\varphi - \varphi_{obs}), C\phi)_{Y_{obs}} \\ &= (V_1(\lambda - \lambda_b), v)_{Y_p} + (C^*V_2(C\varphi - \varphi_{obs}), \phi)_Y, \end{aligned} \quad (20)$$

where ϕ is the solution to the problem:

$$\begin{cases} \frac{\partial \phi}{\partial t} = F'_\varphi(\varphi, \lambda)\phi + F'_\lambda(\varphi, \lambda)v, \\ \phi|_{t=0} = 0. \end{cases} \quad (21)$$

Here $F'_\varphi(\varphi, \lambda) : Y \rightarrow Y$, $F'_\lambda(\varphi, \lambda) : Y_p \rightarrow Y$ are the Fréchet derivatives of F Marchuk et al. (1996) with respect to φ and λ , correspondingly, and C^* is the adjoint operator to C defined by $(C\varphi, \psi)_{Y_{obs}} = (\varphi, C^*\psi)_Y$, $\varphi \in Y$, $\psi \in Y_{obs}$.

Let us consider the adjoint operator $(F'_\varphi(\varphi, \lambda))^* : Y \rightarrow Y$ and introduce the adjoint problem:

$$\begin{cases} \frac{\partial \varphi^*}{\partial t} + (F'_\varphi(\varphi, \lambda))^* \varphi^* = C^* V_2(C\varphi - \varphi_{obs}), \\ \varphi^*|_{t=T} = 0. \end{cases} \quad (22)$$

Then (20) with (21) and (22) gives

$$\begin{aligned} J'(\lambda)v &= (V_1(\lambda - \lambda_b), v)_{Y_p} - (\varphi^*, F'_\lambda(\varphi, \lambda)v)_Y = \\ &= (V_1(\lambda - \lambda_b), v)_{Y_p} - ((F'_\lambda(\varphi, \lambda))^* \varphi^*, v)_{Y_p}, \end{aligned}$$

where $(F'_\lambda(\varphi, \lambda))^* : Y \rightarrow Y_p$ is the adjoint operator to $F'_\lambda(\varphi, \lambda)$. Therefore, the gradient of J is defined by

$$J'(\lambda) = V_1(\lambda - \lambda_b) - (F'_\lambda(\varphi, \lambda))^* \varphi^*. \quad (23)$$

From (20)–(23) we get the optimality system (the necessary optimality conditions, Lions (1968)):

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda) + f, & t \in (0, T), \\ \varphi|_{t=0} = u, \end{cases} \quad (24)$$

$$\begin{cases} \frac{\partial \varphi^*}{\partial t} + (F'_\varphi(\varphi, \lambda))^* \varphi^* = C^* V_2(C\varphi - \varphi_{obs}), \\ \varphi^*|_{t=T} = 0, \end{cases} \quad (25)$$

$$V_1(\lambda - \lambda_b) - (F'_\lambda(\varphi, \lambda))^* \varphi^* = 0. \quad (26)$$

We assume that the system (24)–(26) has a unique solution. The system (24)–(26) may be considered as a generalized model $\mathcal{A}(U) = 0$ with the state variable $U = (\varphi, \varphi^*, \lambda)$, and it contains information about observations.

If the observation operator C is nonlinear, i.e. $C\varphi = C(\varphi)$, then the right-hand side of the adjoint equation (25) contains $(C'_\varphi)^*$ instead of C^* and all the analysis presented below is similar.

To compute the Hessian $\mathcal{H}(\lambda)$ of the cost function (18) one should differentiate (24)–(25) and (23) with respect to λ , following Sect. 3. Then, the action of the Hessian $\mathcal{H}(\lambda)$ on a function $w \in Y_p$ is defined by the successive solutions of the following problems:

$$\begin{cases} \frac{\partial \phi}{\partial t} - F'_\varphi(\varphi, \lambda)\phi = F'_\lambda(\varphi, \lambda)w, & t \in (0, T) \\ \phi|_{t=0} = 0, \end{cases} \quad (27)$$

$$\begin{cases} -\frac{\partial \phi^*}{\partial t} - (F'_\varphi(\varphi, \lambda))^* \phi^* - (F''_{\varphi\varphi}(\varphi, \lambda)\phi)^* \varphi^* = (F''_{\lambda\varphi}(\varphi, \lambda)w)^* \varphi^* - C^* V_2 C \phi, \\ \phi^*|_{t=T} = 0, \end{cases} \quad (28)$$

$$\mathcal{H}(\lambda)w = V_1 w - (F''_{\varphi\lambda}(\varphi, \lambda)\phi)^* \varphi^* - (F''_{\lambda\lambda}(\varphi, \lambda)w)^* \varphi^* - (F'_\lambda(\varphi, \lambda))^* \phi^*. \quad (29)$$

The definition of the Hessian $\mathcal{H}(\lambda)$ by (27)–(29) involves the second-order derivatives of the model operator F with respect to φ and λ .

Numerical examples for computing the Hessian for the parameter estimation problems are presented in Gejadze et al. (2010).

5 Sensitivity Analysis

In the environmental sciences the mathematical models contain parameters which cannot be estimated precisely, because they are used to parametrize some subgrid processes and therefore can not be physically measured. Thus, it is important to be able to estimate the impact of uncertainties on the outputs of the model after assimilation. The optimal solution depends on the parameters, which may contain uncertainties, and for the forecasts it is very important to study the sensitivity of the optimal solution and its functionals with respect to these parameters Marchuk (1995), Cacuci (1981), Dontchev (1983), Griesse and Vexler (2007).

The necessary optimality condition is related to the gradient of the original cost function, thus to study the sensitivity of the optimal solution, one should differentiate the optimality system with respect to imprecisely known parameters. In this case, we come to the second-order adjoint problem Le Dimet et al. (2002). The first studies of sensitivity of the response functions after assimilation with the use of second-order adjoint were done by Le Dimet et al. (1997) for variational data assimilation problem aimed at restoration of initial condition, where sensitivity with respect to model parameters was considered. The equations of the forecast sensitivity to observations in a four-dimensional (4D-Var) data assimilation were derived by Daescu (2008). Based on these results, a practical computational approach was given by Cioaca et al. (2013) to quantify the effect of observations in 4D-Var data assimilation.

Sensitivity of the optimal solution is related to its statistical properties (see Gejadze et al. (2008, 2011, 2013), Shutyaev et al. (2012)). General sensitivity analysis in variational data assimilation with respect to observations for a nonlinear dynamic model was given in Shutyaev et al. (2017)–Shutyaev et al. (2018) to control the initial-value function and the model parameters.

Consider the mathematical model of a physical process that is described by the evolution problem of the form (17):

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda), & t \in (0, T) \\ \varphi|_{t=0} = u. \end{cases} \quad (30)$$

Suppose that for given $u \in X$ and $\lambda \in Y_p$ there exists a unique solution $\varphi \in Y$ to (30).

We introduce the functional

$$J(u) = \frac{1}{2}(V_1(u - u_0), u - u_0)_X + \frac{1}{2}(V_2(C\varphi - \varphi_{obs}), C\varphi - \varphi_{obs})_{Y_{obs}}, \quad (31)$$

where $u_0 \in X$ is a prior initial-value function (background state), $\varphi_{obs} \in Y_{obs}$ is a prescribed function (observational data), Y_{obs} is a Hilbert space (observation space), $C : Y \rightarrow Y_{obs}$ is a linear bounded operator, $V_1 : X \rightarrow X$ and $V_2 : Y_{obs} \rightarrow Y_{obs}$ are symmetric positive definite operators.

Consider the variational data assimilation problem with the aim to identify the initial condition: for given $\lambda \in Y_p$ find $u \in X$ and $\varphi \in Y$ such that they satisfy (30), and on the set of solutions to (30), the functional $J(u)$ takes the minimum value, i.e.

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda), & t \in (0, T) \\ \varphi|_{t=0} = u, \\ J(u) = \inf_v J(v). \end{cases} \quad (32)$$

The necessary optimality condition reduces the problem (32) to the optimality system:

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda), & t \in (0, T) \\ \varphi|_{t=0} = u, \end{cases} \quad (33)$$

$$\begin{cases} -\frac{\partial \varphi^*}{\partial t} - (F'_\varphi(\varphi, \lambda))^* \varphi^* = -C^* V_2(C\varphi - \varphi_{obs}), & t \in (0, T) \\ \varphi^*|_{t=T} = 0, \end{cases} \quad (34)$$

$$V_1(u - u_0) - \varphi^*|_{t=0} = 0 \quad (35)$$

with the unknowns φ, φ^*, u , where $(F'_\varphi(\varphi, \lambda))^*$ is the adjoint to the Frechet derivative of F with respect to φ .

We assume that the system (33)–(35) has a unique solution. The system (33)–(35) may be considered as a generalized model $\mathcal{F}(U, \lambda) = 0$ with the state variable $U = (\varphi, \varphi^*, u)$, and it contains all the available information. All the components of U depend on the parameters $\lambda \in Y_p$, which may contain uncertainties. An important issue is to study the sensitivity of this generalized model with respect to the parameters.

Let us introduce a response function $G(\varphi, u, \lambda)$, which is supposed to be a real-valued function and can be considered as a functional on $Y \times X \times Y_p$. We are interested in the sensitivity of G with respect to λ , with φ and u obtained from the optimality system (33)–(35). As is known Marchuk (1995), Cacuci (1981), Dontchev (1983), sensitivity is defined by the gradient of G with respect to λ , which is a functional derivative:

$$\frac{dG}{d\lambda} = \frac{\partial G}{\partial \varphi} \frac{\partial \varphi}{\partial \lambda} + \frac{\partial G}{\partial u} \frac{\partial u}{\partial \lambda} + \frac{\partial G}{\partial \lambda}. \quad (36)$$

If $\delta\lambda$ is a perturbation on λ , we get from the optimality system:

$$\begin{cases} \frac{\partial \delta \varphi}{\partial t} = F'_\varphi(\varphi, \lambda) \delta \varphi + F'_\lambda(\varphi, \lambda) \delta \lambda, & t \in (0, T) \\ \delta \varphi|_{t=0} = \delta u, \end{cases} \quad (37)$$

$$\begin{cases} -\frac{\partial \delta \varphi^*}{\partial t} - (F'_\varphi(\varphi, \lambda))^* \delta \varphi^* - (F''_{\varphi\varphi}(\varphi, \lambda) \delta \varphi + F''_{\varphi\lambda}(\varphi, \lambda) \delta \lambda)^* \varphi^* = -C^* V_2 C \delta \varphi, \\ \delta \varphi^*|_{t=T} = 0, \end{cases} \quad (38)$$

$$V_1 \delta u - \delta \varphi^*|_{t=0} = 0, \quad (39)$$

and

$$\left(\frac{dG}{d\lambda}, \delta \lambda \right)_{Y_p} = \left(\frac{\partial G}{\partial \varphi}, \delta \varphi \right)_Y + \left(\frac{\partial G}{\partial u}, \delta u \right)_X + \left(\frac{\partial G}{\partial \lambda}, \delta \lambda \right)_{Y_p}, \quad (40)$$

where $\delta\varphi$, $\delta\varphi^*$ and δu are the Gâteaux derivatives of φ , φ^* and u in the direction $\delta\lambda$ (for example, $\delta\varphi = \frac{\partial \varphi}{\partial \lambda} \delta\lambda$).

To compute the gradient $\nabla_\lambda G(\varphi, u, \lambda)$, let us introduce three adjoint variables $P_1 \in Y$, $P_2 \in Y$ and $P_3 \in X$. By taking the inner product of (37) by P_1 , (38) by P_2 and of (39) by P_3 and adding them, we obtain:

$$\begin{aligned} & \left(\delta \varphi, -\frac{\partial P_1}{\partial t} - (F'_\varphi(\varphi, \lambda))^* P_1 - (F''_{\varphi\varphi}(\varphi, \lambda) P_2)^* \varphi^* + C^* V_2 C P_2 \right)_Y + \left(\delta \varphi|_{t=T}, P_1|_{t=T} \right)_X + \\ & + \left(\delta \varphi^*, \frac{\partial P_2}{\partial t} - F'_\varphi(\varphi, \lambda) P_2 \right)_Y + \left(\delta \varphi^*|_{t=0}, P_2|_{t=0} - P_3 \right)_X + \\ & + \left(\delta u, -P_1|_{t=0} + V_1 P_3 \right)_X + \left(\delta \lambda, -(F'_\lambda(\varphi, \lambda))^* P_1 - (F''_{\varphi\lambda}(\varphi, \lambda) P_2)^* \varphi^* \right)_{Y_p} = 0. \end{aligned} \quad (41)$$

Here we put

$$-\frac{\partial P_1}{\partial t} - (F'_\varphi(\varphi, \lambda))^* P_1 - (F''_{\varphi\varphi}(\varphi, \lambda) P_2)^* \varphi^* + C^* V_2 C P_2 = \frac{\partial G}{\partial \varphi},$$

and

$$-P_1|_{t=0} + V_1 P_3 = \frac{\partial G}{\partial u}, \quad P_1|_{t=T} = 0, \quad \frac{\partial P_2}{\partial t} - F'_\varphi(\varphi, \lambda) P_2 = 0, \quad P_2|_{t=0} - P_3 = 0.$$

Hence, we can exclude the variable P_3 by

$$P_3 = P_2|_{t=0}$$

and obtain the initial condition for P_2 in the form:

$$V_1 P_2|_{t=0} = \frac{\partial G}{\partial u} + P_1|_{t=0}.$$

Thus, if P_1, P_2 are the solutions of the following system of equations

$$\begin{cases} -\frac{\partial P_1}{\partial t} - (F'_\varphi(\varphi, \lambda))^* P_1 - (F''_{\varphi\varphi}(\varphi, \lambda) P_2)^* \varphi^* + C^* V_2 C P_2 = \frac{\partial G}{\partial \varphi}, & t \in (0, T) \\ P_1|_{t=T} = 0, \end{cases} \quad (42)$$

$$\begin{cases} \frac{\partial P_2}{\partial t} - F'_\varphi(\varphi, \lambda) P_2 = 0, & t \in (0, T) \\ V_1 P_2|_{t=0} = \frac{\partial G}{\partial u} + P_1|_{t=0}, \end{cases} \quad (43)$$

then from (41) we get

$$\left(\frac{\partial G}{\partial \varphi}, \delta \varphi \right)_Y + \left(\frac{\partial G}{\partial u}, \delta u \right)_X = \left(\delta \lambda, (F'_\lambda(\varphi, \lambda))^* P_1 + (F''_{\varphi\lambda}(\varphi, \lambda) P_2)^* \varphi^* \right)_{Y_p},$$

and the gradient of G is given by

$$\frac{dG}{d\lambda} = (F'_\lambda(\varphi, \lambda))^* P_1 + (F''_{\varphi\lambda}(\varphi, \lambda) P_2)^* \varphi^* + \frac{\partial G}{\partial \lambda}. \quad (44)$$

We get a coupled system of two differential equations (42) and (43) of the first order with respect to time. One equation has a final condition (backward problem) while the other has an initial condition (forward problem) depending on the initial value for the first equation: it is a non-standard problem.

Let us represent the non-standard problem (42)–(43) in an equivalent form:

$$\begin{cases} -\frac{\partial P_1}{\partial t} - (F'_\varphi(\varphi, \lambda))^* P_1 - (F''_{\varphi\varphi}(\varphi, \lambda) P_2)^* \varphi^* + C^* V_2 C P_2 = \frac{\partial G}{\partial \varphi}, & t \in (0, T) \\ P_1|_{t=T} = 0, \end{cases} \quad (45)$$

$$\begin{cases} \frac{\partial P_2}{\partial t} - F'_\varphi(\varphi, \lambda) P_2 = 0, & t \in (0, T) \\ P_2|_{t=0} = v, \end{cases} \quad (46)$$

$$V_1 v - P_1|_{t=0} = \frac{\partial G}{\partial u}. \quad (47)$$

Here we have three unknowns: $v \in X$, $P_1, P_2 \in Y$. Let us write (45)–(47) in the form of an operator equation for v . We define the operator \mathcal{H} by the successive solution of the following problems:

$$\begin{cases} \frac{\partial \phi}{\partial t} - F'_\varphi(\varphi, \lambda)\phi = 0, & t \in (0, T) \\ \phi|_{t=0} = w, \end{cases} \quad (48)$$

$$\begin{cases} -\frac{\partial \phi^*}{\partial t} - (F'_\varphi(\varphi, \lambda))^*\phi^* - (F''_{\varphi\varphi}(\varphi, \lambda)\phi)^*\varphi^* = -C^*V_2C\phi, & t \in (0, T) \\ \phi^*|_{t=T} = 0, \end{cases} \quad (49)$$

$$\mathcal{H}w = V_1w - \phi^*|_{t=0}. \quad (50)$$

Then (45)–(47) is equivalent to the following equation in X :

$$\mathcal{H}v = \mathcal{F} \quad (51)$$

with the right-hand side \mathcal{F} defined by

$$\mathcal{F} = \frac{\partial G}{\partial u} + \tilde{\phi}^*|_{t=0}, \quad (52)$$

where $\tilde{\phi}^*$ is the solution to the adjoint problem:

$$\begin{cases} -\frac{\partial \tilde{\phi}^*}{\partial t} - (F'_\varphi(\varphi, \lambda))^*\tilde{\phi}^* = \frac{\partial G}{\partial \varphi}, & t \in (0, T) \\ \tilde{\phi}^*|_{t=T} = 0. \end{cases} \quad (53)$$

It is easily seen that the operator \mathcal{H} defined by (48)–(50) is the Hessian of the original functional J considered on the optimal solution u of the problem (33)–(35): $J''(u) = \mathcal{H}$. Under the assumption that \mathcal{H} is positive definite, the operator equation (51) is correctly and everywhere solvable in X , i.e. for every \mathcal{F} there exists a unique solution $v \in X$ and

$$\|v\|_X \leq c\|\mathcal{H}\|_X, \quad c = const > 0.$$

Therefore, under the assumption that $J''(u)$ is positive definite on the optimal solution, the non-standard problem (42)–(43) has a unique solution $P_1, P_2 \in Y$.

Based on the above consideration, we can formulate the following algorithm to solve the non-standard problem:

- (1) For $\frac{\partial G}{\partial u} \in X$, $\frac{\partial G}{\partial \varphi} \in Y$ solve the adjoint problem

$$\begin{cases} -\frac{\partial \tilde{\phi}^*}{\partial t} - (F'_\varphi(\varphi, \lambda))^* \tilde{\phi}^* = \frac{\partial G}{\partial \varphi}, & t \in (0, T) \\ \tilde{\phi}^*|_{t=T} = 0 \end{cases} \quad (54)$$

and put

$$\mathcal{F} = \frac{\partial G}{\partial u} + \tilde{\phi}^*|_{t=0}.$$

(2) Find v by solving

$$\mathcal{H}v = \mathcal{F}$$

with the Hessian of the original functional J defined by (48)–(50).

(3) Solve successively the direct and adjoint problems

$$\begin{cases} \frac{\partial P_2}{\partial t} - F'_\varphi(\varphi, \lambda)P_2 = 0, & t \in (0, T) \\ P_2|_{t=0} = v, \end{cases} \quad (55)$$

$$\begin{cases} -\frac{\partial \tilde{P}_1}{\partial t} - (F'_\varphi(\varphi, \lambda))^* \tilde{P}_1 - (F''_{\varphi\varphi}(\varphi, \lambda)P_2)^* \varphi^* + C^* V_2 C P_2 = 0, & t \in (0, T) \\ \tilde{P}_1|_{t=T} = 0, \end{cases} \quad (56)$$

and put

$$P_1 = \tilde{P}_1 + \tilde{\phi}^*.$$

Thus, we obtain $P_1, P_2 \in Y$ as the solutions to the non-standard problem (42)–(43), which determine the sensitivity of the response function with respect to imprecisely known parameters according to (44).

6 Sensitivity with Respect to Observations

In geophysical applications a usual request is the estimation of the sensitivity with respect to observations Langland and Baker (2004), Daescu and Langland (2013), Kalnay et al. (2012), Godinez and Daescu (2009). What will be the impact of an uncertainty on the prediction? It is clear that observations are not directly used in the forward model, they are involved only as a forcing term in the adjoint model. Therefore to apply the general formalism of sensitivity analysis we should apply it not to the model itself but to the optimality system, i.e. the model plus the adjoint model. A very simple example with a scalar ordinary differential equation is given in Le Dimet et al. (2002) showing that the only model is not sufficient to carry out

sensitivity analysis in the presence of data. Differentiating the optimality system will introduce second order derivatives.

Consider the mathematical model governed by the nonlinear evolution problem of the form (17):

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda) + f, & t \in (0, T) \\ \varphi|_{t=0} = u. \end{cases} \quad (57)$$

We suppose that for given $u \in X$, $f \in Y$ and $\lambda \in Y_p$ there exists a unique solution $\varphi \in Y$ to (57) with $\frac{\partial \varphi}{\partial t} \in Y$. The function λ is an unknown model parameter, and we suppose that the initial state u is also unknown, so we will consider joint parameter and state estimation problem.

Let us introduce the cost function as a functional on $X \times Y_p$ in the form

$$J(u, \lambda) = \frac{1}{2} \|V_1^{1/2}(u - u_b)\|_X + \frac{1}{2} \|V_2^{1/2}(\lambda - \lambda_b)\|_{Y_p} + \frac{1}{2} \|V_3^{1/2}(C\varphi - \varphi_{obs})\|_{Y_{obs}}, \quad (58)$$

where $u \in X$, $\lambda_b \in Y_p$ are prior (background) functions, $\varphi_{obs} \in Y_{obs}$ is a prescribed function (observational data), Y_{obs} is a Hilbert space (observation space), $C : Y \rightarrow Y_{obs}$ is a linear bounded operator (observation operator), $V_1 : X \rightarrow X$, $V_2 : Y_p \rightarrow Y_p$ and $V_3 : Y_{obs} \rightarrow Y_{obs}$ are symmetric positive definite bounded operators.

Let us consider the following data assimilation problem with the aim to find the initial value u and the parameter λ : for given $f \in Y$, $\varphi_{obs} \in Y_{obs}$, find $u \in X$, $\lambda \in Y_p$ and $\varphi \in Y$ such that they satisfy (57), and on the set of solutions to (57), the functional $J(u, \lambda)$ takes the minimum value, i.e.

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda) + f, & t \in (0, T) \\ \varphi|_{t=0} = u, \\ J(u, \lambda) = \inf_{w \in X, v \in Y_p} J(w, v). \end{cases} \quad (59)$$

We suppose that the solution of (59) exists. The necessary optimality condition reduces (59) to the optimality system:

$$\begin{cases} \frac{\partial \varphi}{\partial t} = F(\varphi, \lambda) + f, & t \in (0, T), \\ \varphi|_{t=0} = u, \end{cases} \quad (60)$$

$$\begin{cases} \frac{\partial \varphi^*}{\partial t} + (F'_\varphi(\varphi, \lambda))^* \varphi^* = C^* V_3 (C\varphi - \varphi_{obs}), & t \in (0, T) \\ \varphi^*|_{t=T} = 0, \end{cases} \quad (61)$$

$$V_1(u - u_b) - \varphi^*|_{t=0} = 0, \quad (62)$$

$$V_2(\lambda - \lambda_b) - (F'_\lambda(\varphi, \lambda))^* \varphi^* = 0. \quad (63)$$

Here $F'_\varphi(\varphi, \lambda) : Y \rightarrow Y$, $F'_\lambda(\varphi, \lambda) : Y_p \rightarrow Y$ are the Fréchet derivatives of F with respect to φ and λ , correspondingly, and C^* is the adjoint operator to C defined by $(C\varphi, \psi)_{Y_{obs}} = (\varphi, C^*\psi)_Y$, $\varphi \in Y$, $\psi \in Y_{obs}$.

Supposing that the system (60)–(63) has a unique solution φ , $\varphi^* \in Y$, $u \in X$, $\lambda \in Y_p$, we will study the sensitivity of functionals of the optimal solution with respect to the observation data φ_{obs} .

We introduce a response function $G(\varphi, u, \lambda)$, which is supposed to be a real-valued function and can be considered as a functional on $Z = Y \times X \times Y_p$. We are interested in the sensitivity of G with respect to φ_{obs} , with φ , u and λ obtained from the optimality system (60)–(63). By definition, the sensitivity is defined by the gradient of G with respect to φ_{obs} :

$$\frac{dG}{d\varphi_{obs}} = \frac{\partial G}{\partial \varphi} \frac{\partial \varphi}{\partial \varphi_{obs}} + \frac{\partial G}{\partial \lambda} \frac{\partial \lambda}{\partial \varphi_{obs}} + \frac{\partial G}{\partial u} \frac{\partial u}{\partial \varphi_{obs}}, \quad (64)$$

where $\frac{\partial G}{\partial \varphi} : Z \rightarrow Y$, $\frac{\partial G}{\partial \lambda} : Z \rightarrow Y_p$, $\frac{\partial G}{\partial u} : Z \rightarrow X$, and $\frac{\partial \varphi}{\partial \varphi_{obs}}$, $\frac{\partial \lambda}{\partial \varphi_{obs}}$, $\frac{\partial u}{\partial \varphi_{obs}}$ are the Gâteaux derivatives of φ , λ , u with respect to φ_{obs} .

Let $\delta\varphi_{obs}$ be a perturbation on φ_{obs} , then we obtain from the optimality system (60)–(63):

$$\begin{cases} \frac{\partial \delta\varphi}{\partial t} = F'_\varphi(\varphi, \lambda)\delta\varphi + F'_\lambda(\varphi, \lambda)\delta\lambda, & t \in (0, T) \\ \delta\varphi|_{t=0} = \delta u, \end{cases} \quad (65)$$

$$\begin{cases} -\frac{\partial \delta\varphi^*}{\partial t} - (F''_{\varphi\varphi}(\varphi, \lambda))^* \delta\varphi^* - (F''_{\varphi\lambda}(\varphi, \lambda)\delta\varphi)^* \varphi^* = (F''_{\varphi\lambda}(\varphi, \lambda)\delta\lambda)^* \varphi^* \\ \quad - C^* V_3 (C\delta\varphi - \delta\varphi_{obs}), \\ \delta\varphi^*|_{t=T} = 0, \end{cases} \quad (66)$$

$$V_1 \delta u - \delta\varphi^*|_{t=0} = 0, \quad (67)$$

$$V_2 \delta\lambda - (F''_{\lambda\varphi}(\varphi, \lambda)\delta\varphi)^* \varphi^* - (F''_{\lambda\lambda}(\varphi, \lambda)\delta\lambda)^* \varphi^* - (F'_\lambda(\varphi, \lambda))^* \delta\varphi^* = 0, \quad (68)$$

and

$$\left(\frac{dG}{d\varphi_{obs}}, \delta\varphi_{obs} \right)_{Y_{obs}} = \left(\frac{\partial G}{\partial \varphi}, \delta\varphi \right)_Y + \left(\frac{\partial G}{\partial \lambda}, \delta\lambda \right)_{Y_p} + \left(\frac{\partial G}{\partial u}, \delta u \right)_X, \quad (69)$$

where $\delta\varphi$, $\delta\varphi^*$, $\delta\lambda$, δu are the solutions of (65)–(68).

Following the methodology presented in Sect. 5, we obtain the gradient of G through solutions of a non-standard problem.

Let $P_1, P_2 \in Y$, $P_3 \in Y_p$, $P_4 \in X$ be the solutions of the following system of equations

Under the assumption that \mathcal{H} is positive definite, the operator equation (78) is correctly and everywhere solvable in $X \times Y_p$, i.e. for every \mathcal{F} there exists a unique solution $U \in X \times Y_p$ and the estimate is valid:

$$\|U\|_{X \times Y_p} \leq c \|\mathcal{F}\|_{X \times Y_p}, \quad c = \text{const} > 0.$$

Therefore, under the assumption that $J''(u, \lambda)$ is positive definite on the optimal solution, the non-standard problem (70)–(73) has a unique solution $P_1, P_2 \in Y, P_3 \in Y_p, P_4 \in X$.

Based on (70)–(74), we can formulate the following algorithm to compute the gradient of the response function G :

- (1) For $\frac{\partial G}{\partial \lambda} \in Y_p, \frac{\partial G}{\partial \varphi} \in Y, \frac{\partial G}{\partial u} \in X$ solve the adjoint problem

$$\begin{cases} -\frac{\partial \tilde{\phi}^*}{\partial t} - (F'_\varphi(\varphi, \lambda))^* \tilde{\phi}^* = \frac{\partial G}{\partial \varphi}, & t \in (0, T) \\ \tilde{\phi}^*|_{t=T} = 0 \end{cases} \quad (79)$$

and put

$$\mathcal{F} = \left(\frac{\partial G}{\partial u} + \tilde{\phi}^*|_{t=0}, \frac{\partial G}{\partial \lambda} + (F'_\lambda(\varphi, \lambda))^* \tilde{\phi}^* \right)^T.$$

- (2) Find $U = (w, v)^T$ by solving

$$\mathcal{H}U = \mathcal{F}$$

with the Hessian of the original functional J defined by (75)–(77).

- (3) Solve the direct problem

$$\begin{cases} \frac{\partial P_2}{\partial t} - F'_\varphi(\varphi, \lambda)P_2 = F'_\lambda(\varphi, \lambda)v, & t \in (0, T) \\ P_2|_{t=0} = w. \end{cases} \quad (80)$$

- (4) Compute the gradient of the response function as

$$\frac{dG}{d\varphi_{obs}} = V_3 C P_2. \quad (81)$$

The last formula allows us to estimate the sensitivity of the response functions related to the optimal solution after assimilation, with respect to observation data.

7 Application for a Sea Thermodynamics Model

We consider the sea thermodynamics problem in the form Marchuk et al. (1987):

$$\begin{aligned}
 T_t + (\bar{U}, \text{Grad})T - \text{Div}(\hat{a}_T \cdot \text{Grad} T) &= f_T \text{ in } D \times (t_0, t_1), \\
 T &= T_0 \text{ for } t = t_0 \text{ in } D, \\
 -v_T \frac{\partial T}{\partial z} &= Q \text{ on } \Gamma_S \times (t_0, t_1), \quad \frac{\partial T}{\partial n} = 0 \text{ on } \Gamma_{w,c} \times (t_0, t_1), \\
 \bar{U}_n^{(-)} T + \frac{\partial T}{\partial n} &= Q_T \text{ on } \Gamma_{w,op} \times (t_0, t_1), \\
 \frac{\partial T}{\partial n} &= 0 \text{ on } \Gamma_H \times (t_0, t_1),
 \end{aligned} \tag{82}$$

where $T = T(x, y, z, t)$ is an unknown temperature function, $t \in (t_0, t_1)$, $(x, y, z) \in D = \Omega \times (0, H)$, $\Omega \subset R^2$, $H = H(x, y)$ is the function of the bottom relief, $Q = Q(x, y, t)$ is the total heat flux, $\bar{U} = (u, v, w)$, $\hat{a}_T = \text{diag}((a_T)_{ii})$, $(a_T)_{11} = (a_T)_{22} = \mu_T$, $(a_T)_{33} = \nu_T$, $f_T = f_T(x, y, z, t)$ are given functions. The boundary of the domain $\Gamma \equiv \partial D$ is represented as a union of four disjoint parts $\Gamma_S, \Gamma_{w,op}, \Gamma_{w,c}, \Gamma_H$, where $\Gamma_S = \Omega$ (the unperturbed sea surface), $\Gamma_{w,op}$ is the liquid (open) part of vertical lateral boundary, $\Gamma_{w,c}$ is the solid part of the vertical lateral boundary, Γ_H is the sea bottom, $\bar{U}_n^{(-)} = (|\bar{U}_n| - \bar{U}_n)/2$, and \bar{U}_n is the normal component of \bar{U} . The other notations and a detailed description of the problem statement can be found in Agoshkov et al. (2008).

Problem (82) can be written in the form of an operator equation:

$$\begin{aligned}
 T_t + LT &= \mathcal{F} + BQ, \quad t \in (t_0, t_1), \\
 T &= T_0, \quad t = t_0,
 \end{aligned} \tag{83}$$

where the equality is understood in the weak sense, namely,

$$(T_t, \hat{T}) + (LT, \hat{T}) = \mathcal{F}(\hat{T}) + (BQ, \hat{T}) \quad \forall \hat{T} \in W_2^1(D), \tag{84}$$

in this case L, \mathcal{F}, B are defined by the following relations:

$$\begin{aligned}
 (LT, \hat{T}) &\equiv \int_D (-T \text{Div}(\bar{U} \hat{T})) dD + \int_{\Gamma_{w,op}} \bar{U}_n^{(+)} T \hat{T} d\Gamma + \int_D \hat{a}_T \text{Grad}(T) \cdot \text{Grad}(\hat{T}) dD, \\
 \mathcal{F}(\hat{T}) &= \int_{\Gamma_{w,op}} Q_T \hat{T} d\Gamma + \int_D f_T \hat{T} dD, \quad (T_t, \hat{T}) = \int_D T_t \hat{T} dD, \quad (BQ, \hat{T}) = \int_{\Omega} Q \hat{T}|_{z=0} d\Omega,
 \end{aligned}$$

and the functions $\widehat{a}_T, Q_T, f_T, Q$ are such that equality (84) makes sense. The properties of the operator L were studied in Agoshkov et al. (2008).

Problem (82) is linear in T, Q , however, written in the form (83), it is a particular case of the original problem (57), and all the reasoning and the methodology presented in Sect. 6 are easily transferred to the case of problem (83), understood in a weak sense (83).

We consider the data assimilation problem for the sea surface temperature (see Agoshkov et al. (2008)). Suppose that the functions $Q \in L_2(\Omega \times (t_0, t_1))$ and $T_0 \in L_2(D)$ are unknown in problem (82). Let also $T_{obs}(x, y, t) \in L_2(\Omega \times (t_0, t_1))$ be the function on Ω obtained for $t \in (t_0, t_1)$ by processing the observation data, and this function in its physical sense is an approximation to the surface temperature function on Ω , i.e. to $T|_{z=0}$. We admit the case when T_{obs} is defined only on some subset of $\Omega \times (t_0, t_1)$ and denote the indicator (characteristic) function of this set by m_0 . For definiteness sake, we assume that T_{obs} is zero outside this subset.

Consider the data assimilation problem for the surface temperature in the following form: find T_0 and Q such that

$$\begin{cases} T_t + LT = \mathcal{F} + BQ & \text{in } D \times (t_0, t_1), \\ T = T_0, & t = t_0 \\ J(T_0, Q) = \inf_{w,v} J(w, v), \end{cases} \tag{85}$$

where

$$\begin{aligned} J(T_0, Q) = & \frac{\alpha}{2} \int_{t_0}^{t_1} \int_{\Omega} |Q - Q^{(0)}|^2 d\Omega dt + \frac{\beta}{2} \int_D |T_0 - T^{(0)}|^2 dD + \\ & + \frac{1}{2} \int_{t_0}^{t_1} \int_{\Omega} m_0 |T|_{z=0} - T_{obs}|^2 d\Omega dt, \end{aligned} \tag{86}$$

and $Q^{(0)} = Q^{(0)}(x, y, t), T^{(0)} = T^{(0)}(x, y, z)$ are given functions, $\alpha, \beta = const > 0$.

For $\alpha, \beta > 0$ this variational data assimilation problem has a unique solution. The existence of the optimal solution follows from the classic results of the theory of optimal control problems Lions (1968).

The optimality system determining the solution of the formulated variational data assimilation problem according to the necessary condition $\text{grad}J = 0$ has the form:

$$\begin{aligned} T_t + LT = \mathcal{F} + BQ & \text{ in } D \times (t_0, t_1), \\ T = T_0, & t = t_0, \end{aligned} \tag{87}$$

$$\begin{aligned} -(T^*)_t + L^*T^* &= Bm_0(T_{\text{obs}} - T) \text{ in } D \times (t_0, t_1), \\ T^* &= 0, \quad t = t_1, \end{aligned} \tag{88}$$

$$\alpha(Q - Q^{(0)}) - T^* = 0 \text{ on } \Omega \times (t_0, t_1), \tag{89}$$

$$\beta(T_0 - T^{(0)}) - T^*|_{t=0} = 0 \text{ in } D, \tag{90}$$

where L^* is the operator adjoint to L .

Here the boundary-value function Q plays the role of λ from Sect. 6, $\varphi = T$, the operator F has the form $F(T, Q) = -LT + BQ$, and $F'_T = -L$, $F'_Q = B$. Since the operator $F(T, Q)$ is linear in this case and $F''_{TT} = F''_{QT} = F''_{QQ} = 0$, the Hessian \mathcal{H} acting on some $U = (w, \psi)^T$, $w \in L_2(D)$, $\psi \in L_2(\Omega \times (t_0, t_1))$ is defined by the successive solution of the following problems:

$$\begin{cases} \frac{\partial \phi}{\partial t} + L\phi = B\psi, & t \in (t_0, t_1) \\ \phi|_{t=t_0} = w, \end{cases} \tag{91}$$

$$\begin{cases} -\frac{\partial \phi^*}{\partial t} + L^*\phi^* = -Bm_0\phi, & t \in (t_0, t_1) \\ \phi^*|_{t=t_1} = 0, \end{cases} \tag{92}$$

$$\mathcal{H}U = (\beta w - \phi^*|_{t=0}, \alpha\psi - B^*\phi^*)^T. \tag{93}$$

To illustrate the above-presented theory, we consider the problem of sensitivity of functionals of the optimal solution T_0 , Q to the observations T_{obs} . Let us introduce the following response function:

$$G(T) = \int_{t_0}^{t_1} dt \int_{\Omega} k(x, y, t)T(x, y, 0, t)d\Omega, \tag{94}$$

where $k(x, y, t)$ is a weight function related to the temperature field on the sea surface $z = 0$. For example, if we are interested in the mean temperature of a specific region of the sea ω for $z = 0$ in the interval $\bar{t} - \tau \leq t \leq \bar{t}$, then as k we take the function

$$k(x, y, t) = \begin{cases} 1/(\tau \text{mes } \omega) & \text{if } (x, y) \in \omega, \bar{t} - \tau \leq t \leq \bar{t} \\ 0 & \text{else,} \end{cases} \tag{95}$$

where $\text{mes } \omega$ denotes the area of the region ω . Thus, the functional (94) is written in the form:

$$G(T) = \frac{1}{\tau} \int_{\bar{t}-\tau}^{\bar{t}} dt \left(\frac{1}{\text{mes } \omega} \int_{\omega} T(x, y, 0, t)d\Omega \right). \tag{96}$$

Formula (96) represents the mean temperature averaged over the time interval $\bar{t} - \tau \leq t \leq \bar{t}$ for a given region ω . The response functions of this type are of most interest in the theory of climate change (Marchuk (1995), Marchuk et al. (1996)).

In our notations the functional (94) may be written as

$$G(T) = \int_{t_0}^{t_1} (Bk, T) dt = (Bk, T)_Y, \quad Y = L_2(D \times (t_0, t_1)).$$

We are interested in the sensitivity of the response function $G(T)$, obtained for T after data assimilation, with respect to the observation function T_{obs} .

By definition, the sensitivity is given by the gradient of G with respect to T_{obs} :

$$\frac{dG}{dT_{obs}} = \frac{\partial G}{\partial T} \frac{\partial T}{\partial T_{obs}}. \quad (97)$$

Since $\frac{\partial G}{\partial T} = Bk$, then according to the theory presented in Sect. 6, to compute the gradient (97) we need to perform the following steps:

1) For k defined by (95) solve the adjoint problem

$$\begin{cases} -\frac{\partial \tilde{\phi}^*}{\partial t} + L^* \tilde{\phi}^* = Bk, & t \in (t_0, t_1) \\ \tilde{\phi}^*|_{t=t_1} = 0 \end{cases} \quad (98)$$

and put $\Phi = (\tilde{\phi}^*|_{t=0}, B^* \tilde{\phi}^*)^T$.

2) Find $U = (w, v)^T$ by solving $\mathcal{H}U = \Phi$ with the Hessian defined by (91)–(93).

3) Solve the direct problem

$$\begin{cases} \frac{\partial P_2}{\partial t} + LP_2 = Bv, & t \in (t_0, t_1) \\ P_2|_{t=t_0} = w. \end{cases} \quad (99)$$

4) Compute the gradient of the response function as

$$\frac{dG}{dT_{obs}} = m_0 P_2|_{z=0}. \quad (100)$$

The last formula allows us to estimate the sensitivity of the functionals related to the mean temperature after data assimilation, with respect to the observations on the sea surface.

For numerical experiments have used the three-dimensional numerical model of the Baltic Sea hydrothermodynamics developed at the INM RAS on the base of the splitting method Zalesny et al. (2017) and supplied with the assimilation procedure Agoshkov et al. (2008) for the surface temperature T_{obs} with the aim to reconstruct the heat fluxes Q and the initial state T_0 .

The parameters of the considered domain of the Baltic Sea and its geographic coordinates can be described as follows: σ -grid is $336 \times 394 \times 25$ (the latitude, longitude, and depth, respectively). The first point of the "grid C" Zalesny et al. (2017) has the coordinates 9.406° E and 53.64° N. The mesh sizes in x and y are constant and equal to 0.0625 and 0.03125 degrees. The time step is $\Delta t = 5$ minutes. The assimilation procedure worked only during some time windows. To start the assimilation procedure, the function $T^{(0)}$ was taken as a model forecast for the previous time interval.

The Baltic Sea daily-averaged nighttime surface temperature data were used for T_{obs} . These are the data of the Danish Meteorological Institute based on measurements of radiometers (AVHRR, AATSR and AMSRE) and spectroradiometers (SEVIRI and MODIS) Karagali et al. (2012). Data interpolation algorithms were used Zakharova et al. (2013) to convert observations on computational grid of the numerical model of the Baltic Sea thermodynamics. The mean climatic flux obtained from the NCEP (National Center for Environmental Prediction) reanalysis was taken for $Q^{(0)}$.

Using the hydrothermodynamics model mentioned above, which is supplied with the assimilation procedure for the surface temperature T_{obs} , we have performed calculations for the Baltic Sea area where the assimilation algorithm worked only at certain time moments t_0 ; in this case $t_1 = t_0 + \Delta t$. The aim of the experiment was the numerical study of the sensitivity of functionals of the optimal solution T_0 , Q to observation errors in the interval (t_0, t_1) .

We use the discretize-then-optimize approach, and for numerical experiments all the presented equations are understood in a discrete form, as finite-dimensional analogues of the corresponding problems, obtained after approximation. This allows us to consider model equations as a perfect model, with no approximation errors.

Let us present some results of numerical experiments.

The calculation results for $t_0 = 50$ h (600 time steps for the model) are presented in Fig. 1 showing the gradient of the response function $G(T)$ defined by (96) and related to the mean temperature after data assimilation, with respect to the observations on the sea surface, according to (98)–(100). Here $\omega = \Omega$, $\tau = \Delta t$, $\bar{t} = t_1$, $\alpha = \beta = 10^{-5}$ (Fig. 1).

We can see the sub-areas (in red) in which the response function $G(T)$ is most sensitive to errors in the observations during assimilation. The largest values of the gradient of $G(T)$ correspond to the points x, y with a small depth (cf. sea topography, Fig. 2). Thus, the considered functional $G(T)$ of the optimal solution turned out to be the most sensitive to observation errors at surface points near these regions. This result is confirmed by the direct computation of the response function $G(T)$ according to (96) obtained after assimilation, by introducing perturbations into the observation data T_{obs} .

The above studies allow to determine the sea sub-areas in which the response function related to the optimal solution is most sensitive to errors in the observations during variational data assimilation.

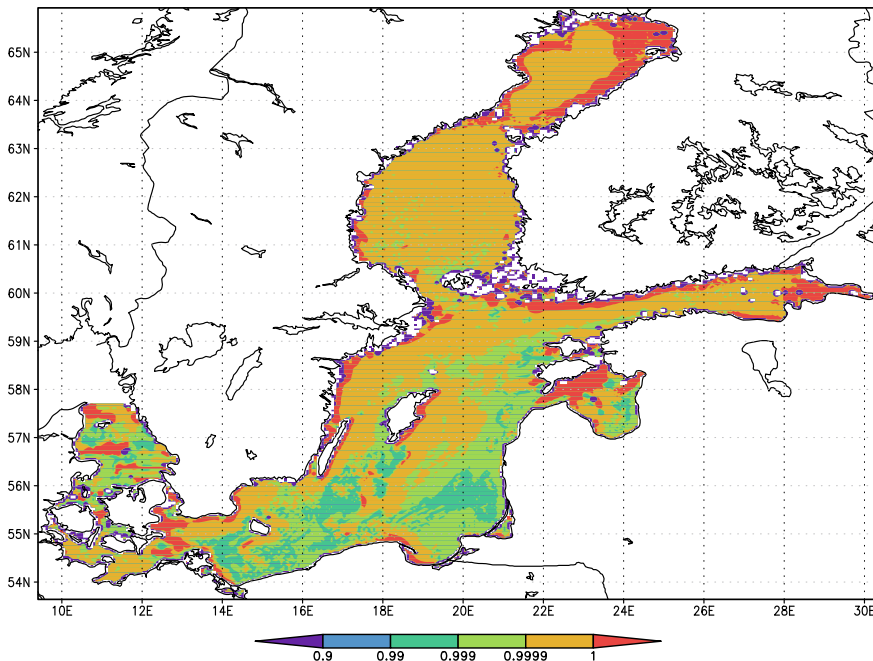


Fig. 1 The gradient of the response function $G(T)$

8 Conclusions

Variational data assimilation is an efficient method in modeling the large-scale geophysical flows, with the main difficulty being linked to the nonlinearity of the governing equations. This method allows to combine the observational data and the model forecasts. From the mathematical point of view, we have to deal with the initial-boundary-value control problems for a nonlinear evolution model governed by partial differential equations. The necessary optimality condition is defined by the optimality system which is based on the gradient of the cost function and involves forward and adjoint equations. To study the variational data assimilation problem as an optimal control problem and to develop efficient algorithms for its numerical solution, second-order information is needed. This is information about the Hessian of the cost functional. To construct the Hessian, it is necessary to differentiate the optimality system and derive a second-order adjoint problem. The investigation of the second-order adjoint equations and the Hessian of the cost function plays an important role in the study of the solvability of the variational assimilation problem, the construction of algorithms for its numerical solution based on the modification of Newton type methods, the identification of model parameters. The Hessian allows to study the sensitivity of the optimal solution and its functionals with respect to observations and uncertainties in parameters.

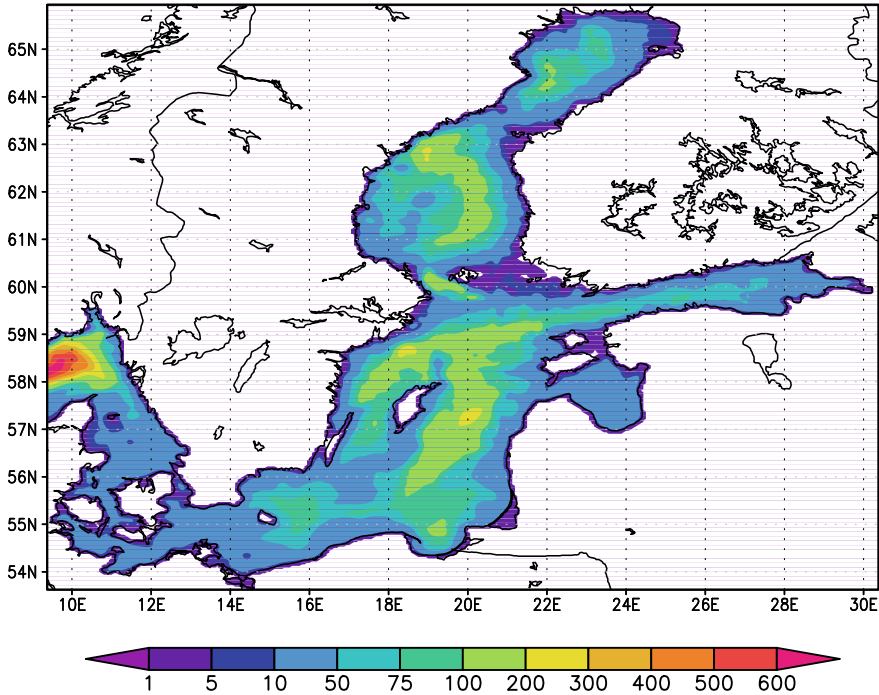


Fig. 2 Baltic Sea topography

Acknowledgements The research (Sections 5–7) was supported by the Russian Science Foundation (project No.20-11-20057).

References

Agoshkov VI, Sheloput TO (2017) The study and numerical solution of some inverse problems in simulation of hydrophysical fields in water areas with ‘liquid’ boundaries. *Russ J Numer Anal Math Modelling* 32:147–164

Agoshkov VI, Parmuzin EI, Shutyaev VP (2008) Numerical algorithm of variational assimilation of the ocean surface temperature data. *Comp Math Math Phys* 48:1371–1391

Agoshkov VI, Parmuzin EI, Zalesny VB, Shutyaev VP et al (2015) Variational assimilation of observation data in the mathematical model of the Baltic Sea dynamics. *Russ J Numer Anal Math Modelling* 30:203–212

Alifanov OM, Artyukhin EA, Rumyantsev SV (1996) *Extreme methods for solving ill-posed problems with applications to inverse heat transfer problems*. Begell House Publishers, Danbury

Asch M, Bocquet M, Nodet M (2016) *Data assimilation: methods, algorithms, and applications*. SIAM, Philadelphia

Bocquet M (2012) Parameter-field estimation for atmospheric dispersion: application to the Chernobyl accident using 4D-Var. *Q J R Meteor Soc* 138:664–681

- Cacuci DG (1981) Sensitivity theory for nonlinear systems: II. Extensions to additional classes of responses. *J Math Phys* 22:2803–2812
- Carrassi A, Bocquet M, Bertino L, Evensen G (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *WIREs Clim Change* 9:1–80
- Chavent G (1983) Local stability of the output least square parameter estimation technique. *Math Appl Comp* 2:3–22
- Cioaca A, Sandu A, de Sturler E (2013) Efficient methods for computing observation impact in 4D-Var data assimilation. *Comput Geosci* 17:975–990
- Daescu DN (2008) On the sensitivity equations of four-dimensional variational (4D-Var) data assimilation. *Mon Weather Rev* 136:3050–3065
- Daescu DN, Langland RH (2013) Error covariance sensitivity and impact estimation with adjoint 4D-Var: Theoretical aspects and first applications to NAVDAS-AR. *Q J R Meteorol Soc* 139:226–241
- Dontchev AL (1983) Perturbations, approximations and sensitivity analysis of optimal control systems. *Lecture notes in control and information sciences*, vol 52. Springer, Berlin
- Fletcher SJ (2017) *Data assimilation for the geosciences: from theory to application*. Elsevier, Amsterdam
- Gejadze I, Le Dimet FX, Shutyaev V (2008) On analysis error covariances in variational data assimilation. *SIAM J Sci Comp* 30:1847–1874
- Gejadze I, Le Dimet FX, Shutyaev V (2010) On optimal solution error covariances in variational data assimilation problems. *J Comput Phys* 229:2159–2178
- Gejadze IYu, Copeland GJM, Le Dimet FX, Shutyaev VP (2011) Computation of the analysis error covariance in variational data assimilation problems with nonlinear dynamics. *J Comput Phys* 230:7923–7943
- Gejadze IYu, Shutyaev VP, Le Dimet FX (2013) Analysis error covariance versus posterior covariance in variational data assimilation. *Q J R Meteor Soc* 139:1826–1841
- Gill PE, Murray W, Wright MH (1981) *Practical optimization*. Academic Press, London
- Godinez HC, Daescu DN, (2009) A second order adjoint method to targeted observations. In: Allen G, Nabrzyski J, Seidel E, van Albada GD, Dongarra J, Sloot PMA (eds) *Computational Science - ICCS 2009*. ICCS, (2009) *Lecture notes in computer science*, vol 5545. Springer, Berlin, Heidelberg, pp 332–341
- Griese R, Vexler B (2007) Numerical sensitivity analysis for the quantity of interest in pde-constrained optimization. *SIAM J Sci Comp* 29:22–48
- Kalnay E, Ota Y, Miyoshi T et al (2012) A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus* 64A:18462
- Karagali I, Hoyer J, Hasager CB (2012) SST diurnal variability in the North Sea and the Baltic Sea. *Remote Sens Environ* 121:159–170
- Kolmogorov AN (1946) On the proof of the method of least squares. *Uspekhi Mat Nauk* 1:57–70
- Langland RH, Baker NL (2004) Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus* 56A:189–201
- Le Dimet FX (1982) A general formalism of variational analysis. Report OK-73091-22-1, CIMMS, Norman
- Le Dimet FX, Talagrand O (1986) Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* 38:97–110
- Le Dimet FX, Ngodock HE, Luong B, Verron J (1997) Sensitivity analysis in variational data assimilation. *J Meteorol Soc Japan* 75:245–255
- Le Dimet FX, Navon IM, Daescu DN (2002) Second-order information in data assimilation. *Mon Weather Rev* 130:629–648
- Lions JL (1968) *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod, Paris
- Marchuk GI, Penenko VV (1978) Application of optimization methods to the problem of mathematical simulation of atmospheric processes and environment. In: *Modelling and optimization of complex systems*. Proceedings of the IFIP-TC7 conference. Springer, Heidelberg, pp 240–252

- Marchuk GI (1995) Adjoint equations and analysis of complex systems. Kluwer, Dordrecht
- Marchuk GI, Dymnikov VP, Zalesny VB (1987) Mathematical models in geophysical hydrodynamics and numerical methods for their realization. Gidrometeoizdat, Leningrad
- Marchuk GI, Agoshkov VI, Shutyaev VP (1996) Adjoint equations and perturbation algorithms in nonlinear problems. CRC Press, New York
- Markov AA (1900) *Ischislenie veroyatnostej*. Imperial Academy of Sciences, St Petersburg
- Navon IM (1998) Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dyn Atmos Oceans* 27:55–79
- Penenko V, Obratsov NN (1976) A variational initialization method for the fields of the meteorological elements. *Soviet Meteorol Hydrol (English transl)* 11:1–11
- Polak E (1997) Optimization: algorithms and consistent approximations. *Appl Math Sci* 124. Springer, New York
- Pontryagin LS, Boltyansky VG, Gamkrelidze RV, Mishchenko EF (1964) The mathematical theory of optimal processes. International Series of Monographs in Pure and Applied Mathematics, vol 55. Pergamon Press, Oxford, New York
- Sasaki Y (1958) An objective analysis based on the variational method. *J Meteorol Soc Japan* 36:77–88
- Schirber S, Klocke D, Pincus R et al (2013) Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *J Adv Model Earth Syst* 5:58–70
- Shutyaev VP, Le Dimet FX, Gejadze IYu et al (2012) Optimal solution error covariance in highly nonlinear problems of variational data assimilation. *Nonlin Processes Geophys* 19:177–184
- Shutyaev V, Le Dimet FX, Shubina E (2017) Sensitivity with respect to observations in variational data assimilation. *Russ J Numer Anal Math Modelling* 32:61–71
- Shutyaev VP, Le Dimet FX, Parmuzin EI (2018) Sensitivity analysis with respect to observations in variational data assimilation for parameter estimation. *Nonlin Processes Geophys* 25:429–439
- Smith PJ, Thornhill GD, Dance SL et al (2013) Data assimilation for state and parameter estimation: application to morphodynamic modelling. *Q J R Meteor Soc* 139:314–327
- Storch RB, Pimentel LCG, Orlande HRB (2007) Identification of atmospheric boundary layer parameters by inverse problem. *Atmos Environ* 41:141–1425
- Sun NZ (1994) Inverse problems in groundwater modeling. Kluwer, Dordrecht
- Talagrand O, Courtier P (1987) Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q J R Meteor Soc* 113:1311–1328
- Thacker WC (1989) The role of the Hessian matrix in fitting models to measurements. *J Geophys Res* 94:6177–6196
- White LW, Vieux B, Armand D, Le Dimet FX (2003) Estimation of optimal parameters for a surface hydrology model. *Adv Water Resour* 26:337–348
- Yuepeng W, Yue C, Navon IM et al (2018) Parameter identification techniques applied to an environmental pollution model. *J Ind Manage Optim* 14:817–831
- Zakharova NB, Agoshkov VI, Parmuzin EI (2013) The new method of ARGO buoys system observation data interpolation. *Russ J Numer Anal Math Modelling* 28:67–84
- Zalesny V, Agoshkov V, Aps R, Shutyaev V et al (2017) Numerical modeling of marine circulation, pollution assessment and optimal ship routes. *J Mar Sci Eng* 5:1–20
- Zhu Y, Navon IM (1999) Impact of parameter estimation on the performance of the FSU global spectral model using its full-physics adjoint. *Mon Weather Rev* 127:1497–1517