Application of the ensemble Kalman filter to environmental data assimilation

Ekaterina Klimova

Institute of Computational Technologies SB RAS, Prosp. Akad. Lavrentyeva, 6, Novosibirsk, 630090, Russia

Abstract. Assessment of the state of the environment with observational data is one of the most urgent modern issues. Such an assessment can be made using forecast models based on data assimilation systems. One of the most popular algorithms for data assimilation is the ensemble Kalman filter, in which the forecast error covariance is estimated using an ensemble of forecasts for perturbed initial fields. Parameter estimation is an important part of atmospheric chemistry modelling. In particular, pollutant emission may be a parameter to be estimated. A single-time estimation based on observations may not give the required accuracy. In this context, the method of ensemble smoothing (EnKS), which uses data from the entire time interval to estimate the parameter at a given time, is becoming increasingly popular. In this paper, we consider a generalization of a previously proposed method called the ensemble π-algorithm, which is a variant of stochastic ensemble Kalman filter. The generalized algorithm is an ensemble smoothing algorithm in which ensemble smoothing is performed for the sample average value and then the ensemble of perturbations is transformed. The proposed algorithm is stochastic. Numerical experiments with a 1-dimensional advection-diffusion model are carried out with the smoothing algorithm proposed in the article.

1. Introduction

Assessment of the state of the environment from observational data is one of the most urgent tasks at present. Such an assessment is made using forecast models based on data assimilation systems. The data assimilation involves the joint use of observations and a mathematical model for optimal estimation of the space-time distribution of the investigated quantities.

There are an enormous number of data assimilation techniques, but, from the point of view of the mathematical formulation of the problem, they all use one of two approaches, variational or dynamical-stochastic. Since observational data are known with errors of a random nature, all methods of assimilation must take into account the statistical properties of measurement errors. The so-called 'noise' of models, which is important for the assessment of the state of the environment, also has a statistical nature. The problem of taking into account the statistical characteristics of forecast and observation errors is naturally solved when applying the dynamic-stochastic approach (the Kalman filter). Numerical realization of the Kalman filter for modern nonlinear models is impossible and, thus, different approximations are currently used. The leading position is taken by the ensemble approach, in which the covariance of forecast errors is estimated using the ensemble of forecasts for perturbed initial fields [1]. The implementation of the ensemble approach also contains technological difficulties, related, in particular, to the large dimension of the matrices under consideration.

The task of parameter estimation is an important part of the problem of atmospheric chemistry modeling. In particular, pollutant emission is the parameter being estimated. The study of the distribution of greenhouse gases in space and time, as well as the assessment of fluxes from the Earth's surface of these gases, is one of the urgent tasks of monitoring the state of the environment.

The parameter of the model can be estimated with successive assimilation in the presence of a large time series of observations with the ensemble Kalman filter. A single-time estimation procedure based on observations may not give the required accuracy. In this connection, the method of ensemble

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

smoothing (EnKS), which uses data from the entire time interval to estimate the values at a given time, is becoming increasingly popular. Now the ensemble Kalman filter and ensemble Kalman smoothing are very popular and are used in atmospheric chemistry modeling. For example, many works are devoted to the estimation of greenhouse gas fluxes. In many works, both the ensemble Kalman filter [2] and ensemble smoothing are applied [3, 4, 5, 6].

In the ensemble Kalman filter, two approaches can be distinguished: the 'stochastic filter' and the 'deterministic filter'. At present a large number of studies are being carried out comparing stochastic and deterministic ensemble filters [1]. Since in the ensemble Kalman filter the optimal estimate is the ensemble average, for problems with the linear model the deterministic and stochastic Kalman filters yield close results.

In this paper, we consider a generalization of the previously proposed ensemble π-algorithm [7], which is a variant of stochastic ensemble Kalman filter. The generalized algorithm is an ensemble smoothing algorithm in which ensemble smoothing is performed for the sample average value, and then the perturbation ensemble is transformed. The proposed algorithm is stochastic. Model numerical experiments for the 1-dimensional advection-diffusion model were carried out. The smoothing algorithm proposed in the article was used in experiments to estimate the model parameter, namely, the emission of a passive impurity.

2. Problems of optimal filtering and smoothing

Let us write the nonlinear dynamical system in the form of the process equation

$$\mathbf{x}_k^t = f(\mathbf{x}_{k-1}^t) + \boldsymbol{\eta}_{k-1}^t \tag{1}$$

and the equation of observation

$$\mathbf{y}_k = h(\mathbf{x}_k^t) + \boldsymbol{\varepsilon}_k^t,$$

where h is, generally speaking, a nonlinear operator, which transforms the forecast values into an observable variable, $\boldsymbol{\eta}_{k-1}^t$ is the vector of 'model noise', $\boldsymbol{\varepsilon}_k^t$ is the observation error vector, $\mathbf{x}_k^t$ is the vector of the estimated variables at time t_k, $\boldsymbol{\varepsilon}_k^t$ and $\boldsymbol{\eta}_{k-1}^t$ are the Gaussian random variables $E[\boldsymbol{\varepsilon}_k^t \left(\boldsymbol{\varepsilon}_k^t\right)^T] = \mathbf{R}_k^t$, $E[\boldsymbol{\eta}_{k-1}^t \left(\boldsymbol{\eta}_{k-1}^t\right)^T] = \mathbf{Q}_{k-1}^t$. We will consider a 'true' value $\mathbf{x}_k^t$.

The task of optimal filtering consists in searching $\min J$ by observations at the time t_K (the final moment of time), the smoothing task is in searching $\min J$ at the time $t_k, (k = 0,...,K)$, where

$$J = E(x_k^a - x_k^t)^T (x_k^a - x_k^t).$$

That is, J is the trace of the estimation error covariance matrix. The solution of the filtration problem in the linear case is the Kalman filter algorithm [8]. In [8], formulas for the optimal smoothing algorithm are also given (in the particular case).

3. Parameter estimation in the data assimilation procedure

Consider the generalization of the equation of the process

$$\mathbf{x}_k^t = f(\mathbf{x}_{k-1}^t, \boldsymbol{\alpha}_{k-1}^t) + \boldsymbol{\eta}_{k-1}^t, \tag{2}$$

and observational data

$$\mathbf{y}_k = h(\mathbf{x}_k^t, \boldsymbol{\alpha}_k^t) + \boldsymbol{\varepsilon}_k^t,$$

where $\boldsymbol{\alpha}_k^t$ is the vector of parameters. We assume that the parameter does not change with time: $\boldsymbol{\alpha}_{k+1}^t = \boldsymbol{\alpha}_k^t$.

Consider the generalized problem of estimating a vector $\mathbf{z} = [\mathbf{x}, \boldsymbol{\alpha}]^T$. Omitting the intermediate calculations, we immediately write down the result of the estimation procedure in general form:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}_{xx} h_x^T (h_x \mathbf{P}_{xx} h_x^T + \mathbf{R})^{-1} [\mathbf{y}_0 - h(\mathbf{x}^f, \boldsymbol{\alpha}^f)] + \mathbf{P}_{x\alpha} h_\alpha^T (h_x \mathbf{P}_{xx} h_x^T + \mathbf{R})^{-1} [\mathbf{y}_0 - h(\mathbf{x}^f, \boldsymbol{\alpha}^f)],$$

$$\boldsymbol{\alpha}^a = \boldsymbol{\alpha}^f + \mathbf{P}_{\alpha x} h_x^T (h_x \mathbf{P}_{xx} h_x^T + \mathbf{R})^{-1} [\mathbf{y}_0 - h(\mathbf{x}^f, \boldsymbol{\alpha}^f)] + \mathbf{P}_{\alpha\alpha} h_\alpha^T (h_x \mathbf{P}_{xx} h_x^T + \mathbf{R})^{-1} [\mathbf{y}_0 - h(\mathbf{x}^f, \boldsymbol{\alpha}^f)].$$

In these formulas $\mathbf{P}_{\alpha x}$ denotes cross-covariance errors of $\mathbf{x}$ and $\boldsymbol{\alpha}$, $\mathbf{P}_{\alpha\alpha}$ is the covariance error matrix of $\boldsymbol{\alpha}$, h_x and h_α are linearized operators at $\mathbf{x}$ and $\boldsymbol{\alpha}$, respectively. If h does not depend on $\boldsymbol{\alpha}$, the estimate of $\mathbf{x}^a$ is carried out according to the same formula as in the usual Kalman filter [9].

In modern studies on the atmospheric chemistry assimilation, this approach is applied to the assessment of greenhouse gas fluxes [2]. At the same time, in most papers the estimated fluxes are considered to be an independent variable, and the model for the propagation of passive gases in the atmosphere is included in the observation operator [3 4, 5, 6]. With this approach, the concentration values themselves are not estimated, and in this formulation the problem will not be optimal.

4. The ensemble Kalman filter. Ensemble smoothing

The derivation of the formulas of the ensemble Kalman filter was first given in Evensen's papers [9]. The stochastic ensemble Kalman filter consists of the ensemble of forecasts $\{\mathbf{x}_k^{f,n}, n = 1, \cdots, N\}$

$$\mathbf{x}_k^{f,n} = f(\mathbf{x}_{k-1}^{a,n}) + \boldsymbol{\eta}_{k-1}^n \tag{3}$$

and the ensemble of analyses $\{\mathbf{x}_k^{a,n}, n = 1, \cdots, N\}$

$$\mathbf{x}_k^{a,n} = \mathbf{x}_k^{f,n} + \mathbf{K}_k \left[\mathbf{y}_k^n + \boldsymbol{\varepsilon}_k^n - h(\mathbf{x}_k^{f,n}) \right]. \tag{4}$$

Ensembles (3) and (4) set a sample of 'true' values, with the sample average value being the optimal estimate, and the deviation from the mean being the ensemble of analysis and forecast errors, respectively. As shown in [8], the probabilistic average is the optimal estimate of the filtration and smoothing problems in the sense of the minimum of the trace of the estimation error covariance matrix. In the ensemble Kalman filter, the ensemble average is an approximation of the probabilistic mean. To implement the ensemble Kalman filter algorithm, it is required to specify an ensemble of observation errors $\{\boldsymbol{\varepsilon}_k^n, n = 1, \cdots, N\}$ as well as an ensemble of forecast errors $\{\mathbf{dx}_k^{f,n} = \mathbf{x}_k^{f,n} - \overline{\mathbf{x}_k^{f,n}}, n = 1, \cdots, N\}$, where $\overline{\mathbf{x}_k^{f,n}} \cong \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_k^{f,n}$ and the model noise ensemble $\{\boldsymbol{\eta}_{k-1}^n, n = 1, \cdots, N\} : E[\boldsymbol{\eta}_{k-1}^n \left(\boldsymbol{\eta}_{k-1}^n \right)^T] = \mathbf{Q}_k$.

The matrix $\mathbf{K}_k$ has the form

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k{}^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k{}^T + \mathbf{R}_k)^{-1},$$

where $\mathbf{P}_k^f$ and $\mathbf{R}_k$ are matrices estimated by the ensemble

$$\mathbf{P}_k^f \triangleq \frac{1}{N-1} \sum_{n=1}^{N} \mathbf{dx}_k^{f,n} \left(\mathbf{dx}_k^{f,n} \right)^T, \mathbf{R}_k \triangleq \frac{1}{N-1} \sum_{n=1}^{N} \boldsymbol{\varepsilon}_k^n \left(\boldsymbol{\varepsilon}_k^n \right)^T,$$

$\mathbf{H}_k$ is the linearized operator $h(\mathbf{x}_k^{f,n})$ with respect to $\overline{\mathbf{x}_k^{f,n}}$:

$$h(\mathbf{x}_k) \cong h(\overline{\mathbf{x}_k^{f,n}}) + \mathbf{H}_k \boldsymbol{\varepsilon}_k^f \quad .$$

The deterministic ensemble Kalman filter (an analysis step) consists of an equation for the mean value

$$\overline{\mathbf{x}_k^{a,n}} = \overline{\mathbf{x}_k^{f,n}} + \mathbf{K}_k \left[\mathbf{y}_k^n - \overline{h(\mathbf{x}_k^{f,n})} \right]$$

and the ensemble of analysis errors with the corresponding covariance matrix satisfies the Kalman filter equation $\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f$ [8].

An algorithm of ensemble Kalman smoothing (EnKS) is proposed in [11]. Let the observation data be given $Y = \{y_k, y_{k+1}, \cdots, y_{k+l}\}$ at times $\{t_k, \cdots, t_{k+l}\}$. If the observation errors are Gaussian random variables and, in addition, the observation errors at different times do not correlate, the EnKS algorithm can be performed sequentially as data is received, using data at a time t_{k+j} to estimate values at a time t_k. At the final time, the results of the ensemble Kalman filter (EnKF) and EnKS are the same [11].

The formula for the ensemble member in the EnKF algorithm has the form

$$\mathbf{x}_k^{a,n} = \mathbf{x}_k^{f,n} + \frac{1}{N-1}\sum_{n=1}^{N} \mathbf{dx}_k^{f,n} \left(\mathbf{dx}_k^{f,n}\right)^T \mathbf{H}^T{}_k (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k{}^T + \mathbf{R}_k)^{-1} \left[\mathbf{y}_k^n + \boldsymbol{\varepsilon}_k^n - h(\mathbf{x}_k^{f,n})\right],$$

where the estimation is carried out at a time t_k according to the observations at the same time.

The formula for the ensemble member in EnKS has the form

$$\mathbf{x}_{k(j)}^{a,n} = \mathbf{x}_{k(j-1)}^{a,n} + \frac{1}{N-1}\sum_{n=1}^{N} \mathbf{dx}_{k(j-1)}^{a,n} \left(\mathbf{dx}_{k+j}^{f,n}\right)^T \mathbf{H}^T{}_{k+j} (\mathbf{H}_{k+j} \mathbf{P}_{k+j}^f \mathbf{H}_{k+j}{}^T + \mathbf{R}_{k+j})^{-1} \left[\mathbf{y}_{k+j}^n + \boldsymbol{\varepsilon}_{k+j}^n - h(\mathbf{x}_{k+j}^{f,n})\right].$$

In this formula the index $k(j)$ means that an estimate is made at time t_k from the data at time t_{k+j}.

As shown in [10], at the time t_k the matrix of ensembles $\mathbf{X}_k = \{x_k^n, n = 1, \cdots, N\}$ after applying the smoothing procedure from the data at times $\{t_{k+1}, t_{k+2}, \cdots, t_{k+l}\}$ can be represented in the form

$$\mathbf{X}_k^{KS} = \left(\prod_{i=k+1}^{k+l} \mathbf{V}_i\right) \mathbf{X}_k^{KF} \ ,$$

where $\mathbf{X}_k^{KF}$ is the result of applying EnKF at the time t_k, $\mathbf{V}_i$ is the matrix of transformations calculated by the ensemble of perturbations and residuals (the difference between the observations and the forecast at the observation point).

In the case where the dynamic system also includes a parameter, the EnKF and EnKS algorithms can evaluate the model parameter in addition to the predicted variable. If the operator $\mathbf{H}_k$ is linear and does not depend on the parameter, the parameter is evaluated independently. In the EnKF algorithm, the parameter estimate has the form

$$\boldsymbol{\alpha}_k^{a,n} = \boldsymbol{\alpha}_k^{f,n} + \frac{1}{N-1}\sum_{n=1}^{N} \mathbf{d}\boldsymbol{\alpha}_k^{f,n} \left(\mathbf{dx}_k^{f,n}\right)^T \mathbf{H}^T{}_k (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k{}^T + \mathbf{R}_k)^{-1} \left[\mathbf{y}_k^n + \boldsymbol{\varepsilon}_k^n - h(\mathbf{x}_k^{f,n})\right].$$

The formula for the ensemble member of the parameter in EnKS is

$$\boldsymbol{\alpha}_{k(j)}^{a,n} = \boldsymbol{\alpha}_{k(j-1)}^{a,n} + \frac{1}{N-1}\sum_{n=1}^{N} \mathbf{d}\boldsymbol{\alpha}_{k(j-1)}^{a,n} \left(\mathbf{dx}_{k+j}^{f,n}\right)^T \mathbf{H}^T{}_{k+j} (\mathbf{H}_{k+j} \mathbf{P}_{k+j}^f \mathbf{H}_{k+j}{}^T + \mathbf{R}_{k+j})^{-1} \left[\mathbf{y}_{k+j}^n + \boldsymbol{\varepsilon}_{k+j}^n - h(\mathbf{x}_{k+j}^{f,n})\right].$$

5. Approaches to optimizing the number of computations in ensemble algorithms for large-dimensional forecast models

The tasks of data assimilation and estimation of parameters in the modeling of the environment are extremely time consuming, requiring large expenditures of computer resources. The use of ensemble algorithms partially allows one to solve this problem, but still the task remains extremely laborious. To carry out the analysis step, algorithms are used to transform the ensemble of forecasts to obtain an ensemble of analyses. One such algorithm is an ensemble π-algorithm [7], which is a stochastic Kalman filter. In this algorithm we perform operations with matrices of the order of the ensemble dimension. This algorithm will be described in more detail in the next section.

To obtain a more easily implemented computer algorithm, the assumption of the stationarity of the time series of forecast errors for an advection - diffusion model with random noise can be used. It can also be assumed that estimation errors in assimilation represent a stationary random sequence. In this case it is possible to use averaging over time instead of averaging over the sample [9]. The idea of estimating the covariance by the time series is proposed in [10], in which the author suggests an approach called 'ensemble optimal interpolation' (EnOI). A similar approach is considered in [12].

6. Ensemble π-algorithm

An ensemble π-algorithm was proposed in [7]. Further, we give the main formulas of this algorithm. We write the equation for the deviation from the mean value $\mathbf{dx}_k^n = \mathbf{x}_k^{a,n} - \overline{\mathbf{x}_k^{a,n}}$, where $\overline{\mathbf{x}_k^{a,n}} = {}^1\!/_N \sum_{n=1}^{N} \mathbf{x}_k^{a,n}$:

$$\mathbf{dx}_k^n = (\mathbf{x}_k^{f,n} - \overline{\mathbf{x}_k^{f,n}}) - \mathbf{P}_k^a \mathrm{H}_k^T \mathbf{R}_k^{-1} \left[\boldsymbol{\varepsilon}_k^n + h(\mathbf{x}_k^{f,n}) - \overline{h(\mathbf{x}_k^{f,n})}\right]. \qquad (5)$$

We express the matrix $\mathbf{P}_k^a$ in terms of $\mathbf{dx}_k^n$:

$$\mathbf{P}_k^a = \frac{1}{N-1}\sum_{n=1}^{N}\mathbf{dx}_k^n\left(\mathbf{dx}_k^n\right)^T. \quad (6)$$

If we substitute the expression (6) into equation (5), we obtain a system of equations with respect to $\mathbf{dx}_k^n$, which can be written in the matrix form as

$$\mathbf{D}^{\mathrm{T}} = \mathbf{F}^{\mathrm{T}} - \mathbf{\Pi}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}, \quad (7)$$

where $\mathbf{D}$ is a matrix of dimension $(L\times N)$, whose columns are vectors $\{\mathbf{dx}_k^n, n=1,\ldots,N\}$, L is the dimension of these vectors, $\mathbf{\Pi}$ is the $(N\times N)$-matrix: $\mathbf{\Pi}^{\mathrm{T}} = \frac{1}{N-1}\mathbf{D}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HF}+\mathbf{E})$, $\mathbf{F}$ is the matrix with columns $\{\mathbf{f}_k^n, n=1,\ldots,N\}$: $\mathbf{f}_k^n = \mathbf{x}_k^{f,n} - \overline{\mathbf{x}_k^{f,n}}$.

It should be emphasized that in the derivation of formula (7) the operator $h(\mathbf{x}_k^{f,n})$ was linearized at $\overline{\mathbf{x}_k^{f,n}}$. Elements of the matrix $\mathbf{\Pi}$ are computed for the matrices $\mathbf{H}_k$ and $\mathbf{R}_k$ for the ensemble of values $\{\mathbf{dx}_k^n, n=1,\ldots,N\}$ and do not depend on the grid node.

For convenience of calculations, the index 'k' will be omitted in the following. It follows from (7) that

$$\mathbf{D}^{\mathrm{T}} = (\mathbf{I}+\mathbf{\Pi}^{\mathrm{T}})^{-1}\mathbf{F}^{\mathrm{T}}, \quad (8)$$

where I is the identity matrix. It is shown in [12] that the matrix $\mathbf{\Pi}$ has the form

$$\mathbf{\Pi}^{\mathrm{T}} = (\mathbf{C}+0.25\mathbf{I})^{\frac{1}{2}} - 0.5\mathbf{I}, \quad (9)$$

$$\mathbf{C} = \frac{1}{N-1}\mathbf{F}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HF}+\mathbf{E}) = \mathbf{C}_1 + \mathbf{C}_2. \quad (10)$$

In these formulas $\mathbf{E}$ is a matrix whose columns are equal to a vector $\boldsymbol{\varepsilon}_k^n$. More detailed calculations are given in [7].

After calculating the matrix $\mathbf{D}$, the covariance matrix $\mathbf{P}_k^a$ and the analysis for the ensemble average are determined:

$$\overline{\mathbf{x}_k^{a,n}} = \overline{\mathbf{x}_k^{f,n}} + \frac{1}{N-1}\mathbf{D}\mathbf{D}^{\mathrm{T}}\mathbf{H}^T\mathbf{R}^{-1}\left[\mathbf{y}_k - \overline{h(\mathbf{x}_k^{f,n})}\right].$$

The ensemble of corresponding analyses in matrix form can be written as

$$\mathbf{X}_k^a = \overline{\mathbf{x}_k^{a,n}} + \mathbf{D},$$

where $\mathbf{X}_k^a$ is the $(L\times N)$ matrix whose columns are vectors $\{\mathbf{x}_k^{a,n}, n=1,\ldots,N\}$.

The ensemble π-algorithm is a stochastic filter in which the ensemble analysis error is obtained using a transformation matrix $(\mathbf{I}+\mathbf{\Pi}^{\mathrm{T}})^{-1}$ (8) independent of the grid node. This approach makes it possible to implement the algorithm locally, while implementing an ensemble π-algorithm requires operations with matrices of the order of the ensemble dimension. As can be seen from the formulas for estimating the parameter in the ensemble algorithm, the matrix $\mathbf{\Pi}$ is the same for estimating the predicted variable and the model parameter. Due to this property, the ensemble π-algorithm can be easily generalized to the case of estimating the parameters of the model.

The covariance matrices of ensemble smoothing have properties analogous to the properties of the ensemble Kalman filter matrices. For this reason, the ensemble π-algorithm can be generalized to the case of ensemble smoothing. Thus, the estimate at time t_k based on the data at time t_{k+j} can be made for the ensemble average value, and the ensemble of perturbations is determined using the

transformation matrix $\mathbf{\Pi}_j$ computed at time t_j. To obtain an ensemble of perturbations, the same matrix $\mathbf{\Pi}_j$ is used to apply the smoothing procedure for a given time interval $\{t_{k+1}, t_{k+2}, \cdots, t_{k+j}\}$. An important property of this algorithm of smoothing is that it will be stochastic as the originally proposed smoothing algorithm [10]. In addition, it can be implemented locally, assessing parameters in a given region.

7. Numerical experiments with a one-dimensional advection-diffusion model

Numerical experiments were carried out with a one-dimensional advection-diffusion model of a passive impurity. The equation of the model has the form

$$\frac{\partial \varphi}{\partial t} + u\frac{\partial \varphi}{\partial x} = b^2 \frac{\partial^2 \varphi}{\partial x^2} + \tilde{g}(x,t), \qquad (11)$$

where φ is the predicted variable, $\tilde{g}(x,t)$ is the source of the passive impurity. The equation was solved by a semi-Lagrangian method, and time-splitting was used. The equation was solved on an interval $(0,1)$, with periodic boundary conditions being specified. 240 grid nodes were considered, $u = 1$, $b^2 = 0.6\times10^{-3}$.

Numerical experiments were carried out with this model using the EnKF and EnKS algorithms based on the ensemble π-algorithm. To implement the algorithm, equation (9) is solved; that is, the square root of the matrix $(\mathbf{C} + 0.25\mathbf{I})$ is calculated. In [7], an approximate estimate of the root of the matrix is proposed. In the present study, an algorithm based on a more general approach is used. One can see from formula (13) that this matrix is nonsymmetric. To calculate the square root of this matrix, an algorithm proposed in [13] based on a triangular Schur decomposition can be used. It can be shown that the real parts of the eigenvalues of the matrix $(\mathbf{C} + 0.25\mathbf{I})$ are positive and, therefore, the algorithm of square root calculation proposed in [13] can be used. In calculating the inverse matrix $(\mathbf{I} + \mathbf{\Pi}^{\mathrm{T}})^{-1}$, it was symmetrized using a standard operation of multiplication by the transpose matrix $(\mathbf{I} + \mathbf{\Pi})(\mathbf{I} + \mathbf{\Pi}^{\mathrm{T}})\mathbf{D}^T = (\mathbf{I} + \mathbf{\Pi})\mathbf{F}^T$. Eigenvalues and eigenvectors of the symmetric matrix $(\mathbf{I} + \mathbf{\Pi})(\mathbf{I} + \mathbf{\Pi}^{\mathrm{T}})$ are used to solve this equation. When calculating the inverse matrix in the π-algorithm, there are no problems caused by the presence of zero eigenvalues and no singular value decomposition is needed, in contrast to the algorithm for a stochastic ensemble Kalman filter proposed in [10].

Consider the finite-difference analog of equation (11):

$$\boldsymbol{\varphi}_{k+1} = \mathbf{A}_k\boldsymbol{\varphi}_k + \mathbf{g}_k,$$

where *k* is the time step number. The given initial values $\boldsymbol{\varphi}_0^t$, $\mathbf{g}_0^t$ were considered 'true values'. To obtain the initial data for the forecast model $\boldsymbol{\varphi}_0^d$, a perturbation was added to the 'true' initial data $\boldsymbol{\varphi}_0^{\mathrm{d}} = \boldsymbol{\varphi}_0^{\mathrm{t}} + \boldsymbol{\delta}, \boldsymbol{\delta} = N(0,s_0)$. *N (a, b)* denotes a random variable distributed according to the normal law with a mathematical expectation equal to *a* and variance equal to *b*. The initial value $\mathbf{g}_0^d$ was assumed to be 0.

To organize numerical experiments, we specified an ensemble of initial fields $\boldsymbol{\varphi}_0^{\mathrm{n}} = \boldsymbol{\varphi}_0^{\mathrm{d}} + \boldsymbol{\delta}^n$, $\delta^n = N(0,s_0), n=1,\cdots,N_{ens}$; $\mathbf{g}_0^{\mathrm{n}} = \mathbf{g}_0^{\mathrm{d}} + \boldsymbol{\delta}_{\mathrm{g}}^{\mathrm{n}}$, $\boldsymbol{\delta}_{\mathrm{g}}^{\mathrm{n}} = N(0,dg_0), n=1,\cdots,N_{ens}$; observations $\mathbf{y}_0 = \boldsymbol{\varphi}_0^{\mathrm{t}} + \boldsymbol{\delta}_0, \boldsymbol{\delta}_0 = N(0,\varepsilon_0)$; ensemble of observations with perturbations $\mathbf{y}_0^n = \mathbf{y}_0 + \boldsymbol{\delta}_0^n, \boldsymbol{\delta}_0^n = N(0,\varepsilon_0), n=1,\cdots,N_{ens}$. The number of the ensemble members is denoted by N_{ens}. The observations were assumed to be known at all grid nodes. The forecast was carried out for N_t = 240 steps in time, the assimilation was carried out at each time step, and the values of the variables $\boldsymbol{\varphi}$ and $\mathbf{g}$ were estimated. The numerical experiments were carried out for $s_0 = \varepsilon_0 = \mathrm{dg}_0 = 0.01$, N_{ens} = 20.

In all numerical experiments $\mathbf{R}=\varepsilon_0^2\mathbf{I}$ was considered. In the analysis at the grid node l, the observations were taken from the interval $(l-id, l+id)$. In the analysis at the grid node l, instead of the matrix $\mathbf{R}$, a matrix $\tilde{\mathbf{R}}=\mathbf{R}\circ\exp[-0.5(\rho_{il}/bc)^2]$ was taken, where ρ_{il} is the distance between the grid node and the observation, '$\circ$' is the sign of the element multiplication. In the experiments the values of $id=5, bc=5\Delta x$ (Δx is the grid spacing) were taken. To prevent divergence of the algorithm, a so-called 'inflation factor' was used. That is, at the forecast step, the ensemble of perturbations multiplied by a constant. In the numerical experiments carried out, the inflation factor was taken equal to 1.3.

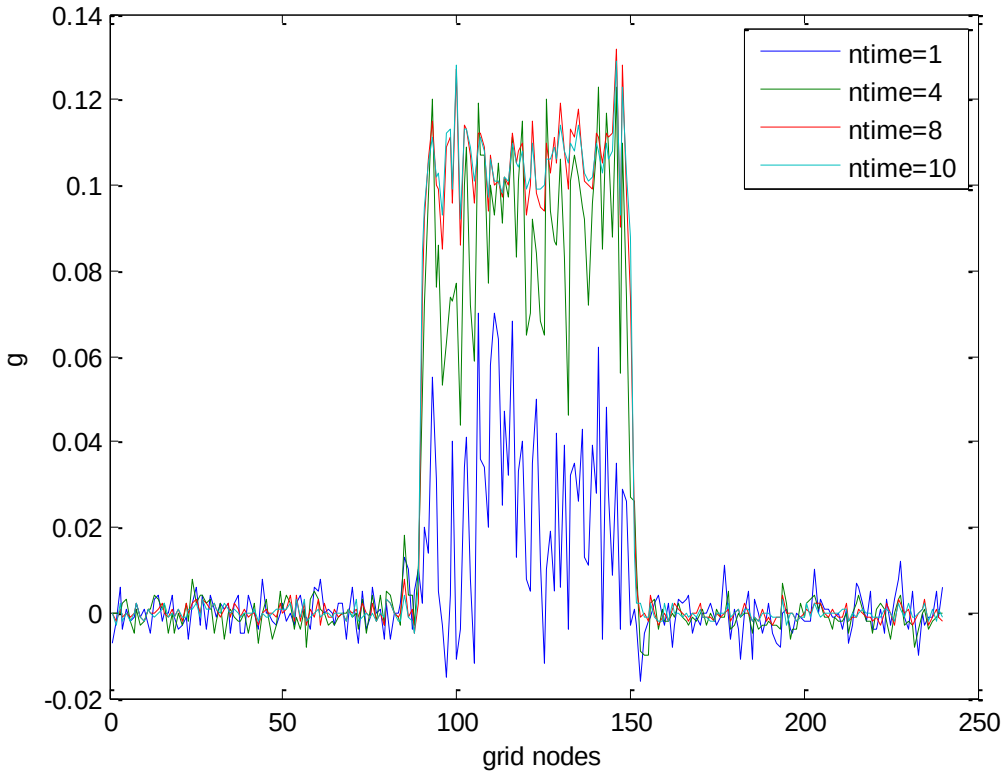

Figure 1. Time behavior of the estimated parameter.

The algorithm of ensemble smoothing was performed with time lags *nt* = 10 and *nt* = 20. That is, to estimate the values at a time t_k, observations were used at times $\{t_k, t_{k+1}, \cdots, t_{k+nt}\}$. The 'true' value of the parameter $\mathbf{g}_k^t$ for all time steps was given as a discrete analog of the function $g_0(x)$, where

$$g_0(x)=\begin{cases} 0.1,\ 0.375\le x\le 0.625 \\ 0,\ 0.375>x\wedge x>0.625. \end{cases}$$

The figures show the results of the numerical experiments. Figure 1 shows the behavior of the estimated parameter $\mathbf{g}$ with respect to time using the EnKF algorithm. The values are given for different times (*ntime*). Figure 2 shows the root-mean square error of the parameter $\mathbf{g}$ estimate for the EnKF and EnKS methods. In this case different values of the time lag *nt* were considered. EnKS1 designates the smoothing algorithm with *nt* = 10, EnKS2 - *nt* = 20. It can be seen from the figures that the assimilation algorithm allows us to estimate a parameter that is not directly measured. At the same time, the accuracy of the estimation at the first time steps increases with the use of ensemble smoothing in comparison with the ensemble Kalman filter.

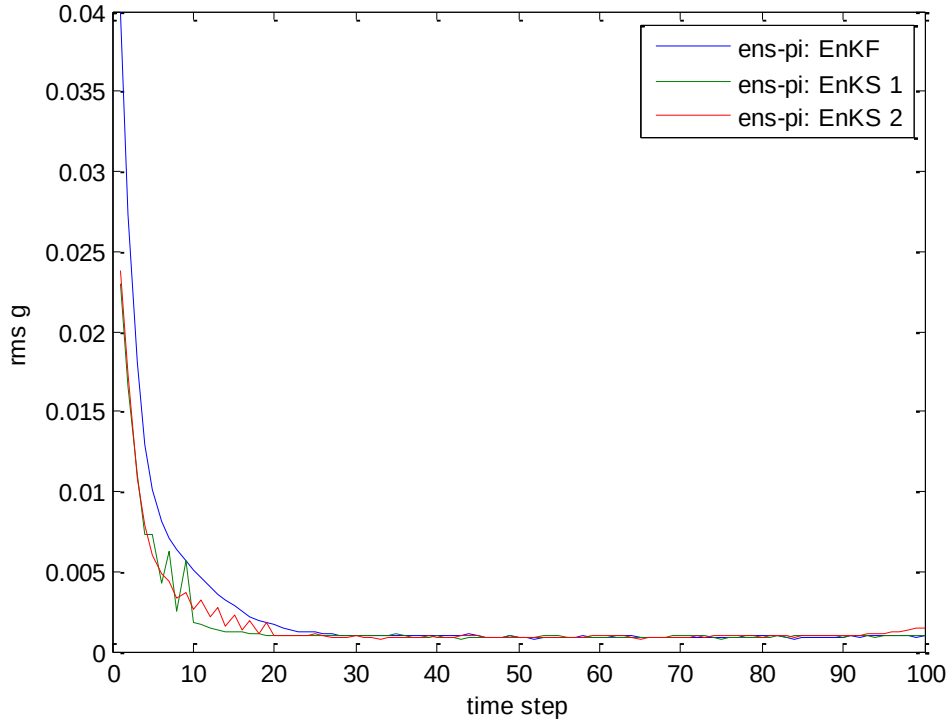

Figure 2. Root-mean square error of parameter estimation.

8. Conclusions

The problem of assessing the state of the environment using observations is currently being solved with the help of data assimilation systems. Atmospheric chemistry models and meteorological fields of wind speed, temperature, etc. are used in these systems. As a mathematical formulation of the problem, the ensemble Kalman filter or the ensemble Kalman smoothing method are increasingly used.

An algorithm based on the algorithms of the ensemble Kalman filter and ensemble Kalman smoothing for estimating the parameters of a model was proposed in this article. It has been shown that the algorithms allow estimating the parameters of the model, while the use of ensemble smoothing makes it possible to improve the accuracy of the estimate by using additional measurement data.

An important application of this algorithm is the problem of estimating emissions of greenhouse gases in the atmosphere. It should be emphasized that solving this large-scale problem requires joint efforts of multiple scientific teams. For example, see papers [4] and [6].

References

[1] Houtekamer H.L and Zhang F 2016 Review of the ensemble Kalman filter for atmospheric data assimilation *Mon. Wea. Rev.* **144** 4489-532

[2] Kang J, Kalnay E, Miyoshi T, Liu J and Fung I 2012 Estimating of surface carbon fluxes with an advanced data assimilation methodology *J. Geophys .Res.* **11** D24101

[3] Feng L, Palmer P I, Bosch H and Dance S 2009 Estimating surface CO2 fluxes from space-borne CO2 dry air mole fraction observations using an ensemble Kalman filter *Atmos. Chem. Phys*. **9** 2619-33

[4] Feng L *et al* 2017 Consistent regional fluxes of CH4 and CO2 inferred from GOSAT proxy XCH4:XCO2 retrievals, 2010-2014 *Atmos. Chem. Phys.* **17** 4781-97

[5] Fraser A, Palmer P I, Feng L, Bosch H, Parker R, Dlugokencky E J, Krummel P B and Langenfelds R L 2014 Estimating regional fluxes of CO2 and CH4 using space-borne observations of XCH4:XCO2 *Atmos. Chem. Phys.* **14** 12883-95
[6] Tsuruta A et al 2017 Global methane emission estimation for 2010-2012 from Carbon tracker Europe-CH4 *Geosci. Model Dev. Discuss.* **10** 1261-87
[7] Klimova E 2012 A suboptimal data assimilation algorithm based on the ensemble Kalman filter *Quart. J. Roy. Meteor. Soc.* **138** 2079-85
[8] Jazwinski A H 1970 *Stochastic Processes and Filtering Theory* (New York: Academic Press)
[9] Krasovskii A A, Beloglazov I N and Chigin G P 1979 *Theory of Correlation-Extreme Navigation Systems* (Moscow: Nauka)
[10] Evensen G 2009 *Data Assimilation. The Ensemble Kalman Filter* (Berlin Heideberg: Spriger-Verlag)
[11] Evensen G, van Leeuwen P J 2000 An ensemble Kalman smoother for nonlinear dynamics *Mon. Wea. Rev.* **128** 1852-67
[12] Klimova E G, Platov G A, Kilanova N V 2014 Development of a system for the assimilation of environmental data based on the Kalman ensemble filter *Computational technologies* **19** 27-37
[13] Björck À, Hammarling S 1983 A Schur method for the square root of a matrix *Linear algebra and its applications* **52/53** 127-40