An algorithm for estimating greenhouse gas fluxes using satellite data for a global transport and diffusion model

M V Platonova\textsuperscript{1} and E G Klimova\textsuperscript{1}  
\textsuperscript{1}The Federal Research Center for Information and Computational Technologies, Novosibirsk, Pr. Lavrentieva 6, Russia, 630090  
E-mail: gumoznaya@gmail.com, klimova@ict.nsc.ru

Abstract. The task of estimating the greenhouse gas fluxes in the atmosphere from satellite data is relevant. This problem is solved based on a data assimilation system using a model of transport and diffusion of an impurity in the atmosphere. The paper discusses an approach to solving the problem of estimating the greenhouse gas fluxes based on the use of the ensemble Kalman filter for the global model of transport and diffusion. The paper proposes a numerical algorithm based on the decomposition of the model domain. The originality of this work is the application of the local ensemble Kalman filter algorithm. The case of this study when the greenhouse gas fluxes are considered as a variable is discussed. Here are the results of the simulated numerical experiments on the implementation of the step of analyzing the data assimilation algorithm in the case of the global model of transfer and diffusion.

1. Introduction

At present, estimating a flux of greenhouse gases in the atmosphere using satellite data is a currently topical issue. This problem is solved based on a data assimilation system using a model of transport and diffusion of an impurity in the atmosphere [1-6]. For modern models with high spatial resolution, this algorithm is difficult to implement because of the high dimensionality of vectors of the predicted variables and observational data [7-8].

This paper describes an approach to solving the problem of estimating the greenhouse gas fluxes, based on the use of the ensemble Kalman filter for the global model of transport and diffusion. In our research, we propose a numerical algorithm based on the decomposition of the model domain. The novelty of the algorithm is the usage of the local ensemble transport Kalman filter locally. The greenhouse gas fluxes from the Earth's surface are considered as an estimated variable. The study presents the results of simulated numerical experiments on the implementation of the step of analyzing the data assimilation algorithm in the global transport and diffusion model.

2. An algorithm for finding estimates of the greenhouse gas fluxes using satellite data

There are many studies devoted to estimating greenhouse gas fluxes using data assimilation methods. All of these studies use an approach called inverse modeling. The concept of this approach is the search for the flux value, using which a forecast is obtained that most precisely approaches the observational data. In a series of publications [1-5] conducted by an international team of authors, the ensemble Kalman filter is used. In these investigations, the Earth’s surface is divided into squares of an equal area, and, accordingly, the average value of the flux over the subdomain is to be estimated.

Below we will consider an algorithm for estimating the average greenhouse gas fluxes in a subdomain. We will obtain an estimate from satellite observations in a given time interval, as in [1-5].
The time interval for which the observational data used are called an "assimilation window". There are various options for the implementation of the algorithm: estimation of the concentration of the greenhouse gases and their fluxes according to observational data [6], and the search for the flux estimates using information about the concentration in observations [1-5]. In the first version, fluxes are an additional parameter and the estimation is obtained using the ensemble Kalman filter. This is an estimate for an extended state vector, which includes the concentration and the greenhouse gases fluxes [6-8]. The second version is easier implemented and has been considered in numerous published works [1-5]. In this paper, the second version of the algorithm for the search for the flux estimates using the information on the concentration of passive impurities in observations is implemented.

Estimating the average fluxes in subdomains $x^a$ with the observations $y_0$ and the forecast $x_f$ is obtained using the standard Kalman filter formulas shown in equations (1), (2) (analysis step) [9-10]:

$$x_a = x_f + K[y_0 - H(x_f)],$$

$$K = P^f H^T (HP^f H^T + R)^{-1}.$$  \hspace{1cm} (1)

$$P^f = Dx_f (Dx_f)^T, \hspace{1cm} K = Dx_f (HDx_f)^T [HDx_f (HDx_f)^T + R]^{-1}.$$  \hspace{1cm} (2)

The ensemble of perturbations of the estimated quantity for the operation of the ensemble Kalman filter [1-5] $Dx_f = \frac{1}{\sqrt{N}}[dx_f^1, \ldots, dx_f^N]^T$ is defined. The matrix $P^f$ is estimated according to the ensemble

$P^f = Dx_f (Dx_f)^T,$

$K = Dx_f (HDx_f)^T [HDx_f (HDx_f)^T + R]^{-1}.$

The operator $H$ includes the simulated forecast at the observation time, interpolation from grid nodes up to observation points, and averaging satellite data vertically with known coefficients (in the case of satellite data). Let the equation of a change in fluxes with time (forecast step) have the form

$$x_f^{n+1} = x_f^n.$$

Observational data on the concentration of greenhouse gases at the time instant $t_n$ is presented as

$$y_0^n = H[f(q_f^n) + x_f^n] + \epsilon_0,$$

where $f$ is a model operator describing a change in the concentration $q_f$ over time. It is included in the observation operator.

It is necessary to specify an ensemble of the errors $Dx_f$ for the implementation of the analysis step. We write down the observation operator in the form

$$H(dx_f^i) = H_i[f(q_f^n) + x_f^n + dx_f^i - f(q_f^n) + x_f^n] = H_i dx_f^i,$$

where $H_i$ is the interpolation operator to the observation point. In the case of satellite data on the greenhouse gas concentrations, the data contain information on the average value in the vertical column: $y_0 = \sum_{l=1}^L \beta_l q_f^0$ is the sum with weights of values at $L$ levels vertically. This summation is also included in the observation operator.

3. A deterministic version of the local data assimilation algorithm

The data assimilation algorithm comprises a sequence of the simulated forecast step and the analysis step. The analysis step is an optimal estimate based on the observational data and forecast. Below we
consider a version of the ensemble Kalman filter, in which the covariance matrix of prediction errors is set at the initial instant of time based on their ensemble of the forecast errors. This covariance matrix will not change in time.

Assume we are given an ensemble of the forecast errors $\mathbf{Dx}_f = \frac{1}{\sqrt{N}}[\mathbf{dx}_f, \cdots, \mathbf{dx}_f]$. The covariance matrix of forecast errors can be represented as $\mathbf{P}_f = \mathbf{Dx}_f \mathbf{Dx}_f^T$. The step of analyzing the data assimilation procedure has the form of (1) - (2). Next, we consider the implementation of the analysis step algorithm based on the deterministic LETKF algorithm [10-11]:

$$
x_a = x_f + \mathbf{Dx}_f \mathbf{\hat{P}}_a \left( \mathbf{H} \mathbf{Dx}_f \right)^T \mathbf{R}^{-1} (y_o - \mathbf{H} x_f), \tag{3}
$$

Then the analysis algorithm becomes local, i.e. it can be implemented independently for any grid node [10-11], and the operations are performed with the ensemble dimension matrices. Such an algorithm can be implemented in subdomains as well as in a simplified deterministic version, in which the ensemble of perturbations is not updated according to the model.

### 4. The general scheme of the numerical experiment

Below we have identified the main stages of the implementation of the algorithm for estimating the greenhouse gas fluxes from the Earth's surface for the global transport and diffusion model using observational data:

1. The assessment is made for a given time interval in which the fluxes are assumed to be constant. In the algorithms, where large amounts of satellite data are processed to estimate epy greenhouse gas fluxes, it is customary to have an estimate for a given time interval (for example, a week), assuming that the values of the fluxes are constant during this period. In this case, the data vector contains all the observational data this period, and the forecast from the transport and diffusion model is included in the interpolation operator.

2. The Earth's surface is divided into regions in which the estimate is obtained. The estimation of the flux values is carried out separately for the given subdomains because a local algorithm is used in the analysis step. This algorithm can be used for each grid node independently. The same approach is used in [1–5]; however, these published works describe an analysis algorithm that is not local.

3. We set the initial (climatic) values of the fluxes in the regions and the covariance matrix of observation errors and the model.

4. Observation data are divided into blocks by time intervals, and then by subdomains.

5. The transfer and diffusion model calculates the forecast for a given initial value of concentrations and fluxes over some time with a given time step. Then, interpolation is made to the observation point and the point at a time at which the observation is made.

6. Fluxes are estimated using formulas (3) - (4).

### 5. Numerical experiments with model data for the analysis step

Numerical experiments were conducted with the model data for a local deterministic data assimilation algorithm. In the experiments, we sought flux estimates using observational data on the concentration of passive impurities. We have implemented the analysis step with different input parameters. A comparative analysis of the accuracy of the estimates obtained is made.

In the process of model experiments, it is believed that prognostic data and observational data are known on the Earth’s surface. The model area is an array of 144x144 grid nodes. The model area is divided into subdomains of the same area i.e. 1000x1000 km. The analysis is locally carried out for each subdomain separately. It is possible to locally carry out the analysis step due to the assimilation
algorithm that we use and its locality properties. As a result, a general assessment of fluxes within each subdomain is made from the estimates obtained.

We have carried out numerical experiments only at the stage of analysis with model data. In this paper, we have not implemented the forecast step. At the beginning of the experiments, the state is set to "true" - we accept it as the actual state of the system. Perturbation is added to "true" for data modeling and forecast modeling. When generating the forecast (the first approximation) and observational data, the disturbance errors were matched with realistic relationships. In the model area, the concentration values were set everywhere and observational data were simulated at all grid points.

The experiment is organized as follows: "true" is not known, it is stored to determine the accuracy of the estimate obtained. At the initial instant of time, the first approximation is set by adding a disturbance to "true" and thus the forecast of the concentration of the passive impurity is simulated: \( x_f \). Then observational data are simulated: disturbances are added to "true", observational data \( y_0 \) contain information about emission. Further, according to formulas (3) - (4), the analysis step is implemented. The resulting estimate \( x_a \) is an estimate of unknown fluxes of the greenhouse gases based on observational data on the concentration of passive pollutants in the atmosphere. An optimal estimate of an unknown parameter is found by calculating the weight coefficients and cross-covariance matrices.

We set error matrices for observations and prediction for working with ensembles. All generated random variables have a normal distribution, zero mean, and a predetermined variance. As errors of the first approximation (forecast at the initial instant of time), we used random variables with a dispersion of 0.08. Experiments were conducted with different values of the variance of observation errors.

The ensemble of concentration errors consists of independent random variables with a distribution of \( N(0,0.1) \). Here and below, \( N(0,\delta) \) is the normal distribution with a mean equal to zero and a given dispersion \( \delta \). The ensemble of emission errors consists of independent random variables with a distribution of \( N(0,0.01) \).

The standard error (root-mean-square error) of the estimate is considered to characterize the quality of the algorithm: \( \text{rms} = \|x_1 - x_a\|^2 \). The second norm of the vector is taken as the norm, \( x_1- "true" \) value of flows, \( x_a \) is the resulting estimate of the greenhouse gas fluxes from the analysis step. Table 1 shows the results of numerical experiments with different distributions of observation errors:

| Distribution of observation errors | \( N(0,0.01) \) | \( N(0,0.001) \) | \( N(0,0.0001) \) |
|-----------------------------------|----------------|----------------|----------------|
| Root Mean Square error            | 1.2457e-04     | 8.1350e-05     | 5.0408e-06     |

Table 1. Results of numerical experiments with model data.

The results of the model experiment show that the data assimilation algorithm makes it possible to reconstruct the flux values from the observational data on the concentration of passive impurities. In this case, the accuracy of restoration depends on the accuracy of the information used.

6. Conclusion
In this paper, we consider the problem of estimating the greenhouse gas fluxes from observational data. This task is currently being solved using data assimilation systems. Models of distribution of impurities in the atmosphere and meteorological fields of the wind speed, temperature, etc. are used to simulate the transport and diffusion of greenhouse gases. The task is very important and relevant to modern realities. A paper highlights the main problems that arise when solving this problem. An approach to solving this problem is presented, based on the analysis of what has been done in the world in this area and the author's experience in solving the problem of data assimilation. The main
features of the algorithm are its locality, the possibility of considering a deterministic version of the algorithm, and the estimation of fluxes from concentration data that are not directly measured.

This paper discusses an efficient flux estimation algorithm based on the ensemble Kalman filter. The results of numerical experiments are presented for testing the analysis step of the algorithm with model data: experimental versions when they search for estimates of the greenhouse gas fluxes are carried out according to the observational data on the concentration of passive impurities.

References
[1] Feng L, Palmer P I, Bosch H and Dance S 2009 Estimating surface CO$_2$ fluxes from space-borne CO$_2$ dry air mole fraction observations using an ensemble Kalman filter Atmospheric chemistry and physics 9 2619-33
[2] Feng L, Palmer P I, Yang Y, Yantosca R M, Kawa S R, Paris J-D, Matsueda H and Machida T 2011 Evaluating a 3-D transport model of atmospheric CO$_2$ using ground-based, aircraft, and space-borne data Atmospheric chemistry and physics 11 2789-2803
[3] Feng L, Palmer P I, Parker R J, Deutscher N M, Feist D G, Kivi R, Morino I and Sussmann R 2016 Estimates of European uptake of CO$_2$ inferred from GOSAT XCO$_2$ retrievals: sensitivity to measurement bias inside and outside Europe Atmospheric chemistry and physics 16 1289-1302
[4] Feng L et al. 2017 Consistent regional fluxes of CH$_4$ and CO$_2$ inferred from GOSAT proxy XCH$_4$: XCO$_2$ retrievals 2010-2014 Atmospheric chemistry and physics 17 4781-97
[5] Fraser A, Palmer P I, Feng L, Bösch H, Parker R, Dlugokencky E J, Krummel P B and Langenfelds R L 2014 Estimating regional fluxes of CO$_2$ and CH$_4$ using space-borne observations of XCH$_4$: XCO$_2$ Atmospheric chemistry and physics 14 12883-95
[6] Kang J, Kalnay E, Miyoshi T, Liu J and Fung I 2012 Estimating of surface carbon fluxes with an advanced data assimilation methodology Journal of geophysical research 117 24101
[7] Klimova E G 2017 Methods of estimation of greenhouse gases concentration in the atmosphere using observations and transport and diffusion model, based on the ensemble Kalman filter All-Russian Conf "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes, SDM (Novosibirsk, Russian Federation); 29 August 2017-31 August 2017; 2033 32851 pp 191-195
[8] Klimova E G 2018 Application of ensemble Kalman filter in environment data assimilation IOP Conf Series: Earth and Environmental Science 211 012049
[9] Evensen G 2009 Data assimilation. The ensemble Kalman filter (Berlin, Heideberg:Springer-Verlag) p 307
[10] Houtekamer P L and Zhang H F 2016 Review of the ensemble Kalman filter for atmospheric data assimilation Monthly Weather Review 144 4489-4532
[11] Hunt B R, Kostelich E J and Szunyogh I 2007 Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter Physica D 230 112-126