Determining the likely localization of methane sources using forecast time series and satellite data

Marina V. Platonova^{1,2}, Ekaterina G. Klimova^{1,2}

¹Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia ²Novosibirsk State University, Novosibirsk, Russia

Abstract

The paper is devoted to the topical problem of determining the sources of methane from observational data. An algorithm based on the statistical optimization method used to estimate a time constant parameter is considered. To implement the algorithm, a variant of ensemble smoothing is used, which is an optimal estimate of the desired parameter based on observational data and forecast for a given time interval. This paper presents the implementation of the algorithm for real observational and forecast data, the results of a three-dimensional transport and diffusion model are taken as a mathematical model, and satellite measurement data are used as observational data. Methane fluxes are estimated in subdomains of the Earth's surface for specified time intervals. The paper contains a mathematical formulation of the problem, a scheme for its numerical implementation. The results of numerical experiments with model and real data are presented.

Keywords

Methane sources, forecast, satellite data.

1. Introduction

The task of searching for methane sources is modern and urgent; interest in solving this problem has been growing more and more recently. International interest is associated with various political and economic factors, including a carbon tax. A common practice is when data assimilation systems are used for such a task [1, 2, 3]. In this work, the problem of probabilistic localization of methane fluxes is solved using the system of data assimilation for the model of transport and diffusion in the atmosphere [4, 5, 6]. The use of a mathematical model and satellite data to solve the problem is currently a relevant method using data assimilation as a basis. In the case of modern models with high spatial resolution, the algorithm is very laborious due to the high dimension of the vectors of predicted variables and observational data [7, 8].

This article discusses an approach to solving this problem based on the decomposition of the model area. Methane sources on the Earth's surface are considered as an estimated parameter. The authors present the results of model numerical experiments with real data on the implementation of a part of the analysis step the data assimilation algorithm in the case of a global model of transport and diffusion.

🛆 gumoznaya@gmail.com (M. V. Platonova); klimova@ict.nsc.ru (E. G. Klimova)

SDM-2021: All-Russian conference, August 24-27, 2021, Novosibirsk, Russia

^{© 02021} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

2. Using time-series forecasts and satellite data to determine the likely localization of methane sources

2.1. The ensemble Kalman filter for methane fluxes estimation

Following [1, 2, 3, 4, 5], we will make an estimate based on the data of satellite observations from a given time interval. Methane flux values will be estimated from observational methane concentrations. This version of the algorithm is considered in many modern works [1, 2, 3, 4, 5].

The estimation of the values of the flow averages over the subdomains from the observational and forecast data is carried out according to the standard formula of the Kalman filter (analysis step) [9, 10]:

$$x_a = x_f + K [y_0 - H(x_f)],$$
(1)

$$K = P^{f} H^{T} \left(H P^{f} H^{T} + R \right)^{-1}.$$
 (2)

Here x_a — estimation of the average flow over the subdomains, x_f is the forecast by the model and y_0 is the observational data. To implement the ensemble Kalman filter, it is necessary to specify an ensemble of perturbations of the estimated quantity [1, 2, 3, 4, 5]:

$$Dx_f = rac{1}{N} \left[dx_1^f, \dots, dx_N^f
ight]^T,$$

where N is the size of the ensemble. Matrix P^{f} is evaluated by the ensemble:

$$P^{f} = Dx_{f} \left(Dx_{f} \right)^{T},$$
$$K = Dx_{f} \left(HDx_{f} \right)^{T} \left[HDx_{f} \left(HDxf \right)^{T} + R \right]^{-1}$$

Operator H includes model prediction at the time of observation, interpolation from grid nodes to observation points. When working with satellite data, the operator includes vertical averaging with known coefficients (using the average kernel). We will assume that the equation of changes in flows from time to time is constant and flows do not change (the forecast step is a change in the variable at the forecast step according to the model), i.e.:

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

n is the number step of the time.

Observational data on the concentration of greenhouse gases at the moment of time t_n can be represented in the form:

$$y_0^n = H\left[f(q_t^n) + x_t^n\right] + \varepsilon_0,$$

where f is the operator of the model, i.e. the model describing the time variation of the concentration q_t^n is included in the observation operator. To implement the analysis step, you need to specify an ensemble of errors. We write the observation operator in the form:

$$H(dx_{f}^{i}) = H_{1}\left[f(q_{t}^{n}) + x_{t}^{n} + dx_{f}^{i} - f(q_{t}^{n}) + x_{t}^{n}\right],$$

where H_1 is the operator of the interpolation to the observation point. In the case of satellite data on greenhouse gas concentrations, the data contains information about the mean in a vertical column:

$$y_0 = \sum_{l=1}^{L} \beta_l q_l^0,$$

the sum with the weights of the values at L levels along the vertical.

2.2. Deterministic version of the LETKF algorithm

Consider a deterministic version of the local data assimilation algorithm. The data assimilation algorithm consists of alternating forecasting steps and model analysis steps. The analysis stage is an optimal estimate based on observational and forecast data [9, 11]. We will consider a variant of the ensemble Kalman filter, in which the covariance matrix of the forecast errors is specified at the initial moment and considered constant over time.

Let us define an ensemble of forecast errors:

$$Dx_f = \frac{1}{N} \left[dx_f^1, \dots, dx_f^N \right]^T$$

The covariance matrix of forecast errors can be represented as $Dx_f(Dx_f)^T$. The analysis step of the assimilation procedure has the form (1)–(2). Consider the implementation of the analysis step algorithm based on the deterministic LETKF algorithm [10, 11]:

$$x_a = x_f + Dx_f P^a (H D x_f)^T R^{-1}, (3)$$

$$P^{a} = \left[(N-1)I + (HDx_{f})^{T} R^{-1} HDx_{f} \right]^{-1}.$$
(4)

In this case, the analysis algorithm becomes local. It can be implemented for any grid node independently [10, 11]; in this case, operations are performed with the ensemble dimension matrices. Additional localization can be performed by element-wise multiplication of the matrix R by a function that decreases with distance. This algorithm can be implemented in a simplified deterministic version in which the ensemble of disturbances is not recalculated according to the model.

Using the formulas of this algorithm, following [1, 2, 3, 4, 5], we will make an estimate based on the data of satellite observations for a given time interval. Methane flux values will be estimated from observational data using methane concentration information. This version of the algorithm is considered in many modern works [1, 2, 3, 4, 5].

2.3. Implementation of the analysis step for real observational and forecast data

In the process of implementing the algorithm for finding the estimate of methane sources, it is possible to distinguish several stages. Note that the estimate is carried out for a given time interval, in which the fluxes are considered constant. In algorithms for processing large amounts of satellite data, it is customary to evaluate for a given time interval (for example, a

week), assuming that the fluxes values are constant during this period. In this case, the data vector contains all observational data from this period, and the forecast from the transport and diffusion model is included in the interpolation operator.

It should be noted that the use of the deterministic variant is possible only in experiments using a 6-hour time interval.

2.4. Surface area decomposition

Due to the specificity of the used algorithm, it is possible to perform the analysis step locally. We assume that the data is known on the surface of the Earth. We have divided the entire surface area of the Earth into subdomains of approximately 1000×1000 km. Further, the work of the formulas of the algorithm will be carried out within each subdomain.

2.5. Specifics of using the data of the MOZART model

The data assimilation algorithm consists of a forecast step and an analysis step. For the forecast step, we used the results of calculations of the mathematical model MOZART-4. This is the MOZART-4 Global Chemical Transport Model, the source code of which is freely available. The results of calculations of this model were provided by colleagues Anatoly A. Lagutin and Egor Yu. Mordvin [12]. This model is autonomous, it only needs meteorological data for the duration of the simulation. The model has a spatial resolution of $2.8 \times 2.8^{\circ}$. The MOZART-4 model has 28 levels in height (from the surface of the Earth to a height of ~ 2 hPa). This model includes 85 types of gases and 12 aerosol components. The content of each component of the model in the atmosphere is found by solving the mass conservation equation taking into account adventive, convective, and diffusion transfer. In addition, there is accounting for the emission component of the underlying surface, aerosols, photochemical reactions and wet/dry deposition. In the MOZART-4 model we used for our calculations, we used a standard set of chemical reactions.

2.6. Specifics of using AIRS data

We used data from the AIRS satellite as observational data for the analysis step. The Atmospheric Infrared Sounder AIRS, launched into low-Earth orbit on May 4, 2002 aboard NASA's Aqua satellite, provides data for monitoring the Earth's atmosphere. AIRS is one of six instruments aboard the Aqua satellite, which is part of NASA's Earth Observing System. AIRS measures all the primary greenhouse gases including carbon dioxide, the largest source of anthropogenic greenhouse gas, carbon monoxide, methane, and ozone.

The main AIRS products used in CH4 extraction are those recoverable using both AIRS and AMSU. The AIRS instrument on NASA's Aqua satellite orbits the Earth from pole to pole approximately fifteen times a day. The AIRS data information products used are divided into series 6-minute granules, and each granule is a file, there are 240 granules for each day. All AIRS information products that were used for the calculations satisfy the measurements by the method of least squares. To properly compare the satellite observations with model simulations, the model data should be convolved using the averaging kernels [12].

3. Model experiments with real data

We divide the surface of the globe into regions for which we will calculate the estimate. Since a local algorithm is used at the analysis step, which can be used for each grid node independently, the procedure for estimating the emission values is carried out separately for the specified subdomains. The same approach is used in works [1, 2, 3, 4, 5], but these works use an analysis algorithm that is not local.

The transport and diffusion model calculates a forecast based on the given initial values of concentrations and fluxes over a time period with a given time step, then interpolation is performed to the observation point and the time at which the observation is made. Further, the search for the estimate is carried out according to formulas (3)-(4).

Numerical experiments were carried out with model data and with real data for the described local deterministic data assimilation algorithm.

It is believed that predictive and observational data are known at the surface of the Earth. The entire area is divided into regions, the analysis is carried out locally for each sub-area separately. The possibility of performing an analysis step locally is due to the properties of the assimilation algorithm used. From the estimates obtained within each sub-area, a general emission estimate is compiled. Over the entire model area, the concentration value was set, observational data were modeled.

The results of the MOZART-4 model and observational data were taken at the same time. We have interpolated the received data into one grid.

We set up observation and forecast error matrices to work with ensembles. All generated random variables have a normal distribution, zero mean and specified variance. As errors in the first approximation (forecast for the initial moment), random variables with a variance of 0.08 were taken.

Figure 1 shows the spatial distribution of the distribution of the methane mixture ratio at an altitude of \sim 1500 m at 00.00 UTC on August 01, 2002.



Figure 1: Spatial distribution of the distribution of the ratio of the mixture of methane at an altitude of \sim 1500 m at 00:00 UTC on August 01, 2002.



Figure 2: Results of model experiments with real data: a -model methane emission in the Siberia subregion; b - an estimate of the localization of methane sources.

To simulate methane emissions, we have chosen one subregion with a size of 1000×1000 km. In this case, the emission is modeled in the region — Siberia (from $50^{\circ}36$ ' to $58^{\circ}48$ ' north latitude and from $78^{\circ}24$ ' to $86^{\circ}48$ ' east longitude). I would like to note that the order of the modeled emission is 0.1e-08.

Data assimilation was carried out at the analysis step for one for one 6-hour interval (from 00:00 UTC August 1, 2002 to 6:00 am August 1, 2002). Results of model experiments with real data are presented on Figure 2. Figure 2, a shows a model methane release in one sub-region (Siberia). Figure 2, b shows an estimate of the localization of methane sources obtained from data on methane concentration.

For a qualitative assessment of the operation of the algorithm for the probabilistic localization of methane sources, let us compare the deviations of the geographic coordinates of the found sources in comparison with the coordinates of the modeled emission. The maximum deviation is $\pm 2^{\circ}$, which is comparable to the degree measure of the grid step in the space of the Mozart-4 model.

Acknowledgments

The authors of the article express their deep gratitude to colleagues who provided materials for the work: professor Anatoly A. Lagutin and assistant professor Egor Yu. Mordvin.

References

 Feng L., Palmer P.I., Bösch H., Dance S. Estimating surface CO2 fluxes from space-borne CO2 dry air mole fraction observations using an ensemble Kalman filter // Atmospheric Chemistry and Physics. 2009. Vol. 9. P. 2619–2633.

- [2] Feng L., Palmer P.I., Yang Y., Yantosca R.M., Kawa S.R., Paris J.-D., Matsueda H., Machida T. Evaluating a 3-D transport model of atmospheric CO2 using ground-based, aircraft, and space-borne data // Atmospheric Chemistry and Physics. 2011. Vol. 11. P. 2789–2803.
- [3] Feng L., Palmer P.I., Parker R.J., Deutscher N.M., Feist D.G., Kivi R., Morino I., Sussmann R. Estimates of European uptake of CO2 inferred from GOSAT XCO2 retrievals: sensitivity to measurement bias inside and outside Europe // Atmospheric Chemistry and Physics. 2016. Vol. 16. P. 1289–1302.
- [4] Feng L. et al. Consistent regional fluxes of CH4 and CO2 inferred from GOSAT proxy XCH4: XCO2 retrievals 2010–2014 // Atmospheric Chemistry and Physics. 2017. Vol. 17. P. 4781–4797.
- [5] Fraser A., Palmer P.I., Feng L., Bösch H., Parker R., Dlugokencky E.J., Krummel P.B., Langenfelds R.L. Estimating regional fluxes of CO2 and CH4 using space-borne observations of XCH4: XCO2 // Atmospheric Chemistry and Physics. 2014. Vol. 14. P. 12883–12895.
- [6] Kang J., Kalnay E., Miyoshi T., Liu J., Fung I. Estimating of surface carbon fluxes with an advanced data assimilation methodology // Journal of Geophysical Research. 2012. Vol. 116. D24101.
- [7] Klimova E.G. An efficient algorithm for stochastic ensemble smoothing // Siberian Journal of Numerical Mathematics. 2020. Vol. 23. No. 4. P. 381–393.
- [8] Klimova E.G. Application of ensemble Kalman filter in environment data assimilation // IOP Conference Series: Earth and Environmental Science. 2018. Vol. 211. P. 012049.
- [9] Evensen G. Data assimilation. The ensemble Kalman filter. Berlin, Heideberg: Spriger-Verlag, 2009.
- [10] Houtekamer P.L., Zhang H.F. Review of the ensemble Kalman filter for atmospheric data assimilation // Monthly Weather Review. 2016. Vol. 144. P. 4489–4532.
- [11] Hunt B.R., Kostelich E.J., Szunyogh I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter // Physica D. 2007. Vol. 230. P. 112–126.
- [12] Mordvin E.Yu., Lagutin A.A. Methane in the atmosphere of Western Siberia. Barnaul: Azbuka, 2016.