

**METHODS OF ESTIMATION OF GREENHOUSE GASES
CONCENTRATION IN THE ATMOSPHERE USING OBSERVATIONS AND
TRANSPORT AND DIFFUSION MODEL, BASED ON THE ENSEMBLE
KALMAN FILTER**

Ekaterina G. Klimova

Institute of Computational Technologies SB RAS, Novosibirsk, Russia

Abstract

Evaluation of the environmental state using observations is one of the most important tasks at present. This assessment is based on data assimilation systems with the involvement of forecast models. The article presents a review of modern methods of greenhouse gases data assimilation including satellite data based on the ensemble Kalman filter application. The main attention is paid to evaluating greenhouse gas fluxes in the data assimilation procedure.

Keywords: data assimilation, satellite data, greenhouse gases, fluxes estimation

МЕТОДЫ ОЦЕНКИ КОНЦЕНТРАЦИИ ПАРНИКОВЫХ ГАЗОВЫХ СОСТАВЛЯЮЩИХ В АТМОСФЕРЕ ПО ДАННЫМ НАБЛЮДЕНИЙ И МОДЕЛИ ПЕРЕНОСА И ДИФФУЗИИ, ОСНОВАННЫЕ НА АНСАМБЛЕВОМ ФИЛЬТРЕ КАЛМАНА

Климова Е.Г.

Институт вычислительных технологий СО РАН, Новосибирск

Оценка состояния окружающей среды по данным наблюдений является одной из наиболее актуальных задач в настоящее время. Такая оценка производится с привлечением прогностических моделей на основе систем усвоения данных. В статье представлен обзор современных методов усвоения данных наблюдений о парниковых газах, в том числе спутниковых, основанных на применении ансамблевого фильтра Калмана. Основное внимание уделяется методам оценки потоков парниковых газов в процедуре усвоения данных.

Ключевые слова: усвоение данных, спутниковые данные, парниковые газы, оценка потоков

Введение. Современное исследование окружающей среды включает в себя моделирование процессов с привлечением математических моделей прогноза погоды и климата, а также моделей распространения примеси в атмосфере. Для проведения математического моделирования состояния окружающей среды необходимо знание измеренных величин не только в точках наблюдений, но и в узлах некоторой регулярной сети. Оценка состояния окружающей среды по данным наблюдений является одной из наиболее актуальных задач в настоящее время. Такая оценка производится с привлечением прогностических моделей на основе систем усвоения данных. Под усвоением данных («data assimilation») принято понимать совместный учет данных и математической модели для наиболее точной оценки пространственно-временного распределения исследуемых величин.

Алгоритмы усвоения данных используются в настоящее время в различных областях, в частности, в задачах прогноза погоды, а также при решении различных задач распространения загрязняющих веществ в атмосфере [16]. Существует огромное количество методик усвоения данных, но, с точки зрения математической постановки задачи все они используют один из двух подходов, вариационный или динамико-стохастический. Поскольку данные наблюдений известны с ошибками, имеющими случайный характер, все методы усвоения должны учитывать статистические свойства ошибок измерений. Статистический характер имеют также так называемые «шумы» моделей, учет которых важен при проведении оценки состояния окружающей среды. Проблема учета статистических характеристик ошибок прогноза и наблюдений естественным образом решается при применении динамико-стохастического подхода (фильтр Калмана). Численная реализация фильтра Калмана для современных нелинейных моделей невозможна, поэтому в настоящее время используются различные приближения. Лидирующее положение занимает ансамблевый подход, при котором ковариации ошибок прогноза оцениваются с помощью ансамбля прогнозов по возмущенным начальным полям. Реализация ансамблевого подхода также содержит технологические сложности, связанные, в частности, с большой размерностью рассматриваемых при этом матриц [2-4, 10-11].

В ансамблевом фильтре Калмана можно выделить два подхода – «стохастический фильтр» и «детерминированный фильтр». В настоящее время проводится большое количество исследований по сравнению стохастических и детерминированных ансамблевых фильтров [12, 19].

Изучение распространения в пространстве и времени парниковых газов, таких как CO_2 и CH_4 , а также оценка потоков с поверхности Земли этих газов представляет одну из актуальных задач мониторинга состояния окружающей среды. Для решения этой задачи принято использовать системы усвоения данных, включающие в себя данные наблюдений и математическую модель распространения газовых составляющих в атмосфере. В данной работе представлен

обзор современных методов усвоения данных о парниковых газах, в том числе спутниковых, основанных на применении ансамблевого фильтра Калмана. Основное внимание уделяется методам оценки потоков парниковых газов в процедуре усвоения данных.

Ансамблевый фильтр Калмана. Вывод формул ансамблевого фильтра Калмана впервые приведен в работе [2]. Запишем нелинейную динамическую систему в виде уравнения процесса

$$\mathbf{x}'_k = f(\mathbf{x}'_{k-1}) + \boldsymbol{\eta}'_{k-1} \quad (1)$$

и уравнения наблюдений

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

где h – оператор, вообще говоря, нелинейный, переводящий значения прогноза в наблюдаемую переменную; $\boldsymbol{\eta}'_{k-1}$ – вектор «шумов модели»; $\boldsymbol{\varepsilon}'_k$ – вектор ошибок наблюдений; \mathbf{x}'_k – вектор оцениваемых переменных в момент времени t_k , $\boldsymbol{\varepsilon}'_k$ и $\boldsymbol{\eta}'_{k-1}$ – Гауссовские случайные переменные: $E[\boldsymbol{\varepsilon}'_k (\boldsymbol{\varepsilon}'_k)^T] = \mathbf{R}'_k$, $E[\boldsymbol{\eta}'_{k-1} (\boldsymbol{\eta}'_{k-1})^T] = \mathbf{Q}'_{k-1}$. Будем считать \mathbf{x}'_k «истинным» значением.

При рассмотрении ансамблевого подхода теоретико-вероятностное осреднение заменяется средним по выборке значением. В этом случае принято считать, что «истина» в начальный момент времени представлена выборкой $\{\mathbf{x}_0 + \boldsymbol{\delta}\mathbf{x}_n, n=1, \dots, N\}$, где \mathbf{X}_0 – известная оценка искомой величины, а выборка $\{\boldsymbol{\delta}\mathbf{x}_n, n=1, \dots, N\}$ моделирует случайные ошибки этой оценки. Предположим, что среднее по выборке значение этой ошибки равно нулю. Изменение «истины» по времени моделируется с помощью уравнения процесса. Тогда оптимальной оценкой будет среднее по выборке значение, а ошибкой оценки – отклонение от среднего значения.

Стохастический ансамблевый фильтр Калмана состоит из ансамбля прогнозов $\{\mathbf{x}_k^{f,n}, n=1, \dots, N\}$

$$\mathbf{x}_k^{f,n} = f(\mathbf{x}_{k-1}^{a,n}) + \boldsymbol{\eta}_{k-1}^n \quad (2)$$

и ансамбля анализов $\{\mathbf{x}_k^{a,n}, n=1, \dots, N\}$

$$\mathbf{x}_k^{a,n} = \mathbf{x}_k^{f,n} + \mathbf{K}_k (\mathbf{y}_k^n + \boldsymbol{\varepsilon}_k^n - h(\mathbf{x}_k^{f,n})). \quad (3)$$

Ансамбли (2) и (3) задают выборку значений «истины», при этом среднее по выборке значение будет являться оптимальной оценкой, а отклонение от среднего – ансамблем ошибок анализа и прогноза, соответственно. Для осуществления ансамблевого варианта алгоритма фильтра Калмана требуется задание ансамбля ошибок наблюдений $\{\boldsymbol{\varepsilon}_k^n, n=1, \dots, N\}$ а также ансамбля ошибок прогноза $\{\mathbf{d}\mathbf{x}_k^{f,n} = \mathbf{x}_k^{f,n} - \overline{\mathbf{x}_k^{f,n}}, n=1, \dots, N\}$, где $\overline{\mathbf{x}_k^{f,n}} \cong \frac{1}{N} \sum_{n=1}^N \mathbf{x}_k^{f,n}$ и ансамбля шумов модели $\{\boldsymbol{\eta}_{k-1}^n, n=1, \dots, N\} : E[\boldsymbol{\eta}_{k-1}^n (\boldsymbol{\eta}_{k-1}^n)^T] = \mathbf{Q}_k$. Матрица \mathbf{K}_k имеет вид

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

где \mathbf{P}_k^f и \mathbf{R}_k – матрицы, оцениваемые по ансамблю

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

\mathbf{H}_k линеаризованный оператор $h(\mathbf{x}_k^{f,n})$ относительно $\overline{\mathbf{x}_k^{f,n}}$:

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

Зададим ошибку анализа $\{\mathbf{d}\mathbf{x}_k^{a,n} = \mathbf{x}_k^{a,n} - \overline{\mathbf{x}_k^{a,n}}, n=1, \dots, N\}$, где $\overline{\mathbf{x}_k^{a,n}} \cong \frac{1}{N} \sum_{n=1}^N \mathbf{x}_k^{a,n}$. Тогда матрица

ковариаций ошибок анализа может быть вычислена по формуле

$$\mathbf{P}_k^a \triangleq \frac{1}{N-1} \sum_{n=1}^N \mathbf{d}\mathbf{x}_k^{a,n} (\mathbf{d}\mathbf{x}_k^{a,n})^T.$$

Формулы (2)-(3) представляют собой стохастический ансамблевый фильтр Калмана [4, 10-11, 17-18].

Детерминированный ансамблевый фильтр Калмана (шаг анализа) состоит из уравнения для среднего значения

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

и оценки ансамбля ошибок анализа так, чтобы соответствующая ковариационная матрица удовлетворяла уравнению фильтра Калмана $\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f$ [13, 19].

Оценка параметров в процедуре усвоения данных. Рассмотрим уравнение процесса в виде

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

данные наблюдений

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

где $\boldsymbol{\alpha}'_k$ - вектор параметров. Будем считать, что параметр не меняется со временем: $\boldsymbol{\alpha}'_{k+1} = \boldsymbol{\alpha}'_k$.

Рассмотрим обобщенную задачу оценивания вектора $\mathbf{z} = [\mathbf{x}, \boldsymbol{\alpha}]^T$ [1, 14]. Опуская промежуточные выкладки, запишем сразу результат процедуры оценивания в общем виде [1]:

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

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

В этих формулах $\mathbf{P}_{\alpha x}$ - кросс-ковариации ошибок \mathbf{x} и $\boldsymbol{\alpha}$, $\mathbf{P}_{\alpha\alpha}$ - ковариационная матрица ошибок $\boldsymbol{\alpha}$. h_x и h_α - линеаризованные операторы по \mathbf{x} и $\boldsymbol{\alpha}$, соответственно. В случае, если h от $\boldsymbol{\alpha}$ не зависит, оценка \mathbf{x}^a проводится по той же формуле, что и в обычном фильтре Калмана.

Современные направления в работах по оценке потоков парниковых газов по спутниковым данным. Существует большое количество работ, посвященных оценке потоков парниковых газов с помощью процедуры усвоения данных. Все эти работы так или иначе используют подход, называемый «inverse modelling», когда ищутся значения потоков, с использованием которых получается прогноз, наилучшим образом приближающийся к данным наблюдений. Выделим серию работ, проводимых международным коллективом авторов, в которых используется ансамблевый фильтр Калмана [5-9]. В этих работах проводится разбиение поверхности Земли на квадраты равной площади (1000 км на 1000 км) и предполагается, что требуется оценить среднее по подобласти значение потока. Оценка значений средних по подобластям потоков \mathbf{x}^a по данным наблюдений \mathbf{y}_0 и прогнозу \mathbf{x}^f производится по стандартной формуле фильтра Калмана [14]:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K} [\mathbf{y}_0 - H(\mathbf{x}^f)],$$

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

Для реализации ансамблевого фильтра Калмана задается ансамбль возмущений оцениваемого параметра [2-4]:

$$\Delta \mathbf{x}^f = \frac{1}{\sqrt{N}} [\Delta x_1, \dots, \Delta x_N]^T,$$

Матрица \mathbf{P}^f оценивается по ансамблю

$$\mathbf{P}^f = \Delta \mathbf{x}^f (\Delta \mathbf{x}^f)^T,$$

$$\mathbf{K}_e = \Delta \mathbf{x}^f (\Delta \mathbf{y})^T [\Delta \mathbf{y} (\Delta \mathbf{y})^T + \mathbf{R}]^{-1},$$

$$\Delta \mathbf{y} = H(\mathbf{x}^f + \Delta \mathbf{x}^f) - H(\mathbf{x}^f),$$

где оператор H включает в себя прогноз по модели к моменту наблюдения, интерполяцию из узлов сетки в точки наблюдений, а также, в случае спутниковых данных, осреднение по вертикали с известными коэффициентами («average kernel»), N - число элементов ансамбля.

В серии работ группы авторов [5-9] используется модель GEOS-Chem (3-D chemistry transport model) и спутниковые данные GOSAT. Кроме того, используются дополнительные наблюдения сети GLOBALVIEW, а также данные из других источников.

Промежуток времени, за который используются данные наблюдений, называется окном усвоения. В работе [5] рассматривается окно усвоения, состоящее из 12 циклов по 8 дней, и производится оценка средних за 8 дней значений потоков.

С точки зрения математической постановки задачи выделим следующие моменты:

1. производится оценка потоков без уточнения концентраций; решение задачи в такой постановке не будет оптимальным;
2. поскольку оператор H включает в себя математическую модель распространения примесей данные наблюдений являются асинхронными и, строго говоря, данная задача является задачей сглаживания, а не фильтрации [3-4].

С точки зрения практической реализации для решения данной задачи требуется задание матрицы \mathbf{R} – матрицы ковариаций ошибок наблюдений и модели, начальные (климатические) значения потоков.

В работе [20], посвященной оценке потоков CH_4 , рассматривается аналогичная постановка задачи и производится сравнительный анализ зависимости результатов от целого набора параметров. Как отмечается в работе [20], результат оценивания зависит от используемой модели, а также от параметризаций приземного и пограничного слоя в этих моделях.

Следует отметить работу [16], которая также посвящена применению ансамблевого фильтра Калмана для оценки потоков CO_2 по данным AIRS. С математической точки зрения, главным отличием этой работы является то, что рассматривается процедура совместной оценки концентрации и потоков вместе с метеорологическими параметрами. Авторы используют сравнительно небольшое окно усвоения (6 часов), в качестве метода используется алгоритм LETKF – детерминированный ансамблевый фильтр Калмана [13]. Решение задачи совместного оценивания концентрации и эмиссии является оптимальным с математической точки зрения. Однако задача становится технически гораздо более сложно реализуемой.

Заключение. Задача оценка потоков парниковых газов с поверхности Земли в настоящее время решается с помощью систем усвоения данных. При этом используются модели распространения примеси в атмосфере и метеорологические поля скорости ветра, температуры и т.д. для моделирования переноса и диффузии парниковых газов.

В качестве математической постановки задачи все чаще используется ансамблевый фильтр Калмана. Следует подчеркнуть, что реализация такой масштабной задачи требует усилий как правило нескольких научных коллективов. В качестве примера можно рассмотреть современные работы 2017 года [8] и [20]. В первой работе принимали участие представители 16 организаций из 9 стран, во второй – 17 организаций из 10 стран.

ЛИТЕРАТУРА

- [1] Справочник по теории автоматического управления под ред. А.А. Красовского. М.: Наука, 1987. 711 с.
- [2] Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics // Journal Geophysical Research. 1994. V. 99. P. 10143-10162.
- [3] Evensen, G. The ensemble Kalman filter: theoretical formulation and practical implementation // Ocean Dynamics. 2003. V. 53. P. 343-367.
- [4] Evensen, G. Data assimilation. The ensemble Kalman filter. Berlin Heidelberg: Springer-Verlag, 2009. 307 p.
- [5] Feng L., P.I.Palmer, H.Bosch, and S.Dance Estimating surface CO_2 fluxes from space-borne CO_2 dry air mole fraction observations using an ensemble Kalman filter // Atmospheric chemistry and physics. 2009. V. 9. P. 2619-2633.

- [6] Feng L. et al. Evaluating a 3-D transport model of atmospheric CO₂ using ground-based, aircraft, and space-borne data // *Atmospheric chemistry and physics*. 2011. V. 11. P. 2789-2803.
- [7] Feng L. et al. Estimates of European uptake of CO₂ inferred from GOSAT X_{co2} retrievals: sensitivity to measurement bias inside and outside Europe // *Atmospheric chemistry and physics*. 2016. V. 16. P.1289-1302.
- [8] 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. V. 17. P. 4781-4797.
- [9] Fraser A. et al. Estimating regional fluxes of CO₂ and CH₄ using space-borne observations of XCH₄:XCO₂ // *Atmospheric chemistry and physics*. 2014. V. 14. P. 12883-12895.
- [10] Houtekamer, H.L. Mitchell, Data assimilation using an ensemble Kalman Filter technique // *Monthly Weather Review*. 1998. V. 126. P. 796-811.
- [11] Houtekamer, P.L., Mitchell, H.L. Ensemble Kalman Filtering // *Quarterly Journal of the Royal Meteorological Society*. 2005. V. 131. P. 1-23.
- [12] Houtekamer, H.L. Zhang, F. Review of the ensemble Kalman filter for atmospheric data assimilation // *Monthly Weather Review*. 2016. V. 144. P. 4489-4532.
- [13] Hunt, B.R., Kostelich, E.J. Szunyogh, I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter // *Physica D*. 2007. V. 230. P. 112-126.
- [14] Jazwinski, A.H. Stochastic processes and filtering theory. New York: Academic Press, 1970. 376 p.
- [15] Kalnay, E. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge Univ. Press, 2002. 328 p.
- [16] Kang J. et al. Estimating of surface carbon fluxes with an advanced data assimilation methodology // *Journal of geophysical research*. 2012. V.116. D24101, doi:10.1029/2012JD018259.
- [17] Klimova, E. G. A data assimilation technique based on the pi-algorithm // *Russian Meteorology and Hydrology*. 2008. V. 33. P. 143-150.
- [18] Klimova, E. A suboptimal data assimilation algorithm based on the ensemble Kalman filter // *Quarterly Journal of the Royal Meteorological Society*. 2012. V. 138, P. 2079-2085.
- [19] Tippett, M.K., J.L. Anderson, C.Y. Bishop, T.m. Hamill, and J.S. Whitaker. Ensemble square root filters // *Monthly Weather Review*. 2003. V. 131. P. 1485-1490.
- [20] Tsuruta A. et al. Global methane emission estimation for 2010-2012 from Carbon tracker Europe-CH₄ v.10. // *Geoscientific model development*. 2017. V. 10. P. 1261-1287.