

ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ ПО ОЦЕНКЕ СИСТЕМАТИЧЕСКОЙ ОШИБКИ МОДЕЛИ В ЗАДАЧЕ УСВОЕНИЯ ДАННЫХ О КОНЦЕНТРАЦИИ ПАССИВНОЙ ПРИМЕСИ*

Н. В. КИЛАНОВА, Е. Г. КЛИМОВА

Институт вычислительных технологий СО РАН, Новосибирск, Россия

e-mail: kilanova@mail.ru, klimova@ict.nsc.ru

An error estimation algorithm for a problem of advection and diffusion of a passive pollution in data assimilation procedure is proposed. Results of the experiments have shown the applicability of the proposed algorithm to the problem of a passive pollution data assimilation. Properties of the algorithm are tested with the help of numerical experiments with the model data.

Введение

Для получения информации о пространственно-временном распределении концентрации пассивной примеси в атмосфере необходимо использовать систему усвоения данных измерений с привлечением математической модели переноса и диффузии примеси. Задачей усвоения данных называется задача совместного учета прогностической модели и данных наблюдений для получения оптимальной в некотором смысле оценки искомых полей. Одним из подходов при решении задачи усвоения является применение алгоритма фильтра Калмана [1].

Для того чтобы алгоритм фильтра Калмана работал эффективно, требуется адекватное задание статистических характеристик ошибок модели и данных наблюдений. Оценка и учет ошибок модели менее изучены в этом алгоритме. В классической постановке алгоритма фильтра Калмана предполагается, что ошибка модели является гауссовским белым шумом. Вообще говоря, математическое ожидание ошибки модели (называемое систематической ошибкой) не равно нулю. Оценка систематической ошибки модели является в настоящее время одной из актуальных задач в проблеме усвоения данных. Этой проблеме посвящена данная работа, в которой предлагается алгоритм оценки систематической ошибки на примере модели переноса и диффузии пассивной примеси в процедуре усвоения данных.

Прогноз полей концентрации дается с помощью полулагранжевой модели переноса и диффузии пассивной примеси для Северного полушария. Для реализации алгоритма

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 04-05-64481).

© Институт вычислительных технологий Сибирского отделения Российской академии наук, 2006.

фильтра Калмана требуется вычисление матриц очень высокой размерности, что в настоящее время невозможно ни на одной ЭВМ. Поэтому в работе рассматривается подход, в котором используется упрощенная модель для расчета матриц ковариаций ошибок прогноза [2–4].

1. Модель переноса и диффузии пассивной примеси для Северного полушария

Предварительная оценка полей концентрации примеси дается с помощью модели переноса и диффузии пассивной примеси:

$$\begin{aligned} & \frac{\partial s}{\partial t} + \frac{u}{a \cos \varphi} \frac{\partial s}{\partial \lambda} + \frac{v}{a} \frac{\partial s}{\partial \varphi} + (w - w_g) \frac{\partial s}{\partial z} = \\ & = \frac{1}{a^2 \cos^2 \varphi} \frac{\partial}{\partial \lambda} \left(k_1 \frac{\partial s}{\partial \lambda} \right) + \frac{1}{a^2 \cos \varphi} \frac{\partial}{\partial \varphi} \left(k_2 \cos \varphi \frac{\partial s}{\partial \varphi} \right) + \frac{\partial}{\partial z} \left(k_3 \frac{\partial s}{\partial z} \right) + f, \\ & s(\lambda, \varphi, z, t) = s(\lambda + 2\pi, \varphi, z, t), \\ & s = s_G \text{ при } \varphi = 0, \\ & \nu \frac{\partial s}{\partial z} = \alpha (s - s_0) \text{ при } z = b + h, \\ & \nu \frac{\partial s}{\partial z} = 0 \text{ при } z = H. \end{aligned}$$

Здесь λ — долгота; φ — широта; z — высота над уровнем моря; s — концентрация примеси; $\mathbf{V} = (u, v, w - w_g)$ — вектор скорости ветра с компонентами в направлениях λ, φ, z соответственно; w_g — скорость гравитационного оседания; a — средний радиус Земли; k_1, k_2 — коэффициенты турбулентного обмена в горизонтальном и k_3 — вертикальном направлениях; f — функция, описывающая размещение и мощность источников; $b(\lambda, \varphi)$ — функция, описывающая рельеф подстилающей поверхности; h — высота приземного слоя и H — верхняя граница расчетной области; параметр α характеризует взаимодействие атмосферы с подстилающей поверхностью [5].

Задача рассматривается в области

$$D_t = G \times [0, T],$$

где $G = S \times [b + h, H]$; $S = \{0 \leq \lambda \leq 2\pi; 0 \leq \varphi \leq \pi/2\}$.

Для решения поставленной задачи используется метод расщепления по физическим процессам [6]. В результате на каждом шаге по времени последовательно решаются задача переноса примеси по траекториям и задача турбулентной диффузии. При решении задачи переноса примеси по траекториям используется полулагранжева схема с применением монотонизирующей процедуры, позволяющей избежать искусственных экстремумов в приближенном решении [7, 8]. Задача турбулентной диффузии аппроксимируется по схеме Кранка — Николсона. К полученному разностному уравнению применяется метод расщепления по направлениям, для решения системы разностных уравнений используется метод факторизации.

Подробнее описание модели и численных экспериментов с системой усвоения для этой модели приведено в [2–4].

2. Усвоение данных наблюдений в задаче переноса и диффузии пассивной примеси

Используемый в задаче оценивания полей концентрации пассивной примеси алгоритм усвоения данных основан на теории оптимальной фильтрации Калмана. Постановка задачи оптимальной фильтрации подробно изложена в [9], описание алгоритма фильтра Калмана можно найти в [1, 10–14]. Дискретный алгоритм фильтра Калмана в линейном случае изложен в [2].

Предположим, что модель переноса и диффузии в дискретном по времени и пространству виде записывается следующим образом:

$$x_k = A_{k-1}x_{k-1}. \quad (1)$$

Здесь x_k — n -вектор прогнозируемых значений концентрации примеси в момент времени t_k , его размерность равна количеству узлов сетки: $n = n_\lambda n_\varphi n_z$, где n_λ — количество узлов сетки по долготе, n_φ — по широте, n_z — по высоте.

Пусть “истинное” значение концентрации в момент времени t_k описывается векторным уравнением

$$x_k^t = A_{k-1}x_{k-1}^t + \varepsilon_{k-1}^t, \quad (2)$$

где ε_k — случайный вектор “шумов” модели, удовлетворяющий условиям

$$E\varepsilon_k^t = 0, \quad E(\varepsilon_k^t)(\varepsilon_l^t)^T = Q_k\delta_{kl}.$$

Вектор данных наблюдений представим в виде

$$y_k^o = M_k x_k^t + \xi_k^o, \quad (3)$$

где y_k^o — m -вектор наблюдений в момент времени t_k ; M_k — $(n \times m)$ -матрица; ξ_k^o — случайный m -вектор ошибок наблюдений, удовлетворяющий условиям

$$E\xi_k^o = 0, \quad E(\xi_k^o)(\xi_l^o)^T = R_k\delta_{kl}, \quad E(\xi_k^o)(\varepsilon_l^t)^T = 0.$$

Оптимальная оценка концентрации x_k^a по данным наблюдений (3) и модели переноса и диффузии примеси (1) ищется из условия минимума следа матрицы ковариаций ошибок оценивания. Решением этой задачи является алгоритм фильтра Калмана [9]:

$$x_k^f = A_{k-1}x_{k-1}^a; \quad (4)$$

$$P_k^f = A_{k-1}P_{k-1}^a A_{k-1}^T + Q_{k-1}; \quad (5)$$

$$K_k = P_k^f M_k^T (M_k P_k^f M_k^T + R_k)^{-1}; \quad (6)$$

$$P_k^a = (I - K_k M_k) P_k^f; \quad (7)$$

$$x_k^a = x_k^f + K_k (y_k^o - M_k x_k^f); \quad (8)$$

$$P_k^f = E(x_k^f - x_k^t)(x_k^f - x_k^t)^T, \quad P_k^a = E(x_k^a - x_k^t)(x_k^a - x_k^t)^T.$$

Здесь x_k^f — прогноз концентрации по модели; x_{k-1}^a — вектор проанализированных значений в момент времени t_{k-1} , т. е. искомая нами оценка, которая получена на $(k-1)$ -м шаге по

времени; оценку x_{k-1}^a называют оценкой анализа; P_k^f , P_k^a — матрицы ковариации ошибок прогноза и анализа соответственно; K_k — весовая матрица.

Таким образом, алгоритм фильтра Калмана состоит из шага прогноза, когда по прогностической модели вычисляется предварительная оценка концентрации по формуле (4) и рассчитываются матрицы ковариации ошибок прогноза по (5). Затем вычисляется весовая матрица K_k по формуле (6). Далее идет шаг анализа, который заключается в получении непосредственно искомой оценки анализа по (8) и матрицы ковариации ошибок анализа по формуле (7).

В моменты времени, когда наблюдения недоступны, считаем, что оценка анализа совпадает с прогнозом по модели, а матрица ковариации ошибок анализа равна матрице ковариации ошибок прогноза.

Реализация алгоритма фильтра Калмана на ЭВМ в полной постановке невозможна, так как для современных глобальных моделей порядок ковариационных матриц достигает миллиона. Одним из способов решения этой проблемы является использование упрощенных моделей для расчета матриц ковариаций ошибок прогноза. Такой алгоритм называют субоптимальным алгоритмом фильтра Калмана [12–15].

3. Численные эксперименты с субоптимальным алгоритмом фильтра Калмана

В классической постановке алгоритма фильтра Калмана в формуле (2) сделано предположение, что ошибки модели ε_{k-1}^t являются белым шумом. В последнее время в ряде работ их авторы отступают от этого упрощающего предположения, полагая, что ошибка модели состоит из белого шума и ошибки с ненулевым математическим ожиданием, называемой систематической ошибкой модели.

Рассматривался вариант субоптимального алгоритма в применении к задаче оценки систематической ошибки модели. Считаем, что “истинное” поле концентрации можно задать в виде

$$\begin{aligned}x_k^t &= Ax_{k-1}^t + G_{k-1}q_{k-1}^t + \varepsilon_{k-1}^t, \\q_k^t &= q_{k-1}^t,\end{aligned}$$

где A — оператор модели; x_k^t — “истинная” концентрация метана в момент времени t_k ; q_{k-1}^t — “истинная” систематическая ошибка модели в момент времени t_{k-1} ; матрица G_{k-1} в общем случае неквадратная; ошибки ε_{k-1}^t обладают свойствами

$$E\varepsilon_k^t = 0; \quad E(\varepsilon_k^t)(\varepsilon_l^t)^T = Q_k\delta_{kl}.$$

Прогноз концентрации по модели x_k^f (предварительная оценка поля концентрации) задавался следующим образом:

$$x_k^f = Ax_{k-1}^f + G_{k-1}q_{k-1}^f; \quad (9)$$

$$\Delta x_k = A\Delta x_{k-1} + G_{k-1}\Delta q_{k-1}; \quad (10)$$

$$q_k^f = q_{k-1}^f; \quad (11)$$

$$\Delta q_k = \Delta q_{k-1}. \quad (12)$$

Здесь q_{k-1}^f — прогноз систематической ошибки модели в момент времени t_{k-1} . Параллельно с полями x_k^f, q_k^f вычислялись значения Δx_k и Δq_k , где $\Delta x_k = x_k^t - x_k^f$ — отклонение оценки концентрации от “истинного” значения концентрации, а $\Delta q_k = q_k^t - q_k^f$ — отклонение оценки систематической ошибки модели от “истинной” систематической ошибки.

В такой постановке задачи правую часть $G_{k-1}q_{k-1}^t$ можно рассматривать как эмиссию. Результаты численных экспериментов по оценке эмиссии в задаче переноса и диффузии пассивной примеси на примере атмосферного газа метана приведены в [3, 4].

Алгоритм оценки параметра q_{k-1} состоит в том, что вместо исходного вектора x в алгоритме фильтра Калмана оценивается расширенный вектор состояния (x, q) [16]. В этом случае шаг анализа алгоритма фильтра Калмана запишется в виде

$$x_k^a = x_k^f + P_{xx,k}^f M_k^T (M_k P_k^f M_k^T + R_k)^{-1} (y_k^o - M_k x_k^f); \quad (13)$$

$$q_k^a = q_k^f + P_{xq,k}^f M_k^T (M_k P_k^f M_k^T + R_k)^{-1} (y_k^o - M_k x_k^f); \quad (14)$$

$$\Delta x_k = \Delta x_k - P_{xx,k}^f M_k^T (M_k P_k^f M_k^T + R_k)^{-1} (y_k^o - M_k x_k^f); \quad (15)$$

$$\Delta q_k = \Delta q_k - P_{xq,k}^f M_k^T (M_k P_k^f M_k^T + R_k)^{-1} (y_k^o - M_k x_k^f), \quad (16)$$

где $P_{xx,k}^f$ — матрица ковариации ошибок прогноза оценки концентрации; $P_{xq,k}^f$ — матрица кроссковиариаций ошибок прогноза оценки концентрации и оценки систематической ошибки модели.

В данном алгоритме при вычислении ковариационных матриц использовалось предположение о том, что вероятностное осреднение можно заменить осреднением по времени [17, 18]. Таким образом, в формулах (17), (18) черта сверху обозначает осреднение по времени:

$$P_{xx,k}^f = \overline{\Delta x_k (\Delta x_k)^T} = \frac{1}{N-1} \sum_{i=1}^N \Delta x_i (\Delta x_i)^T; \quad (17)$$

$$P_{xq,k}^f = \overline{\Delta x_k (\Delta q_k)^T} = \frac{1}{N-1} \sum_{i=1}^N \Delta x_i (\Delta q_i)^T. \quad (18)$$

В численных экспериментах расчеты проводились на сетке $2.5^\circ \times 2.5^\circ$ по горизонтали и пятнадцати уровнях по вертикали, шаг по времени $\Delta t = 15$ мин. Были использованы реальные данные о полях скорости ветра, температуры и давления и модельные данные о концентрации метана [19]. Расчеты проводились для временного периода 48 ч с усвоением каждые 12 ч.

В первом эксперименте в моменты наблюдений моделировались данные наблюдений по формуле (3) таким образом, что в каждый момент усвоения задавалась информация в полосе. На рис. 1 приведено распределение данных наблюдений. При этом каждая вертикальная полоса соответствует области, в которой задаются данные в 12, 24, 36 и 48 часов.

Положим, что в начальный момент времени истинное значение систематической ошибки постоянно для всей области решения задачи. В эксперименте задавалось $q_0^t = 0.0009$. Кроме того, задано отклонение оценки параметра от истинного значения $\Delta q_0 = 0.0009 \times N(0, 1) \times 0.1$, где $N(0, 1)$ — нормально распределенная случайная величина с нулевым математическим ожиданием и дисперсией, равной 1.

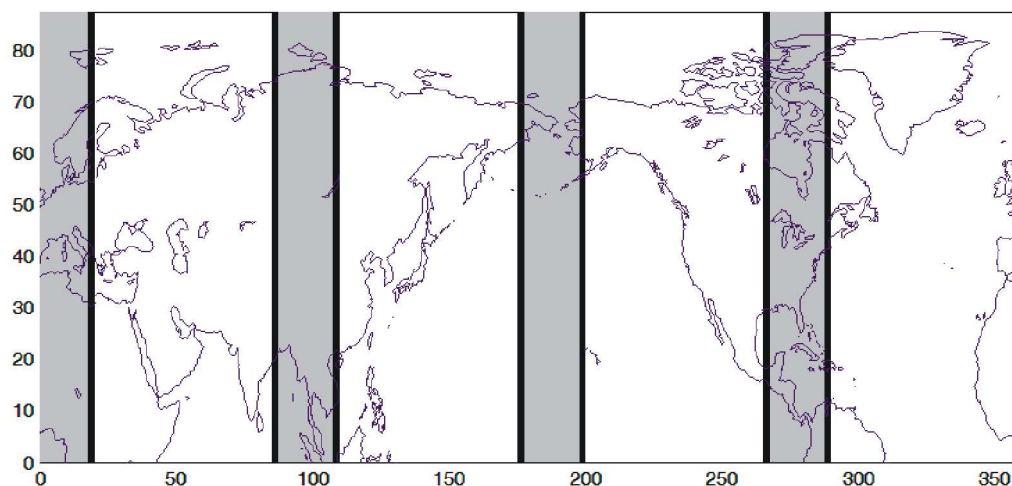
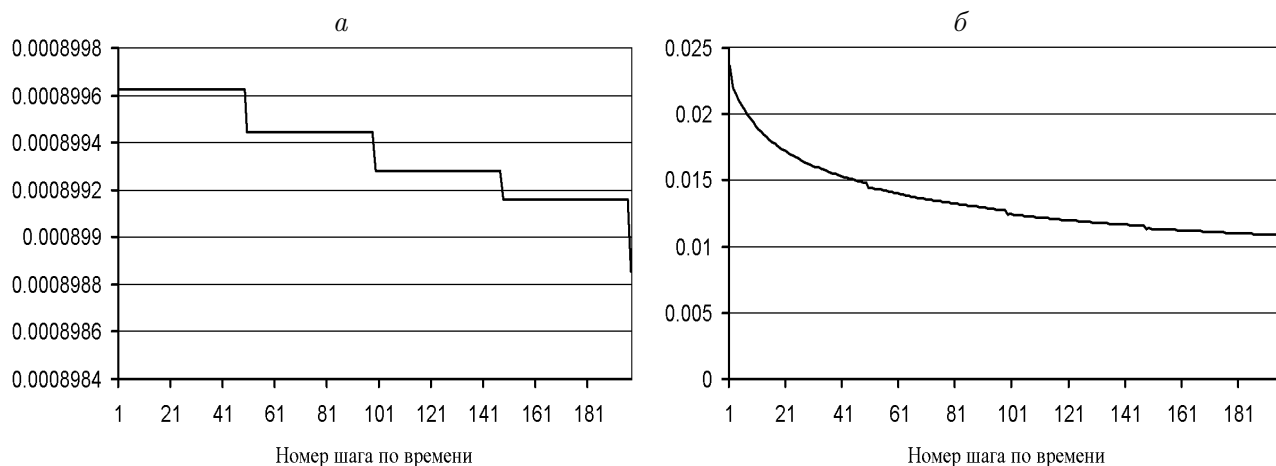


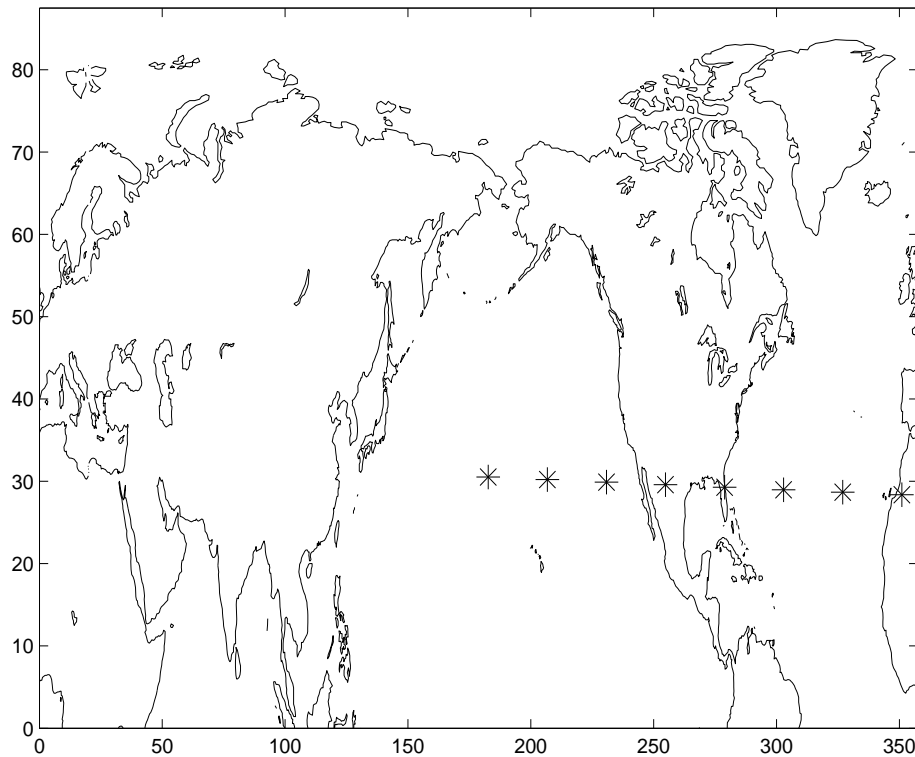
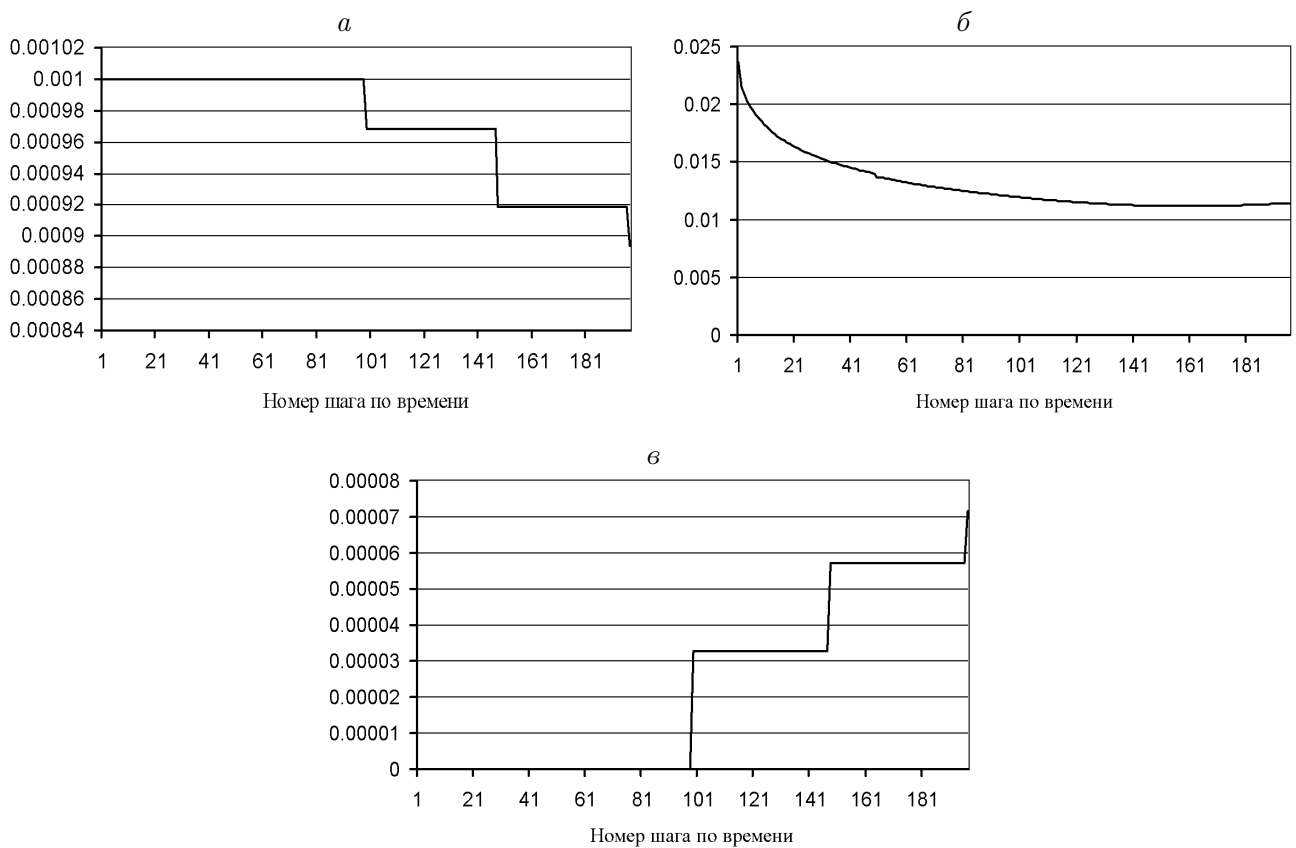
Рис. 1. Моделируемое распределение данных наблюдений.

Рис. 2. Среднеквадратическая ошибка: параметра q_k , ppm (а) и оценки концентрации, ppm (б). Эксперимент 1.

Результаты эксперимента приведены на рис. 2, а и б. Из рис. 2, а видно, что заданная в начальный момент времени величина $\Delta q_0 = q_0^t - q_0^f$ убывает в процессе усвоения, на рис. 2, б отражено убывание среднеквадратической ошибки концентрации.

Во втором численном эксперименте данные наблюдений моделировались в точках реальных спутниковых наблюдений. Координаты точек наблюдений и значения концентраций в них размещены на сайте [20]. На рис. 3 представлено распределение данных наблюдений, которые участвовали в усвоении в момент времени $t = 12$ ч. По вертикали данные моделировались на четырех верхних уровнях, а именно на высотах 13 660, 16 250, 21 374 и 25 290 м.

Данные наблюдений моделировались таким образом, что невязка в точках наблюдений составляла 0.001. По формулам (9)–(12) определяли предварительную оценку концентрации и систематической ошибки. В моменты усвоения по (13)–(16) получена оценка анализа концентрации и систематической ошибки модели. Результаты этого эксперимента приведены на рис. 4.

Рис. 3. Распределение данных наблюдений в момент времени $t = 12$ ч.Рис. 4. Среднеквадратическая ошибка: параметра q_k , ррм (а); оценки концентрации, ррм (б); оценка значения систематической ошибки модели q_k , ррм (в). Эксперимент 2.

Как видно из рис. 4, *a* и *б*, среднеквадратические ошибки оцениваемого параметра q_k и среднеквадратическая ошибка концентрации убывают после каждого усвоения. Следует отметить, что убывание среднеквадратической ошибки концентрации в эксперименте 1 (см. рис. 2, *б*) регулярнее и больше по значению, чем в эксперименте 2 (см. рис. 4, *б*). Это естественным образом объясняется тем фактом, что в каждом усвоении в первом эксперименте участвовало значительно большее количество точек наблюдений, чем в эксперименте 2. Из рис. 4, *в* видно, что восстановление значения систематической ошибки производится с каждым усвоением, при этом в начальный момент времени значение систематической ошибки равно нулю. В эксперименте с модельными данными показано, что при усвоении происходит убывание ошибки как самой концентрации, так и оцениваемого параметра.

Таким образом, в работе предлагается алгоритм оценки ошибки модели по данным наблюдений в процедуре усвоения, основанной на фильтре Калмана. Результаты проведенных экспериментов показали применимость предложенного алгоритма в задаче усвоения данных о пассивной примеси. Свойства алгоритма проверены с помощью численных экспериментов с модельными данными.

Список литературы

- [1] GHIL M., MALANOTTE-RIZZOLLI P. Data assimilation in meteorology and oceanography // *Advances in Geophysics*. 1991. Vol. 33.
- [2] Климова Е.Г., Киланова Н.В. Усвоение данных наблюдений в задаче переноса и диффузии пассивной примеси // *География и природные ресурсы*. 2004 (Спецвыпуск): Тр. Междунар. конф. ENVIRROMIS-2004. Новосибирск, 2004. С. 175–180.
- [3] Киланова Н.В., Климова Е.Г. Оценка полей концентрации метана над Северным полушарием по данным измерений и модели переноса и диффузии пассивной примеси // *Вычисл. технологии*. 2005 (Спецвыпуск): Тр. Междунар. конф. CITES-2005. Новосибирск, 2005. С. 132–137.
- [4] Климова Е.Г., Киланова Н.В. Усвоение данных наблюдений в задаче совместного оценивания концентрации и эмиссии пассивной примеси // *Матер. V Междунар. симп. "Контроль и реабилитация окружающей среды"*. Томск, 2006. С. 113–115.
- [5] МАРЧУК Г.И., АЛОЯН А.Е. Глобальный перенос примеси в атмосфере // *Изв. РАН. Физика атмосферы и океана*. 1995. Т. 31, № 5. С. 597–606.
- [6] МАРЧУК Г.И. Математическое моделирование в проблеме окружающей среды. М.: Наука, 1982. 319 с.
- [7] ТОЛСТЫХ М.А. Полулагранжева модель атмосферы с высоким пространственным разрешением для численного прогноза погоды // *Метеорология и гидрология*. 2001. № 4. С. 5–16.
- [8] VERMEJO R., STANFORTH A. The conversion of semi-Lagrangian advection scheme to quasi-monotone scheme // *Mon. Wea. Rev.* 1992. Vol. 120. P. 2622–2632.
- [9] JAZWINSKI A.H. Stochastic processes and filtering theory. N.Y.: Acad. Press, 1970. 377 p.

- [10] MENARD R., COHN S.E., CHANG L.-P., LYSTER P.M. Assimilation of stratospheric chemical tracer observations using a Kalman filter. Pt 1: Formulation // Mon. Wea. Rev. 2000. Vol. 128. P. 2654–2671.
- [11] MENARD R., CHANG L.-P. Assimilation of stratospheric chemical tracer observations using a Kalman filter. Pt 2: χ^2 -validated results and analysis of variance and correlation dynamics // Mon. Wea. Rev. 2000. Vol. 128. P. 2672–2686.
- [12] Климова Е.Г. Упрощенные модели для расчета ковариационных матриц в алгоритме фильтра Калмана // Метеорология и гидрология. 2000. № 6. С. 18–30.
- [13] Климова Е.Г. Модель для расчета ковариаций ошибок прогноза в алгоритме фильтра Калмана, основанная на полных уравнениях // Метеорология и гидрология. 2001. № 11. С. 11–21.
- [14] Климова Е.Г. Численные эксперименты по усвоению метеорологических данных с помощью субоптимального фильтра Калмана // Метеорология и гидрология. 2003. № 10. С. 54–67.
- [15] TODLING R., COHN S. Suboptimal schemes for atmospheric data assimilation based on the Kalman filter // Mon. Wea. Rev. 1996. Vol. 124. P. 2530–2557.
- [16] БАЛАКРИШНАН А.В. Теория фильтрации Калмана. М.: Мир, 1988. 168 с.
- [17] СОНЕЧКИН Д.М. Динамико-стохастический подход к задаче объективного анализа данных разнородных метеорологических наблюдений // Тр. Гидрометцентра СССР. 1976. Вып. 181. С. 54–76.
- [18] ЯГЛОМ А.М. Корреляционная теория стационарных случайных функций. Л.: Гидрометеоиздат, 1981. 280 с.
- [19] КРУПЧАТНИКОВ В.Н., КРЫЛОВА А.И. Численное моделирование распределения метана по данным наблюдений на поверхности Земли // Оптика атмосферы и океана. 2000. Т. 13, № 6-7. С. 622–626.
- [20] <http://haloedata.larc.nasa.gov>

*Поступила в редакцию 1 августа 2006 г.,
в переработанном виде — 8 августа 2006 г.*