

## Использование комплекса моделей разной сложности в задачах обратного моделирования процессов переноса и трансформации примесей в атмосфере

А. В. Пененко<sup>\* \*\*</sup>, П. Н. Антохин<sup>\*\*\* \*\*</sup>,  
А. А. Бакланов<sup>\*\*</sup>, В. В. Пененко<sup>\*</sup>

*Задачи применения интегрированных моделей для оценки и прогнозирования качества воздуха требуют разработки алгоритмов усвоения данных. Использование сложных в вычислительном плане моделей в алгоритмах обратного моделирования, которые часто требуют многократного решения прямых и сопряженных задач, может быть затруднительно с практической точки зрения. Рассмотрен алгоритм, задействующий две модели: детальную и упрощенную. Детальная модель применяется в режиме прямого моделирования, а для уточнения информации о ключевых для моделирования параметрах по данным наблюдений (здесь, в частности, источников выбросов) используется упрощенная модель. При этом возникает задача подстройки (обучения) упрощенной модели для воспроизведения поведения детальной модели. Алгоритм протестирован на региональном сценарии оценки качества воздуха для Сибирского региона.*

**Ключевые слова:** обратное моделирование, комплекс моделей, перенос-трансформация примесей, идентификация источников выбросов, мезомасштабные модели, оператор чувствительности.

DOI: 10.52002/0130-2906-2025-6-67-78

### Введение

Интенсивно развивающиеся технологии проведения наблюдений, которые стали доступны в последнее десятилетие, требуют внимания и дают мощный импульс для разработки новых методов совмещения данных измерений с моделями для улучшения прогноза динамики и оценки состояния атмосферы. В частности, в последние десятилетия значительно улучшились и расширились методы и возможности мультиплатформенных (наземных, авиационных и спутниковых на основе методов дистанционного

<sup>\*</sup> Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук; e-mail: [aleks@otmgrp.sscs.ru](mailto:aleks@otmgrp.sscs.ru) (Пененко Алексей Владимирович).

<sup>\*\*</sup> Российский государственный гидрометеорологический университет; e-mail: [arp@iao.ru](mailto:arp@iao.ru) (Антохин Павел Николаевич).

<sup>\*\*\*</sup> Институт оптики атмосферы им. В. Е. Зуева Сибирского отделения Российской академии наук.

зондирования) наблюдений за загрязнением воздуха и атмосферными параметрами, а также их оперативная доступность в реальном времени. Классические методы усвоения однотипных данных не всегда применимы для таких систем, поэтому одним из ключевых современных трендов совершенствования систем прогноза и моделирования атмосферных процессов является разработка новых методов использования разного типа данных наблюдений в моделях, включая алгоритмы усвоения и интегрирования данных, методы коррекции смещений (*bias correction methods*), методы искусственного интеллекта на основе машинного обучения, в том числе искусственных нейронных сетей, гибридные методы [7]. Это приводит к пересмотру концепций моделирования и важности усвоения данных (как химических, так и метеорологических измерений) в системах оперативного прогноза и интеграции/совмещения наблюдений и моделей для долгопериодной оценки качества воздуха и осаждения примесей. Более подробный обзор по теме можно найти в [5, 6], а также в других работах авторов.

Современные интегрированные модели качества воздуха (например, WRF-Chem [11, 16], COSMO-ART [12], Enviro-HIRLAM [5]) имеют высокую сложность. В этих многомерных нестационарных моделях учитываются процессы переноса и трансформации примесей в атмосфере, а также взаимодействие между примесями и метеорологическими переменными. Задачи применения этих моделей для оценки и прогнозирования качества воздуха требуют разработки для них алгоритмов усвоения данных мониторинга [6]. В частности, новое поколение геостационарных спутников позволяет обеспечить беспрецедентный обзор качества атмосферного воздуха из космоса. С учетом уже имеющихся и запланированных миссий еще более насущной является разработка эффективных систем усвоения данных для интегрированных моделей, которые оптимально учитывают мультиплатформенные наблюдения и спутниковую информацию. При этом даже для решения задач прямого моделирования с помощью интегрированных моделей, когда все необходимые параметры моделей считаются заданными, требуются существенные вычислительные ресурсы. Таким образом, использование сложных в вычислительном плане моделей в алгоритмах обратного моделирования, которые часто требуют многократного решения прямых и сопряженных задач, может быть обременительно с практической точки зрения.

В подходе авторов к решению задач совместного использования моделей и данных наблюдений (т. е. задач обратного моделирования) параметры моделей делятся на априори “заданные” и на “функции неопределенности”, которые и применяются для подстройки к имеющимся данным наблюдений. Задачи усвоения данных формализуем как последовательности связанных обратных задач на наборе временных интервалов, называемых окнами усвоения. Каждая из обратных задач последовательности имеет искомые переменные и наборы доступных данных измерений. Результаты решения предыдущих задач из последовательности используются как априорная информация для последующих.

Рассмотрим обратную задачу в окне усвоения. Для ее решения используется подход на основе операторов чувствительности и ансамблей решений сопряженных уравнений, восходящий к работам Г. И. Марчука [1, 2].

Достоинством подхода является возможность эффективного распараллеливания на современных вычислительных системах. Генерируемые алгоритмом агрегаты можно использовать для последующего уточнения решения методами машинного обучения. Наличие кластерной версии позволяет применять алгоритм к региональным трехмерным задачам обратного моделирования, включая задачи усвоения данных. Детали можно найти в [3, 4] и в более ранних работах авторов. Другие подходы к идентификации источников выбросов и усвоению данных, а также примеры их применения можно найти в [8—10, 14, 15].

Целью настоящей работы является развитие алгоритма обратного моделирования на основе двух моделей: детальной и упрощенной. Детальная модель используется в режиме прямого моделирования, а для уточнения информации о ключевых для моделирования параметрах по данным наблюдений (в частности, источников выбросов) взята упрощенная модель. При этом также возникает задача подстройки (обучения) упрощенной модели так, чтобы результаты ее работы без учета данных измерений были в некотором роде сходны с результатами работы детальной модели. Для совместного использования моделей потребовалось разработать процедуры отображения для их основных переменных — функций состояния, неопределенности и параметров. В качестве функции неопределенности в данном исследовании выбраны функции источников, так как они во многом определяют результат моделирования качества воздуха. Данные измерений заданы в виде “снимков” интегральных концентраций по высоте (total column). Для проведения численных экспериментов выбран сценарий для Сибирского региона. Данные синтетических измерений генерировались с помощью модели WRF-Chem с истинными источниками, заданными согласно базе EDGAR.

Работа имеет следующую структуру. В п. 2.1 приведена постановка задачи обратного моделирования. В п. 2.2 построен алгоритм оценки функции состояния для пары моделей. В п. 2.3 рассмотрены вопросы взаимных отображений переменных пары моделей. В п. 2.4 описан сценарий обратного моделирования. В разделе 3 обсуждаются полученные результаты.

## 2. Методы и алгоритмы

### 2.1. Постановка задачи

Упрощенная модель для описания распространения  $N_c$  реагирующих химических веществ рассматривается в области  $T = [0, T]$ , где — достаточно гладкая аппроксимация ограниченной прямоугольной области  $[0, X] \times [0, Y] \times [0, Z]$  в  $\mathbf{R}^3$ ,  $T > 0$ , — граница области :

$$\frac{d}{dt} (\text{diag}( \mathbf{u} ) \mathbf{c} + P_l(t, \mathbf{x})) = \mathbf{c}(t, \mathbf{x}) + f_l + r_l, \quad (\mathbf{x}, t) \in [0, T], \quad (1)$$

$$\mathbf{n} \cdot (\text{diag}( \mathbf{u} ) \mathbf{c}) = 0, \quad (\mathbf{x}, t) \in \text{out} [0, T], \quad (2)$$

$$\mathbf{c} = \mathbf{c}_l, \quad (\mathbf{x}, t) \in \text{in} [0, T], \quad (3)$$

$$\mathbf{c} = \mathbf{c}_l^0, \quad \mathbf{x} \in \text{in}, \quad t = 0, \quad (4)$$

$$l = 1, \dots, N_c, \quad (5)$$

где  $t \in [0, T]$  — время;  $\mathbf{x}$  — пространственные координаты;  $N_c$  — число рассматриваемых веществ;  $c_l = c_l(\mathbf{x}, t)$  — концентрация  $l$ -го вещества в точке  $(\mathbf{x}, t)$ ;  $\mathbf{c} = \mathbf{c}(\mathbf{x}, t)$  — вектор, состоящий из элементов  $c_l(\mathbf{x}, t)$ ,  $l = 1, \dots, N_c$ , который называется функцией состояния,  $L = \{1, \dots, N_c\}$ . Функции  $c_l(\mathbf{x}, t) \in \mathbf{R}^3$  соответствуют коэффициентам диффузии;  $\text{diag}(\mathbf{a})$  — диагональная матрица с вектором  $\mathbf{a}$  на диагонали;  $\mathbf{u}(\mathbf{x}, t) \in \mathbf{R}^3$  — вектор скорости ветра. Скорости должны удовлетворять условиям неразрывности сплошной среды.  $\mathbf{u}^{(in)}$  и  $\mathbf{u}^{(out)}$  — части границы области  $\Omega$ , на которых вектор  $\mathbf{u}(\mathbf{x}, t)$  соответственно направлен внутрь области  $\Omega$  или равен нулю, либо направлен из области  $\Omega$ ;  $\mathbf{n}$  — вектор внешней нормали. Функции  $c_l(\mathbf{x}, t)$ ,  $c_l^0(\mathbf{x})$  описывают соответственно граничные и начальные условия;  $f_l(\mathbf{x}, t)$  — известная аргонная функция источника;  $r_l(\mathbf{x}, t)$  — неизвестная функция источника. В данном случае  $c_l(\mathbf{x}, t)$  соответствует “фоновой” концентрации  $l$ -й примеси за границами области. Элементы операторов продукции и деструкции  $P_l$ ,  $D_l: [0, T] \times \mathbf{R}_+^{N_c} \rightarrow \mathbf{R}_+$  определены моделью трансформации.

В качестве функции неопределенности будем рассматривать вектор-функцию  $\mathbf{q} = \{r_l(\mathbf{x}, t)\}_{l=1}^{N_c}$ , имеющие смысл источников выбросов. Остальные параметры считаются заданными и обозначаются как  $\mathbf{v} = \{\mathbf{u}(\mathbf{x}, t), \{c_l(\mathbf{x}, t)\}_{l=1}^{N_c}, \{c_l^0(\mathbf{x})\}_{l=1}^{N_c}, \{D_l\}_{l=1}^{N_c}\}$ .

Оператором прямой задачи назовем оператор  $\mathcal{H}[\mathbf{q}; \mathbf{v}]$ , который при заданных параметрах  $\mathbf{v}$  отображает искомую функцию распределения источников выбросов  $\mathbf{q} \in Q$  в решение задачи (1)–(5). Задачу вычисления  $\mathcal{H}[\mathbf{q}; \mathbf{v}]$  по  $\mathbf{q} \in Q$  назовем прямой. Здесь  $Q$  обозначает множество предполагаемых искомым функций источников выбросов. В нашем случае источники выбросов  $\mathbf{q}$  из  $Q$  считаются стационарными ( $r_l(\mathbf{x}, t) = r_l(\mathbf{x})$ ), неотрицательными ( $r_l(\mathbf{x}, t) > 0$ ); они выбрасывают определенный набор веществ  $L_{src} \subseteq L$ , ( $r_l(\mathbf{x}, t) = 0, l \notin L_{src}$ ). Источники предполагаются только в “приземном” слое модели. Пусть в  $Q$  существует некоторая истинная функция источника, обозначим ее  $\mathbf{q}^{(*)} \in Q$ , и  $\mathbf{q}^{(*)} = \mathcal{H}[\mathbf{q}^{(*)}; \mathbf{v}]$  — соответствующее решение прямой задачи с функцией источника  $\mathbf{q}^{(*)}$ .

Определим оператор наблюдений  $H(\cdot)$ , который связывает функцию состояния модели с результатами наблюдений, и сформулируем обратную задачу по решению уравнения

$$\mathbf{I} = H[\mathbf{q}^{(*)}; \mathbf{v}] + \mathbf{I} \tag{6}$$

относительно  $\mathbf{q}^{(*)}$ , где  $\mathbf{I}$  — известные данные измерений;  $\mathbf{I}$  — функция из некоторого множества допустимых значений, которая представляет оценку ошибок результатов и моделей измерений. Обычно  $\mathbf{I}$  неизвестна, но есть частичная информация о ней, например норма в некотором пространстве. Пусть данные измерений имеют вид пространственных “снимков”  $\mathbf{I}^{(m)}$  интегральных (total-column) концентраций для заданного набора веществ  $L_{meas}$ , сделанных в моменты времени  $t_m = \{t_m^m\}_{m=1}^M \in [0, T]$ :

$$\int_0^z c_l^{(*)}(x, y, z, t_m^m) dz = I_l^m(x, y) | (x, y) \in [0, X] \times [0, Y], l \in L_{meas} \subseteq L.$$

Детальную модель рассматриваем в виде оператора прямой задачи с неизвестным внутренним устройством, который отображает заданные параметры  $\hat{\mathbf{v}}$  и функцию неопределенности  $\hat{\mathbf{q}} = \hat{\mathbf{Q}}$  на функцию состояния  $\hat{\mathbf{I}} = \hat{\mathbf{I}}[\hat{\mathbf{q}}; \hat{\mathbf{v}}]$ . В данной работе оператор вычисляется с помощью модели WRF-Chem 4.2.1 [11, 16], но может использоваться и любая другая подходящая модель. Требуется найти  $\hat{\mathbf{q}}^{(*)}; \hat{\mathbf{v}}$  такое, что

$$\mathbf{I} = \hat{H}[\hat{\mathbf{q}}^{(*)}; \hat{\mathbf{v}}] \quad \mathbf{I}, \quad (7)$$

где  $\hat{H}$  — оператор измерений для детальной модели. Далее для краткости детальную модель будем называть WRF, а упрощенную — IMDAF.

## 2.2. Идентификация источников для пары моделей

Для нахождения источника применим алгоритм решения обратной задачи (6) для данных  $\mathbf{I}$ , получая результат  $\mathbf{q}^{(*)}$ . Для решения обратной задачи используется модификация алгоритма на основе операторов чувствительности и ансамблей сопряженных уравнений для модели (1)—(5) из [3, 4] и других работ авторов. Особенностью алгоритма является использование системы функций для предварительного проектирования данных (извлечения характерных особенностей). В данной версии алгоритма система проектирования имеет два параметра:  $\text{Snapshot}^{(x)}$  и  $\text{Snapshot}^{(y)}$ , которые определяют пространственное разрешение рассматриваемых данных. Для любого изображения  $x$  и  $y$  находятся в диапазонах  $0, \dots, \text{Snapshot}^{(x)} - 1$  и  $0, \dots, \text{Snapshot}^{(y)} - 1$  соответственно.

Так как априорная информация IMDAF отличается от априорной информации WRF, то требуется некоторая процедура подстройки (обучения или “калибровки”) IMDAF для того, чтобы решения прямых задач обеих моделей соответствовали друг другу с учетом процедур отображения переменных ( $\hat{\mathbf{I}}[0, \hat{\mathbf{v}}] = \hat{\mathbf{I}}[\mathbf{q}, \mathbf{v}]$ ). Рассмотрим более слабое соответствие между переменными:  $\hat{\mathbf{I}}^{(0)} = \hat{H}[0, \hat{\mathbf{v}}] = H[\mathbf{q}, \mathbf{v}]$ . Для установления этого соответствия применим алгоритм решения обратной задачи (6) для данных  $\hat{\mathbf{I}}^{(0)}$ , получая результат  $\mathbf{q}^{(*)}$ .

Для того чтобы удалить из результата решения обратной задачи с моделью IMDAF вклад разницы между моделями, будем искать значение функции неопределенности для модели WRF в виде

$$\mathbf{q}^{(*)} = (\mathbf{q}^{(*)} \quad \underline{\mathbf{q}}^{(*)}).$$

Выберем так, чтобы получить соотношение

$$\mathbf{I} = \hat{H}[(\mathbf{q}^{(*)} \quad \underline{\mathbf{q}}^{(*)}), \hat{\mathbf{v}}].$$

Выберем его в предположении, что для  $\hat{\mathbf{I}}$  выполняется следующее соотношение:

$$\hat{\mathbf{I}}[\mathbf{q}, \hat{\mathbf{v}}] = \hat{\mathbf{I}}[0, \hat{\mathbf{v}}] + \hat{\mathbf{I}}[\mathbf{q}, \hat{\mathbf{v}}] - \hat{\mathbf{I}}[0, \hat{\mathbf{v}}].$$

Это соотношение, в частности, будет выполняться для модели (1)—(5) в отсутствие слагаемых, отвечающих за трансформацию примесей ( $P_l(t, \mathbf{x}) = 0, \mathcal{I}_l(t, \mathbf{x}) = 0$ ). Рассмотрим разность

$$\begin{aligned} & \mathbf{I} - \hat{\mathbf{I}}^{(0)} \hat{H} [(\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}), \hat{\mathbf{v}}] \hat{H} \hat{\mathbf{v}} \\ & = \hat{H} [(\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}), \hat{\mathbf{v}}] \hat{H} [0, \hat{\mathbf{v}}] \\ & \hat{H} [(\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}), \hat{\mathbf{v}}] \hat{H} \hat{\mathbf{v}} = (\hat{H} [(\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}), \hat{\mathbf{v}}] \hat{H}^{(0)}). \end{aligned}$$

Следовательно, уточняющий коэффициент и оценку решения задачи идентификации источника определим из соотношения

$$= \frac{\|\mathbf{I} - \hat{\mathbf{I}}^{(0)}\|}{\|\hat{H} [(\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}), \hat{\mathbf{v}}] \hat{H}^{(0)}\|}, \hat{\mathbf{q}}^{(1)} = \gamma^{(1)} (\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}). \quad (8)$$

Таким образом, приходим к следующему алгоритму (рис. 1):

1. Вычисляем фоновое решение по модели WRF:  $\hat{\varphi}^{(0)} = \hat{H} [0, \hat{\mathbf{v}}]$  и генерируем на его основе данные измерений

$$\hat{\mathbf{I}}^{(0)} = \hat{H} \hat{\varphi}^{(0)}.$$

2. Решаем (параллельно) обратные задачи по идентификации источников для данных измерений  $\mathbf{I}$  и  $\hat{\mathbf{I}}^{(0)}$ , получая решения  $\mathbf{q}^{(1)}$  и  $\underline{\mathbf{q}}^{(1)}$  соответственно.

3. Вычисляем значение  $\hat{H} [(\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}), \hat{\mathbf{v}}]$ .

4. Определяем коэффициент коррекции  $\gamma^{(1)}$  согласно (8).

5. В качестве решения рассматриваем

$$\hat{\mathbf{q}}^{(1)} = \gamma^{(1)} (\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}), \hat{\varphi}^{(1)} = \hat{H} [(\mathbf{q}^{(1)} - \underline{\mathbf{q}}^{(1)}), \hat{\mathbf{v}}].$$

### 2.3. Отображение переменных моделей

Для модели WRF-Chem 4.2.1 [11, 16] взята та же конфигурация, что и в работе [3]. Для отображения переменных модели WRF на параметры  $\mathbf{u}$  и  $\mathbf{v}$  модели IMDAF использована та же процедура, что и в работе [3]. Для задания начальных и граничных условий газовых и аэрозольных компонентов в модели WRF берут результаты глобальных моделей, например MOZART4, WACCM, CAM-Chem, GEOS-Chem и др. Эти модели генерируют трехмерные поля концентраций для различных химических веществ. Множество узлов грубой сетки IMDAF является подмножеством узлов сетки WRF. При переходе от сетки WRF к сетке IMDAF пространственная

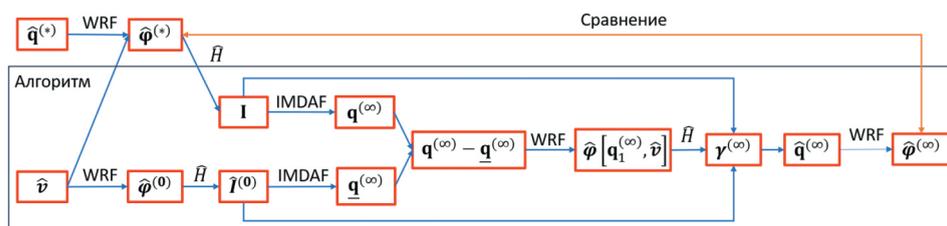


Рис. 1. Схема работы алгоритма.

и временная области остаются теми же, но количество узлов сетки по пространству и времени уменьшается в каждом направлении в требуемое количество раз.

Для задания данных измерений, начальных и граничных значений сопоставим функции состояния моделей. Размерности функции состояния в модели WRF-Chem —  $млн^{-1}$ , в модели IMDAF —  $мг/м^3$ . Для сопоставления значения функций состояния моделей мы использовали соотношения

$$\hat{c}_X = \frac{p}{RT} \frac{M_X}{1000} \hat{c}_X^W$$
, где  $\hat{c}_X^W$  ( $млн^{-1}$ ) — концентрация вещества  $X$  в модели WRF;  $\hat{c}_X$  ( $мг/м^3$ ) — концентрация вещества  $X$  в модели IMDAF;  $M_X$  ( $г/моль$ ) — молярная масса вещества  $X$ ;  $p = 101325$  Па,  $R = 8,314$  Дж/(моль К),  $T = 298,15$  К.

Как и в случае с параметрами адвекции-диффузии, начальное состояние отображается на грубую сетку через исключение значений в узлах, не попадающих в грубую сетку. Значения берутся для начальных значений решения  $\hat{c}^i[0, \hat{v}^i]$ . Для задания фоновых значений  $\hat{c}^i$  используются минимальные значения функции состояния  $\hat{c}^i[0, \hat{v}^i]$  во всей пространственно-временной области. В операторе измерений  $\hat{H}$  также используются преобразованные значения  $\hat{c}^i[\mathbf{q}, \hat{v}^i]$  на более грубой сетке.

$$I_{ij}^m = \sum_{i=1}^{Nlev} c^{(m)}(x_i, y_j, z_i) \quad i = 1, \dots, Nlong, j = 1, \dots, Nlat,$$

где  $(x_i, y_j, z_i)$  — координаты узла грубой сетки;  $Nlong, Nlat, Nlev$  — количество узлов грубой сетки по широте, долготе и высоте;  $c^{(m)}$  — индекс измеряемого вещества.

Восстановленные источники обратно отображаются с сетки модели IMDAF на сетку WRF. В модели IMDAF источник предполагается объемным и имеет размерность  $мг/(с \cdot м^3)$ , размерность функции площадного источника модели WRF —  $моль/(км^2 \cdot ч)$ . При переходе с грубой сетки на тонкую восстановленные функции источников сначала линейно интерполируются на тонкую сетку, после этого проводится преобразование значений. Если в узле сетки IMDAF задана некоторая скорость изменения массы газа на единицу объема, то сначала она пересчитывается в терминах производства в слое высоты  $h$  ячейки определенного количества молекул в секунду в предположении, что все молекулы производятся в узле сетки, как в точечном источнике. При переходе на сетку WRF разыскивается ближайший узел новой сетки и предполагается, что все молекулы образуются в этом узле, и значение пересчитывается, исходя из площади соответствующей ячейки. Мощности источников сопоставляются по формуле

$$\hat{q}_X = \frac{1000Sh}{\hat{S}} \cdot 3600 \cdot \frac{q_X}{M_X},$$

где  $M_X$  ( $г/моль$ ) — молярная масса вещества  $X$ ;  $S$  ( $м^2$ ) — площадь ячейки сетки IMDAF;  $h = 1$  м — высота слоя работы объемного источника;  $\hat{S}$  ( $м^2$ ) — площадь ячейки горизонтальной сетки WRF;  $\hat{q}_X$  ( $моль/(км^2 \cdot ч)$ ) — мощность источника  $X$  в модели WRF;  $q_X$  ( $мг/(с \cdot м^3)$ ) — мощность источника  $X$  в модели IMDAF.

### 2.4. Сценарий обратного моделирования

В численных экспериментах рассматривалась область, изображенная на рис. 2. По данным базы EDGAR на поверхности задаются постоянные на интервале истинные источники  $\text{NO}_2$  модели WRF. В качестве модели трансформаций используется RADM2 and Made/sorgamaerosols (chem\_opt = 2).

Далее с помощью модели WRF рассчитывались изменения полей концентраций на интервале 0 ч 14 февраля — 18 ч 16 февраля 2024 г. Сетка модели WRF составляет  $397 \times 32 \times 79 \times 99$  для  $N_t \times N_{lev} \times N_{long} \times N_{lat}$ . По данным измерения концентраций  $\text{NO}_2$  ( $L_{\text{meas}} = \{\text{NO}_2\}$ ) вычислялись данные интегральных измерений по высоте (total-column), которые измеряются в моменты  $t^m$ ,  $m = 1, \dots, 5$ , через равные промежутки времени, исключая границы. Рассмотрены системы проектирования с параметрами пространственного разрешения  $\left\{ \begin{matrix} (x) \\ \text{Snapshot} \end{matrix} \right\} \left\{ \begin{matrix} (y) \\ \text{Snapshot} \end{matrix} \right\} \{8, 8, 16, 16, 18, 18, 20, 20\}$ . Начальное предположение о распределении источников в обратной задаче нулевое ( $\mathbf{q}^{(0)} = 0$ ).

В модели (1)—(5) рассматривается взаимодействие между  $N_c = 5$  химическими соединениями  $\{\text{O}_3, \text{NO}, \text{NO}_2, \text{O}_2, \text{O}_3\text{P}\}$ . Константы скоростей фотохимических реакции зависят от времени аналогично [3]. Сетка модели IMDAF  $986 \times 5 \times 39 \times 49 \times 16$  для  $N_t \times N_c \times N_{lev} \times N_{long} \times N_{lat}$ . Количество неизвестных  $39 \times 49$ . Общее количество данных измерений в обратной

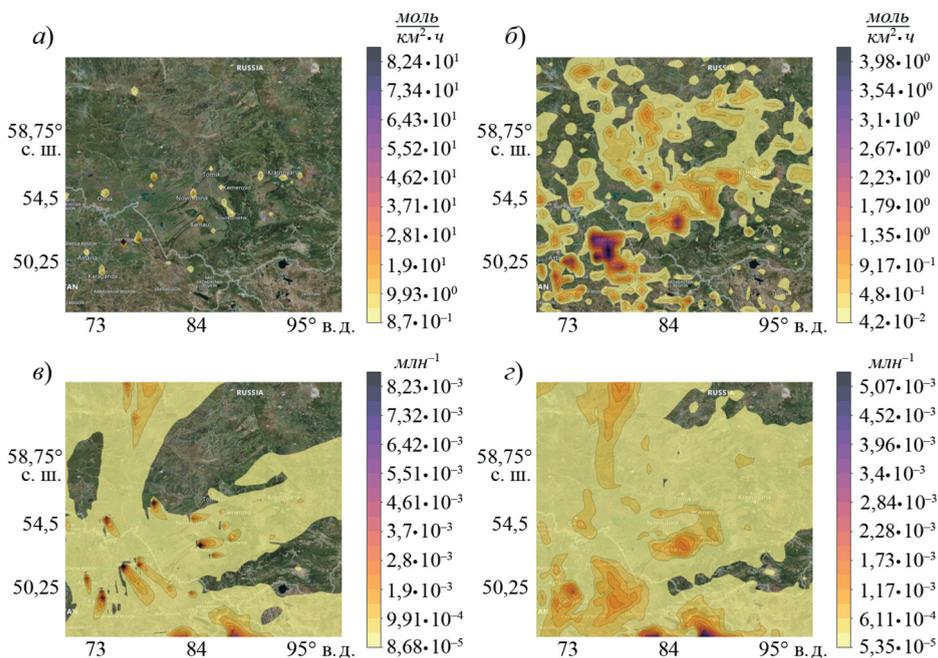


Рис. 2. Значение истинного источника  $\hat{\text{NO}}_2^{(*)}$  (a) и результат его восстановления  $\hat{\text{NO}}_2^{(*)}$  (б), а также значение истинной функции состояния  $\hat{\text{NO}}_2^{(*)}$  (в) и восстановленной функции состояния  $\hat{\text{NO}}_2^{(*)}$  (г) для  $\left\{ \begin{matrix} (x) \\ \text{Snapshot} \end{matrix} \right\} \left\{ \begin{matrix} (y) \\ \text{Snapshot} \end{matrix} \right\} = 20 \times 20$  в момент времени 17 ч 40 мин 15 февраля 2024 г.

задаче 39 49 5. В численных экспериментах учитывается <sup>(x)</sup> Snapshot <sup>(y)</sup> Snapshot  
 5 агрегатов данных.

### 3. Результаты

На рис. 2 приведено сравнение истинных решений для источников и функций состояния с восстановленными источниками. Можно отметить, что в восстановленном источнике теряется его детальная точечная структура, и вместо локализованных источников они восстанавливаются как объединенные в кластеры и сглаженные. Тем не менее общая локализация кластеров источников для кластеров населенных пунктов и промышленных районов региона прослеживается. Авторы считают, что эта разница между истинным и восстановленным решениями вызвана, во-первых, различием в моделях, так как данные измерений генерируются моделью WRF и восстанавливаются моделью IMDAF, и, во-вторых, свойством задачи, в которой информация от источника к измерениям передается посредством процессов адвекции-диффузии-реакции и усреднений по высоте, теряющих информацию о точности источников. В предыдущих работах авторов [3, 4] проведены эксперименты с восстановлением по данным измерений, сгенерированных моделью IMDAF. При этом в работе [3] для получения истинного решения к полю из точечных источников также дополнительно применяли свертку с гауссовым распределением. В обоих случаях при использовании достаточно большого количества функций проектирования восстановление источника получалось практически точным (с относительными ошибками порядка  $10^{-11}$  [3]).

Чтобы гарантировать точечный характер восстановленных источников, можно явно разыскивать не функцию источника, распределенного в пространстве, а наборы координат и мощностей отдельных источников [13]. Ранее авторы статьи также представили алгоритм, когда функция распределенного источника с помощью нейросети, обученной на основе оператора чувствительности, преобразуется в набор точечных источников, и задача сводится к определению их мощностей при фиксированных локализациях. Как вариант, для задания локализаций можно использовать известные базы данных источников с последующим уточнением их мощностей.

Анализ работ других авторов [8—10, 14, 15], посвященных идентификации источников выбросов на основе данных измерений, показал, что результаты восстановления, которые определяются в том числе и как уточнения (increments) по отношению к известным базам источников выбросов, также имеют сглаженный характер. Следовательно, результаты, полученные авторами статьи, качественно согласуются с аналогичными результатами других авторов.

На рис. 3 приведена зависимость от модельного времени абсолютных ошибок восстановления функции состояния  $^{-p}$  алгоритмом для различных значений параметра <sup>(x)</sup> Snapshot <sup>(y)</sup> Snapshot :

$$^{-p} \quad \left\| \hat{q}_{\text{NO}_2}^{(x)}(\cdot, t^p) - \hat{q}_{\text{NO}_2}^{(y)*}(\cdot, t^p) \right\|_{3D},$$

$$p = 0, \dots, N_t - 1.$$

Для вычисления нормы ошибки используется евклидова норма для сеточных функций.

В таблице представлены относительная ошибка восстановления функции состояния за весь временной интервал

$$= \frac{\| \hat{q}_{NO_2}^{(x)} - \hat{q}_{NO_2}^{(y)} \|_{4D}}{\| \hat{q}_{NO_2}^{(y)} \|_{4D}},$$

где  $\| \cdot \|_{4D}$  — евклидова норма для сеточных функций, и время решения обратной задач, измеренное во временах решения прямой задачи.

Расчеты проводились на двух ЭВМ. Обратная задача в системе IMDAF решалась на кластере NKS-1P Сибирского суперкомпьютерного центра (ССКЦ) с использованием трех узлов Intel Xeon Gold 6248 R (каждый имеет 2 процессора 24 ядра 2 вычислительных потока, 3,00 ГГц, 384 Гб оперативной памяти). Общее количество используемых ядер — 144. Узлы соединены с помощью Cluster Interconnect Omni-Path со скоростью 100 Гбит/с. Модель WRF запускалась на сервере с характеристиками 2 Intel(R) Xeon(R) Gold 5220 CPU @ 2.20GHz (всего 36 ядер 2 вычислительных потока), 768 Гб оперативной памяти.

Решение прямой задачи для модели WRF требует примерно 1800 с 0,5 ч (на 64 вычислительных потоках). Для сравнения также проведено решение обратной задачи в системе IMDAF на сервере. Расчет прямой задачи в

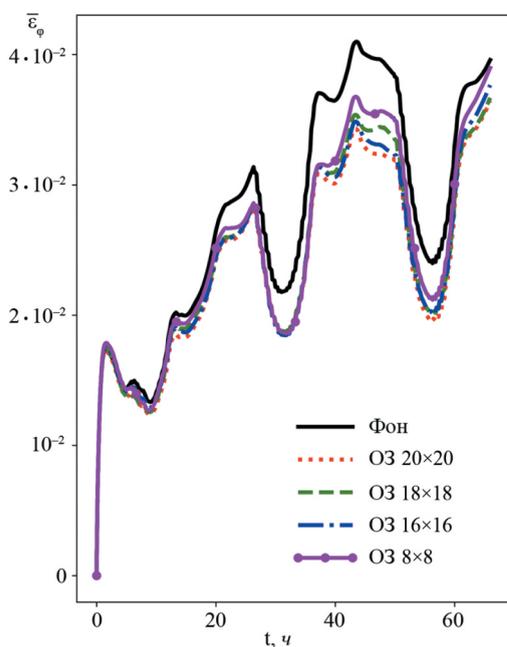


Рис. 3. Зависимость абсолютной ошибки оценки функции состояния модели WRF для разных Snapshot<sup>(x)</sup> и Snapshot<sup>(y)</sup> и для решения без использования данных (Фон); O3 — обратная задача.

модели IMDAF занимает примерно 10 с (на одном вычислительном потоке). При этом решение обратной задачи для параметров систем проектирования 8 8 занимает 13469 с 3,74 ч. На ССКЦ решение прямой задачи в среднем требует тех же 10 с (на одном вычислительном потоке), в то время как решение обратной задачи для 8 8 требует 9006 с 2,5 ч, для 16 16 — 19321 с 5,4 ч и для 20 20 — 31236 с 8,7 ч. Если предположить, что при использовании для решения обратных задач модели WRF время вычислений, измеренное в решении прямых задач, будет тем же, то такой расчет при параметрах 16 16 потребовал бы 2013 1800 с 1006 ч. В разработанной версии алгоритма для его работы требуется (параллельно) решить две обратных задачи в системе IMDAF и последовательно три прямых

**Относительная ошибка восстановления функции состояния  
и время решения обратной задачи (ОЗ), измеренное во временах решения  
прямой задачи**

Параметр системы проектирования	Относительная ошибка восстановления функции состояния	Время решения обратной задачи в прямых задачах на ССКЦ
Прямая задача	0,35	1
ОЗ 8 8	0,32	967
ОЗ 16 16	0,30	2013
ОЗ 18 18	0,31	1992
ОЗ 20 20	0,30	2756

задачи для модели WRF. Приближенное время вычислений для 16 16 составляет  $5,4 ч + 3 \cdot 0,5 ч = 6,9 ч$ , что в 9,6 раза меньше реального времени (66 ч).

В дальнейшем планируется рассмотреть конфигурацию алгоритма, где найденные значения функций неопределенности для фонового решения будут непосредственно включаться в процедуру решения обратной задачи в качестве априорных значений (в виде  $f$ ). В этом случае обратные задачи по восстановлению фона и уточнению источников придется решать последовательно, однако это позволит более точно учесть нелинейный характер задачи. Также, так как фоновое решение полностью доступно, при восстановлении фоновых значений можно использовать оператор измерений, который получает четырехмерный массив изменения концентраций химического вещества. В работе алгоритм рассматривается в одном окне усвоения. В будущем планируется применить его в составе алгоритма усвоения данных на последовательности окон усвоения.

Таким образом, подход использования нескольких моделей для обратного моделирования процессов переноса и трансформации примесей показал свою работоспособность. Определены пути его дальнейшего развития.

Разработка алгоритма решения обратных задач для трехмерных моделей выполнена в рамках государственного задания Института вычислительной математики и математической геофизики СО РАН FWNM-2025-0003, в части использования модели WRF-Chem и задания региональных сценариев обратного моделирования — при поддержке гранта Российского научного фонда 23-77-30008.

## Литература

- Марчук Г. И.** Математическое моделирование в проблеме окружающей среды. — М., Наука, 1982, 319 с.
- Марчук Г. И.** О постановке некоторых обратных задач. — Доклады АН СССР, 1964, т. 156, № 3, с. 503—506.
- Пененко А. В., Гочаков А. В., Антохин П. Н.** Алгоритм усвоения данных на основе оператора чувствительности для трехмерной модели переноса и трансформации примесей в атмосфере. — Оптика атмосферы и океана, 2024, т. 37, № 9, с. 719—728.
- Пененко А. В., Пененко В. В., Цветова Е. А.** Алгоритмы усвоения данных для моделей химии атмосферы. — Известия РАН. Физика атмосферы и океана, 2025 (принята в печать).

- 5. Baklanov A., Korsholm U. S., Nuterman R., Mahura A., Nielsen K. P., Sass B. H., Rasmussen A., Zakey A., Kaas E., Kurganskiy A., Sorensen B., and Gonzalez-Aparicio I.** The Enviro-HIRLAM online integrated meteorology–chemistry modelling system: Strategy, methodology, developments, and applications (v. 7.2). — *Geosci. Model Dev.*, 2017, vol. 10, pp. 2971–2999; doi: 10.5194/gmd-10-2971-2017.
- 6. Bocquet M., Elbern H., Eskes H., Hirtl M., Zabkar R., Carmichael G. R., Flemming J., Inness A., Pagowski M., Perez Camano J. L., Saide P. E., San Jose R., Sofiev M., Vira J., Baklanov A., Carnevale C., Grell G., and Seigneur C.** Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled chemistry meteorology models. — *Atmos. Chem. Phys.*, 2015, vol. 15, pp. 5325–5358; doi: 10.5194/acp-15-5325-2015.
- 7. Cheng S., Quilodran-Casas C., Ouala S., Farchi A., Liu V., Tandeo P., Fablet R., Lucor D., Iooss B., Brajard J., Xiao D., Janjic T., Ding W., Guo Y., Carrassi A., Bocquet M., and Arcucci R.** Machine learning with data assimilation and uncertainty quantification for dynamical systems: A review. — *IEEE/CAA J. of Automatica Sinica. Institute of Electrical and Electronics Engineers (IEEE)*, 2023, vol. 10, No. 6, pp. 1361–1387; doi: 10.1109/JAS.2023.123537.
- 8. Cheng X., Hao Z., Zang Z., Liu Z., Xu X., Wang S., Liu Y., Hu Y., and Ma X.** A new inverse modeling approach for emission sources based on the DDM-3D and 3DVAR techniques: An application to air quality forecasts in the Beijing–Tianjin–Hebei region. — *Atmos. Chem. Phys.*, 2021, vol. 21, No. 18, pp. 13747–13761; doi: 10.5194/acp-21-13747-2021.
- 9. Dammers E., Tokaya J., Mielke C., Hausmann K., Griffin D., McLinden C., Eskes H., and Timmermans R.** Can TROPOMI NO<sub>2</sub> satellite data be used to track the drop in and resurgence of NO<sub>x</sub> emissions in Germany between 2019–2021 using the multi-source plume method (MSPM)? — *Geosci. Model Develop.*, 2024, vol. 17, No. 12, pp. 4983–5007; doi: 10.5194/gmd-17-4983-2024.
- 10. Fortems-Cheiney A., Broquet G., Potier E., Plauchu R., Berchet A., Pison I., Denier van der Gon H., and Dellaert S.** CO anthropogenic emissions in Europe from 2011 to 2021: Insights from Measurement of Pollution in the Troposphere (MOPITT) satellite data. — *Atmos. Chem. Phys.*, 2024, vol. 24, No. 8, pp. 4635–4649; doi: 10.5194/acp-24-4635-2024.
- 11. Grell G. A., Peckham S., Schmitz R., Mckeen S., Frost G., Skamarock W., and Eder B.** Fully coupled “online” chemistry within the WRF model. — *Atmos. Environ.*, 2005, vol. 39, No. 37, pp. 6957–6975; doi: 10.1016/j.atmosenv.2005.04.027.
- 12. Knote C., Brunner D., Vogel H., Allan J., Asmi A., Aijala M., Carbone S., Gon H., Jimenez J., Kiendler-Scharr A., Mohr C., Poulain L., Prevot A., Swietlicki E., and Vogel H.** Towards an online-coupled chemistry-climate model: Evaluation of trace gases and aerosols in COSMO-ART. — *Geosci. Model Develop.*, 2011, vol. 4, No. 4, pp. 1077–1102; doi: 10.5194/gmd-4-1077-2011.
- 13. Mamonov A. V. and Tsai Y.-H. R.** Point source identification in nonlinear advection-diffusion-reaction systems. — *Inverse Probl.*, 2013, vol. 15, No. 4, p. 35009; doi: 10.1088/0266-5611/29/3/035009.
- 14. Rey-Pommier A., Chevallier F., and Ciais P., Broquet G., Christoudias T., Kushta J., Hauglustaine D., and Sciare J.** Quantifying NO<sub>x</sub> emissions in Egypt using TROPOMI observations. — *Atmos. Chem. Phys.*, 2022, vol. 22, No. 17, pp. 11505–11527; doi: 10.5194/acp-22-11505-2022.
- 15. Savas D., Dufour G., Coman A., Siour G., Fortems-Cheiney A., Broquet G., Pison I., Berchet A., and Bessagnet B.** Anthropogenic NO<sub>x</sub> emission estimations over east China for 2015 and 2019 using OMI satellite observations and the new inverse modeling system CIF-CHIMERE. — *Atmosphere*, 2023, vol. 14, No. 1, p. 154; doi: 10.3390/atmos14010154.
- 16. Skamarock W. C., Klemp J. B., Dudhia J., Gill D. O., Barker D. M., Duda M. G., Huang X.-Y., Wang W., and Powers J. G.** A Description of the Advanced Research WRF Version 4. /NCAR Technical Note. — Boulder, Colorado, NCAR, 2019; doi: 10.6084/m9.figshare.7369994.v4.