

УДК 519.6, 551.46

АЛГОРИТМЫ УСВОЕНИЯ ДАННЫХ ДЛЯ МОДЕЛЕЙ ХИМИИ АТМОСФЕРЫ

© 2025 г. А. В. Пененко^{a, b, *}, В. В. Пененко^a, Е. А. Цветова^a

^aИнститут вычислительной математики и математической геофизики СО РАН,
пр. Лаверентьева, 6, Новосибирск, 630090 Россия

^bЮгорский государственный университет, ул. Чехова, 16, Ханты-Мансийск, 628012 Россия

*email: aleks@ommgp.sgcc.ru

Поступила в редакцию 29.07.2024 г.

После доработки 15.01.2025 г.

Принята к публикации 05.03.2025 г.

Задача усвоения данных для нестационарных моделей переноса и трансформации примесей рассматривается как последовательность связанных обратных задач о восстановлении пространственно-временной структуры функций состояния с учетом поступающих в ходе моделирования данных измерений. Усвоение данных осуществляется вместе с идентификацией дополнительной искомой функции источника, которую мы называем функцией неопределенности модели. Целью работы является краткий исторический обзор и представление актуальной версии алгоритма усвоения данных для моделей химии атмосферы на основе операторов чувствительности и ансамблей решений сопряженных уравнений с демонстрацией работы алгоритма для трехмерной модели с нелинейным оператором измерений в одном из сценариев моделирования для Байкальского региона.

Ключевые слова: усвоение данных, модели переноса и трансформации примесей, сопряженные уравнения, оператор чувствительности, обратное моделирование, данные измерений

DOI: 10.31857/S0002351525030091

ВВЕДЕНИЕ

Академик Г.И. Марчук, известный ученый и организатор науки, создал крупную школу российских математиков в нескольких областях знаний. Пожалуй, самая многочисленная часть его учеников и продолжателей исследует проблемы вычислительной математики и решает задачи физики атмосферы и океана, охраны окружающей среды и климата. В частности, эта тематика исследований продолжает развиваться двумя ветвями, в ИВМ РАН и ИВМиМГ СО РАН (бывшем ВЦ СО АН СССР). В настоящей статье мы даем краткий экскурс в нашу общую историю и систематизируем текущее состояние исследований новосибирской ветви школы Г.И. Марчука по усвоению данных наблюдений и смежным вопросам.

Отдавая должное оригинальности и значимости теоретических и прикладных результатов

наших коллег из ИВМ РАН, рекомендуем заинтересованным читателям ряд монографий: по проблемам динамики Земной системы и климата [Марчук и др., 1984; Марчук и др., 1987; Дымников и Залесный, 2019]; по тематике сопряженных уравнений, оптимального управления и вариационного усвоения данных [Марчук и др., 1993; Шутяев, 2001; Агошков, 2003, и т.д.]; по проблемам окружающей среды [Алоян, 2008]. Это далеко не полный список публикаций, на который хочется обратить внимание специалистов. Есть еще многочисленные статьи, содержащие оригинальные результаты, в том числе и с зарубежными коллегами, в известных российских и зарубежных рецензируемых журналах.

На первом этапе работы в Новосибирске Г.И. Марчук инициировал в институте работы по созданию численных методов и алгоритмов для моделирования гидродинамических процессов, сначала в атмосфере, а затем и в океа-

не. Большое внимание было уделено актуальнейшим вопросам прогнозирования погоды численными методами. Здесь был достигнут большой прогресс, поскольку модели, разработанные в ВЦ СО АН СССР были переданы в Западно-Сибирское Управление ГМС. Впоследствии, как естественное развитие тематики по совместному решению задач атмосферы и океана, появилась возможность обратиться к вопросам исследования климата. С этой целью была организована группа специалистов, составивших позже ядро сотрудников ИВМ.

Проблема охраны окружающей среды находилась в сфере научных интересов Г.И. Марчука на протяжении всей его жизни. Будучи директором, он создал в нашем институте специальную лабораторию, основным направлением деятельности которой была разработка численных моделей и методов для решения природоохранных задач. Он сам непосредственно участвовал в развитии тематики и здесь его вклад бесценен [Марчук, 1964, 1974, 1982, 1992]. Особо следует отметить такие разделы, как сопряженные уравнения, вариационное усвоение данных наблюдений, обратные задачи, теория чувствительности. На протяжении многих лет мы развиваем концепцию природоохранного прогнозирования и проектирования, опираясь на основные идеи Г.И. Марчука в этой области.

На начальном этапе работы лаборатории усилия концентрировались на создании моделей гидродинамики атмосферы, их конкретных реализациях и приложениях к проблемам городов и промышленных регионов. Здесь следует упомянуть большое влияние работ Г.И. Марчука в области вычислительной математики [Марчук, 1974, 1988] в части постановок задач и разработке численных алгоритмов. Безусловно перспективным являлось использование методов расщепления, без которых и сейчас не обходится практически ни одна сложная современная модель.

Природоохранные задачи предусматривают оценку качества среды, и, в первую очередь, атмосферного воздуха, который связан с загрязнением от источников естественного и антропогенного характера. Для этой цели система уравнений гидродинамики дополнялась уравнением распространения примеси. Последняя сначала рассматривалась обобщенно, как трассер, переносимый воздушными пото-

ками, рассчитанными по гидродинамической модели. Впоследствии к гидродинамической модели добавились модели химической кинетики, которые были созданы в сотрудничестве с коллегами из Института химической кинетики и горения Сибирского отделения АН СССР. Отработка методологии решения задач атмосферной химии осуществлялась на примерах процессов превращения веществ, типичных для атмосферы индустриально нагруженных городов и регионов: трансформации серу- и азотсодержащих веществ и фотохимического окисления метана при различных вариантах энерго- и теплоснабжения [Пененко и др., 1984; Пененко и Алоян, 1985; Бажин и др., 1991; Пененко и др., 1997].

Решение совместных задач динамики и химии атмосферы [Baklanov et al., 2014; Baklanov et al., 2018] предъявляет дополнительные требования ко всем алгоритмам реализации, в том числе и алгоритмам усвоения данных [Vocquet et al., 2015]. Перечислим лишь некоторые особенности этого класса задач.

- Химический состав атмосферы может быстро меняться, поэтому интерес представляют текущие и будущие состояния системы (оценка текущего состояния и прогноз “химической погоды”) в “реальном времени”.
- Разнообразии химических механизмов и их нелинейный характер; различные временные масштабы процессов порождают “жесткость” систем уравнений химической кинетики.
- Существенное влияние на поведение системы оказывают не только начальные данные, но и параметры моделей и источники примесей.
- Высокая размерность ($\approx 10^7$) современных моделей атмосферной химии, которая определяется пространственным разрешением их дискретных аналогов и количеством рассматриваемых веществ, вынуждает заботиться о вычислительной эффективности алгоритмов.
- Данных измерений обычно существенно меньше, чем искомым переменных. Кроме того, ошибка измерений концентраций может быть достаточно велика.

Вариационный формализм, восходящий в метеорологии к работам Ю. Сасаки [Sasaki, 1958, 1970], оказался полезным инструментом для создания комплексной системы моделирования. Для реализации комплекса природоохранных моделей в ВЦ СО АН СССР был разработан системный

подход, основанный на вариационном принципе [Пененко, 1975, 1981, 2009], объединяющий все элементы технологии математического моделирования. В формулировке вариационного принципа присутствуют несколько основных компонентов. Это модели процессов, представленные соответствующими системами уравнений с краевыми и начальными условиями (они входят в виде интегрального тождества); данные наблюдений и модели наблюдений; априорная информация обо всех искомым элементах системы моделирования; целевые критерии прогнозирования и проектирования, например, функциональные ограничения на качество окружающей среды, санитарно-экологические стандарты качества жизни и здоровья населения агломераций и др. С позиций вариационного принципа численные модели являются ограничениями на класс функций и связями между параметрами и функциями состояния.

В природоохранных задачах часто требуется использование комбинированных методов прямого и обратного моделирования. Прямое моделирование представляет модели различной степени сложности; методы обратного моделирования обеспечивают совместное использование моделей и данных наблюдений и, следовательно, реализуют более высокие системные уровни проблемы в целом. Основной функционал для организации системы моделирования формируется так, чтобы учесть в нем все модели и доступные данные наблюдений, а также минимизировать влияние неопределенностей, которые в них имеются.

Вариационный подход сводит задачи обратного моделирования к задачам минимизации целевого функционала. Для минимизации целевого функционала можно использовать как классические градиентные, так и различные метаэвристические алгоритмы (см., например, [Penenko et al., 2022]). Для построения градиентов целевых функционалов обычно используются сопряженные уравнения и теория чувствительности. Применение теории чувствительности позволяет связать вариации целевого функционала с вариациями параметров, входящих в рассматриваемую систему [Пененко, 1975; Марчук и Пененко, 1979; Marchuk and Penenko, 1981; Пененко и др., 1985].

Вариационные алгоритмы усвоения данных и первые примеры их реализации для задач фи-

зики атмосферы были представлены в [Пененко и Образцов, 1976; Marchuk and Penenko, 1981]. В работе 1976 г. был предложен вариационный алгоритм согласования полей метеоэлементов. Публикация 1981 г. появилась в монографии по материалам доклада и итогам конференции “Monsoon Dynamics”, которая состоялась в 1977 г. в Нью Дели. Работа посвящена применению теории возмущений и теории управления в численных моделях динамики атмосферы и океана. В численном примере было показано восстановление поля геопотенциала на сфере при отсутствии части данных наблюдений, относящихся к океану. Примечательно, что публикация появилась в главе “Некоторые важные методы математического моделирования (Some important mathematical modeling techniques)”.

Вариационное усвоение данных привлекло внимание зарубежных коллег [Le Dimet and Talagrand, 1986]. Впоследствии вариационное усвоение данных (3DVAR и 4DVAR) широко распространилось по миру и было внедрено в оперативную практику прогнозирования погоды в Европейском центре среднесрочных прогнозов (ECMWF) в Рединге (Великобритания) и в других научных центрах. Алгоритмы усвоения данных [Vocquet et al., 2015; Carrassi et al., 2018] широко используются для оперативного и квазиоперативного контроля качества воздуха [Elbern et al., 2014; Silver et al., 2015; Nguyen et al., 2018; Xing et al., 2020; Mijling, 2020; Nguyen and Soulhas, 2021].

Общая структура технологии моделирования, основанная на вариационном подходе, его подробное описание и построение с его помощью согласованных конечно-разностных версий математических моделей имеется в [Пененко, 1975, 1981]. Для построения численных схем мы применяем методы декомпозиции и расщепления, обеспечивающие им свойства суммарной аппроксимации и допускающие параллельную организацию вычислений [Penenko et al., 2014]. В частности, были разработаны и протестированы алгоритмы быстрого усвоения данных с вариационным усвоением на отдельных этапах схемы расщепления. В случае контактных измерений такая схема позволяет реализовывать безытерационные (т.н. “прямые” алгоритмы) [Пененко и Пененко, 2014; Пененко и др., 2016, 2018].

В современной версии алгоритмов мы рассматриваем задачу усвоения данных для нестационарных моделей в более общей формулировке: как последовательность связанных обратных задач восстановления пространственно-временной структуры функций состояния с учетом различных наборов данных измерений, поступающих в процессе моделирования. Ключевой особенностью этого класса задач является совпадение модельного времени и времени сбора данных. Усвоение данных осуществляется с помощью идентификации некоторой внутренней искомой функции модели, которую мы называем функцией неопределенности. Для решения обратных задач могут использоваться любые подходящие методы.

Одной из основных неопределенностей при изучении качества воздуха является информация об источниках (и стоках) химических веществ [Elbern et al., 2007; Vocquet et al., 2015]. Поэтому в качестве основной задачи обратного моделирования будем рассматривать обратную задачу идентификации источников [Марчук, 1982; Пененко и др., 1985; Pudykiewicz, 1998; Десятков и др., 1999; Issartel, 2003, 2005; Mamonov and Tsai, 2013; Turbelin et al., 2014; Kumar et al., 2015; Bieringer et al., 2017].

Системы мониторинга качества атмосферного воздуха различаются по своему временному и пространственному охвату, составу наблюдаемых химических веществ и точности получаемых данных. С помощью различных систем мониторинга собирается огромное количество разнородных данных, поэтому повышение эффективности использования этих данных является научной проблемой [WMO, 2015, 2018; Sokhi et al., 2022]. Следовательно, другим важным компонентом постановок задач обратного моделирования наряду с моделями изучаемых процессов, являются операторы (модели) измерений, связывающие измеряемые характеристики с некоторыми содержательными характеристиками атмосферы, например, с распределением концентраций примесей в некоторой пространственной области. Различным способом, измерительному оборудованию и сетям мониторинга с их режимами работы могут соответствовать различные операторы измерений. В частности, нелинейные операторы измерений могут возникать в задачах обработки данных дистанционного зондирования [Зуев и Зуев, 1992].

Наряду с оптимизационным подходом для решения обратных задач мы развиваем методика с использованием операторов чувствительности и ансамблей решений сопряженных уравнений [Пененко и др., 2019; Penenko, 2020]. Этот подход объединяет идеи из работ Г.И. Марчука, который сформулировал подход к обратным задачам на основе ансамблей сопряженных уравнений [Марчук, 1964; Marchuk and Penenko, 1981], идеи методов “смягчения” (mollification) [Murio, 1993] и концепцию извлечения структуры из данных (“image to structure operator” [Le Dimet et al., 2015]). Г.И. Марчук [Марчук, 1964] предложил объединить основанные на сопряженных задачах интерпретации отдельных элементов данных измерений в одно уравнение. В случае большого количества элементов данных (например, спутниковых снимков высокого разрешения) прямое применение подхода усложняется, поскольку требуется решить большое количество сопряженных задач, равное количеству рассматриваемых элементов данных (например, пикселей) и требуется агрегация данных. Ранее алгоритмы идентификации источников на основе ансамблей сопряженных уравнений были применены к линейной модели переноса примесей в [Issartel, 2005] и к нелинейной модели переноса-трансформации с точечными источниками и измерениями в [Mamonov and Tsai, 2013]. Кластерная реализация алгоритма на основе ансамблей сопряженных уравнений для линейной задачи идентификации источника в масштабе города была представлена в [Панасенко и Старченко, 2009]. Обзор других методов, основанных на ансамблях сопряженных уравнений, можно найти в [Bieringer et al., 2017].

Оператор чувствительности собирается из функций чувствительности, соответствующих агрегатам данных измерений, и позволяет связать вариацию искоемых величин (функций неопределенности) с вариацией наблюдаемых переменных модели. Для вычисления операторов чувствительности обратной задачи используются наборы согласованных, в смысле тождества Лагранжа, численных схем. Ансамблевая конструкция оператора позволяет естественно комбинировать различные типы данных измерений в одном операторном уравнении [Penenko et al., 2021], приводить к единому виду квазилинейных операторных уравнений широкий спектр обратных задач, использовать специали-

зированные алгоритмы для решения получающихся операторных уравнений [Penenko, 2020], естественно распараллеливать алгоритмы и организовывать их выполнение на высокопроизводительных кластерных системах [Penenko and Rusin, 2022]. Отдельно отметим возможность оценки областей наблюдаемости (информативность данных) на основе анализа свойств операторов чувствительности [Penenko, 2020; Penenko et al., 2021].

С помощью оценок информативности данных на основе операторов чувствительности и методов машинного обучения также был разработан гибридный алгоритм, позволяющий уточнять решение обратной задачи за счет включения дополнительной априорной информации [Penenko et al., 2024]. Это новое направление исследований Новосибирской школы соответствует активно развивающимся в настоящее время методам обратного моделирования на основе машинного обучения [Arridge et al., 2019; Liu et al., 2022; Cheng et al., 2023].

Для компьютерной реализации технологии моделирования в ИВМиМГ СО РАН разработаны и поддерживаются три разномасштабных и многоцелевых программных комплекса: AT-MES (Atmospheric Transport and dispersion Models of global and regional scales for climatic Environmental Studies) [Penenko and Tsvetova, 2008, Penenko et al., 2012, Пененко и др., 2015a,b]; MMAD&IT (Mesoscale Model of Atmospheric Dynamics and Impurity Transport) [Ryanova et al., 2014; Пененко и др., 2015b]; и IMDAF (Inverse Modeling and Data Assimilation Framework) с применением к задачам химии атмосферы [Penenko, 2020; Penenko et al., 2021, Penenko and Rusin, 2022; Penenko et al., 2024]. Завершаем наш краткий обзор утверждением, что разрабатываемые в школе Г.И. Марчука алгоритмы, модели и методы соответствуют идеям осуществления в будущем полного предсказательного моделирования Земной системы [Brunet, 2015; Mahura et al., 2024].

В следующих разделах статьи представлены новые результаты наших исследований, касающиеся решения обратных задач на основе операторов чувствительности и ансамблей решений сопряженных уравнений. Мы рассматриваем новые версии алгоритма для задачи усвоения данных для нелинейных моделей адвекции-диффузии-реакции примесей в атмосфере с учетом нелинейных операторов измере-

ний. Эта версия является развитием [Penenko et al., 2024b] предыдущих вариантов с линейными операторами измерений и трехмерной моделью и [Penenko and Rusin, 2023] с нелинейным оператором измерений.

МЕТОДЫ И АЛГОРИТМЫ

Постановки задач

Математическая модель для описания пространства N_c реагирующих химических веществ рассматривается в области $\Omega_T = \Omega \times [0, T]$, где Ω — достаточно гладкая аппроксимация ограниченной прямоугольной области $[0, X] \times [0, Y] \times [0, Z]$ в \mathbb{R}^3 , $T > 0$, $\partial\Omega$ — граница области Ω , и $[\underline{t}, \bar{t}] \subset [0, T]$:

$$\begin{aligned} \frac{\partial \varphi_l}{\partial t} - \nabla \cdot (\text{diag}(\mu_l) \nabla \varphi_l - \mathbf{u} \varphi_l) + P_l(t, \varphi) \varphi_l = \\ = \Pi_l(t, \varphi) + f_l + r_l, \quad (\mathbf{x}, t) \in \Omega \times [\underline{t}, \bar{t}], \end{aligned} \quad (1)$$

$$\mathbf{n} \cdot (\text{diag}(\mu_l) \nabla \varphi_l) = 0, \quad (\mathbf{x}, t) \in \Gamma^{(out)} \subset \partial\Omega \times [\underline{t}, \bar{t}], \quad (2)$$

$$\varphi_l = \alpha_l, \quad (\mathbf{x}, t) \in \Gamma^{(in)} \subset \partial\Omega \times [\underline{t}, \bar{t}], \quad (3)$$

$$\begin{aligned} \varphi_l = \varphi_l^0, \quad \mathbf{x} \in \Omega, t = \underline{t}, \\ l = 1, \dots, N_c \end{aligned} \quad (4)$$

где $t \in [0, T]$ — время, \mathbf{x} — пространственные координаты, N_c — число рассматриваемых веществ, $\varphi_l = \varphi_l(\mathbf{x}, t)$ — концентрация -того вещества в точке $(\mathbf{x}, t) \in \Omega_T$, φ — вектор, состоящий из элементов $\varphi_l(\mathbf{x}, t)$, $l = 1, \dots, N_c$, который называется функцией состояния, $L = \{1, \dots, N_c\}$. Вместо индекса далее мы будем также использовать обозначение соответствующего соединения. Функции $\mu_l(\mathbf{x}, t) \in \mathbb{R}^3$ соответствуют коэффициентам диффузии, $\text{diag}(\mathbf{a})$ — диагональная матрица с вектором \mathbf{a} на диагонали, $\mathbf{u}(\mathbf{x}, t) \in \mathbb{R}^3$ — вектор скорости ветра. Скорости должны удовлетворять условиям неразрывности сплошной среды. $\Gamma^{(in)}$ и $\Gamma^{(out)}$ — части границы области $\partial\Omega_T$, на которых вектор $\mathbf{u}(\mathbf{x}, t)$, соответственно, направлен внутрь области Ω_T или равен нулю, либо направлен из области Ω_T ; \mathbf{n} — вектор внешней нормали. Функции $\alpha_l(\mathbf{x}, t)$, $\varphi_l^0(\mathbf{x})$ описывают, соответственно, граничные и начальные условия, $f_l(\mathbf{x}, t)$ — известная *a priori*

функция источника; $r_l(\mathbf{x}, t)$ — неизвестная функция источника. В данном случае $\alpha_l(\mathbf{x}, t)$ соответствует “фоновой” концентрации l -той примеси за границами области. Элементы операторов продукции и деструкции $P_l, \Pi_l: [0, T] \times \mathbb{R}_+^{N_c} \rightarrow \mathbb{R}_+$ определены моделью трансформации. Выражения для операторов P, Π , полученные из кинетических уравнений для счетных концентраций, переписаны для работы с массовыми концентрациями.

Все параметры модели условно делятся на “заданные” и функции неопределенности, за счет которых можно уточнять математическую модель. Обозначим через \mathbf{q} вектор неопределенностей задачи — это вектор параметров модели, за счет которых предполагается решать задачи обратного моделирования. Вектор \mathbf{q} принадлежит некоторому заданному множеству Q . Здесь Q обозначает множество предполагаемых искомым функций источников выбросов. В нашем случае источники выбросов \mathbf{q} из Q предполагаются стационарными ($r_l(\mathbf{x}, t) = r_l(\mathbf{x})$), неотрицательными ($r_l(\mathbf{x}, t) > 0$); они выбрасывают определенный набор веществ $L_{src} \subset L$, ($r_l(\mathbf{x}, t) = 0, l \notin L_{src}$). Дополнительно известно, что источники находятся только в “приземном” слое модели.

Задачу определения функций φ из (1)–(4) по известным \mathbf{q} будем называть прямой. Обозначим ее решение через $\varphi[\mathbf{q}]$. Пусть $\varphi^{(*)} = \varphi[\mathbf{q}^{(*)}]$ — решение прямой задачи, соответствующее некоторому “истинному” значению функции неопределенности $\mathbf{q}^{(*)} \in Q$. Множество решений прямой задачи обозначим $\Phi = \{\varphi[\mathbf{q}] | \mathbf{q} \in Q\}$.

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

$$H\left(\varphi\left[\mathbf{q}^{(*)}\right]\right) = \mathbf{I} + \delta I, \quad (5)$$

относительно $\mathbf{q}^{(*)}$, где \mathbf{I} — известные данные измерений, δI — функция из некоторого множества допустимых значений, которая представляет оценку ошибок результатов и моделей измерений. Обычно она неизвестна, но известна частичная информация о ней, например,

норма в некотором пространстве. Пусть данные измерений имеют вид пространственных “снимков” полей концентраций $\varphi^{(*)}$, полученных определенным способом в моменты времени $\Theta = \{\theta^m\}_{m=1}^M \subset [0, T]$.

Определим множество W окон усвоения $[\underline{t}^{[k]}, \bar{t}^{[k]}] \subset [0, T]$, $k=1 \dots N_W$, которые не пересекаются, и $\bar{t}^{[k]} = \underline{t}^{[k+1]}$, $\underline{t}^{[1]} = 0$. Задача усвоения данных формулируется как последовательность связанных обратных задач на W . Искомые источники выбросов \mathbf{q} предполагаются постоянными в окнах усвоения. Для поиска источников используются данные измерений, попадающие в окно усвоения: $\{I_l^m | \theta^m \in [\underline{t}^{[k]}, \bar{t}^{[k]}]\}$. Функция состояния при переходе из одного окна усвоения в другое считается непрерывной: $\varphi(\bar{t}^{[k]}) = \varphi(\underline{t}^{[k+1]})$. В качестве начального приближения для решения в $k+1$ -ом окне усвоения используется результат решения на предыдущем окне $q^{[k+1](0)} = q^{[k](\infty)}$.

Задача усвоения данных, в отличие от обратной задачи, нацелена на численное решение в потенциально неограниченном по времени интервале ($T \rightarrow \infty$), соответствующем процессу мониторинга. Задача усвоения данных является более общей по отношению к обратной задаче (которая является ее частным случаем при $N_W = 1, [\underline{t}^{[1]}, \bar{t}^{[1]}] = [0, T]$). Алгоритм усвоения данных состоит из двух компонентов: алгоритма решения обратных задач на окнах усвоения и способа переноса результатов решения с одного окна усвоения на следующее для формирования априорной информации.

Решение обратных задач на основе операторов чувствительности с нелинейными операторами измерений

Схематично представим общую постановку обратной задачи. Пусть

$$L(\varphi^{(*)}) = \mathbf{q}^{(*)}, \quad (6)$$

$$H(\varphi^{(*)}) = \mathbf{I} + \delta I, \quad (7)$$

где (6) обозначает модель процессов (1)–(4), φ — функция состояния модели, \mathbf{q} — функция нео-

пределенности обратной задачи, (7) — это модель измерений с нелинейным оператором измерений $H(\varphi)$, \mathbf{I} — результат измерений, $\delta\mathbf{I}$ — погрешность измерений. В обратной задаче требуется, зная операторы L, H и данные измерений I , определить \mathbf{q} . Будем предполагать, что $\varphi^{(*)}$ определяется из (6) единственным образом.

Для некоторого оператора A определим оператор разделенных разностей такой, что

$$\bar{V}A[x^{(2)}, x^{(1)}](x^{(2)} - x^{(1)}) = A[x^{(2)}] - A[x^{(1)}],$$

где $\bar{V}A[x^{(2)}, x^{(1)}]$ является линейным. Используя определение операторов разделенных разностей $\bar{V}H[\varphi^{(2)}, \varphi^{(1)}]$, можно для произвольной пробной функции h написать

$$\langle \bar{V}H[\varphi^{(2)}, \varphi^{(1)}](\varphi^{(2)} - \varphi^{(1)}), h \rangle_{\mathbf{I}} = \langle I^{(2)} - I^{(1)}, h \rangle_{\mathbf{I}},$$

где $H(\varphi^{(n)}) = I^{(n)}, n=1,2$ и $\langle \dots, \rangle_{\mathbf{I}}$ — некоторое скалярное произведение в пространстве результатов измерений. Отсюда

$$\langle \varphi^{(2)} - \varphi^{(1)}, \bar{V}H[\varphi^{(2)}, \varphi^{(1)}]^* h \rangle_{\varphi} = \langle I^{(2)} - I^{(1)}, h \rangle_{\mathbf{I}}, \quad (8)$$

где $\langle \dots, \rangle_{\varphi}$ — скалярное произведение в пространстве функций состояния модели.

Аналогично, из (6) можно для $L(\varphi^{(n)}) = q^{(n)}, n=1,2$ и произвольной ψ получить

$$\langle \bar{V}L[\varphi^{(2)}, \varphi^{(1)}](\varphi^{(2)} - \varphi^{(1)}), \psi \rangle_q = \langle q^{(2)} - q^{(1)}, \psi \rangle_q,$$

где $//$ — скалярное произведение в пространстве функций неопределенности. Отсюда следует

$$\langle \varphi^{(2)} - \varphi^{(1)}, \bar{V}L[\varphi^{(2)}, \varphi^{(1)}]^* \psi \rangle_{\varphi} = \langle q^{(2)} - q^{(1)}, \psi \rangle_q. \quad (9)$$

Выбирая $\psi = \psi[\varphi^{(2)}, \varphi^{(1)}; h]$ из решения сопряженного уравнения

$$\bar{V}L[\varphi^{(2)}, \varphi^{(1)}]^* \psi = \bar{V}H[\varphi^{(2)}, \varphi^{(1)}]^* h, \quad (10)$$

и объединяя (8) и (9), мы получим соотношение чувствительности между данными измерений и искомыми параметрами (например, источниками выбросов):

$$\langle I^{(2)} - I^{(1)}, h \rangle_{\mathbf{I}} = \langle q^{(2)} - q^{(1)}, \psi[\varphi^{(2)}, \varphi^{(1)}; h] \rangle_q. \quad (11)$$

В случае линейного оператора измерений [Penenko, 2020; Penenko A. et al., 2021], оператор разделенных разностей $\bar{V}H[\varphi^{(2)}, \varphi^{(1)}]$ не зависит от $\varphi^{(n)}, n=1,2$, то есть $\bar{V}H[\varphi^{(2)}, \varphi^{(1)}] = H$. Так как значение функции состояния определяется согласно (11) значением функции неопределенности, то можно записать $\varphi^{(n)} = \varphi[q^{(n)}], n=1,2$ и, следовательно,

$$\psi[q^{(2)}, q^{(1)}; h] := \psi[\varphi[q^{(2)}], \varphi[q^{(1)}]; h].$$

Соотношение чувствительности (11) является основным агрегатом для формирования оператора чувствительности по общей схеме разрабатываемого подхода. Для этого зададим некоторый набор функций проектирования данных измерений $U = \{h^{(\xi)}\}_{\xi \in \Xi}$, и при опреде-

ленных $\mathbf{q}^{(n)}, n=1,2$ получим соответствующий ансамбль решений сопряженных уравнений $\{\psi[q^{(2)}, q^{(1)}; h^{(\xi)}]\}_{\xi \in \Xi}$. С их помощью можно определить операторы “извлечения структуры из данных” (аналогично [Le Dimet et al., 2015]) и “чувствительности обратной задачи”

$$P_U I = \sum_{\xi \in \Xi} \langle I, h^{(\xi)} \rangle_{\mathbf{I}} e^{(\xi)},$$

$$M_U[q^{(2)}, q^{(1)}]z = \sum_{\xi \in \Xi} \langle z, \psi[q^{(2)}, q^{(1)}; h^{(\xi)}] \rangle_q e^{(\xi)},$$

где $e^{(\xi)}$ — элементы канонического базиса \mathbb{R}^{Ξ} . Для них верно соотношение

$$M_U[q^{(2)}, q^{(1)}](q^{(2)} - q^{(1)}) = P_U I^{(2)} - P_U I^{(1)}.$$

Использование системы проектирования позволяет, во-первых, контролировать размерность операторов чувствительности за счет агрегирования и, во-вторых, организовать априорную фильтрацию шума в данных. Элементы ансамбля решений сопряженных уравнений можно вычислять параллельно, что способствует эффективной реализации алгоритма усвоения данных на современных вычислительных

системах. В качестве нелинейных операторов измерений также могут использоваться алгоритмы сжатия данных и алгоритмы подавления шума на основе нейросетей типа энкодер-декодер.

Пусть выбрано некоторое вспомогательное значение функции неопределенности q , тогда для решения $q^{(*)}$ обратной задачи (6), (7) верно квазилинейное соотношение

$$M_U [q^{(*)}, q] (q^{(*)} - q) = P_U I - P_U H(\varphi[q]) + P_U \delta I, \quad (12)$$

которое и используется для поиска $q^{(*)}$ и исследования свойств обратной задачи. Для решения уравнения (12) на основе разложения на линейную и нелинейную части

$$M_U [q, q] (q^{(*)} - q) = P_U I - P_U H(\varphi[q]) + w, \quad (13)$$

$$w = P_U \delta I - (M_U [q^{(*)}, q] - M_U [q, q]) (q^{(*)} - q) \quad (14)$$

используется алгоритм типа Ньютона-Канторовича из [Penenko, 2020]. Алгоритм допускает эффективное распараллеливание, в том числе и на системах с распределенной памятью [Penenko and Rusin, 2022], что позволяет применять его для решения трехмерных задач [Penenko and Rusin, 2023; Penenko et al., 2024b]. Начальное приближение и результат работы алгоритма обозначим через $q^{(0)}$ и $q^{(\infty)}$.

СЦЕНАРИЙ ОБРАТНОГО МОДЕЛИРОВАНИЯ

Для численных экспериментов рассмотрим сценарий усвоения данных для Байкальского региона аналогичный представленному в работе [Penenko et al., 2024b] с тем отличием, что здесь выбран другой вид искомого источника и рассматривается нелинейный оператор измерений.

Предполагаем, что измеряются концентрации озона O_3 ($L_{meas} = \{O_3\}$); измерения заданы в виде мгновенных двумерных “снимков” полей концентрации по всей области. Как и в [Penenko, Rusin 2023], в качестве оператора измерений рассмотрим оператор, аналогичный интегральной концентрации (“total-column”) и соответствующий процессам ослабления излучения в атмосфере:

$$H^{(m)}[\varphi](x, y) = I_a \exp\left(-\int_0^Z C_a \varphi_{l^{(m)}}(x, y, z, t^{(m)}) dz\right) \quad (15)$$

для заданных констант I_a и C_a , моментов времени $t^{(m)} \in [0, T]$ и вещества $l^{(m)}$.

Для обработки данных типа двумерных “снимков” используем следующую систему проектирования:

$$h^{(\xi)} = C\left(X, \theta_x^{(\xi)}, \theta_{Snapshot}^{(x)}, x\right) C\left(Y, \theta_y^{(\xi)}, \theta_{Snapshot}^{(y)}, y\right),$$

где $\xi = 1, \dots, \Xi_{Snapshot}$, $C(X, \theta, \Theta, x)$ — элементы системы ступенчатых функций на интервале $[0, X]$:

$$C(X, \theta, \Theta, x) = \begin{cases} \frac{\Theta}{X}, x \in \left[\frac{X\theta}{\Theta}, \frac{X(\theta+1)}{\Theta}\right] \\ 0, x \notin \left[\frac{X\theta}{\Theta}, \frac{X(\theta+1)}{\Theta}\right] \end{cases}$$

Пусть количество снимков равно $N_{Snapshot}$. Система проектирования данных имеет два параметра: $\theta_{Snapshot}^{(x)}$ и $\theta_{Snapshot}^{(y)}$, которые определяют пространственное разрешение рассматриваемых данных. Для любого изображения θ_x и θ_y находятся, соответственно, в диапазонах $0, \dots, \theta_{Snapshot}^{(x)} - 1$ и $0, \dots, \theta_{Snapshot}^{(y)} - 1$. Следовательно, $\Xi_{Snapshot} = N_{Snapshot} \times \theta_{Snapshot}^{(x)} \times \theta_{Snapshot}^{(y)}$.

В алгоритме типа Ньютона-Канторовича решения (13) используются производная $M_U [q, q]$. Следовательно, для вычисления правой части (10) нам потребуются действия производных оператора (15) на функции проектирования $\nabla H^{(m)}[\varphi]^* h^{(\xi)}$:

$$\nabla H^{(m)}[\varphi]^* h^{(\xi)} = -C_a H^{(m)}[\varphi](x, y) C\left(X, \theta_x^{(\xi)}, x\right) C\left(Y, \theta_y^{(\xi)}, y\right) \delta\left(t - t^{(m)}\right) \delta\left(1 - I^{(m)}\right),$$

где δ — соответствующие дельта-функции по времени и химическим веществам. Заметим, что обе части (15) не зависят от вертикальной координаты. Левая часть сопряженной задачи (10), определяемая моделью процессов, представлена в работах [Penenko et al., 2020; Penenko and Rusin, 2022].

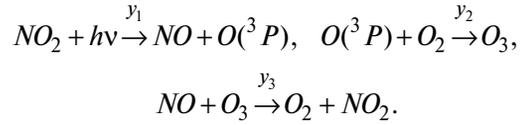
Измерения проводятся через каждые 4 ч интервала моделирования (исключая начальную и конечную точки), то есть количество снимков $N_{Snapshot} = 5$. Рассмотрены три системы проектирования с параметрами пространственного разрешения

$$\theta_{Snapshot}^{(x)} \times \theta_{Snapshot}^{(y)} \in \{4 \times 4, 16 \times 16, 22 \times 22\},$$

и, следовательно, с различным количеством функций проектирования.

Для тестирования алгоритма рассматривалась задача усвоения данных для области, соответствующей Байкальскому региону и изображенной на рис. 1а (как и в [Penenko et al., 2024b]). Временной интервал соответствует интервалу 24.07.2019 00:00 — 24.07.2019 24:00. Скорости ветра и коэффициенты диффузии вычислены с помощью региональной модели WRF 4.2.1 [Skamarock et al., 2019]. В модели (1)–(4) рассматривается взаимодействие между $N_c = 5$ хими-

ческими соединениями $\{O_3, NO, NO_2, O_2, (^3P)\}$ согласно модели химии атмосферы из [Hundsdoerfer and Verwer, 2013]:



Первый коэффициент скорости реакции y_1 зависит от времени суток, а y_2 и y_3 являются постоянными:

$$y_1(t) = \begin{cases} 10^{-5} e^{\sec(t)}, & 4 \times 3600 < t < 20 \times 3600, \\ 10^{-40}, & \text{иначе} \end{cases}$$

$$\sec(t) = \left(\sin \left(\frac{\pi}{16} (\bar{t}_h - 4) \right) \right)^{0.2},$$

$$\bar{t}_h = t_h - 24 \text{int} \left(\frac{t_h}{24} \right), \quad t_h = \frac{t}{3600},$$

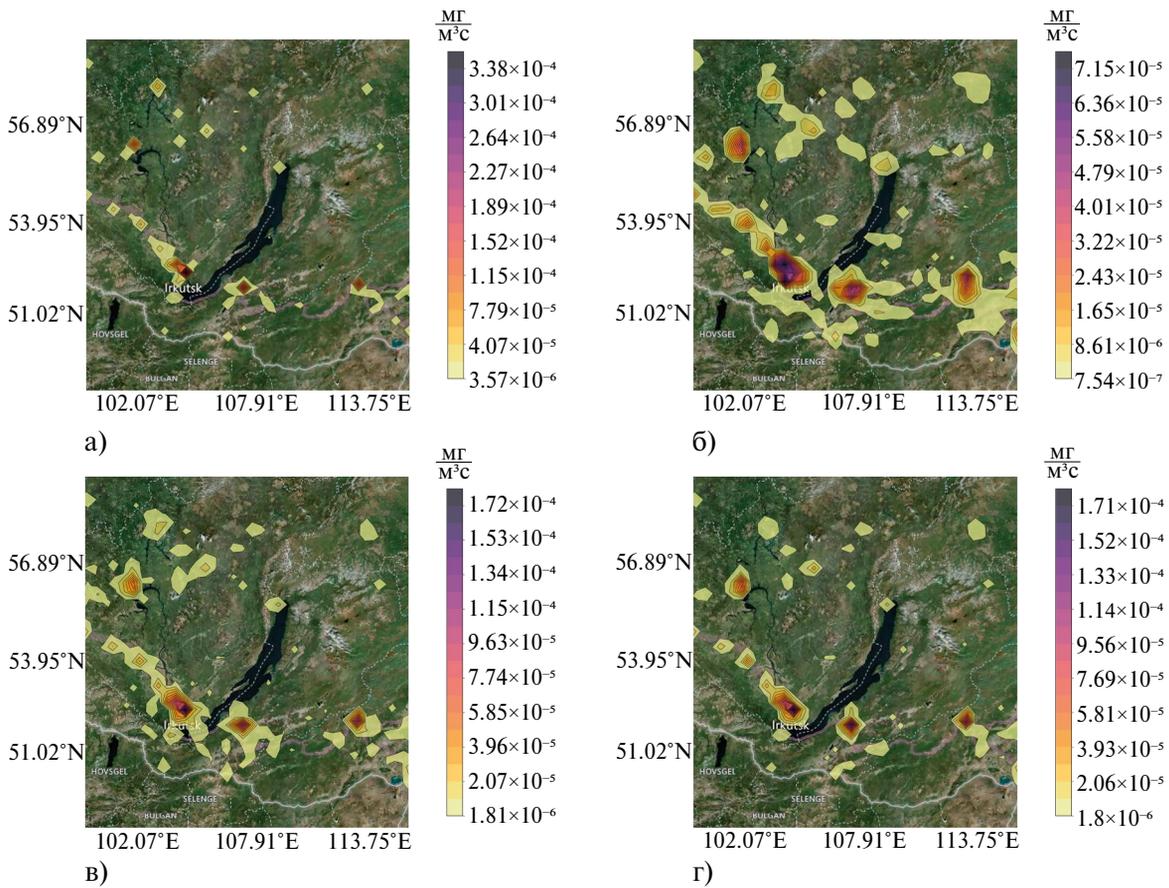


Рис. 1. “Истинная” функция источников $r_{NO}^{(*)}$ — (а); результаты восстановления $r_{NO}^{(\infty)}$ для параметров системы проектирования 22x22 в различных окнах усвоения: (б) — [0.8x3600]; (в) — [8x3600, 16x3600]; (г) — [16x3600, 24x3600]

$$y_2 = 1.87 \cdot 10^{-14}, \quad y_3 = 10^{-16},$$

и $\text{int}\left(\frac{t_h}{24}\right)$ — это целая часть $\frac{t_h}{24}$. Концентрация O_2 считается постоянной.

В сценариях мы полагаем, что “истинная” функция распределения источников $q^{(*)}$ состоит из стационарных источников, распределенных по горизонтали согласно рис. 1а), и расположенных по вертикали только на первом слое. Источники имеют различные неотрицательные мощности и выбрасывают только NO (т.е. $L_{src} = \{NO\}$). Следовательно, задача поиска $q^{(*)} = \{r_l^{(*)}(x, t)\}_{l \in \{O_3, NO, NO_2, O_2, O^3P\}}$ сводится к поиску функции $r_{NO}^{(*)}$.

Соответствующий источнику NO результат работы алгоритма решения обратной задачи или усвоения данных будем обозначать через $r_{NO}^{(\infty)}$. Для простоты начальные концентрации всех веществ, кроме O_2 и O_3 заданы нулевыми: $\varphi_{O_3}^0 = 0.05 \frac{M\mathcal{C}}{M^3}$, $\varphi_{O_2}^0 = 284202 \frac{M\mathcal{C}}{M^3}$, $\varphi_{NO}^0 = \varphi_{NO_2}^0 = \varphi_{O_3P}^0 = 0$. На границе $\Gamma^{(in)}$, где скорости ветра направлены внутрь области, параметр краевого условия соответствует фоновым концентрациям $\alpha_l = \varphi_l^0$.

Начальное предположение о распределении источников в обратной задаче и в первом окне усвоения нулевое ($q^{(0)} = 0$). Соответствующее решение прямой задачи будем называть “фоно-

вым” $\varphi^{(0)} = \varphi[q^{(0)}]$. Рассматривается два размера окон усвоения: 4 ч и 8 ч.

Количество точек сетки по времени, пространству и химическим веществам $N_t \times N_x \times N_y \times N_z \times N_c = 711 \times 43 \times 44 \times 11 \times 5$. Общее количество доступных данных измерений в обратной задаче $N_x \times N_y \times N_{Snapshot} = 43 \times 44 \times 5$.

РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

На рис. 1 представлена “истинная” функция источников $r_{NO}^{(\infty)}$ и результаты ее восстановления в различных окнах усвоения для параметров системы проектирования 22x22.

На рис. 2 представлен результат восстановления функции источников $r_{NO}^{(\infty)}$ в финальном окне усвоения для различных параметров системы проектирования: 4x4 и 16x16. На рис. 3 представлены абсолютные ошибки восстановления источников $r_{NO}^{(\infty)}$ и функций состояния $\varphi_{NO}^{(\infty)}$ в зависимости от модельного времени алгоритмом решения обратной задачи (IP) и алгоритмом усвоения данных с окнами усвоения 8 ч (8h) и 4 ч (4h). Для обратных задач стационарный источник разыскивается на всем интервале, поэтому ошибка со временем не меняется. Для задач усвоения ошибки в окнах усвоения отличаются. На рис. 4 представлены относительные ошибки восстановления на всем интервале тех же характеристик, что и на рис. 3. На рис. 5 представлено время решения задач обратного моде-

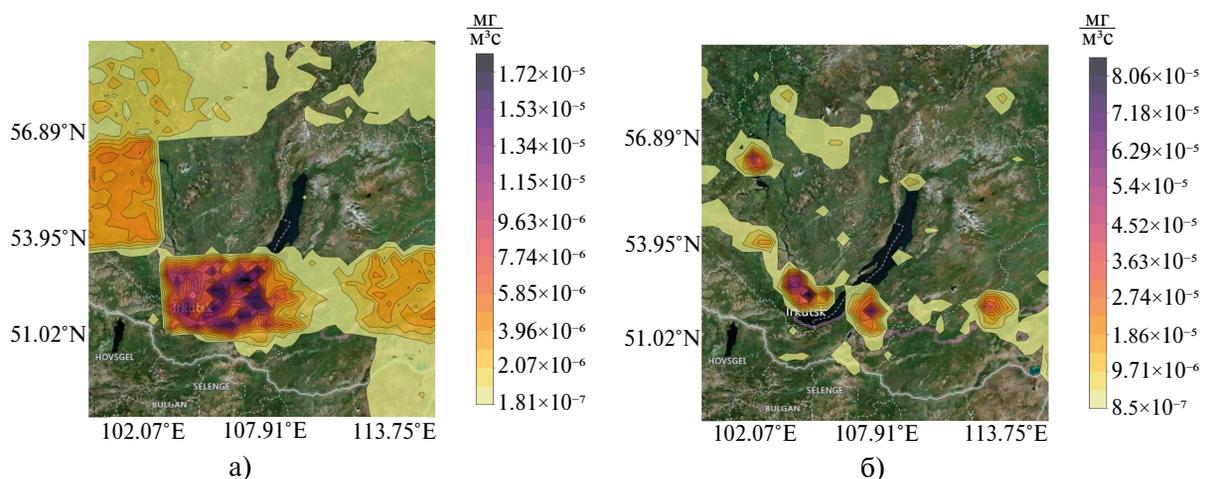


Рис. 2. Результаты восстановления в финальном окне усвоения [16x3600,24x3600] для параметров системы проектирования 4x4 — (а); 16x16 — (б)

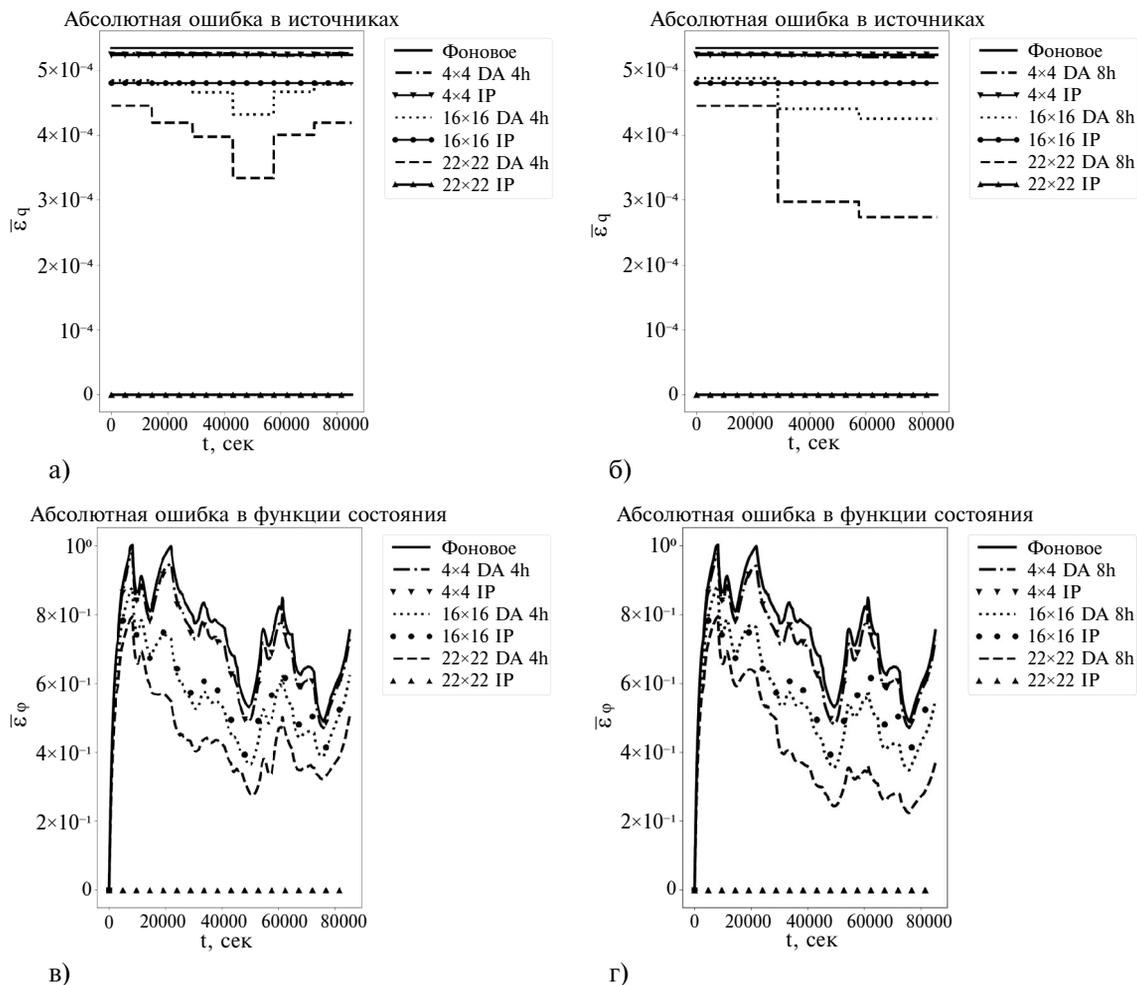


Рис. 3. Абсолютная ошибка в зависимости от времени: в источниках $r_{NO}^{(\infty)}$ — (а), (б); в функциях состояния $\varphi_{NO}^{(\infty)}$ — (в), (г); для обратных задач (IP); для задач усвоения данных (DA) с окнами усвоения 8 ч (8h) и 4 ч (4h). Кривая “фоновое” соответствует решению прямой задачи $\varphi^{(0)}$ без учета данных измерений

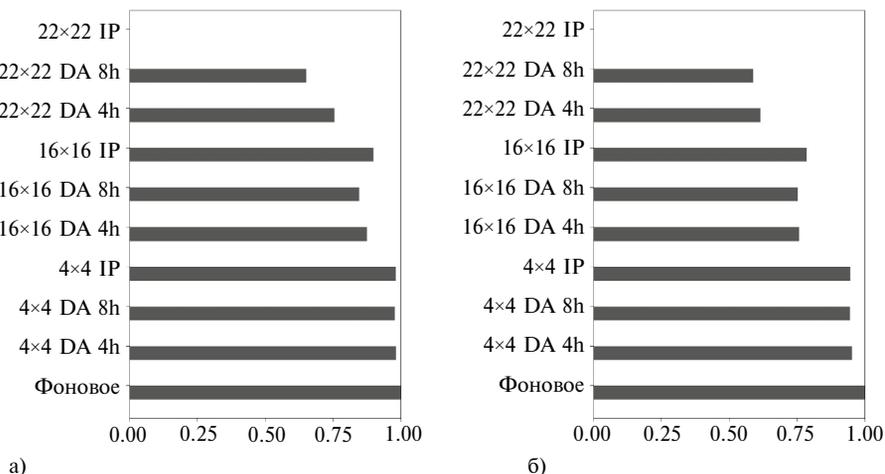


Рис. 4. Относительная ошибка восстановления для обратных задач (IP) и задач усвоения данных (DA) с окнами усвоения 8 ч (8h) и 4 ч (4h): (а) — источника $r_{NO}^{(\infty)}$; (б) — функции состояния $\varphi_{NO}^{(\infty)}$

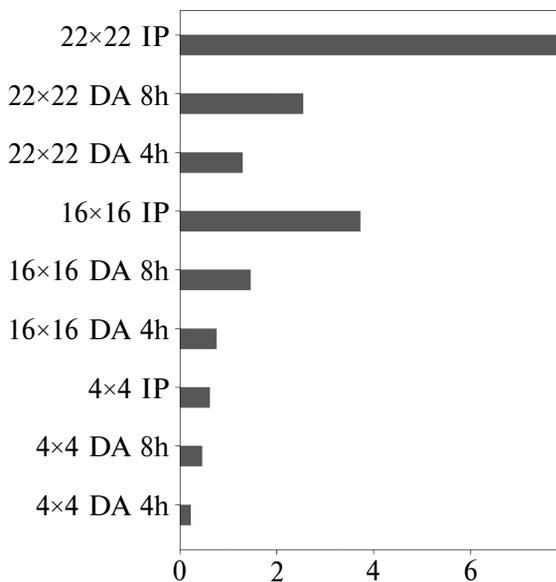


Рис. 5. Относительное время вычислений, измеренное во временах решения прямой задачи

лирования, измеренное во временах решения прямой задачи. Среднее время решения прямой задачи составляет примерно 7 с. Время вычислений измерено в Сибирском суперкомпьютерном центре на гибридном кластере NKS-1P компании РСК, Москва, Россия, с использованием трех узлов Intel Xeon Gold 6248R (каждый имеет 2 процессора \times 24 ядра \times 2 потока, 3.00 ГГц, 384 ГБ оперативной памяти). Общее количество используемых ядер составляет 144. Узлы соединены с помощью Cluster Interconnect Omni-Path со скоростью 100 Гбит/с.

Анализируя рис. 4, мы можем отметить, что обратная задача решается практически точно с максимальным числом рассмотренных параметров системы проектирования 22×22 (аналогичное поведение наблюдалось и для линейного оператора [Penenko et al., 2024b]). В этом случае количество искомым $N_x \times N_y \times 1 \times |L_{src}| = 1892$, меньше, чем количество использованных данных $\Xi_{Snapshot} = \theta_{Snapshot}^{(x)} \times \theta_{Snapshot}^{(y)} \times N_{Snapshot} = 2420$. Для параметра 16×16 значение $\Xi_{Snapshot} = 1280$, и точность восстановления значительно падает. Точность алгоритма усвоения при решении той же задачи (22×22) ниже, однако на ее решение, согласно рис. 5, требуется меньше времени вычислений. Аналогичная картина наблюдается при сравнении результатов для окон усвоения 8 ч и 4 ч: точность восстановления для меньшего

окна усвоения несколько ниже, чем для 8-часового окна, при этом время вычислений значительно ниже. Для примера 16×16 алгоритмы усвоения данных оказались точнее алгоритма решения обратной задачи. Предположительно, это связано с локальными стационарными точками алгоритма решения обратной задачи. Для примера 4×4 все алгоритмы показывают похожие результаты. Сравнивая рис. 1 и 2, можно заключить, что восстановление для примера 4×4 качественно отличается от “истинного” решения, и в нем отчетливо проявляется система проектирования. Восстановление в ходе усвоения данных для 22×22 и 16×16 больше соответствует “истинному” по локализации источников, однако отличается по мощности, что и приводит к относительно высоким количественным значениям ошибок восстановления. Точность восстановления функции состояния в относительных величинах несколько выше, чем при восстановлении источника (рис. 4). Для окна усвоения 4 ч, в отличие от окна усвоения 8 ч, на рис. 3 можно заметить, что ошибка восстановления источника сначала убывает, а затем возрастает. Такое отличие предположительно вызвано соотношением между длиной окна усвоения и интервалом между измерениями.

Таким образом, предлагаемая версия алгоритма эффективно работает в целом, но сильно зависит от разрешения системы проектирования. Последнее, однако, позволяет построить соответствующую стратегию при решении практических задач, выбирая между точностью решения и временем получения результата.

Что касается расширения области приложений разработанного алгоритма, то можно предположить, что в качестве нелинейного оператора измерений может также использоваться оператор удаления шума из данных, например, реализованный в виде предобученной нейросети, натренированной на удаление определенного вида шума. Так как стандартные средства позволяют дифференцировать нейросетевые операторы, то их сравнительно легко включить в алгоритм, что и планируется сделать.

Новая тематика в природоохранном прогнозировании и проектировании связана с концепцией цифрового двойника природного процесса. Она предполагает разработку системы, включающей взаимосвязанные модели из-

учаемого процесса и двухсторонние информационные связи между моделями и прототипом. Процессы переноса и трансформации примесей в атмосфере могут, в частности, рассматриваться как прототип для цифрового двойника. Так как состояние этого объекта динамично меняется и не является в полной мере наблюдаемым, то актуальными становятся задачи уточнения параметров моделей по имеющимся разнородным данным мониторинга, в том числе и в режиме усвоения данных (т.е. в процессе получения данных).

ЗАКЛЮЧЕНИЕ

Подход с использованием ансамблей решений сопряженных уравнений на первый взгляд представляется альтернативой вариационному, однако они дополняют друг друга, расширяя возможности и увеличивая гибкость методики в целом. В основе обоих подходов лежит соотношение чувствительности. Оба подхода используют формулировки со слабыми ограничениями, когда в постановках задач в том или ином виде могут учитываться “любые неопределенности”, начиная от недостатка знаний о “точном” описании изучаемых процессов и заканчивая возможными ошибками компьютерной реализации. Все вместе обеспечивает согласование всех объектов технологии прямого и обратного моделирования.

В работе представлены алгоритмы усвоения данных для трехмерной модели адвекции-диффузии-реакции примесей в атмосфере с нелинейными операторами измерений на основе операторов чувствительности и ансамблей решений сопряженных уравнений. Алгоритмы реализованы и протестированы в рамках разрабатываемой системы обратного моделирования Inverse Modeling Data Assimilation Framework (IMDAF) на задачах оценки качества воздуха для Байкальского региона. Таким образом, можно заключить, что реализованный алгоритм расширяет область применимости алгоритмов на основе операторов чувствительности.

ФИНАНСИРОВАНИЕ

Разработка алгоритма усвоения данных для трехмерных моделей переноса и трансформации примесей выполнена в рамках государственного задания Министерства науки и выс-

шего образования РФ для ИВМиМГ СО РАН FWNM-2022-0003, учет нелинейных измерений выполнен в рамках государственного задания Министерства науки и высшего образования РФ (тема “Аналитическое и численное исследование обратных задач об определении параметров источников атмосферного или водного загрязнения и (или) параметров среды”, код темы: FENG-2023-0004).

СПИСОК ЛИТЕРАТУРЫ

- Агошков В.И.* Методы оптимального управления и сопряженных уравнений в задачах математической физики. М.: ИВМ РАН, 2003. 255 с.
- Алоян А.Е.* Моделирование динамики и кинетики газовых примесей и аэрозолей в атмосфере. М.: Наука, 2008. 415 с.
- Бажин Н.М., Пененко В.В., Алоян А.Е., Скубневская Г.И.* Математическое моделирование распространения загрязнений в атмосфере с учетом химической трансформации // Успехи химии. 1991. Т. 60. № 3.
- Десятков Б.М. и др.* Определение некоторых характеристик источника аэрозольных примесей путем решения обратной задачи их распространения в атмосфере // Оптика атмосферы и океана. 1999. Т. 12. № 2. С. 136–139.
- Дымников В.П., Залесный В.Б.* Основы вычислительной геофизической гидродинамики. М.: ГЕОС, 2019. 448 с.
- Зуев В.Е., Зуев В.В.* Дистанционное оптическое зондирование атмосферы. СПб.: Гидрометеоиздат, 1992. 232 с.
- Марчук Г.И.* О постановке некоторых обратных задач // Докл. АН СССР. 1964. Т. 156. № 3. С. 503–506.
- Марчук Г.И.* Численное решение задач динамики атмосферы и океана. Л.: Гидрометеоиздат, 1974. 304 с.
- Марчук Г.И., Пененко В.В.* Исследование чувствительности дискретных моделей динамики атмосферы и океана // Изв. АН СССР. Физика атмосферы и океана. 1979. Т. 15. № 11. С. 1123–1131.
- Марчук Г.И.* Математическое моделирование в проблеме окружающей среды. М.: Наука, 1982. 319 с.
- Марчук Г.И., Дымников В.П., Залесный В.Б., Лыков В.Н., Галин В.Я.* Математическое моделирование общей циркуляции атмосферы и океана Л.: Гидрометеоиздат, 1984. 320 с.
- Марчук Г.И., Дымников В.П., Залесный В.Б.* Математические модели в геофизической гидродинамике и численные методы их реализации. Л.: Гидрометеоиздат, 1987. 287 с.

- Марчук Г.И.* Методы расщепления. М.: Наука, 1988. 263 с.
- Марчук Г.И.* Сопряженные уравнения и анализ сложных систем. М.: Наука, 1992. 334 с.
- Марчук Г.И., Агошков В.И., Шутяев В.П.* Сопряженные уравнения и методы возмущений в нелинейных задачах математической физики. М.: Наука, 1993. 223 с.
- Панасенко Е.А., Старченко А.В.* Определение городских районов-загрязнителей атмосферного воздуха по данным наблюдений // Оптика атмосферы и океана. 2009. Т. 22. № 3. С. 279–283.
- Пененко В.В.* Вычислительные аспекты моделирования динамики атмосферных процессов и оценки влияния различных факторов на динамику атмосферы / В кн. Некоторые проблемы вычислительной и прикладной математики. Новосибирск: Наука, 1975. С. 61–75.
- Пененко В.В., Образцов Н.Н.* Вариационный метод согласования полей метеорологических элементов // Метеорология и гидрология. 1976. № 11. С. 1–11.
- Пененко В.В.* Методы численного моделирования атмосферных процессов. Л.: Гидрометеиздат, 1981. 352 с.
- Пененко В.В., Алоян А.Е., Бажин Н.М., Скубневская Г.И.* Численная модель гидрометеорологического режима и загрязнения промышленных районов // Метеорология и гидрология. 1984. № 4. С. 5–15.
- Пененко В.В., Алоян А.Е.* Модели и методы для задач охраны окружающей среды. Новосибирск: Наука, 1985. 224 с.
- Пененко В.В., Рапута В.Ф., Панарин А.В.* Планирование эксперимента в задаче определения положения и мощности источника примеси // Метеорология и гидрология. 1985. № 11. С. 15–22.
- Пененко В.В., Цветова Е.А., Скубневская Г.И., Дульцева Г.Г., Кейко А.В.* Численное моделирование химической кинетики и переноса загрязняющих примесей в атмосфере индустриальных регионов // Химия устойчивого развития. 1997. № 5. С. 535–539.
- Пененко В.В.* Вариационные методы усвоения данных и обратные задачи для изучения атмосферы, океана и окружающей среды // Сибирский журнал вычислительной математики. 2009. Т. 12. № 4. С. 421–434.
- Пененко А.В., Пененко В.В.* Прямой метод вариационного усвоения данных для моделей конвекции-диффузии на основе схемы расщепления // Вычислительные технологии. 2014. Т. 19. № 4. С. 69–83.
- Пененко В.В., Цветова Е.А., Пененко А.В.* Развитие вариационного подхода для прямых и обратных задач гидротермодинамики и химии атмосферы // Изв. РАН. Физика атмосферы и океана. 2015. Т. 51. № 3. С. 358–367.
- Пененко В.В., Цветова Е.А., Пененко А.В.* Методы совместного использования моделей и данных наблюдений в рамках вариационного подхода для прогнозирования погоды и качества состава атмосферы // Метеорология и гидрология. 2015. № 6. С. 13–24.
- Пененко А.В., Пененко В.В., Цветова Е.А.* Последовательные алгоритмы усвоения данных в моделях мониторинга качества атмосферы на базе вариационного принципа со слабыми ограничениями // Сибирский журнал вычислительной математики. 2016. Т. 19. № 4. С. 401–418.
- Пененко А.В., Мукатова Ж.С., Пененко В.В., Гочаков А.В., Антохин П.Н.* Численное исследование прямого вариационного алгоритма усвоения данных в городских условиях // Оптика атмосферы и океана. 2018. Т. 31. № 6. С. 456–462.
- Пененко В.В., Пененко А.В., Цветова Е.А., Гочаков А.В.* Методы исследования чувствительности модели качества атмосферы и обратные задачи геофизической гидротермодинамики // Прикладная механика и техническая физика. 2019. Т. 60. № 2. С. 238–246.
- Шутяев В.П.* Операторы управления и итерационные алгоритмы в задачах вариационного усвоения данных. М.: Наука, 2001. 238 с.
- Arridge S. et al.* Solving inverse problems using data-driven models // Acta Numerica. 2019. V. 28. P. 1–174.
- Baklanov A. et al.* Online coupled regional meteorology chemistry models in Europe: current status and prospects // Atmospheric Chemistry and Physics. 2014. V. 14. № 1. P. 317–398.
- Baklanov A. et al.* From urban meteorology, climate and environment research to integrated city services // Urban Climate. 2018. V. 23. P. 330–341.
- Bieringer P. E. et al.* Paradigms and commonalities in atmospheric source term estimation methods // Atmospheric Environment. 2017. V. 156. P. 102–112.
- Bocquet M. et al.* Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models // Atmospheric Chemistry and Physics. 2015. V. 15. № 10. P. 5325–5358.
- Carrasi A. et al.* Data assimilation in the geosciences: An overview of methods, issues, and perspectives // Wiley Interdisciplinary Reviews: Climate Change. 2018. V. 9. 10.1002/wcc.535.
- Cheng S. et al.* Machine learning with data assimilation and uncertainty quantification for dynamical systems: A Review // IEEE/CAA J. of Automatica Sinica. 2023. V. 10. № 6. P. 1361–1387.

- Elbern H. et al.* Emission rate and chemical state estimation by 4-dimensional variational inversion // *Atmospheric Chemistry and Physics Discussions*. 2007. V. 7. № 1. P. 1725–1783.
- Elbern H. et al.* Data assimilation in atmospheric chemistry and air quality/ In: *Advanced Data Assimilation for Geosciences*. Oxford University Press, 2014. P. 507–534. 10.1093/acprof:oso/9780198723844.003.0022.
- Hundsdoerfer W., Verwer J.G.* Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations. Berlin Heidelberg: Springer, 2013. 479 p.
- Le Dimet F.X., Talagrand O.* Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects // *Tellus*. 1986. V. 38 A. P. 97–110.
- Le Dimet F.-X., Souopgui I., Titaud O., Shutyaev V., and Hussaini M. Y.* Toward the assimilation of images // *Nonlinear Processes in Geophysics*. 2015. V. 22. № 1. P. 15–32. doi: 10.5194/npg-22-15-2015.
- Issartel J.-P.* Rebuilding sources of linear tracers after atmospheric concentration measurements // *Atmospheric Chemistry and Physics*. 2003. V. 3. P. 2111–2125.
- Issartel J.-P.* Emergence of a tracer source from air concentration measurements, a new strategy for linear assimilation // *Atmospheric Chemistry and Physics*. 2005. V. 5. № 1. P. 249–273.
- Kumar P. et al.* Reconstruction of an atmospheric tracer source in an urban-like environment // *Journal of Geophysical Research: Atmospheres*. 2015. V. 120. № 24. P. 12589–12604.
- Liu X. et al.* Data-Driven Machine Learning in Environmental Pollution: Gains and Problems // *Environmental Science and Technology*. 2022. V. 56. № 4. P. 2124–2133.
- Mahura A. et al.* Towards seamless environmental prediction — development of Pan-Eurasian EXperiment (PEEX) modelling platform// *Big Earth Data*. 2024. V. 8:2. P. 189–230.
DOI: 10.1080/20964471.2024.2325019
- Mamonov A.V., Tsai Y.-H.R.* Point source identification in nonlinear advection-diffusion-reaction systems // *Inverse Problems*. 2013. V. 29. № 3. P. 35009.
- Marchuk G.I., Penenko V.V.* Application of perturbation theory to problem of simulation of atmospheric processes / In: *Monsoon dynamics* (Eds. J. Lighthill & P.R.Pearce). Cambridge: Cambridge University Press, 1981. P. 639–655.
- Mijling B.* High-resolution mapping of urban air quality with heterogeneous observations: a new methodology and its application to Amsterdam // *Atmospheric Measurement Techniques*. 2020. V. 8. P. 4601–4617.
- Murio D.A.* The mollification method and the numerical solution of ill-posed problems / D.A. Murio, John Wiley & Sons, Inc., 1993.
- Nguyen C., Soulhac L., Salizzoni P.* Source Apportionment and Data Assimilation in Urban Air Quality Modelling for NO₂: The Lyon Case Study // *Atmosphere*. 2018. V. 9.
doi:10.3390/atmos9010008.
- Nguyen C.V., Soulhac L.* Data assimilation methods for urban air quality at the local scale // *Atmospheric Environment*. 2021. V. 253. P. 118366.
- Penenko V., Tsvetova E.* Orthogonal decomposition methods for inclusion of climatic data into environmental studies // *Ecological Modelling*. 2008. V. 217. № 3–4. P. 279–291.
- Penenko V., Baklanov A., Tsvetova E., Mahura A.* Direct and inverse problems in a variational concept of environmental modeling // *Pure and Applied Geophysics*. 2012. V. 169. № 3. P. 447–465.
- Penenko V.V., Tsvetova E.A., Penenko A.V.* Variational approach and Euler’s integrating factors for environmental studies // *Computers & Mathematics with Applications*. 2014. V. 67. № 12. P. 2240–2256.
- Penenko A.* Convergence analysis of the adjoint ensemble method in inverse source problems for advection-diffusion-reaction models with image-type measurements// *Inverse Problems & Imaging*, 2020. V. 14. № 5. P. 757–782.
- Penenko A., Penenko V., Tsvetova E., Gochakov A., Pyanova E., and Konopleva V.* Sensitivity operator framework for analyzing heterogeneous air quality monitoring systems // *Atmosphere*. 2021. V. 12. № 12. P. 1697.
- Penenko A., Rusin E.* Parallel implementation of a sensitivity operator-based source identification algorithm for distributed memory computers // *Mathematics*. 2022. V. 10. № 23. P. 4522.
- Penenko A.V., Konopleva V.S., Penenko V.V.* Inverse modeling of atmospheric chemistry with a differential evolution solver: inverse problem and data assimilation// *IOP Conference Series: Earth and Environmental Science*. 2022. V. 10. P. 012015
- Penenko A.V., Rusin E.V.* Numerical study of a sensitivity operator-based emission sources identification algorithm with nonlinear measurement operator // *Proc. SPIE, 29th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*. 2023.
doi: 10.1117/12.2690857.
- Penenko A., Emelyanov M., Rusin E., Tsybenova E., Shablyko V.* Hybrid Deep Learning and Sensitivity Operator-Based Algorithm for Identification of Localized Emission Sources // *Mathematics*. 2024. V. 12 № 1. P. 78.
- Penenko A.V., Gochakov A.V., Antokhin P.N.* Data assimilation algorithm based on the sensitivity operator for a three-dimensional model of transport and transfor-

- mation of atmospheric contaminants // *Atmos. Ocean Opt.* 2024. V. 37. P. 822–832.
- Pyanova E.A., Penenko V.V., Faleychik, L.M.* Simulation of atmospheric dynamics and air quality in the Baikal region // *Proc. SPIE, 20th International Symposium on Atmospheric and Ocean Optics. Atmospheric Physics.* 2014. V. 9292.
doi:10.1117/12.2074998.
- Pudykiewicz J.A.* Application of adjoint tracer transport equations for evaluating source parameters // *Atmospheric Environment.* 1998. V. 32. № 17. P. 3039–3050.
- Sasaki Y.K.* An objective analysis based on the variational method // *J. Meteor. Soc. Japan.* 1958. V. 36. P. 77–88.
- Sasaki Y.* Some basic formalisms in numerical variational analysis // *Mon. Wea Rev.* 1970. V. 98. P. 875–883.
- Skamarock W.C. et al.* A Description of the Advanced Research WRF Version 4. 2019.
- Silver J.D. et al.* Multi-species chemical data assimilation with the Danish Eulerian hemispheric model: system description and verification // *Journal of Atmospheric Chemistry.* 2015. V. 73. № 3. P. 261–302.
- Sokhi R.S. et al.* Advances in air quality research — current and emerging challenges // *Atmospheric Chemistry and Physics.* 2022. V. 22. № 7. P. 4615–4703.
- Turbelin G., Singh S.K., Issartel J.-P.* Reconstructing source terms from atmospheric concentration measurements: Optimality analysis of an inversion technique // *Journal of Advances in Modeling Earth Systems.* 2014. V. 6. № 4. P. 1244–1255.
- Xing J. et al.* Data Assimilation of Ambient Concentrations of Multiple Air Pollutants Using an Emission-Concentration Response Modeling Framework // *Atmosphere.* 2020. V. 11. № 12. P. 1289.
- WMO. Seamless prediction of the Earth system: from minutes to months. Geneva, Switzerland: World Meteorological Organization, 2015. WMO-No 1156. 471 p.
- WMO. Guide to Instruments and Methods of Observation. Geneva, Switzerland: WMO. 2018. P. 506–541.

DATA ASSIMILATION ALGORITHMS FOR ATMOSPHERIC CHEMISTRY MODELS

© 2025 A. V. Penenko^{1,2,*}, V. V. Penenko¹, E. A. Tsvetova¹

¹*Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
Ac. Lavrentiev av., 6, Novosibirsk, 630090 Russia*

²*Yugra State University, Chekhov str., 16, Khanty-Mansiysk, 628012 Russia*

*e-mail: aleks@ommgp.sccc.ru

The problem of data assimilation for nonstationary models of impurity transfer and transformation is considered as a sequence of related inverse problems of restoring the spatiotemporal structure of state functions, taking into account the measurement data received during modeling. Data assimilation is carried out together with the identification of an additional desired source function, which we call the uncertainty function of the model. The purpose of the work is a brief historical overview and presentation of an up-to-date version of the data assimilation algorithms for atmospheric chemistry models based on sensitivity operators and ensembles of solutions to adjoint equations. A demonstration of the algorithm for a three-dimensional model with a nonlinear measurement operator is given in a modeling scenario for the Baikal region.

Keywords: data assimilation, transport and transformation models, adjoint equations, sensitivity operator, inverse modeling, measurement data