

**SDM-2025**

*Федеральный Исследовательский центр  
Информационных и Вычислительных  
Технологий*



**Современные технологии обработки  
пространственных данных  
в задачах мониторинга эмиссии парниковых газов**

---

*Климова Екатерина Георгиевна*

# I. Постановка задачи. Краткий обзор.

## Введение

---

- Одной из важных задач мониторинга природных явления является задача получения значений параметров окружающей среды на некоторой регулярной сетке.
- В настоящее время принято решать такие задачи с привлечением всех имеющихся данных наблюдений а также математической модели интересующего нас процесса.
- Математическая постановка задачи входит в набор задач так называемого обратного моделирования.
- Решение задачи обратного моделирования для заданной модели процесса и набора данных наблюдений циклически по времени является задачей усвоения данных.
- В проблему обратного моделирования включена также задача оценки параметров модели.

# Введение (продолжение).

## Задача обратного моделирования

---

Задачей **обратного моделирования** называют оценку неизвестного вектора параметров большой размерности  $\mathbf{x} \in R^M$  по вектору данных наблюдений  $\mathbf{y} \in R^P$  (\*). Предположим, что

$$\mathbf{y} = M(\mathbf{x}) + \delta ,$$

где  $M$  - известный оператор прогноза-наблюдений,  $\delta$  - случайный шум с заданной функцией распределения.

Под задачей **усвоения данных** принято понимать последовательное по времени оценивание неизвестной величины по данным наблюдений (\*).

(\*). Nakamura G., Potthast R. *Inverse Modeling*. IOP Publishing. 2015. 509 pp. [doi.org/10.1088/978-0-7503-1218-9](https://doi.org/10.1088/978-0-7503-1218-9).

# Введение (продолжение)

---

Современные подходы к задаче усвоения данных включают в себя

- первичную обработку данных наблюдений, контроль и т.д.;
- математическую постановку задачи;
- разработку системы усвоения, практическая реализация задачи.

## Введение (продолжение)

---

- Для решения задачи усвоения данных используются математические постановки задачи, при этом в мире в настоящее время применяется два подхода – вариационный (3DVAR-4DVAR) и динамико-стохастический (ансамблевый фильтр Калмана; ансамблевое сглаживание).

# Современные методы усвоения данных, основанные на ансамблевом фильтре Калмана

## *Ensemble Kalman filter - EnKF*

*(Evensen, 1994; Mitchell, Houtekamer, 1998)*

**Ансамбль  
начальных  
полей и  
прогнозов:**

$$x_f^{0(i)} = \hat{x} + \Delta x_0^{(i)}, \quad i = 1, \dots, N$$

$$x_{k+1}^{f(i)} = \Phi_k(x_k^{a(i)}) + \xi_k^{(i)},$$

**Оценка  
ковариацион  
ных  
матриц:**

$$P_k^f H_k^T = \frac{1}{N-1} \sum_{i=1}^N (x_k^{f(i)} - \overline{x_k^{f(i)}})(H_k x_k^{f(i)} - \overline{H_k x_k^{f(i)}})^T,$$

$$H_k P_k^f H_k^T = \frac{1}{N-1} \sum_{i=1}^N (H_k x_k^{f(i)} - \overline{H_k x_k^{f(i)}})(H_k x_k^{f(i)} - \overline{H_k x_k^{f(i)}})^T,$$

**Ансамбль  
«анализов»:**

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1},$$

$$x_k^{a(i)} = x_k^{f(i)} + K_k (y_k^{0(i)} - H_k x_k^{f(i)}),$$

$$y_k^{0(i)} = y_k^0 + r_k^{(i)}$$

$$\overline{x_{k+1}} = \frac{1}{N} \sum_{i=1}^N x_{k+1}^{(i)}$$

# Вариационная постановки задачи усвоения данных: алгоритмы 3DVAR/4DVAR

---

В алгоритме анализа данных, называемом в литературе 3DVAR (3-Dimensional VARiational) ищется значение вектора  $x_a$ , доставляющее минимум функционалу

$$2J(x) = (x - x_b)^T B^{-1}(x - x_b) + [y_0 - H(x)]^T R^{-1}[y_0 - H(x)].$$

В этой формуле  $B$  и  $R$  - ковариационные матрицы ошибок прогноза  $x_b$  и наблюдений  $y_0$ , соответственно. Оператор  $H$ , вообще говоря, нелинейный, переводит прогностические значения в точки наблюдений (и наблюдаемые переменные).

Минимум функционала есть решение уравнения

$$\nabla J(x_a) = 0.$$

Если градиент функционала записывается в следующем виде:

$$\nabla J(x) = B^{-1}(x - x_b) + H^T R^{-1} H(x - x_b) - H^T R^{-1} \{y_0 - H(x_b)\},$$

тогда решение задачи имеет вид

$$x_a = x_b + (B^{-1} + H^T R^{-1} H) \{y_0 - H(x_b)\}.$$

# Вариационная постановки задачи усвоения данных: алгоритмы 3DVAR/4DVAR

---

Алгоритм 4DVAR, оперативно работающий в системах усвоения ведущих мировых прогностических центров, представляет собой обобщение 3DVAR на пространственно-временной случай. 4DVAR позволяет использовать наблюдения из некоторого временного интервала  $(t_n - t_0)$ . Считается, что на этом интервале заданы наблюдения  $\{y_i^0, i = 0, \dots, N\}$ . Ищется минимум функционала

$$J[x(t_0)] = \frac{1}{2} [x(t_0) - x^b(t_0)]^T B_0^{-1} [x(t_0) - x^b(t_0)] + \frac{1}{2} \sum_{i=0}^N [H(x_i) - y_i^0]^T R^{-1} [H(x_i) - y_i^0]$$

при условии, что

$$x(t_n) = M_n[x(t_0)].$$

Минимум функционала в 4DVAR ищется с помощью итерационных методов (например, метод квази-Ньютона). Для реализации итерационного процесса требуется оценить градиент функционала  $J(x)$ , который, в свою очередь, вычисляется с помощью сопряженной линеаризованной модели

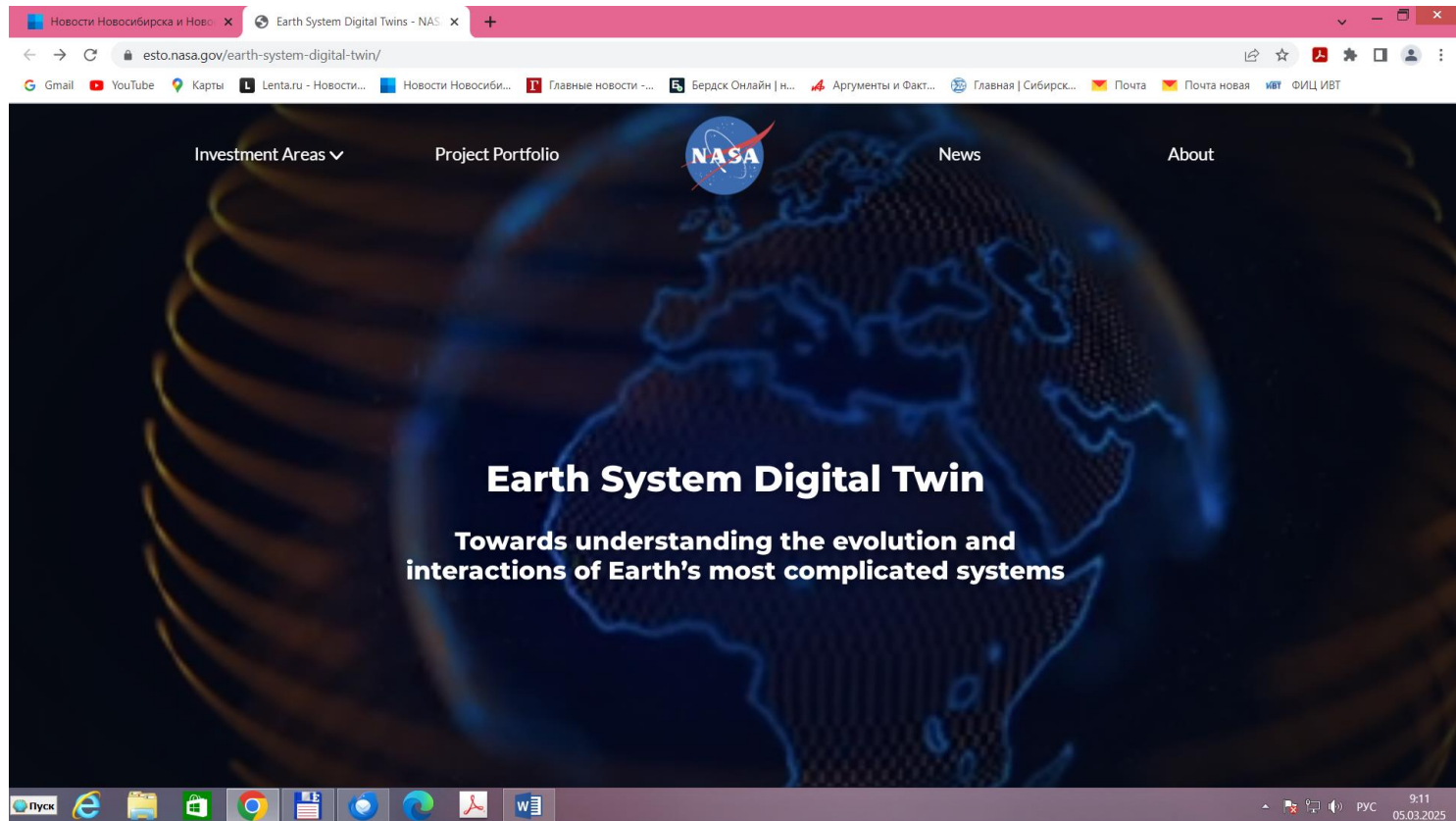
# Примеры применения систем усвоения данных

---

1. Архив данных реанализа (Reanalysis). ERA5 - 5-я версия реанализа ECMWF: 30 км по горизонтали, 137 уровней по вертикали, часовые данные анализа, система усвоения ECMWF, 4DVAR .  
<https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5>
2. Современные оперативные системы усвоения метеорологических данных.
3. Системы оценки потоков парниковых газов (глобальные модели атмосферы, наземные данные, данные спутникового зондирования).
4. Copernicus Atmospheric Monitoring Service (CAMS) – архив данных о химических составляющих атмосферного воздуха.

# Применение систем усвоения данных в современных задачах описания состояния окружающей среды: использование усвоения данных при разработке цифровых двойников окружающей среды:

---



# Применение систем усвоения данных в современных задачах описания состояния окружающей среды: использование усвоения данных при разработке цифровых двойников окружающей среды:

---

**AMS Annual Meeting, 2025:**

## **What Is the Role and Opportunity of Data Assimilation in the Earth System Digital Twin?**

Amos Stephen Lawless. University of Reading & National Centre for Earth Observation, Reading United Kingdom

Sarah L Dance. University of Reading & National Centre for Earth Observation, Reading United Kingdom

### **Abstract**

Many initiatives are currently underway to build **digital twins of the Earth system**, including NASA's Earth System Digital Twins efforts and the European Union's Destination Earth program. ....

Furthermore, where components of the Earth system are coupled in a digital twin (e.g. linking atmospheric models to hydrology), care must be taken to link the different temporal and spatial information from observations appropriately. Similar challenges are currently being addressed in **data assimilation** for standard numerical weather prediction. We will highlight where current research in **data assimilation** could be informative for the design of digital twins and discuss potential avenues for further research.

# Применение усвоения данных в машинном обучении

---

## **Data Learning: Integrating Data Assimilation and Machine Learning**

Caterina Buizza <sup>b</sup>, César Quilodrán

Casas <sup>a</sup>, Philip Nadler <sup>a</sup>, Julian Mack <sup>a</sup>, Stefano Marrone <sup>a f</sup>, Zainab Titus <sup>c</sup>, Clémence Le Cornec <sup>d</sup>, Evelyn Heylen <sup>e</sup>, Tolga Dur <sup>a</sup>, Luis Baca Ruiz <sup>a g</sup>, Claire Heaney <sup>c</sup>, Julio Amador Díaz Lopez <sup>a h</sup>, K.S. Sesh Kumar <sup>a</sup>, Rossella Arcucci <sup>a c</sup>

*Journal of Computational Science, Volume 58, February 2022, 101525*

**Data Assimilation (DA)** is the approximation of the true state of some physical system by combining observations with a dynamic model. DA incorporates observational data into a prediction model to improve forecasted results. These models have increased in sophistication to better fit application requirements and circumvent implementation issues. Nevertheless, these approaches are incapable of fully overcoming their unrealistic assumptions. **Machine Learning (ML)** shows great capability in approximating nonlinear systems and extracting meaningful features from high-dimensional data. ML algorithms are capable of assisting or replacing traditional forecasting methods. However, the data used during training in any Machine Learning (ML) algorithm include numerical, approximation and round off errors, which are trained into the forecasting model. Integration of ML with DA increases the reliability of prediction by including information with a physical meaning. **This work provides an introduction to Data Learning, a field that integrates Data Assimilation and Machine Learning to overcome limitations in applying these fields to real-world data.** The fundamental equations of DA and ML are presented and developed to show how they can be combined into Data Learning. We present a number of Data Learning methods and results for some test cases, though the equations are general and can easily be applied elsewhere.

## **II. Современные подходы к задаче оценки потоков парниковых газов с поверхности Земли по данным измерений**

---

### **Актуальность.**

- Необходимость противодействия климатическим изменениям сохраняет актуальность в мире.
- В России продолжается активное развитие климатической политики. Ключевое значение при этом приобретают данные о выбросах и поглощениях парниковых газов: на них опираются климатические цели и отчетность стран и компаний, они являются основой для реализации климатических проектов, на них строятся углеродные рынки.

# Методы оценки потоков парниковых газов (на примере метана).

---

- В работах используются глобальные и региональные модели переноса и диффузии. Метеорологические данные получены на основе современных моделей прогноза погоды и климата.
- В качестве данных измерений используются данные наземных наблюдательных сетей, а также данные космического зондирования.
- Математические постановки задач вариационного и динамико-стохастического подходов эквивалентны (в линейном гауссовском случае).
- Основное различие алгоритмов состоит в технологии их реализации.

# Проблемы реализации алгоритмов усвоения данных в задаче оценки потоков парниковых газов с поверхности Земли

---

$y=F(x, p)$ ,  $F$  – модель,  $p$  - параметр

$y$ - данные измерений о концентрации,  $x$  – эмиссия

Оценка эмиссии с помощью описания физических процессов:

*“bottom-up constraint on surface fluxes”*

Определение эмиссии по данным наблюдений о концентрации с привлечением химической транспортной модели:

*“top-down constraint on surface fluxes”*

# Методы оценки потоков парниковых газов (на примере метана).

## Алгоритмы, основанные на вариационном подходе (4DVAR, «Inverse modelling»)

---

*Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015*

*Joannes D. Maasackers, Daniel J. Jacob, Melissa P. Sulprizio, Tia R. Scarpelli, Hannah Nesser, Jian-Xiong Sheng, Yuzhong Zhang,*

*Monica Hersher, A. Anthony Bloom, Kevin W. Bowman, John R. Worden, Greet Janssens-Maenhout, and Robert J. Parker*

*Atmos. Chem. Phys., 19, 7859–7881, <https://doi.org/10.5194/acp-19-7859-2019>, 2019.*

Решается задача поиска минимума

$$J(x) = (x - x_a)^T S_a^{-1} (x - x_a) + \gamma (y - F(x))^T S_0^{-1} (y - F(x))$$

Где  $x$  - оцениваемая переменная,

$S_a$  - матрица ковариаций ошибок первого приближения,

$y$  - данные измерений

$F(x)$  - перевод оцениваемой величины в точку наблюдения (с использованием модели)

$S_0$  - матрица ковариаций ошибок наблюдений

# Методы оценки потоков парниковых газов (на примере метана).

## Алгоритмы, основанные на вариационном подходе (4DVAR, «Inverse modelling»)

---

*Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015*

*Joannes D. Maasakkers et al. Atmos. Chem. Phys., 19, 7859–7881, 2019.*

Если представить

$$y = F(x),$$

$$F(x) = Kx + c,$$

$$K = \partial y / \partial x$$

То оптимальное решение имеет вид

$$\hat{x} = x_a + S_a K^T (KS_a K^T + \frac{S_0}{\gamma})^{-1} (y - Kx_a)$$

Матрица  $K$  оценивается с помощью расчетов по возмущенным начальным данным и глобальной транспортной модели.

# Методы оценки потоков парниковых газов (на примере метана). Алгоритмы, основанные на вариационном подходе (4DVAR, «Inverse modelling»)

---

*Wetland emission and atmospheric sink changes explain methane growth in 2020*

*Shushi Peng, Xin Lin, Rona L. Thompson, Yi Xi, Gang Liu, Didier Hauglustaine, Xin Lan, Benjamin Poulter, Michel Ramonet, Marielle Saunois, Yi Yin, Zhen Zhang, Bo Zheng & Philippe Ciais*

*Nature volume 612, pages477–482 (2022)*

Оцениваются средние за неделю значения потоков метана на сетке 1.9 на 3.75

$$J(x) = 0.5(x - x^b)^T B^{-1}(x - x^b) + 0.5(y - H(x))^T R^{-1}(y - H(x))$$

H – оператор наблюдений, переводящий значения первого приближения в точку наблюдений, включая прогноз по модели.

Для минимизации функционала используется итерационный метод.

# Методы оценки потоков парниковых газов (на примере метана).

## Алгоритмы, основанные на ансамблевом фильтре Калмана.

*Estimating surface CO<sub>2</sub> fluxes from space-borne CO<sub>2</sub> dry air mole fraction observations using an ensemble Kalman filter. Feng L, Palmer P I, Bosch H and Dance S 2009 Atmospheric chemistry and physics 9 pp 2619-33*

1. Оценка значений средних по подобластям потоков  $\mathbf{x}^a$  по данным наблюдений  $\mathbf{y}_0$  и прогнозу  $\mathbf{x}_f$  производится по стандартной формуле фильтра Калмана (шаг анализа):

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{K}[\mathbf{y}_0 - H(\mathbf{x}_f)],$$

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

2. Для реализации ансамблевого фильтра Калмана задается ансамбль возмущений оцениваемой величины  $\mathbf{D}\mathbf{x}_f = \frac{1}{\sqrt{N}} [\mathbf{d}\mathbf{x}_f^1, \dots, \mathbf{d}\mathbf{x}_f^N]^T$ . Матрица  $\mathbf{P}^f$  оценивается по ансамблю

$$\mathbf{P}^f = \mathbf{D}\mathbf{x}_f (\mathbf{D}\mathbf{x}_f)^T.$$

# Методы оценки потоков парниковых газов (на примере метана).

## Алгоритмы, основанные на ансамблевом фильтре Калмана.

---

3. Уравнение изменения потоков по времени (шаг прогноза) имеет вид

$$\mathbf{x}_f^{n+1} = \mathbf{x}_f^n .$$

4. Оператор «наблюдений»  $\mathbf{H}$  включает в себя прогноз по модели к моменту наблюдения, интерполяцию из узлов сетки в точки наблюдений, а также, в случае спутниковых данных, осреднение по вертикали с известными коэффициентами.

Данные наблюдений о концентрации парниковых газов в момент времени  $t_n$  представим в виде

$$\mathbf{y}_0^n = H[f(\mathbf{q}_t^n) + \mathbf{x}_t^n] + \boldsymbol{\varepsilon}_0 ,$$

где  $f$  – оператор модели, т.е., модель, описывающая изменение по времени концентрации  $\mathbf{q}$ , входит в оператор наблюдений.

# Методы оценки потоков парниковых газов (на примере метана).

## Алгоритмы, основанные на ансамблевом фильтре Калмана.

---

*Estimation of CH<sub>4</sub> emission based on an advanced 4D-LETKF assimilation system. Bisht Jagat S.H. et al., Geoscientific Model Development, 2023, vol. 16, p. 1823-1838*

В алгоритме LETKF на шаге анализа выполняется поиск оценки только для среднего по ансамблю значения, после чего вычисляется ансамбль значений анализов. Ниже представлены формулы, описывающие этап анализа алгоритма LETKF, для заданного момента времени  $k$ :

$$x_k^a = x_k^f + Dx_k^f \tilde{P}^a (HDx_k^f)^T R_k^{-1} [y_0^k - h(x_k^f)],$$

$$\tilde{P}^a = [(N-1)I + (HDx_k^f)^T R_k^{-1} HDx_k^f]^{-1},$$

где  $I$  – единичная матрица,  $N$  – размерность ансамбля,  $Dx_k^f$  – ансамбль ошибок прогноза, нормированных на  $\sqrt{N-1}$ ,  $H$  – линеаризованный оператор наблюдений

# Глобальный Реанализ парниковых газов CAMS (Copernicus Atmospheric Monitoring Service)

<https://ads.atmosphere.copernicus.eu/>

---

- *Massart, S., A. Agusti-Panareda, I. Aben, A. Butz, F. Chevallier, C. Crevoisier, R. Engelen, C. Frankenberg, and O. Hasekamp, 2014: Assimilation of atmospheric methane products into the MACC-II system: from SCIAMACHY to TANSO and IASI, Atmos. Chem. Phys., 14, 6139-6158, <https://doi.org/10.5194/acp-14-6139-2014>*
- *Bergamaschi, P., and Coauthors, 2009: Inverse modeling of global and regional CH<sub>4</sub> emissions using SCIAMACHY satellite retrievals, J. Geophys. Res., 114, D22301, <https://doi.org/10.1029/2009JD012287>*

## Методика оценки эмиссии:

- TM5-4DVAR Inverse Modeling System.
- Рассматриваются вариант нелинейного алгоритма 4DVAR.
- Оцениваемый вектор включает концентрацию и эмиссию метана.
- Система производит оценку для периода 14 месяцев; данные «усваиваются» блоками: все данные за 3 часовой период.
- В качестве начальных значений эмиссии используются априорные оценки с помощью “bottom-up inventories”.
- Рассматриваются 3 группы источников:
  1. болота и рисовые поля;
  2. “biomass burning”;
  3. все остальные.

Данные о потоках CH<sub>4</sub>: сетка 2° × 3°; 1990-2022; 6 час/день; ежемесячно

# III. Система усвоения данных, основанная на ансамблевом фильтре Калмана для оценки потоков парниковых газов с поверхности Земли

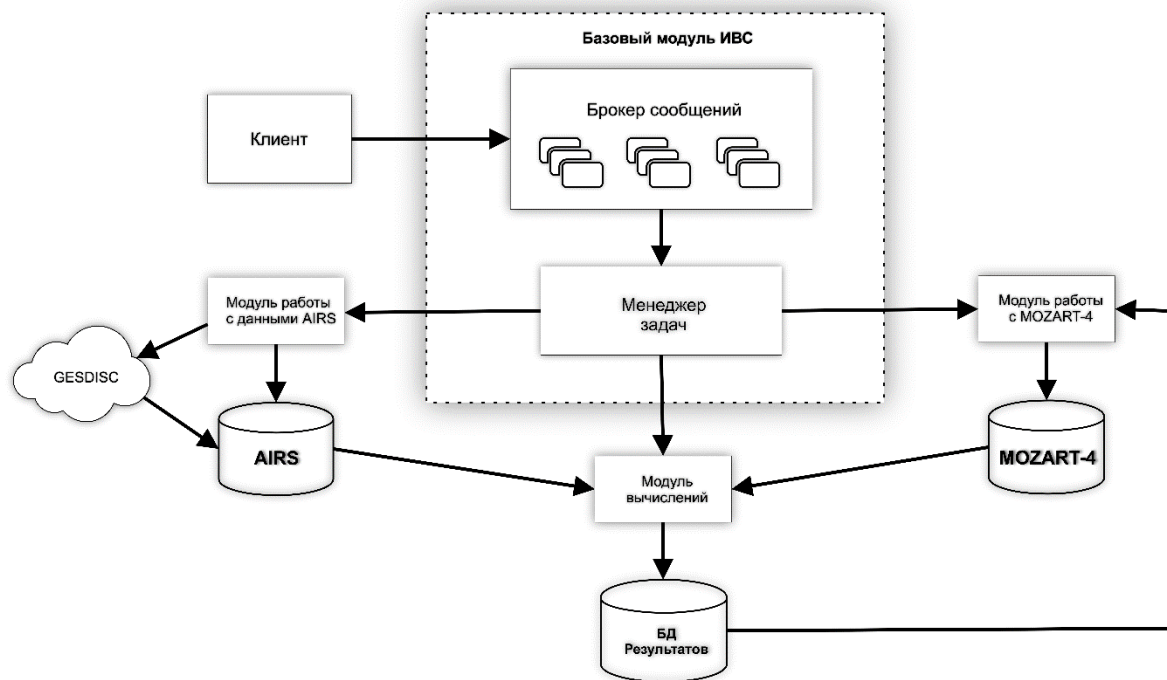
---

В настоящее время в ФИЦ ИВТ проводится разработка системы усвоения спутниковых данных, основанной на ансамблевом подходе, для оценки потоков парниковых газов с поверхности Земли, а также реализация алгоритмов усвоения данных в виде информационно-вычислительной системы.

Особенности разрабатываемой системы:

- Алгоритм EnKF (LETKF).
- В качестве транспортной химической модели используется модель MOZART-4 (сетка по горизонтали 2.8 градуса, 28 сигма-уровней по вертикали).
- В модели информация об эмиссии содержится в данных о приземной концентрации.
- Оценка эмиссии производится для 7-дневного периода для областей 1000 км на 1000 км.

# Информационно-вычислительная система для решения задачи усвоения данных при моделировании окружающей среды (совместно с В.Д. Котлером, М.В. Платоновой).



- Система состоит из нескольких ключевых компонентов.
- В блоке баз данных системы усвоения данных предусмотрено хранение информации, полученной как из спутниковых наблюдений, так и результатов расчетов математической модели. Использование баз данных обеспечивает возможность проведения статистических анализов, а также оперативный доступ к данным.

*Блок-схема системы «Оценка потоков парниковых газов по спутниковым данным»*

# **Система усвоения данных, основанная на ансамблевом фильтре Калмана для оценки потоков парниковых газов с поверхности Земли: возможности**

---

- В разрабатываемой системе предусмотрена возможность использования данных из разных источников, а также других транспортных моделей.
- С помощью системы возможно проведение исследований по оценке свойств новых алгоритмов усвоения данных в условиях, максимально приближенных к реальности.
- Требуется анализ возможностей системы по проведению экспертных оценок временной динамики эмиссии метана с заданных территорий.

## Заключительные выводы

---

- Проблема парниковых газов требует наряду с развитием сети измерений разработки систем оценивания потоков парниковых газов с поверхности Земли с помощью современных технологий обработки данных.
- Разработка таких систем требует создания больших научных коллективов для проведения междисциплинарных исследований а также высокопроизводительных компьютеров.
- Существование готовой программной продукции в данной области не отменяет необходимости разработки такой системы и проведения исследований в области новых методов и алгоритмов.

*Спасибо за внимание!*

