

Применение современных алгоритмов искусственного интеллекта в задачах наук о Земле

Михаил Криницкий, к.т.н.

Лаборатория машинного обучения в науках о Земле МФТИ,
Институт океанологии РАН



Михаил Криницкий

к.т.н.

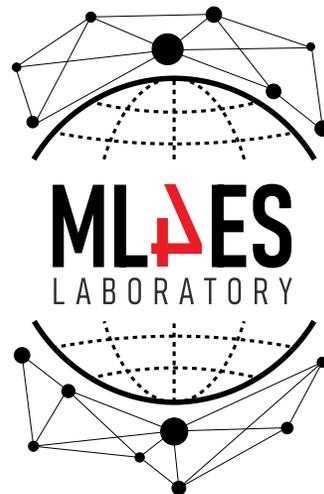
Лаборатория машинного обучения в науках о Земле МФТИ
Институт океанологии им. П.П. Ширшова РАН

Что мы делаем:

Разрабатываем, адаптируем, анализируем поведение алгоритмов машинного обучения и глубокого обучения в прикладных и фундаментальных задачах наук о Земле

krinitsky.ma@phystech.edu

<https://t.me/mkrinitskiy>



Задачи наук о Земле

ФУНДАМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

- Изучаем физические, геобиохимические, ... процессы в геосферах (атмосфера, океан, суша, литосфера, стратосфера, *etc.*) и их взаимодействие;
- Разрабатываем новые средства и методы измерений, экспедиционных и мониторинговых наблюдений;
- Разрабатываем новые способы извлечения знаний, сведений и информации о процессах, объектах и их характеристиках;

Задачи наук о Земле

ПРИКЛАДНЫЕ ИССЛЕДОВАНИЯ

- Разрабатываем и внедряем **новые способы моделирования процессов**, в т.ч. для отдельных регионов, акваторий;
- Разрабатываем и внедряем **новые методы валидации моделей**;
- Разрабатываем **новые способы применения моделей** на практике: в прогнозах, контуре оценки рисков.

Науки о Земле в 2026 г. – науки о данных

Науки о Земле – сфера Data Science

ECMWF* (октябрь 2022 г.):

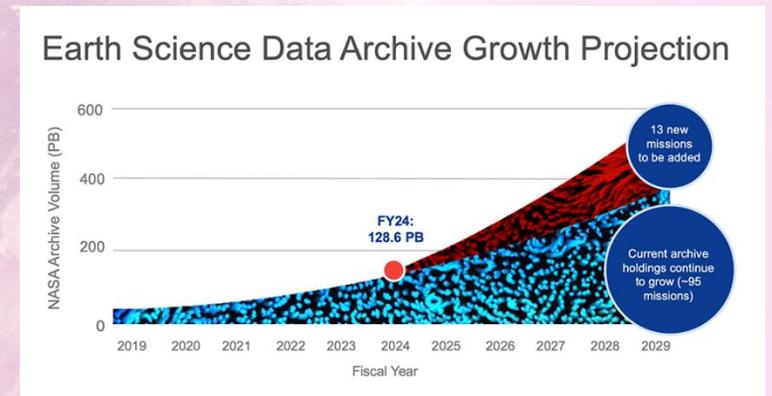
- **689 PB** в системе хранения и распространения MARS:
 - 510 PB первичных данных, 188 PB вторичных (резервные копии)
- ~ **10%** - данные наблюдений**, **90%** - модельные

NASA Earthdata** (2024 год):

- > **128 PB** доступных данных
- Прогноз на 2030 – **600 PB**

Copernicus Data Space Ecosystem*** (июнь 2025 г.):

- **80 PB** в системе хранения и распространения CDSE:
- > **100 PB** к 2028 г.



*Wilkinson, R., Mleczo, M. M., Brewin, R. J. W., Gaston, K. J., Mueller, M., Shutler, J. D., ... & Anderson, K. (2024). Environmental impacts of earth observation data in the constellation and cloud computing era. *Science of The Total Environment*, 909, 168584.

**<https://www.earthdata.nasa.gov/about/open-science>

***<https://earthobservations.org/about-us/news/from-open-data-to-empowering-global-insights-with-the-copernicus-data-space-ecosystem>

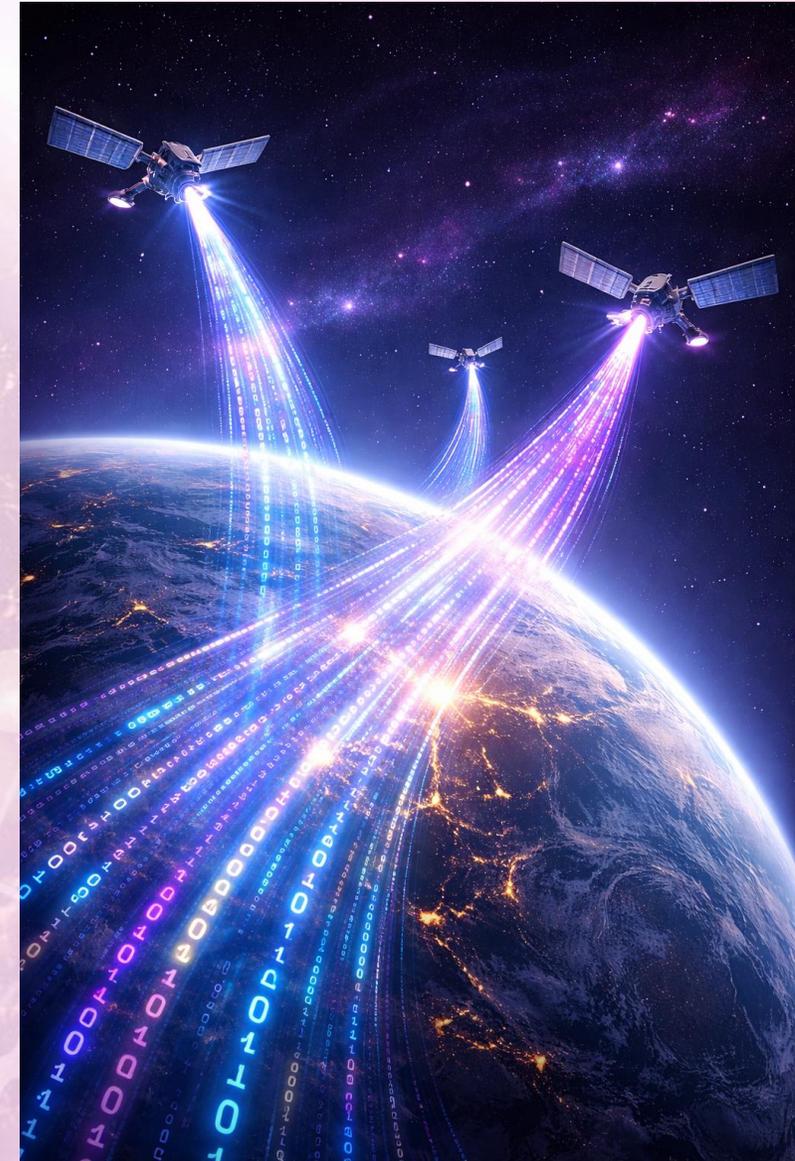
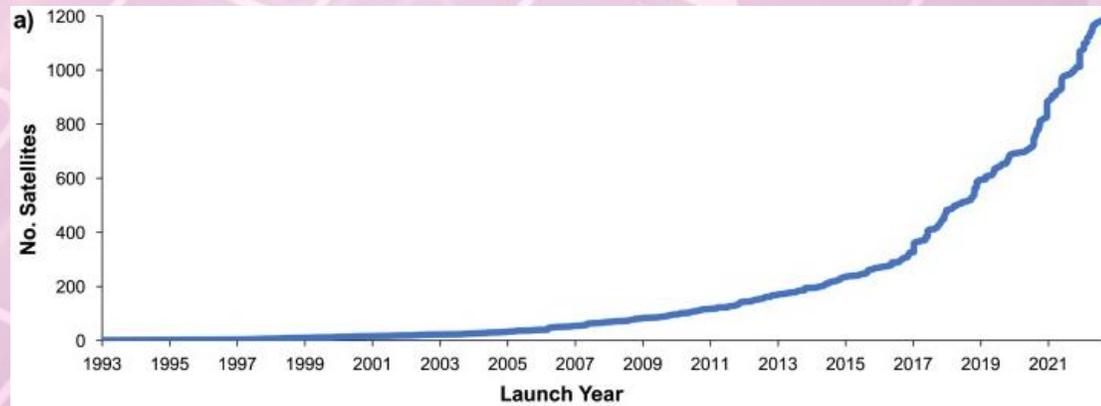
Скорость накопления данных в ЕС

ECMWF* (октябрь 2022 г.):

- +287 ТБ в сутки

Данные дистанционного зондирования:
более 100 ТБ в сутки*

Объемы данных RS будут расти до эксабайт**

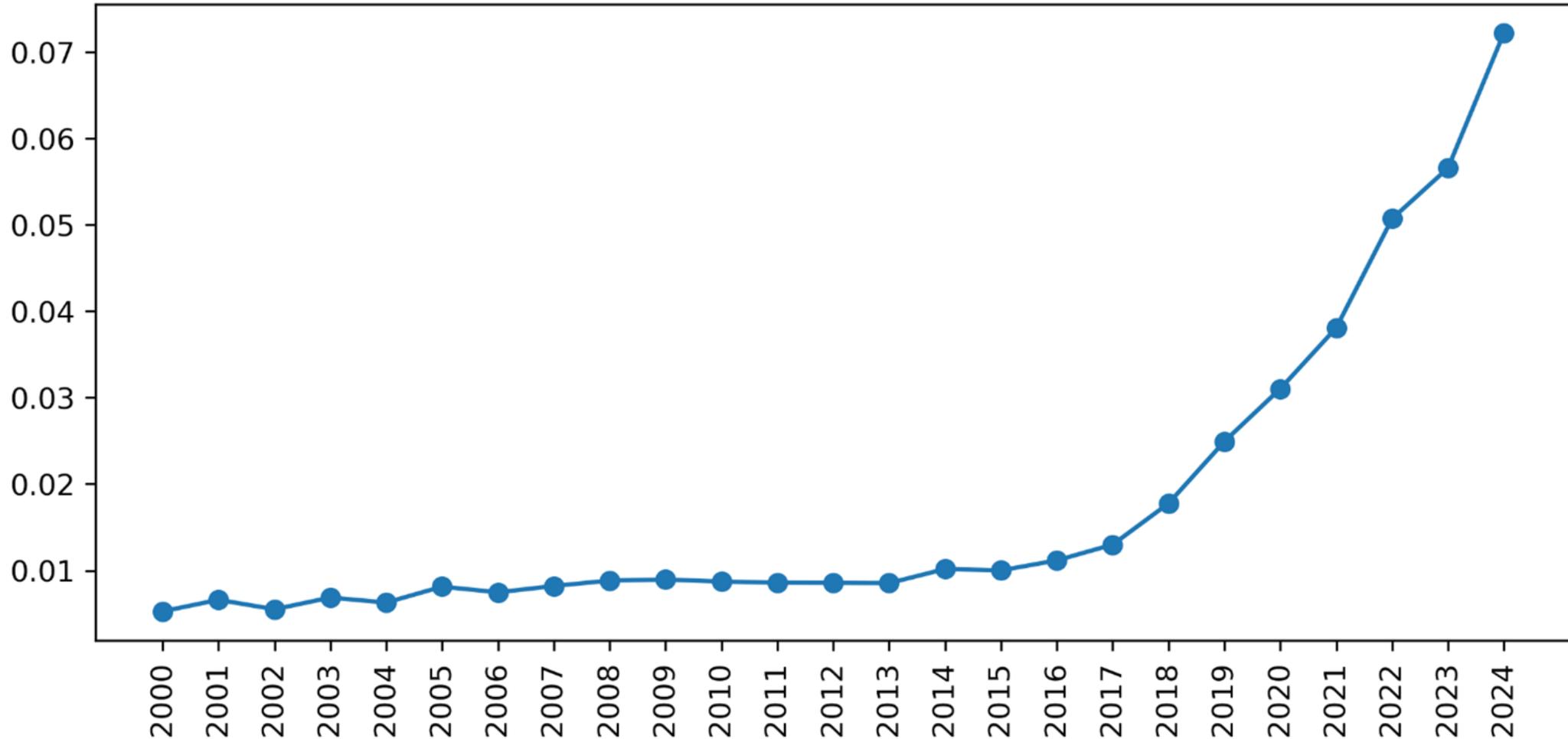


*<https://www.reorbit.space/resources/articles/earth-observation-today-rapid-data-from-space-unlocking-new-business-opportunities>

** Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems*, 51, 47-60.

DS/ML/AI in ES – неизбежность

Documents by year,
relative to total EARTH&ENVI documents

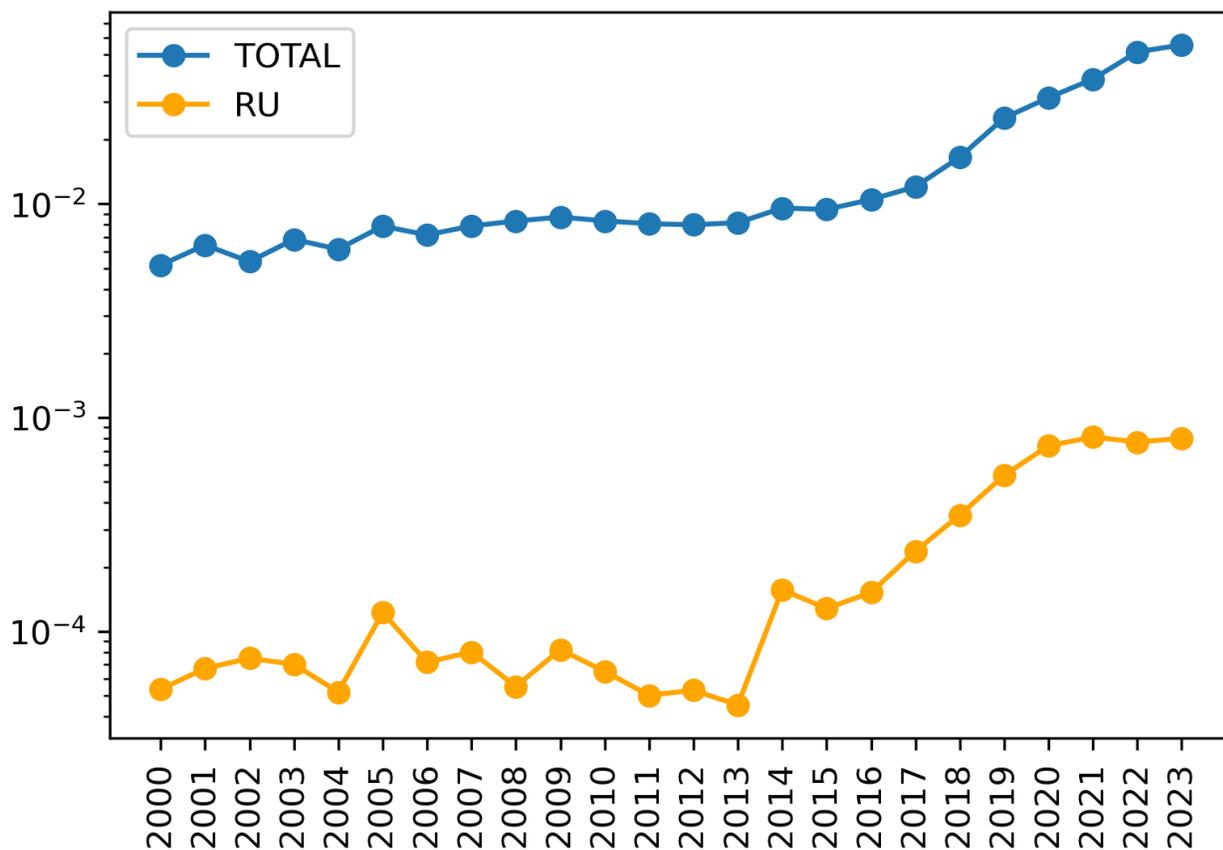


Осторожно, демотиватор

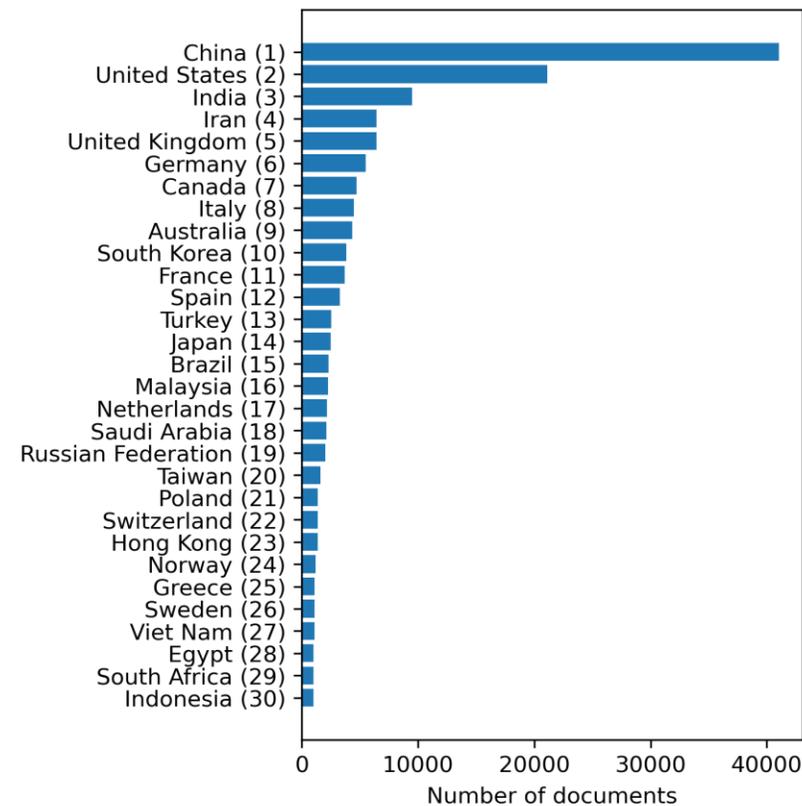
МО в науках о Земле

Статьи в рецензируемых журналах по тематике “Environmental sciences” и “Earth sciences” с применением методов машинного обучения. По данным Scopus.

Documents by year,
relative to total EARTH&ENVI documents

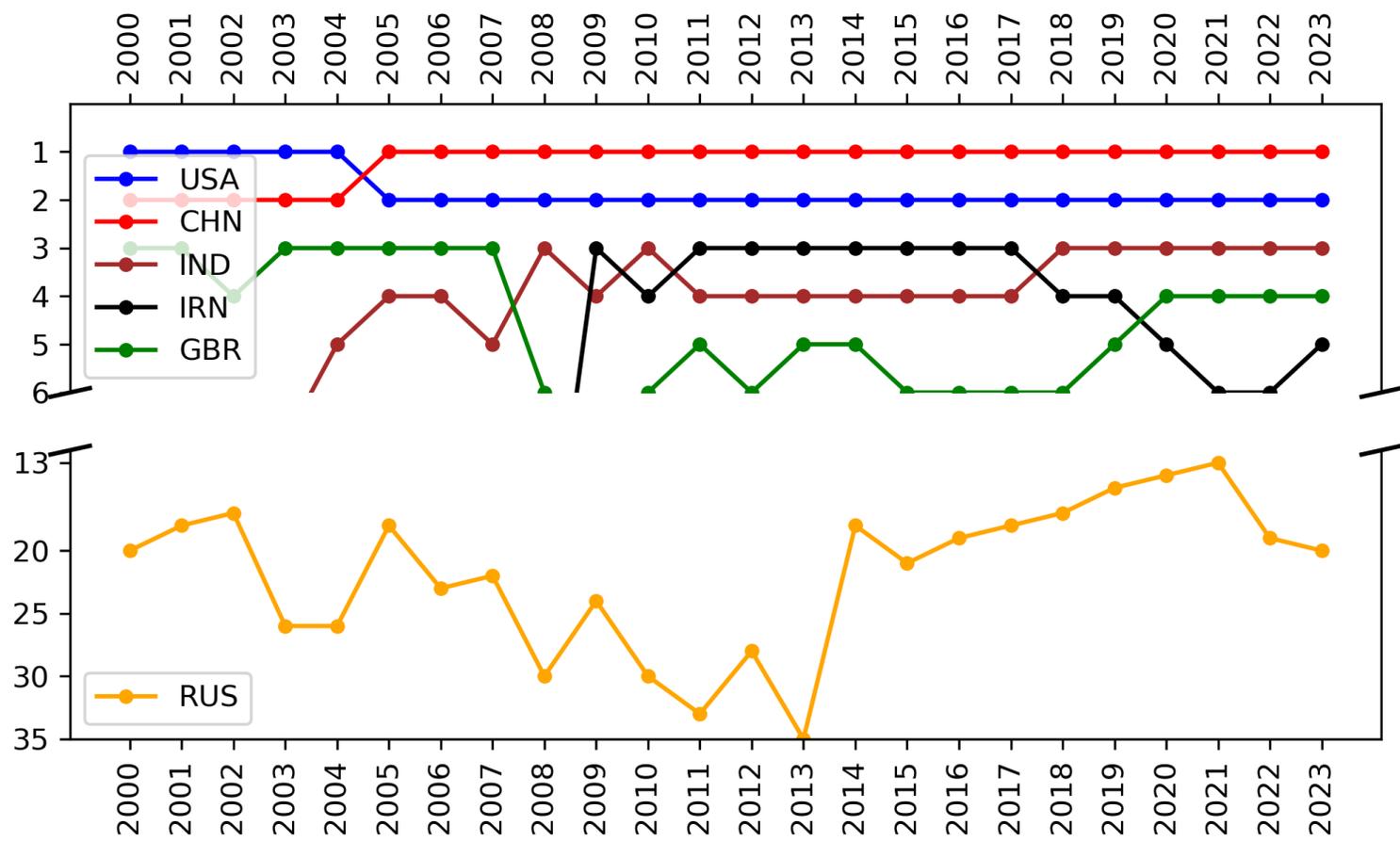


Documents, 2000-2023
AI&ML in EART&ENVI



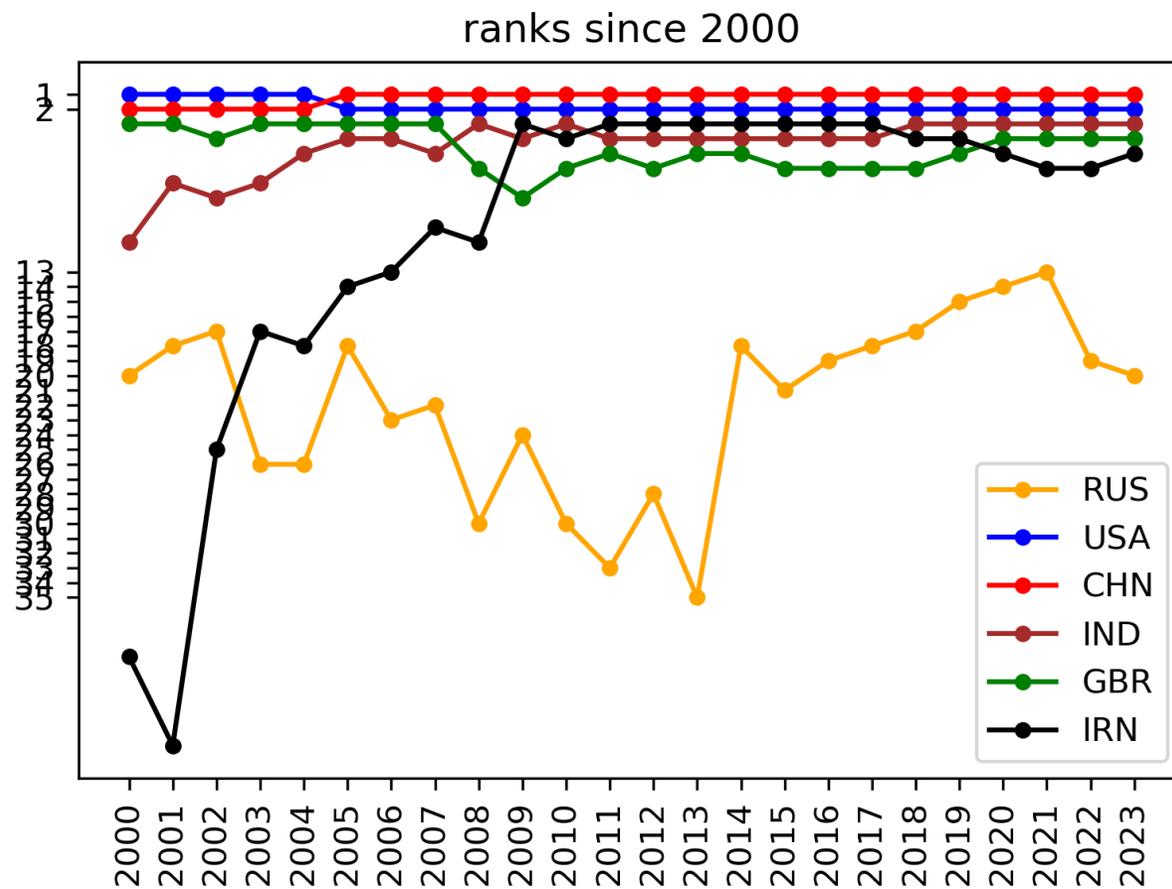
МО в науках о Земле

Статьи в рецензируемых журналах по тематике “Environmental sciences” и “Earth sciences” с применением методов машинного обучения. По данным Scopus.



МО в науках о Земле

Статьи в рецензируемых журналах по тематике “Environmental sciences” и “Earth sciences” с применением методов машинного обучения. По данным Scopus.



Что такое ИИ?

Что такое искусственный интеллект (по версии 2026 г.)?

- Определение ИИ (версия 1956 г.)

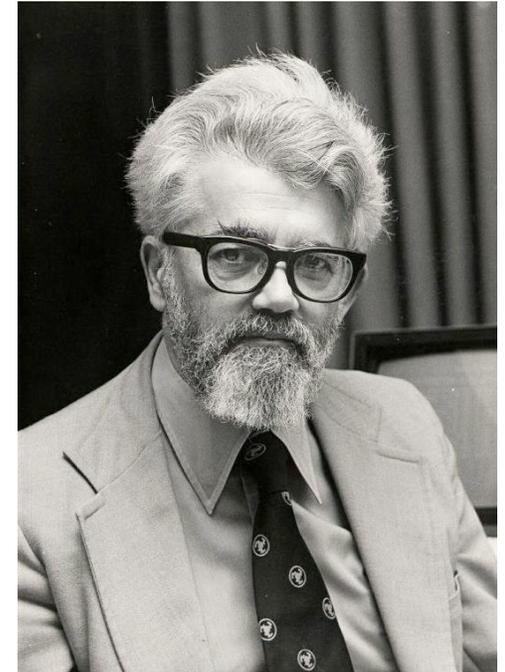
Искусственный интеллект (ИИ)

это наука и инженерная технология создания интеллектуальных машин, и в особенности интеллектуальных компьютерных программ. ИИ связан со сходной задачей использования компьютеров для понимания человеческого интеллекта, но не обязательно ограничивается биологически правдоподобными методами.

Murió John McCarthy

Дартмутский семинар по вопросам искусственного интеллекта,
Дартмутский колледж, ГанOVER, Нью-Гемпшир, США, 1956 г.

ИИ - научное направление, в рамках которого ставятся и решаются задачи аппаратного или **программного моделирования** тех **видов человеческой деятельности, которые традиционно считаются интеллектуальными**

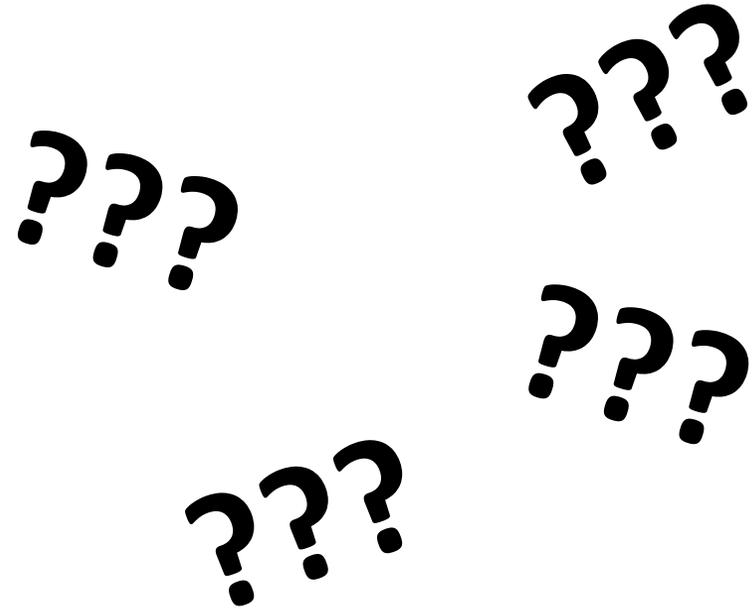


Что такое ИИ (по версии 2026 г.)?

- Более **70** определений интеллекта (2007 г.)*

Intelligence measures an agent's ability to achieve goals in a wide range of environments.

Интеллект – мера способности агента достигать целей в широком диапазоне типов окружения



Что такое ИИ (по версии 2026 г.)?

- Определение интеллекта (версия 2019 г.)

*The **intelligence** of a system is a measure of its **skill-acquisition efficiency** over a scope of tasks, with respect to priors, experience, and generalization difficulty.*

Интеллект системы – это мера **эффективности ее способности приобретать навыки** в различных задачах при условии заданных априорных предпосылок, опыта и сложности необходимых обобщений.

???

???

???

???



Francois Chollet,
программист (Google),
исследователь ИИ

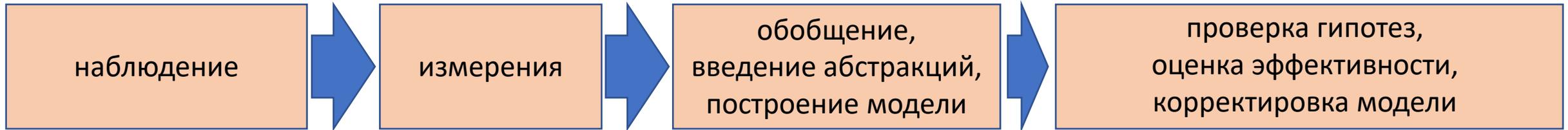
Что такое ИИ (по версии 2023 г.)?

- Строгого определения ИИ или даже просто интеллекта – нет
- При этом выделяют три вида ИИ:
 - Слабый ИИ (AI) – решает конкретные задачи, поставленные человеком
 - Общий ИИ (AGI) – умеет решать задачи, исследовать, ставить задачи
 - Супер-ИИ (ASI) – способности и возможности непознаваемы человеком
- Сегодня мы говорим о **СЛАБОМ ИИ**

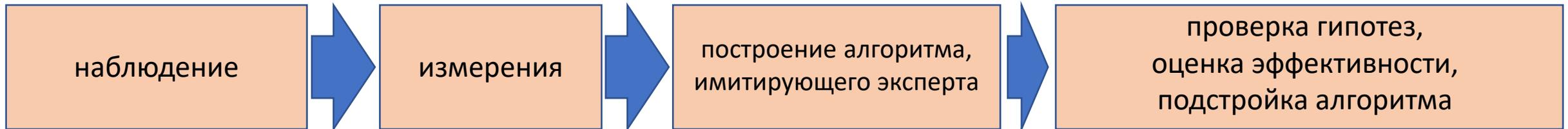
**Как обычно решают задачи
в естественных науках?**

ML/DL/AI: место в методологии

«классический» путь развития физических моделей («из первых принципов»)



эвристический способ решения задач (решения, основанные на правилах)



статистический способ (науки, основанные на данных, МО, Data Science)



методы машинного обучения

**Какого рода
бывают исследования?**

Фундаментальные, поисковые исследования

Прикладные исследования



Generative modeling

Self-supervised modeling

Unsupervised learning

Data structure exploration

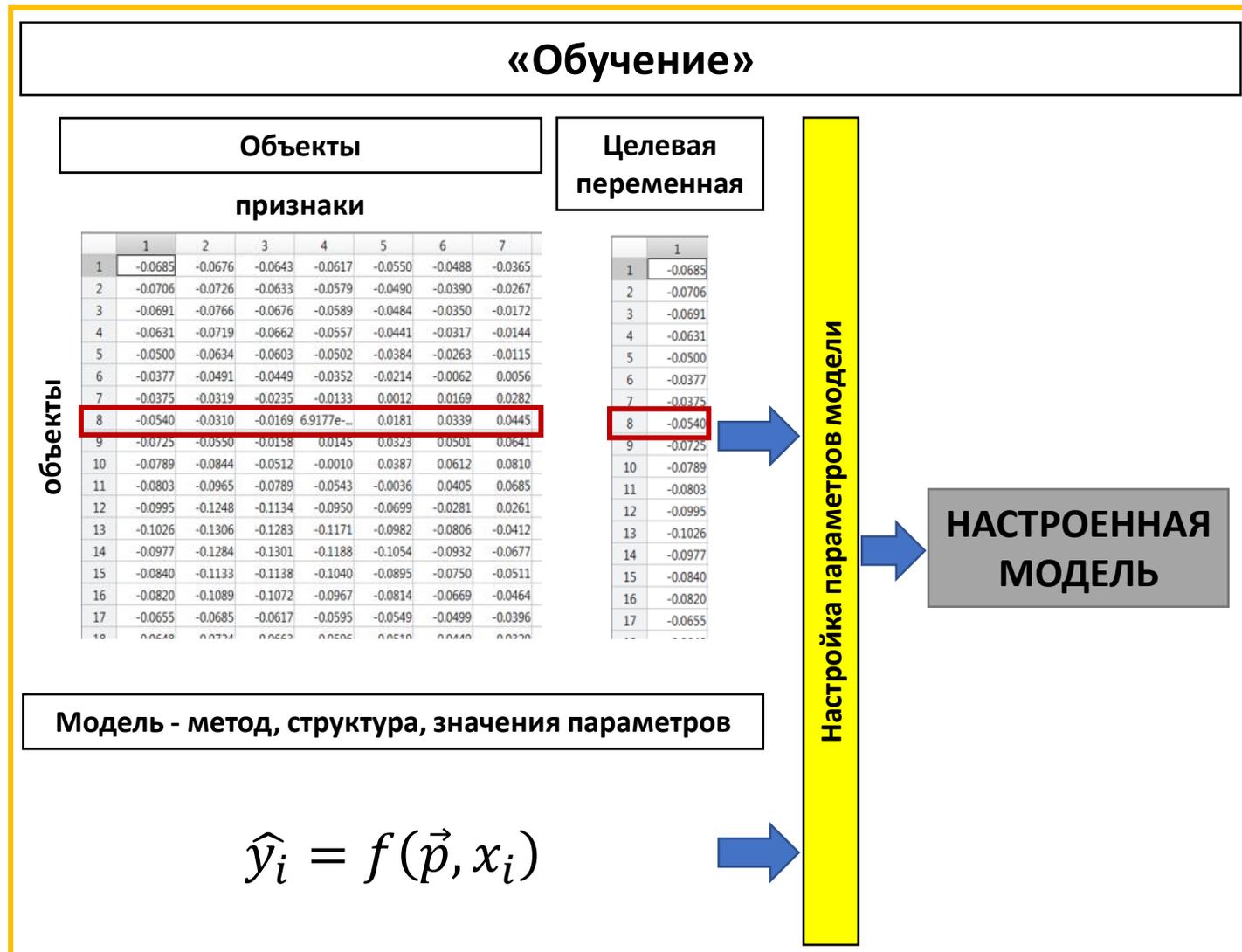
Dimensionality reduction

Supervised learning with monitoring data

Supervised learning with remote sensing data

Supervised learning with simulated data

Контролируемое обучение: типичная схема

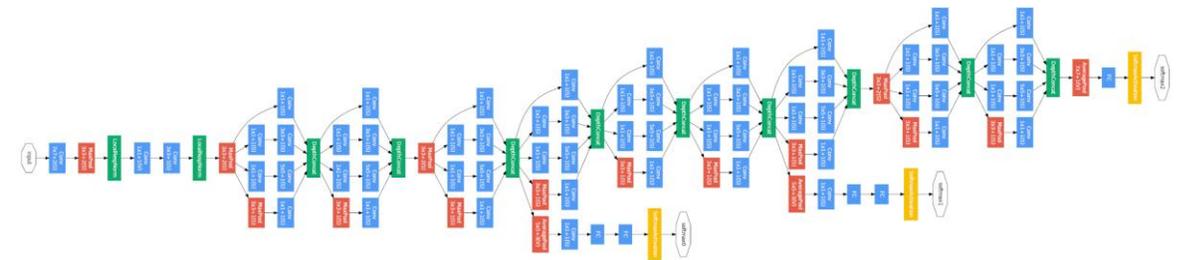
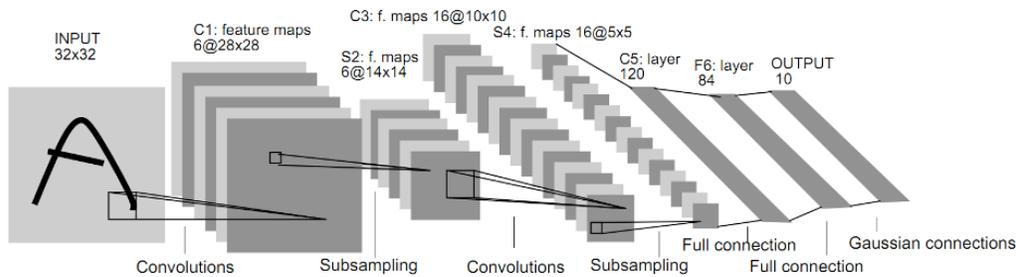
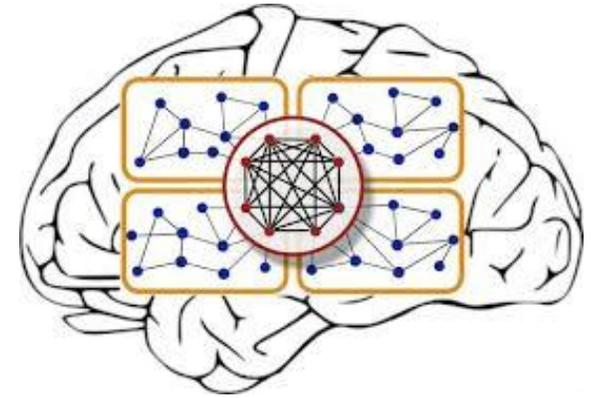
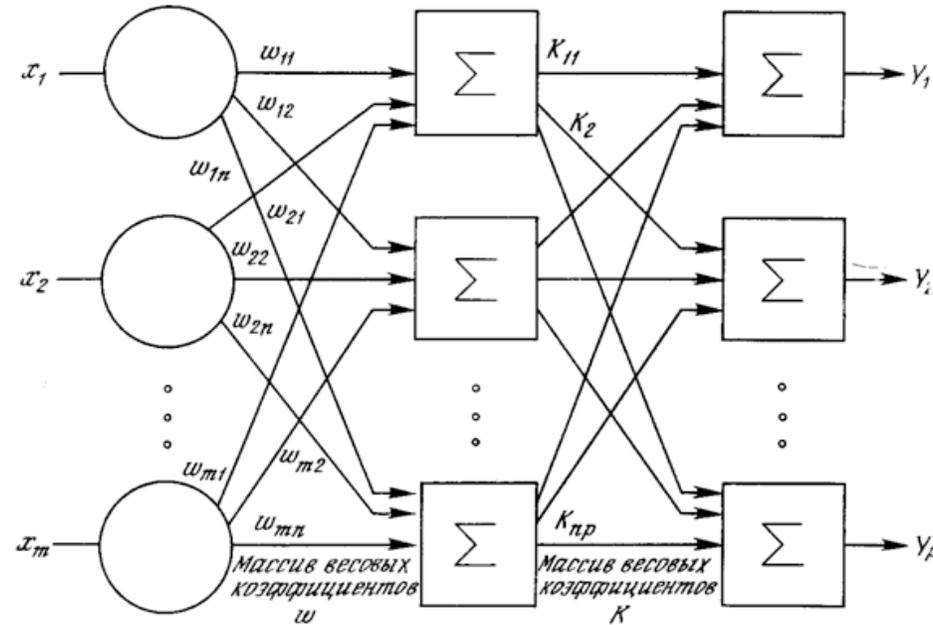
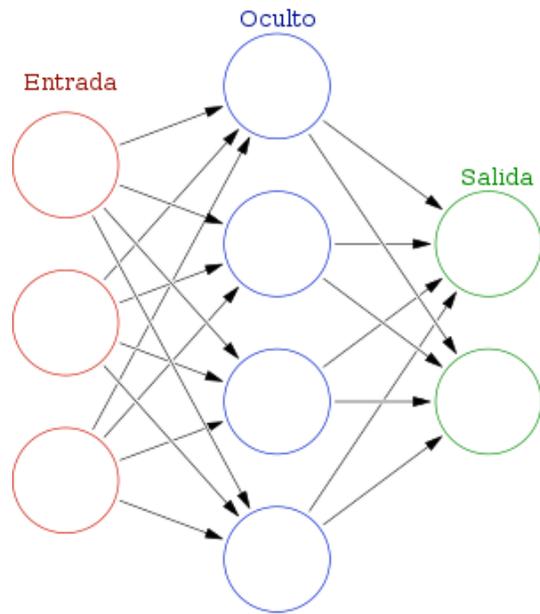


Контролируемое обучение

НЕОБХОДИМЫЕ КОМПОНЕНТЫ

- Обучающий набор данных (верное соответствие $y_i \leftrightarrow x_i$)
- Модель МО (предположение о виде связи между целевой переменной и признаками)
- Способ настройки модели МО (алгоритм обучения)
- Способ сохранения и восстановления структуры и полученных настроек модели
- Способ применения модели МО на новых данных (вычисления целевой переменной)
- Способ оценки качества модели МО

Искусственные нейронные сети

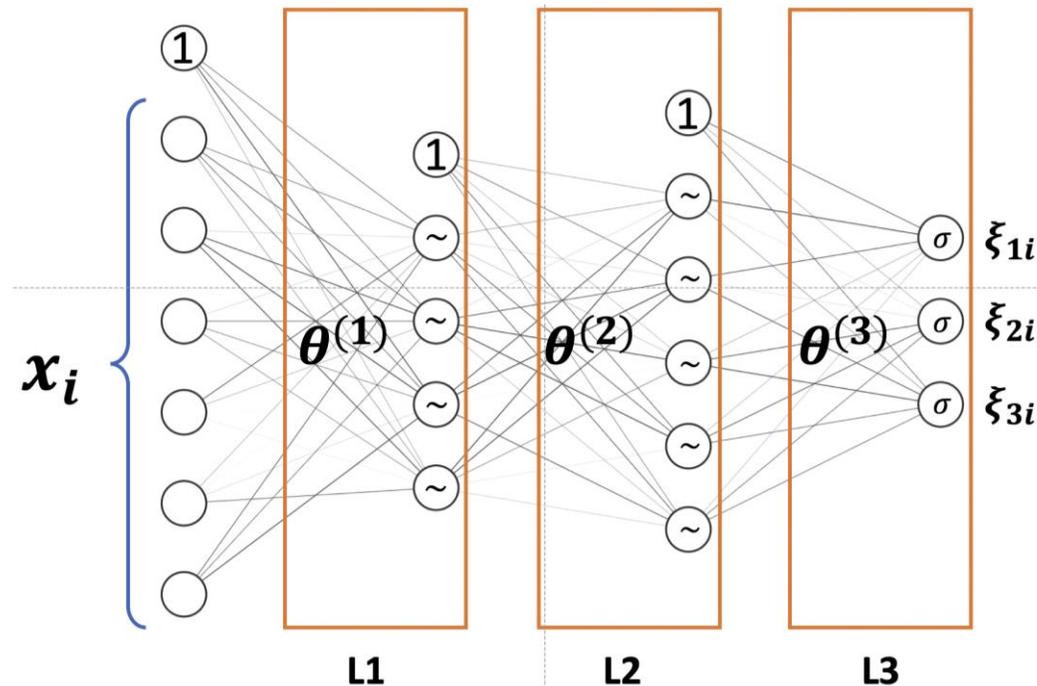


Искусственные нейронные сети

- ИНС - функция

$$\xi = \mathcal{F}(\theta, x)$$

$$\xi_i = \sigma \left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi \left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi \left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i \right) \right) \right)$$

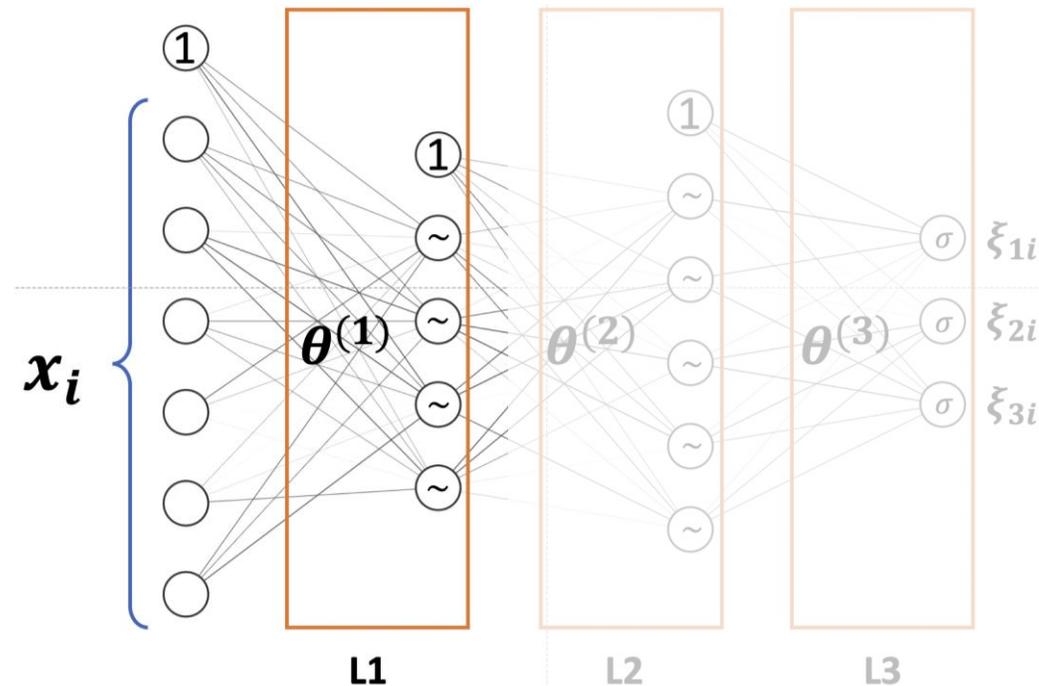


Искусственные нейронные сети

- ИНС - функция

$$\xi = \mathcal{F}(\theta, x)$$

$$\xi_i = \sigma \left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi \left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi \left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i \right) \right) \right)$$

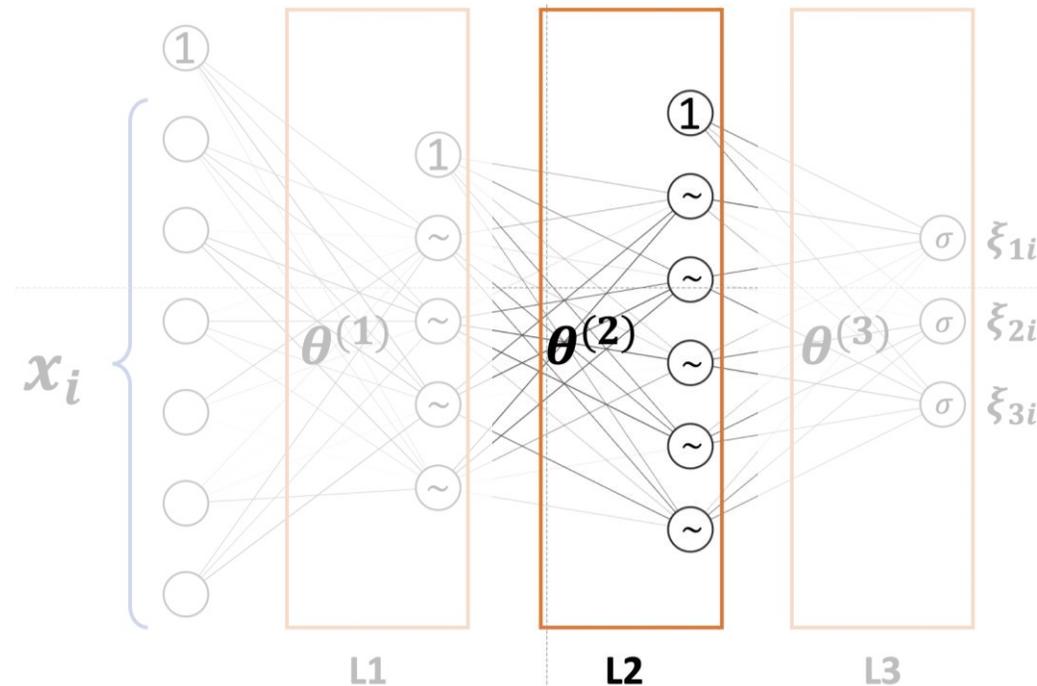


Искусственные нейронные сети

- ИНС - функция

$$\xi = \mathcal{F}(\theta, x)$$

$$\xi_i = \sigma \left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi \left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi \left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i \right) \right) \right)$$

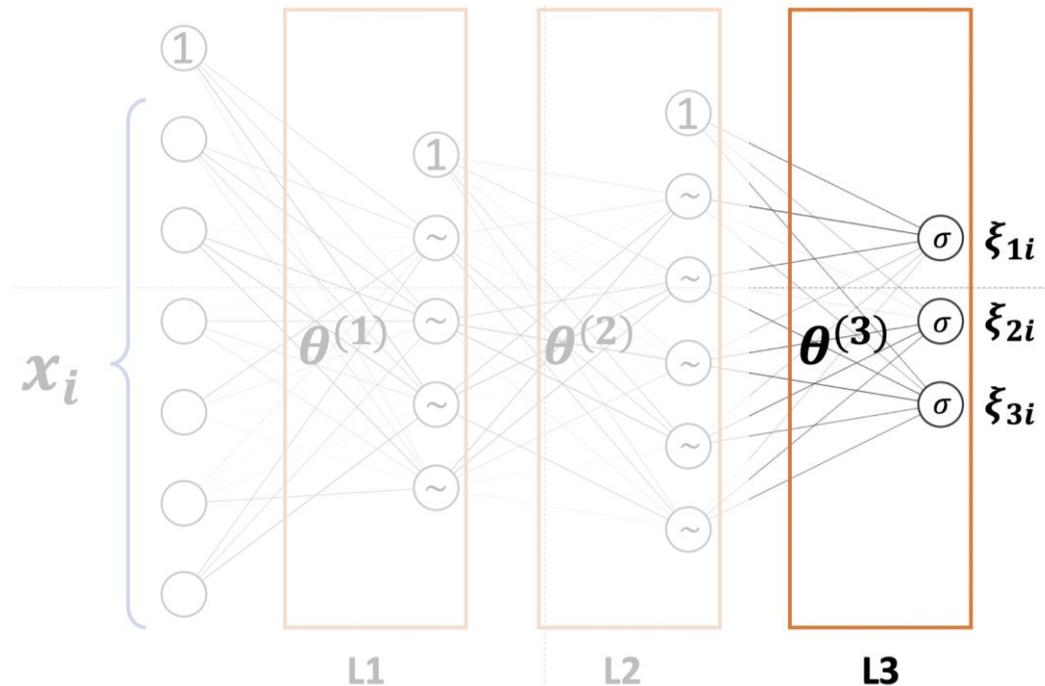


Искусственные нейронные сети

- ИНС - функция

$$\xi = \mathcal{F}(\theta, x)$$

$$\xi_i = \sigma \left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi \left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi \left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i \right) \right) \right)$$

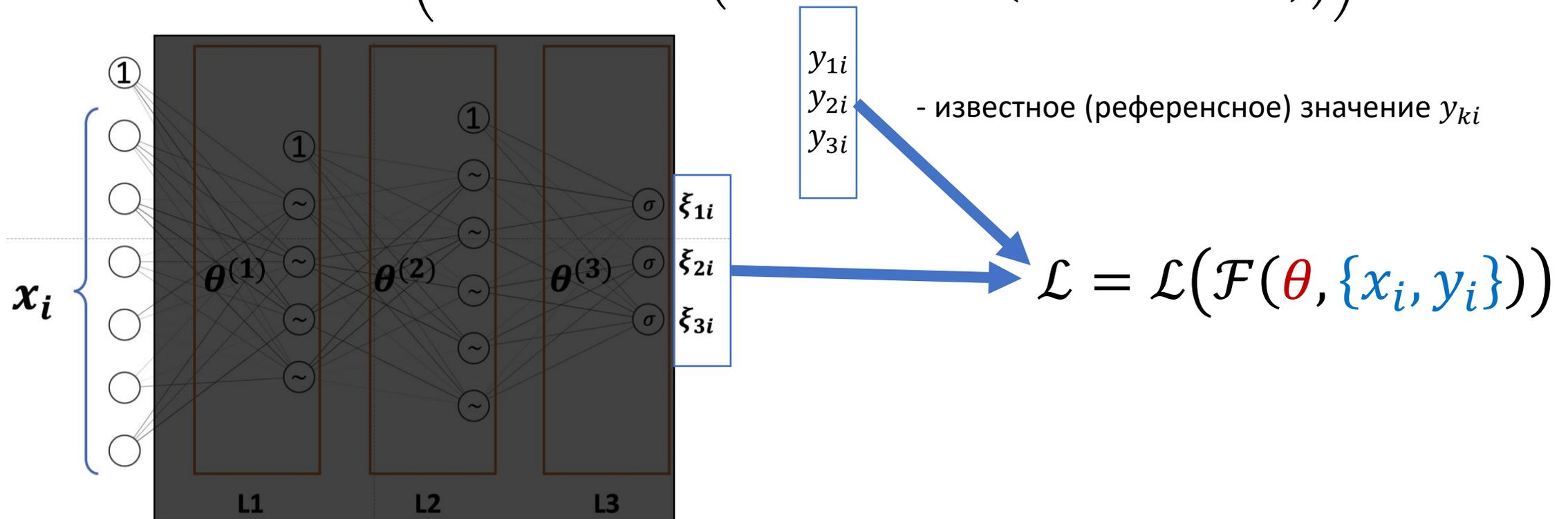


Искусственные нейронные сети

- ИНС - функция

$$\xi = \mathcal{F}(\theta, x)$$

$$\xi_i = \sigma \left(\theta_0^{(3)} + \theta^{(3)} \cdot \phi \left(\theta_0^{(2)} + \theta^{(2)} \cdot \phi \left(\theta_0^{(1)} + \theta^{(1)} \cdot x_i \right) \right) \right)$$



Искусственные нейронные сети

- ИНС – сложная нелинейная дифференцируемая функция

$$\xi_i = \mathcal{F}(\theta, x_i)$$

ξ_i - оценка целевой переменной для объекта, описываемого x_i

$$\mathcal{L} = \mathcal{L}(\mathcal{F}(\theta, \{x_i, y_i\}))$$

\mathcal{L} - функция потерь, функция ошибки, ...

$\nabla_{\theta} \mathcal{L} \Rightarrow$ оптимизация $\mathcal{F}(\theta, x)$
для решения задачи \mathcal{L}

Искусственные нейронные сети

- ИНС – сложная нелинейная дифференцируемая функция

x_i - признаки объекта (ковариаты, предикторы)

x_i - может быть:

- многомерным
- распределенным в пространстве
- описывающим эволюцию во времени

ξ_i - оценка целевой переменной для объекта, описываемого x_i

ξ_i - может быть:

- многомерной
- распределенной в пространстве
- описывающей эволюцию во времени



ПРИМЕРЫ

Фундаментальные, поисковые исследования

Прикладные исследования



Generative modeling

Self-supervised modeling

Unsupervised learning

Data structure exploration

Dimensionality reduction

Supervised learning with monitoring data

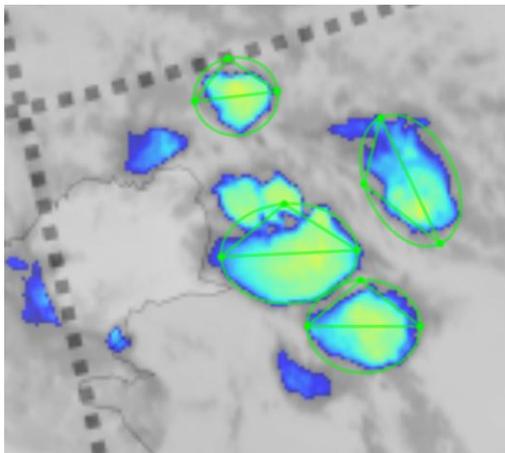
Supervised learning with remote sensing data

Supervised learning with simulated data

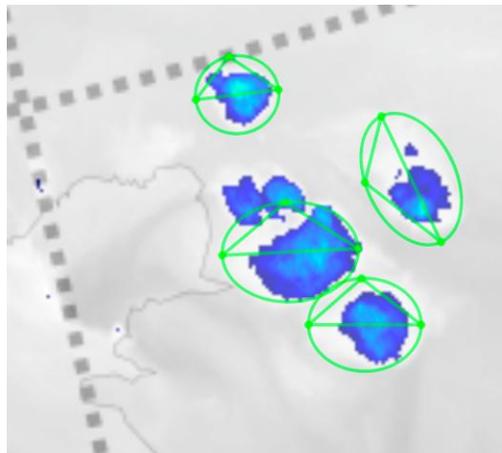
ИИ для обнаружения феноменов

Обнаружение мезомасштабных конвективных систем

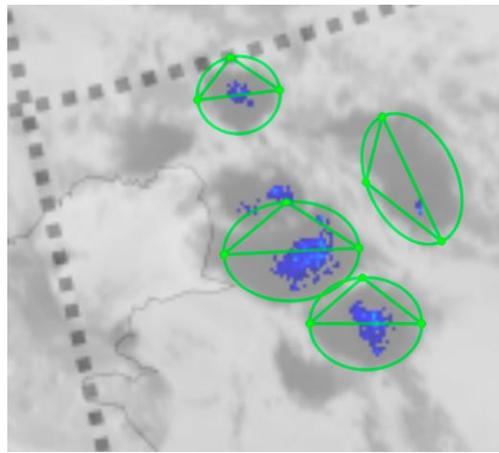
ch9 (10.8 μm)



ch5 (6.25 μm)



ch5 - ch9



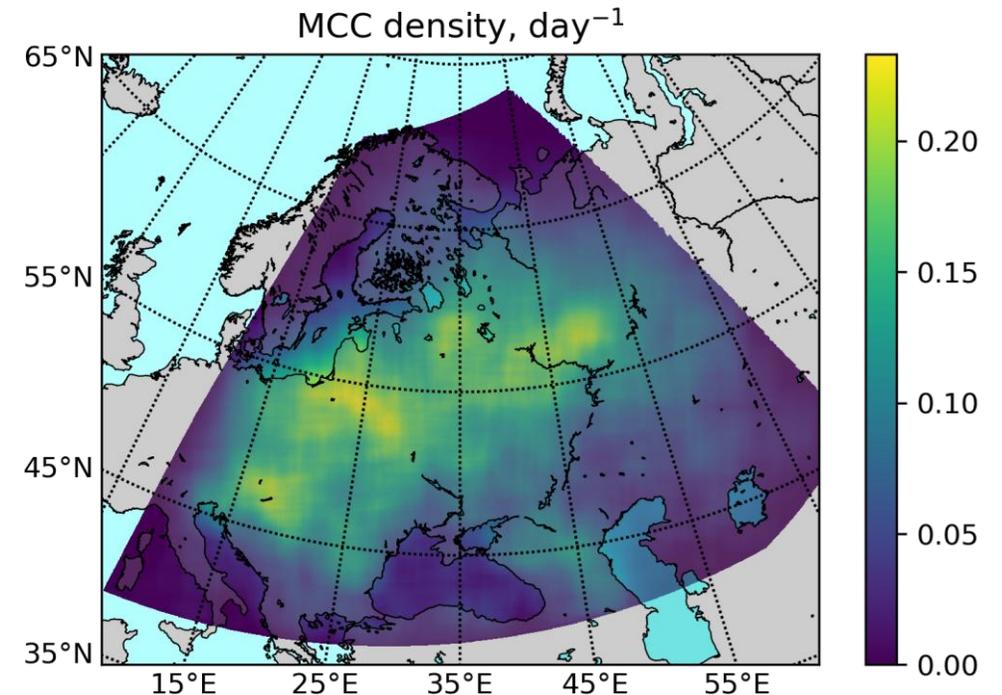
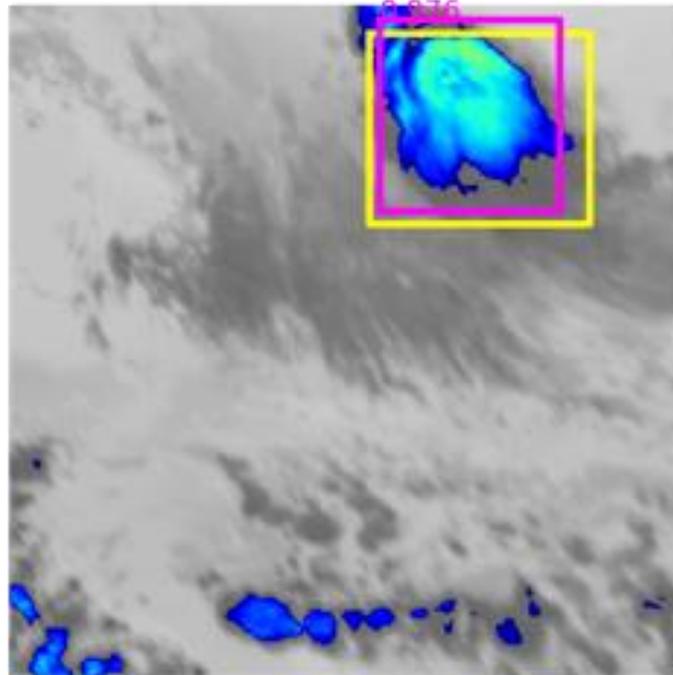
Meteosat (MSG4) data,
European territory of Russia

ИИ для обнаружения феноменов

Обнаружение мезомасштабных конвективных систем



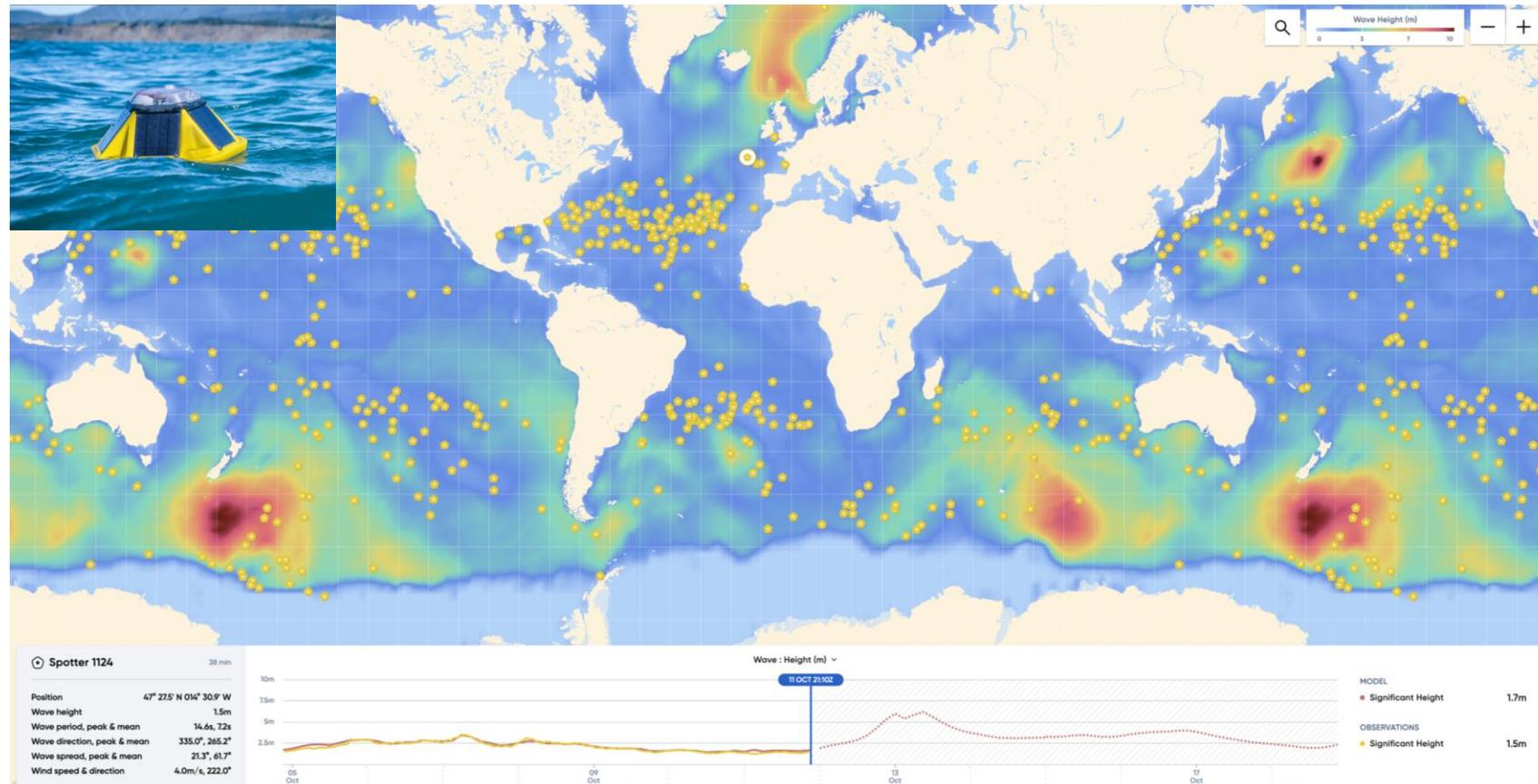
ch9 (10.8 μm)



Анализ и интерпретация данных наблюдений

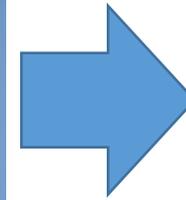
Характеристики ветрового волнения по данным навигационного радара

Дрейфующие буи Spotter



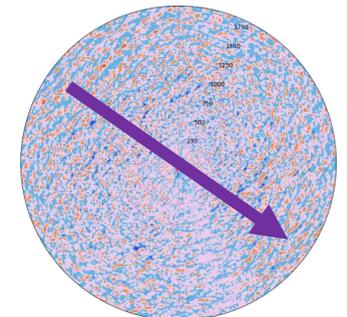
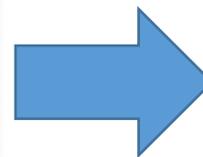
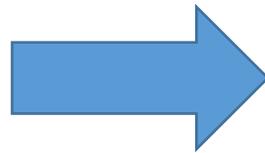
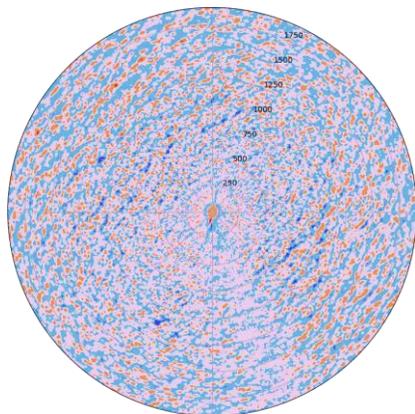
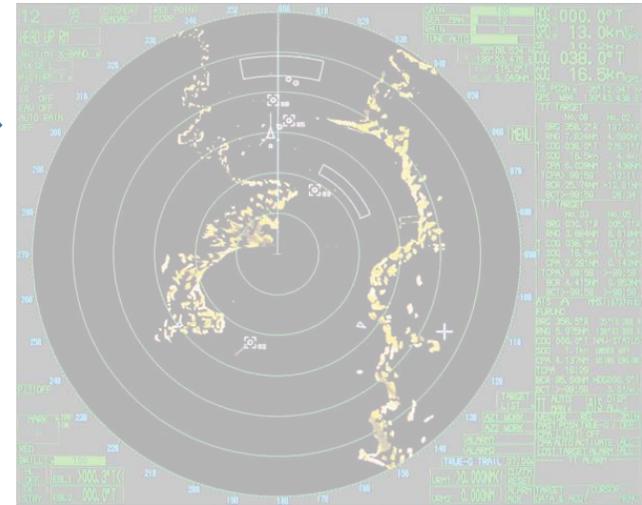
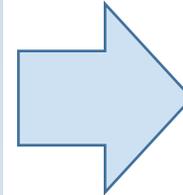
ИИ для измерений

Оценка характеристик ветрового волнения



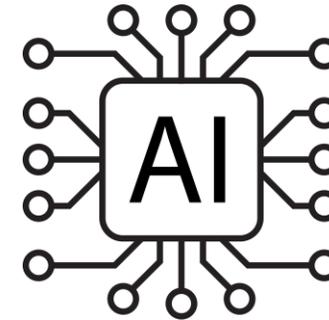
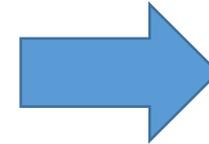
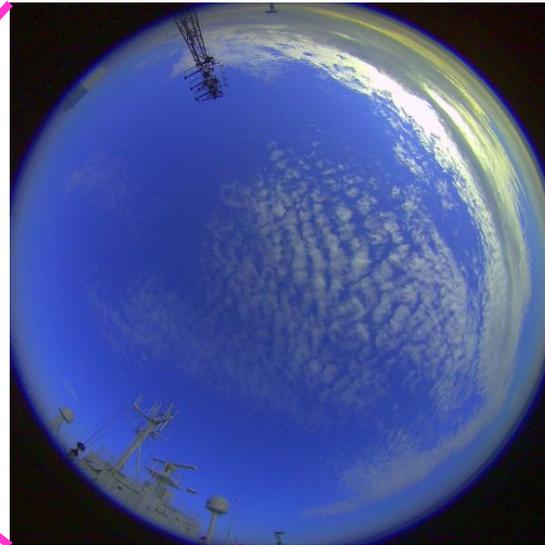
ИИ для измерений

Оценка характеристик ветрового волнения



ИИ для измерений

Характеристики облачности по данным цифровой фотосъемки небосвода



- Балл общей облачности
 - Типы облаков
- Высота нижней границы облачности
- Поток приходящей коротковолновой радиации

Krinitkiy, M.; Aleksandrova, M.; Verezemskaya, P.; Gulev, S.; Sinitsyn, A.; Kovaleva, N.; Gavrikov, A. On the Generalization Ability of Data-Driven Models in the Problem of Total Cloud Cover Retrieval. *Remote Sens.* **2021**, *13*, 326. <https://doi.org/10.3390/rs13020326>

Krinitkiy, M.; Koshkina, V.; Borisov, M.; Anikin, N.; Gulev, S.; Artemeva, M. Machine Learning Models for Approximating Downward Short-Wave Radiation Flux over the Ocean from All-Sky Optical Imagery Based on DASIO Dataset. *Remote Sens.* **2023**, *15*, 1720. <https://doi.org/10.3390/rs15071720>

Анализ и интерпретация данных наблюдений

Возможности и перспективы:

- Автоматизация мониторинговых наблюдений (облачности, волнения, *etc.*);
- Исключение человеческого фактора из процесса наблюдений;
- Расширение сети наблюдений («малой кровью»)

Анализ и интерпретация данных наблюдений

Вызовы:

- Малые объемы обучающих данных (???), высокая стоимость получения данных;
- Шум в исходных данных, зависимость достоверности данных от чистоты эксперимента, показаний наблюдателя;
- **Вопрос доверия:** экспертные наблюдения все еще считаются более достоверными;
- **Вопрос интерпретируемости результатов ИИ:** решение эксперта объяснимо, решение ИИ - нет
- **Вопрос применимости/обобщения/воспроизводимости:** эффективность стат. моделей (ИИ) падает при смене условий съемки, региона, камеры, радара, *etc.*

Спасибо за внимание

Михаил Криницкий, к.т.н.

МФТИ, Институт океанологии РАН, МГУ