



# Искусственный интеллект в науках о Земле

Михаил Криницкий

К.Т.Н.

Зав. лабораторией машинного обучения в науках о Земле МФТИ  
с.н.с. Института океанологии РАН им. П.П. Ширшова

PREVIOUSLY

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, МАШИННОЕ ОБУЧЕНИЕ, ГЛУБОКОЕ ОБУЧЕНИЕ

## ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



## MACHINE LEARNING

Machine learning begins to flourish.



## DEEP LEARNING

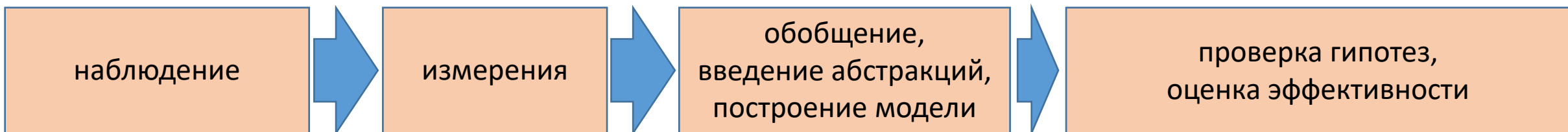
Deep learning breakthroughs drive AI boom.



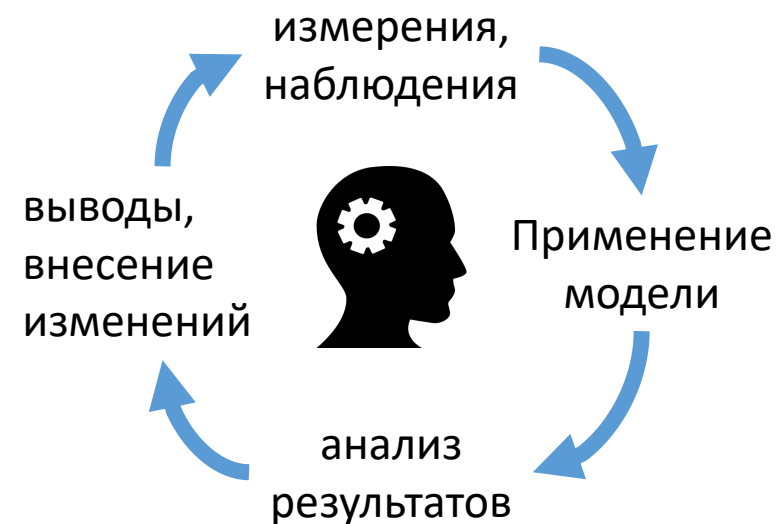
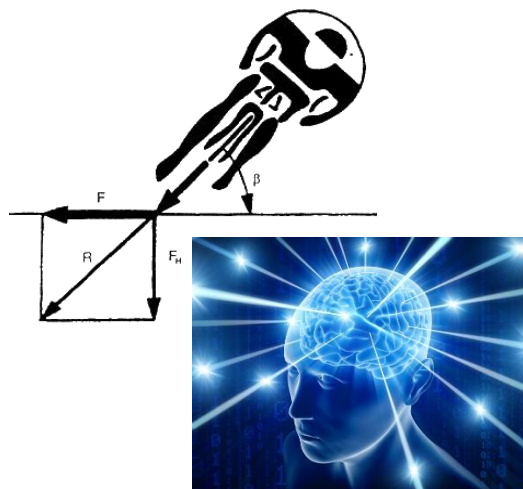
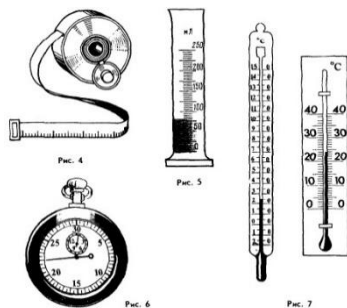
PREVIOUSLY

# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

## КАК проводятся физические исследования?



by Frits Ahlefeldt



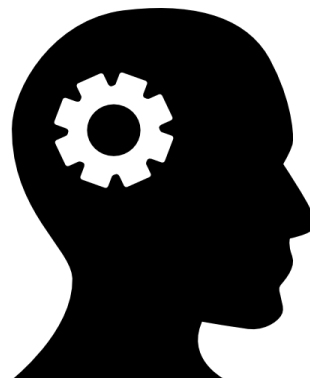
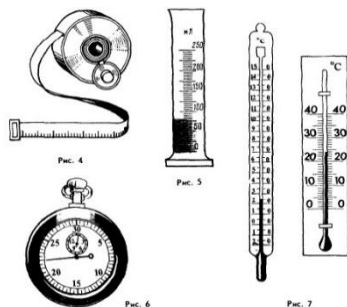
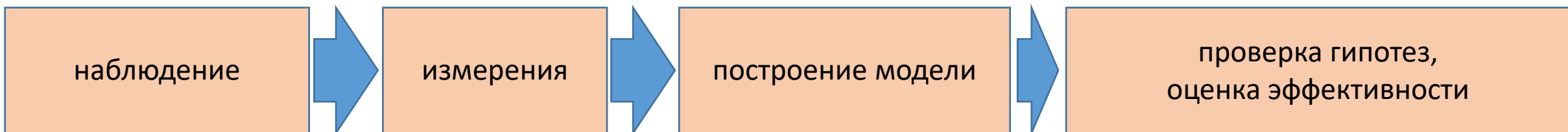
Настоящая наука начинается с тех пор, как начинают измерять.  
Точная наука немыслима без меры.  
Д.И. Менделеев

PREVIOUSLY

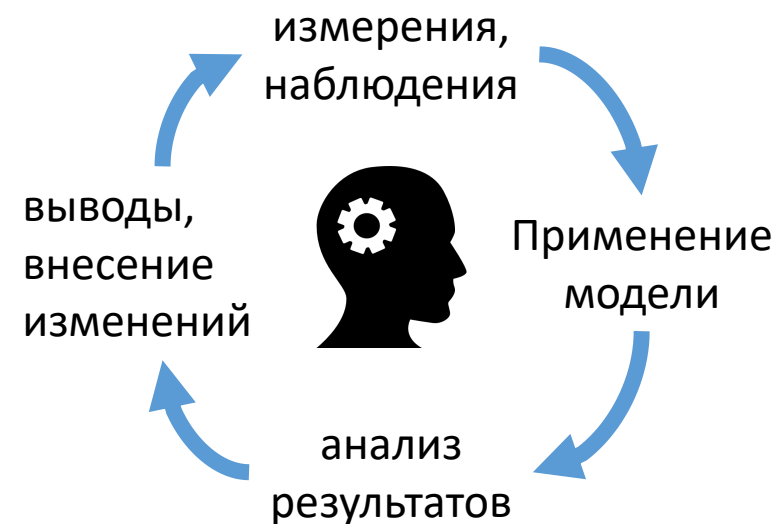
# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит

все равно строим модель



обобщение ?  
введение абстракций ?



# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит  
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы)  $\{y_i\}$
- Данные измерений  $\{x_i\}$
- $\mathcal{F}: \mathcal{X} \rightarrow \mathcal{Y}$
- Обобщающая. Достоверная (в каком смысле?)

Применимая.

PREVIOUSLY

# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (**человеку**) непонятно, что происходит  
**все равно строим модель**

## КАК?

Методы машинного обучения

Искусственный интеллект

Теория Вапника-Червоненкиса

Статистическая теория восстановления  
зависимостей по эмпирическим данным

Машинный интеллект

PREVIOUSLY

# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

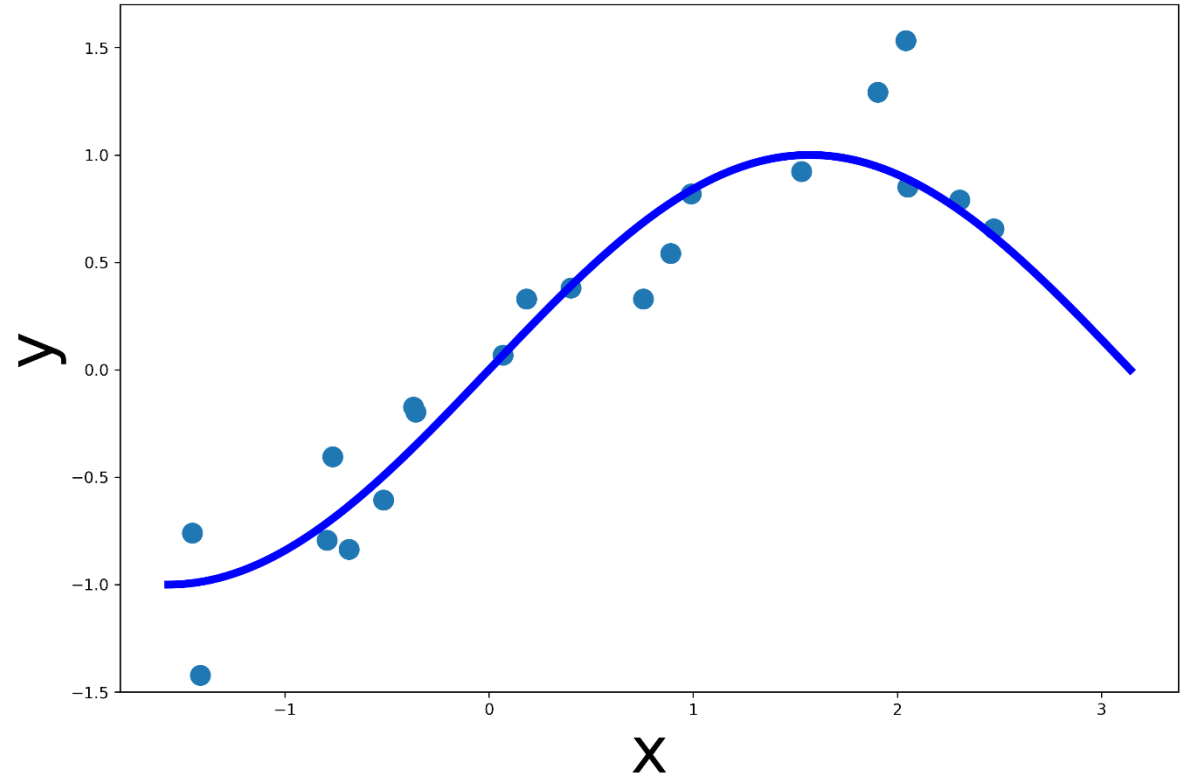
строим модель для решения задачи

типы задач:

○ «Обучение с учителем»

- восстановление регрессии

**что я хочу?** – значение ***y***



PREVIOUSLY

# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

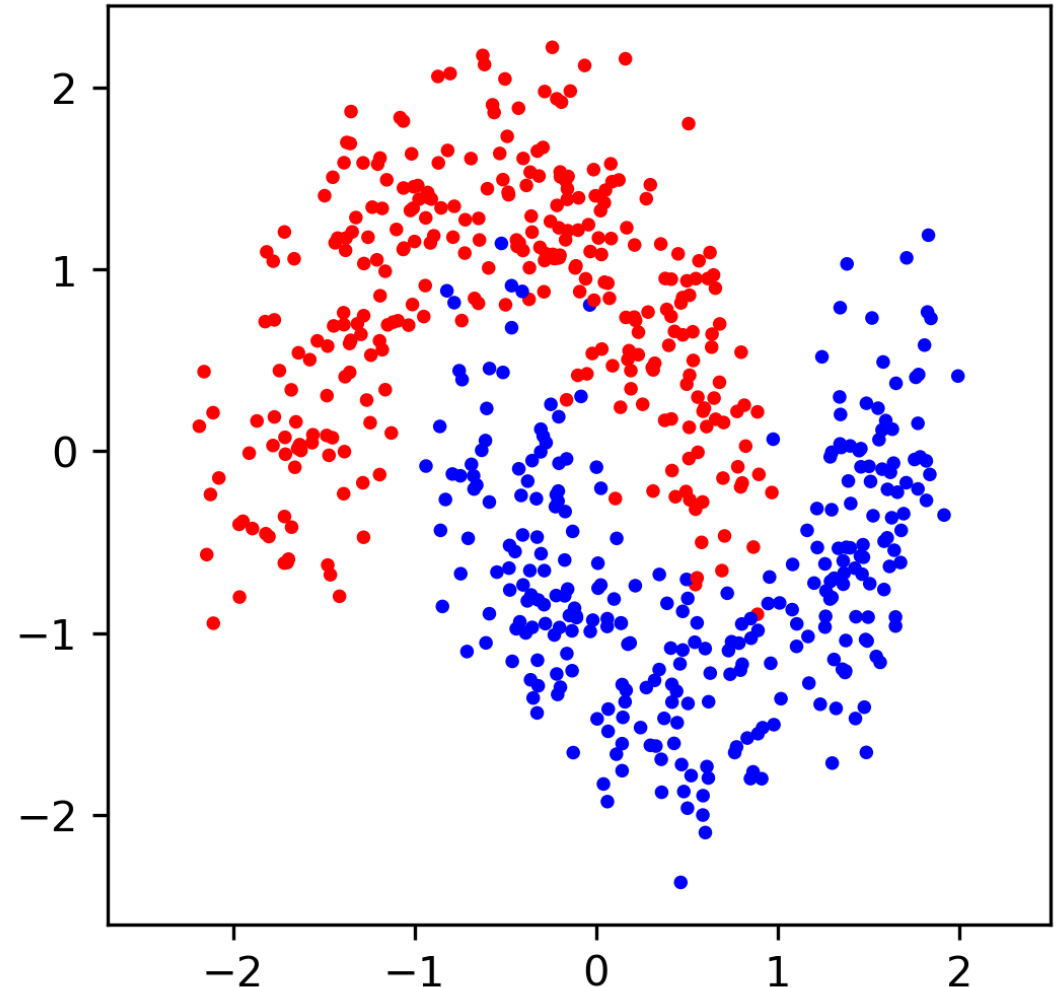
строим модель для решения задачи

типы задач:

○ «Обучение с учителем»

- восстановление регрессии
- классификация

**что я хочу?** — метку класса  
(красный или синий?)





# Понятия в МО

- Объекты/события
- Признаковое описание объектов/событий  $\mathbf{x}$  – случайная величина
- Реализация признакового описания для  $i$ -го объекта/события  $\mathbf{x}_i$
- Целевая переменная  $y$  – случайная величина
- Реализация целевой переменной для  $i$ -го объекта/события  $y_i$
- Множество возможных векторов признакового описания  $\mathbf{X}$
- Множество возможных значений (исходов) целевой переменной  $\mathbf{Y}$
- Отображение  $\mathcal{F}: \mathbf{X} \rightarrow \mathbf{Y}$  – модель МО, иногда статистической или вероятностной природы

# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

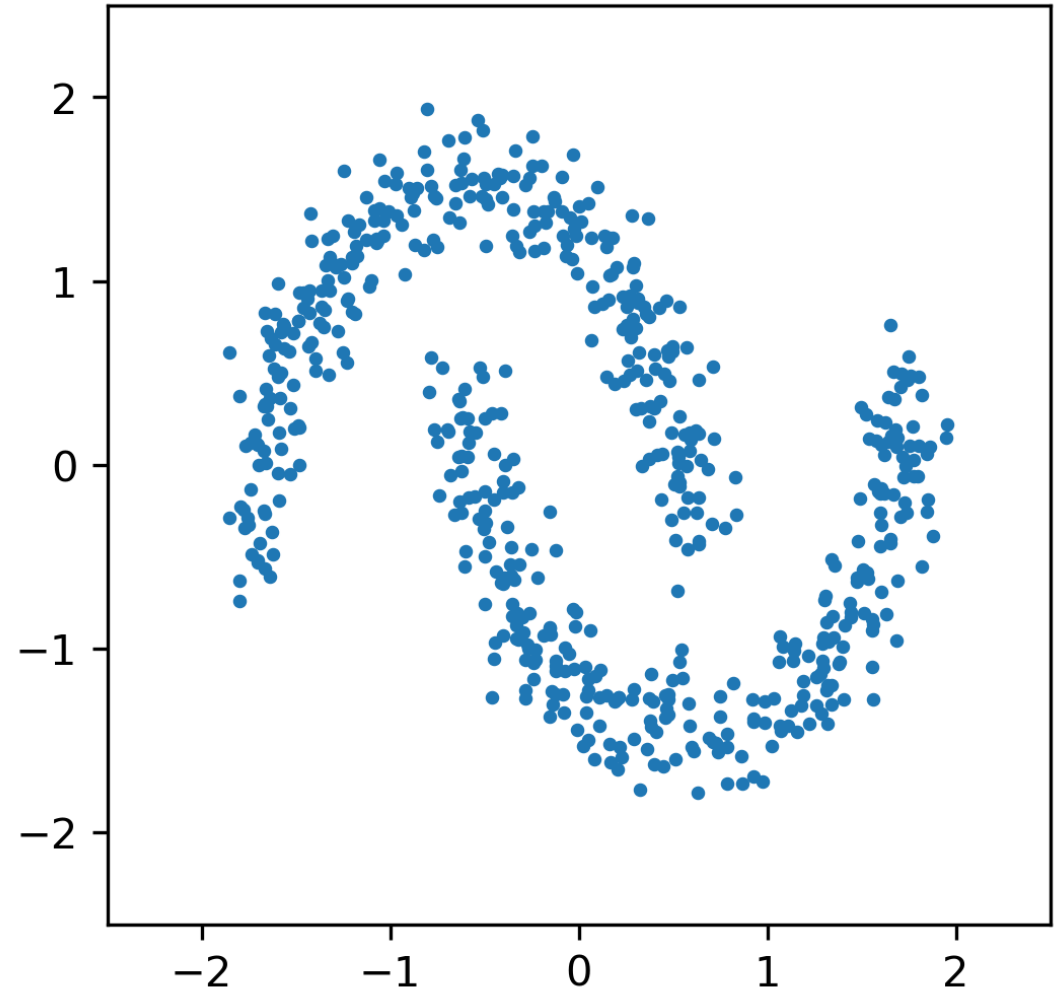
строим модель для решения задачи

типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - поиск структуры в данных

**что я хочу?**

- метки групп
- знать, есть ли группы?
- сколько групп?



# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

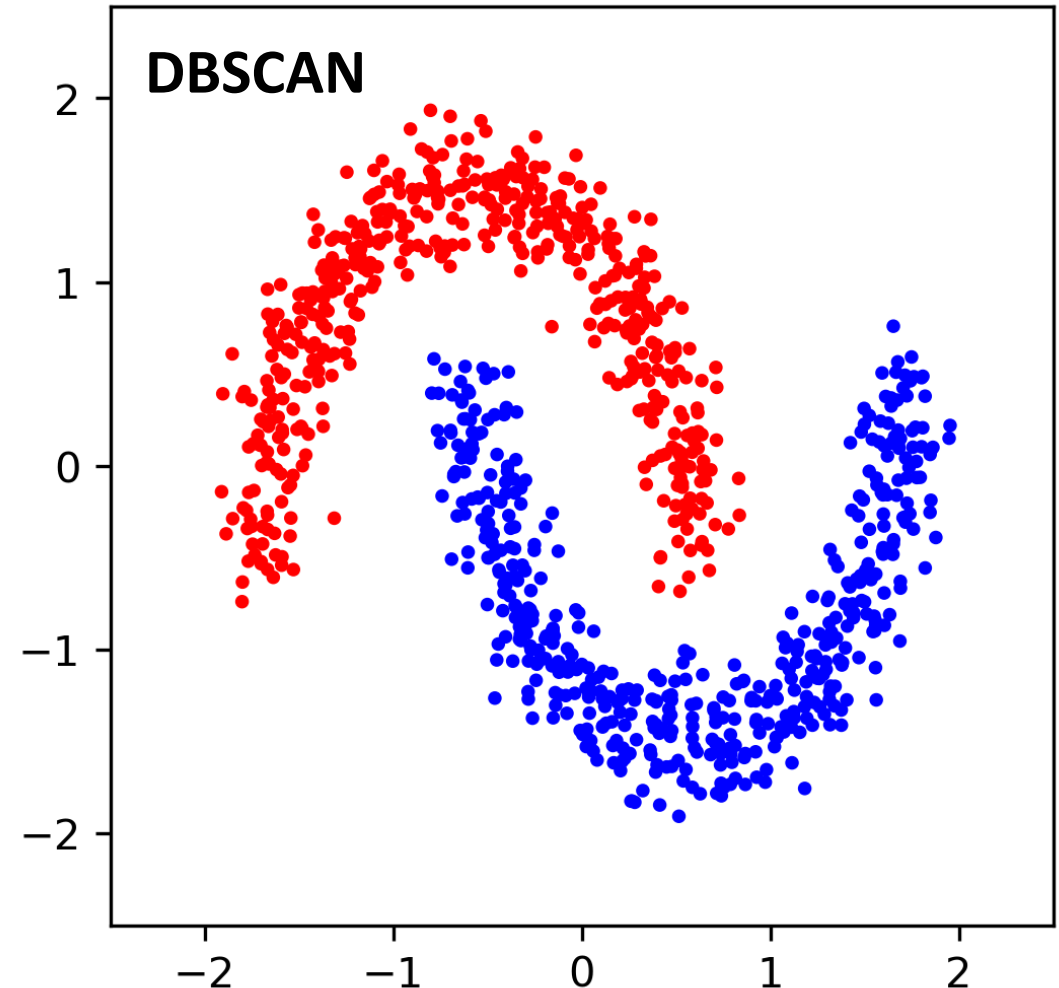
строим модель для решения задачи

типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - кластеризация

**что я хочу?**

- метки групп
- знать, есть ли группы?
- сколько групп?



# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

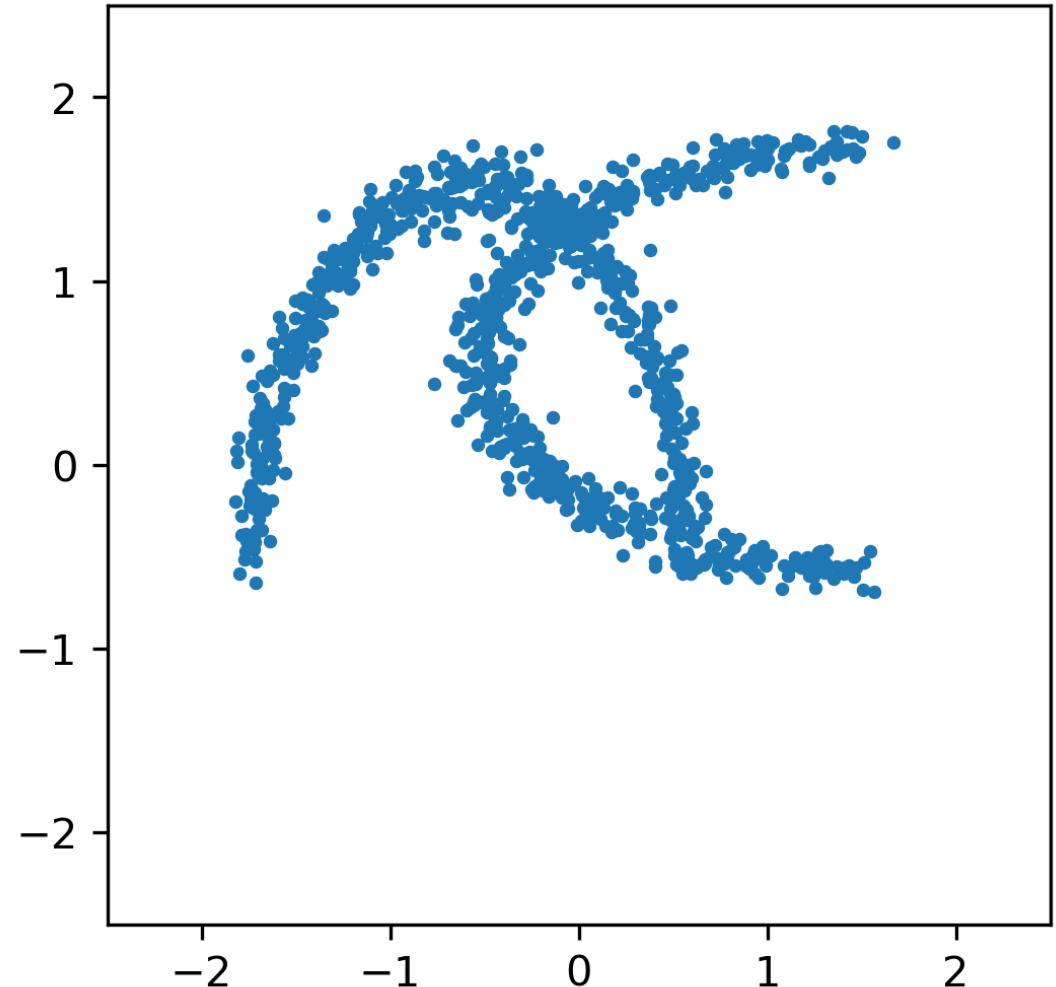
типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - кластеризация

**Всегда ли есть решение?**

хоть какое-нибудь

**ДА**



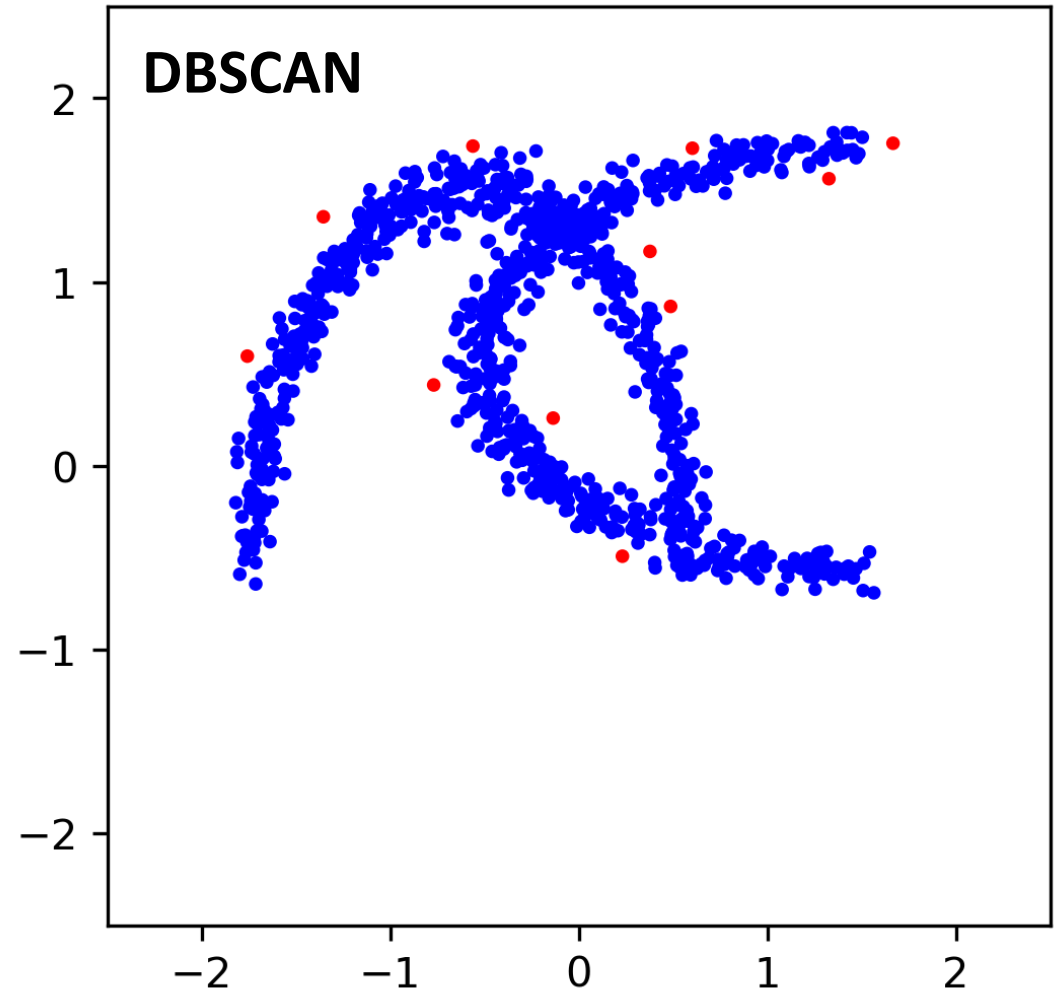
# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - кластеризация

Всегда ли есть решение,  
**которое мне понравится?**



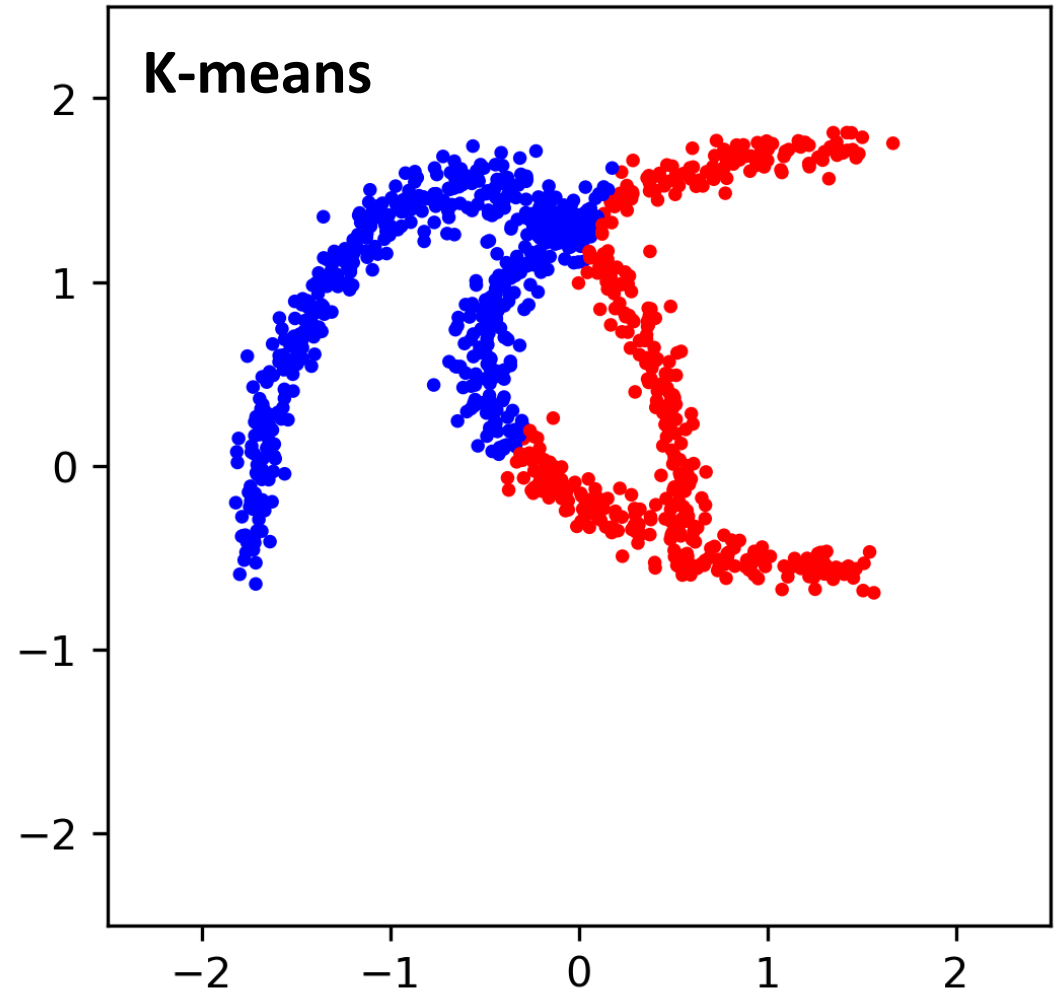
# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - кластеризация

Всегда ли есть решение,  
**которое мне понравится?**



# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

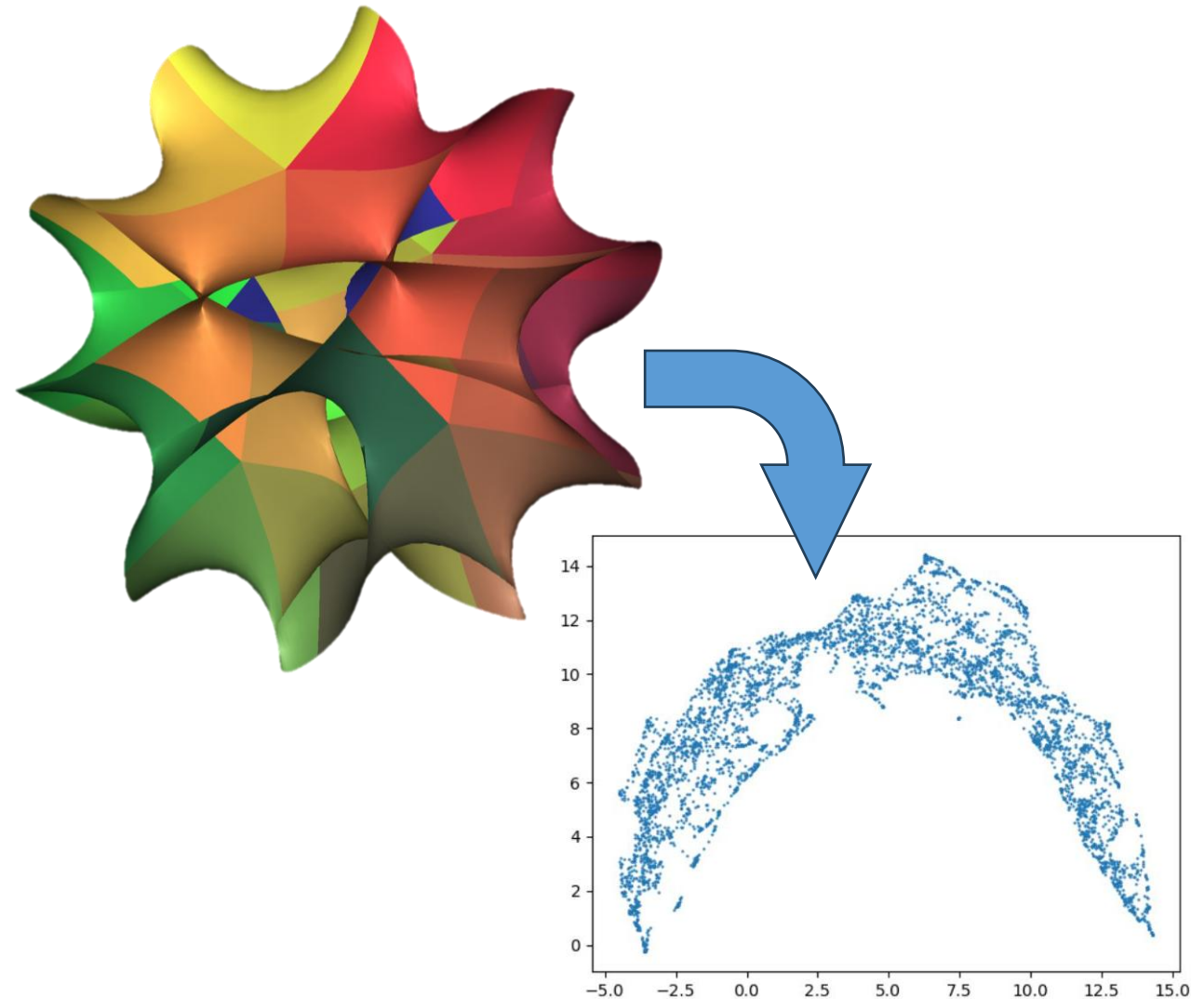
○ «Обучение без учителя»

- снижение размерности

**что я хочу?**

признаковое описание

сниженной размерности



# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

○ «Обучение без учителя»

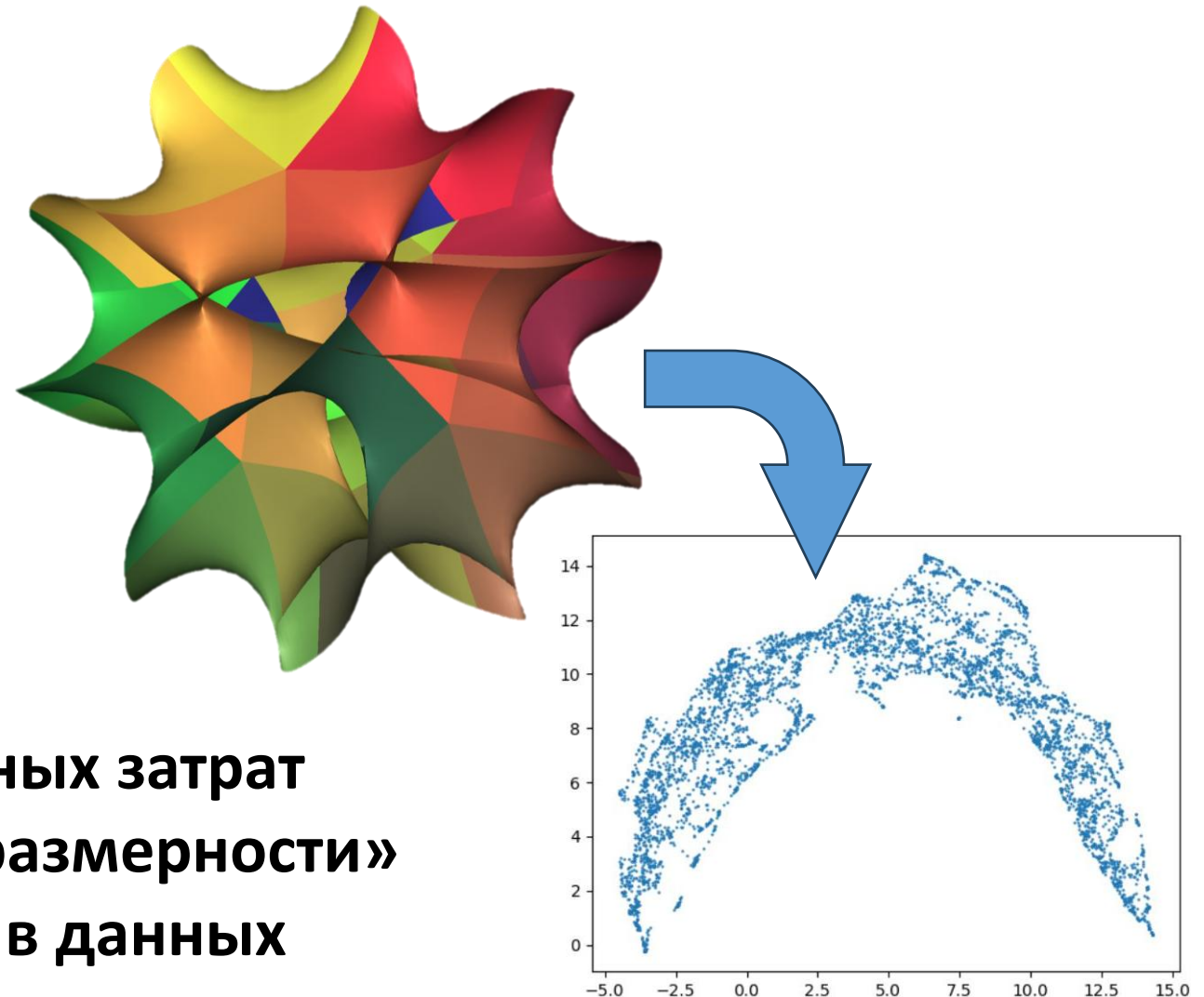
- снижение размерности

**что я хочу?**

признаковое описание  
сниженной размерности

**зачем?**

- визуализация данных
- снижение вычислительных затрат
- борьба с «проклятием размерности»
- снижение уровня шума в данных





# ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

- «Обучение с учителем»
  - восстановление регрессии
  - классификация
- «Обучение без учителя»
  - Кластеризация
  - Снижение размерности
  - Аппроксимация распределения данных
- Другие задачи: смежные, редкие, специальные.
  - самоконтролируемое обучение
  - с частичным привлечением учителя
  - обучение с подкреплением
  - выучивание меры различия (дистанции)
  - ...

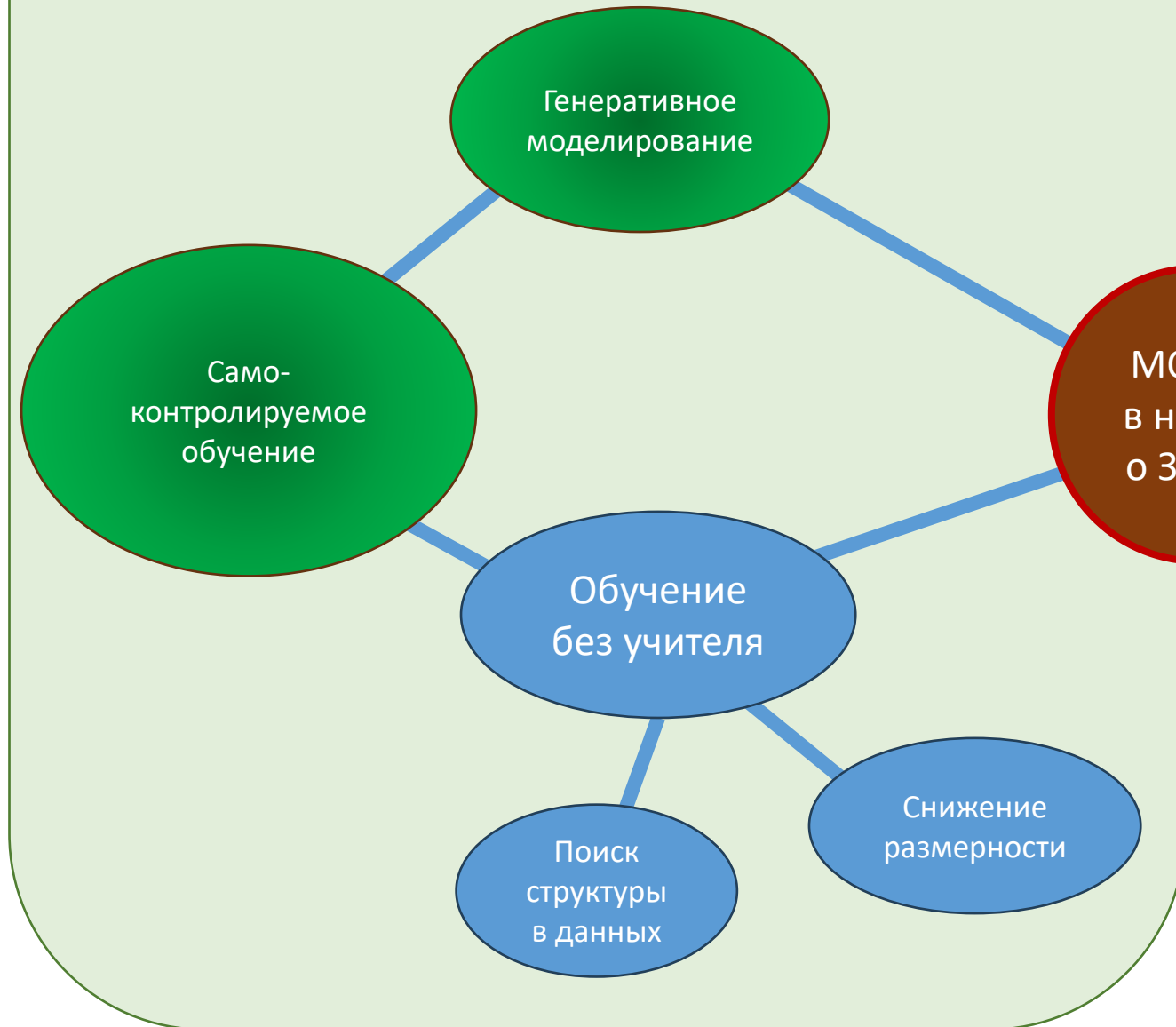


# Примеры ML/DL/AI в задачах наук о Земле



## Поисковые исследования

## Прикладные исследования



# Анализ результатов моделирования

## Прикладные

- Статистический даунскейлинг
- Краткосрочный и сверхкраткосрочный нейросетевой прогноз (погоды, погоды в океане, ледовой обстановки, рядов натурных измерений)

## Поисковые

(Поиск структуры в данных)

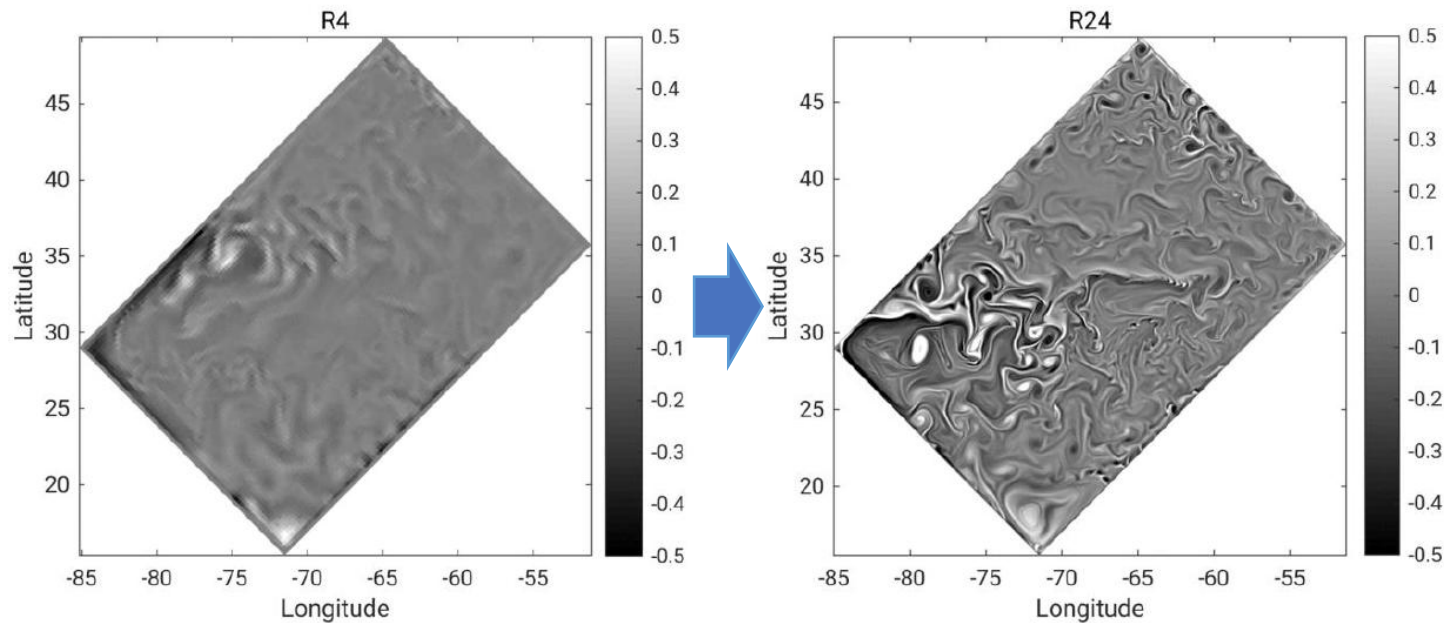
- Выявление экстремальных и аномальных событий, объектов
- Оценка тенденций их возникновения

(Генеративные модели)

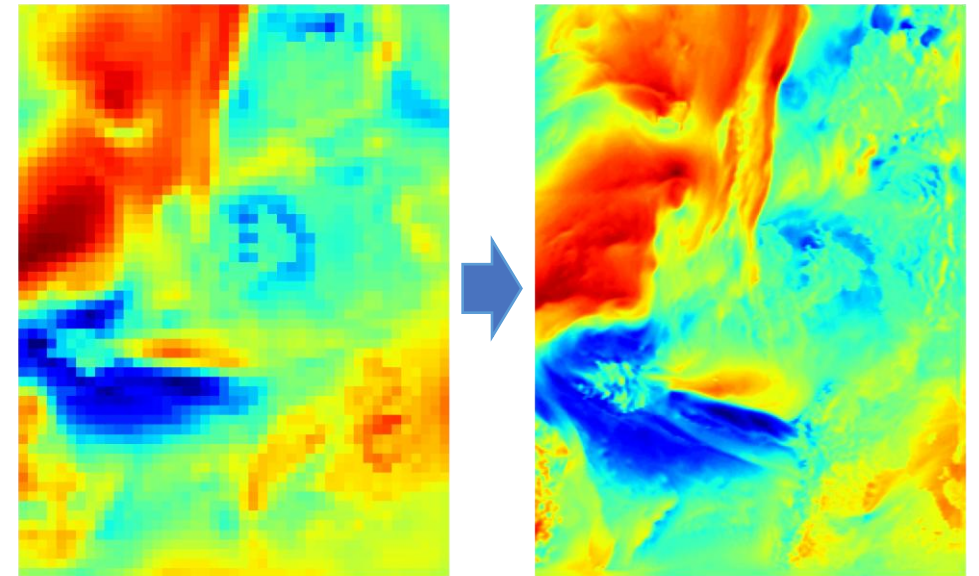
- Нейросетевая оценка качества гидродинамического моделирования

# Анализ результатов моделирования

- Статистический даунскейлинг



Масштабирование динамики течений

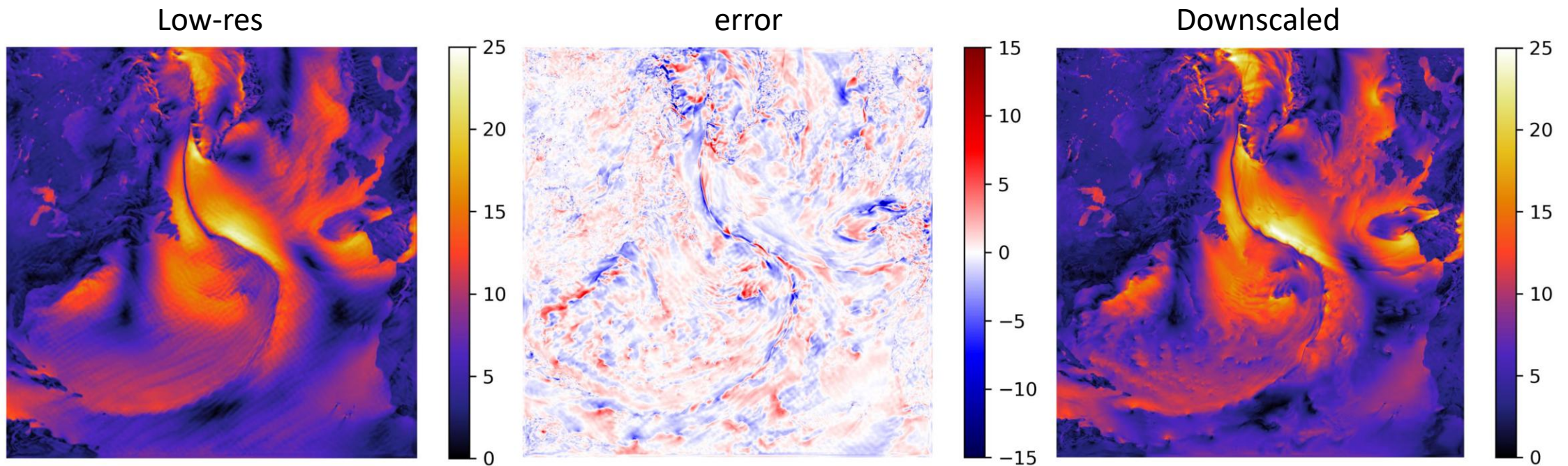


Масштабирование скорости ветра



# Анализ результатов моделирования

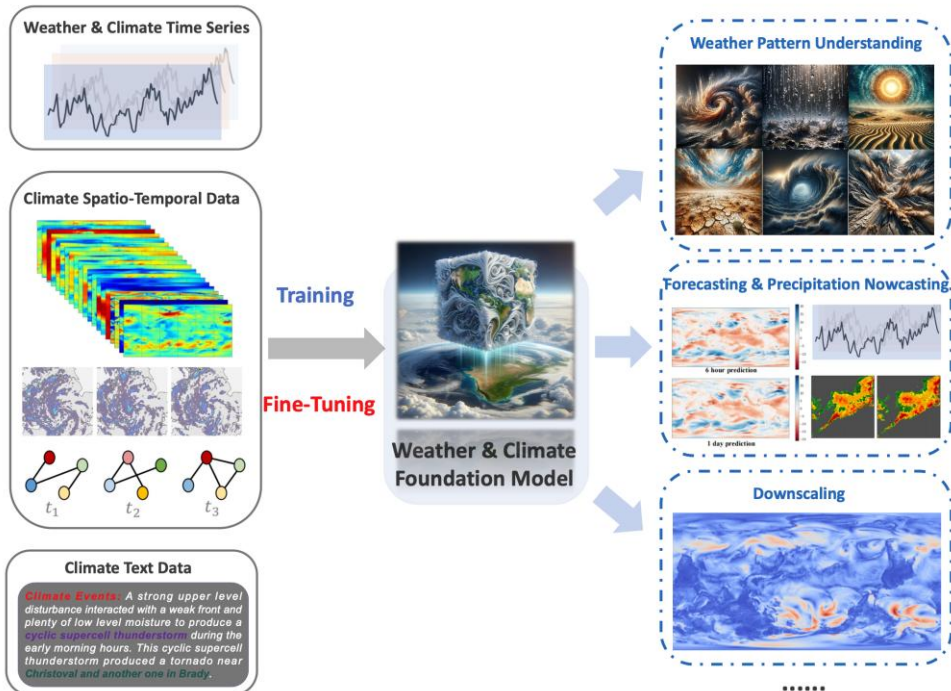
- Статистический даунскейлинг



Масштабирование скорости ветра

# Анализ результатов моделирования

- Статистический даунскейлинг с использованием фундаментальных нейросетевых моделей



- |   |          |
|---|----------|
| 1. Microsoft ClimaX <sup>1,*</sup> – Jan'2023, UCLA (USA)   | 1.40625° |
| 2. FengWu <sup>2,*</sup> – Apr'2023, China (6 организаций)  | 0.25°    |
| 3. PanGu <sup>3,*</sup> – July'2023, Huawei (China)         | 0.25°    |
| 4. FuXi <sup>4,*</sup> – Jun'2023, Fudan University (China) | 0.25°    |
| 5. FourCastNet <sup>5,*</sup> – Feb'2022, NVIDIA            | 0.25°    |
| 6. GraphCast <sup>6,*</sup> – Nov'2023, Google              | 0.25°    |
| 7. W-MAE <sup>7,*</sup> – Apr'2023, UEST (China)            | 0.25°    |



## Поисковые исследования

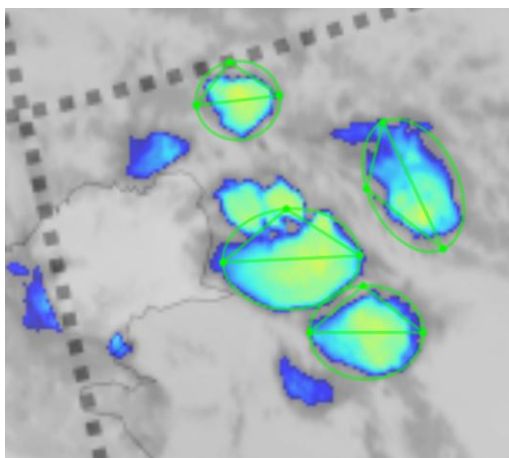
## Прикладные исследования



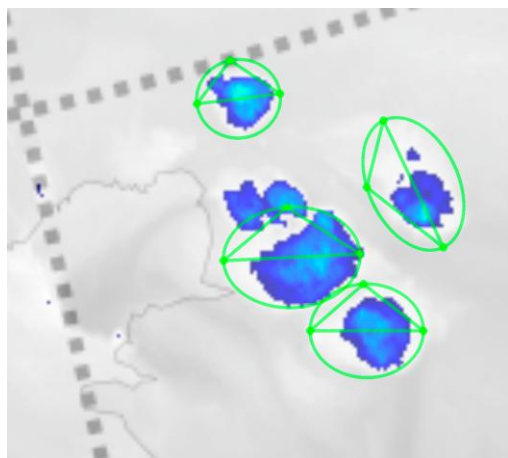
# Контролируемое обучение: пример

Идентификация мезомасштабных конвективных систем

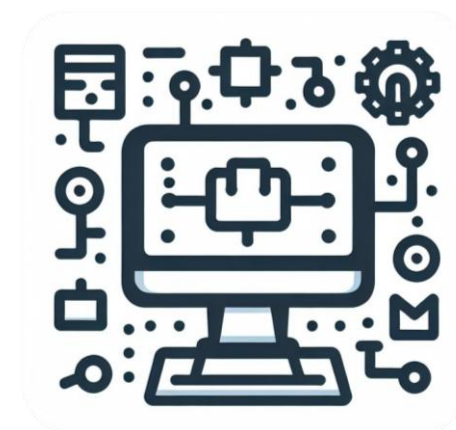
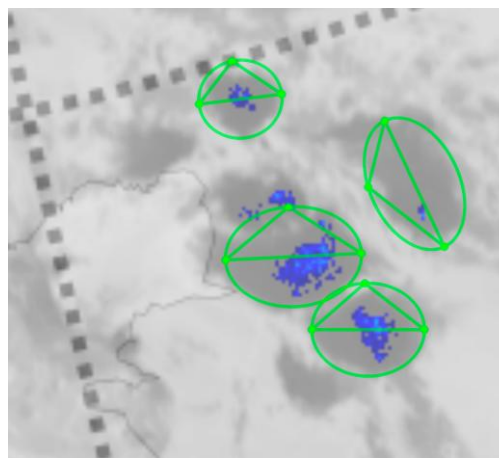
$ch9$  ( $10.8 \mu m$ )



$ch5$  ( $6.25 \mu m$ )



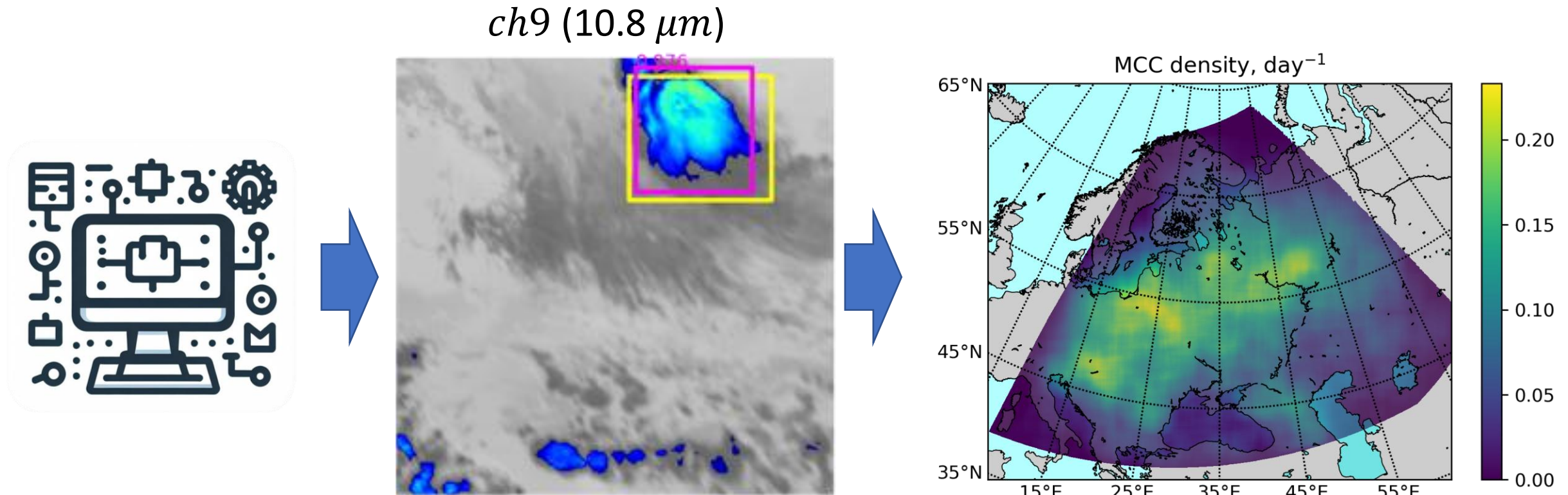
$ch5 - ch9$



Данные ДЗЗ (Meteosat, MSG4),  
Европейская территория России

# Контролируемое обучение: пример

## Идентификация мезомасштабных конвективных систем



## Поисковые исследования



## Прикладные исследования



# Анализ и интерпретация данных наблюдений

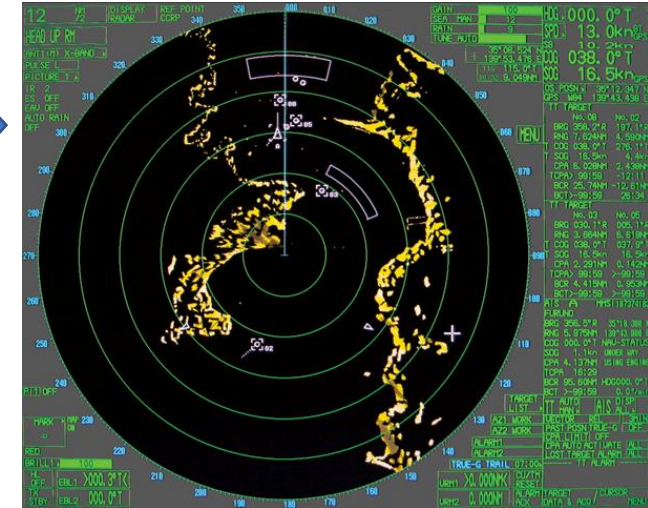
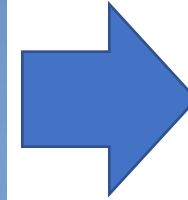
## Прикладные

- Определение характеристик ветрового волнения по данным судового радара
- Определение характеристик облачности по данным визуальной съемки
- Прогноз характеристик возвратной миграции нерки
- Определение реактоспособности субстрата в активных центрах гидролаз для определения эффективности активации различных систем: комплексов бактериальной металло- $\beta$ -лактамазы NDM-1 и L1 с антибиотиком имипенема, а также комплексов капралактама и капралактона с липазой CALB



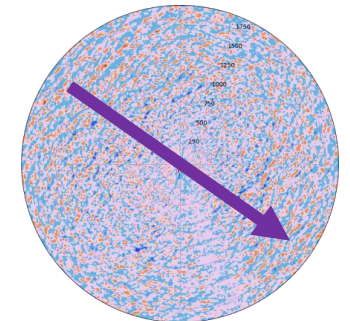
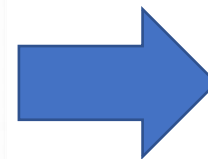
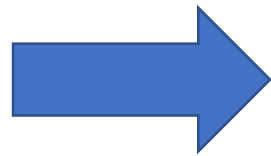
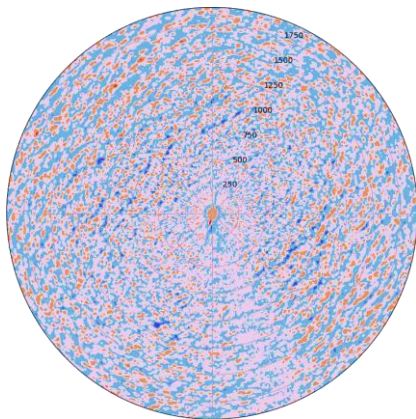
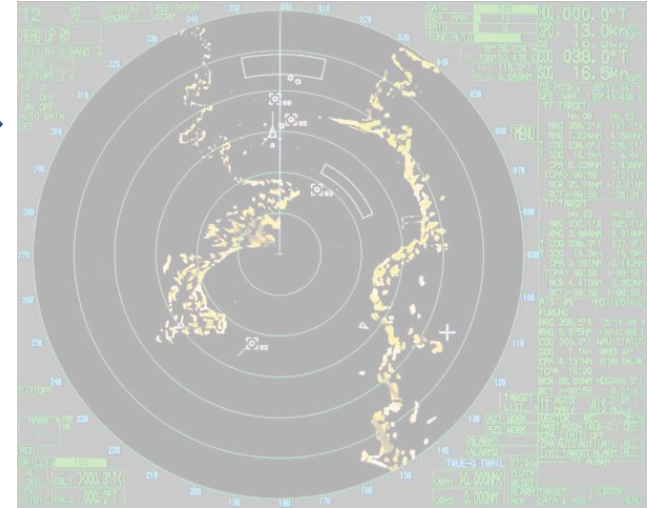
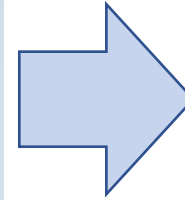
# Анализ и интерпретация данных наблюдений

## Характеристики ветрового волнения по данным навигационного радара



# Анализ и интерпретация данных наблюдений

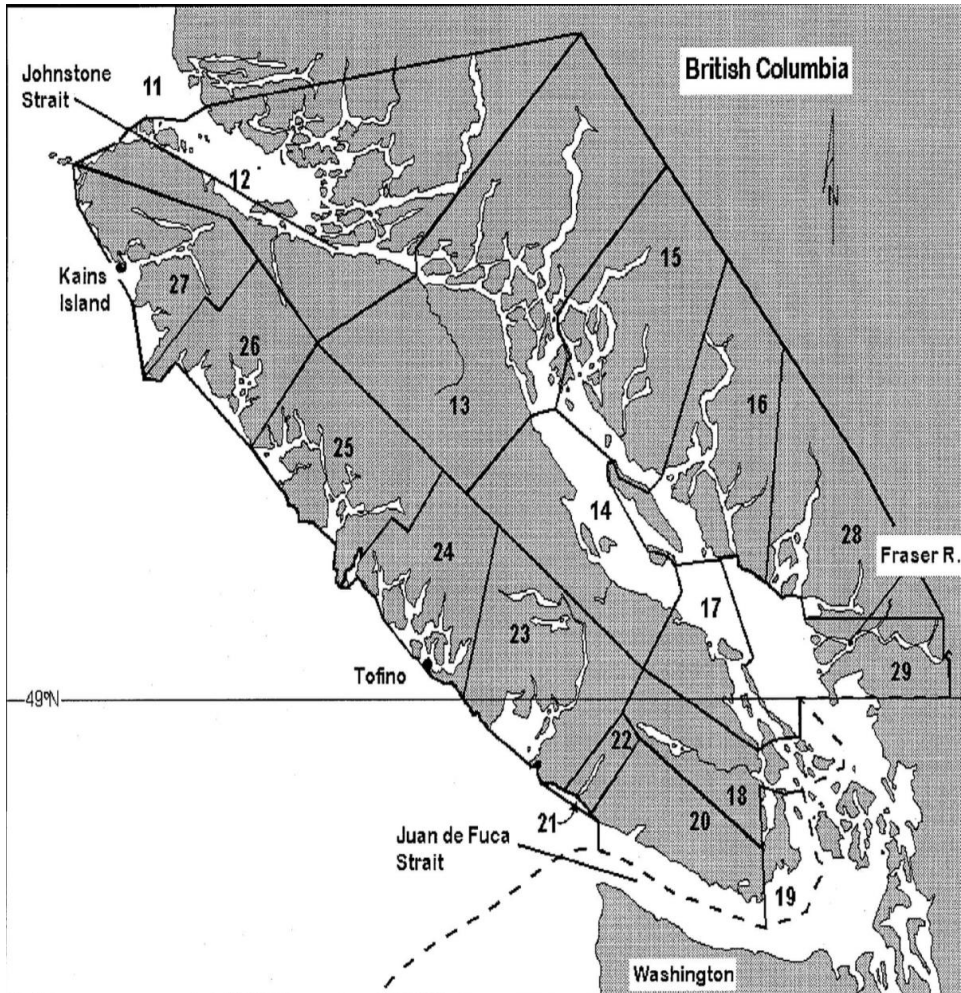
Характеристики ветрового волнения по данным навигационного радара





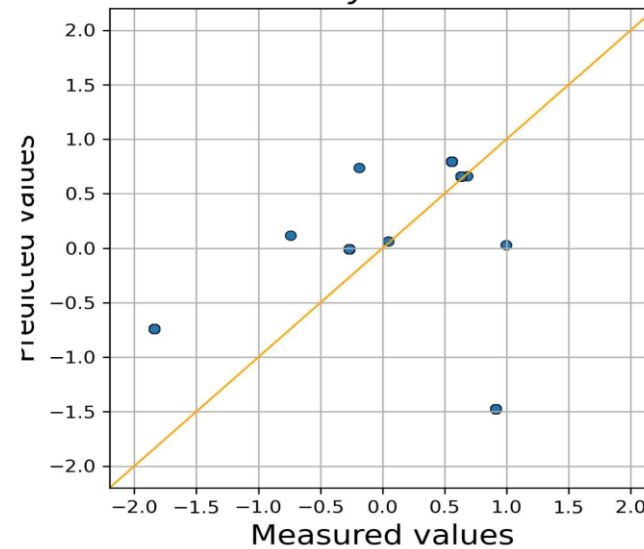
# Прогноз рядов измерений

Прогноз характеристик возвратной миграции нерки в устье р. Фрейзер



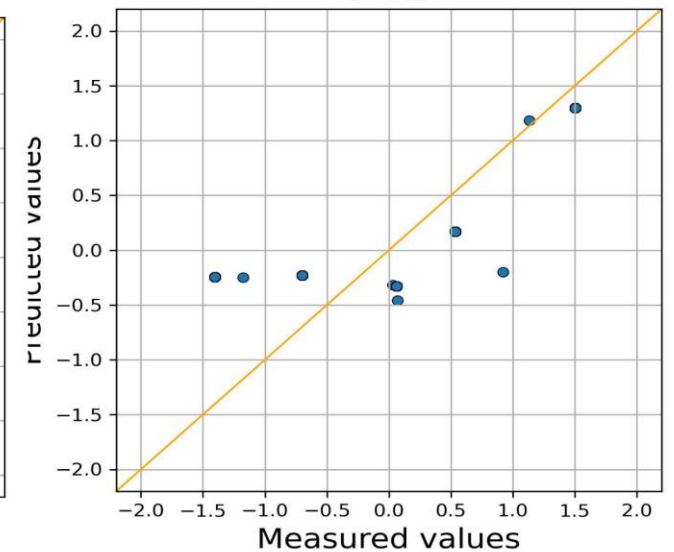
дата медианного возврата

*Early Stuart*



доля северного отклонения

*NDR*





# Перспективные типы задач: прикладные

- Анализ данных дистанционного зондирования (спутникового базирования, судового, БПЛА) - решение прикладных задач в постановке контролируемого обучения
- Анализ данных полевых измерений
  - заполнение пропусков во временных рядах, в пространственно распределенных данных
  - восстановление климатических рядов по косвенным измерениям
- Статистическое масштабирование геофизических полей
- Статистический прогноз (краткосрочный: погоды, ледовой обстановки, рядов измерений)

# Перспективные типы задач: поисковые

- **Выявление паттернов** (динамических, пространственных) в климатических и погодных данных
- **Выявление аномалий** в климатических, погодных данных
- **Предварительное самоконтролируемое обучение** нейросетевых моделей для последующего решения широкого круга прикладных задач
- (Обучение?) исследование свойств **фундаментальных** климатических, погодных **моделей**
- **Усвоение данных** в моделировании атмосферы, океана
- **Выучивание** нейросетевой **меры качества** воспроизведения динамики климата, динамики атмосферы, океана
- **Идентификация ДУЧП** климатических моделей с использованием нейросетей
- **Внедрение физических ограничений** в нейросетевое моделирование климатических процессов в атмосфере, океане

# Неконтролируемое обучение

## Выявление структуры состояний стратосферного полярного вихря

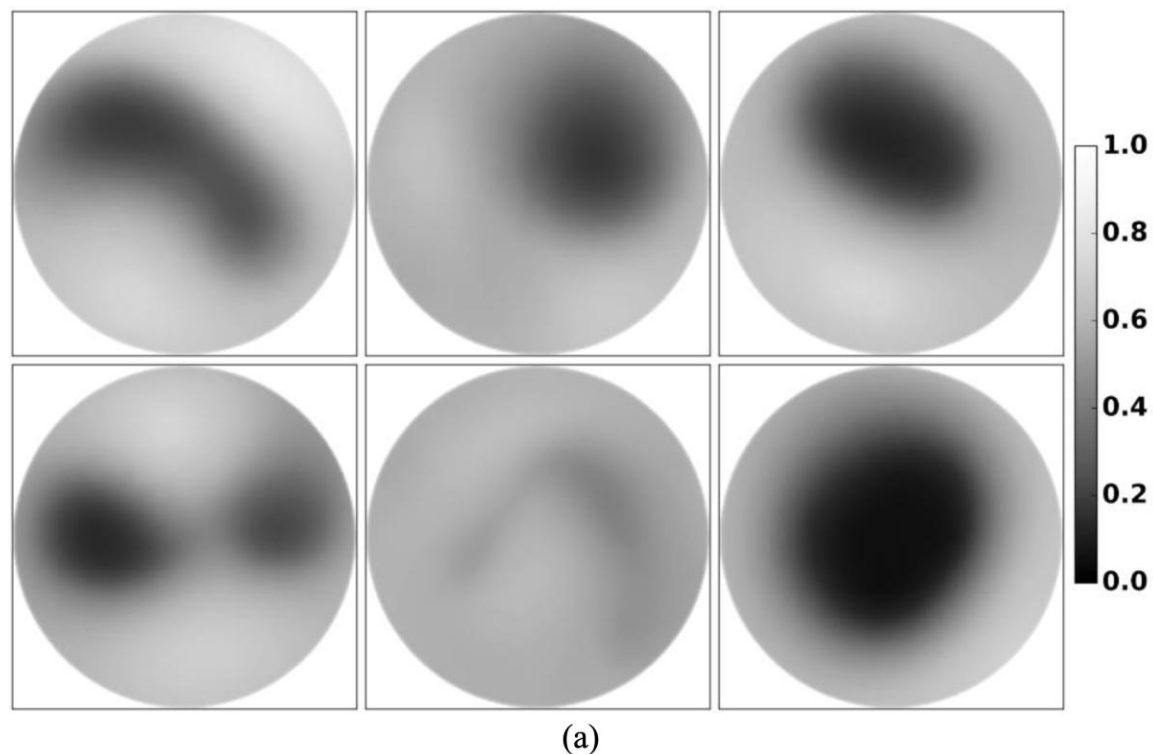
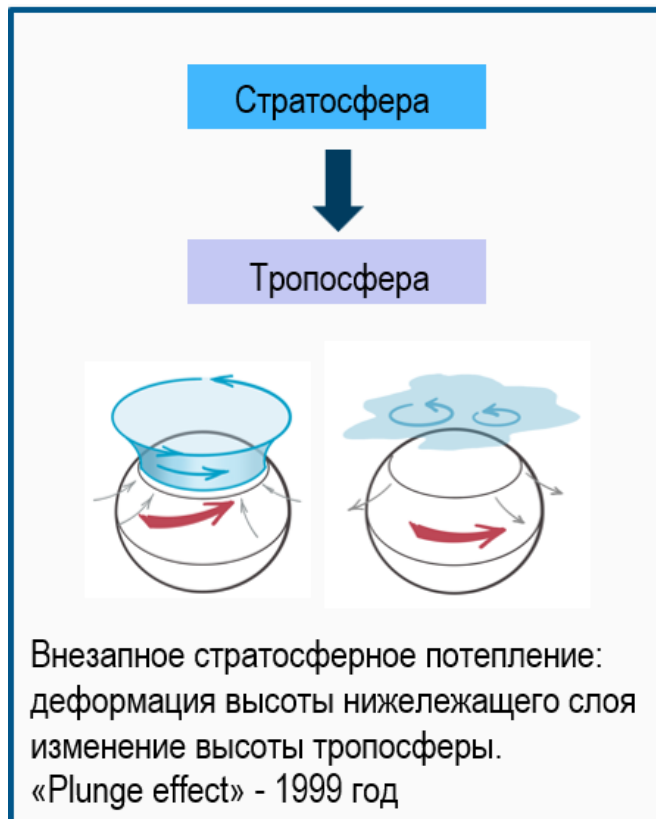
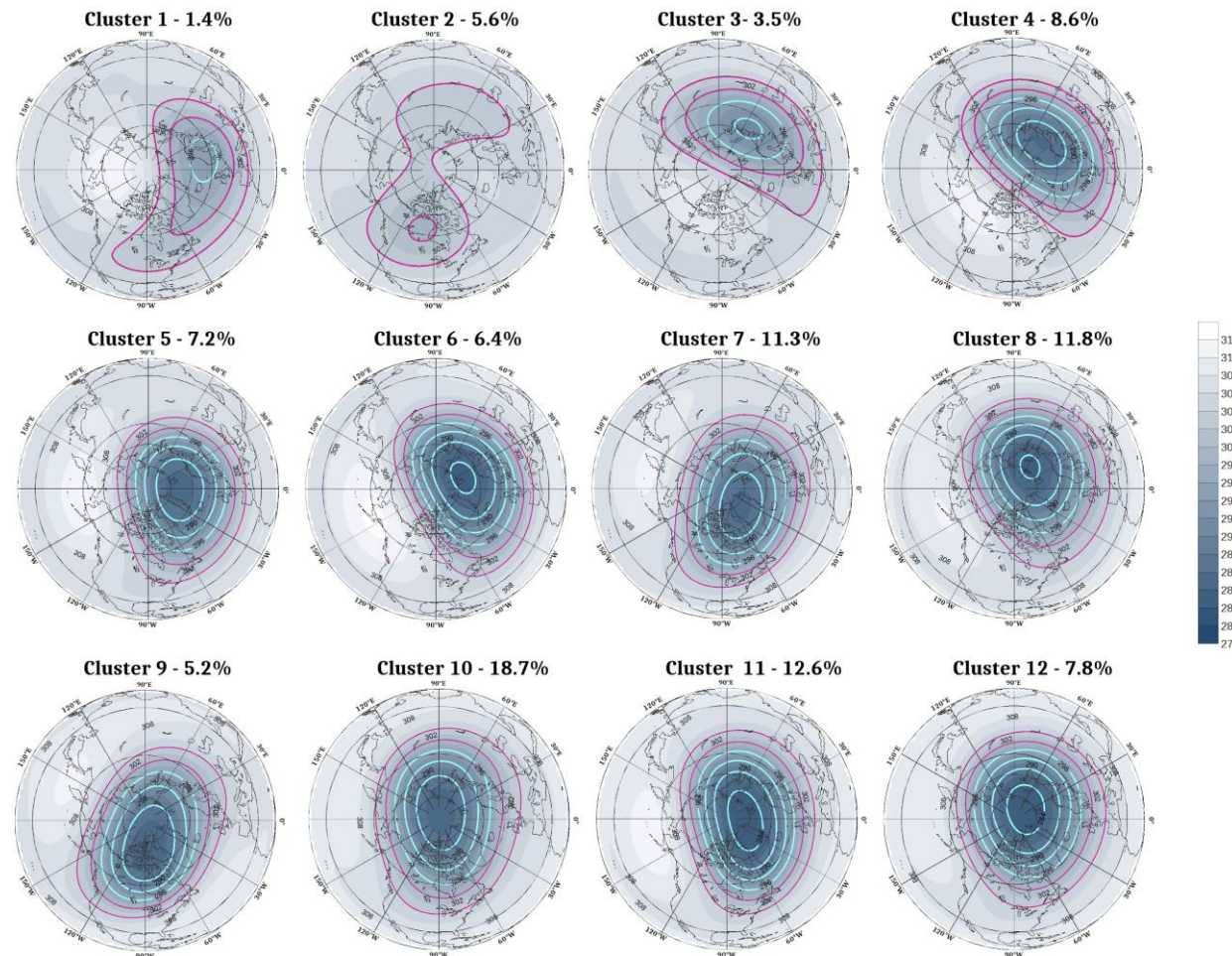


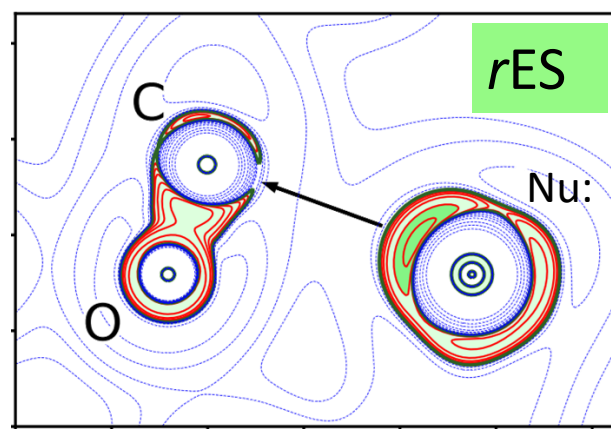
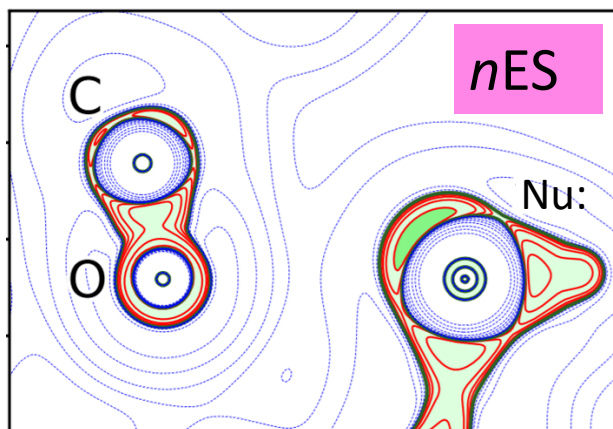
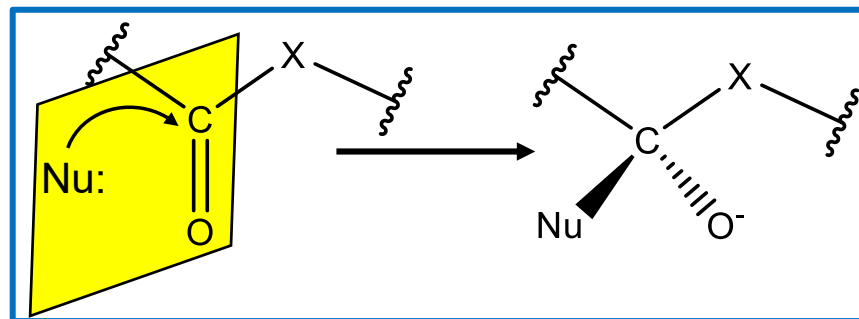
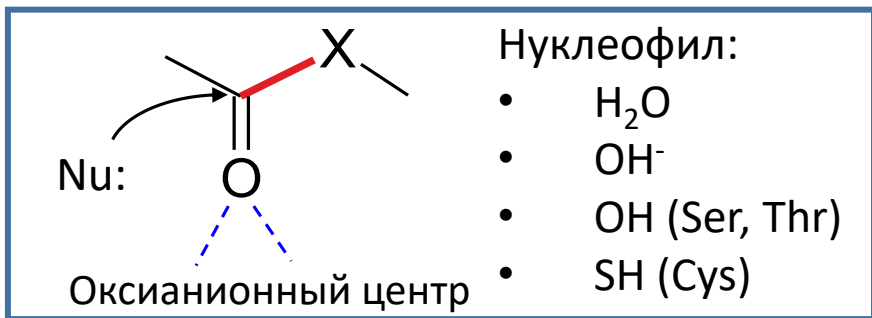
Figure 1: (a) Examples from the dataset of PV states (HGT values only, normalized)

# Неконтролируемое обучение

## Выявление структуры состояний стратосферного полярного вихря



# Активация субстрата



$\nabla^2 \rho > 0$  – область деконцентрации ЭП  
 $\nabla^2 \rho < 0$  – область концентрации ЭП



## Цель работы:

Разработка свёрточной нейронной сети, проводящей бинарную классификацию наличия активации субстрата в активных центрах гидролаз, и ее применение для определения эффективности активации различных систем: комплексов бактериальной металло-β-лактамазы NDM-1 и L1 с антибиотиком имипенема, а также комплексов капралактама и капралактона с липазой CALB.

# Контролируемое обучение

Реактоспособность субстрата по картам лапласиана эл. плотности

