

Машинное обучение для детализированного моделирования погоды и климата в мегаполисах

Варенцов М.И.^{1-4*}, Криницкий М.А.^{5,6,1},
Коспанов А.Р.^{2,3}, Степаненко В.М.^{1,2}

(1) МГУ имени М.В. Ломоносова, Научно-исследовательский вычислительный центр

(2) МГУ имени М.В. Ломоносова, Географический факультет

(3) Гидрометцентр России

(4) Институт физики атмосферы им. А.М. Обухова РАН

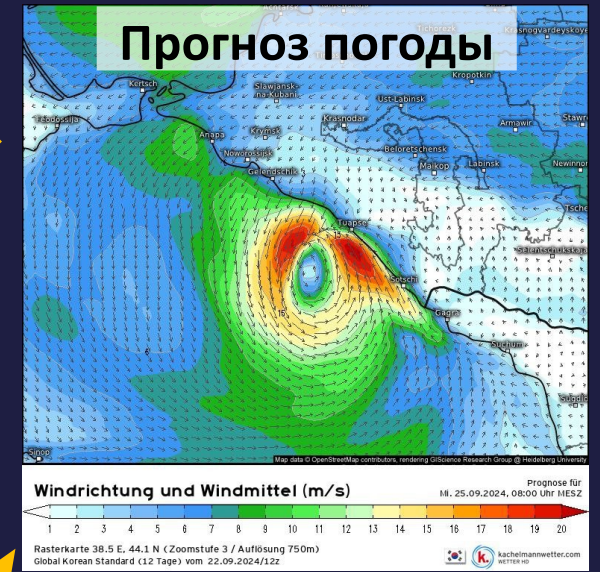
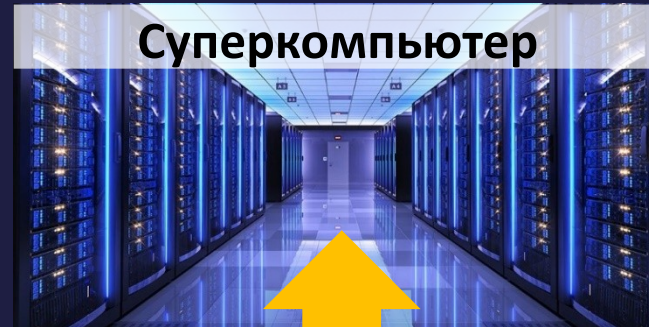
(5) Институт океанологии им. П.П. Ширшова РАН

(6) Московский физико-технический институт

*mikhail.varentsov@srcc.msu.ru



Моделирование погоды и климата



Conservation of momentum:

$$\frac{\partial \vec{V}}{\partial t} = -(\vec{V} \cdot \nabla) \vec{V} - \frac{1}{\rho} \nabla p - \vec{g} - 2\vec{\Omega} \times \vec{V} + \nabla \cdot (k \nabla \vec{V}) - \vec{F}_d$$

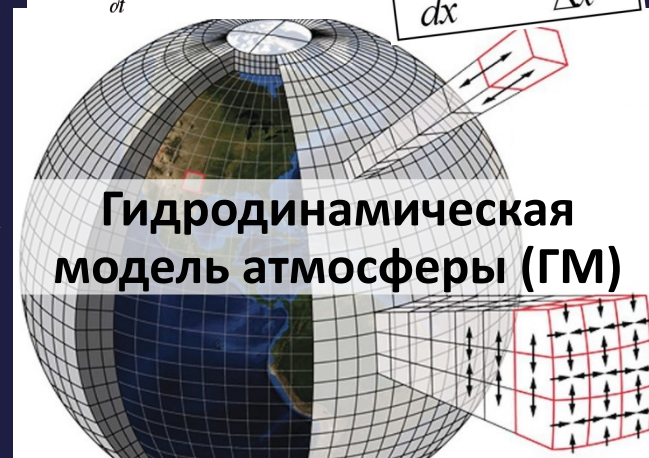
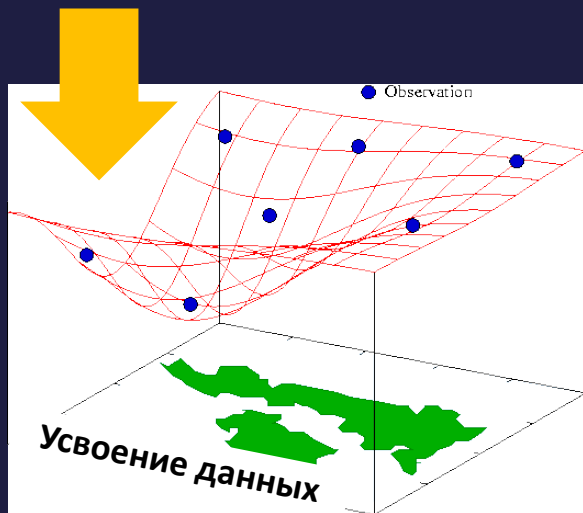
Conservation of energy:

$$\rho c_p \frac{\partial T}{\partial t} = -\rho c_p (\vec{V} \cdot \nabla) T - \nabla \cdot \vec{R} + \nabla \cdot (k \nabla T) + C + S$$

Conservation of mass:

$$\frac{\partial \rho}{\partial t} = -(\vec{V} \cdot \nabla) \rho - \rho (\nabla \cdot \vec{V})$$

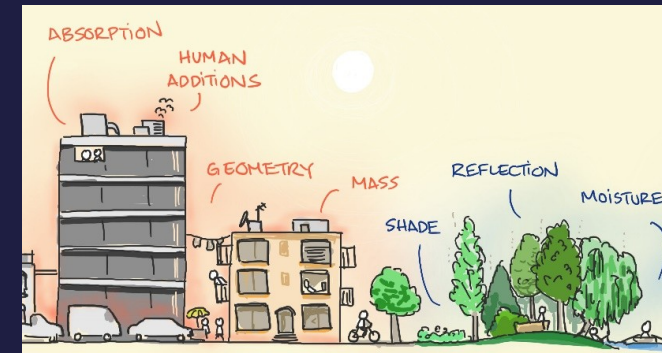
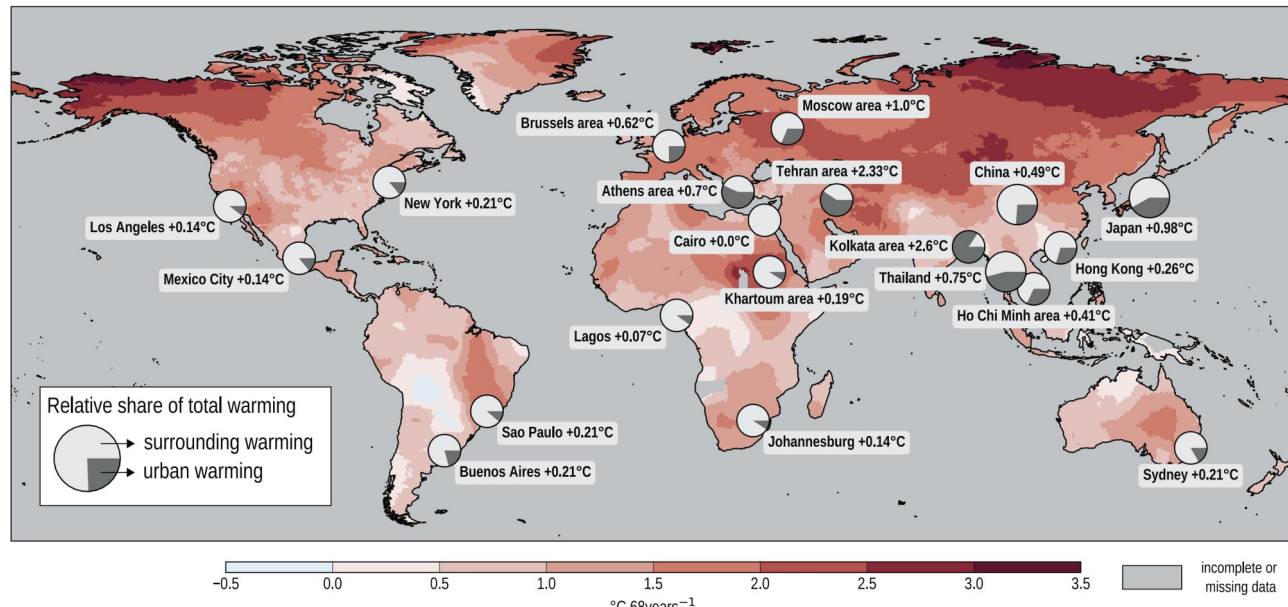
$$\frac{df_{1.5}}{dx} \approx \frac{f_2 - f_1}{\Delta x}$$



Моделирование погоды и климата для урбанизированных территорий

- Города особо уязвимы к погодно-климатическим экстремумам
- Города сами по себе влияют на атмосферные процессы и местный климат
- Города являются «горячими точками» глобального потепления (IPCC AR6, 2021)

(a) Trend in global surface air temperature (CRU TS, 1950-2018)

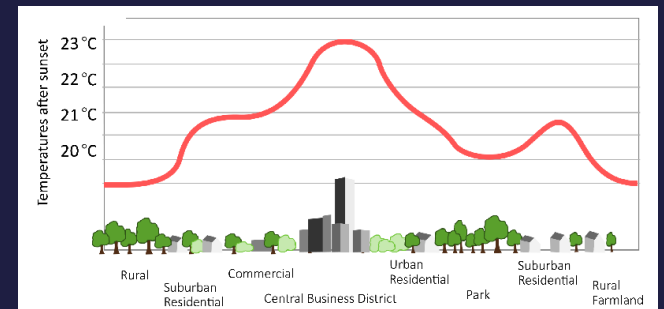


Особенности подстилающей

поверхности



Эффект городского острова тепла



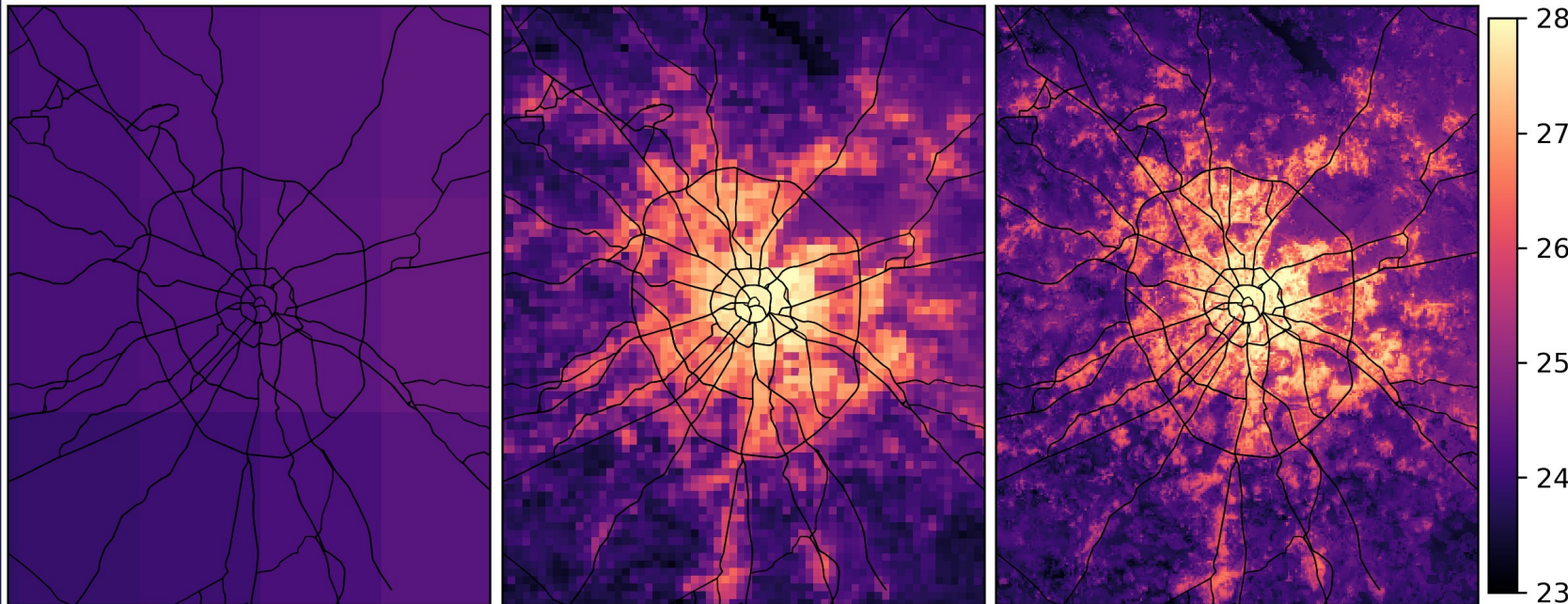
Влияние на осадки и опасные явления погоды



Влияние на термический комфорт населения

Моделирование погоды и климата для урбанизированных территорий

Температура в Москве во время волны жары в июне 2021 г. (Varentsov et al., 2023)

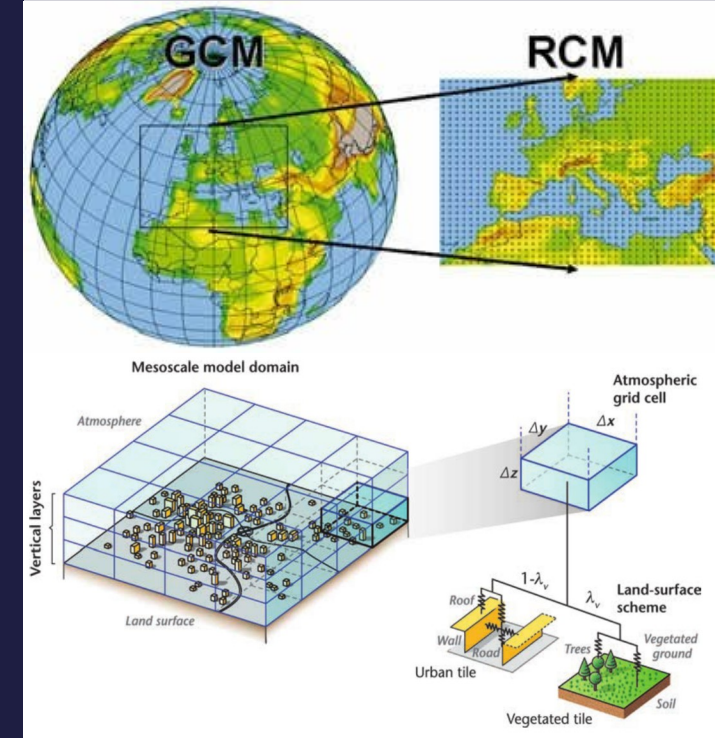


Результаты глобального моделирования (реанализ ERA5, шаг сетки ≈ 30 км)

Региональное моделирование с шагом сетки **1 км**
 ≈ 100 процессоро-часов на 1 модельный день

Региональное моделирование с шагом сетки **250 м**
 ≈ 1000 процессоро-часов на 1 модельный день

Динамический даунскейлинг



Чем тут может помочь ИИ?

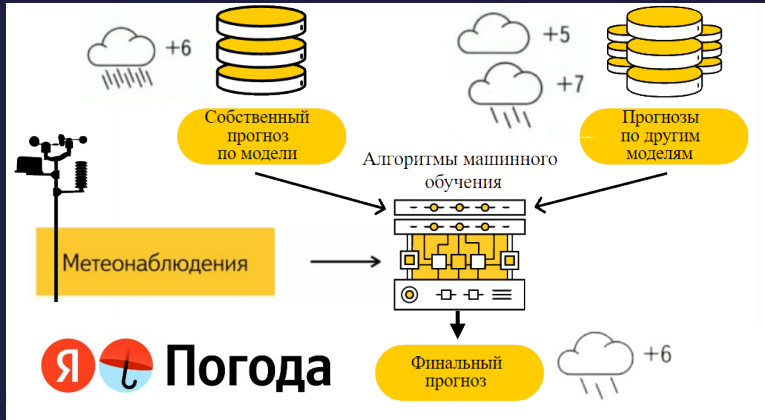


ИИ в науках о Земле



ИИ в моделировании погоды и климата

ИИ для уточнения прогнозов гидродинамических моделей



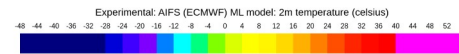
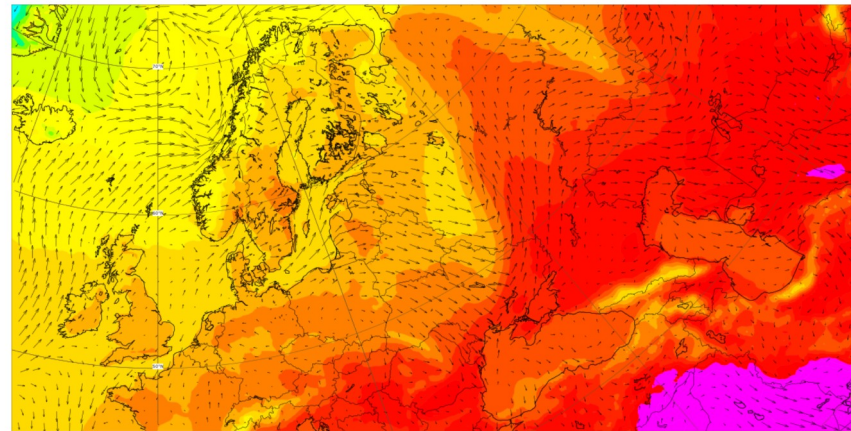
ИИ как альтернатива гидродинамическим моделям?

Графовые искусственные нейронные сети, обученные на данных ERA5 и прогноза IFS ECMWF

AIFS (ECMWF)

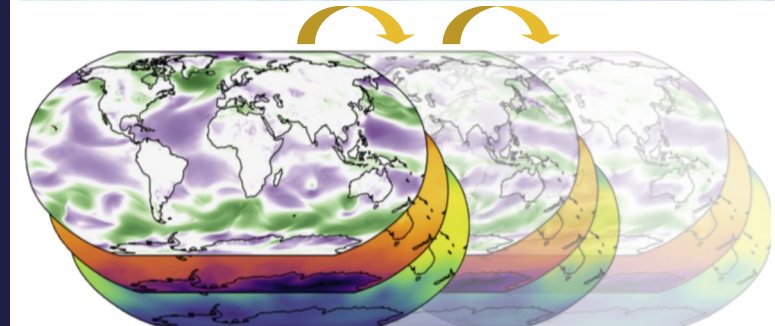
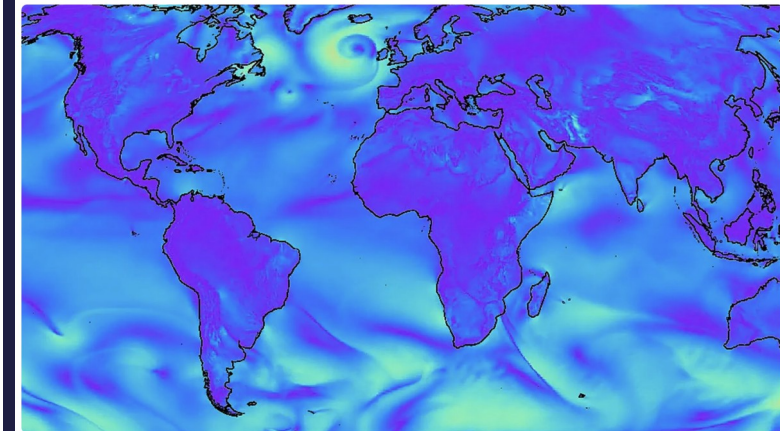
Experimental: AIFS (ECMWF) ML model: 2 m temperature and 10 m wind

Base time: Wed 19 Jun 2024 12 UTC Valid time: Thu 20 Jun 2024 12 UTC (-24h) Area : North East Europe



GraphCast: AI model for faster and more accurate global weather forecasting

GraphCast (Google)

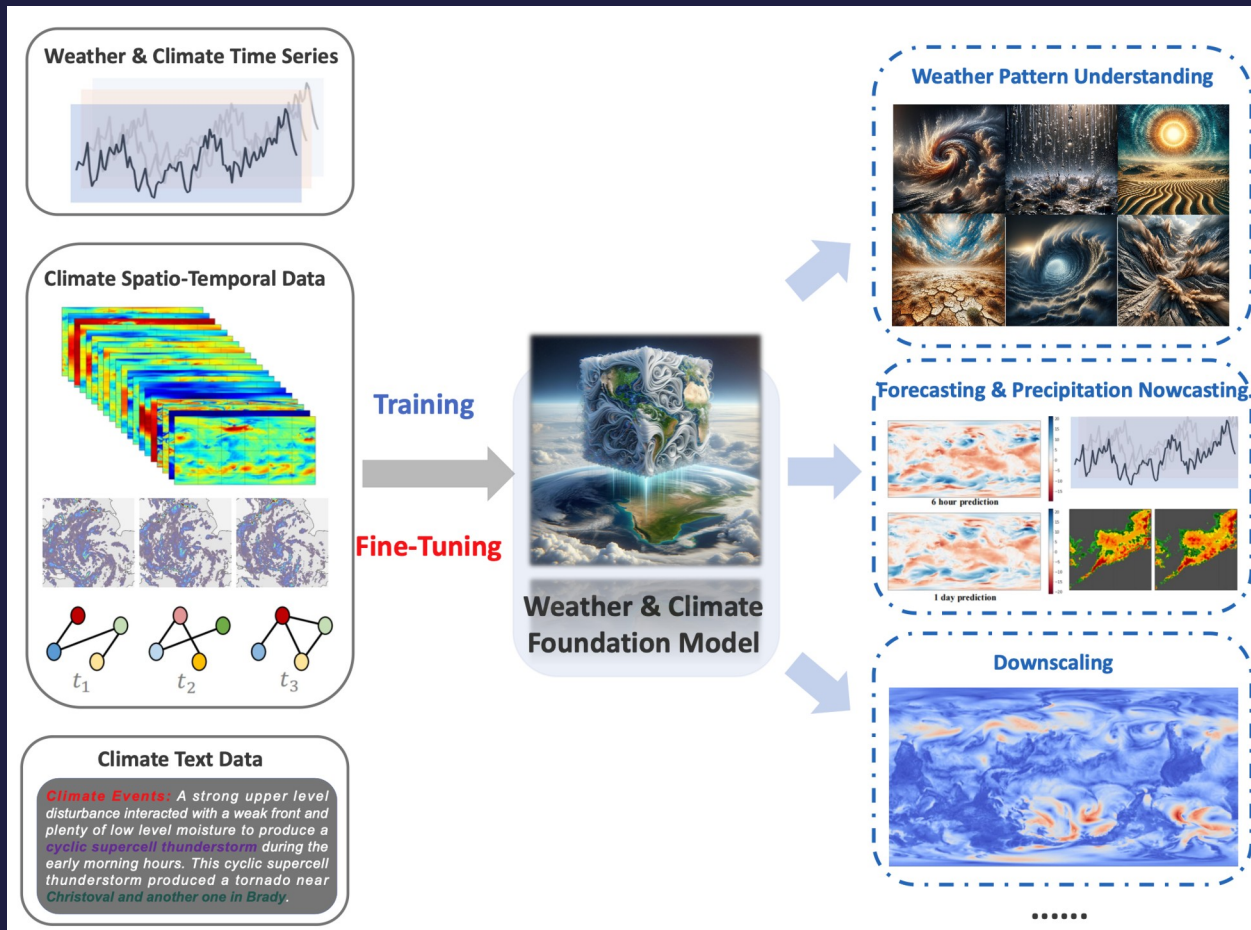


¹Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ... & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science*, eadi2336

²<https://www.ecmwf.int/en/newsletter/178/news/aifs-new-ecmwf-forecasting-system>

ИИ для моделирования погоды и климата

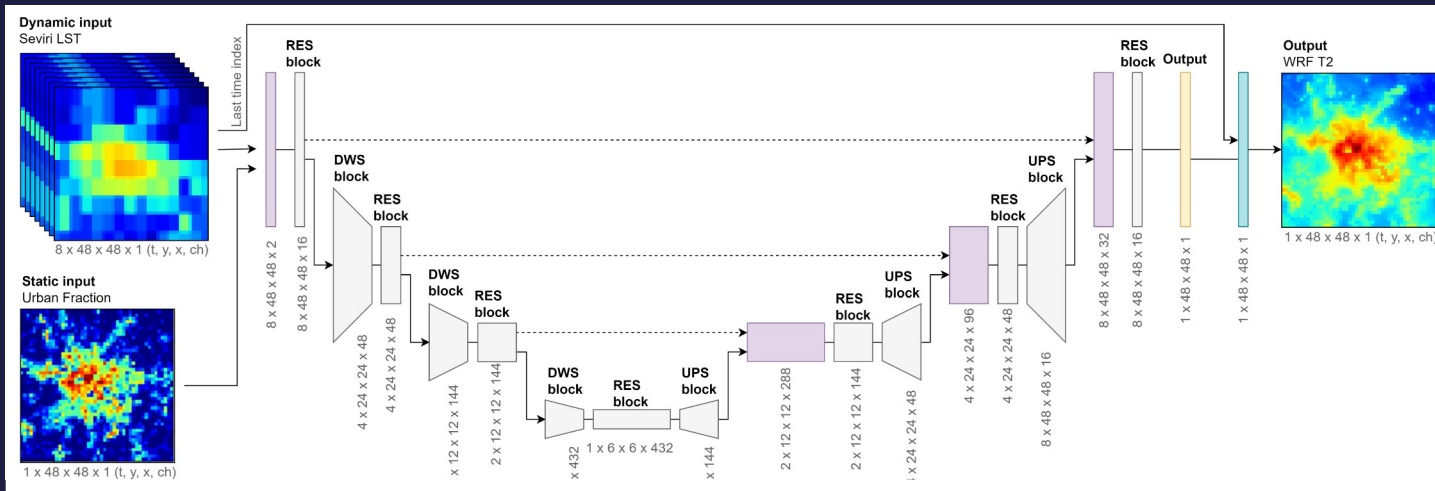
Фундаментальные модели для климатических и погодных данных



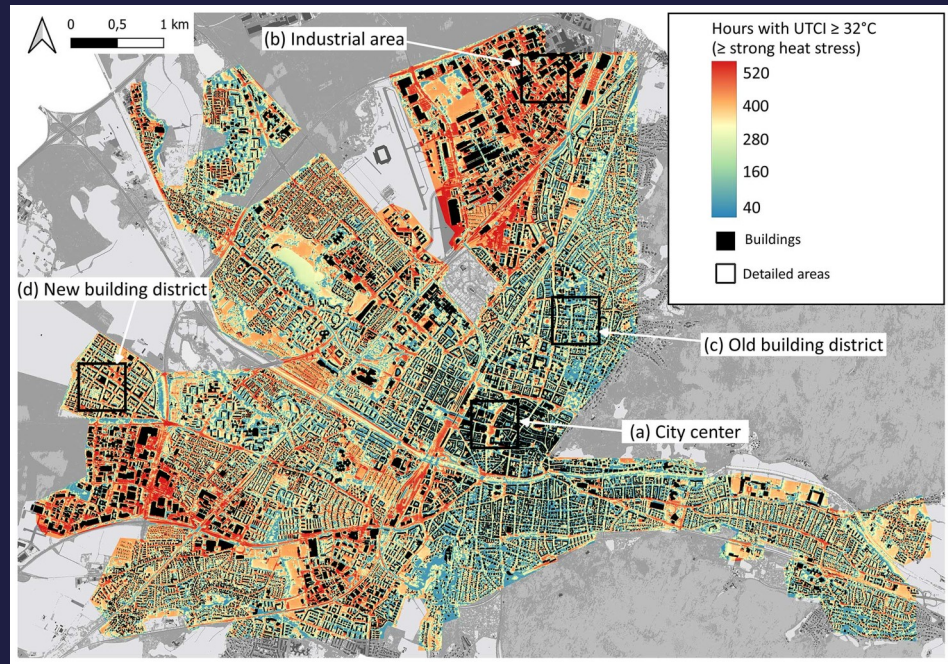
Модель	Разрешение	Архитектура
Microsoft ClimaX ^{1,*} – Jan'2023, UCLA (USA)	1.40625°	T
FengWu ^{2,*} – Apr'2023, China (6 организаций)	0.25°	T
PanGu ^{3,*} – July'2023, Huawei (China)	0.25°	T
FuXi ^{4,*} – Jun'2023, Fudan University (China)	0.25°	T
FourCastNet ^{5,*} – Feb'2022, NVIDIA	0.25°	FNO
GraphCast ^{6,*} – Nov'2023, Google	0.25°	G
W-MAE ^{7,*} – Apr'2023, UEST (China)	0.25°	T

Chen, S., Long, G., Jiang, J., Liu, D., & Zhang, C. (2023). Foundation models for weather and climate data understanding: A comprehensive survey. *arXiv preprint arXiv:2312.03014*.

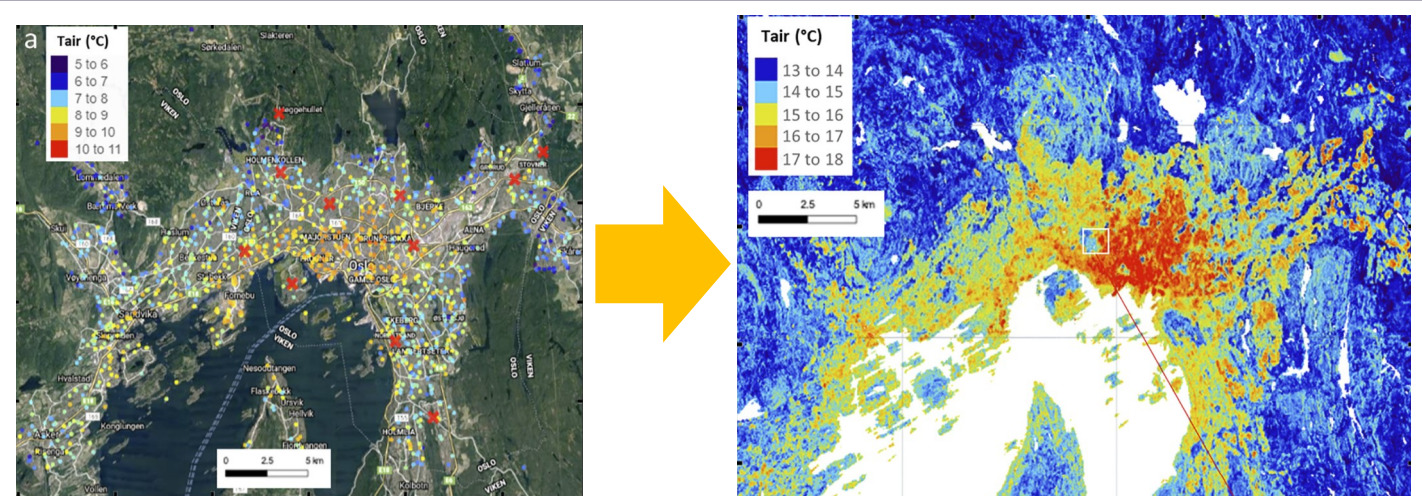
ИИ в городской метеорологии



Afshari et al. (2023) Statistical Downscaling of SEVIRI Land Surface Temperature to WRF Near-Surface Air Temperature Using a Deep Learning Model

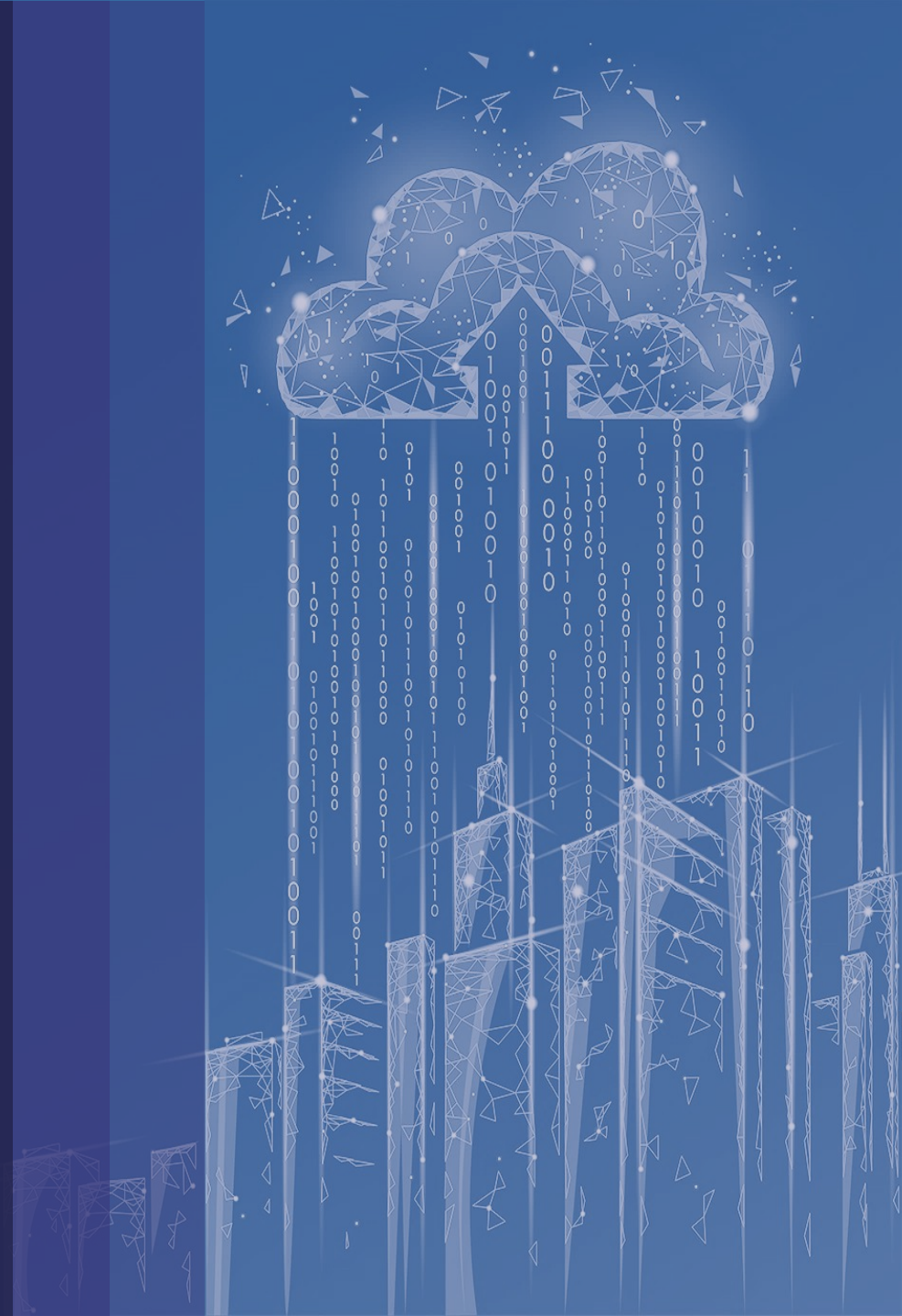


Briegel et al (2024). High-resolution multi-scaling of outdoor human thermal comfort and its intra-urban variability based on machine learning

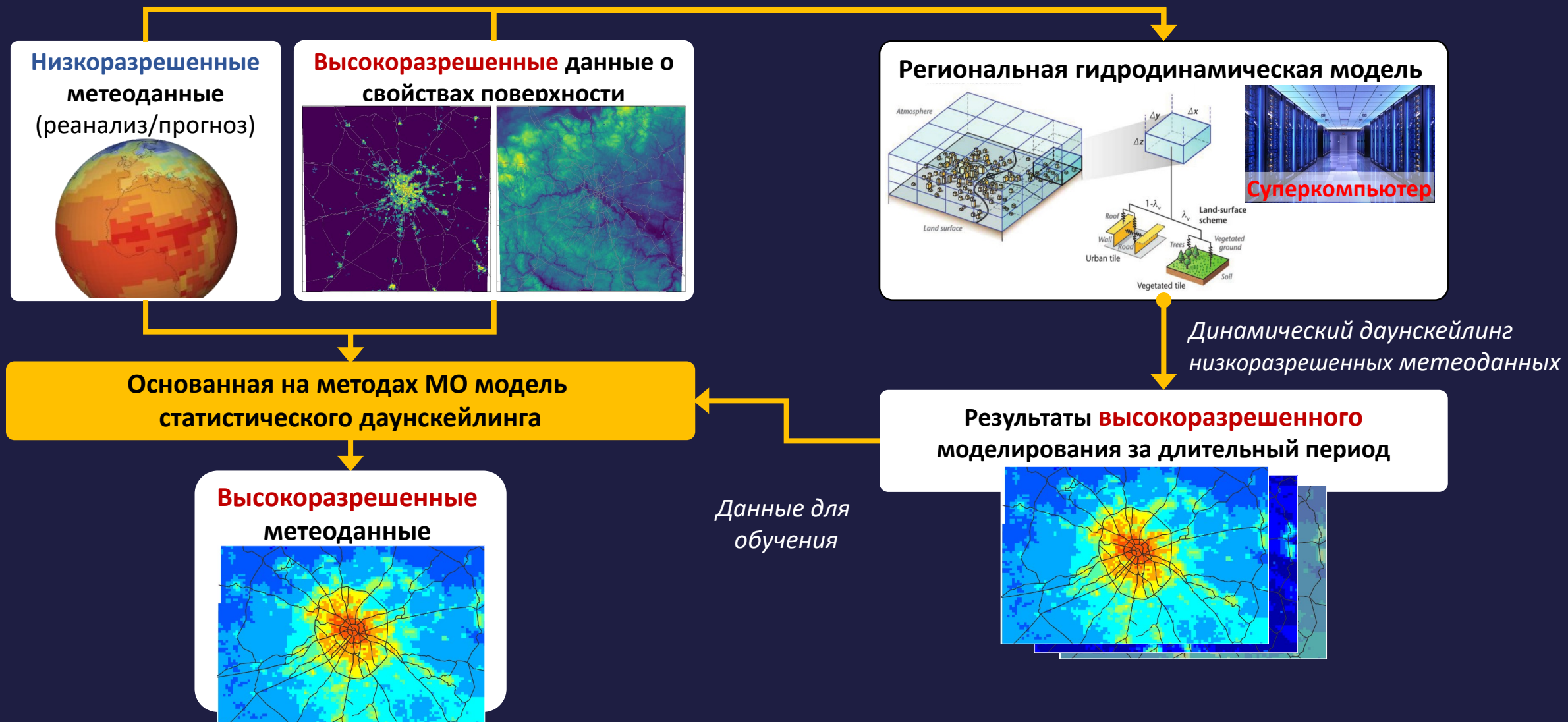


Venter et al. (2020) Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data

Наши исследования



Общая идея исследования



Локальная постановка задачи

Даунскейлинг «в точку»: аппроксимация интенсивности острова тепла в центре Москвы

❑ **Целевая переменная:**
разность температуры между центром Москвы (Балчуг) и средней температурой по 9 загородным метеостанциям

❑ **Предикторы:**
средние по Московскому региону характеристики метеорологического режима (реанализ ERA5 и/или наблюдения на загородных метеостанциях)

❑ **Формализация задачи МО:**

$$\Delta = \left(\underset{\substack{\text{↑} \\ \text{Моментальное} \\ \text{значении предикторов}}}{1}, \dots, 1 \left(1 \mid \dots - \right), 2(\dots), \dots, \dots \right)$$

Значения за предшествующих моментов времени

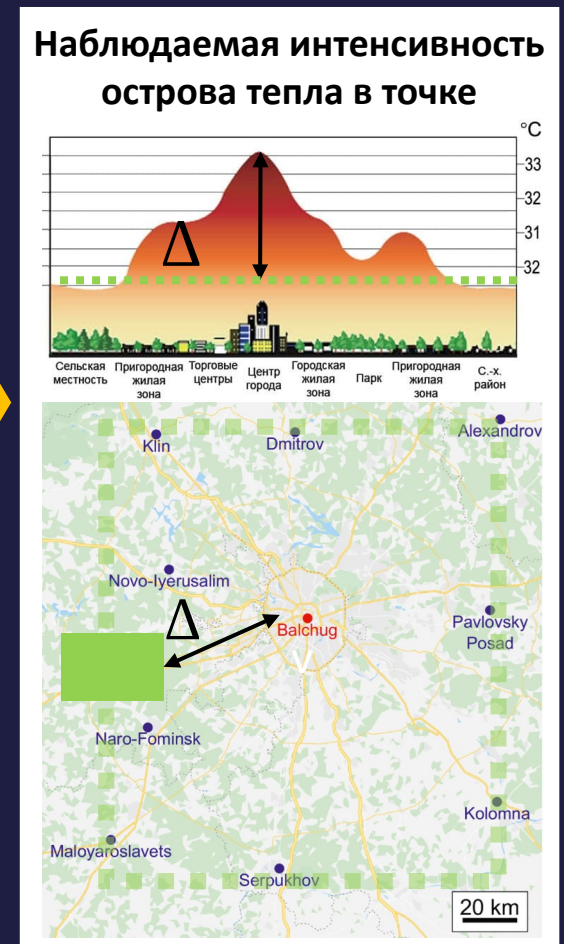
❑ **Данные для обучения:**

- Временные ряды с дискретность 3 ч за 2001-2021 гг.
- 5-блочная кросс-валидация с отношением тренировочной и тестовой выборки 4:1

Низкоразрешенные метеоданные (реанализ и/или наблюдения)

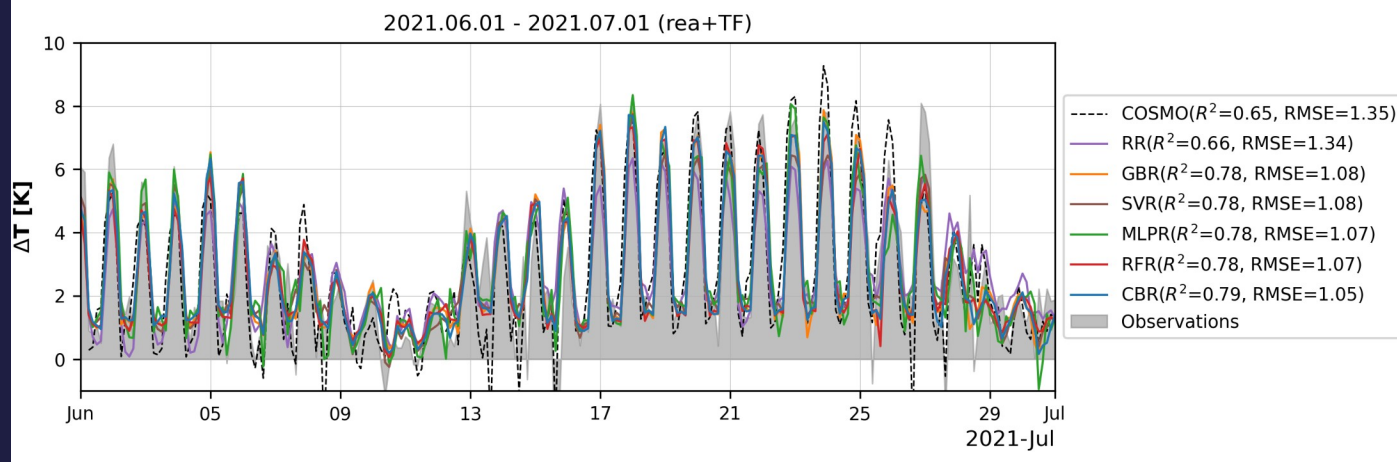
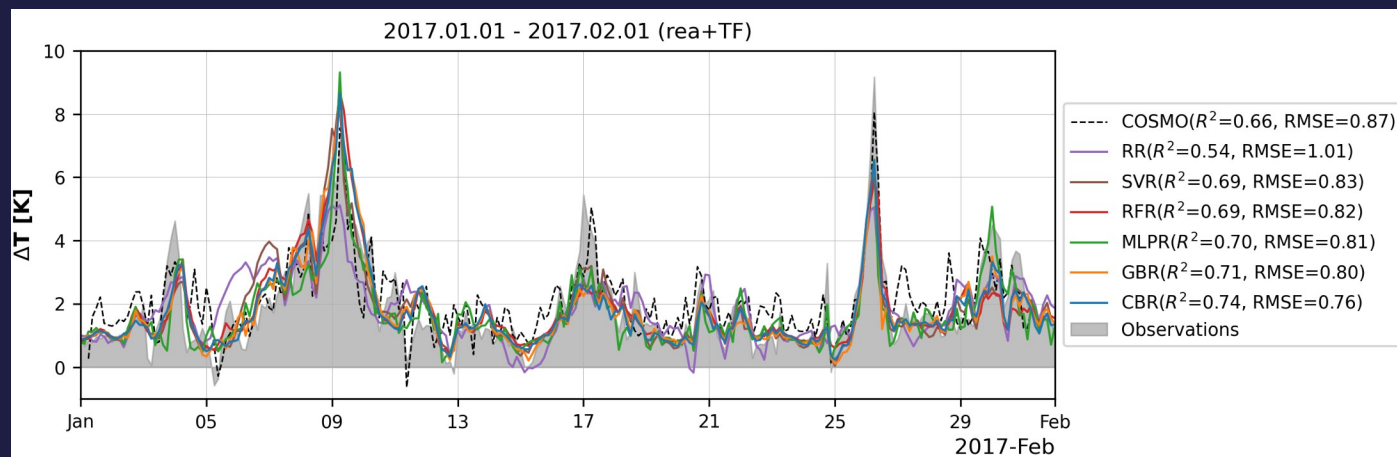
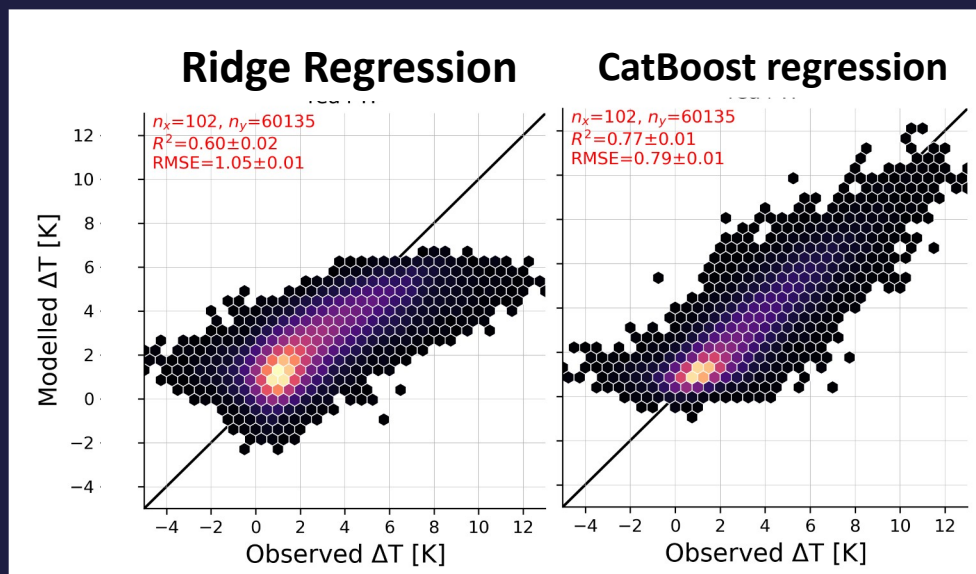
Метеорологические предикторы	
t2m	Температура воздуха на 2 м, °C
rh2m	Отн. влажность на 2 м, %
vel10m	Скорость ветра на 10 м, м/с
tcc	Доля общей облачности
lcc	Доля нижней облачности
WF	Фактор погоды, эмпирическая функция скорости ветра и облачности [Oke, 1998]
sp	Атмосферное давление, гПа
blh	Высота погранслоя, м
str	Длинноволновый баланс, Вт/м ²
ssr	Коротковолновый баланс, Вт/м ²
strd	Приходящая коротковолновая радиация, Вт/м ²
ssrd	Приходящая длинноволновая радиация, Вт/м ²
tp	Сумма осадков за 3 ч, мм

Наблюдения и реанализ (для t2m, rh2m, vel10m, tcc, lcc, WF)
Только реанализ (для sp, blh, str, ssr, strd, ssrd, tp)



Локальная постановка задачи

Даунскейлинг «в точку»: аппроксимация интенсивности острова тепла в центре Москвы



Open Access Article

Machine Learning for Simulation of Urban Heat Island Dynamics Based on Large-Scale Meteorological Conditions

Climate 2023, 11(10), 200;

<https://doi.org/10.3390/cli11100200>



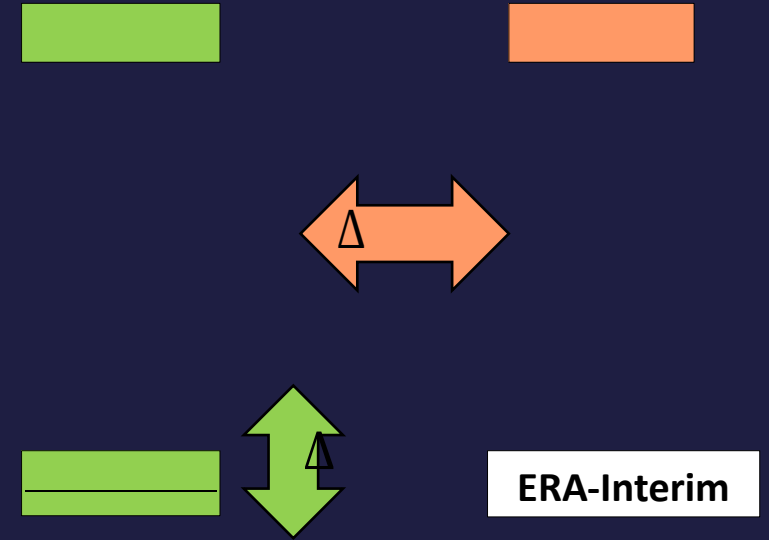
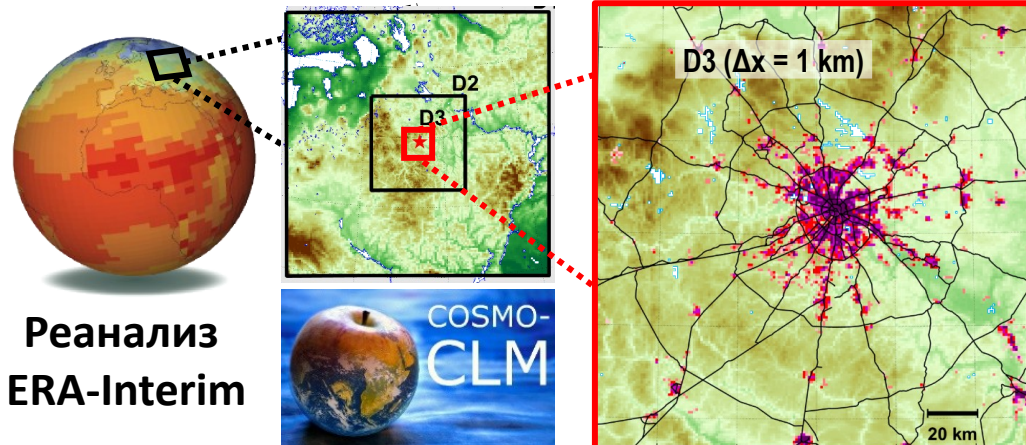
Нелокальная постановка задачи

Аппроксимация локальных аномалий как пространственно-временные поля

Исходные данные: результаты численных экспериментов с региональной климатической моделью COSMO-CLM

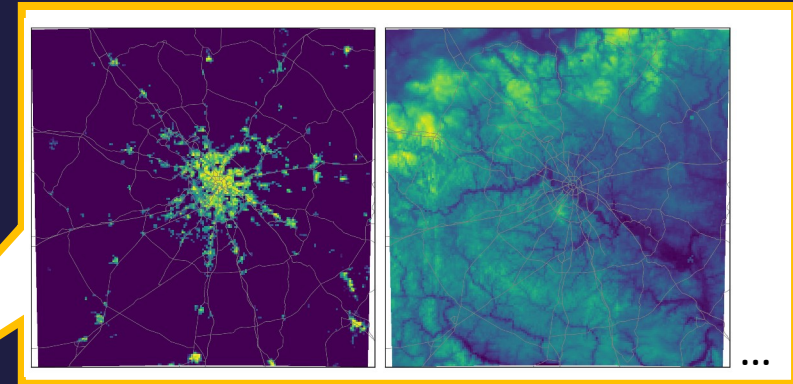
- ✓ Динамический даунскейлинг реанализа ERA-Interim
- ✓ Целевая область моделирования с шагом сетки 1 км (160x160 км)*
- ✓ Эксперименты с учетом влияния города и без него (URB/noURB)* для 10 летних сезонов, 2007-16 гг. (Varentsov et al., 2018)
- ✓ Данные с дискретностью 1 час по времени

* пока используется только 5% от доступного массива данных



Квазилокальный подход

- Целевая переменная: аномалия температуры ΔT



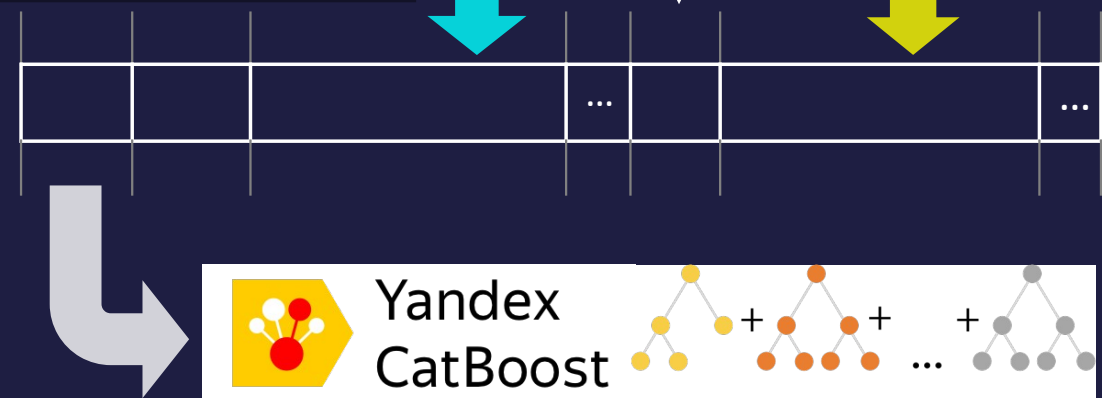
- Формализация задачи МО



- Метеорологические предикторы:** сглаженные до разрешения реанализа ERA-Interim результаты эксперимента noURB

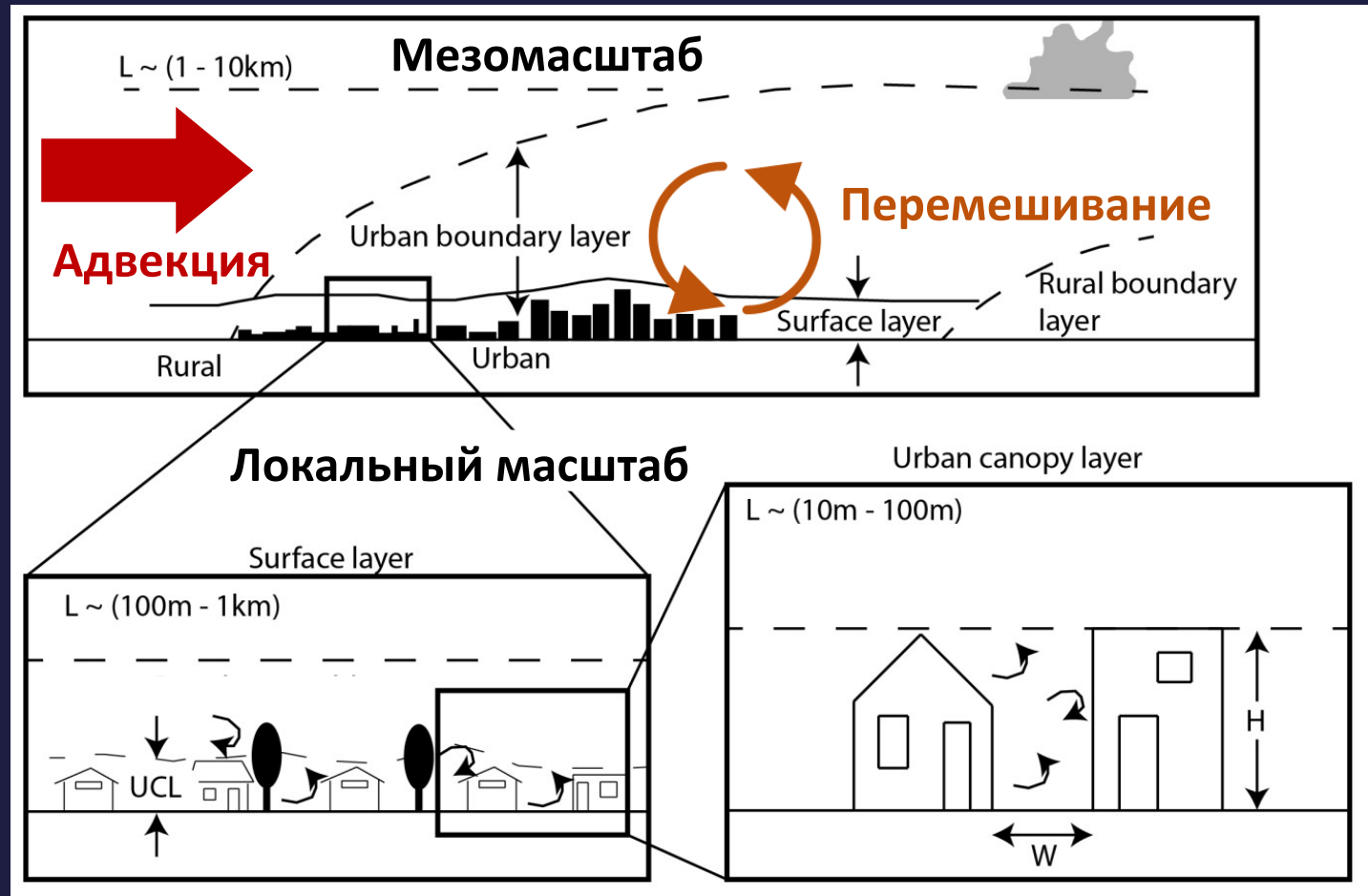
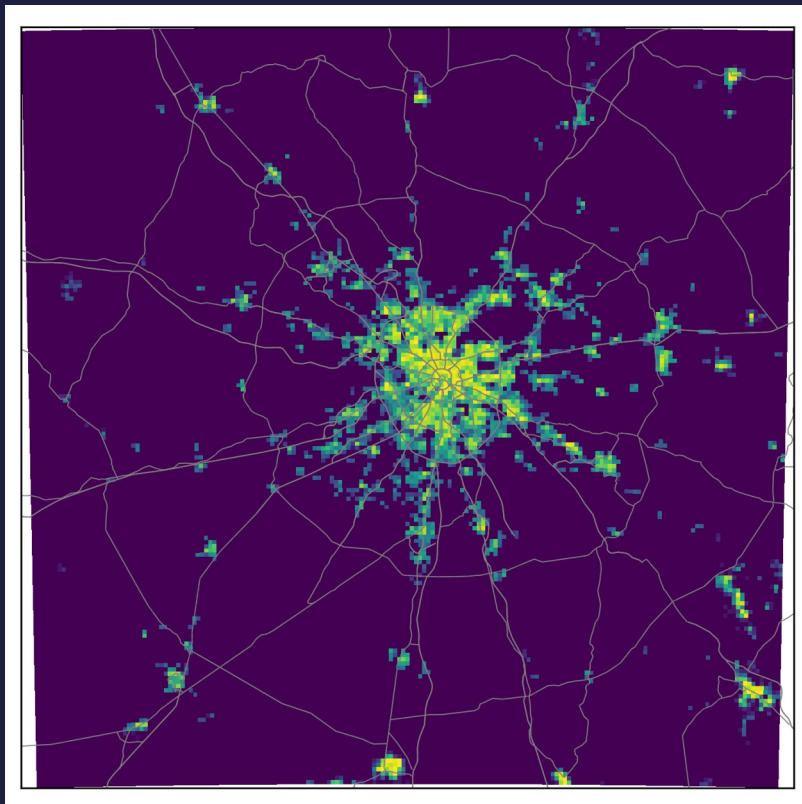
- Пространственные предикторы:** задаваемые в COSMO свойства поверхности и городской среды

- Модель МО:** CatBoost Regression



Пространственные предикторы

Локальные значения
пространственного предиктора

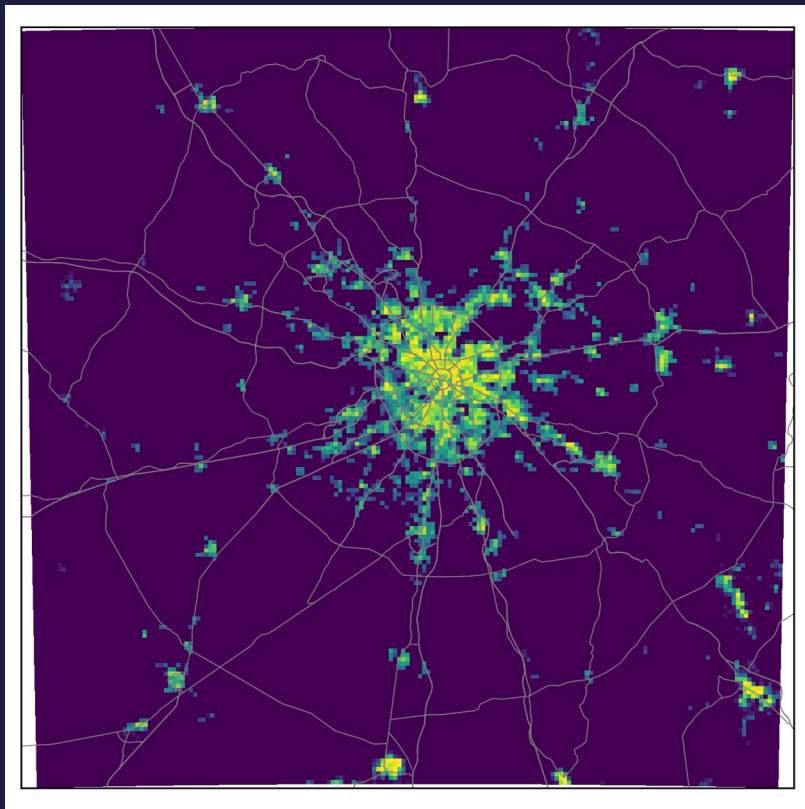


Пространственные предикторы

Локальные значения
пространственного предиктора

Учет перемешивания:
статический изотропный фильтр

Учет адвекции:
эластичная трансформация +
динамический направленный фильтр



Эксперимент SF_loc

Эксперимент SF_dif

Эксперимент SF_adv

Результаты: аппроксимация

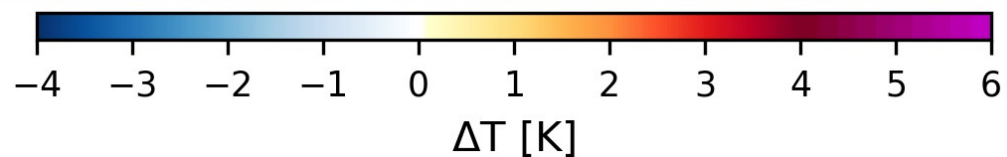
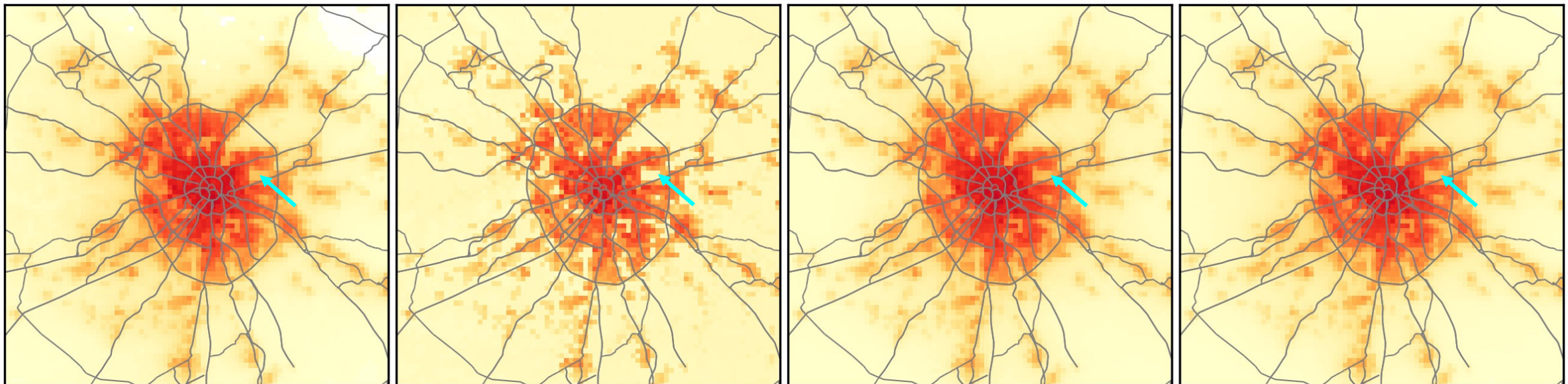
Средняя аномалия температуры за лето 2010 г.

COSMO (ground truth)

ML (SF_loc)

ML (SF_dif)

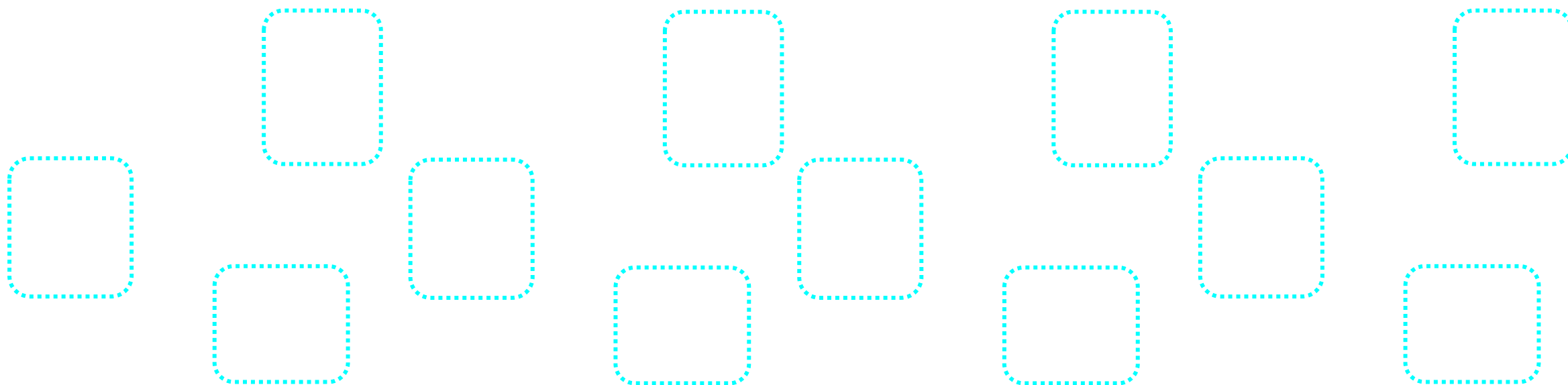
ML (SF_adv)



Результаты: аппроксимация

Моменты времени с ярко выраженной адвекцией острова тепла

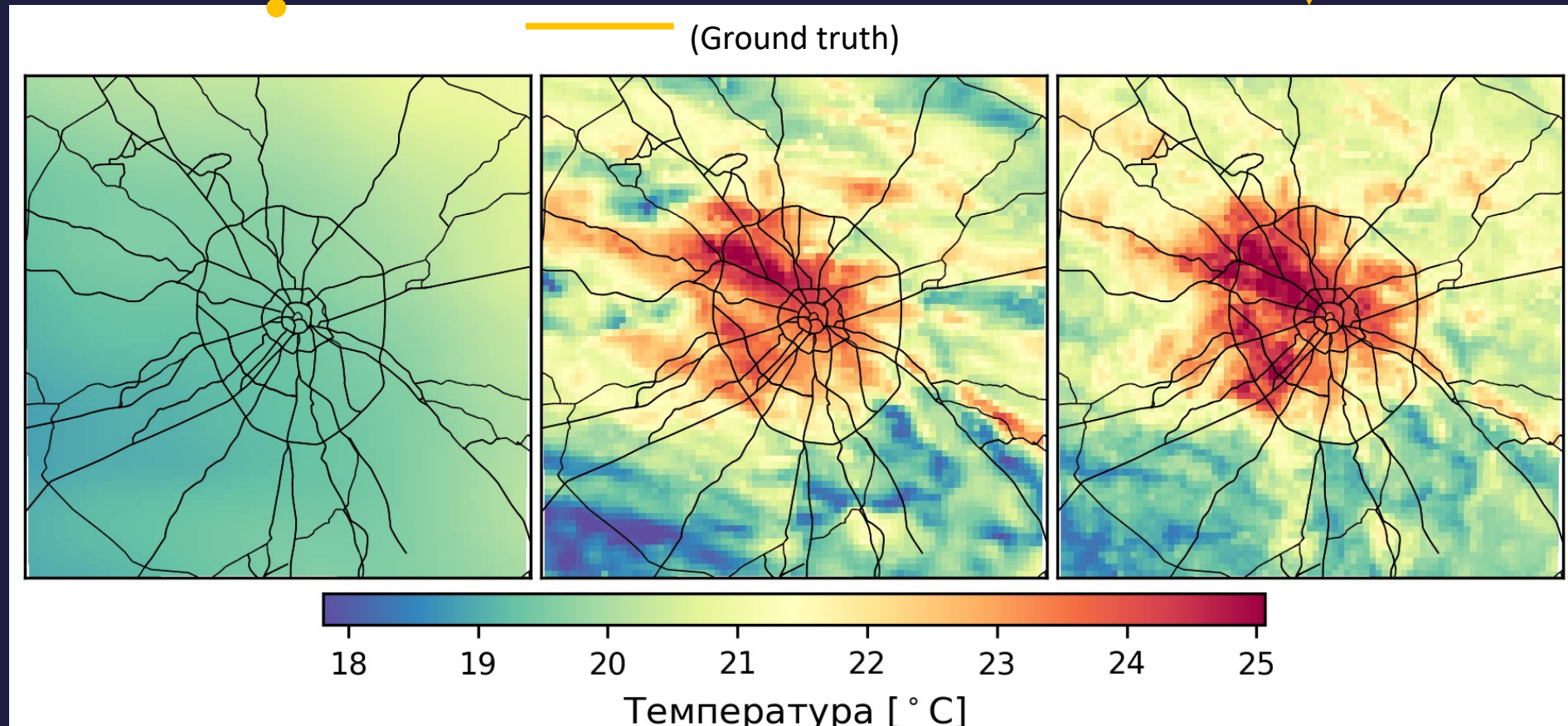
COSMO (ground truth)



Результаты: модель даунскейлинга



«Тропическая ночь»
накануне самого
жаркого дня в истории
Москвы, 29.07.2010



Вычислительные аспекты

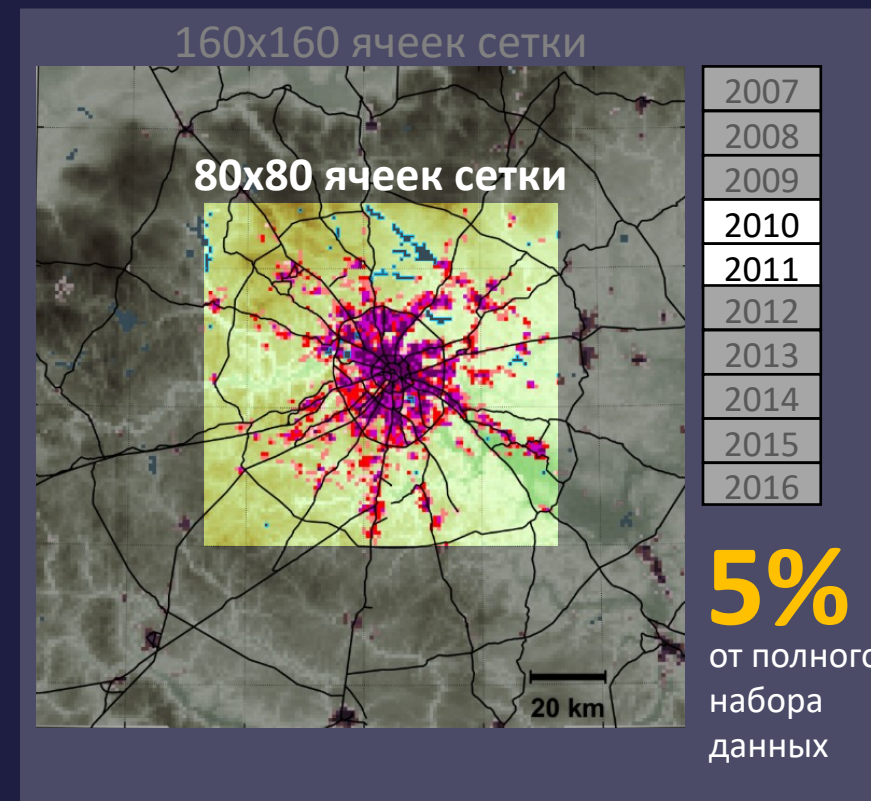
- ✓ Пока используется только **5%** от полного набора данных (1600 точек по пространству × 5904 точек по времени)
- ✓ **75 признаков** в эксперименте SF_adv
- ✓ **2.8×10^9** значений (**≈14 Gb**) в массиве данных
- ✓ Скорость обучения модели CatBoost ($n_estimators = 1000$) при использовании GPU для одной обучающей выборки:
 - **25 минут** на персональном компьютере
 - **35 секунд** на узле МГУ-270 (ускорение в 50 раз!)



Intel Core i7-11800H, 24 GB RAM,
GeForce RTX 3050 Ti Laptop GPU



1 узел СК МГУ-270:
8 GPU, 128 CPU, 2 TB RAM




Результаты

- ✓ Методы машинного обучения перспективны в городской метеорологии для детализации низкоразрешенных метеоданных для городов.
- ✓ «Классические» модели МО позволяют воспроизводить основные закономерности пространственно-временной динамики локальных аномалий в городах на основе крупномасштабных метеорологических предикторов и детализированных данных о подстилающей поверхности.
- ✓ Среди «классических» моделей МО алгоритм CatBoost выделяется как по метрикам качества, так и по вычислительной эффективности, особенно при использовании GPU.
- ✓ Фундаментальный интерес представляет сравнение «классических» моделей МО в квазилокальной постановке со свёрточными нейросетями
- ✓ **Дальнейшее развитие разрабатываемых моделей требует высокопроизводительных вычислений.**

Open Access Article

Machine Learning for Simulation of Urban Heat Island Dynamics Based on Large-Scale Meteorological Conditions

Climate 2023, 11(10), 200;
<https://doi.org/10.3390/cli11100200>



Moscow University Physics Bulletin **79**(8), x (2024)

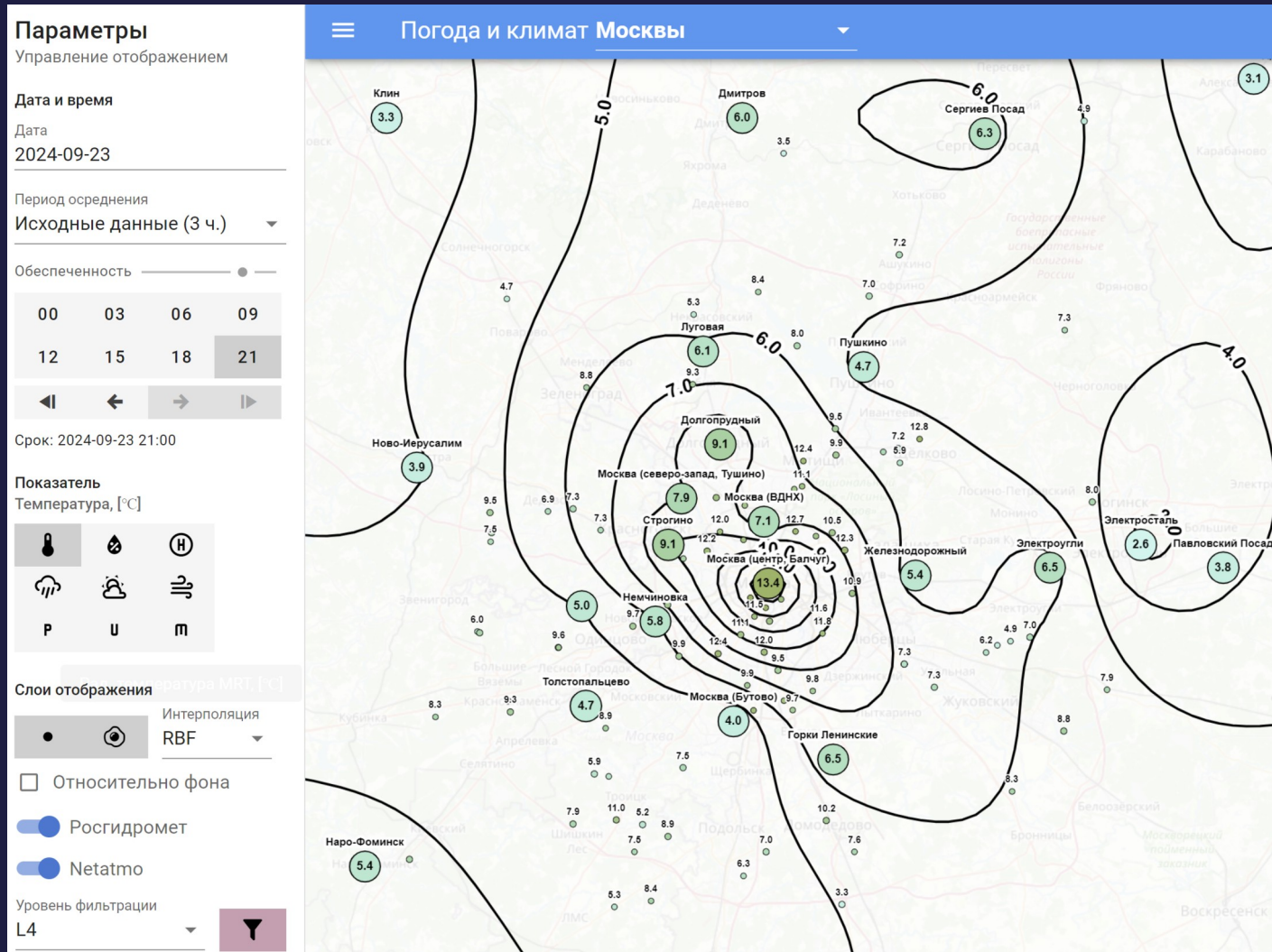
Machine learning in environmental sciences

Approximation of spatial and temporal variability of the urban heat island in Moscow using machine learning

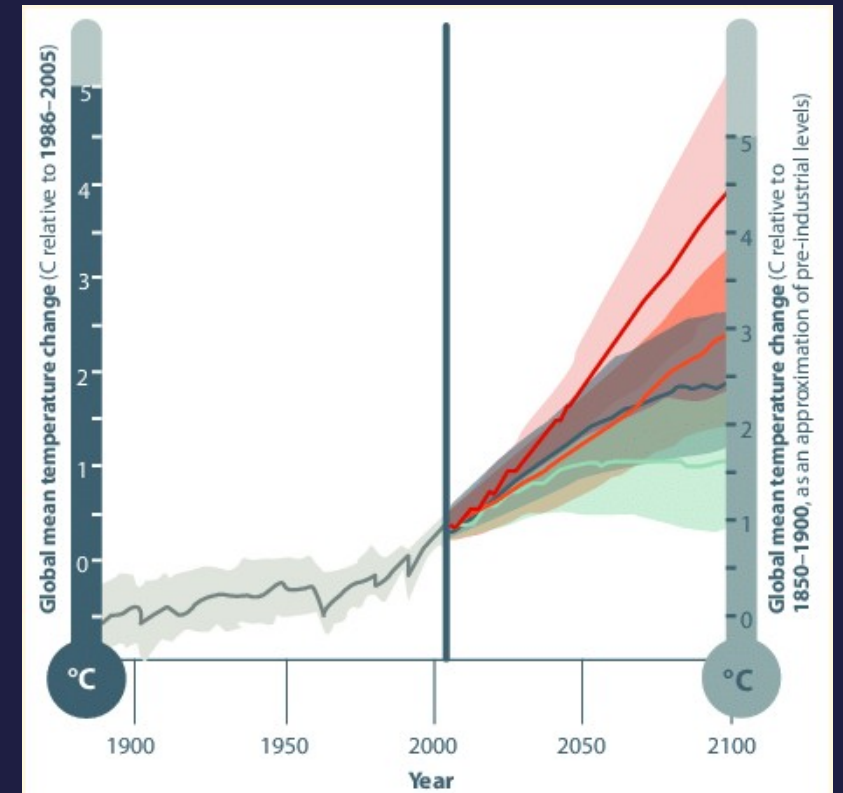
M.I. Varentsov,^{1,2,3,*} M.A. Krinitskiy,^{4,5,1} and V.M. Stepanenko^{1,2}

Перспективы практического применения

Интеграция в погодно-климатические сервисы



Детализация ожидаемых изменений климата и их влияния на население



**Спасибо за внимание!
Вопросы?**

mikhail.varentsov@srcc.msu.ru



Дополнительные слайды

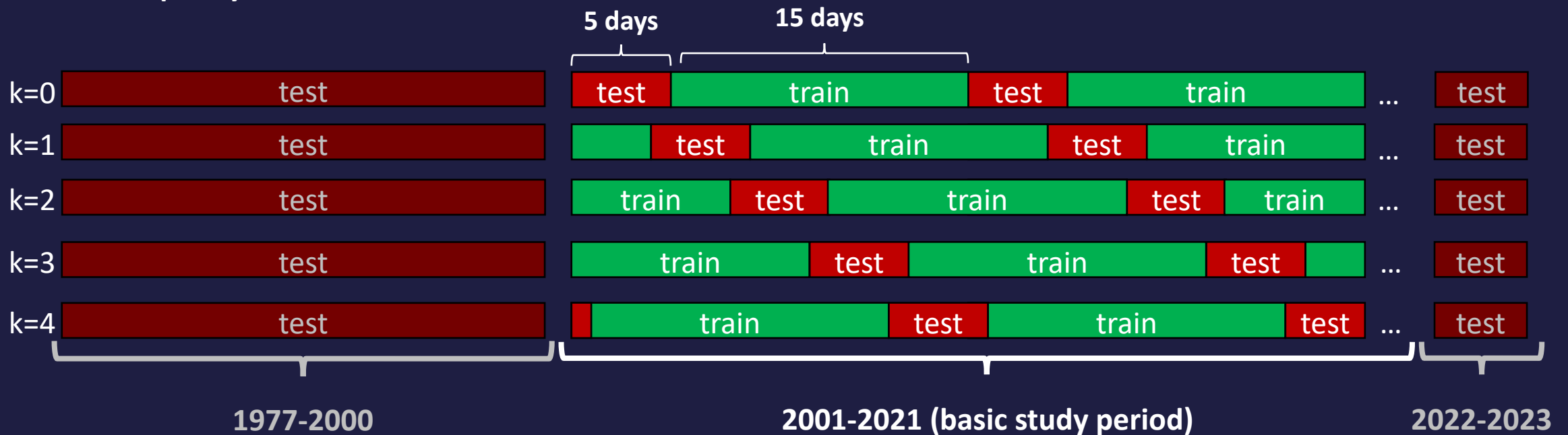
Этап №1: кросс-валидация

□ The dataset:

- time series with 3-hour step for 47 years (1977-2021)
- 21 years (2001-2021) are used to train the models and for the major part of analysis

□ Train-test split using blocked k-fold method with 5-days blocks and 4:1 train-to-test ratio

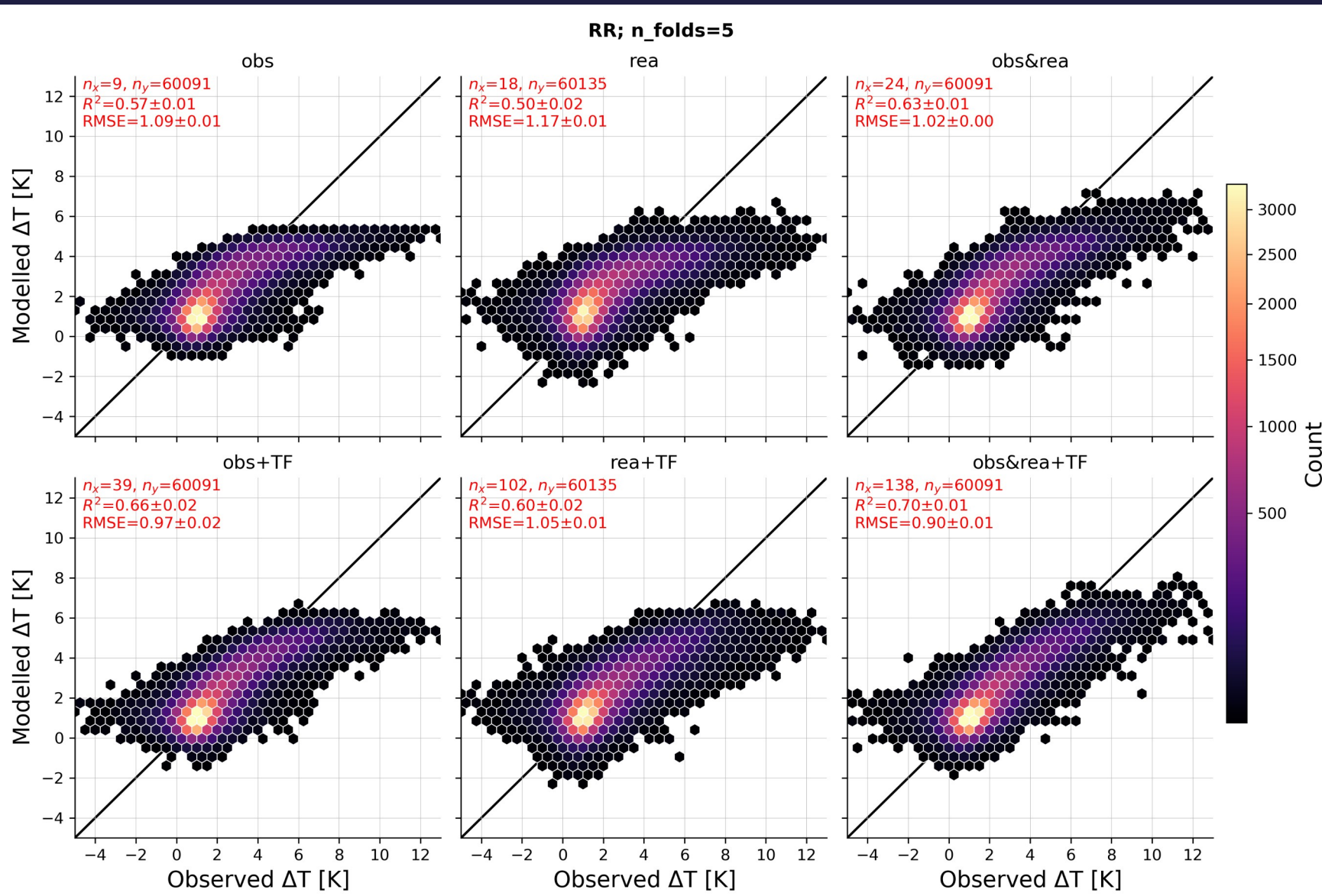
□ Model quality metrics: RMSE, ME, R²



Этап №1: результаты (RR)

Предикторы
только за
текущий i-й
момент
времени

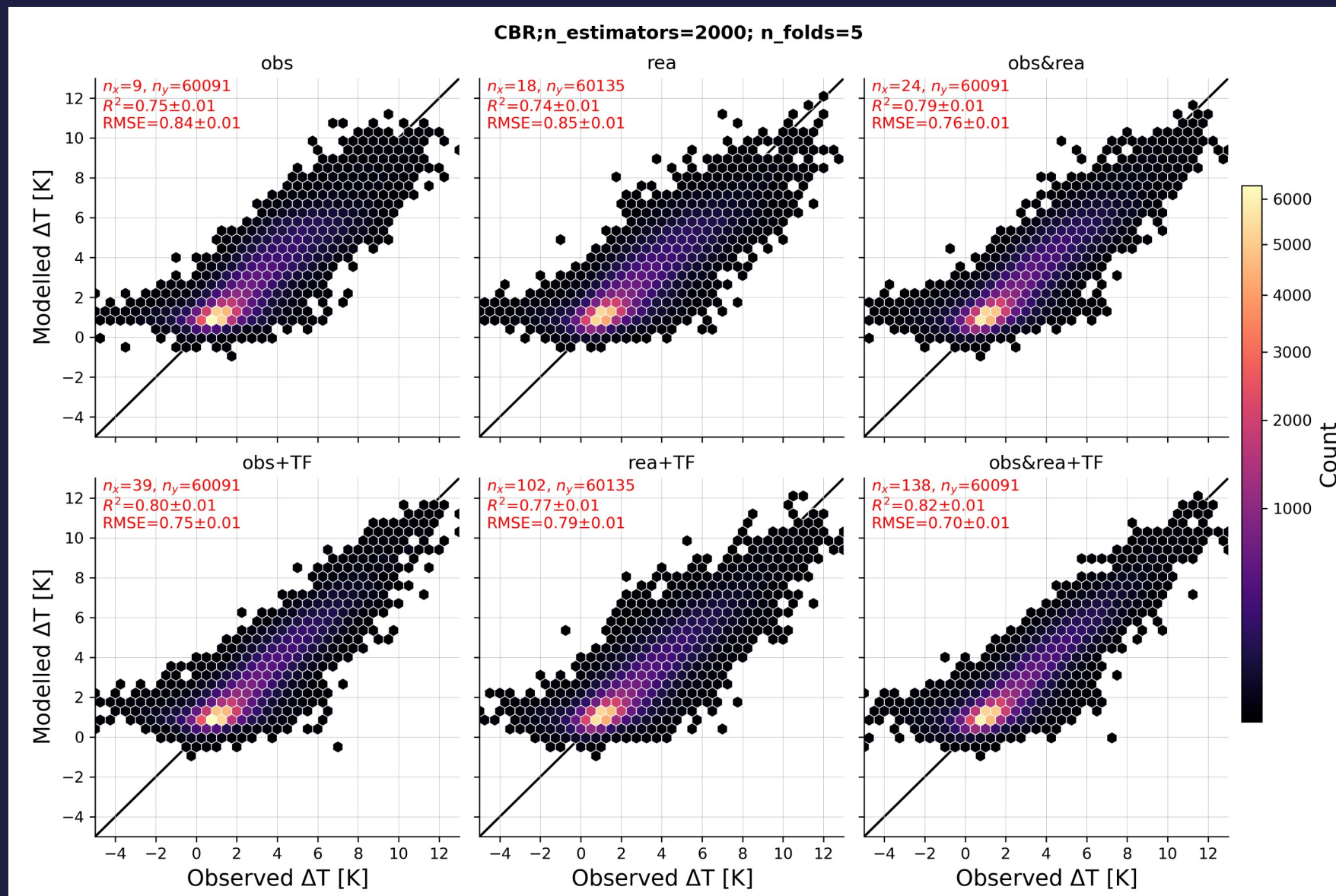
С
дополнительны
ми признаками на
основе предикторов
за предшествующие
моменты времени



Этап №1: результаты (CVR)

Предикторы
только за
текущий i-й
момент
времени

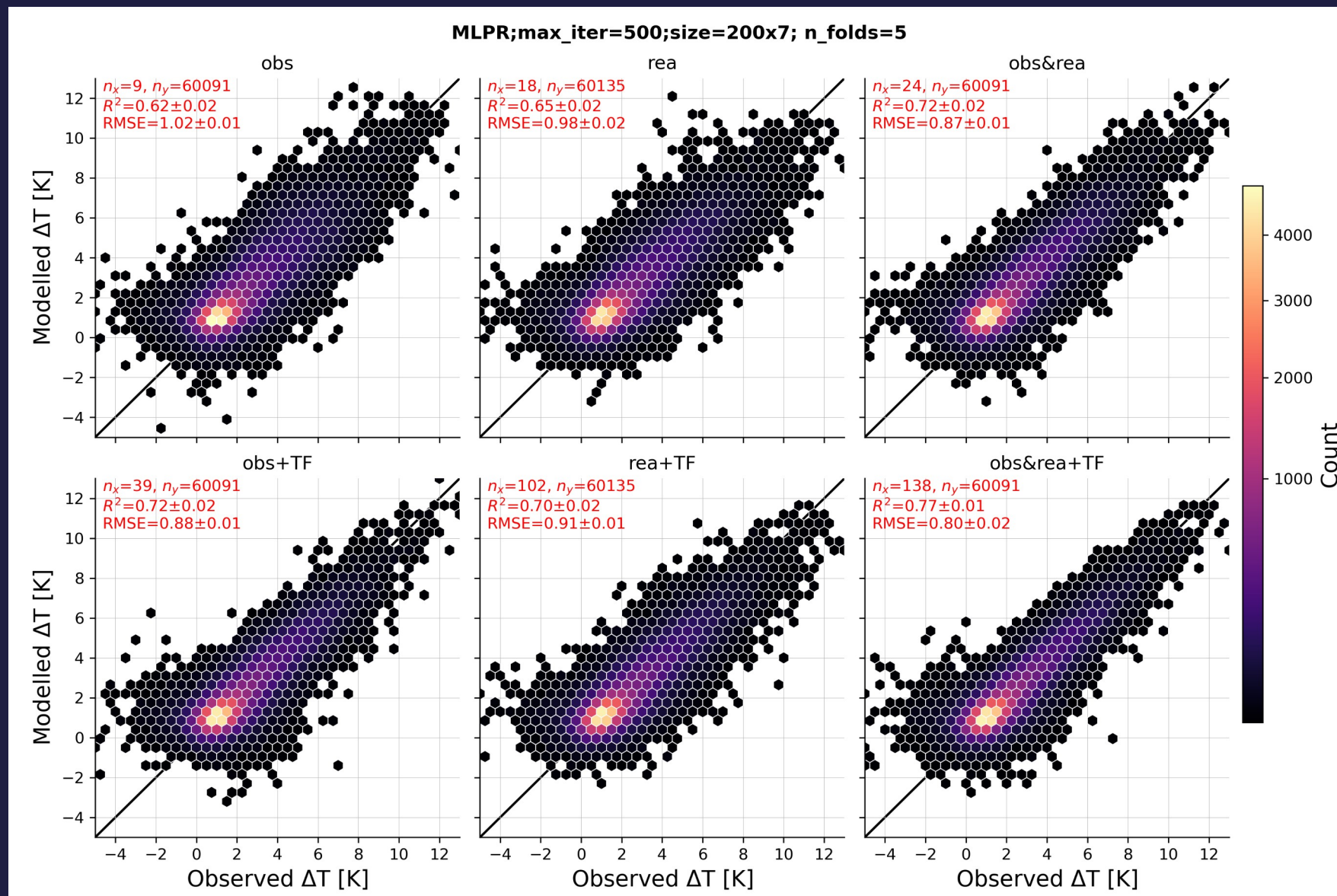
С
дополнительны
ми признаками на
основе предикторов
за предшествующие
моменты времени



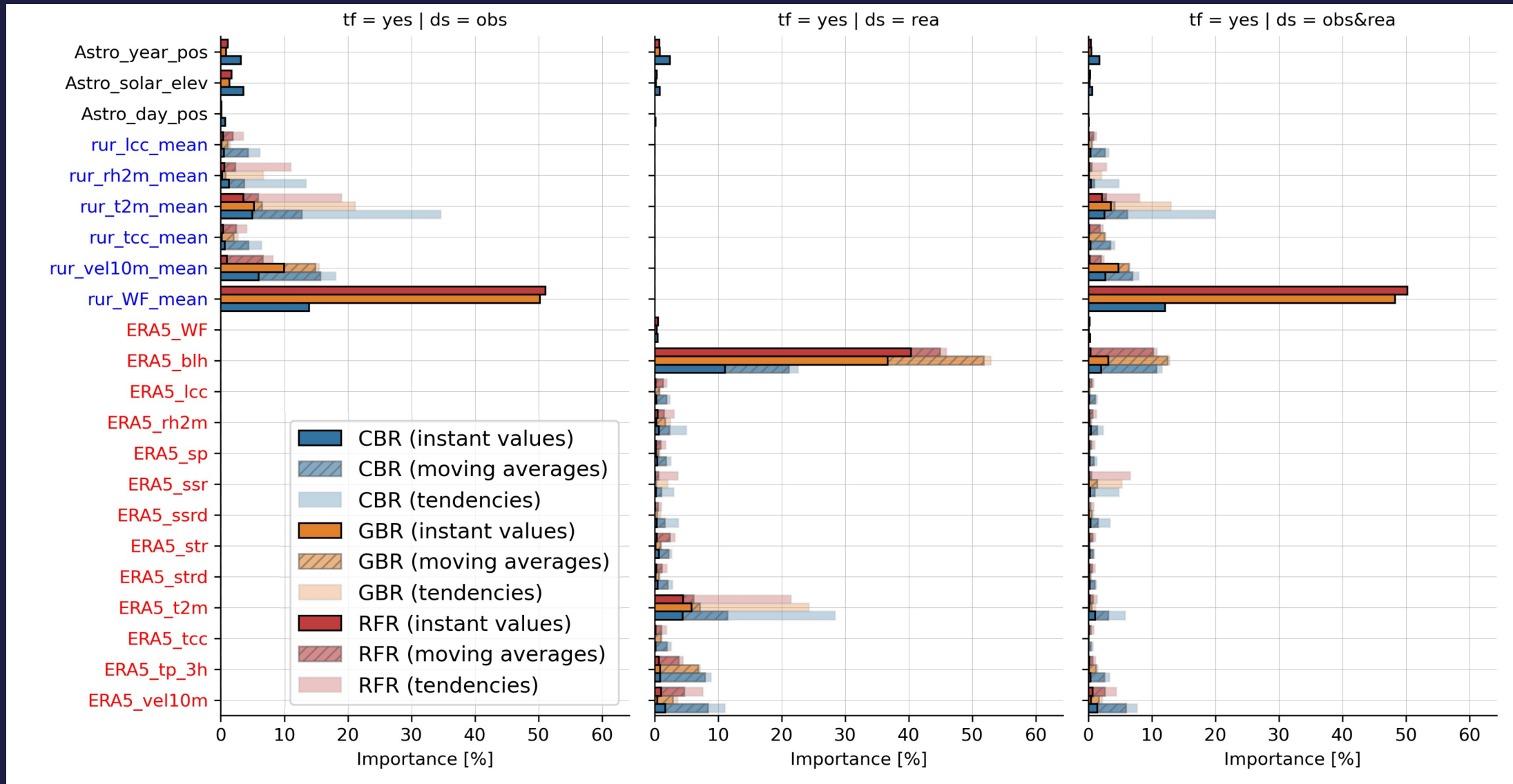
Этап №1: результаты (MLPR)

Предикторы
только за
текущий i-й
момент
времени

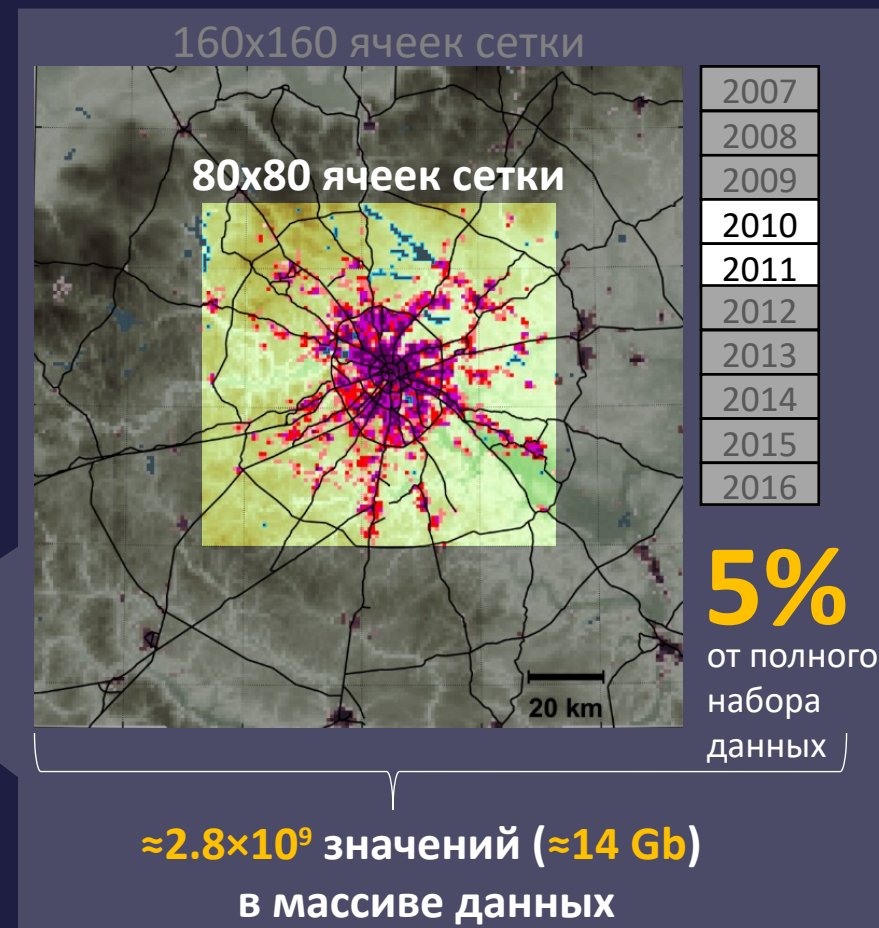
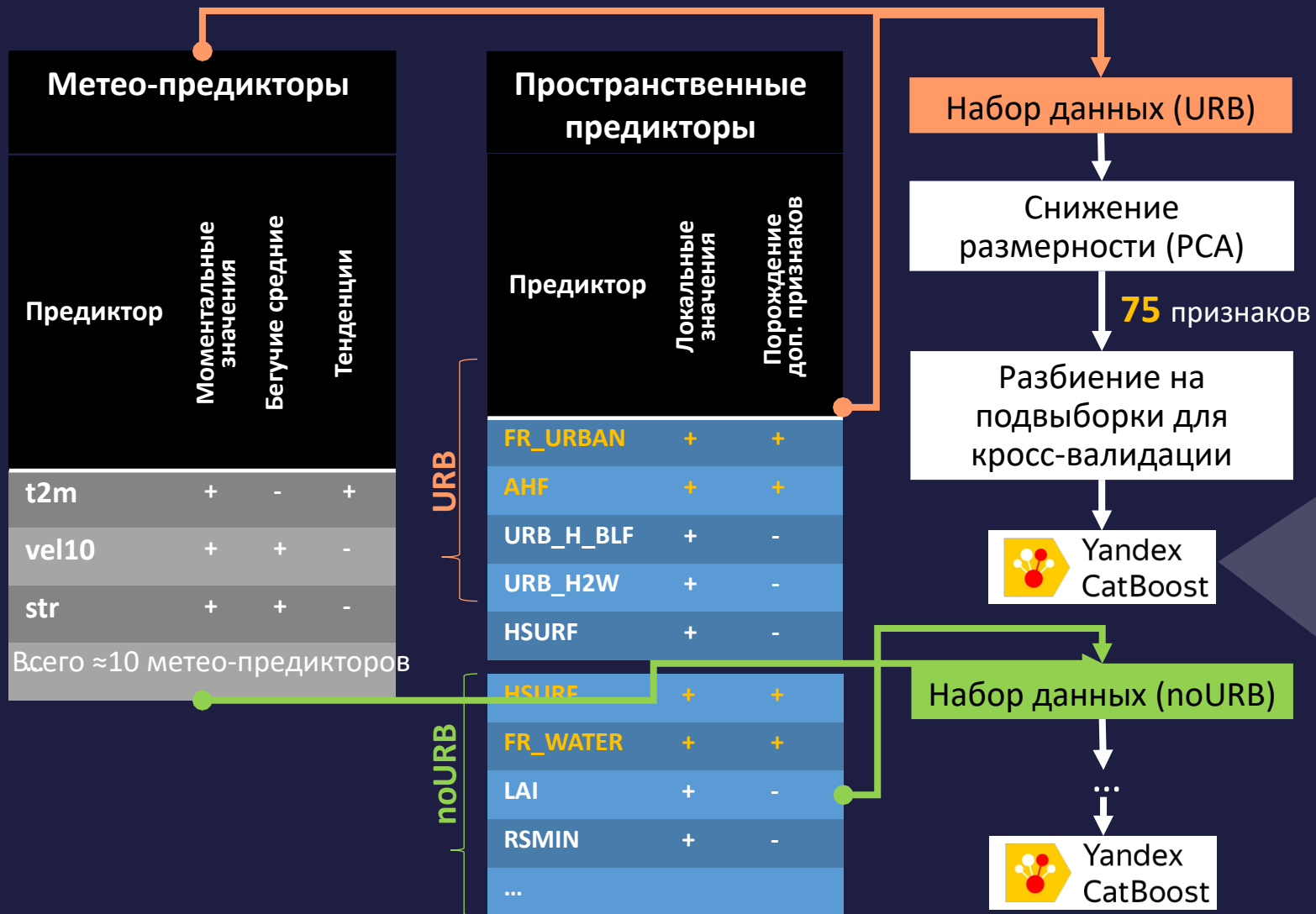
С
дополнительны
ми признаками на
основе предикторов
за предшествующие
моменты времени



Этап №1: значимость предикторов



Этап №2: данные для обучения



«Квазилокальный» и «нелокальный» подходы

