

Двадцать третья международная конференция «Современные проблемы
дистанционного зондирования Земли из космоса»

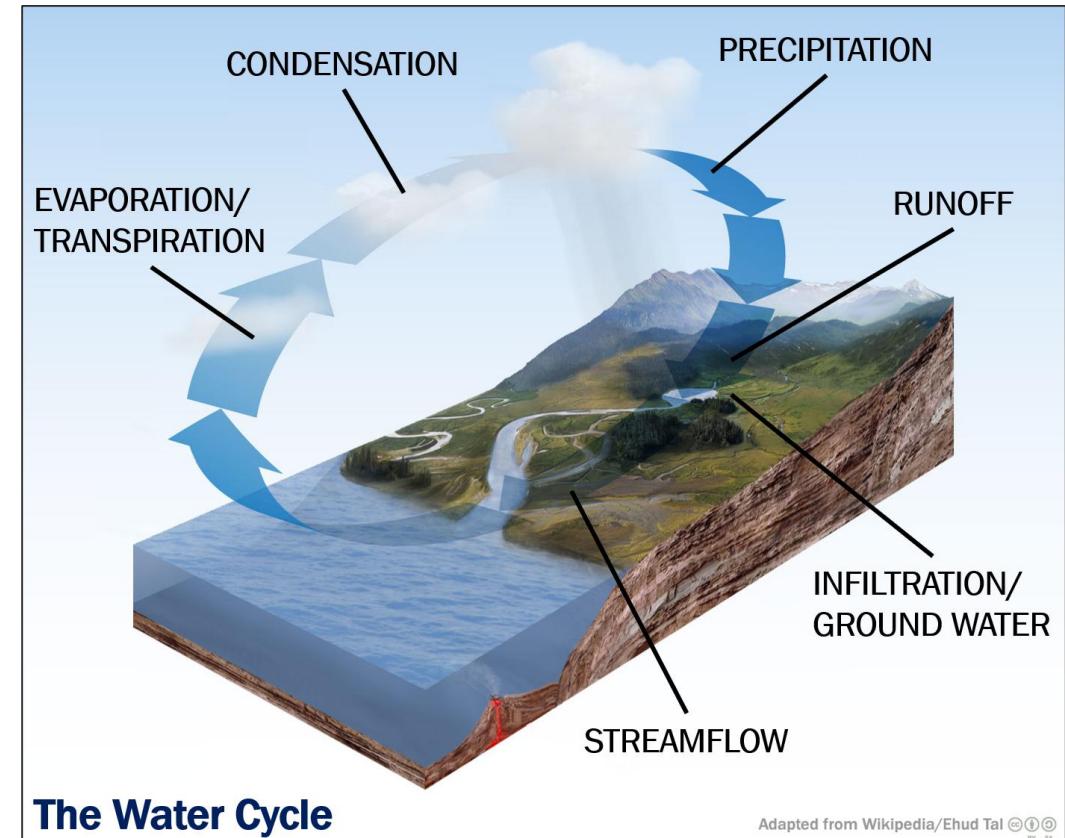
10-14 ноября 2025 г., ИКИ РАН

Коррекция модельных осадков INM-CM5-0 над Северной Евразией с использованием данных IMERG и моделей машинного обучения

Бахрамхан Я.О., НИУ ВШЭ, ИКИ РАН
Акперов М.Г., ИФА РАН

Актуальность исследования

Оценка атмосферных осадков необходима для различных сфер жизнедеятельности, в частности для сельского хозяйства, природопользования и строительства. Однако поля осадков, получаемые с помощью численных моделей погоды и климата, содержат систематические ошибки, которые проявляются как во временном, так и в пространственном масштабах. Для коррекции систематических ошибок применяют различные методы, включая статистические и динамические подходы и методы машинного обучения. По сравнению со статистическим методом алгоритмы машинного обучения учитывают нелинейные связи между модельными данными и данными реанализа. По сравнению с динамическим методом, модели машинного обучения не требуют значительных вычислительных ресурсов.



weather.gov

Adapted from Wikipedia/Ehud Tal CC BY-SA

Выбор территории, данных и методов

Исходные данные:

- Поля среднесуточных осадков климатической модели INM-CM5-0, полученных в историческом эксперименте проекта CMIP6 за 2007–2014 гг.
- Поля среднесуточных осадков мультиспутникового продукта IMERG (англ. Integrated Multi-satellitE Retrievals for GPM), в котором поля осадков получены на основе данных инфракрасных и пассивных микроволновых датчиков, а также наблюдений с метеорологических станций.

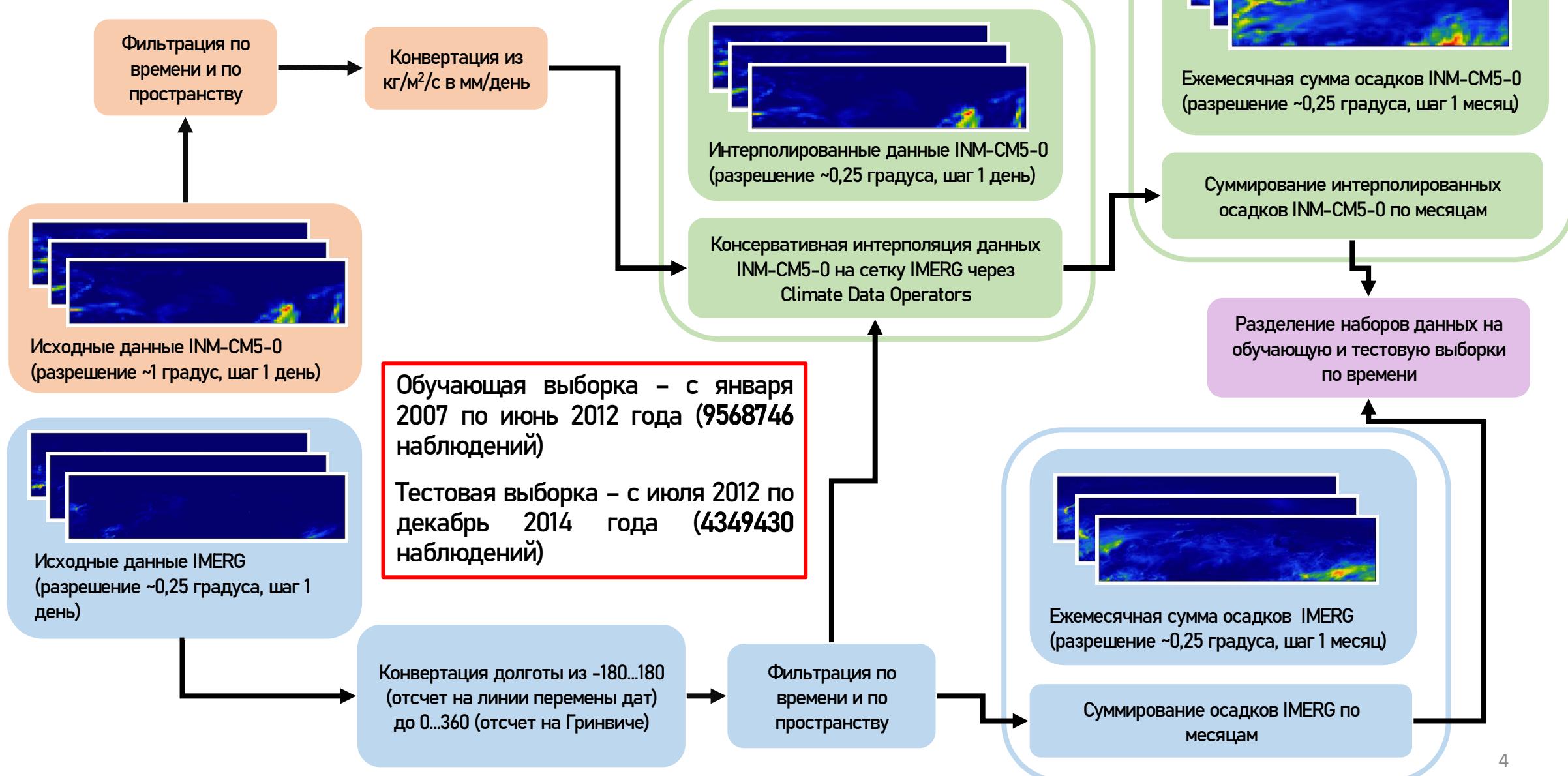
Выбранные методы машинного обучения для коррекции атмосферных осадков:

- Линейная регрессия
- Градиентный бустинг
- Случайный лес

Территория – Северная Евразия в пределах в прямоугольнике 40° – 85° с.ш, 0° – 180° в.д и 180° – 160° з.д.



Алгоритм предобработки данных

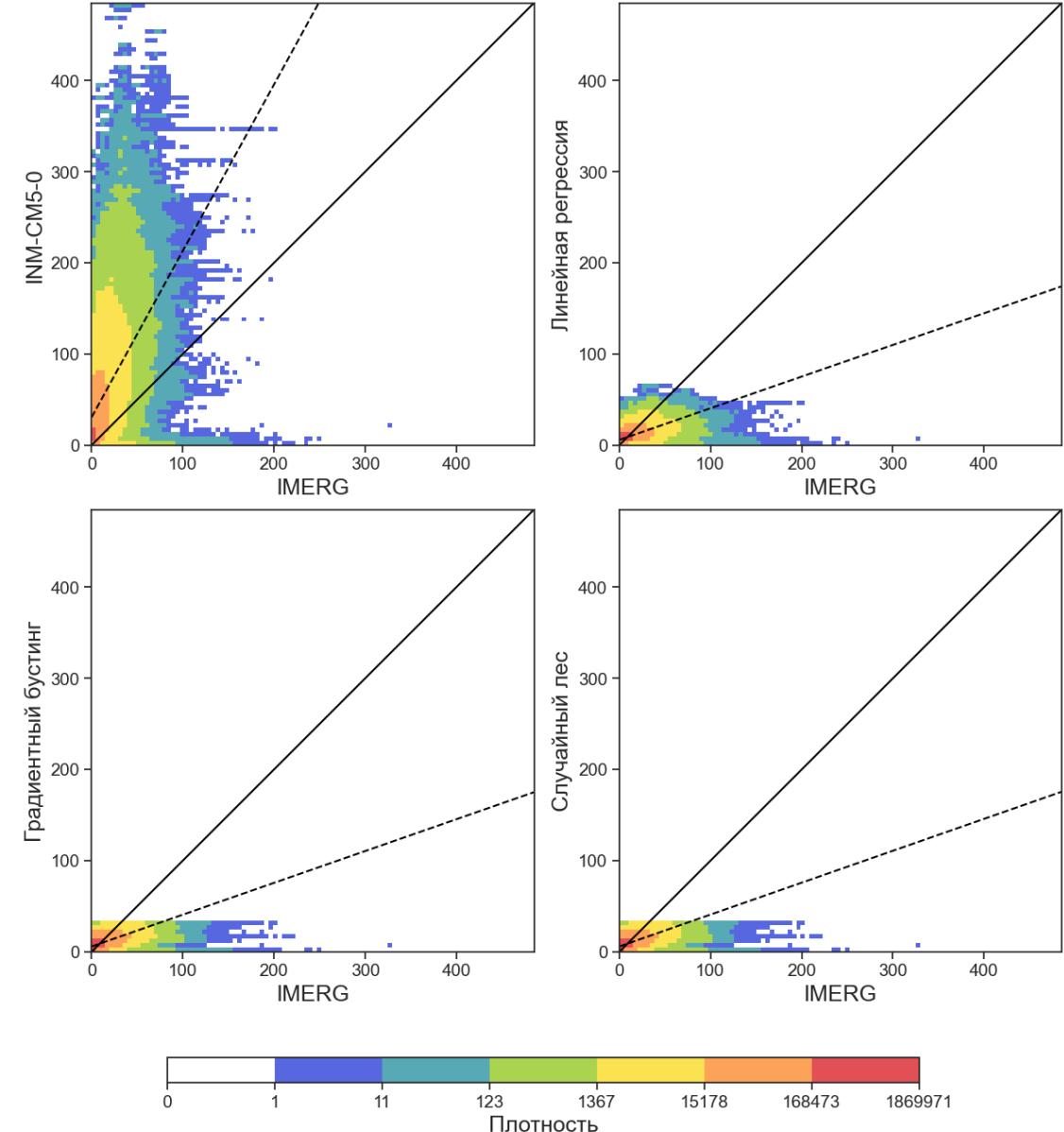


Результаты. Анализ метрик и распределения

Все метрики в результате обучения всех трех моделей улучшились. Наибольшее улучшение показали корень от среднеквадратической ошибки (RMSE) и средняя абсолютная ошибка (MAE). В среднем RMSE понизился на 81,5% от исходного значения, а MAE - на 82,3%. Коэффициент корреляции незначительно увеличился. R^2 поменял свой знак и теперь модели коррекции объясняют около четверти дисперсии данных IMERG.

Предсказания моделей слабо отклоняются от значений IMERG для нулевых и умеренных сумм осадков и сильно для экстремальных осадков.

| Модель | RMSE | MAE | r | R^2 |
|---------------------------|--------|--------|-------|---------|
| Исходные данные INM-CM5-0 | 57,461 | 40,729 | 0,504 | -20,626 |
| Линейная регрессия | 10,726 | 7,246 | 0,504 | 0,246 |
| Градиентный бустинг | 10,624 | 7,167 | 0,518 | 0,261 |
| Случайный лес | 10,618 | 7,166 | 0,519 | 0,262 |

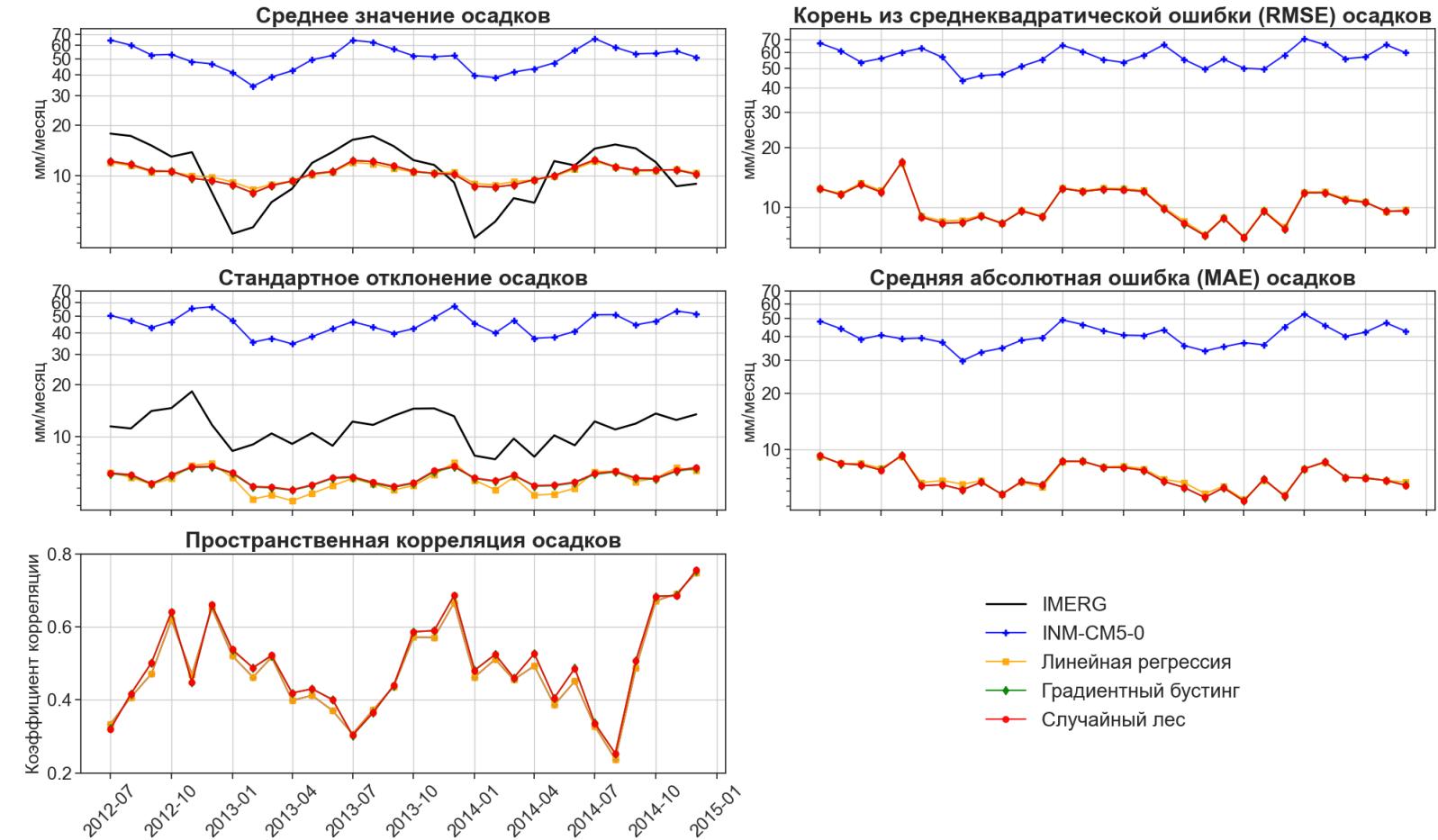


Временной анализ метрик

Среднее значение ежемесячной суммы осадков для Северной Евразии переоценено для зимних месяцев и недооценено для летних, хотя скорректированные предсказания повторяют тренд повышения суммы осадков летом и ее снижения зимой.

Стандартное отклонение занижено у всех моделей, в особенности у линейной регрессии.

Корреляция скорректированных осадков с IMERG обладает сезонностью – ее максимум (около 0,55-0,7) приходится на осень, а минимум (0,3 – 0,45) наблюдается летом, и она практически совпадает с корреляцией осадков INM-CM5-0 с осадками по IMERG.

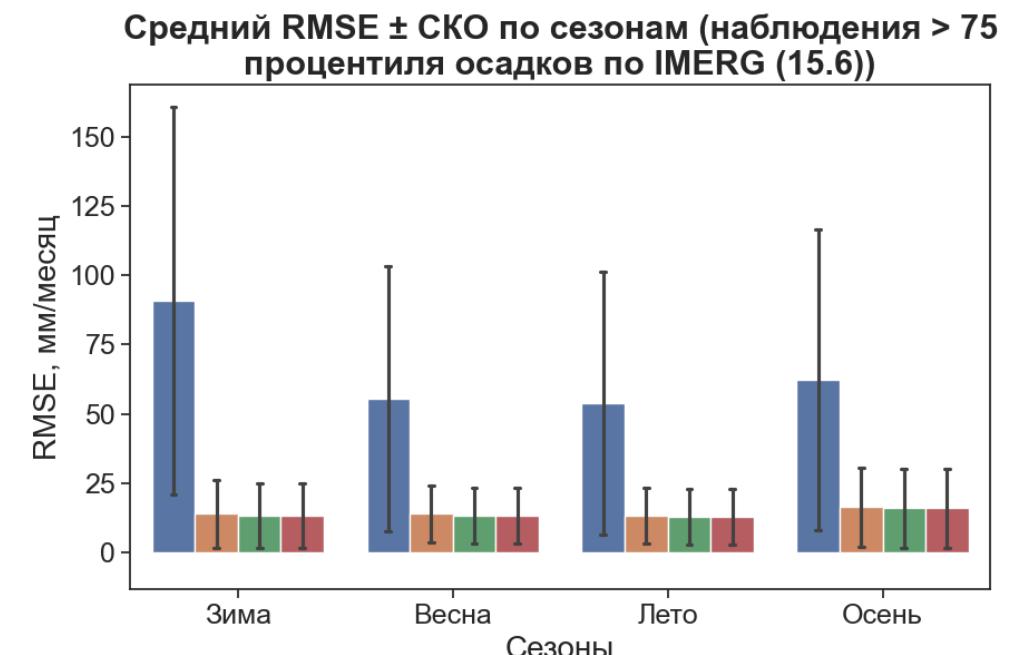
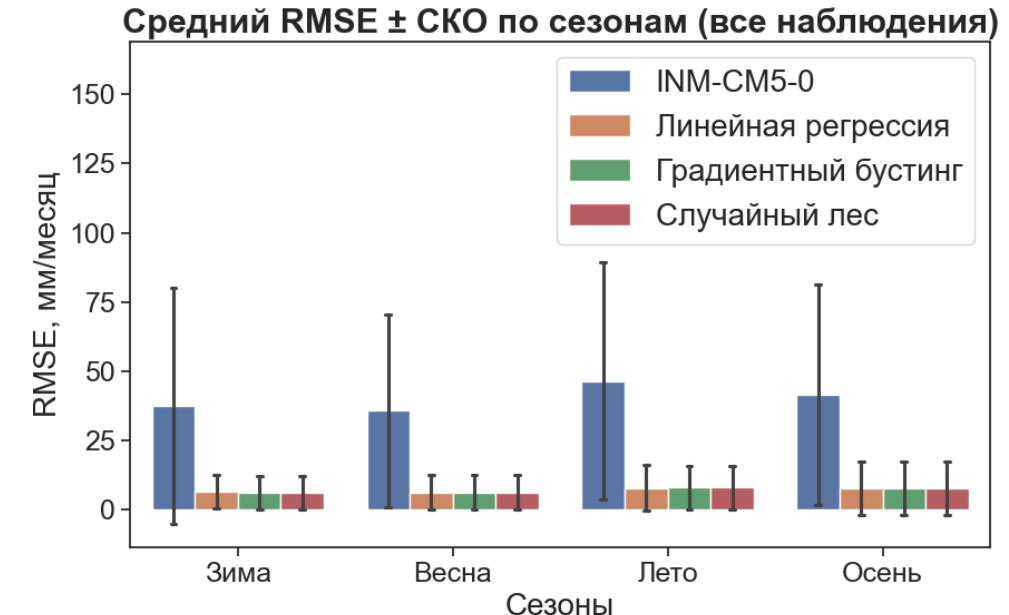


Сезонный анализ метрик и анализ экстремальных осадков

Средний RMSE, а также его стандартное отклонение по всем сезонам уменьшился в результате коррекции как умеренных, так и экстремальных сумм осадков.

Моделям коррекции удалось уменьшить разрыв между средним RMSE в сезоны с обильными дождевыми осадками (летом и осенью) и средним RMSE зимой и весной.

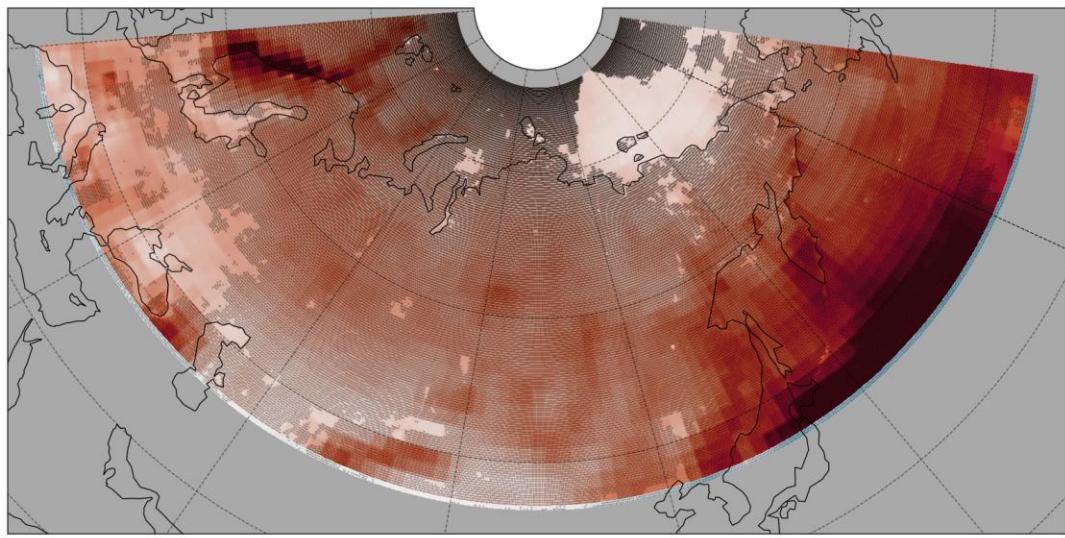
В случае с экстремальными осадками у среднего RMSE между осадками по INM-CM5-0 и по IMERG наблюдается повышенное значение в зимнее время, в то время как для среднего RMSE между скорректированными осадками и IMERG пик приходится на осень.



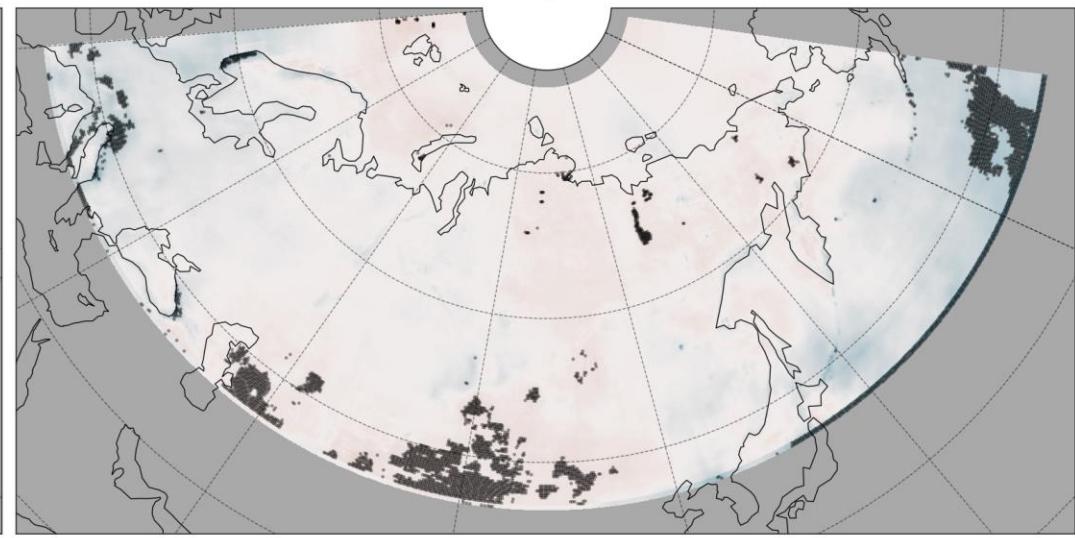
Результаты. Пространственный анализ метрик

Отклонение среднего

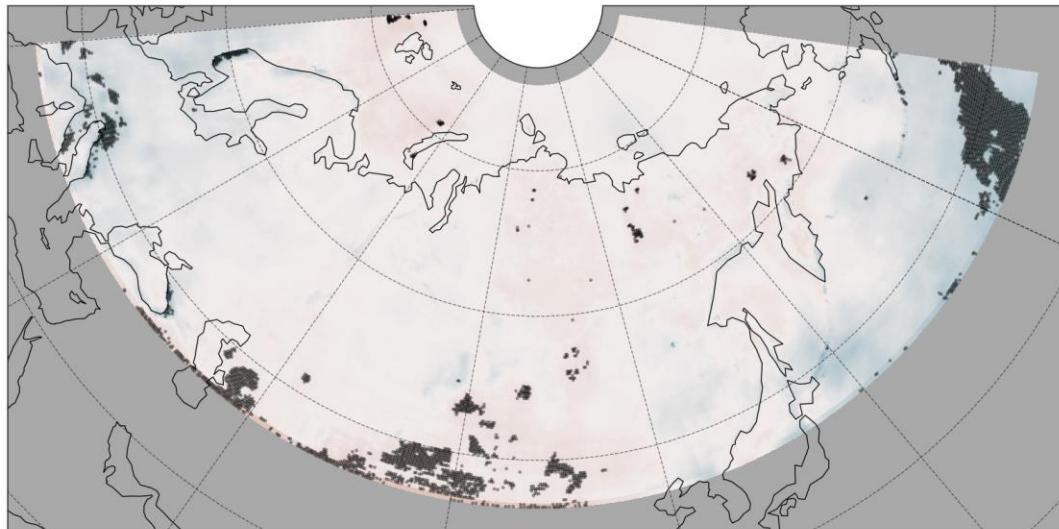
INM-CM5-0



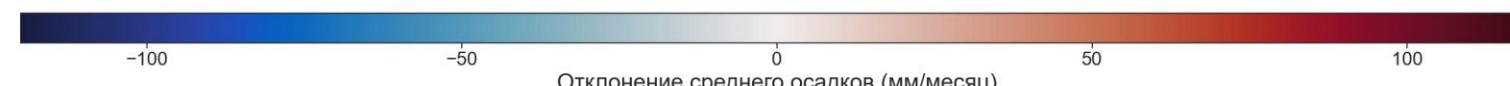
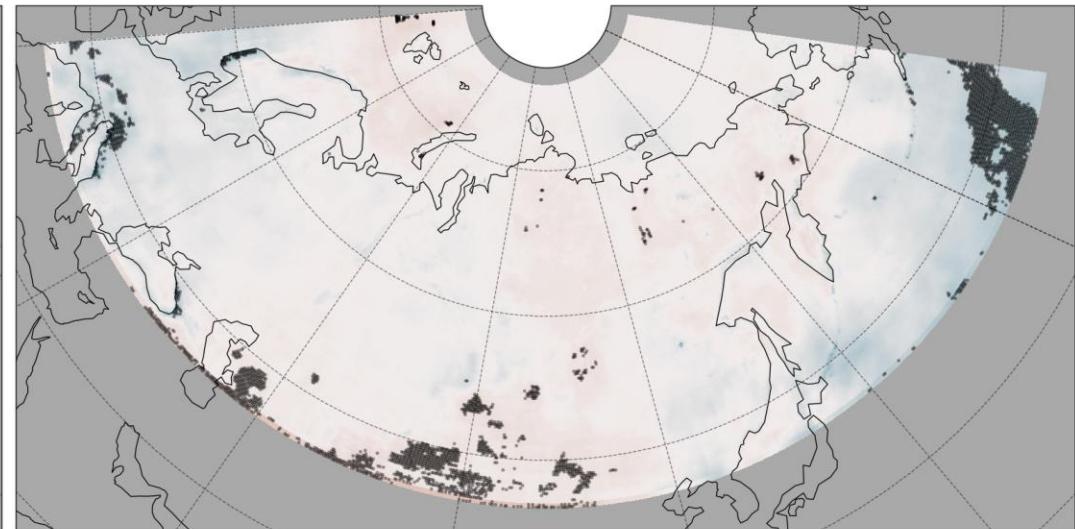
Линейная регрессия



Градиентный бустинг

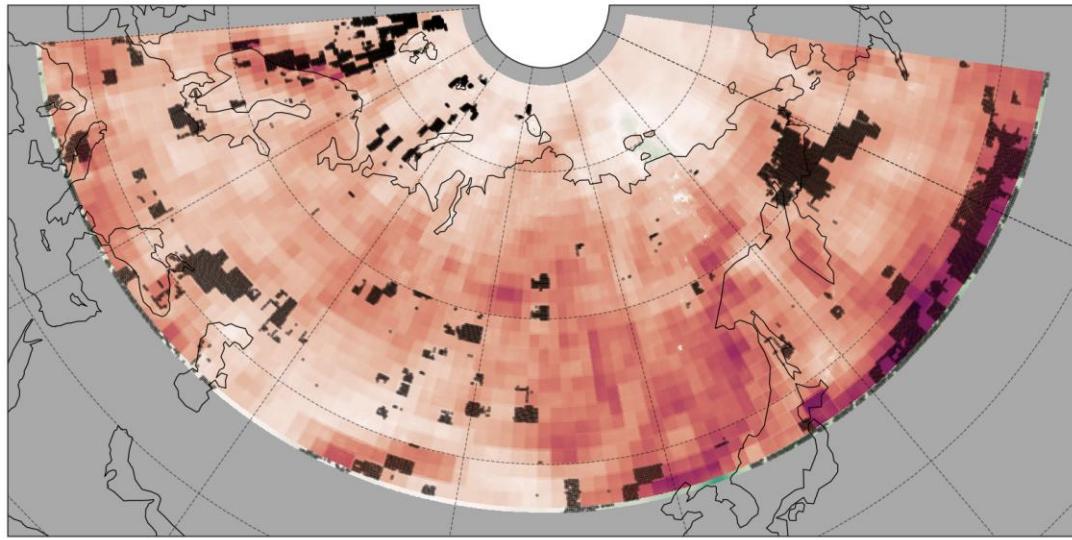


Случайный лес

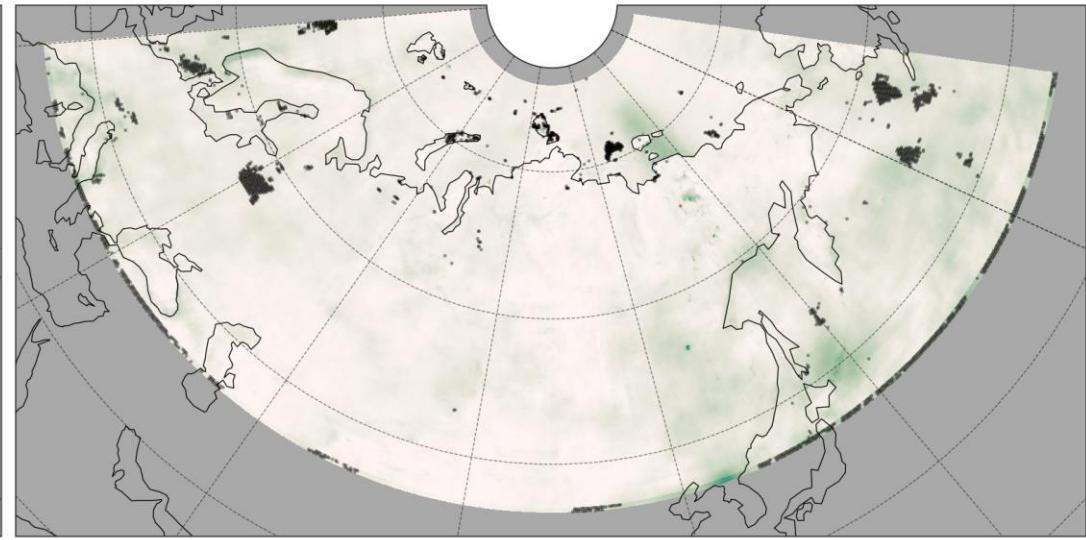


Отклонение стандартного отклонения

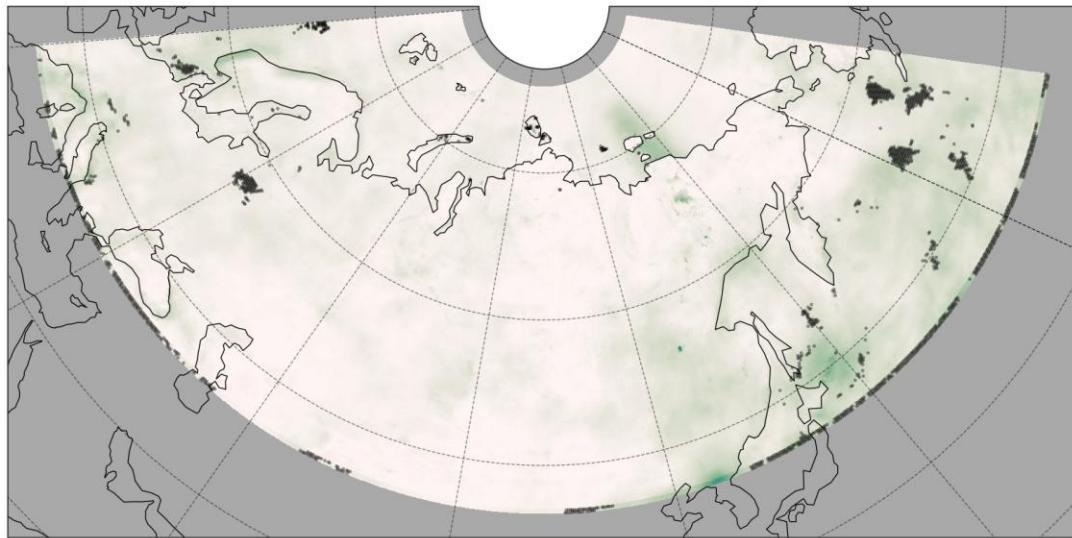
INM-CM5-0



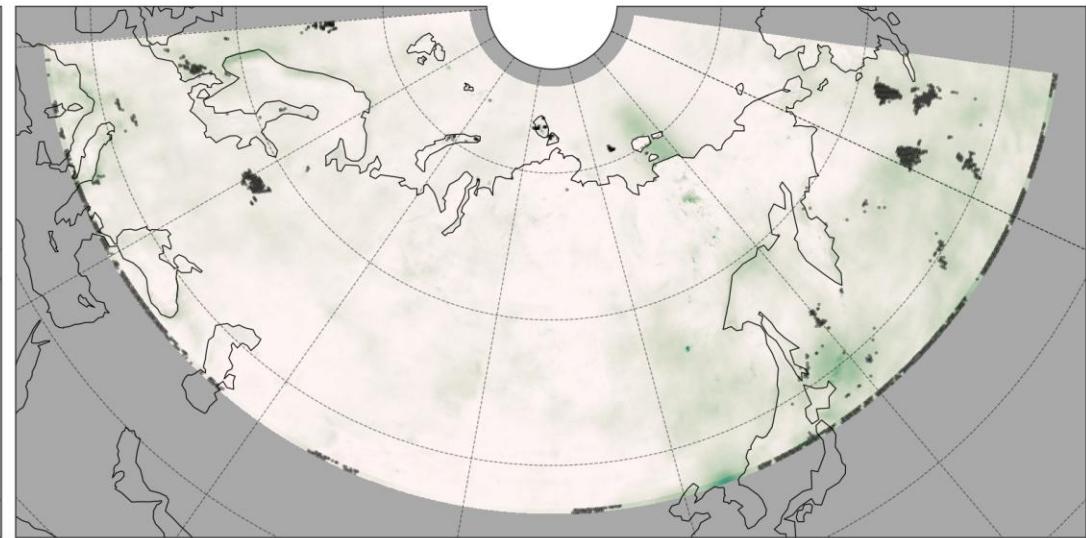
Линейная регрессия



Градиентный бустинг

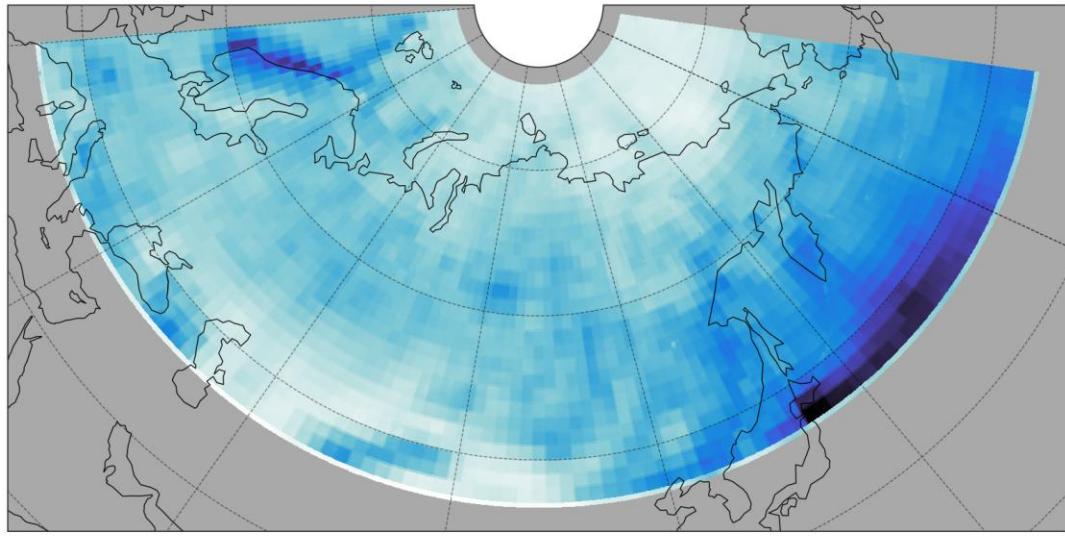


Случайный лес

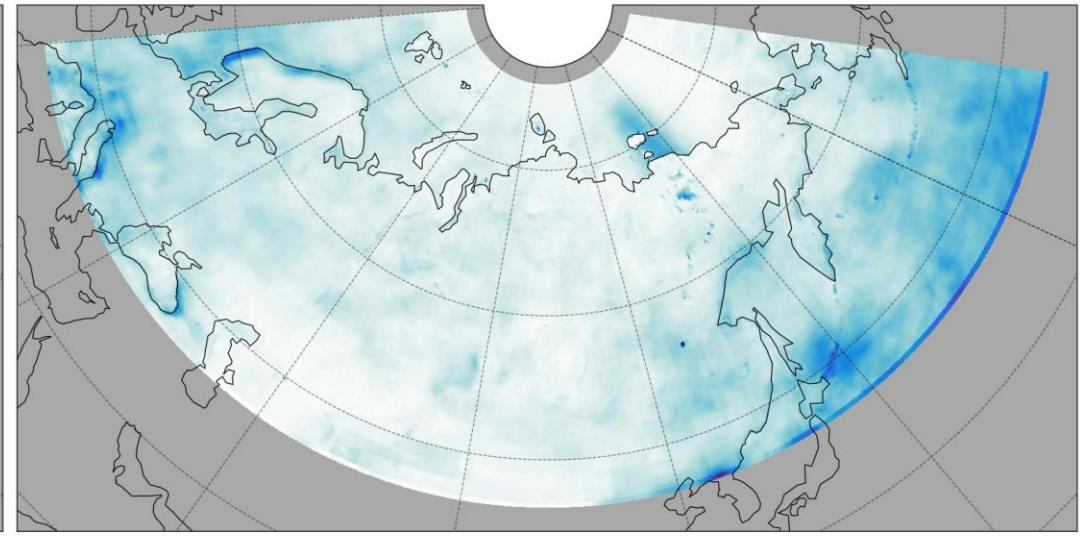


RMSE осадков

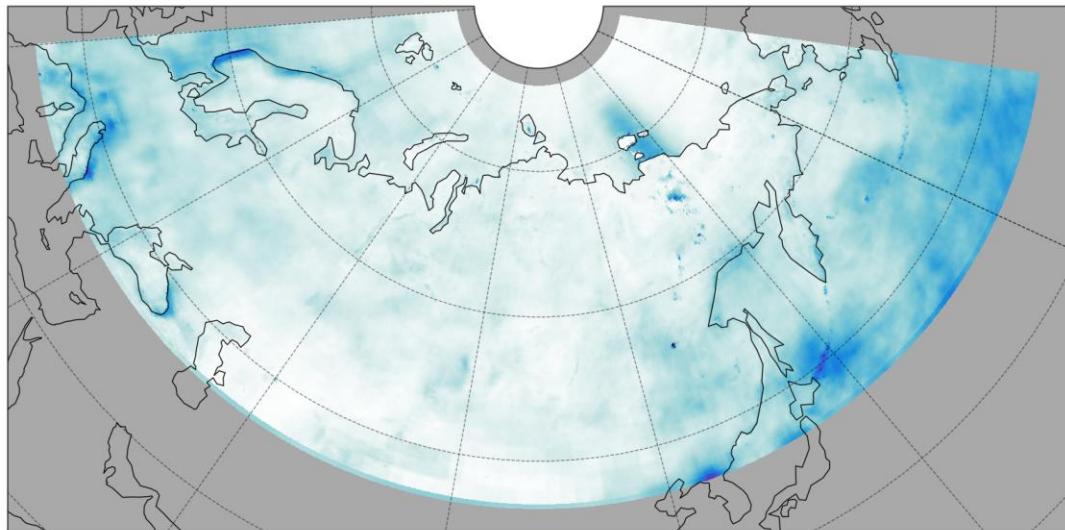
Исходные данные



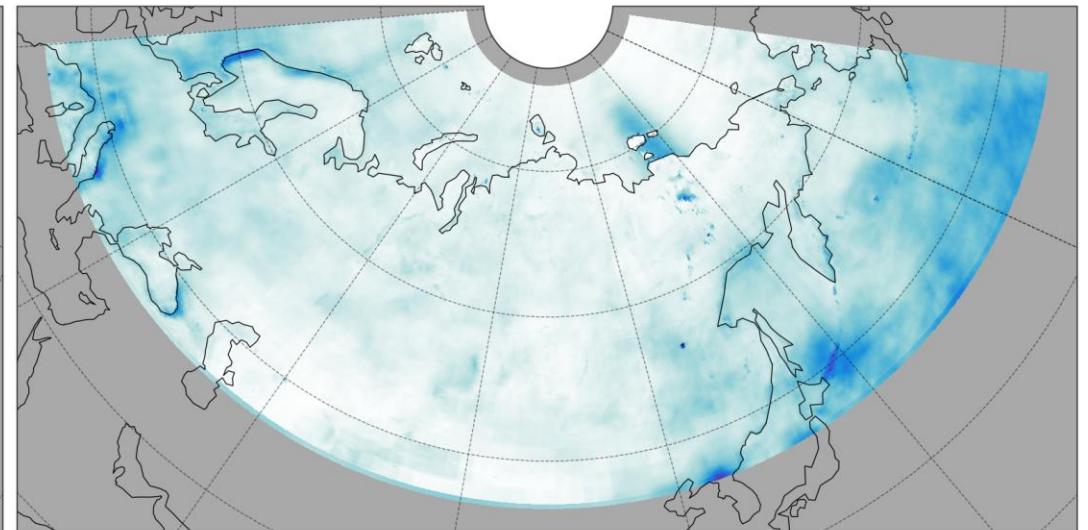
Линейная регрессия



Градиентный бустинг

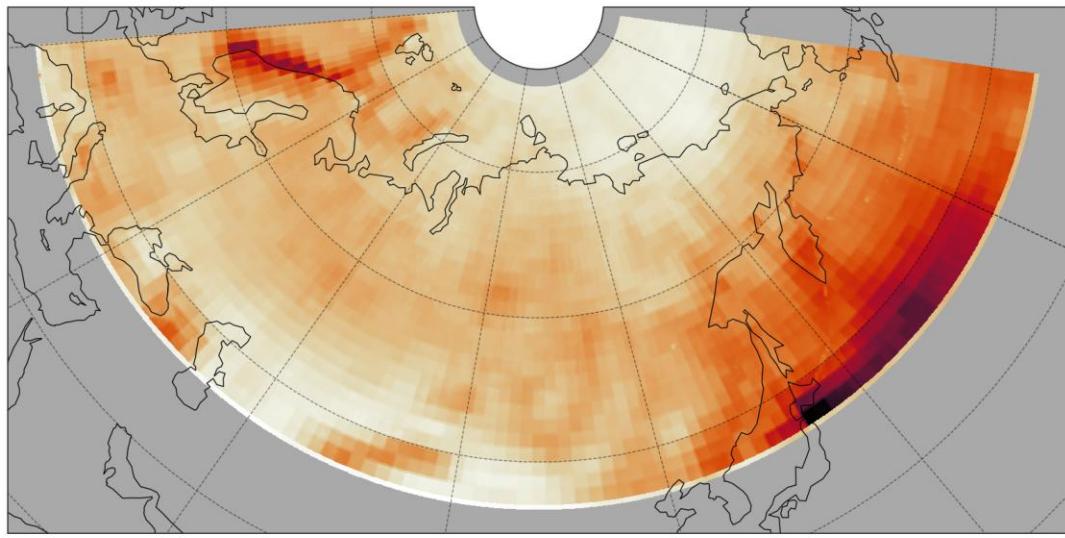


Случайный лес

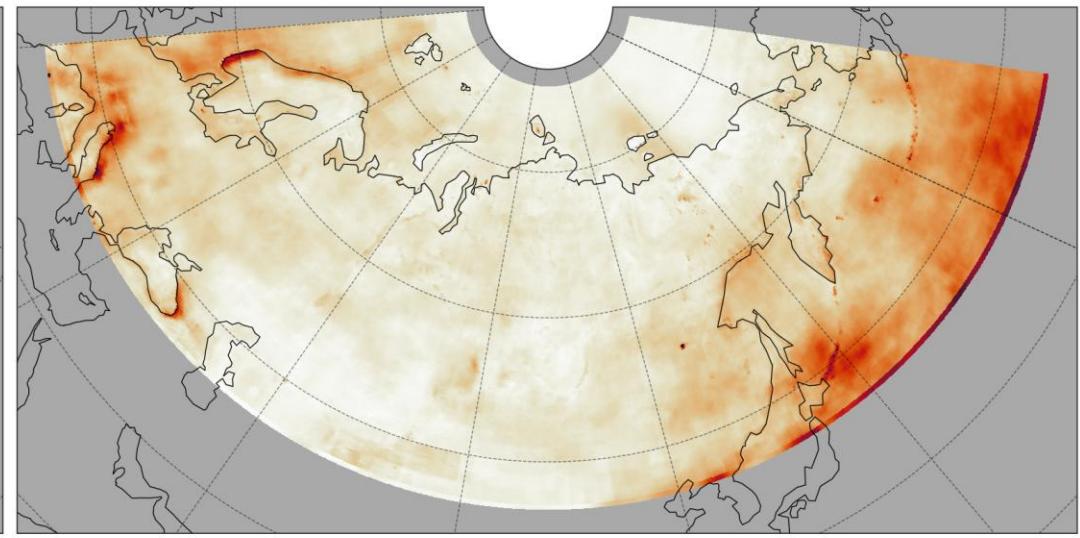


МАЕ осадков

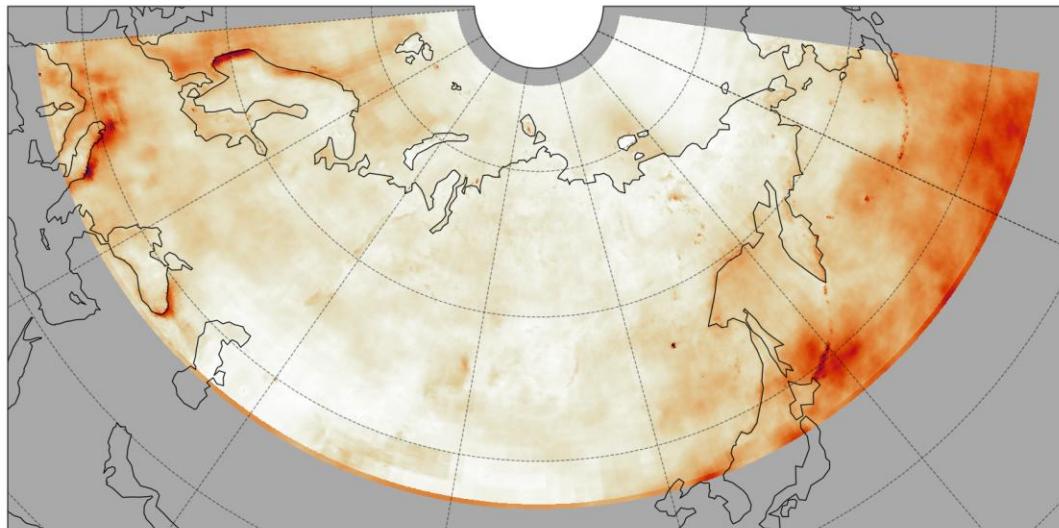
Исходные данные



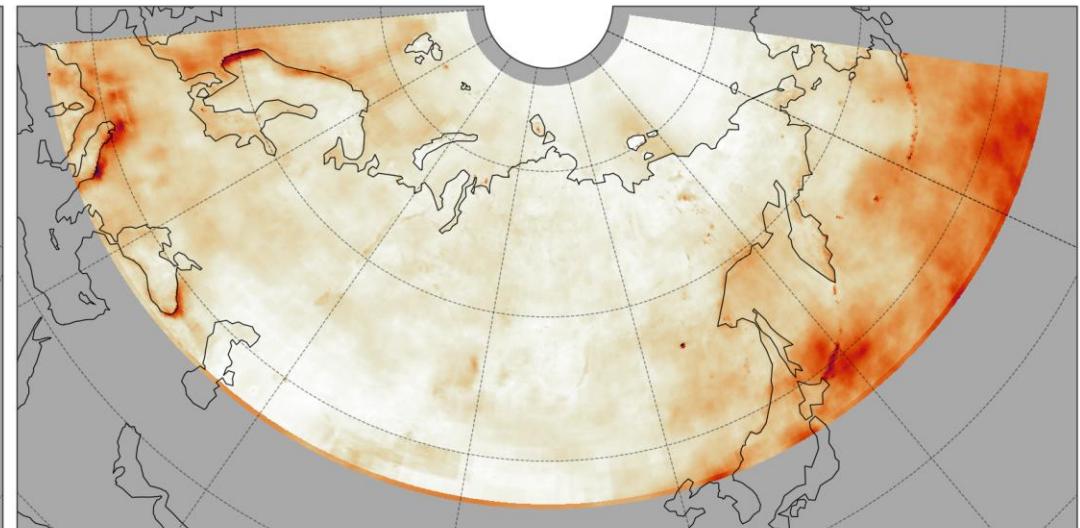
Линейная регрессия



Градиентный бустинг

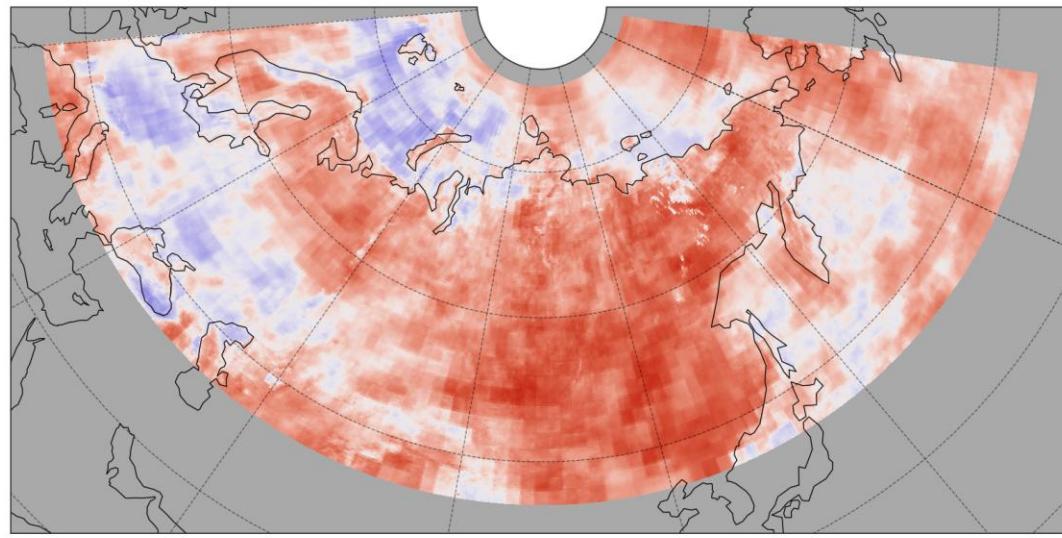


Случайный лес

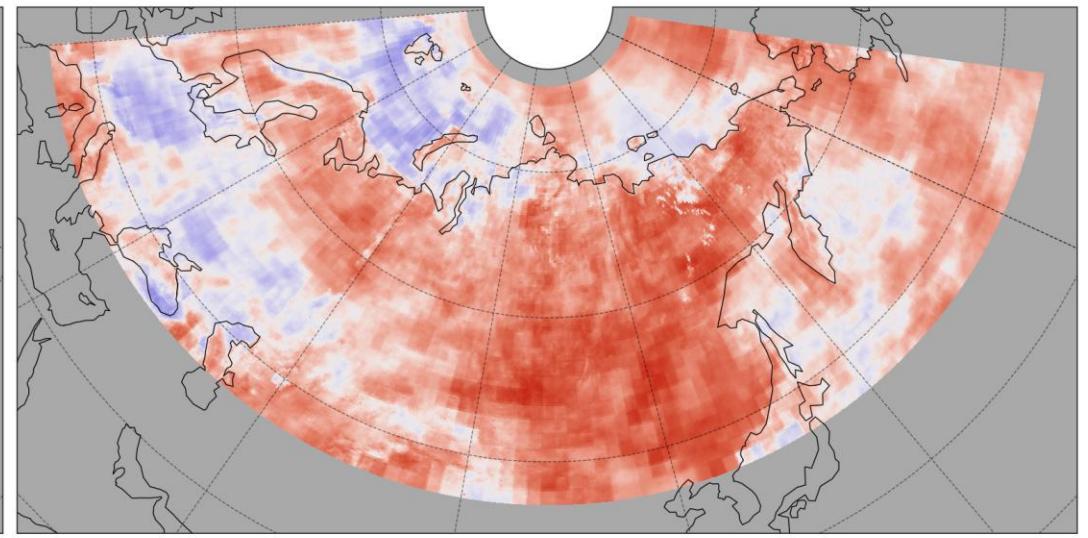


Корреляция осадков

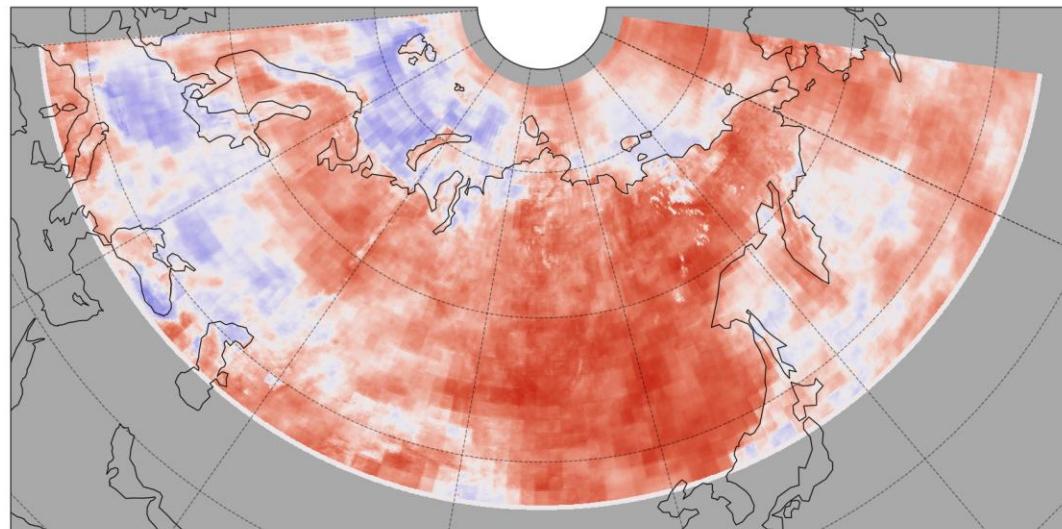
Исходные данные



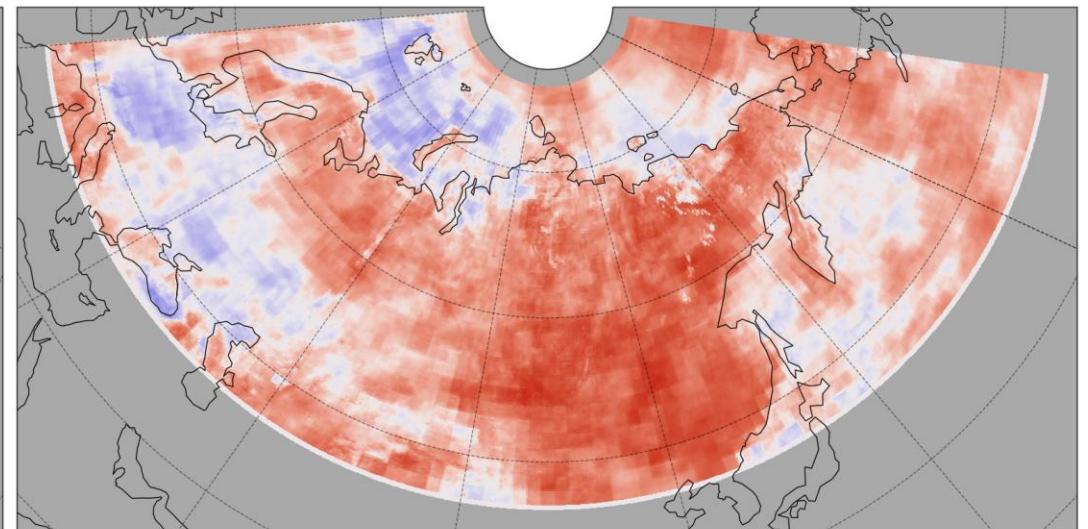
Линейная регрессия



Градиентный бустинг



Случайный лес



Сравнение с аналогичными исследованиями

| Исследование | Что корректировалось | Использованные методы | Результаты |
|----------------------|--|--|---|
| Seo and Ahn (2023) | ежедневное количество летних (май-сентябрь) осадков в Южной Корее из модели WRF | <ul style="list-style-type: none"> Квантильное преобразование (QM) Модель Long Short-Term Memory (LSTM) | RMSE: WRF_RAW = 1.10; WRF_QM = 0.17; WRF_LSTM = 0.69 R ² : WRF_RAW = 0.201; WRF_QM = 0.230; WRF_LSTM = 0.451 |
| Xu et al. (2019) | ежемесячные суммы осадков для территории Китая, полученные из ансамбля моделей North American Multi-Model Ensemble | Вейвлет-преобразование исходных данных + обучение модели опорных векторов (WSVM) и случайного леса (WRF) | Снижение RMSE на 18–40 мм (21–33%) Медиана корреляции по WSVM/WRF достигает 0.65 |
| Sengoz et al. (2023) | суммарные суточные осадки на Северную Америку из 8 моделей численного прогноза погоды | <ul style="list-style-type: none"> Сверточная нейросеть Полносвязная нейросеть | MAE: 0.79/0.78 мм/день (улучшение на 17%) RMSE: 2.53/2.54 мм/день (улучшение на 3%) |
| Li et al. (2023) | 3-часовые прогнозы осадков из модели FGOALS-f3-L на бассейн реки Чжуцзян | <ul style="list-style-type: none"> Классическое квантильное преобразование Улучшенный метод на основе QM (MPTT-QM) | MPTT-QM увеличил корреляцию с 0.02 (FGOALS) до 0.15 (для 14-дневного прогноза) и 0.13 (для 90-дневного) |
| Vandal et al. (2017) | ежедневные осадки на северо-восточную часть США, полученные из модели CCSM4 | <ul style="list-style-type: none"> PCA + Support Vector Regression (PCASVR) Elastic-Net ... | Elastic-Net – одна из самых лучших моделей с RMSE = 0.65 мм/день и корреляцией = 0.64 PCASVR – одна из самых худших моделей с RMSE = 1.10 мм/день и корреляцией = 0.33 |

Заключение

С помощью моделей машинного обучения успешно проведена коррекция ежемесячных сумм осадков, которая привела к значительному улучшению MAE и RMSE и к небольшому росту коэффициентов корреляции и детерминации. В целом предсказания всех трех моделей, использованных для коррекции – линейная регрессия, градиентный бустинг и случайный лес – схожи друг с другом, поэтому для коррекции ежемесячных сумм осадков возможно применение всех проанализированных в настоящем исследовании алгоритмов.

Для улучшения существующей методики коррекции рассматривается включение других метеорологических переменных, полученных моделью INM-CM5-0, в алгоритмы коррекции, например температуры или относительной влажности. Так модели коррекции будут находить зависимости между переменными, что способствует улучшению результата. Подбор модельных данных INM-CM5-0 и IMERG за больший интервал необходим для получения алгоритма коррекции, устойчивого к шумам и аномалиям и способного к большей обобщающей способности. Особый случай вызывают экстремальные осадки, для которых систематические смещения оказываются сильнее, чем у умеренных осадков. Для их коррекции возможно применение нейронных сетей, учитывающие сложные нелинейные взаимодействия между модельными осадками и осадками по наблюдениям.

Список литературы

1. Pradhan, R., Markonis, Y., Godoy, M., Villalba-Pradas, A., Andreadis, K., Nikolopoulos, E., Papalexiou, S., Rahim, A., Tapiador, F., & Hanel, M. (2022). Review of GPM IMERG performance: A global perspective. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2021.112754>.
2. Haider, A., Song, W., Wang, Y., Xiang, K., Chen, L., Feng, T., Linghu, S., & Alam, M. (2024). Validation of CRUTS v4.08, ERA5-Land, IMERG v07B, and MSWEP v2.8 Precipitation Estimates Against Observed Values over Pakistan. *Remote. Sens.*, 16, 4803. <https://doi.org/10.3390/rs16244803>.
3. Jiang, S., Wei, L., Ren, L., Zhang, L., Wang, M., & Cui, H. (2022). Evaluation of IMERG, TMPA, ERA5, and CPC precipitation products over mainland China: Spatiotemporal patterns and extremes. *Water Science and Engineering*. <https://doi.org/10.1016/j.wse.2022.05.001>.
4. Xie, W., Yi, S., Leng, C., Xia, D., Li, M., Zhong, Z., & Ye, J. (2022). The evaluation of IMERG and ERA5-Land daily precipitation over China with considering the influence of gauge data bias. *Scientific Reports*, 12. <https://doi.org/10.1038/s41598-022-12307-0>.
5. Mahmoud, M., Mohammed, S., Hamouda, M., & Mohamed, M. (2020). Impact of Topography and Rainfall Intensity on the Accuracy of IMERG Precipitation Estimates in an Arid Region. *Remote. Sens.*, 13, 13. <https://doi.org/10.3390/rs13010013>.
6. Xiong, W., Tang, G., Wang, T., , Z., & Wan, W. (2022). Evaluation of IMERG and ERA5 Precipitation-Phase Partitioning on the Global Scale. *Water*. <https://doi.org/10.3390/w14071122>.
7. Mahmoud, M.T.; Mohammed, S.A.; Hamouda, M.A.; Dal Maso, M.; Mohamed, M.M. Performance of the IMERG Precipitation Products over High-latitudes Region of Finland. *Remote Sens.* 2021, 13, 2073. <https://doi.org/10.3390/rs13112073>
8. Seo, G.-Y.; Ahn, J.-B. Comparison of Bias Correction Methods for Summertime Daily Rainfall in South Korea Using Quantile Mapping and Machine Learning Model. *Atmosphere* 2023, 14, 1057. <https://doi.org/10.3390/atmos14071057>
9. Xu, L., Chen, N., Zhang, X., Chen, Z., Hu, C., & Wang, C. (2019). Improving the North American multi-model ensemble (NMME) precipitation forecasts at local areas using wavelet and machine learning. *Climate Dynamics*, 53, 601-615. <https://doi.org/10.1007/s00382-018-04605-z>.
10. Sengoz, C., Ramanna, S., Kehler, S., Goomer, R., & Pries, P. (2023). Machine learning approaches to improve north american precipitation forecasts. *IEEE Access*, 11, 97664-97681.
11. Li, X.; Wu, H.; Nanding, N.; Chen, S.; Hu, Y.; Li, L. Statistical Bias Correction of Precipitation Forecasts Based on Quantile Mapping on the Sub-Seasonal to Seasonal Scale. *Remote Sens.* 2023, 15, 1743. <https://doi.org/10.3390/rs15071743>
12. Vandal, T., Kodra, E., & Ganguly, A. (2017). Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theoretical and Applied Climatology*, 137, 557-570. <https://doi.org/10.1007/s00704-018-2613-3>.