

Восстановление пропущенных значений в спутниковых снимках методами машинного обучения на примере данных температуры поверхности Земли

Сарафанов М.И., Казаков Э.Э.



Введение

Актуальность

Спутниковые снимки стали неотъемлемой частью исследований в науках о Земле. Однако, использовать получаемые спутниковые снимки во многих случаях затруднительно, или даже невозможно, по причине большого количества пропусков.

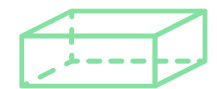
Пропуски в спутниковых снимках препятствуют процессам мониторинга природных объектов и осложняют анализ полученных данных. Для решения проблемы заполнения пропусков существует множество подходов. Ниже представлена одна из некоторых классификаций методов заполнения пропусков в пространственных данных:



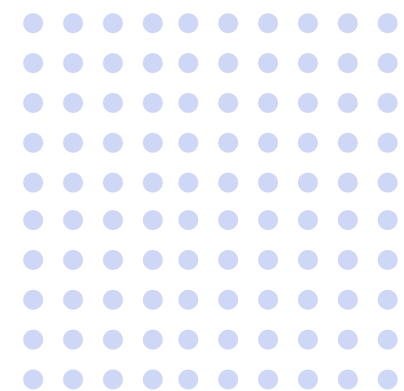
Методы на основе анализа временных рядов;



Методы, использующие пространственную информацию;



Подходы с использованием пространственно-временных закономерностей.



Цель работы

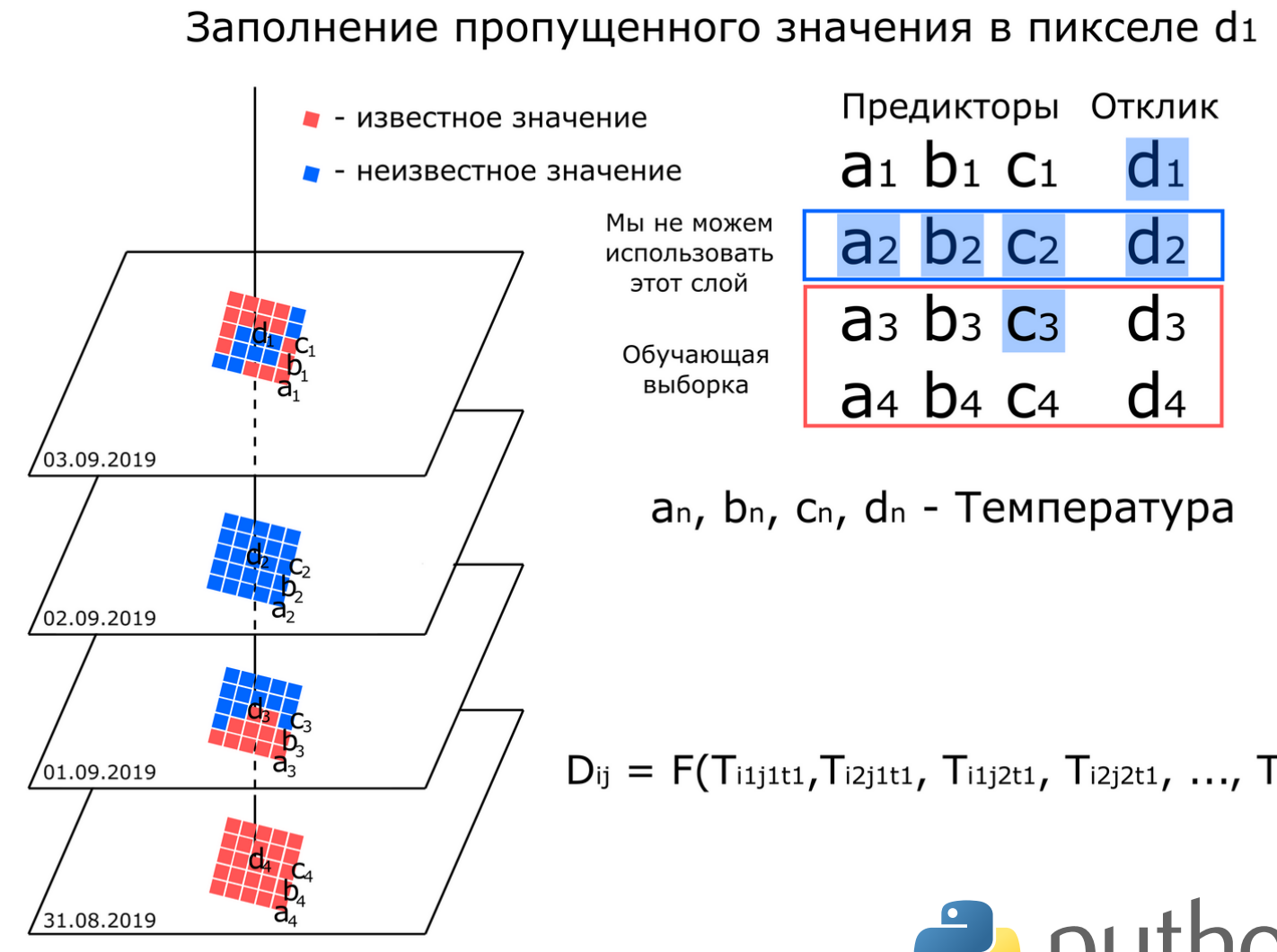
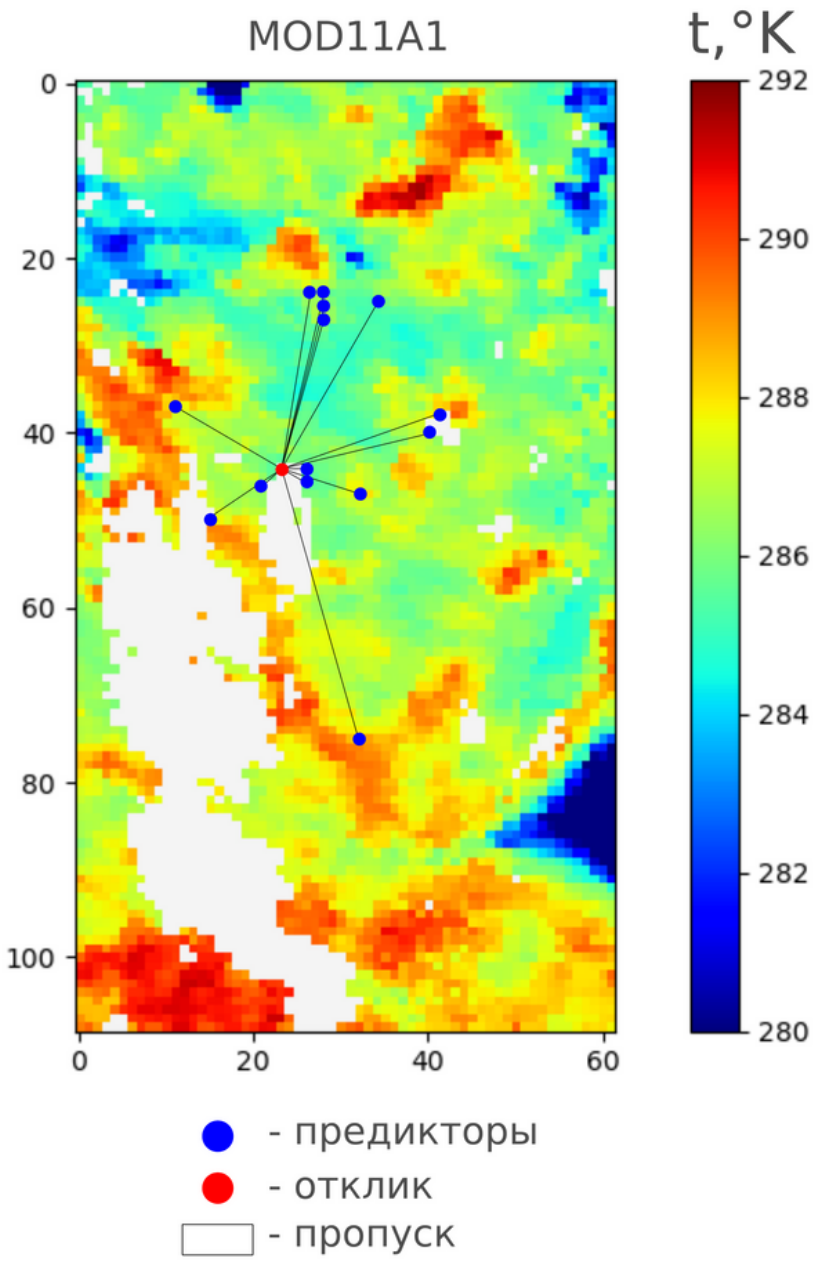


В рамках нашего исследования мы предлагаем метод восстановления данных в пропусках. Верификация модели была произведена на данных температуры поверхности Земли, получаемых сенсором MODIS (спутник Terra) и SLSTR (Sentinel-3). Использовались однодневные композиты (продукт MOD11A1), мгновенные данные MODIS (продукт MOD11_L2) и мгновенные данные Sentinel (продукт Sentinel-3 LST)

Методы

Мы предполагаем, что существует возможность восстановить значения в пропуске, используя известные значения на том же снимке, основываясь на наблюдениях этих взаимосвязей в прошлые периоды. Аппроксимирующую функцию можно построить с помощью алгоритмов машинного обучения. Так, можно использовать **метод опорных векторов**, ЛАССО регрессию, случайный лес или k-ближайших соседей.

Предикторами могут быть выбраны все известные пиксели на снимке, 100 случайных известных пикселей, или некоторое количество (в рамках данной работы - 40) ближайших к пропуску пикселей из того же типа земель, что и пропущенный пиксель.



РЕЗУЛЬТАТЫ

Валидация осуществлялась на 3х продуктах:

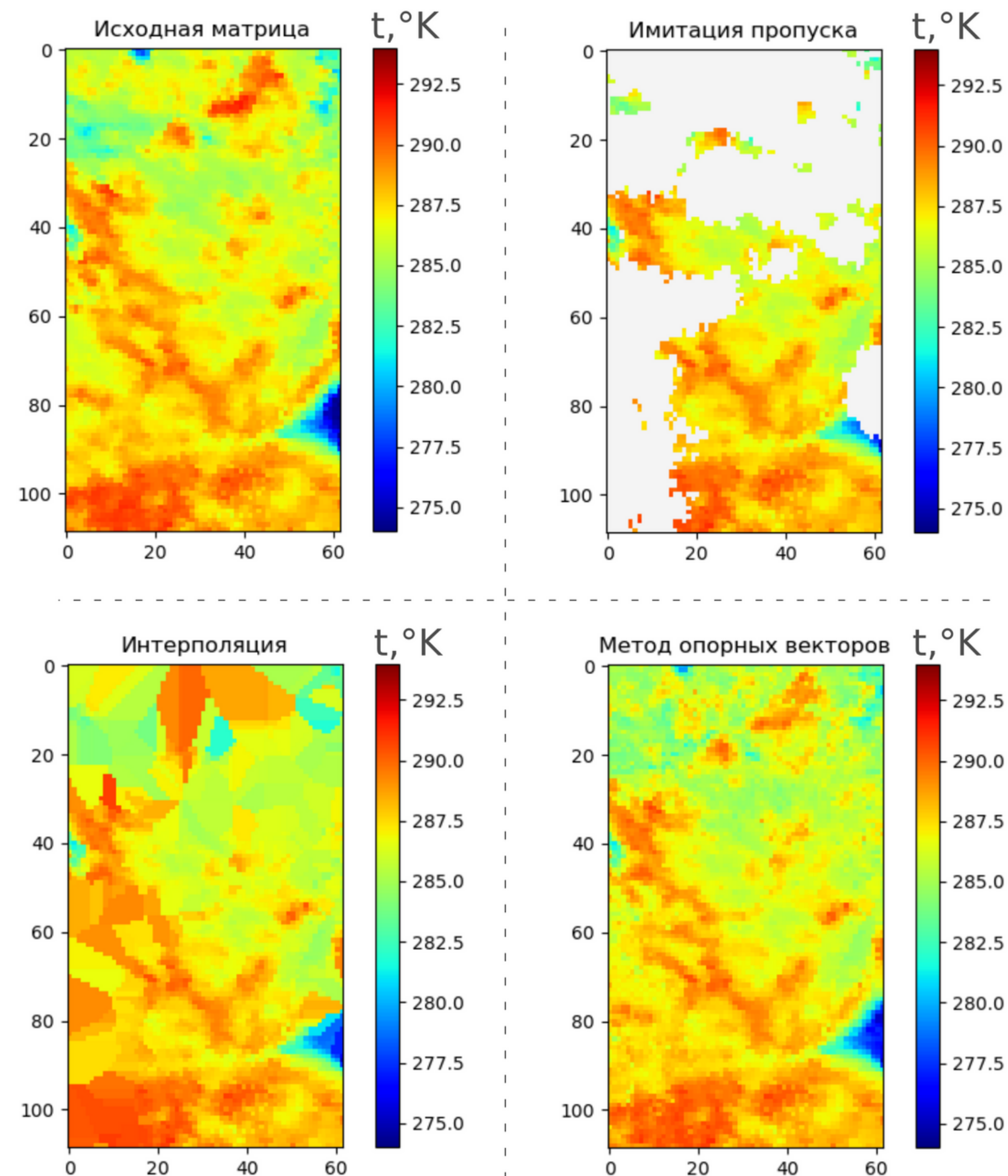
- 1) MOD11A1 (однодневные композиты с сенсора MODIS)
- 2) MOD11_L2 (мгновенные данные с сенсора MODIS)
- 3) Sentinel-3 LST (мгновенные данные с сенсора SLSTR Sentinel-3)

3 тестовые территории:

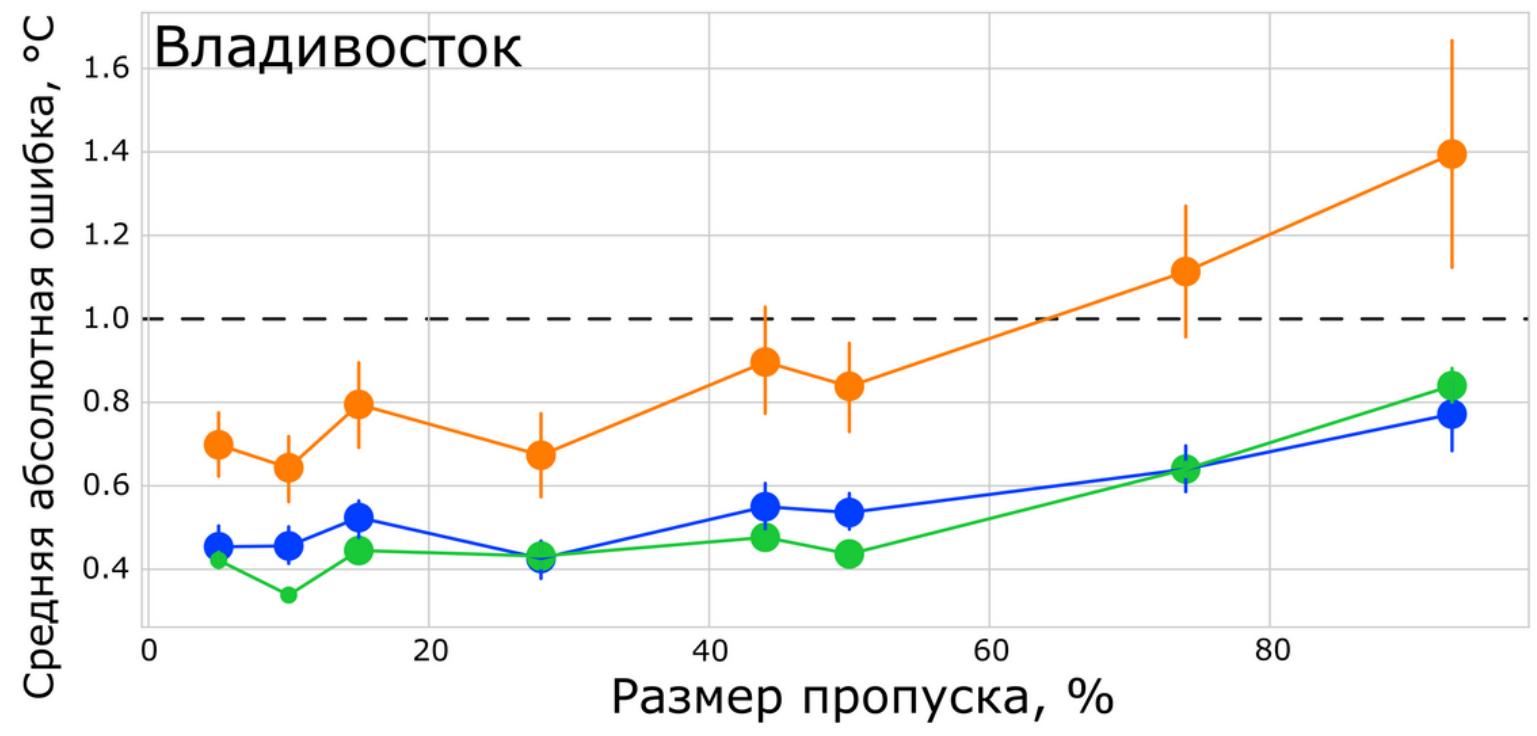
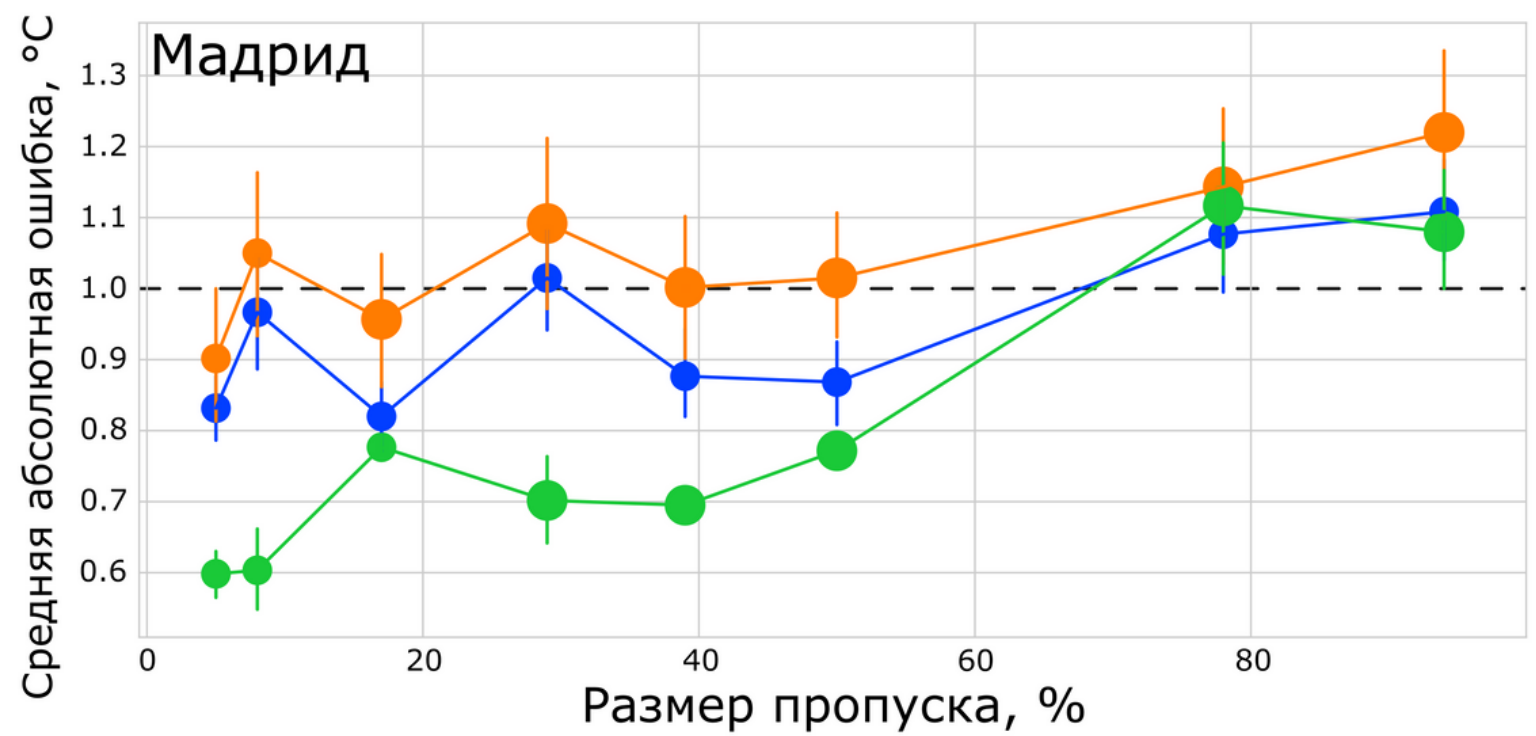
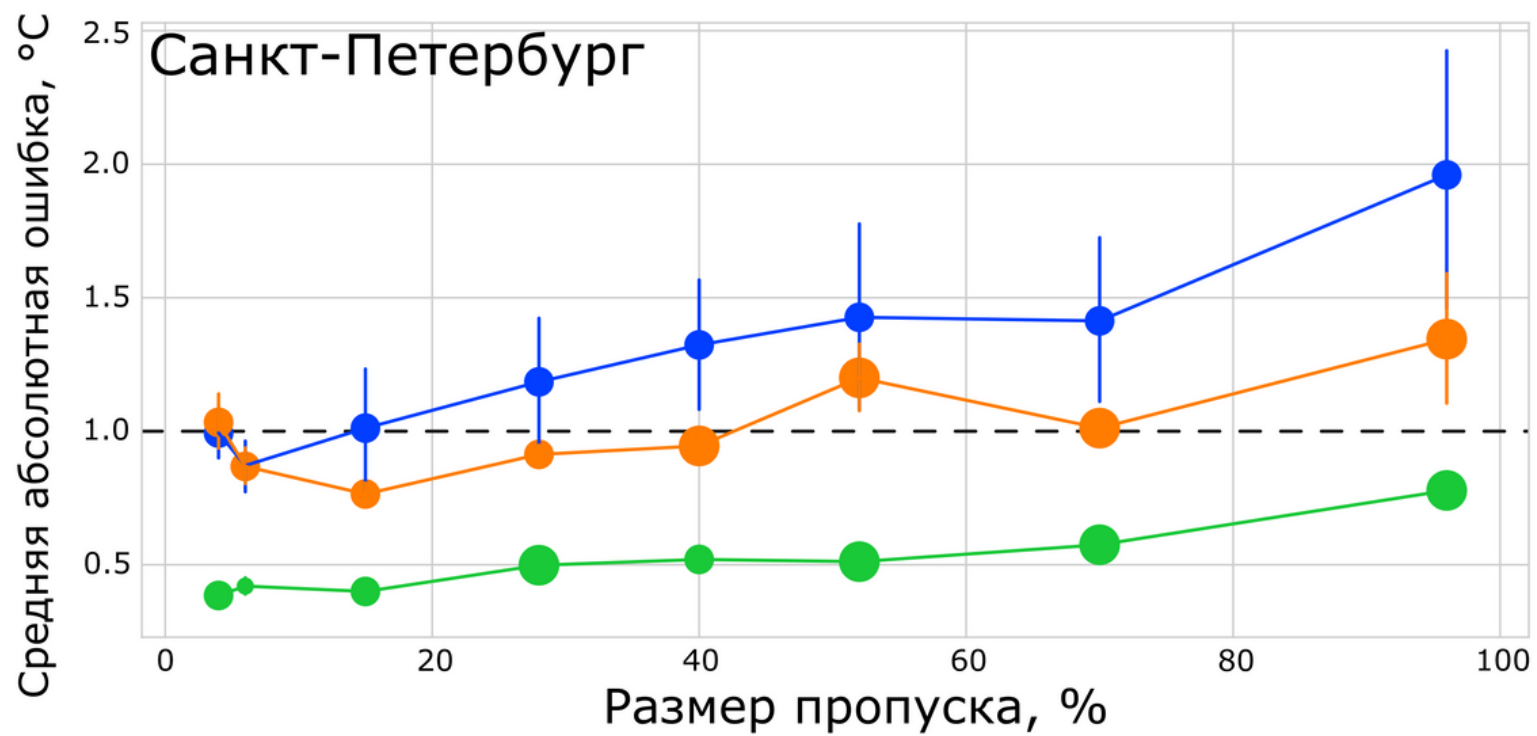
- a) Санкт-Петербург
- b) Мадрид
- c) Владивосток

Различные типы пропусков

Для каждого продукта на каждой тестовой территории выбирались 6 снимков, в которых генерировались 8 типов пропусков различной формы и размеров (от 4 до 96% процентов на снимке).



Зависимость величины ошибки от размера пропуска и размаха значений в нём

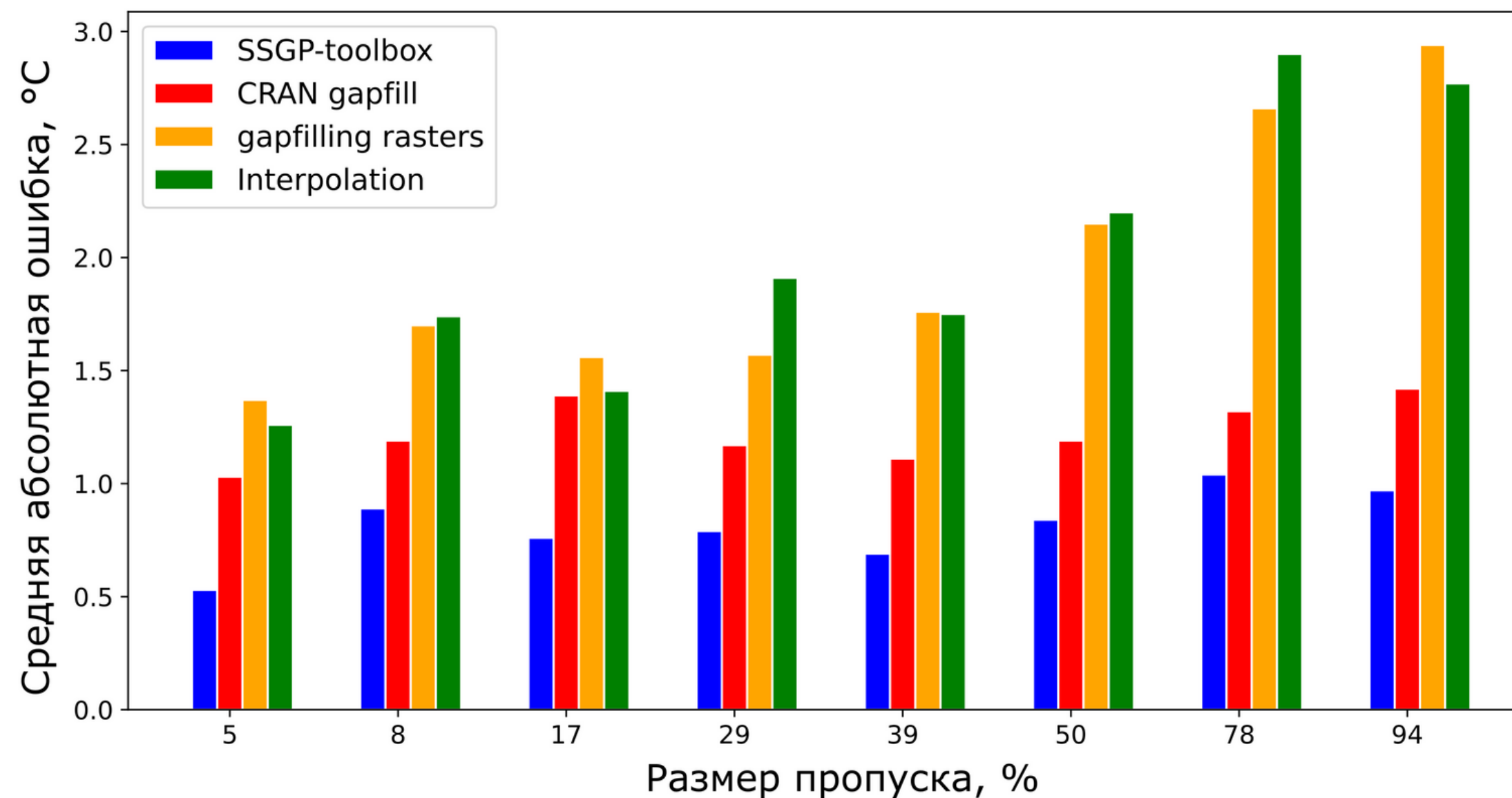


Продукт	Размах значений в пропуске, °C
● Sentinel-3 LST	● <8
● MOD11_L2	● 8-16
● MOD11A1	● >16

"Усы" показывают стандартную ошибку

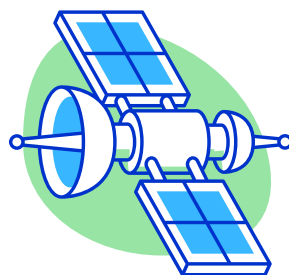
Сравнение с аналогами

Среди open-source решений мы производили сравнение с алгоритмами восстановления пропусков на языке R - "CRAN gapfill" и "gapfilling rasters"

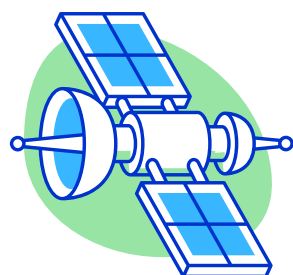


На рисунке показаны результаты сравнения алгоритмов восстановления пропусков на снимках продукта MOD11A1 для Мадрида. Из графика видно, что предложенный нами алгоритм (Simple Spatial Gapfilling Processor toolbox (SSGP-toolbox)) превосходит в точности своих конкурентов.

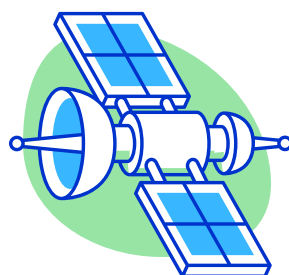
Заключение



Был реализован модуль для заполнения пропусков в данных дистанционного зондирования (ДДЗ) на языке Python и выложен в открытый доступ через репозиторий github;

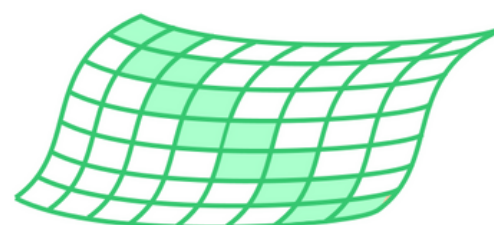


Алгоритм превосходит в точности рассмотренные аналоги. При этом "CRAN garfill" считается одним из наиболее точных решений для задачи восстановления пропусков в ДДЗ;



Алгоритм позволяет восстанавливать значения в температурных полях со средним значением средней абсолютной ошибки менее чем 1 градус, при этом оценивает неопределенность прогноза;

Simple Spatial Gapfilling Processor - toolbox



<https://github.com/Dreamlone/SSGP-toolbox>

