

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОЦЕНКИ ЭКСТРЕМАЛЬНО НИЗКИХ ТЕМПЕРАТУР В ПРИЗЕМНОМ СЛОЕ ВОЗДУХА В ЗИМНЕЕ ВРЕМЯ ПО ДАННЫМ СПУТНИКОВЫХ СВЧ НАБЛЮДЕНИЙ

Мордвин Е.Ю., Лагутин А.А., Волков Н.В.
Алтайский государственный университет, Барнаул, Россия

Актуальность

Важнейшей характеристикой климатической системы является температура приземного слоя воздуха (**T2M**), в частности, ее возможные экстремальные значения.

В труднодоступных регионах редкая сеть наземных пунктов наблюдения.

Спутниковые СВЧ радиометры восстанавливают **T2M** даже при наличии облачности и имеют приемлемое пространственное и временное разрешение.

Однако для получения хорошей точности **T2M** требуется коррекция алгоритма или результатов алгоритма.

Цели и задачи

Цель работы: разработка метода коррекции данных о температуре приземного слоя воздуха, полученных в результате работы стандартного алгоритма MIRS v3.0 в зимний период времени с использованием методов машинного обучения.

Задачи:

1. Формирование архива данных MIRS v3.0.
2. Формирование архива метеонаблюдений и климатических данных ERA5 и GFS.
3. Подбор регрессоров и обучение моделей.
4. Разработка программного комплекса для оперативной коррекции информации о **T2M** по данным MIRS v3.0.

Алгоритм MIRS

MIRS v3.0 (Microwave Integrated Retrieval System)

Обработка: AMSU, MHS, SSMI/S, ATMS (спутники серии NOAA, METOP, DMSR, NPP/JPSS).

Разработан в NOAA (National Oceanic and Atmospheric Administration) для внутреннего использования и работы с данными, полученными в режиме прямого вещания (**Direct Broadcast, DB**).

В основе MIRS – метод одномерного вариационного поиска (**1DVAR**), который представляет собой итеративный физический алгоритм.

Приборы:

AMSU/MHS (Advanced Microwave Sounding Unit/Microwave Humidity Sounder)

Спутники серии **NOAA** (с 1998 года) и **MetOp** (с 2006 года), всего **7**.

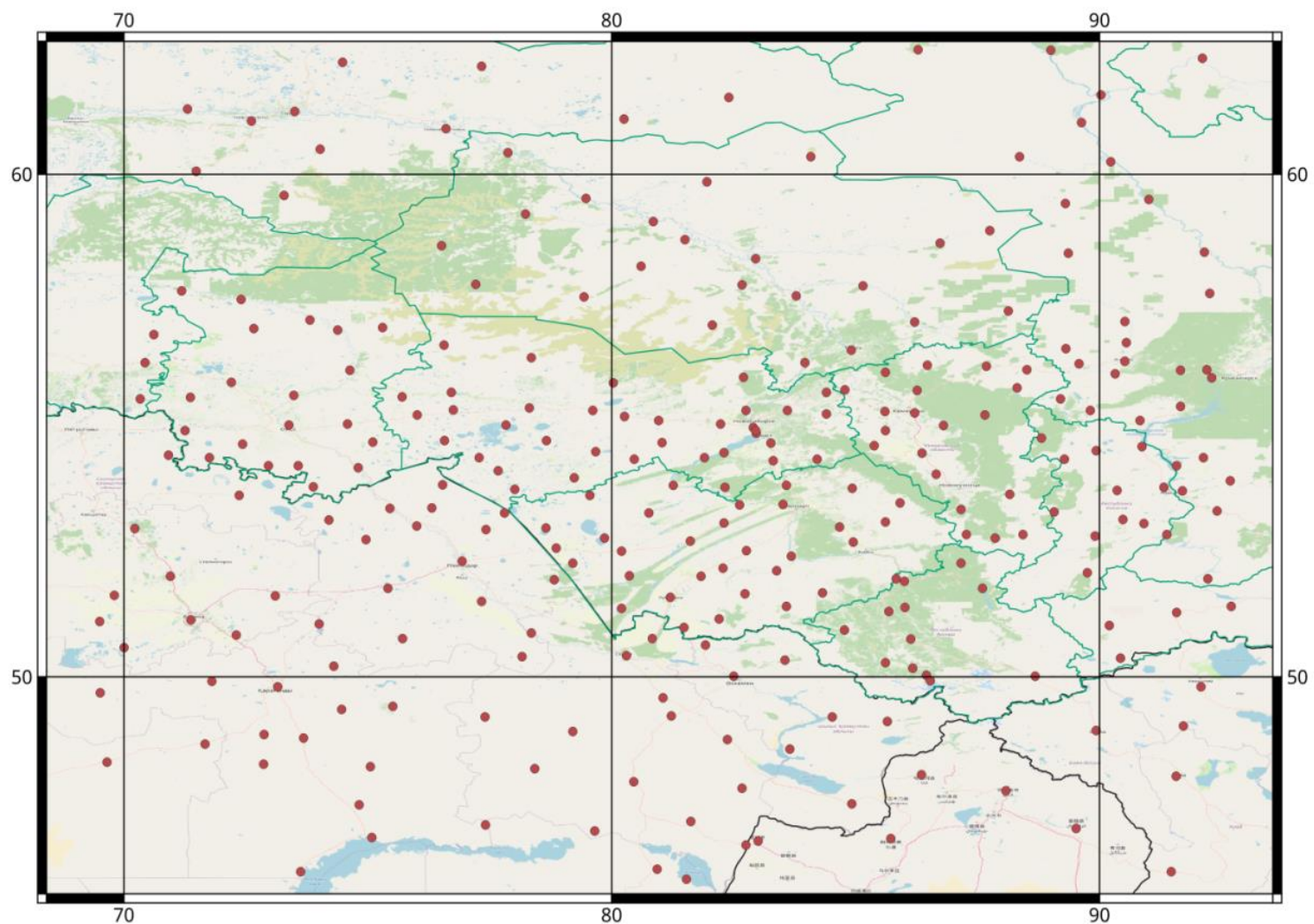
Диапазон **23÷190** ГГц.

ATMS (Advanced Technology Microwave Sounder)

Спутники: **Suomi-NPP**, **NOAA-20** и **NOAA-21**.

22 спектральных канала в диапазоне **23÷190** ГГц

Пространственное распределение метеостанций на территории юга Западной Сибири



Выборка данных

- 250 метеостанций;
- период октябрь 2022 - март 2023 + декабрь 2020 г;
- ~60000 наблюдений для каждого спутника.

Данные разделены на диапазоны по высоте:

+ > 200 м; 200-500м; 500-800 м; > 800 м.

+ обучение/проверка в соотношении 75/25.

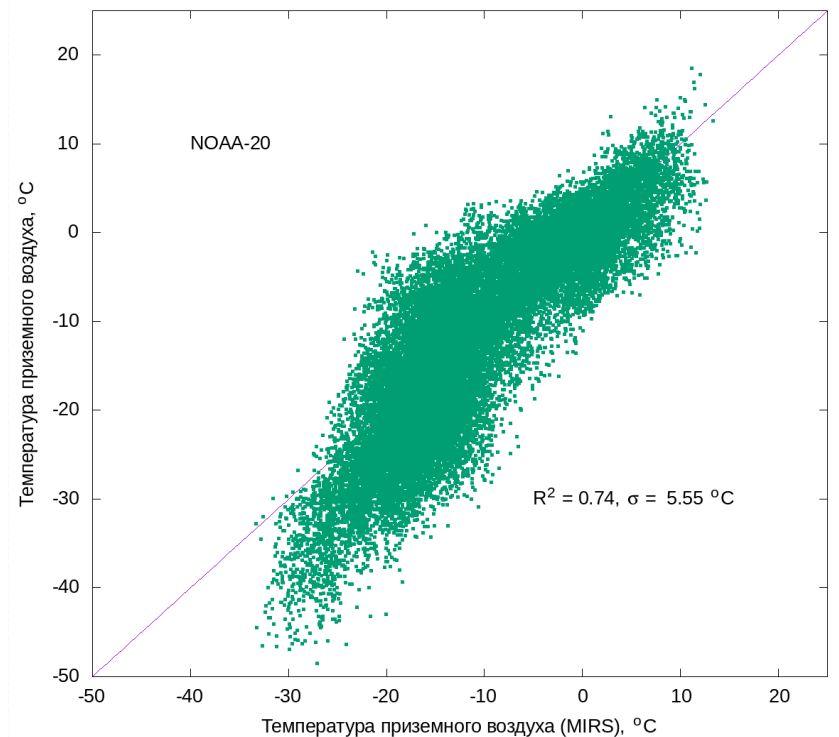
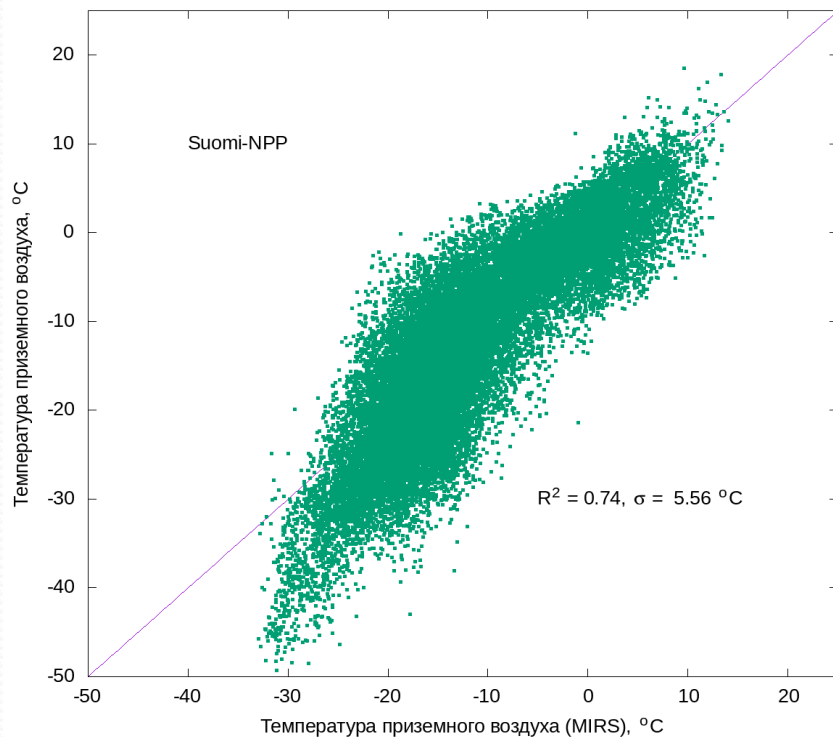
Источники данных

- T_2M – искомая величина, температура приземного слоя воздуха по данным наблюдений ВМО;
- $TAIR$ – температура приземного слоя воздуха по данным MIRS;
- $TSURF$ – температура подстилающей поверхности по данным MIRS;
- TPW – полное влагосодержание атмосферы по данным MIRS;
- $\Delta PSURF$ – вариация атмосферного давления на уровне поверхности, найденная как разница между климатической нормой (по данным ERA5) и давлением на момент пролета спутника (по данным GFS);
- HGT – высота поверхности относительно уровня моря по данным GFS;
- $T_{2M_{MIN}}$ – климатический минимум температуры приземного слоя воздуха за последние 20 лет, полученный по данным ERA5;
- U_{10m} – зональная компонента скорости ветра по данным GFS;
- V_{10m} – меридиональная компонента скорости ветра по данным GFS.

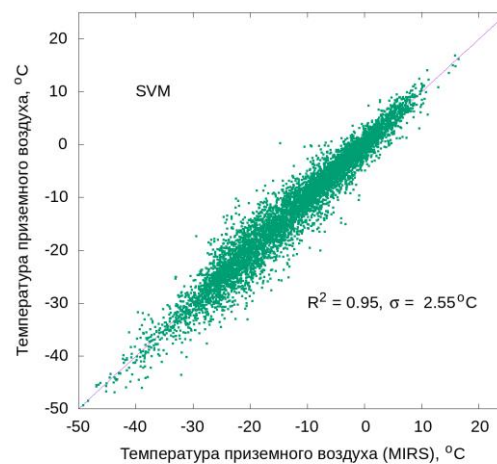
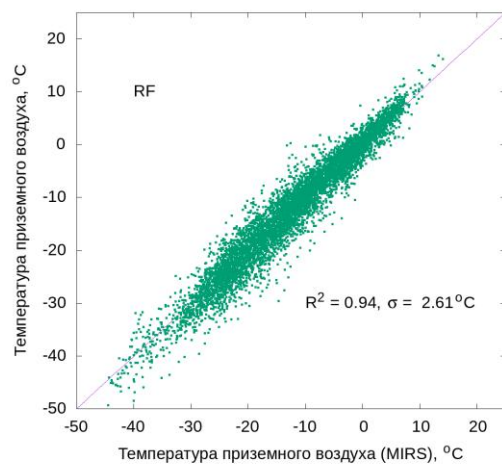
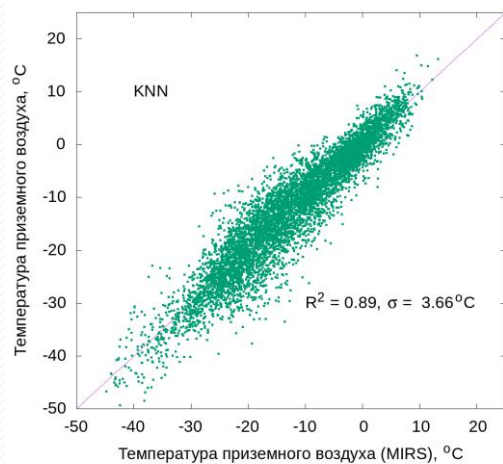
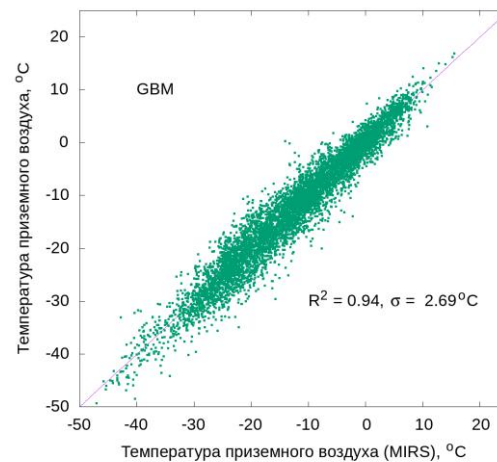
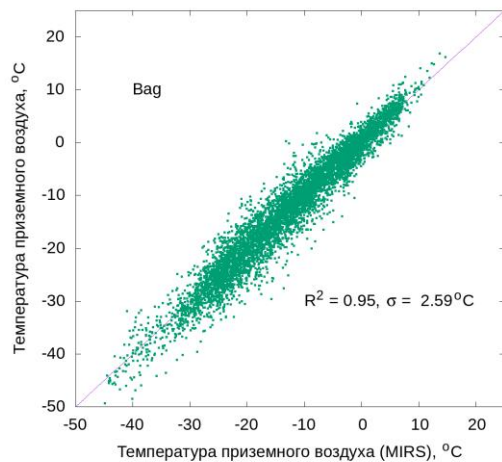
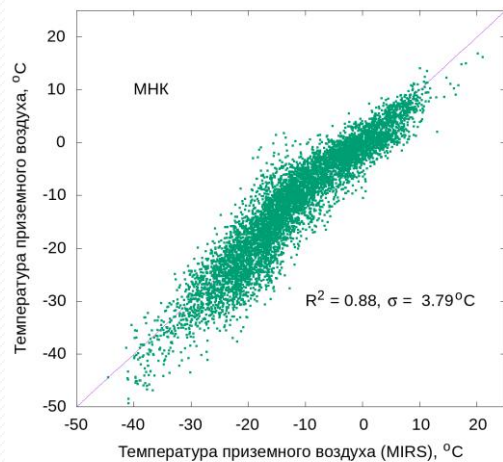
Методы

- **МНК** – метод наименьших квадратов;
- **KNN** – k-ближайших соседей;
- **SVM** – метод опорных векторов;
- **Bag** – бэггинг;
- **GBM** – градиентный бустинг;
- **RF** – случайный лес.

Диаграммы рассеяния для T2M по данным алгоритма MIRS и результатов наземных метеонаблюдений для метеостанций, находящихся на высотах 0-200 м.



Диаграммы рассеяния для скорректированной T2M по данным ATMS/Suomi-NPP и наземных наблюдений метеостанций, находящихся на высотах 0-200 м.



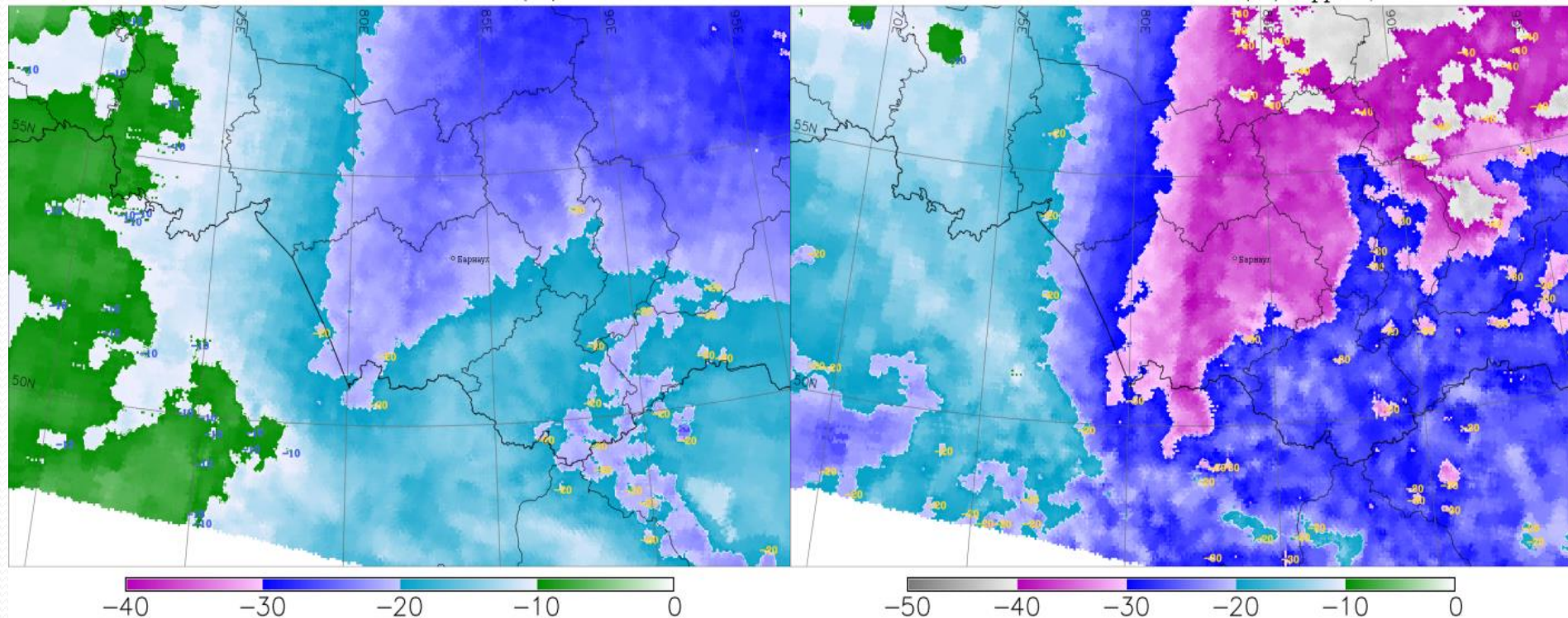
Проверка качества полученных моделей

Группа	МНК		KNN		Bag		RF		GBM		SVM	
	R ²	σ, °C	R ²	σ, °C	R ²	σ, °C	R ²	σ, °C	R ²	σ, °C	R ²	σ, °C
ATMS/Suomi-NPP												
> 200 м	0.88 / 3.79		0.89 / 3.66		0.95 / 2.59		0.94 / 2.61		0.94 / 2.69		0.95 / 2.55	
200-500 м	0.79 / 4.77		0.80 / 4.57		0.88 / 3.61		0.88 / 3.60		0.89 / 3.44		0.87 / 3.78	
500-800 м	0.74 / 4.96		0.79 / 4.49		0.83 / 4.01		0.84 / 3.97		0.85 / 3.85		0.85 / 3.83	
< 800 м	0.73 / 5.28		0.80 / 4.52		0.84 / 4.08		0.84 / 4.05		0.85 / 3.85		0.85 / 3.90	
ATMS/NOAA-20												
> 200 м	0.87 / 3.75		0.89 / 3.62		0.94 / 2.69		0.94 / 2.60		0.95 / 2.62		0.95 / 2.51	
200-500 м	0.79 / 4.73		0.81 / 4.51		0.88 / 3.62		0.87 / 3.72		0.88 / 3.51		0.87 / 3.72	
500-800 м	0.73 / 4.91		0.79 / 4.52		0.81 / 4.12		0.82 / 3.99		0.84 / 3.95		0.84 / 3.91	
< 800 м	0.73 / 5.21		0.80 / 4.57		0.83 / 4.18		0.84 / 4.08		0.85 / 3.91		0.85 / 3.89	

Температура приземного слоя воздуха по данным ATMS/Suomi-NPP на 21 января 2023 года 20:29 UTC.

Оригинал

Коррекция



Результаты и выводы

1. Показано, что спутниковые оценки температуры приземного слоя воздуха по данным алгоритма MIRS (СВЧ-радиометры ATMS/Suomi-NPP и ATMS/NOAA-20) завышены относительно результатов наземных наблюдений, причем для температур ниже $-30\text{ }^{\circ}\text{C}$ разница может достигать до $15\text{ }^{\circ}\text{C}$.
2. Создан программный комплекс для коррекции данных MIRS, который внедрен в оперативной цепочку обработки спутниковых данных в лаборатории космического мониторинга и вычислительных технологий АлтГУ (совместно с ФИЦ ИВТ).
3. Показано, что параметры качества полученных моделей, основанных на деревьях решений, позволяют достичь полуторакратного уменьшения дисперсии относительно методов MNK и KNN.
4. Установлено, что лучший результат показал метод опорных векторов, для которого $R^2 = 0.95$, а $\sigma = 2.5\text{ }^{\circ}\text{C}$. В частности SVM имеет меньшую разницу между восстановленными данным и тестовыми значениями в области низких температур (менее $-30\text{ }^{\circ}\text{C}$).



Спасибо за внимание!