



SMAP sea surface salinity improvement in the Arctic region using machine learning approaches

Savin A.
Krinitskiy M.
Osadchiev A.

DLCP, 2023

Почему соленость морской воды важна?

Соленость + температура → глобальная система течений

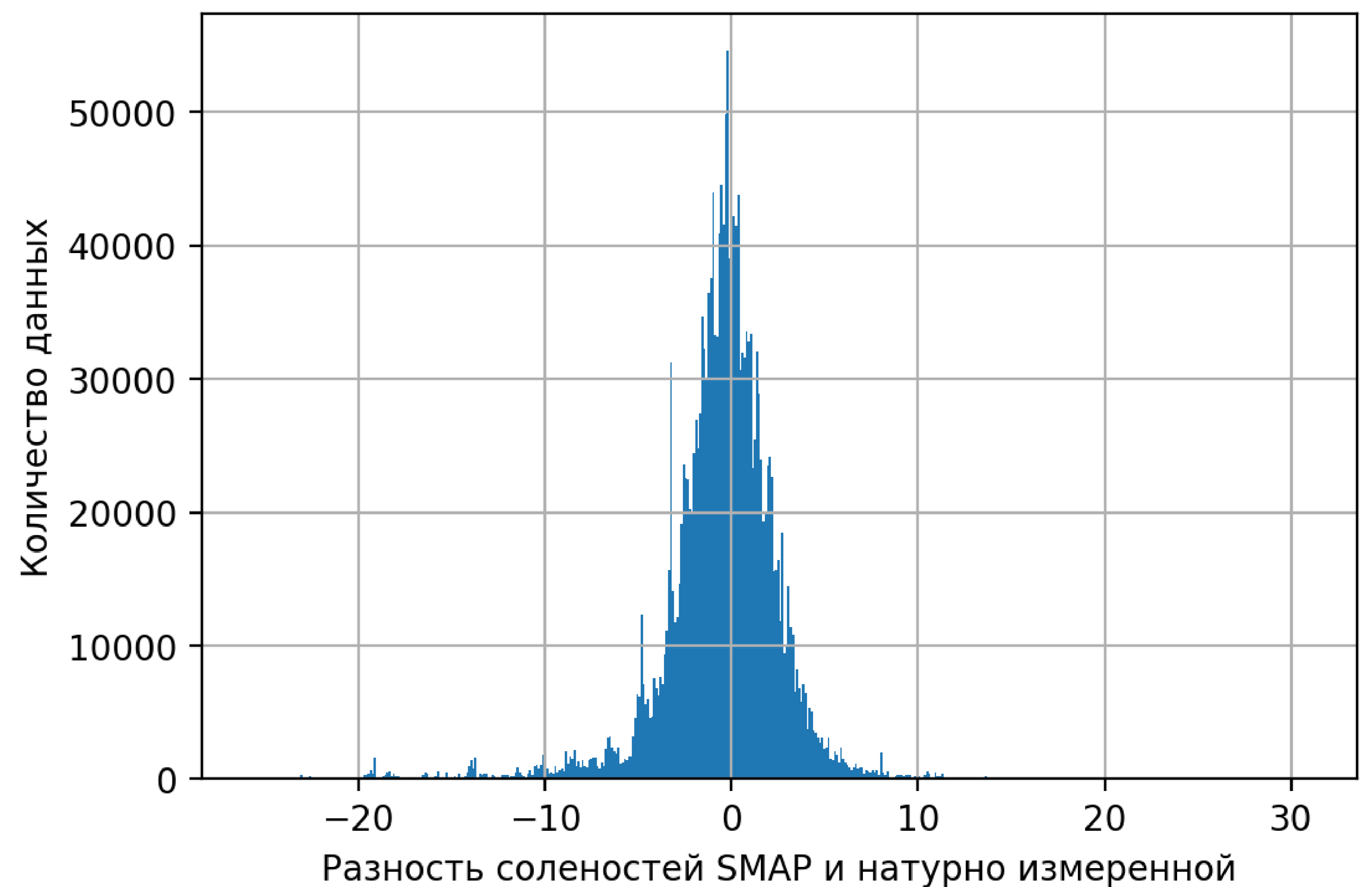
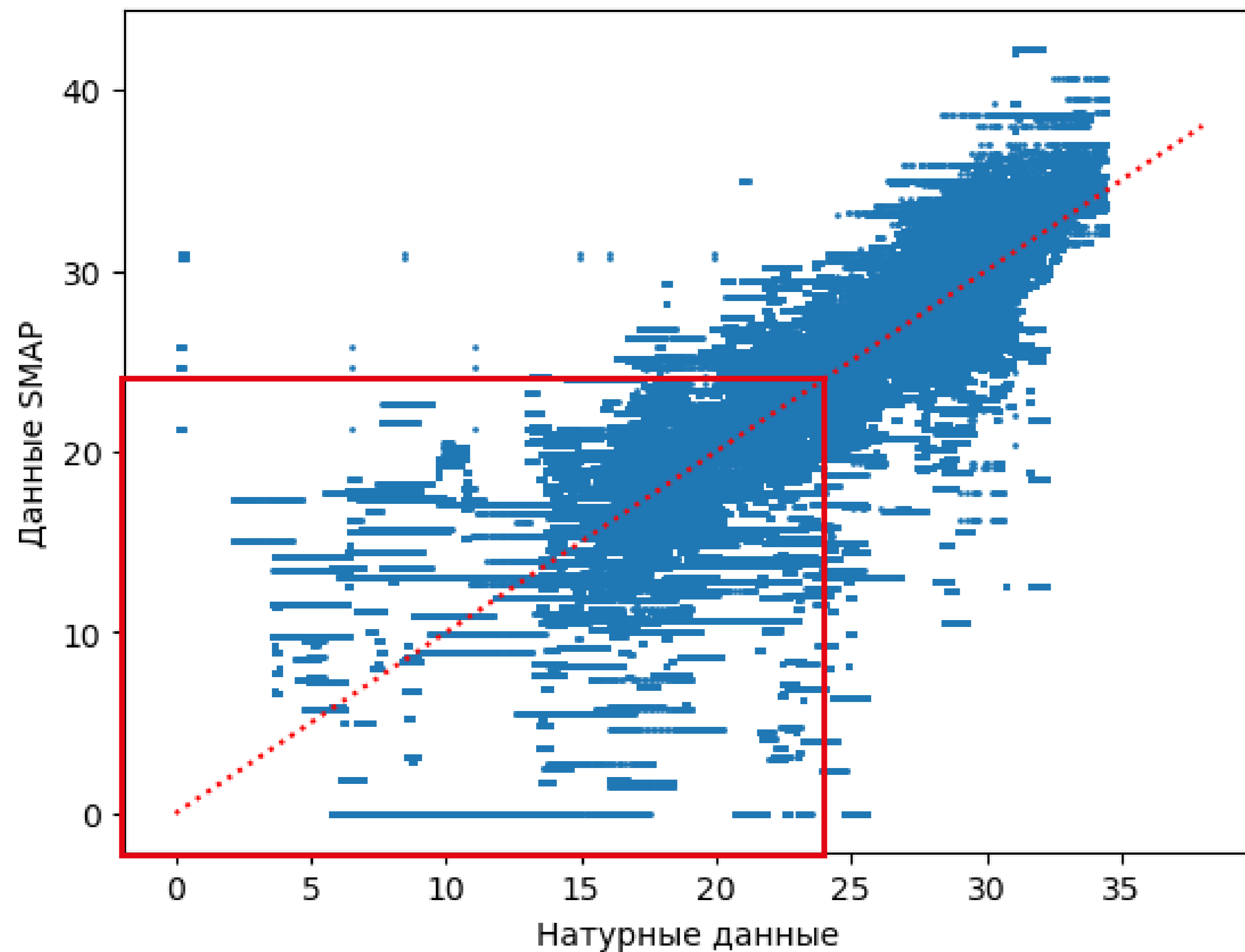
Соленость как индикатор физико-химических процессов

Соленость как определяющая характеристика физических переменных

Проблемы

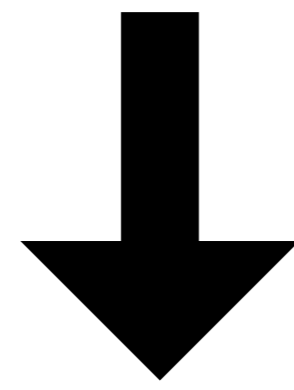
Особенности региона — низкие температура и соленость, влияние речного стока

corr = 0.82, RMSE = 3.21

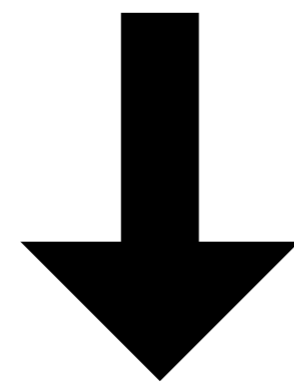


Основные существующие решения

Построение и/или улучшение модели
диэлектрической проницаемости



Коэффициенты отражения, излучения



Радиояркостьная температура

Alexandre Supply, Jacqueline Boutin
“New insights into SMOS sea surface salinity
retrievals in the Arctic Ocean”

RMSE: 1.46 -> 1.28

Машинное обучение

Eunna Jang , Young Jun Kim , Jungho Im &
Young-Gyu Park (2021)

“Improvement of SMAP sea surface salinity
in river-dominated oceans using
machine learning approaches”

RMSE: 0.774 -> 0.203, 0.556

Цель работы

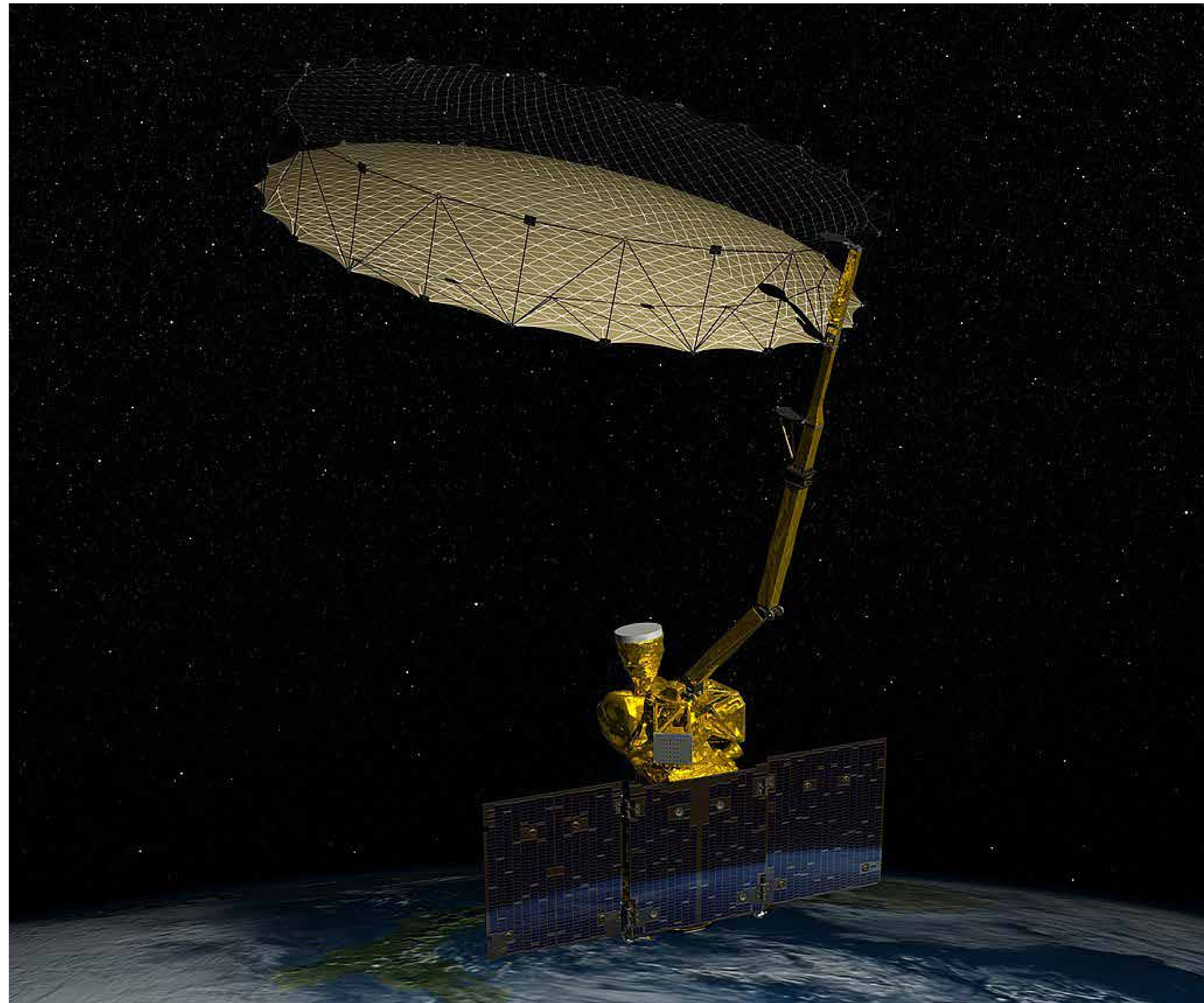
Улучшение алгоритма спутниковой солености поверхности океана для морей российской Арктики по спутниковым данным и данным натурных измерений с использованием методов машинного обучения

Спутниковые данные

Спутник SMAP,
Данные RSS V4.0 уровня L2,
NASA PO.DAAC

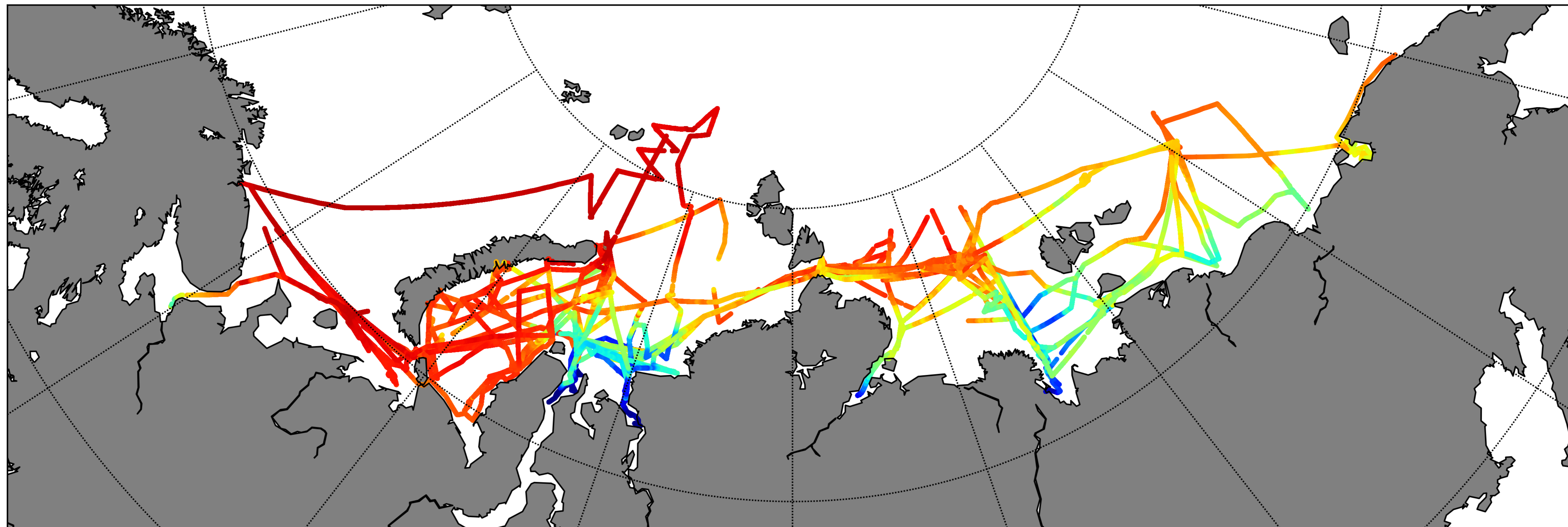
Сторонние данные

Данные Канадского метеоро-
логического центра (СМС),
кросс-калиброванного мульти-
платформенного анализа
ветра (ССМР),
Национальных центров эколо-
гического прогнозирования
Соединенных Штатов (NOAA
NCER)



Натурные данные

Данные собраны в экспедициях ИО РАН за 2015-2021 гг.



5

10

15

20

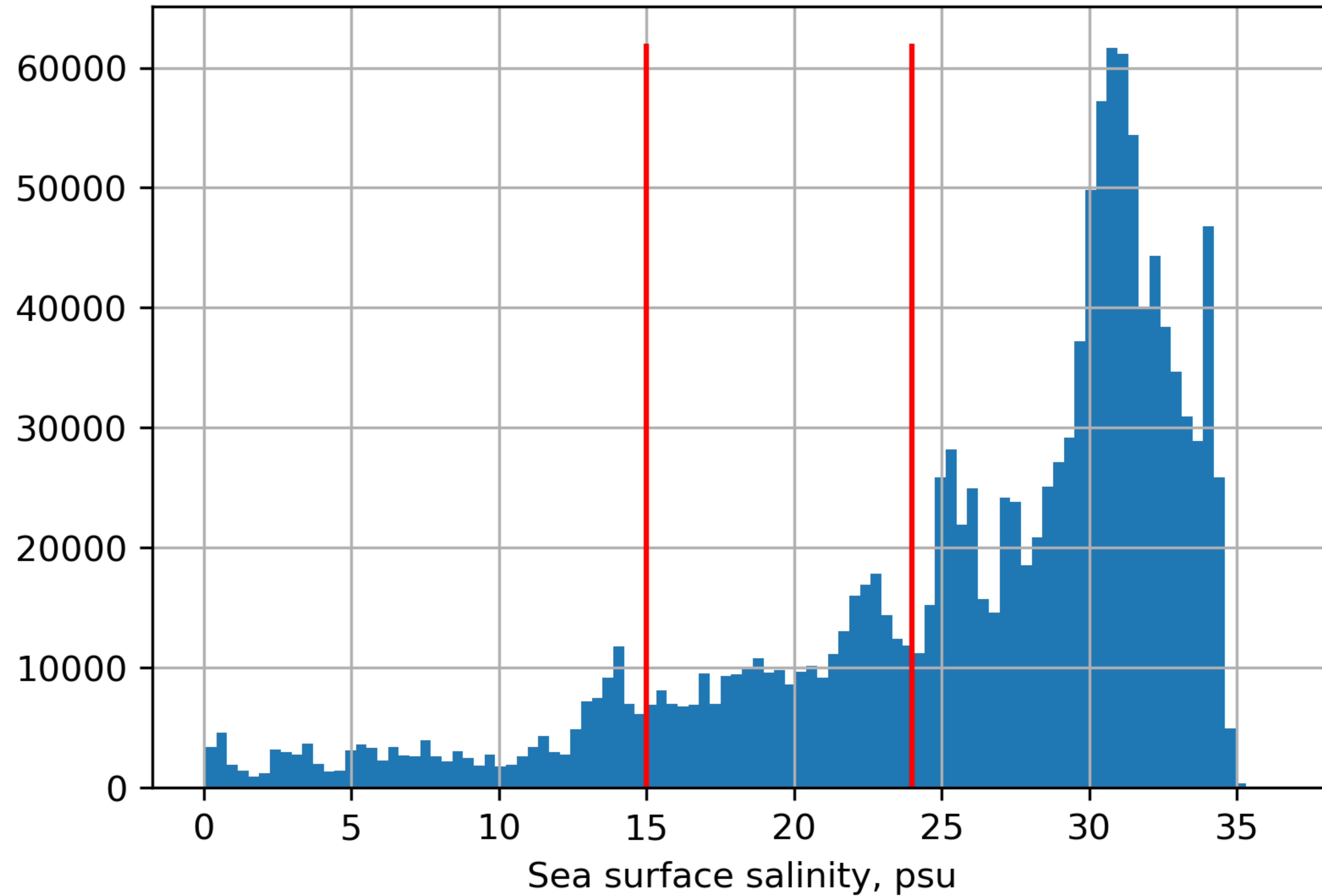
25

30

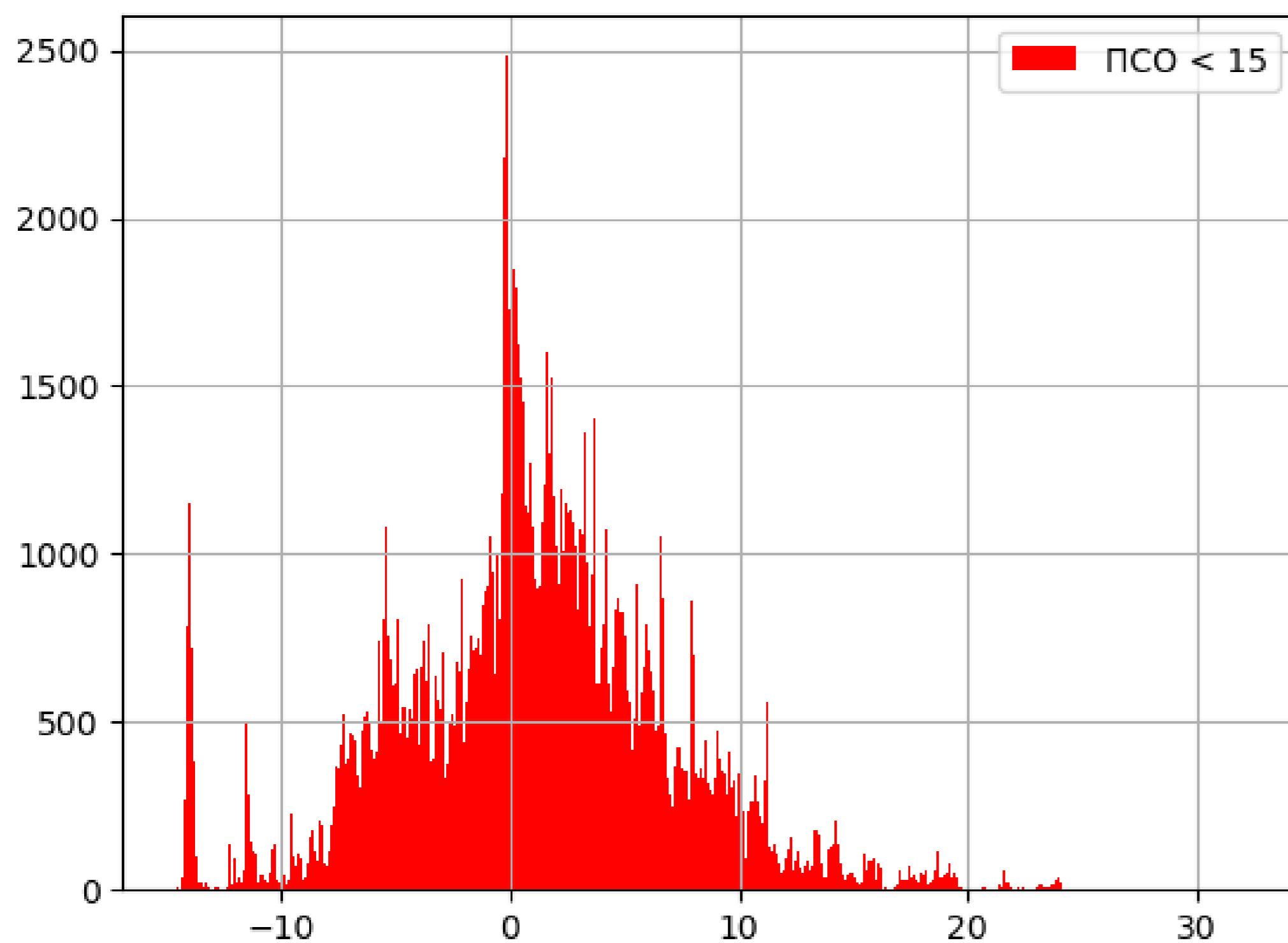
35

Соленость, епс

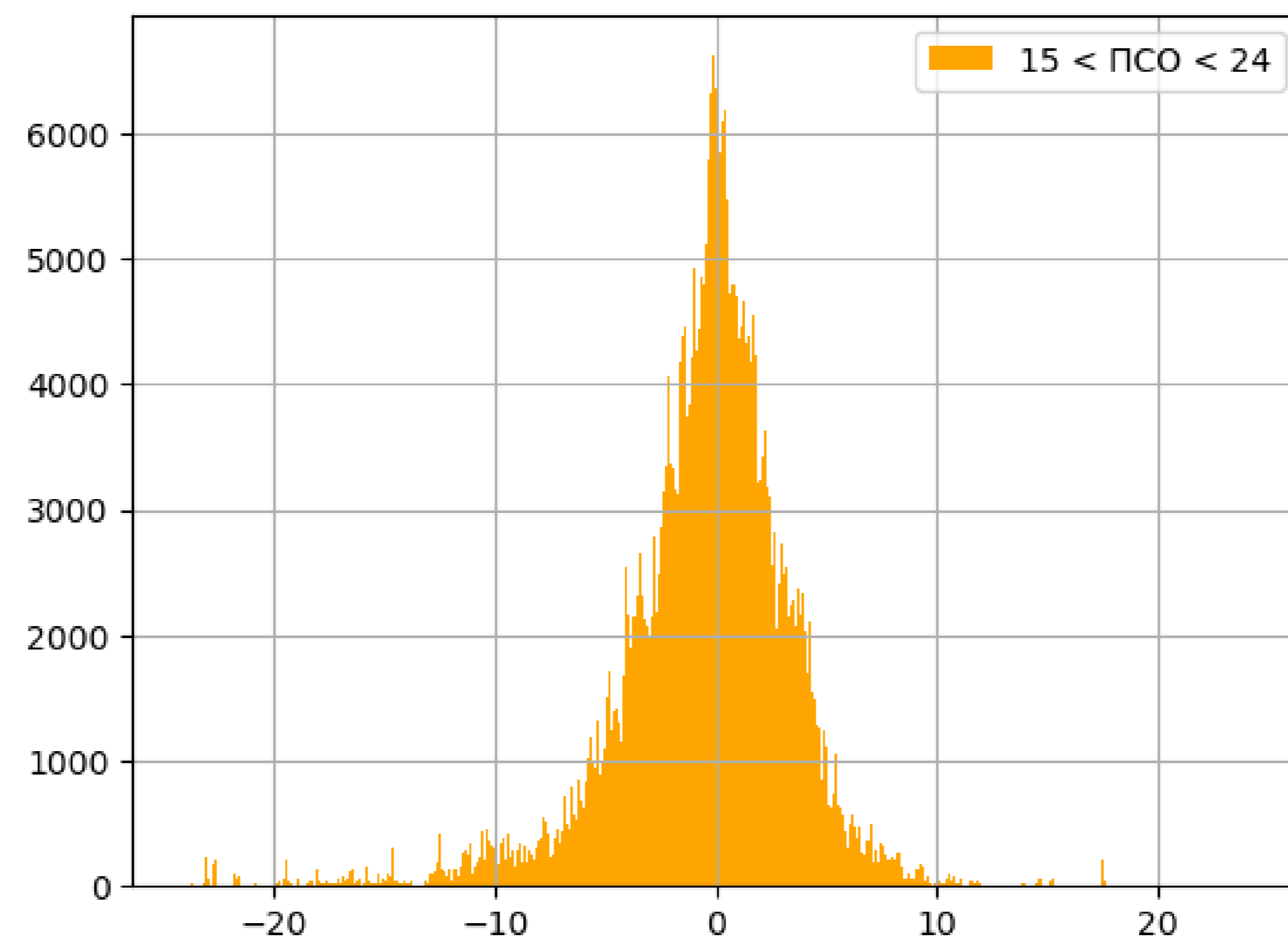
Натурные данные



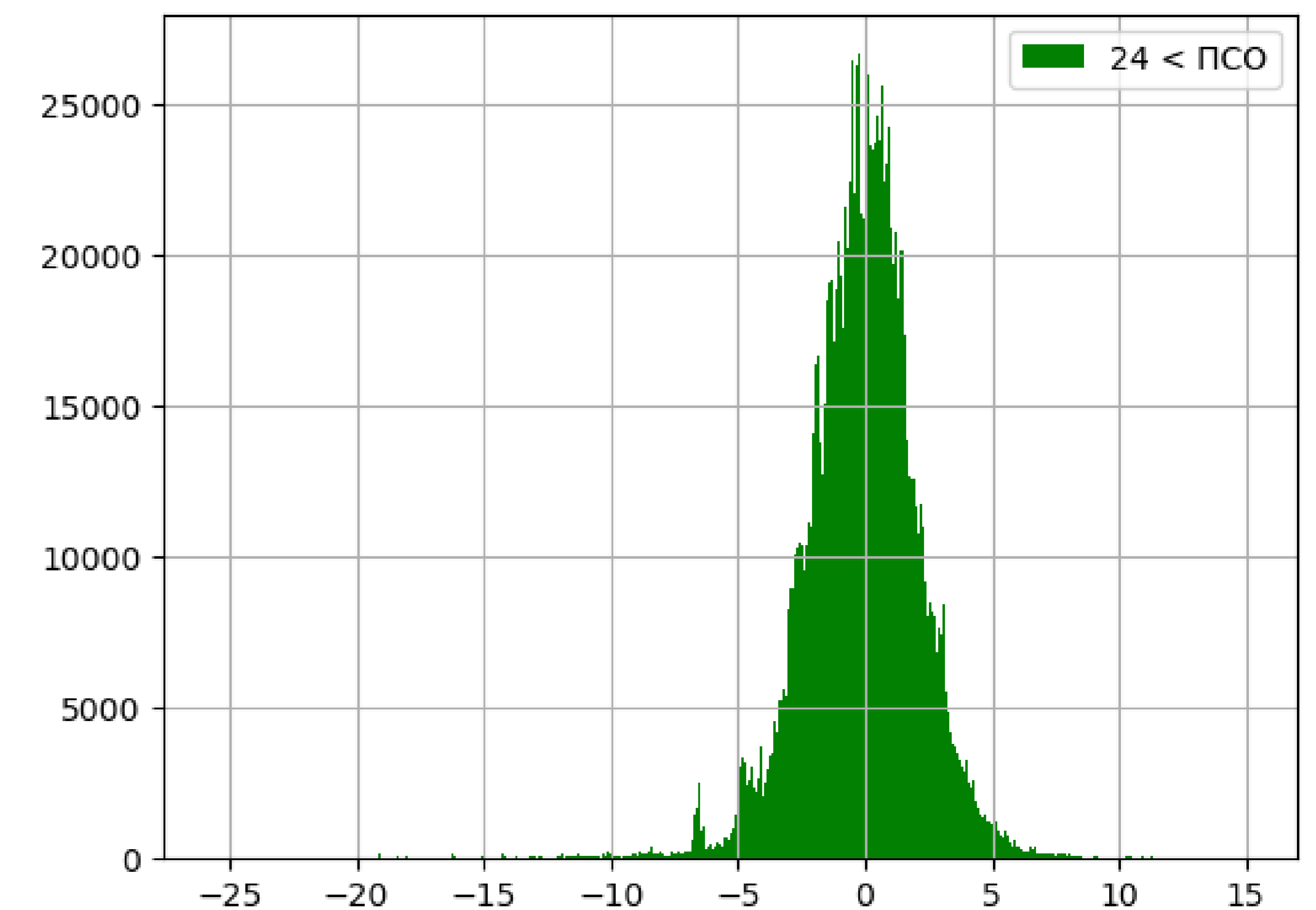
Ошибки спутниковой солености по выделенным классам



RMSE = 6.06



RMSE = 4.04



RMSE = 2.41

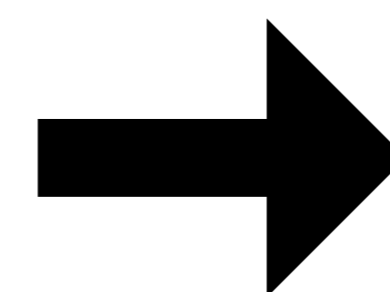
corr = 0.82, RMSE = 3.21

Методология

Спутниковые данные сопоставляются с натурными по принципу пространственной (10 км) и временной (3 часа) близости

Признаковое описание

Радиояркостьная температура, SSS star, поток солнечной радиации, скорость и направление ветра, температура поверхности океана, доля земли и доля льда в спутниковом снимке, зенитный и азимутальный углы Солнца для рассматриваемой точки



Поверхностная
соленость океана

Модели классические

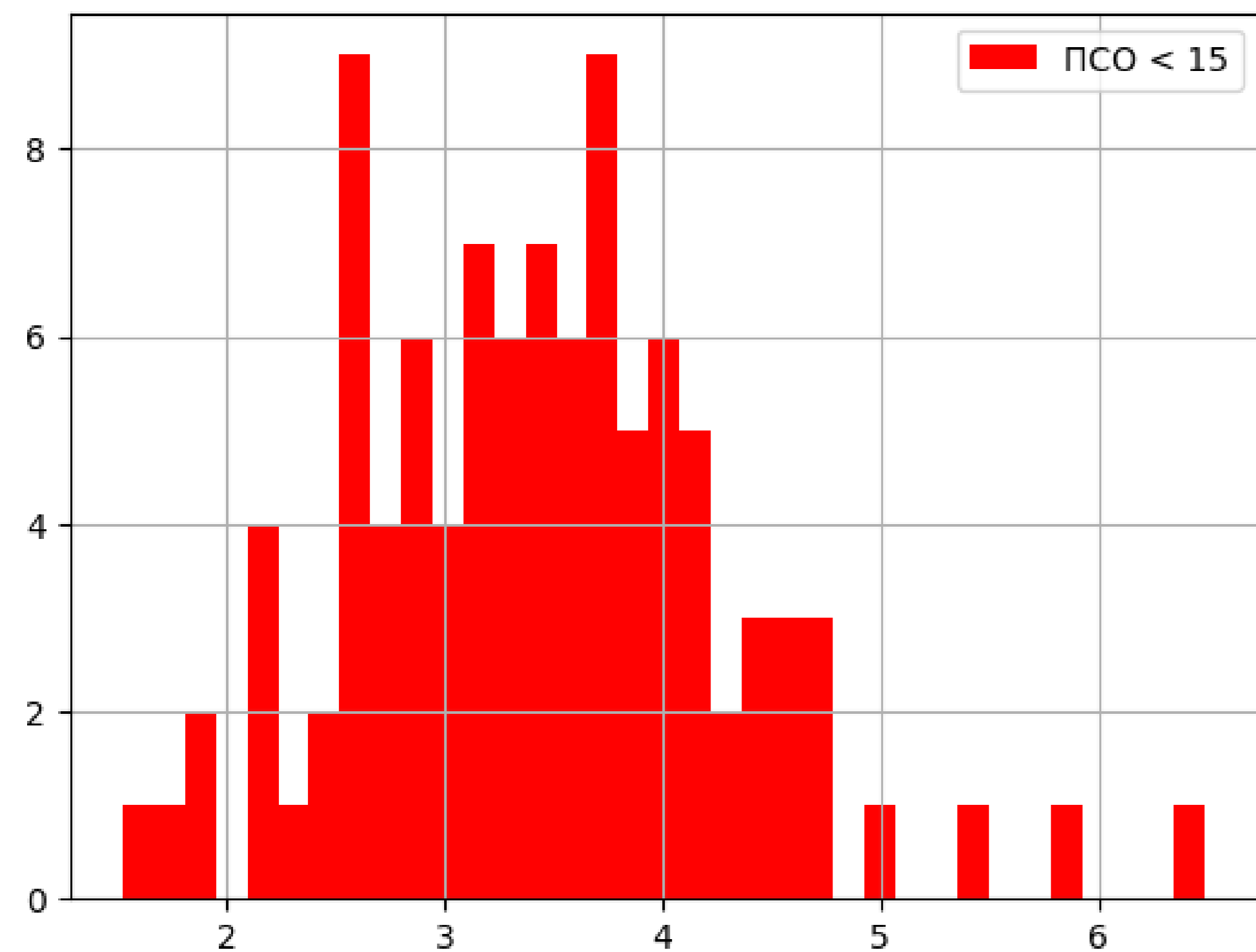
Классические модели Random Forest, Gradient Boosting Catboost

Исследуются двухуровневые (classification + regression) и одноуровневые (regression) модели.

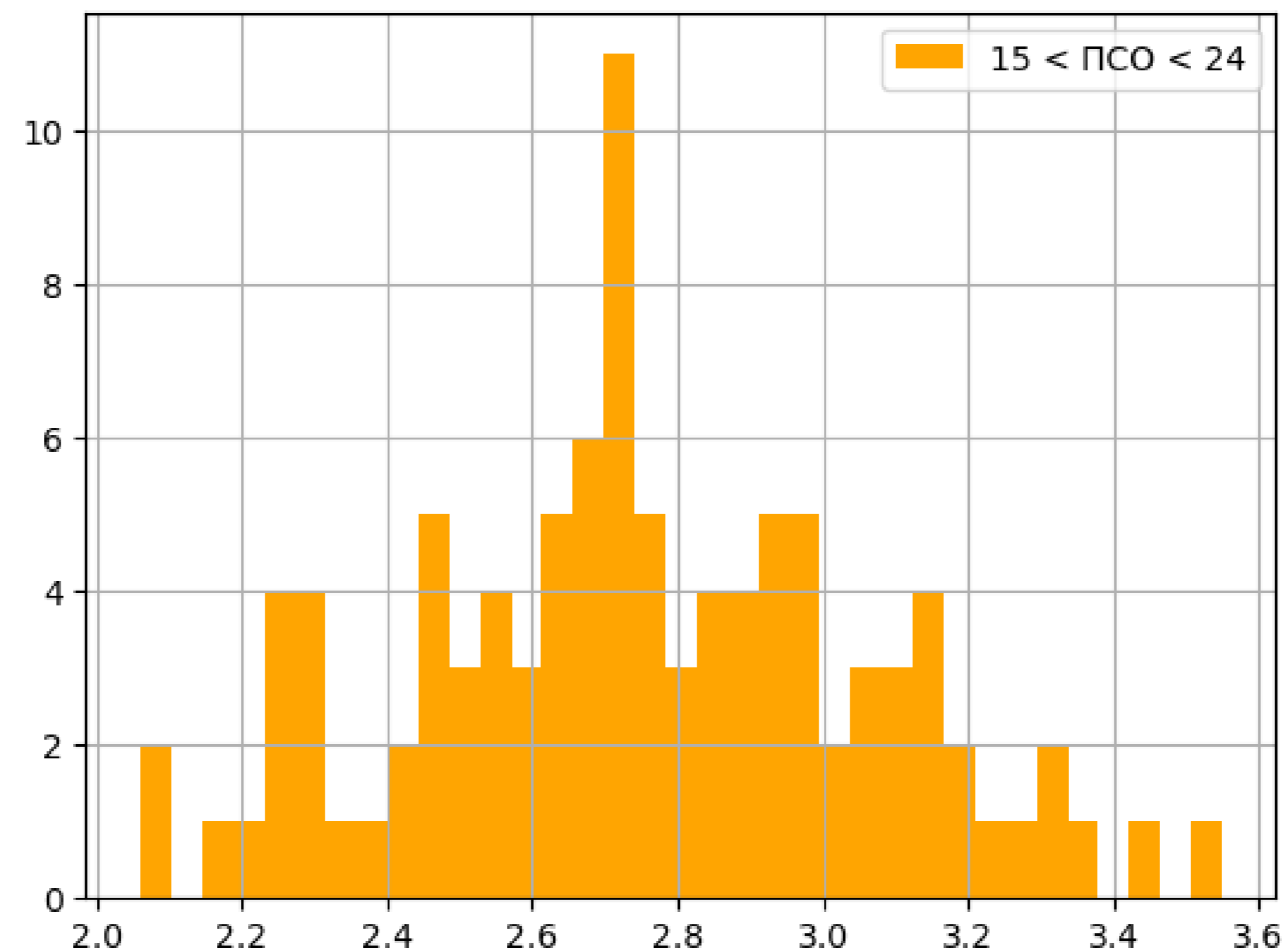
Результаты

	Random Forest	Gradient Boosting	Было:
одноуровневая модель	4.39 +- 1.29 psu	3.97 +- 1.09 psu	6.06 psu 4.04 psu 2.41 psu
	3.43 +- 0.50 psu	2.75 +- 0.31 psu	
	2.07 +- 0.30 psu	1.73 +- 0.20 psu	
двухуровневая модель	3.95 +- 0.98 psu	3.43 +- 0.86 psu	
	3.34 +- 0.52 psu	3.67 +- 0.66 psu	
	2.04 +- 0.29 psu	1.86 +- 0.28 psu	
Общее	2.63 +- 0.32 psu	2.20 +- 0.24 psu	3.21 psu

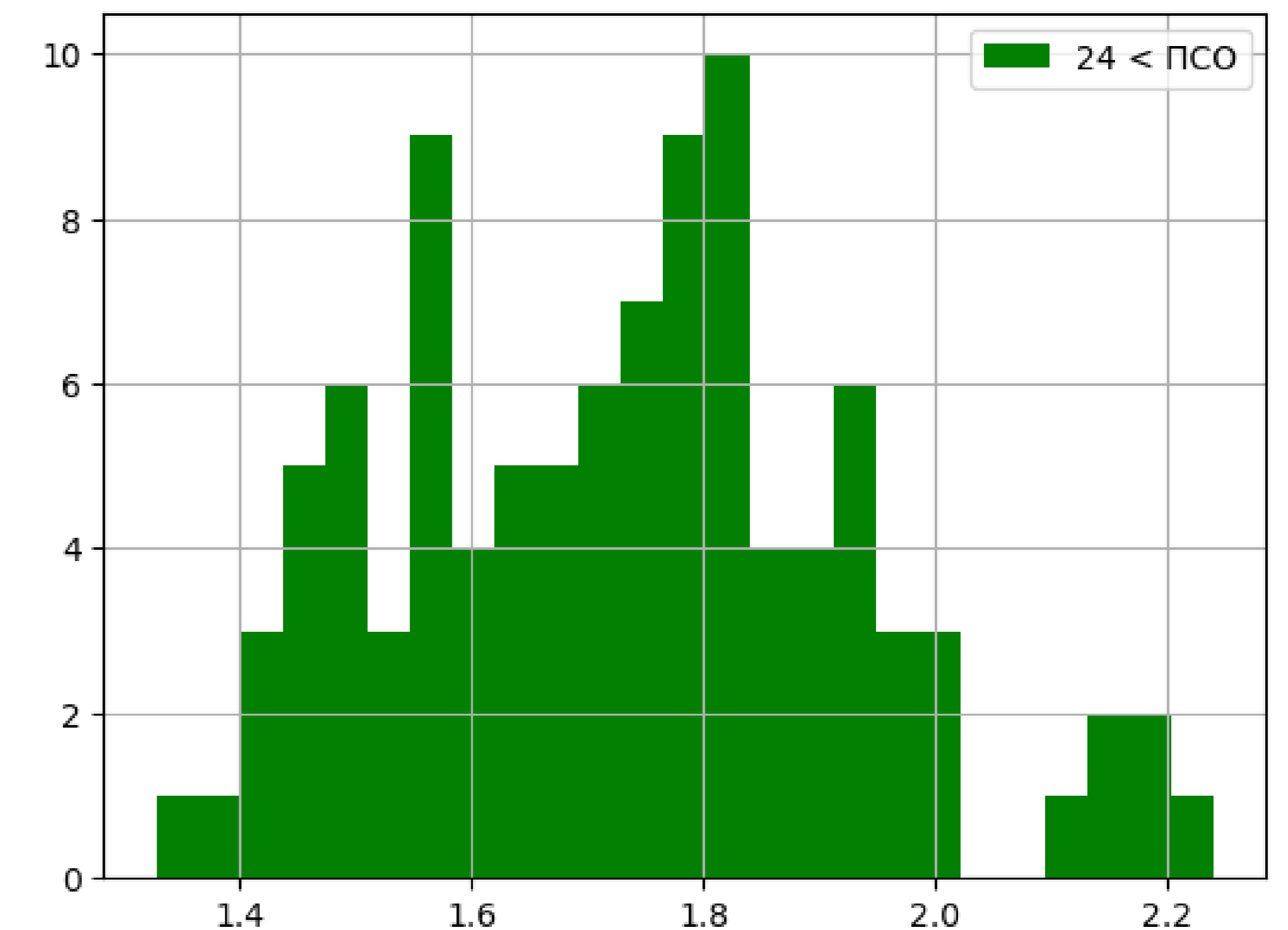
Распределение ошибок лучших моделей



3.43 +- 0.86 psu

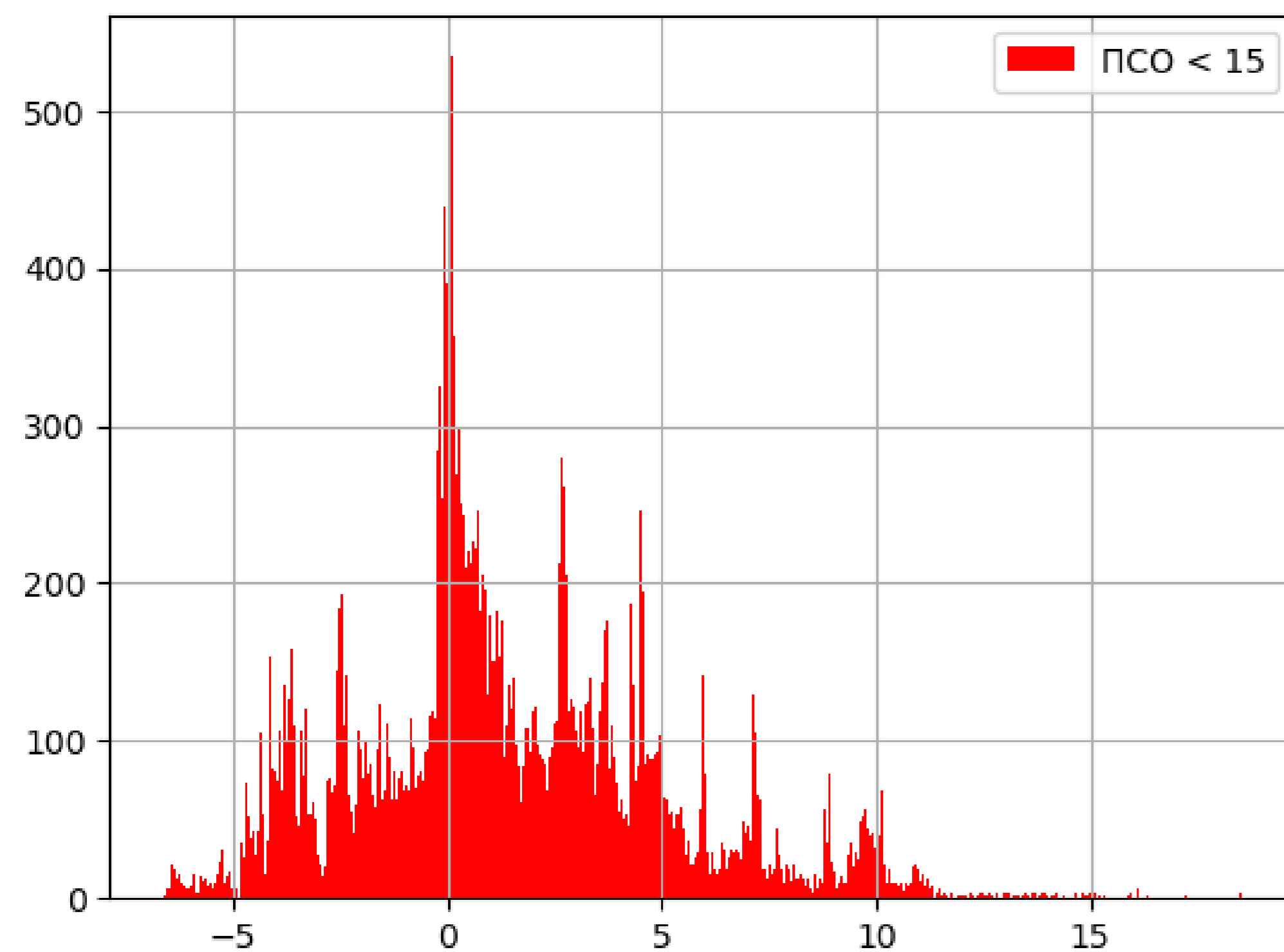


2.75 +- 0.31 psu

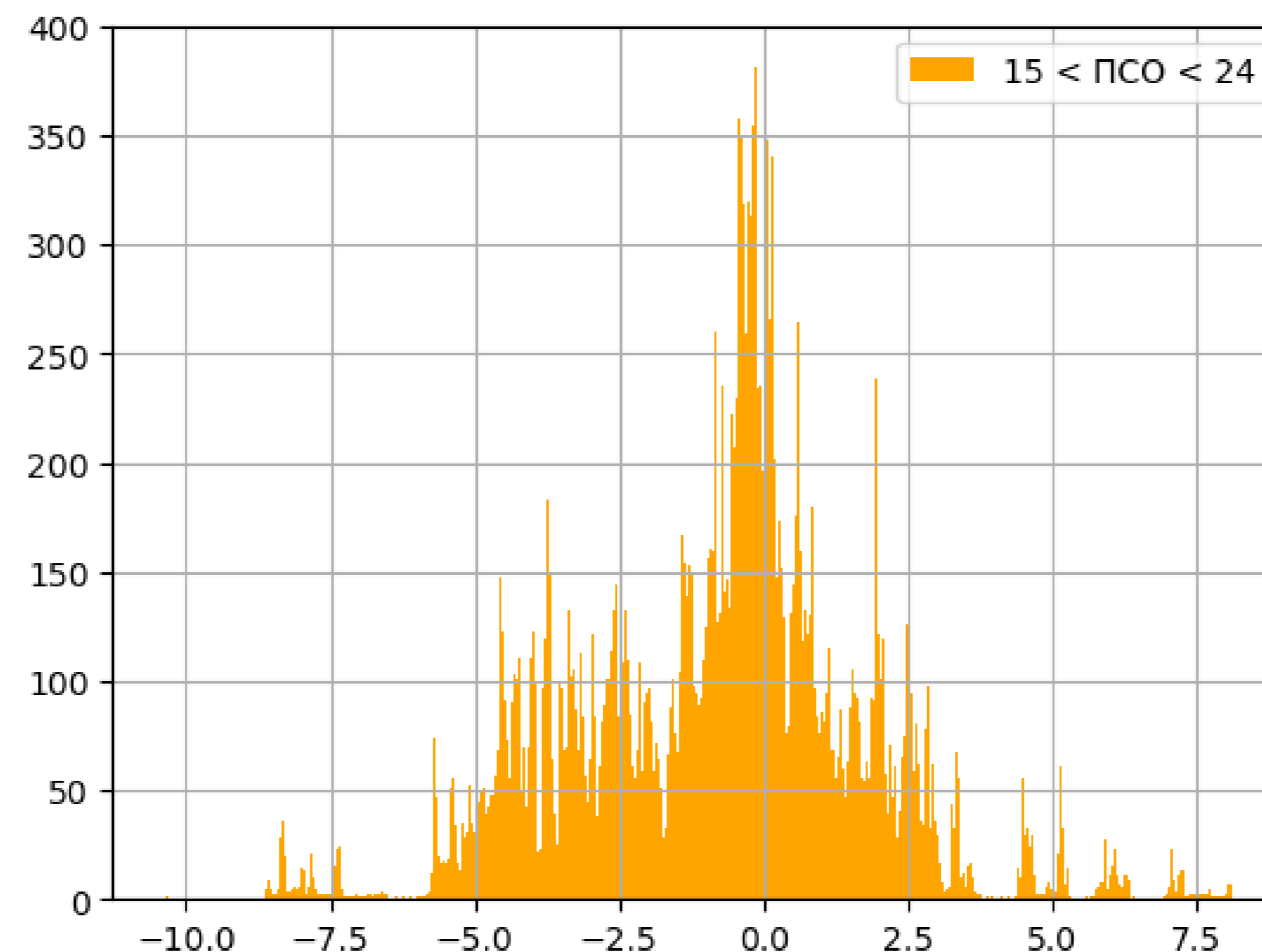


1.73 +- 0.20 psu

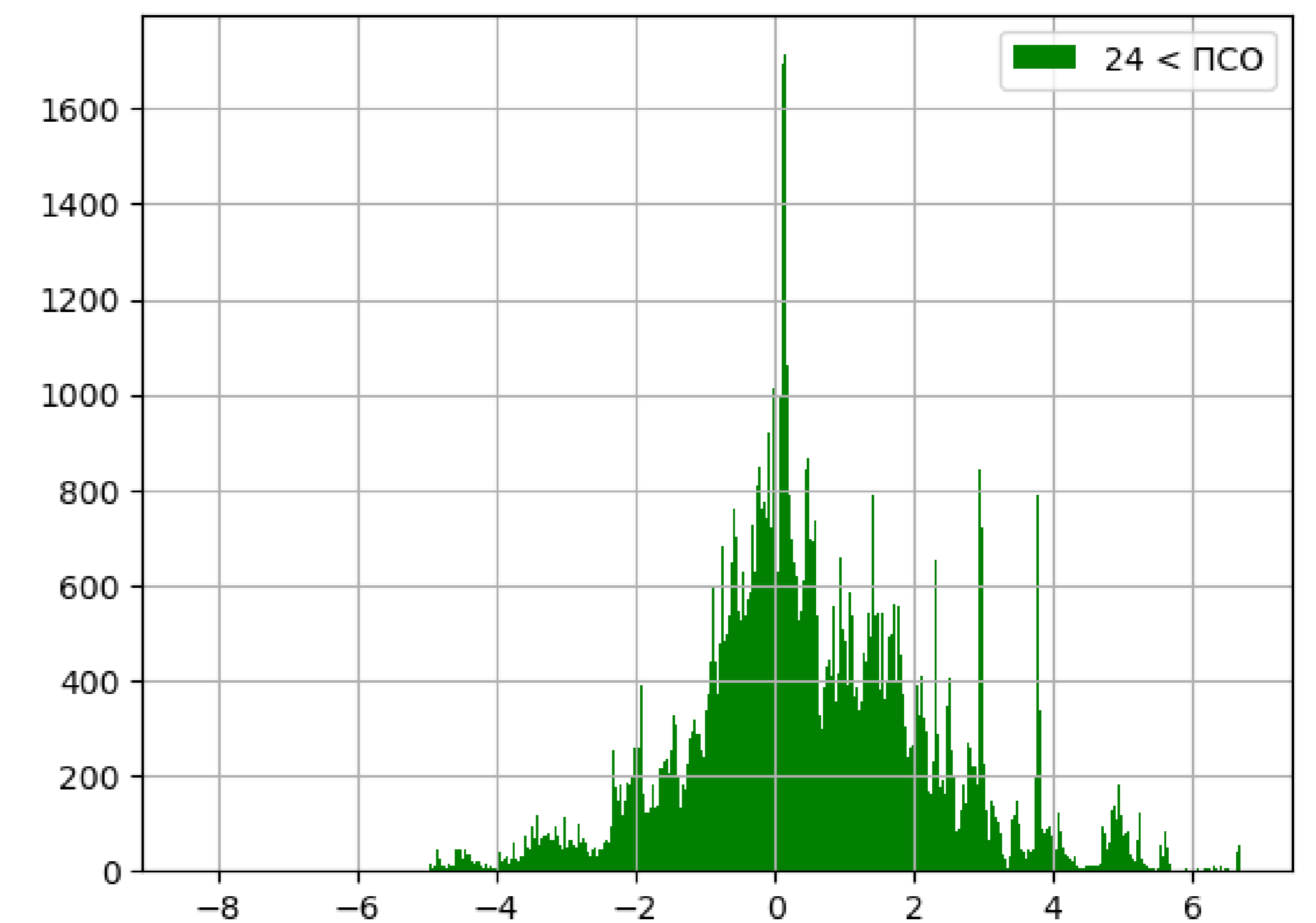
Характерное распределение ошибок в конкретном запуске



RMSE = 3.39
(6.06)



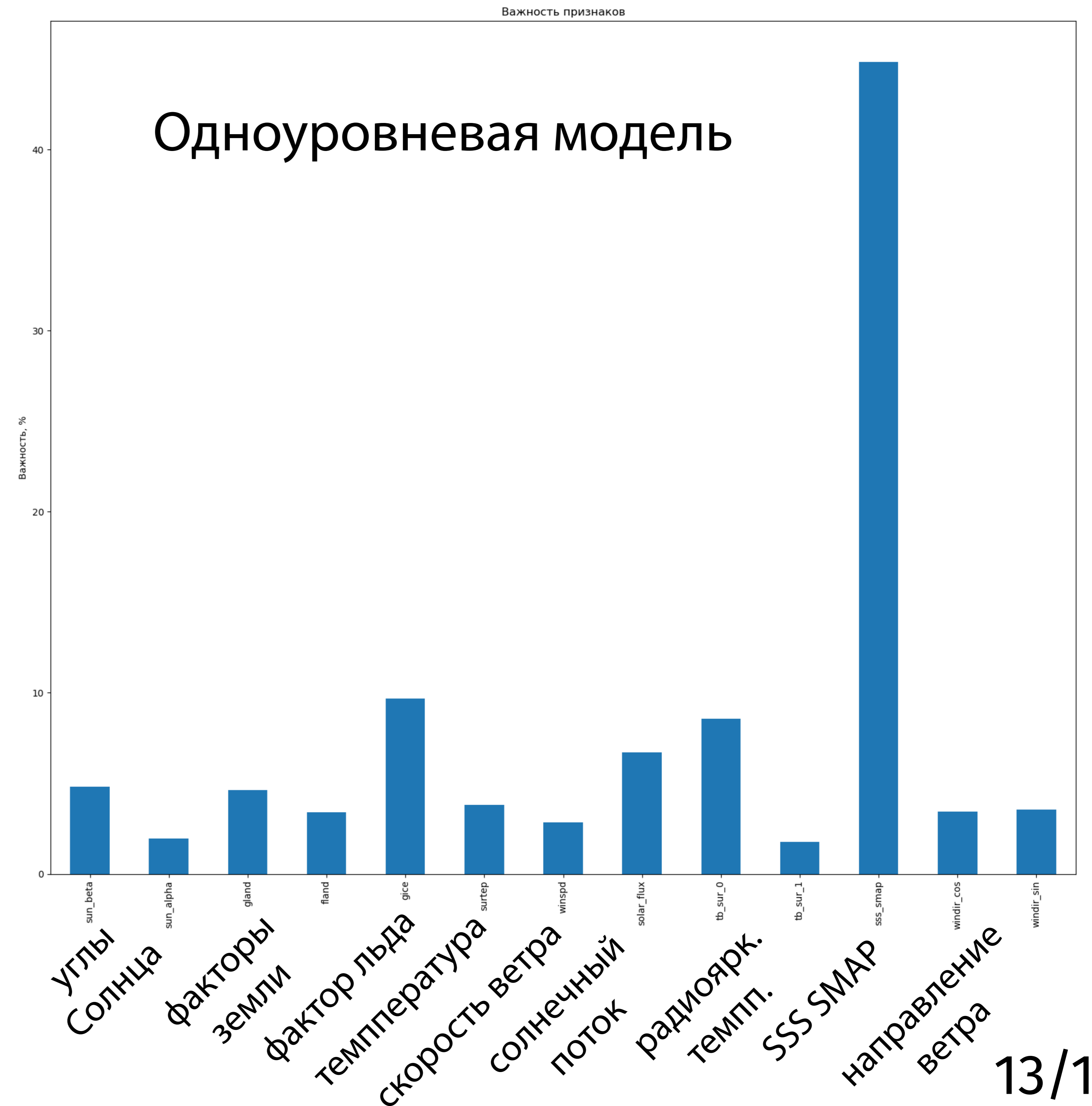
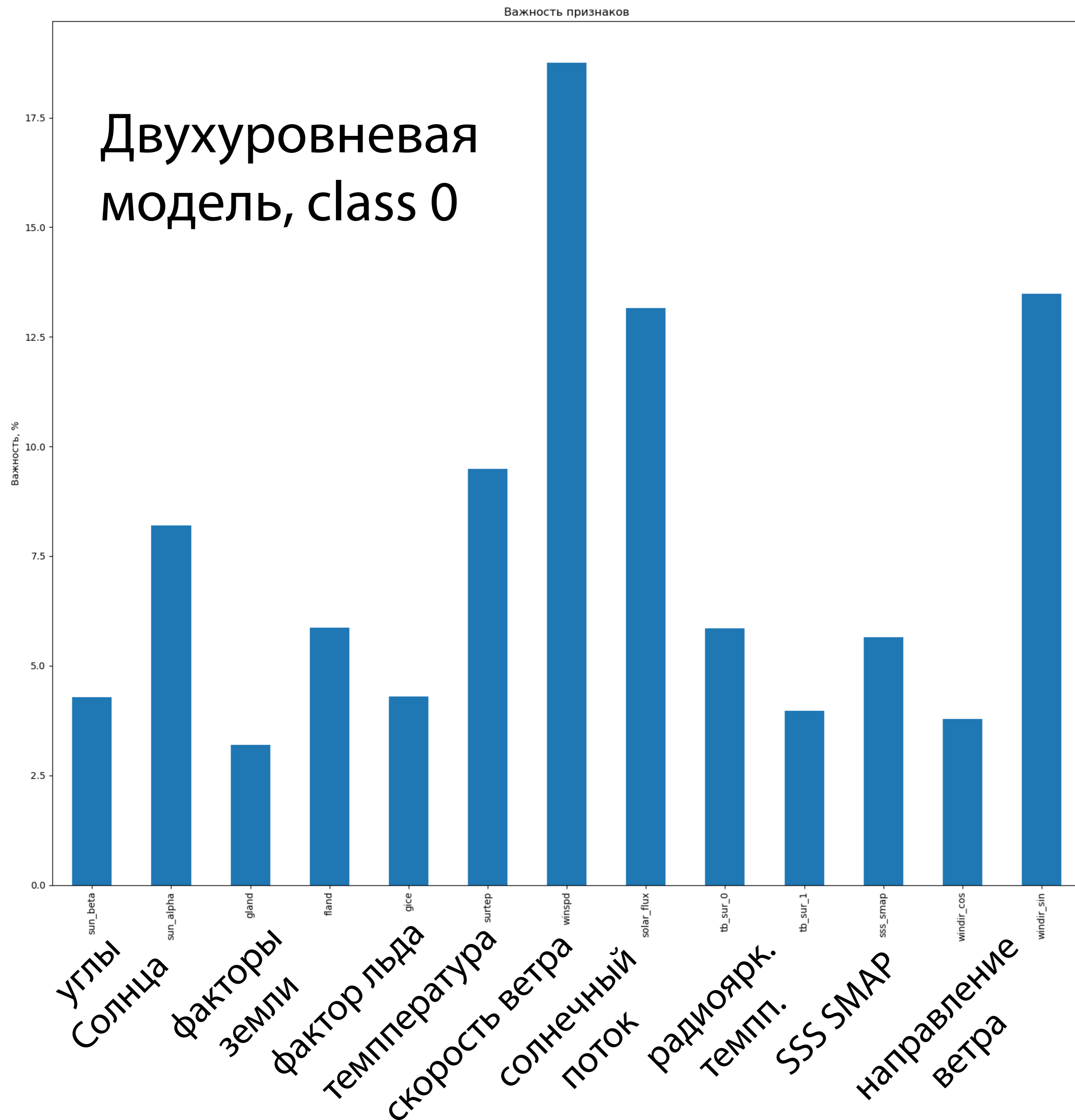
RMSE = 2.63
(4.04)



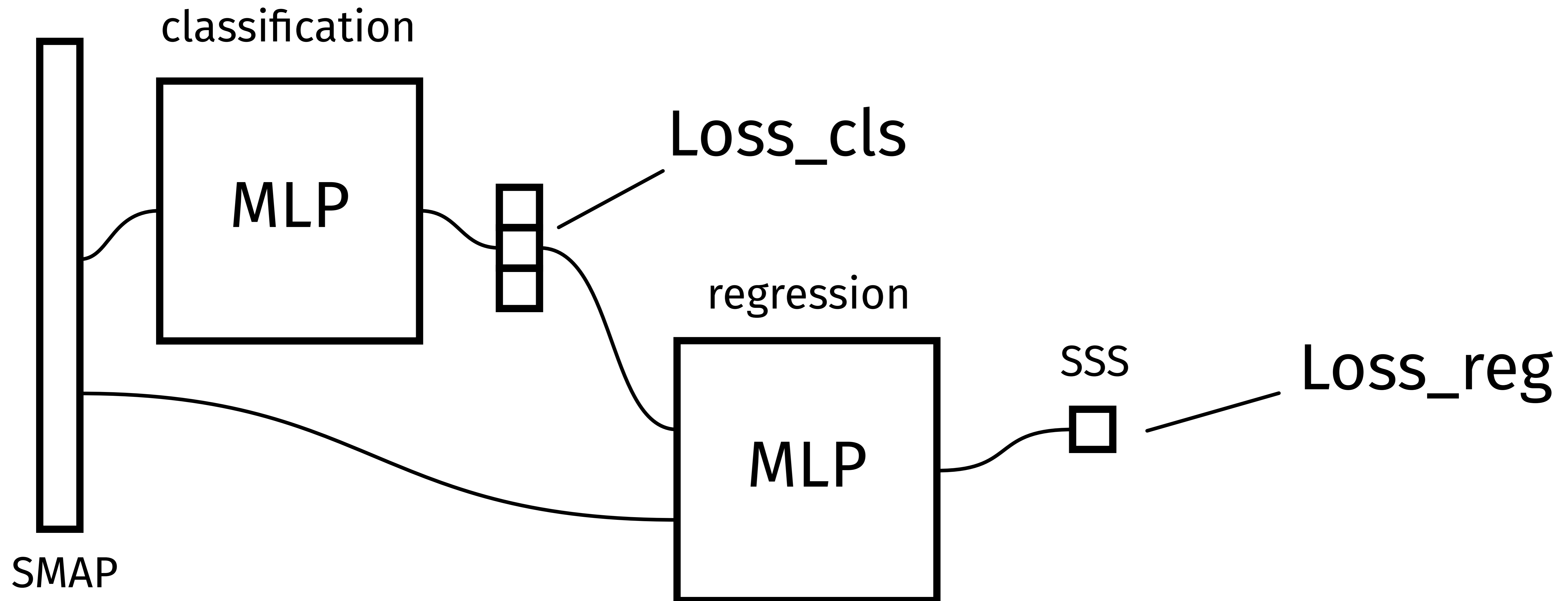
RMSE = 1.83
(2.41)

corr = 0.89 (0.82), RMSE = 2.11 (3.21)

Значимость признаков лучших моделей



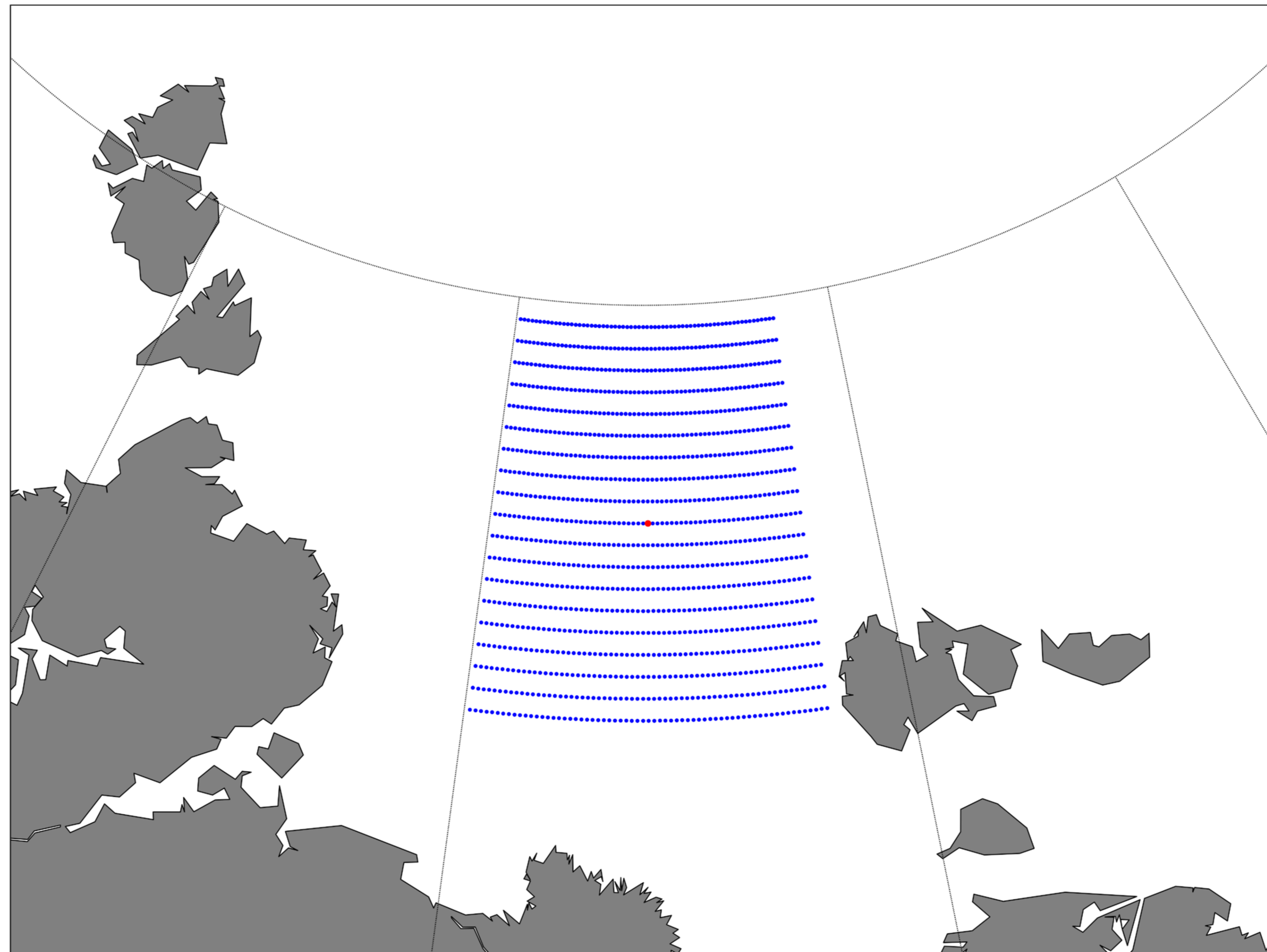
Модель Artificial Neural Network



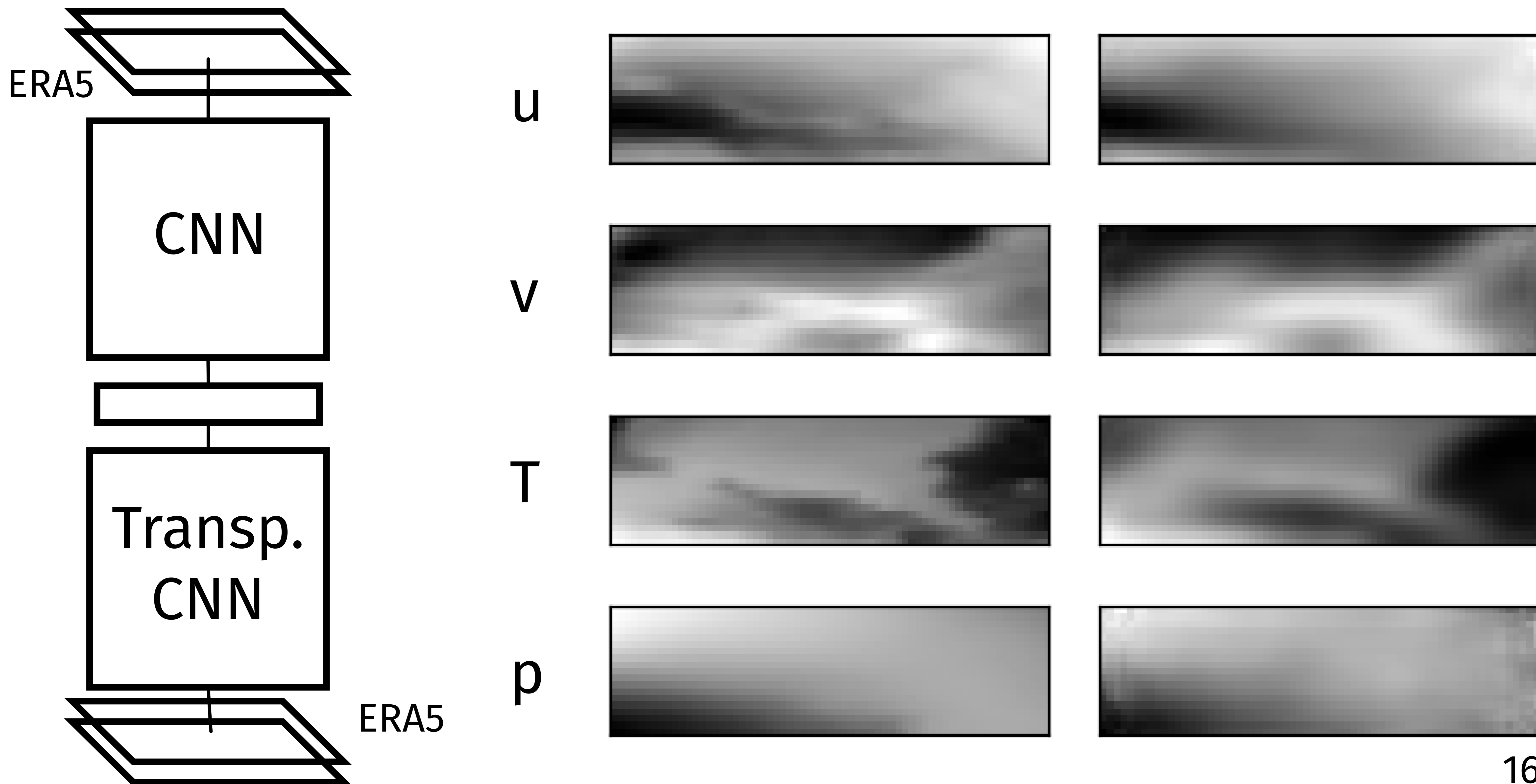
Результат:
2.85 +- 0.43 psu

Усложнение модели данными ERA5

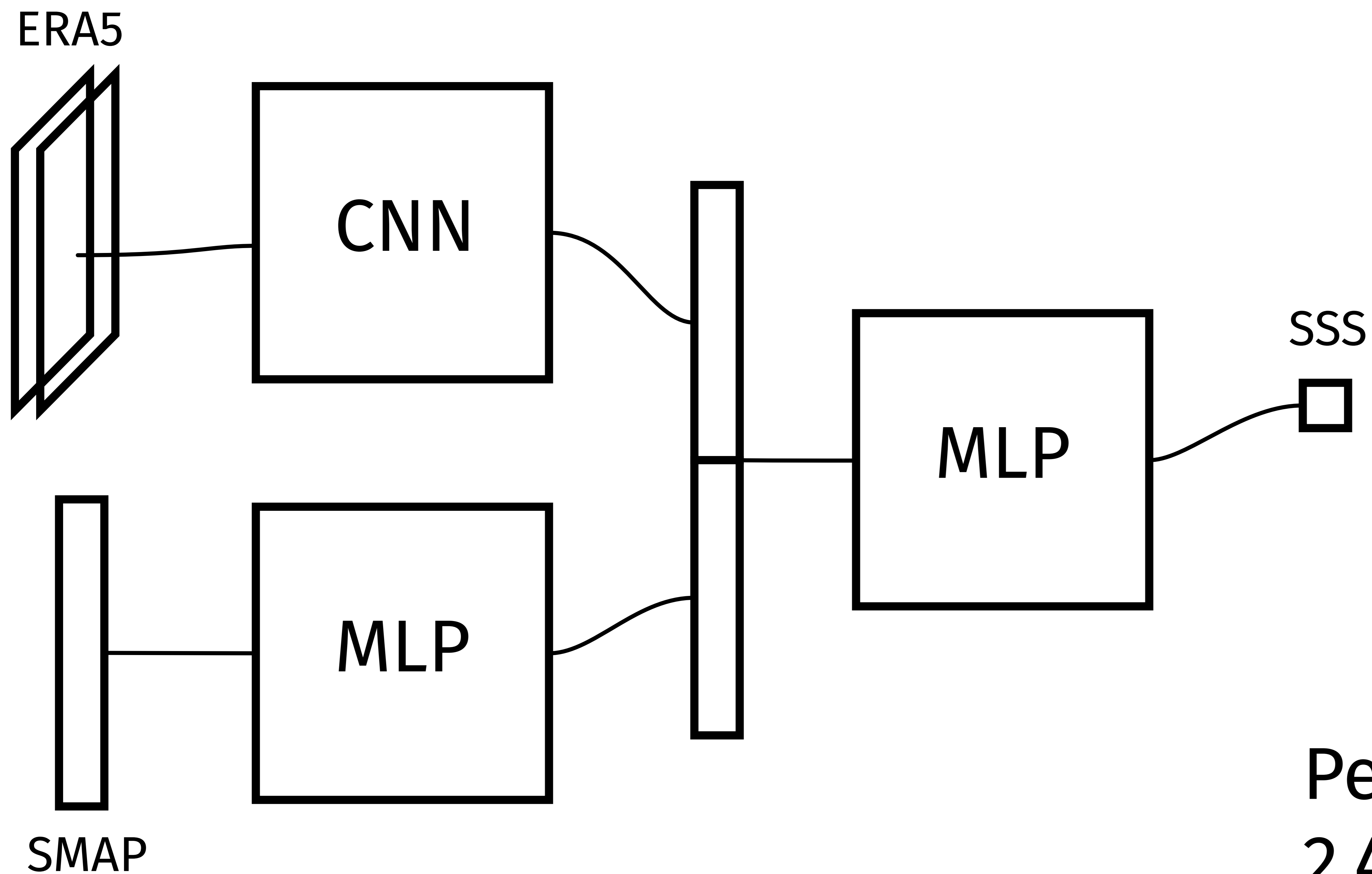
Берутся двумерные данные о поле ветра на высоте 10м, давления над морской поверхностью и температуре на высоте 2м в окрестности рассматриваемой точки



Предобучение CNN в подходе автокодировщика



Комплексная модель



Результат:
2.45 psu - ?

Выводы

- Все рассмотренные модели машинного обучения повысили качество спутниковой солености относительно стандартного алгоритма SMAP
- Наибольшая точность на данный момент достигнута при использовании модели Gradient Boosting Catboost, причем на низких значения солености лучше проявляет себя двухуровневая модель, в то время как на средних и высоких — одноуровневая
- Нейросетевые модели (пока) не дают существенного прироста по качеству и нуждаются в более тонкой настройке
- Наименьшее качество достигнуто на низких соленостях, что можно относительно малым количеством данных измерений, а также высокой стратификацией опресненных вод

Спасибо за внимание!