



ПРОГНОЗИРОВАНИЕ ХАРАКТЕРИСТИК ВОЗРАТНОЙ МИГРАЦИИ НЕРКИ В УСТЬЕ РЕКИ ФРЕЙЗЕР С ПРИМЕНЕНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

Mikhail Borisov^{1,2}

Mikhail Krinitskiy^{1,2}

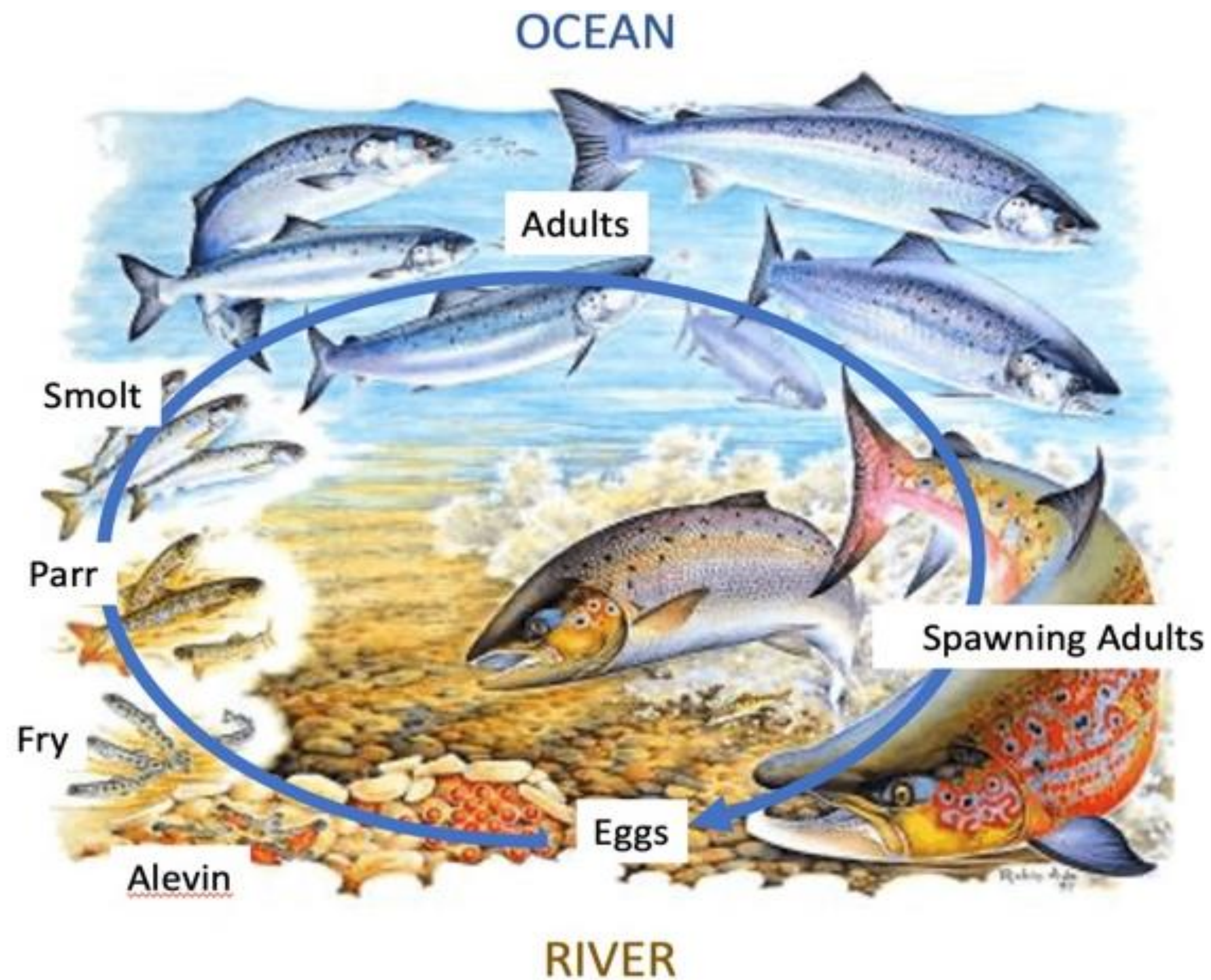
¹**Moscow Institute of Physics and Technology**

²**Shirshov Institute of Oceanology, Russian Academy of Sciences**

Введение

Oncorhynchus nerka также известная как нерка, проходит сложный жизненный цикл в разных водоемах.

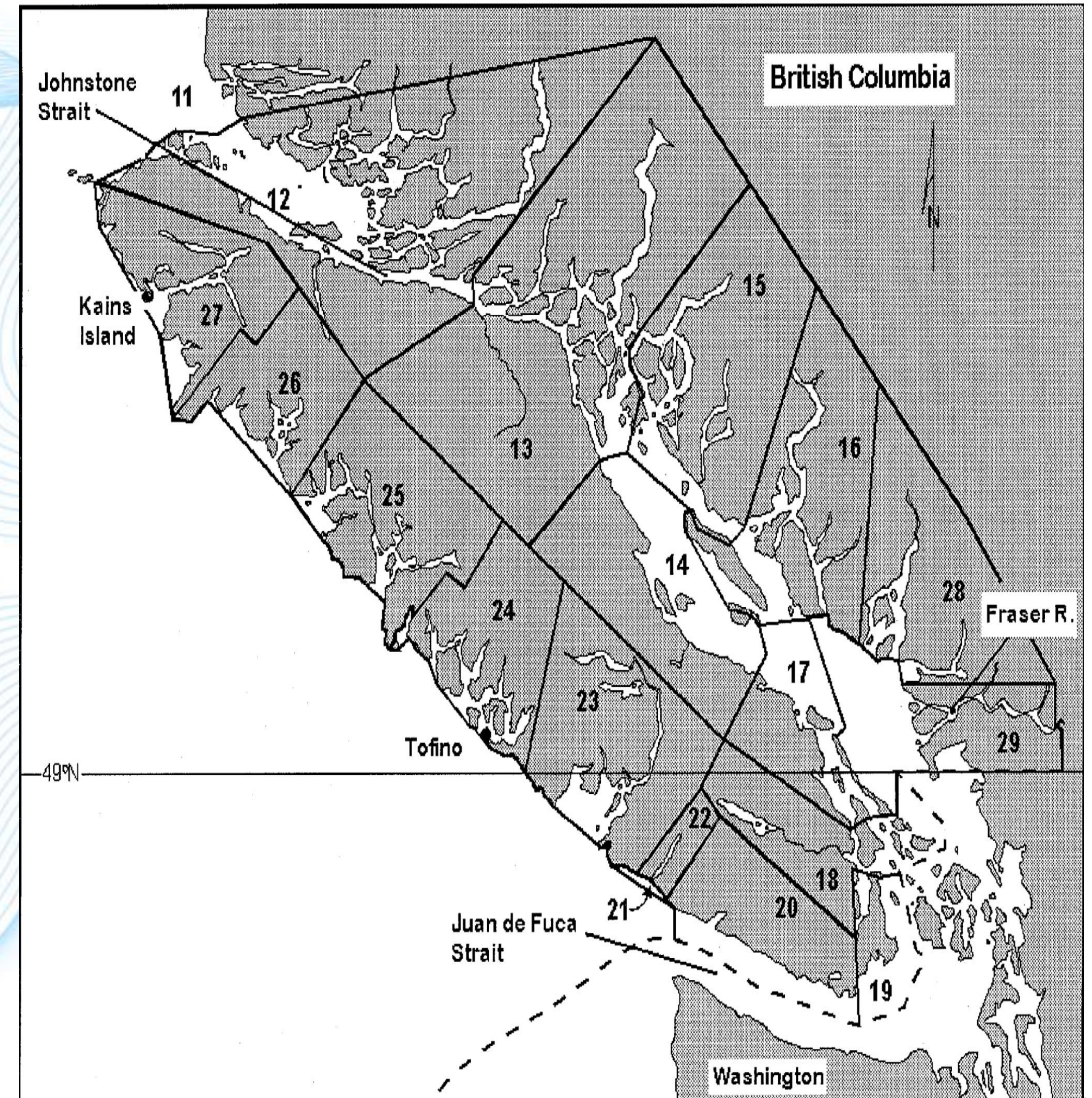
Рыба начинает свою жизнь в пресноводных реках и озерах, где она вылупляется из икры. После нескольких месяцев молодь мигрирует в океан, где проводит от одного до четырех лет, набирая вес и готовясь к размножению. Взрослые особи возвращаются в родные пресноводные водоемы для нереста – возвратная миграция нерки.

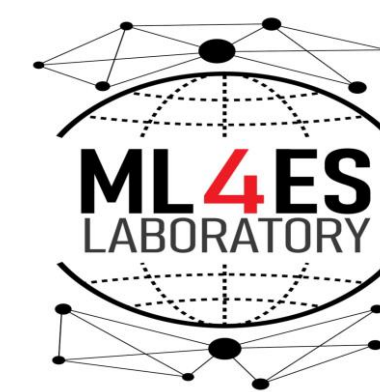


Введение

Возвращаясь в реку Фрайзер для нереста, нерка (*Oncorhynchus nerka*) сталкивается с выбором маршрута вдоль острова Ванкувер. Эти маршруты включают путь через пролив Джонстона на севере и пролив Хуан-де-Фука на юге.

1. Дата возвращения первой половины популяции нерки в реку Фрейзер служит ключевым индикатором для определения промысловых сезонов.
2. Процент особей, выбирающих северный маршрут миграции вокруг острова Ванкувер, напрямую влияет на квоты вылова между США и Канадой.
3. Точные прогнозы данных параметров имеют решающее значение для устойчивого управления рыбными запасами и сохранения популяции нерки.





Цель исследования

В данном исследовании предполагается, что поведение нерки связано с химическим составом морской воды, температурой поверхности моря и динамикой течений в верхних слоях океана.

Таким образом цель исследования заключается в создании и обучении модели, способной прогнозировать характеристики возвратной миграции нерки на основе факторов окружающей среды.

Были решены следующие задачи:

- Сбор и предварительная обработка данных
- Создание обучающей выборки
- Обучение моделей
- Оптимизация выбора контрольных точек
- Определение лучшей модели

Характеристики возвратной миграции

Целевые переменные:

1. Медианная дата возвратной миграции для подвида Chilko
2. Медианная дата возвратной миграции для подвида Early Stuart
3. Доля северного отклонения

Медианная дата возвратной миграции характеризует момент, когда 50% возвращающихся нерки из северо-восточной части Тихого океана прибывают в устье пролива Хуан-де-Фука между Канадой и США. Доля северного отклонения отражает процентный коэффициент ухода рыбы на северный маршрут в течение года. Данные получены с 1940х годов.

Медианная дата была преобразована согласно формуле для учета периодичности явления. Затем каждая переменная была нормализована.

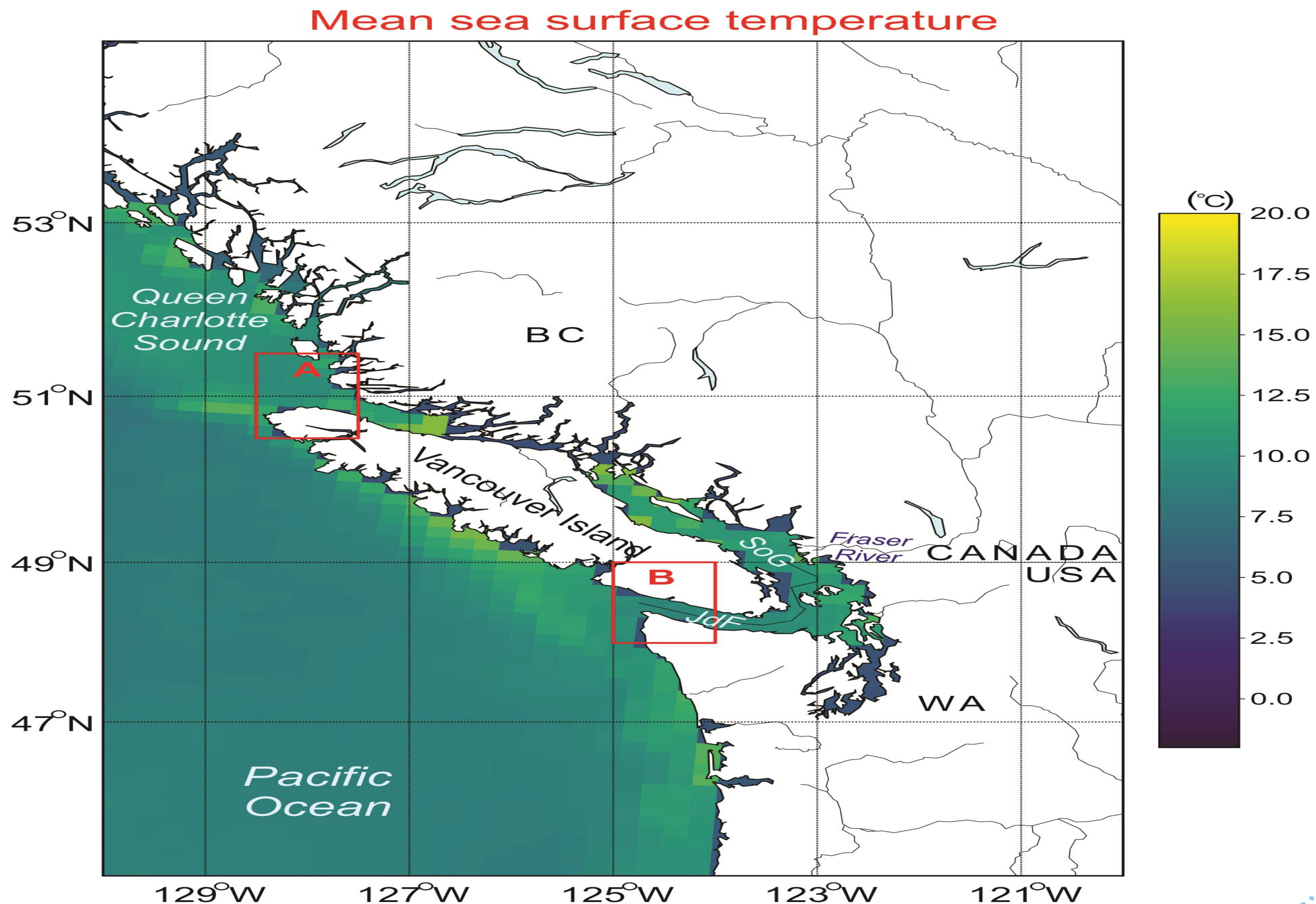
$$RT_{cos} = \cos\left(\frac{\pi * \text{порядковый день}}{\text{дней в году}}\right)$$

$$RT_{cos} = \frac{RT_{cos} - \text{mean}(RT_{cos})}{\sigma(RT_{cos})}$$

$$NDR_{std} = \frac{NDR - \text{mean}(NDR)}{\sigma(NDR)}$$

Данные Glorys2v4

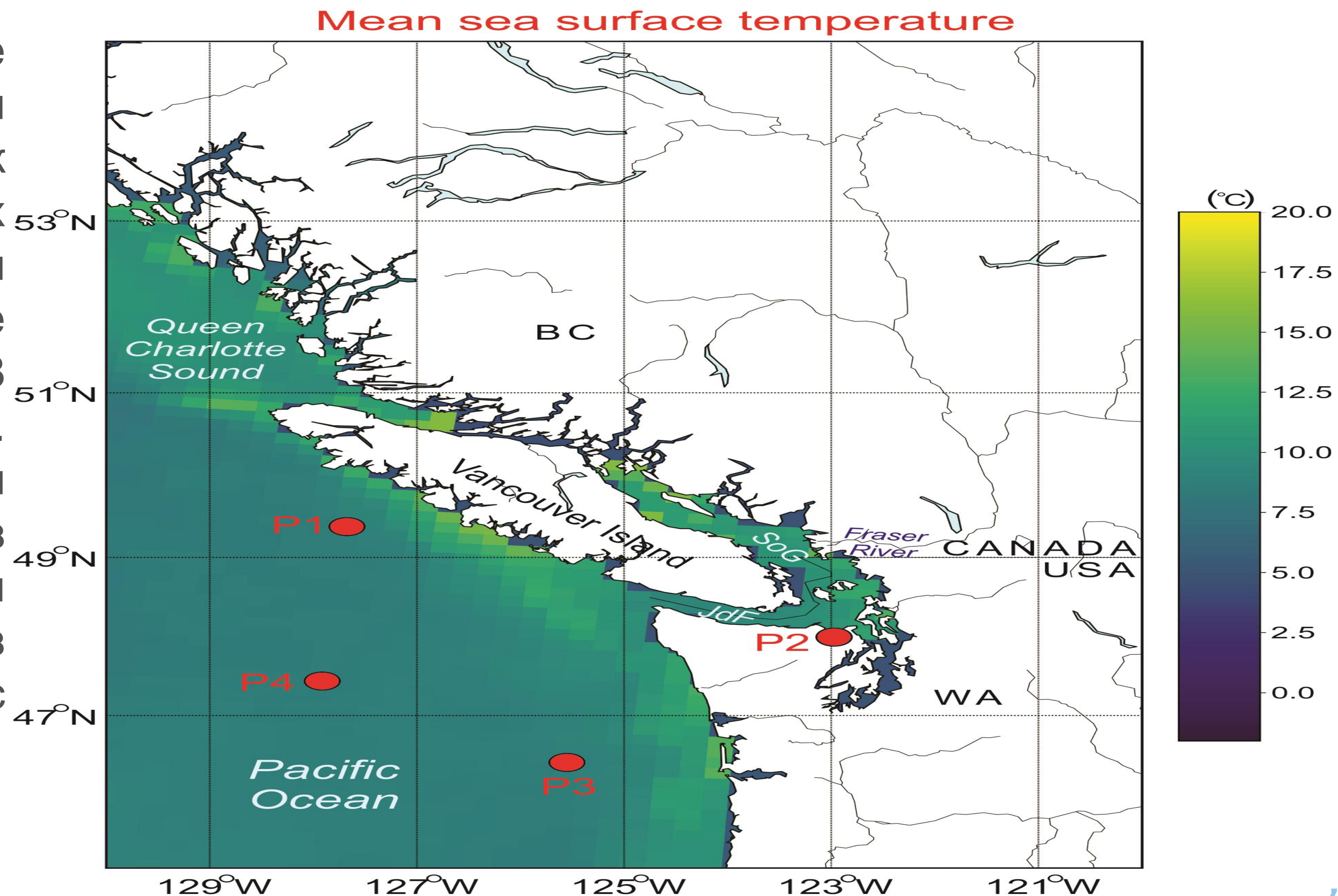
- 1992-2016 ежемесячно
- Зональная скорость ветра и течений
- Меридиональная скорость ветра и течений
- Соленность
- Температура



Обучающая выборка

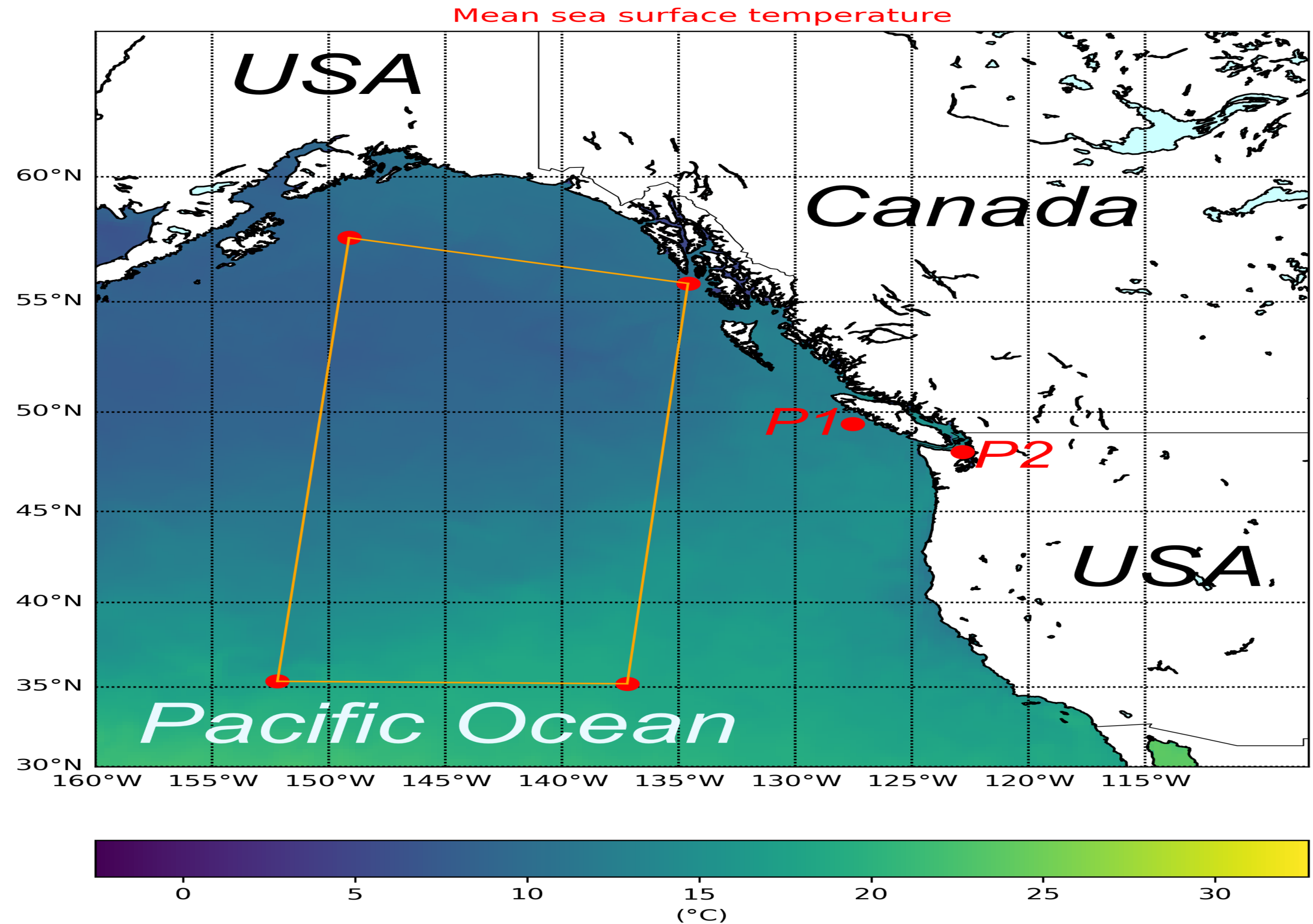
Признаковое описание составлялось из значений геофизических характеристик океана в контрольных точках за 7 месяцев. Таким образом получили признаковое описание размером 168 признаков на год. Аугментация выборки производилась 10000 раз путем добавлением доли случайного числа из нормального распределения с маленькой дисперсией

$$\mu = 0, \sigma = 0.001$$

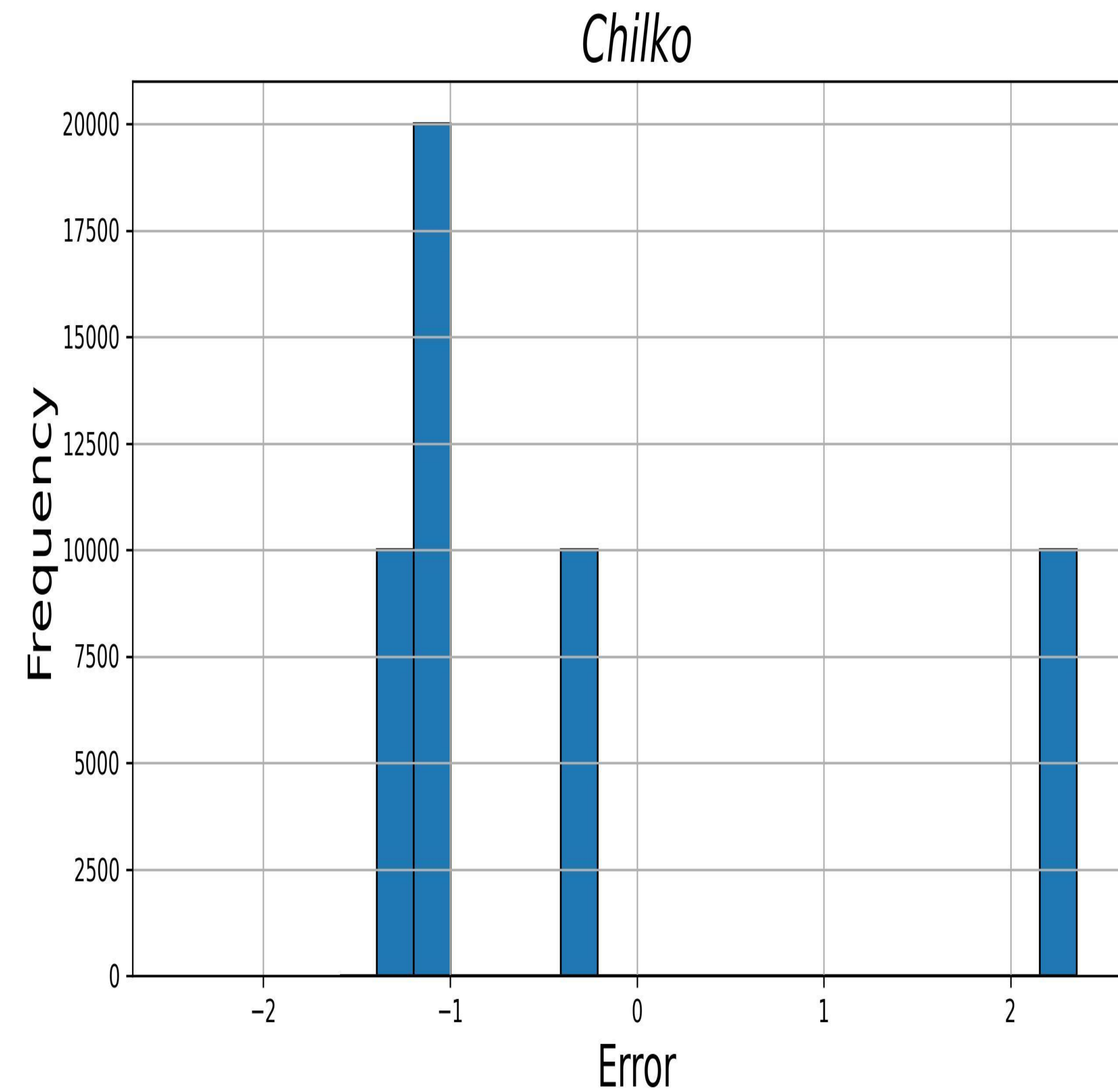
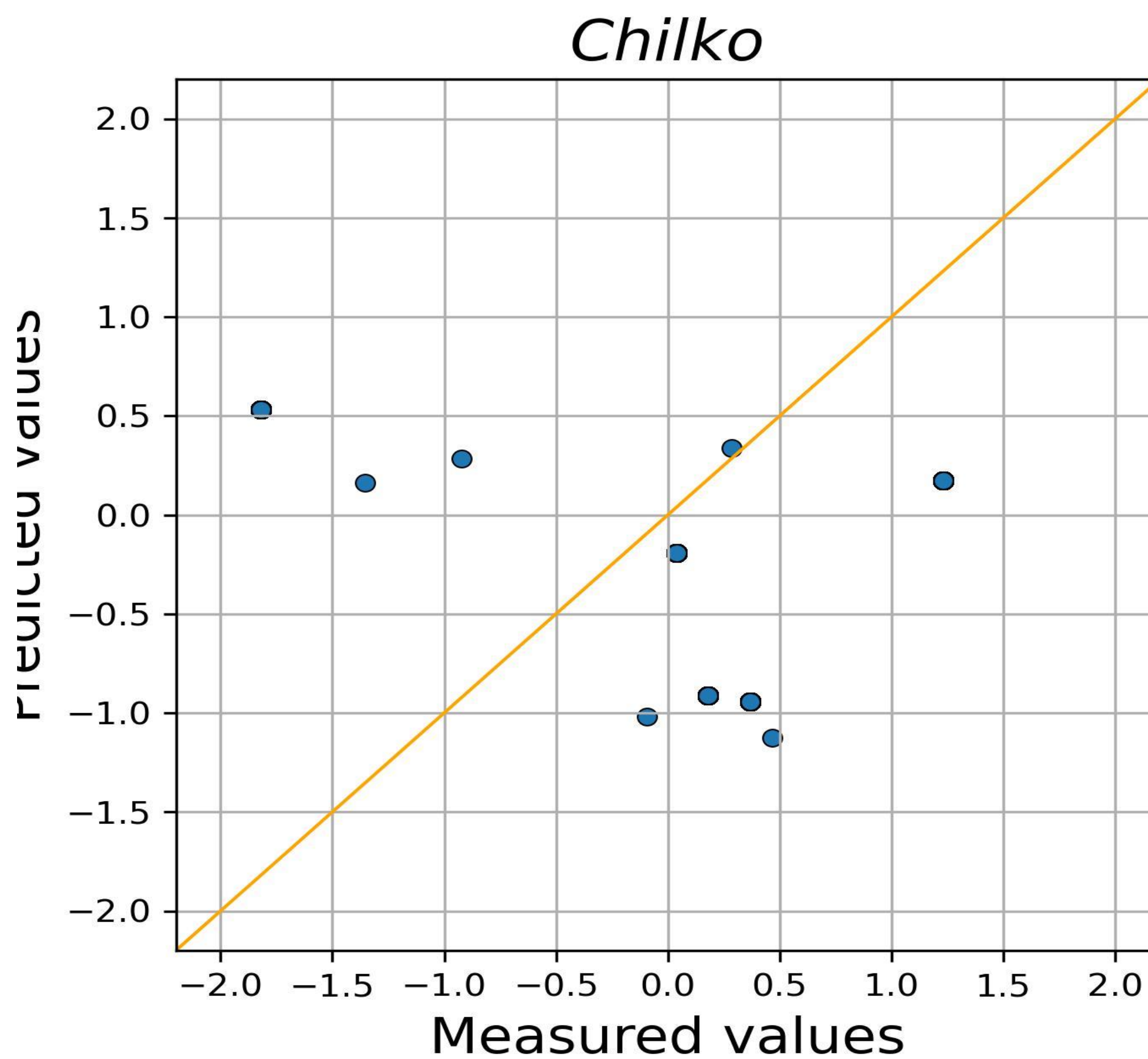


Методы

В качестве моделей использовались :
Гребневая и линейная регрессии, случайный лес, градиентный бустинг.
Для оптимизации выбора контрольных точек были зафиксированы первые две, оставшиеся выбирались из большого диапазона точек при помощи оптимизации гиперпараметров с использованием Optuna

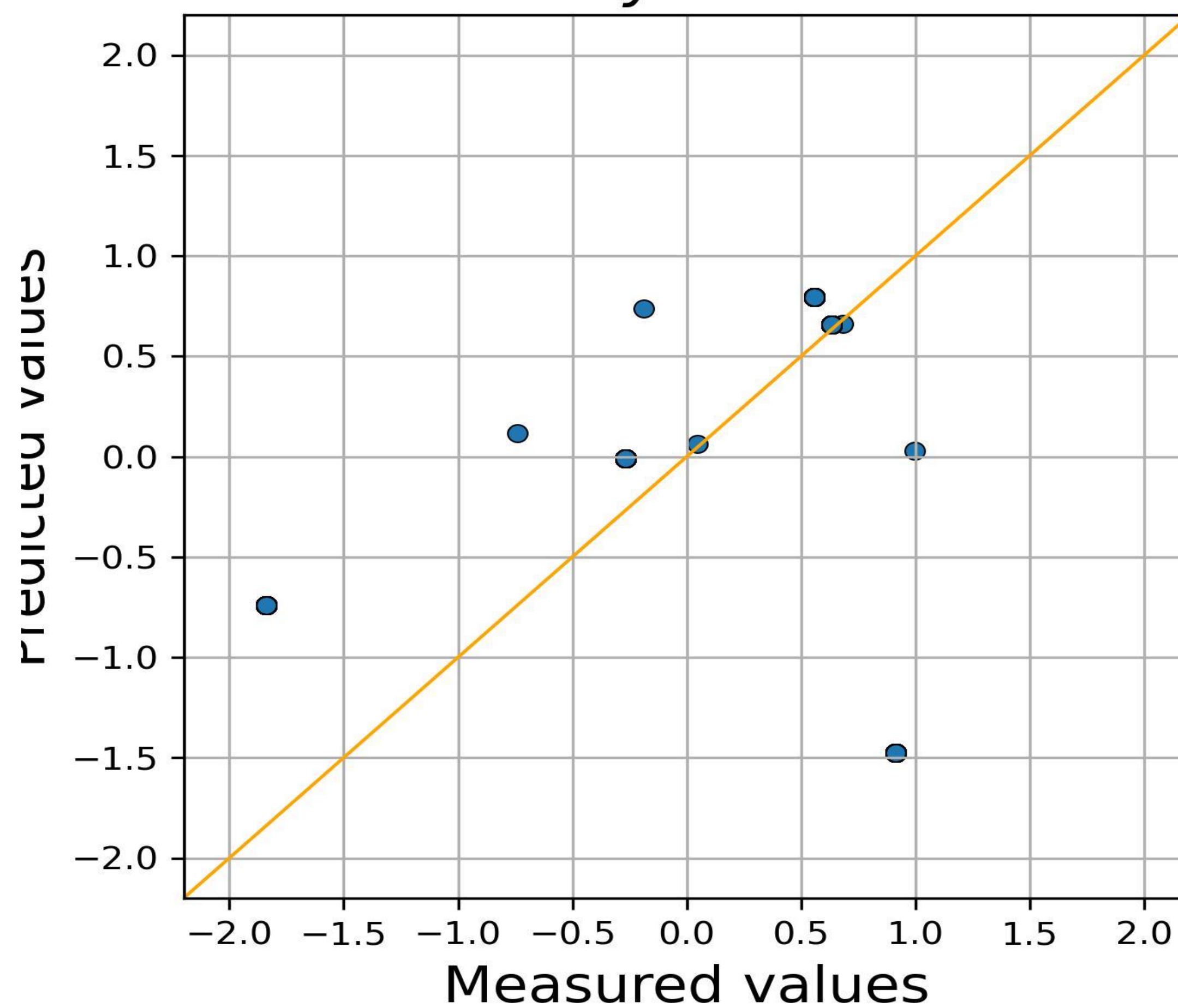


Результаты Chilko

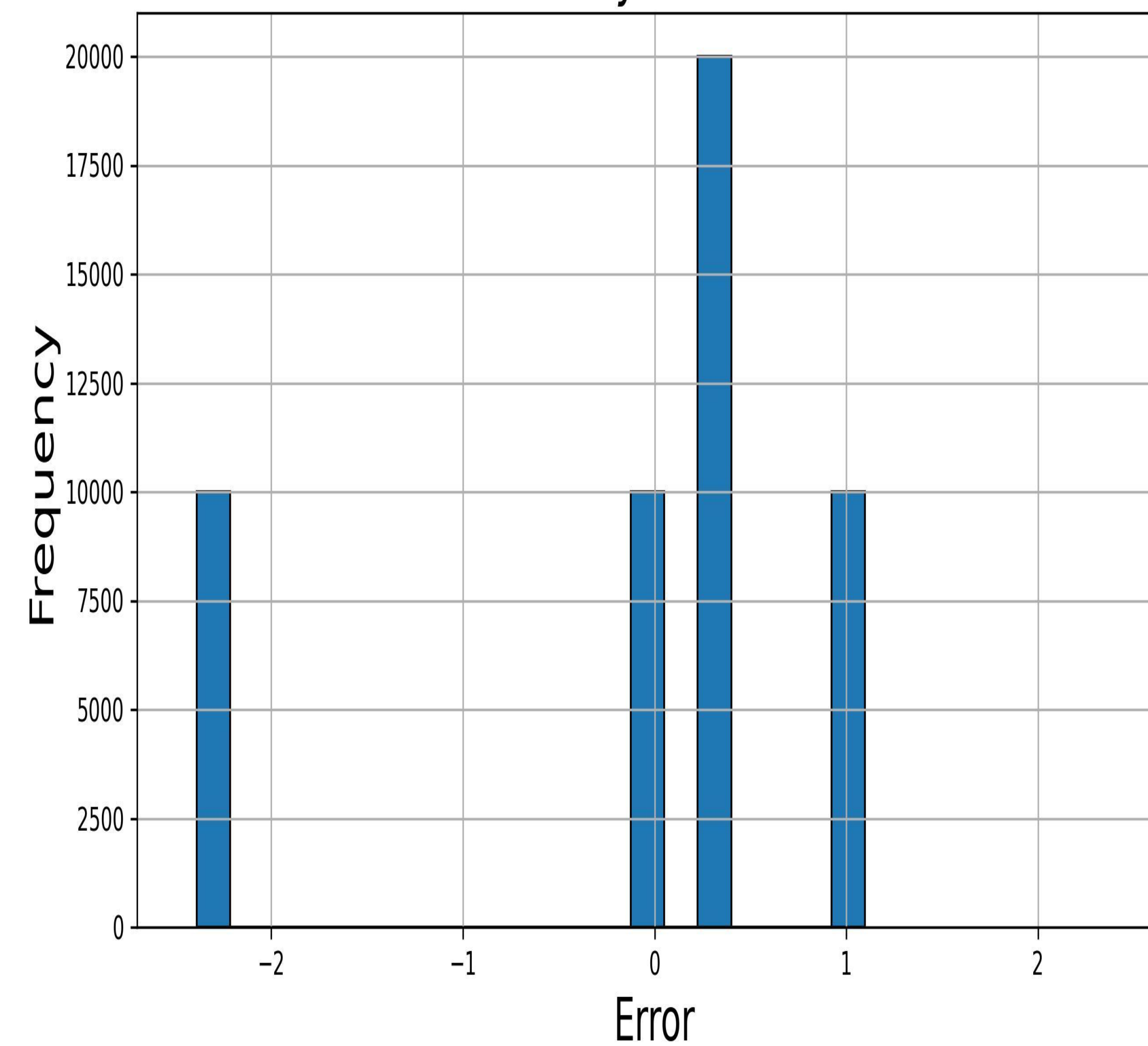


Результаты Early Stuart

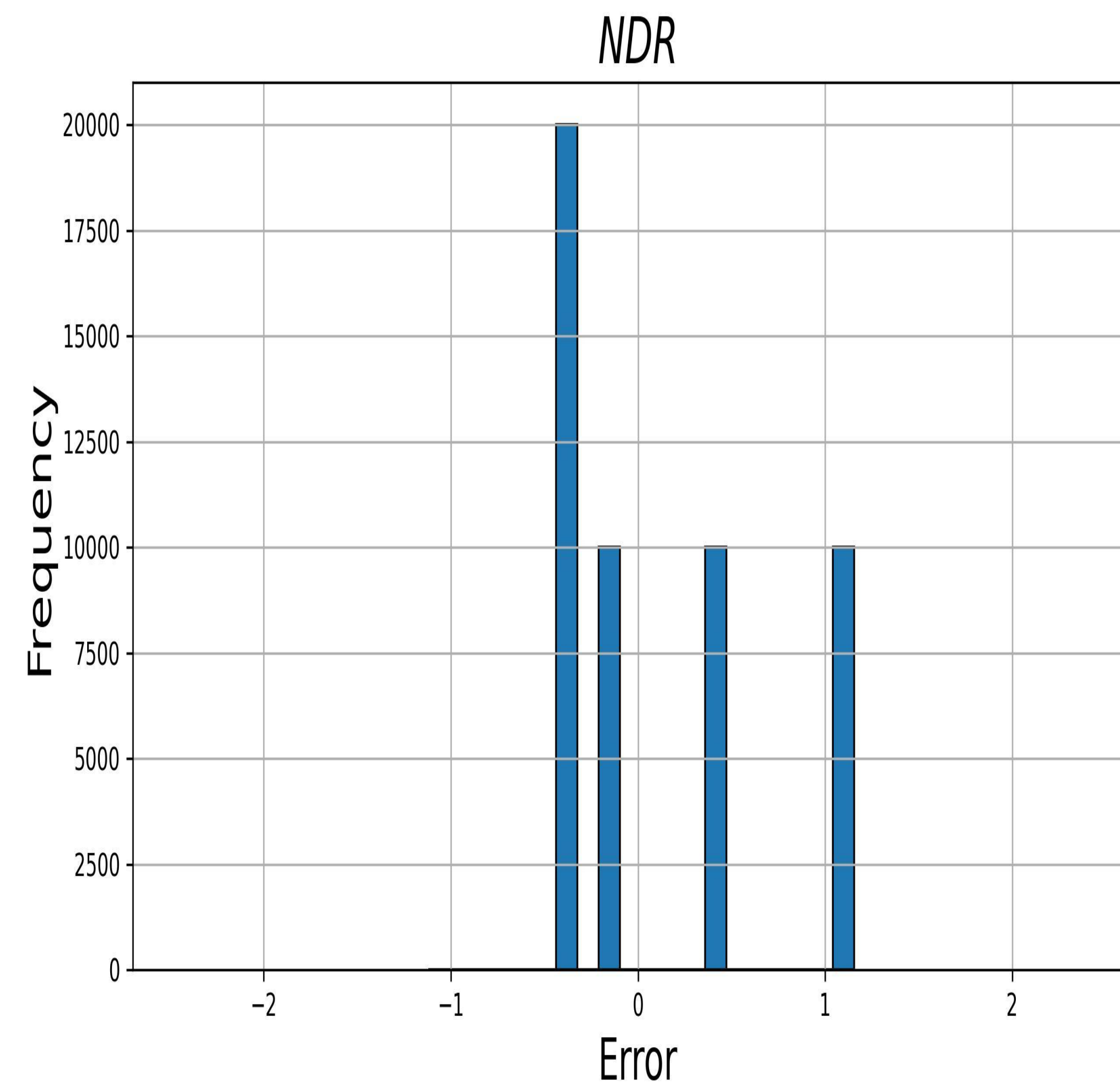
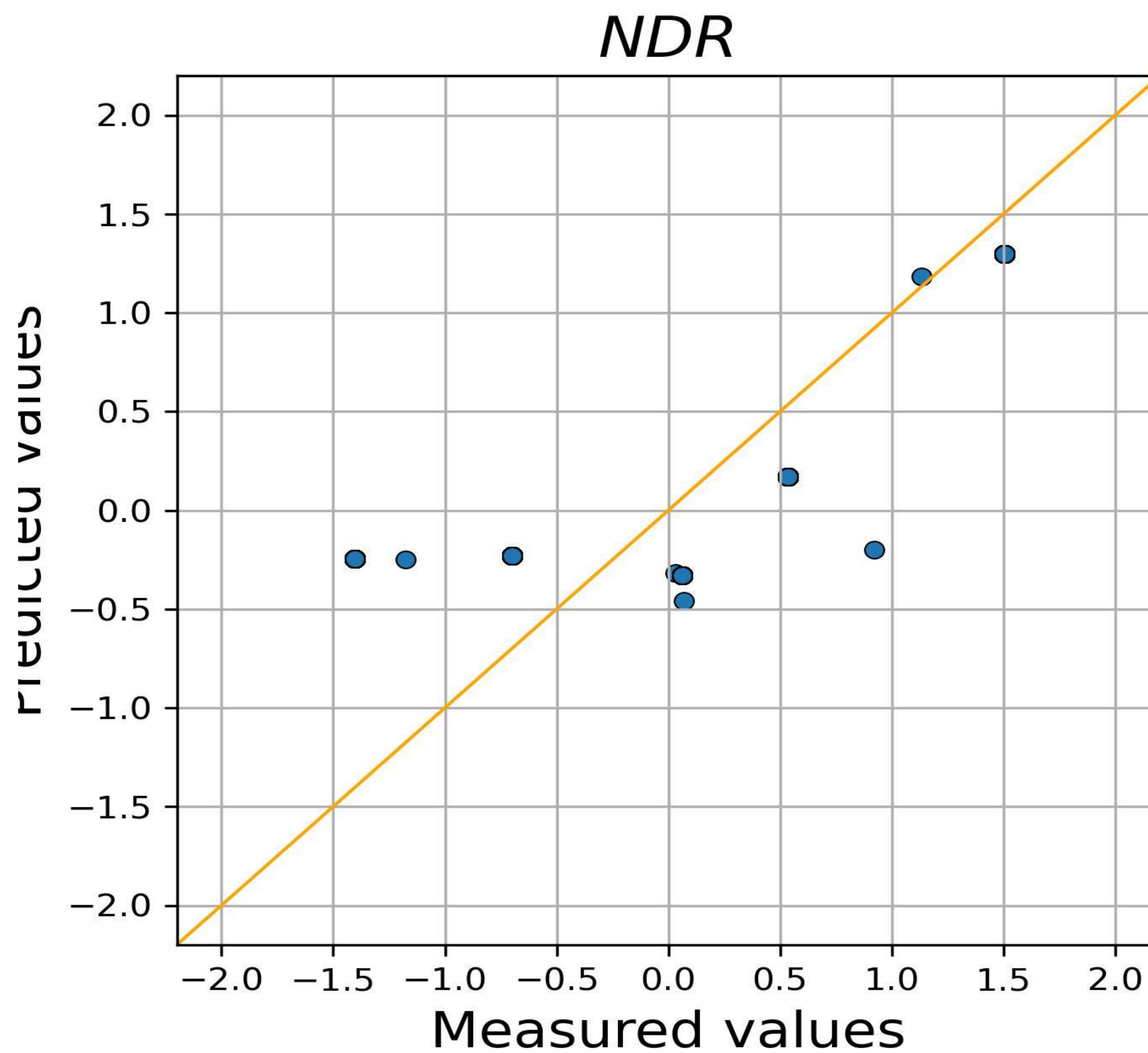
Early stuart

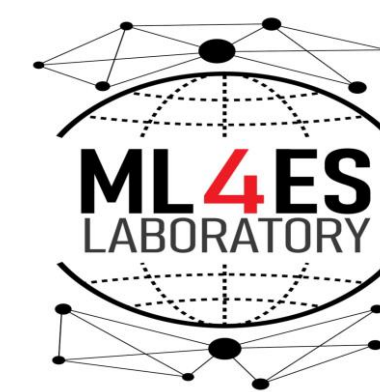


Early stuart



Результаты NDR





Заключение

- Использованы данные реанализа Glorys2v4
- Проведена аугментация и нормализация данных
- Применены классические модели машинного обучения
- Проведена оптимизация выбора контрольных точек



Развитие исследования

- Использование реанализов ERA-5, NARR, CMSS CNOR с большим временным диапазоном.
- Применение Momentum Contrast нейронной сети для извлечения признакового описания в одномерный вектор
- Применение ConvLSTM сетей для прогнозирования без предварительной обработки данных