



Четырнадцатая международная Школа-семинар
«Спутниковые методы и системы исследования Земли»

Восстановление спутниковых данных о концентрации хлорофилла с использованием методов глубокого обучения

Глуховец Д., Курилов О.,
Подосенов А., Ханков И.

25 мая – 1 июня 2026 года, Таруса

Введение

Цель: заполнение пропусков в полученных по данным спутниковых сканеров цвета распределениях концентрации хлорофилла, возникающих из-за облачности, с использованием методов глубокого обучения.

Задачи:

1. Выбор данных и регионов
2. Выбор и обоснование методов
3. Обучение на искусственных пропусках
4. Заполнение пропусков
5. Оценка ошибок

Методы

Для заполнения пространств под облаками использованы технологии сверточных сетей. В классическую архитектуру U-Net , которая учится распознавать пространственные узоры (пятна, фронты, градиенты) и переносит их на пропущенные участки встроены механизм внимания (attention) который в ближнем пространстве и времени учитывает концентрацию хлорофилла и механизм памяти LSTM (Long Short-Term Memory) который учитывает изменения в пределах 12 дневного окна. Для более быстрого и устойчивого обучения модели кадры со спутника обрабатываются простой сверточной сетью которая делает грубое предзаполнение элементами разного масштаба (узкие филаменты, обширные цветения).

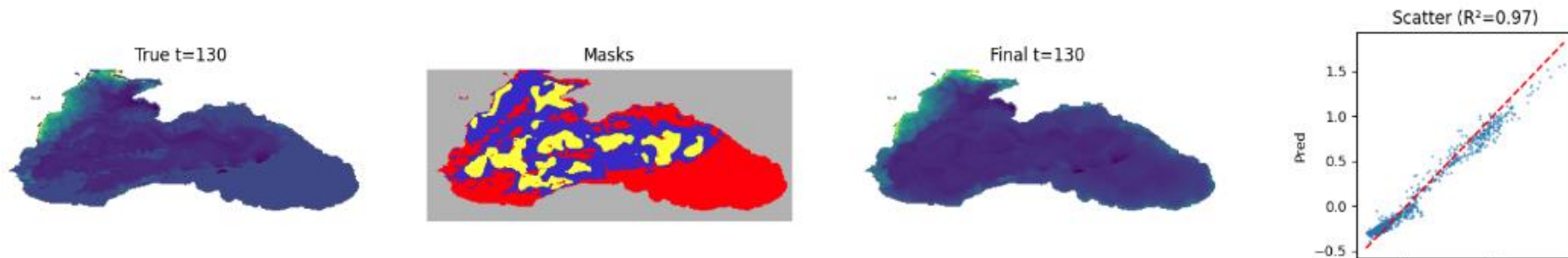
Расчетные мощности и среда

Сценарий	Ресурсы	Время и затраты
Одна карта / проверка месяца	Обычный ПК или ноутбук, среда Python, готовые веса модели, входной .nc	минуты на карту; покупка GPU не нужна
Обработка всего архива Карского моря	ПК с 16 ГБ RAM или Google Colab; папка с .nc, климатология, чекпоинты моделей	десятки минут на GPU или до нескольких часов на CPU
Переобучение под другой регион / сенсор	GPU желательно: T4/L4/A100 или локальная видеокарта; заново собрать train/val/test	часы-дни экспериментов; основные затраты — обучение и валидация

На выходе: восстановленный Chlor_a, карта риска ошибки. Главное ограничение: при переносе на новый регион или новый набор сенсоров модель лучше переобучить и заново проверить на искусственных пропусках. Для применения готовой модели GPU не обязателен; GPU нужен в основном для переобучения и массовых экспериментов.

Искусственные облака

Для обучения сети и определения качества восстановления над чистой водой в кадрах использовано искусственное “заоблачивание” ~30%.



Результаты по Черному морю

В данных L3 со спутника Aqua Modis Черное море представлено матрицами 144x360.

В результате обучения модели на данных за 4 года и предсказании модели в кадрах под искусственными облаками 5 года получены метрики $R^2 = 0.87$ и MAE = 0.036 (в log10 шкале)

Карское море: данные и разбиение по годам

Работали со среднемесячными картами для теплого сезона. Годы разделены

так, чтобы модель не видела будущие карты во время обучения

Обучение
2006–2018

Валидация
2019–2021

Тест
2022–2024

Всего
76 карт

Входные поля:

Chlor_a — хлорофилл (стд)

K24 — хлорофилл (региональный)

ag_443 — показатель поглощения OPOB

I2_flags — флаги качества

месяц — сезонная информация

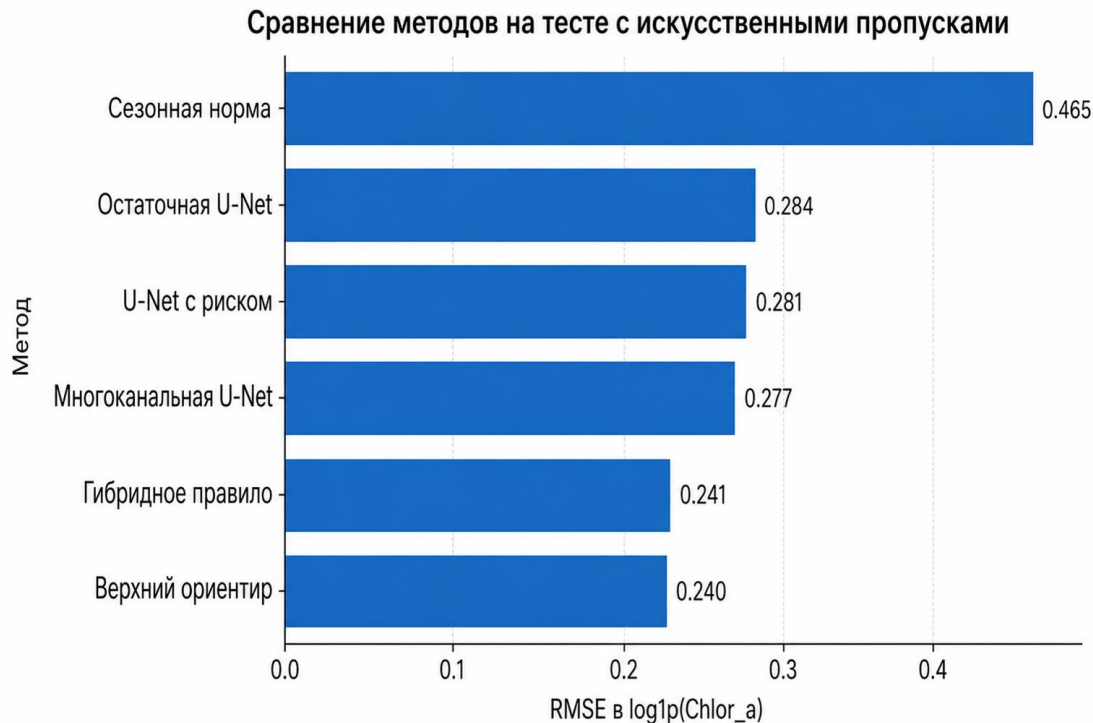
Проверка качества:

Чтобы измерить ошибку, мы скрывали часть известных пикселей.

Сравнение методов

Метод	Что использует	Зачем нужен
Сезонная норма	медианная карта для месяца	простая опорная точка
U-Net (расчет поправки)	сезонная норма + локальная поправка	не учить сезонность заново
Многоканальная U-Net	Chlor_a, K24, ag_443	добавить связанные поля
U-Net с риском	карта + оценка неопределенности	найти случаи, в которых модель ненадежна
Гибридное правило	при высоком риске используем более устойчивую модель	финальный алгоритм

Гибридное правило почти достигает верхнего ориентира

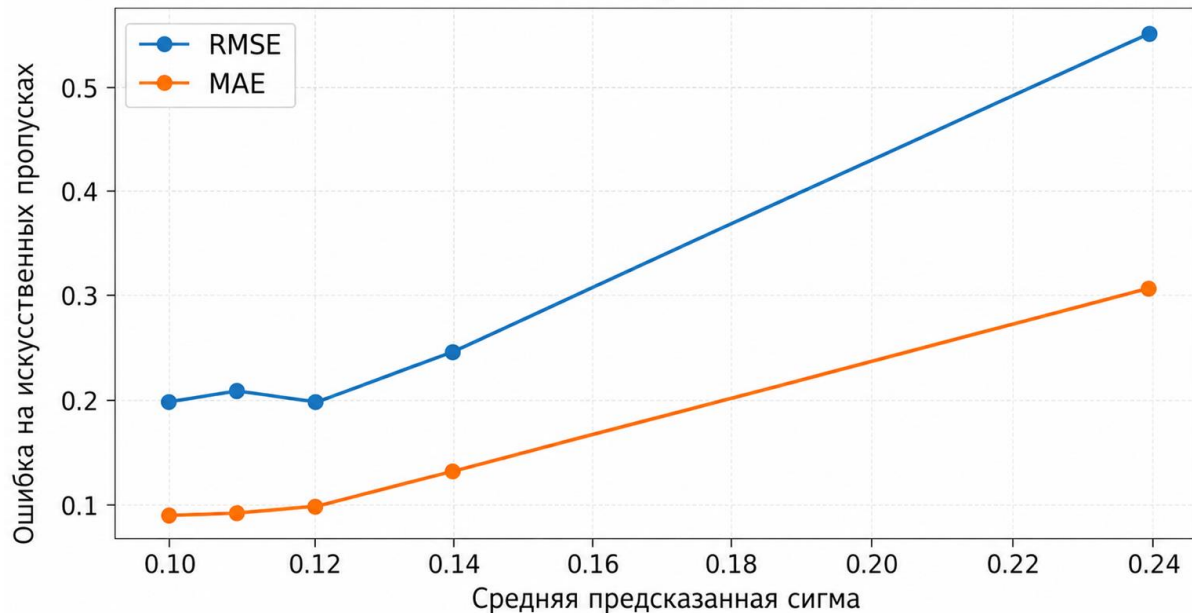


Гибрид — лучший рабочий метод: RMSE 0.241 вместо 0.465 у сезонной нормы.

Верхний ориентир 0.240 — это то насколько хорошо мы могли бы восстановить карту, если бы заранее безошибочно знали, какой из двух вариантов лучше выбрать.

Оценка риска предсказывает ошибки

Надёжность оценки: ошибка растёт с увеличением предсказанной сигмы

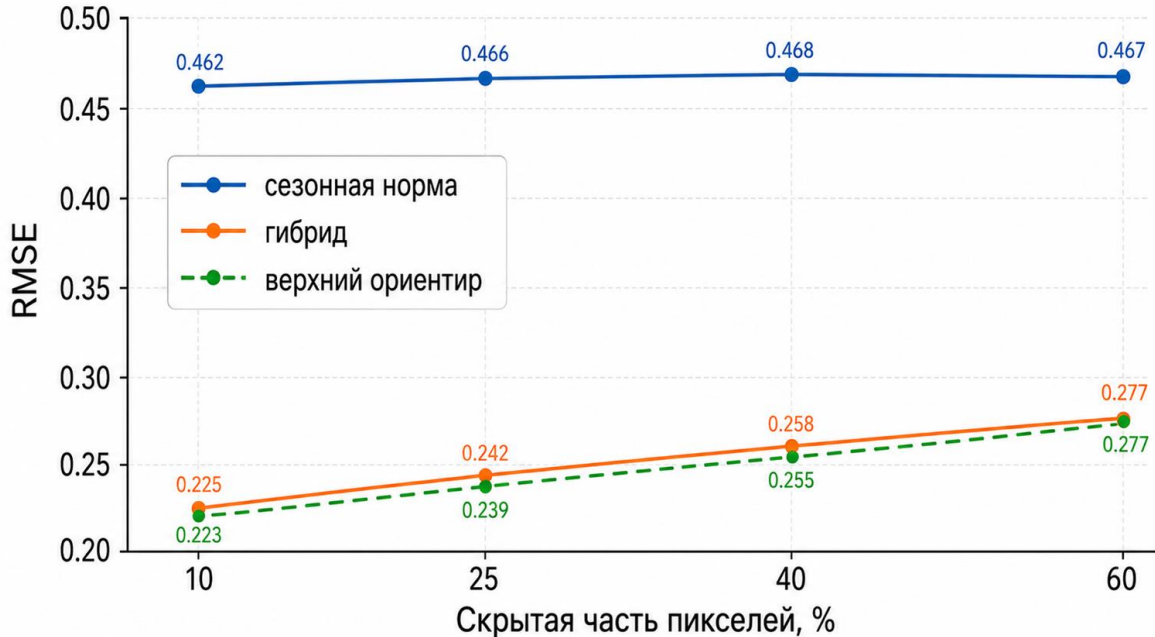


Модель заполняет пропуски и показывает, где восстановлению следует доверять с большей осторожностью.

Эта оценка включает переключение на более устойчивую модель.

Метод устойчив при росте площади пропусков

Сравнение методов на искусственных пропусках

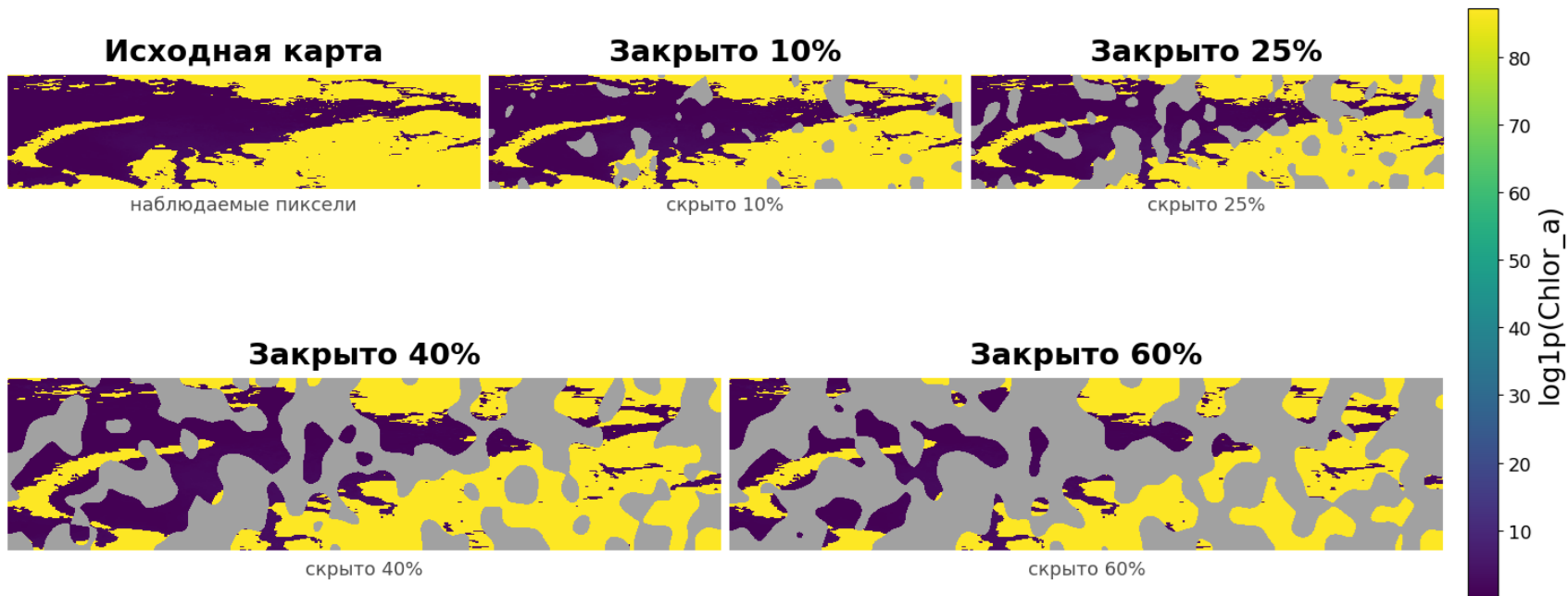


Проверили четыре уровня: от 10% до 60% скрытых пикселей.

Гибридная схема ухудшается плавно и остаётся рядом с верхним ориентиром.

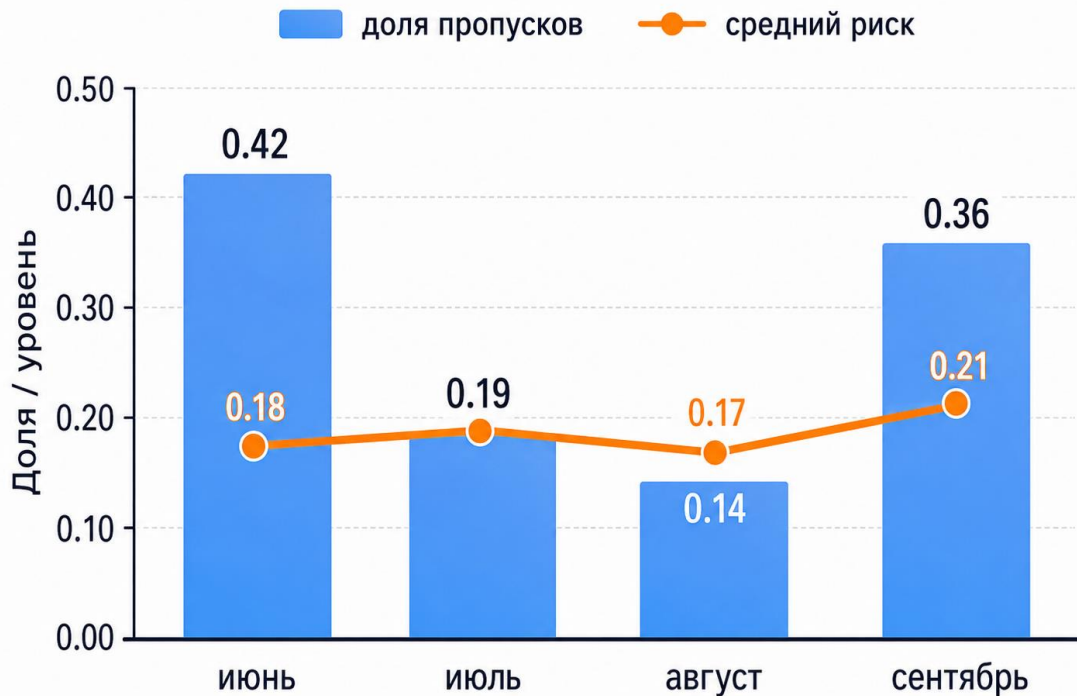
Примеры искусственных пропусков разного покрытия

Искусственные пропуски для проверки модели: июль 2022



Сложность восстановления зависит от месяца

Сезонная изменчивость пропусков и риска

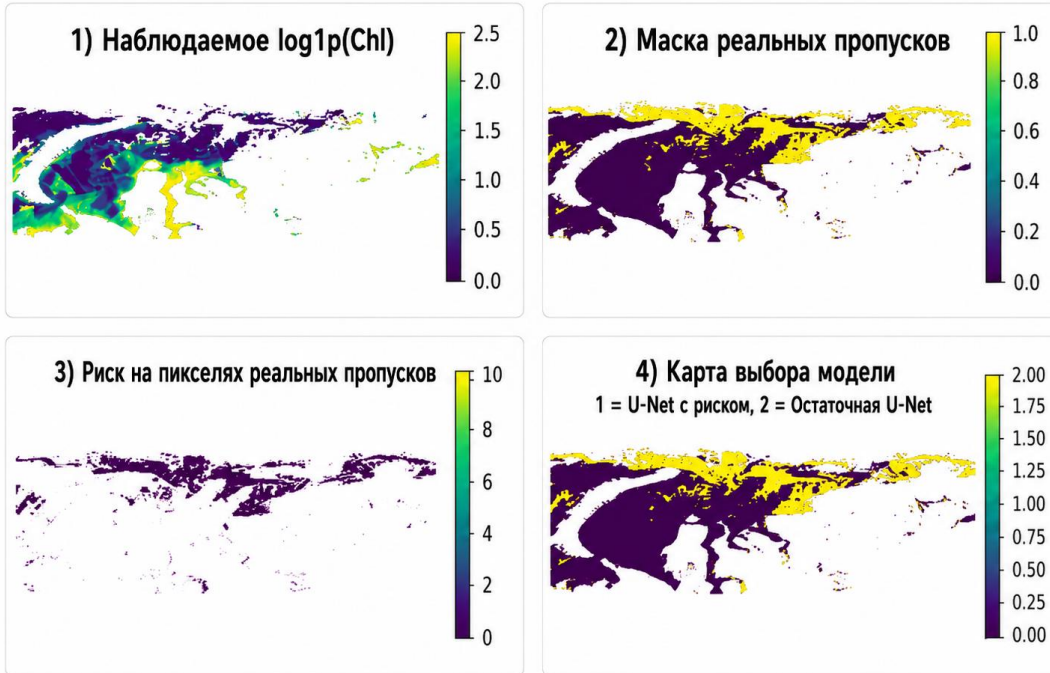


На реальных пропусках нет правильного ответа, поэтому важны контроль домена и карта риска.

Август обычно самый полный месяц. Июнь и сентябрь дают больше пропусков; сентябрь ещё и самый рискованный.

Пример переключения модели

Сентябрь 2023: переключение модели при высоком риске



Карта за сентябрь 2023 года получила самую высокую среднюю оценку риска: 0.316. По заранее выбранному порогу модель переключилась на устойчивое восстановление.

Внутри рабочей морской области после заполнения не осталось пустых пикселей.

Результаты

**RMSE на искусственных
пропусках
0.241**

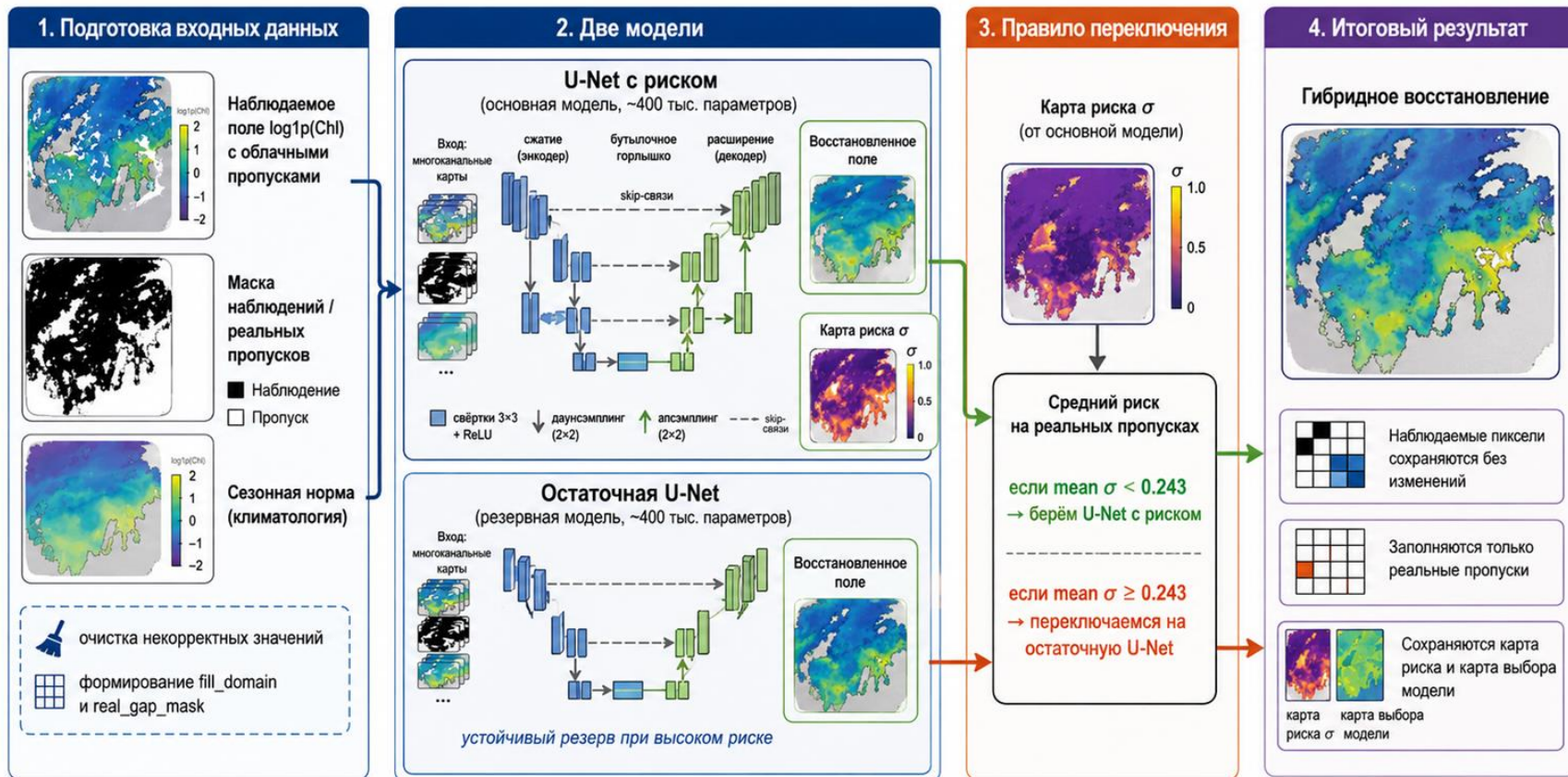
**Верхний ориентир
0.240**

**Объем датасета
76**

Итоговый алгоритм заполняет облачные пропуски, показывает, где восстановление рискованно, и автоматически выбирает более устойчивый вариант. Ошибка снижена на 48% относительно сезонной нормы.

В перспективе можно применить тот же проверенный контур к Баренцеву морю.

Схема работы контура



Планы на будущее

- добавление данных о ТПМ (непрерывные GHRSSST)
- обучение на двумерной модели роста и распространения фитопланктона?