

«Исследование методов интерполяции данных по температуре
поверхности моря, полученных по данным дистанционного
зондирования Земли»

Сорокин Андрей Евгеньевич, НИУ «МИЭТ»

Введение

Постановка проблемы

Одной из ключевых проблем при анализе температуры поверхности моря (ТПМ) является нерегулярность и неполнота исходных данных, получаемых со спутников.

Актуальность

Использование геостатистического метода кригинга [1,2] (позволяющего не только интерполировать данные на регулярную сетку, но и оценивать точность восстановления) и детерминированного метода обратных взвешенных расстояний [3] критически важно для повышения качества прогнозов.

Научная новизна

Впервые выполнено сравнение методов универсального кригинга и обратных взвешенных расстояний для восстановления ТПМ с оценкой влияния геометрии пропусков.

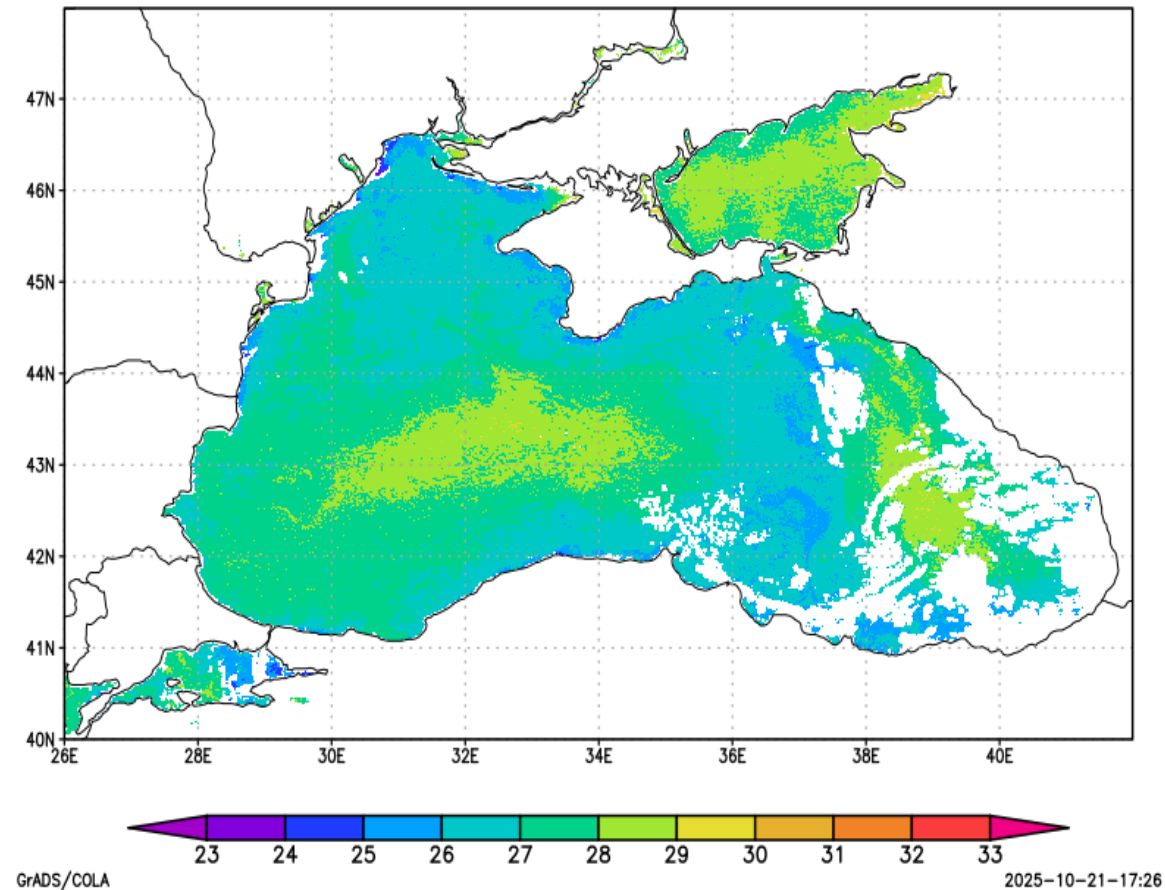


Рис. 1 – Пример пропусков данных

Цели и задачи исследования

Цель работы

Исследование, реализация и сравнительный анализ эффективности методов обычного и универсального кригинга для интерполяции полей температуры поверхности моря (ТПМ) на регулярную сетку.

Основные задачи

- Реализовать алгоритмы обычного, универсального кригинга и метода обратных взвешенных расстояний для восстановления пропущенных значений ТПМ;
- Для оценки точности методов провести серию твин-экспериментов на реальных данных спутника Aqua, представляющих собой мгновенные поля (снимки), пересчитанные на регулярную сетку размером 1600x800 узлов с шагом 0.01x0.01 градуса, покрывающую акватории Черного, Азовского и Мраморного морей [4];
- Оценить влияние размера области пропуска на качество восстановления;
- Исследовать влияние геометрии пропусков;
- Определить области применения методов.

Обычный кригинг (ОК)

Построили экспериментальную вариограмму по формуле:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (z(x_i + h) - z(x_i))^2, \text{ где:}$$

$\gamma(h)$ — значение экспериментальной вариограммы для расстояния h .

h — вектор расстояния между парой точек данных.

$N(h)$ — количество пар точек данных, разделенных вектором h .

$z(x_i)$ — измеренное значение температуры в точке x_i .

$z(x_i + h)$ — измеренное значение температуры в точке, удаленной от x_i на вектор h .

Подбор теоретической модели

Таблица 3 – Теоретические вариограммы
для обычного кригинга

- Гауссова модель

$$\gamma(h) = c_0 + c \left(1 - \exp \left(-\frac{3h^2}{a^2} \right) \right)$$

- Сферическая модель

$$\gamma(h) = \begin{cases} c_0 + c \left[\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right], & h \leq a \\ c + c_0, & h > a \end{cases}$$

- Экспоненциальная модель

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + (c - c_0) \left[1 - \exp \left(-\frac{3h}{a} \right) \right], & h \neq 0 \end{cases}$$

Модель вариограммы	Коэффициент детерминации, R ²
Гауссова	0.8941
Сферическая	0.8538
Экспоненциальная	0.7859

Система уравнений обычного кригинга

Решили систему уравнений:
$$\begin{cases} \sum_{j=1}^{n(x)} \omega_j \gamma(x_i, x_j) + \mu = \gamma(x_i, x_0), & i = 1, \dots, n \\ \sum_{j=1}^{n(x)} \omega_j = 1 \end{cases},$$

где:

- ω_j - искомые веса;
- γ - значение вариограммы;
- μ - искомый параметр Лагранжа;
- n - количество точек в окрестности прогнозируемой точки.

Построили оценку в прогнозируемой точке: $Z^*(x_0) = \sum_{i=1}^n \omega_i Z(x_i)$

Универсальный кригинг (УК)

Представим $Z(x)$ в виде:

$$Z(x) = m(x) + R(x),$$

$m(x) = \sum_{k=0}^k a_k f_k(x)$, где $f_k(x)$ – базисная функция;

$$E[R(x)] = 0.$$

В качестве тренда была взята линейная базисная функция: $m(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$

Построили вариограмму ошибок: $\gamma_R(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (R(x_i + h) - R(x_i))^2$

Приблизили вариограмму теоретической и посчитали коэффициенты c_0 , c , a

Универсальный кригинг

Таблица 4 – Сравнение теоретических вариограмм для универсального кригинга

Модель вариограммы	Коэффициент детерминации, R ²
Гауссова	0.8903
Сферическая	0.8858
Экспоненциальная	0.8186

Переход от вариограммы к ковариации

$$C_R(h) = c - \gamma_R(h) + c_0$$

Решили систему уравнений:

$$\left\{ \begin{array}{l} \sum_{j=1}^{n(x)} \omega_j C_R(x_i - x_j) + \sum_{k=0}^K \mu_k f_k(x_i) = C_R(x_i - x), i = 1, \dots, n(x) \\ \sum_{j=1}^{n(x)} \omega_j = 1 \\ \sum_{j=1}^{n(x)} \omega_j f_k(x_j) = f_k(x), \quad k = 0, \dots, K \end{array} \right.$$

Построили оценку в прогнозируемой точке:

$$Z^*(x_0) = \sum_{i=1}^n \omega_i Z(x_i)$$

Метод обратных взвешенных расстояний

Математическая постановка задачи: имеется набор из N опорных точек $\{x_i | i = 1, 2, \dots, N\}$ с известными значениями в этих точках $Z(x_i)$. Интерполированное значение $Z(x_0)$ рассчитывается по следующей формуле:

$$Z(x_0) = \frac{\sum_{i=1}^N w_i Z(x_i)}{\sum_{i=1}^N w_i},$$

где:

- N – количество опорных точек (известных значений), участвующих в расчете;
- $Z(x_i)$ – измеренное значение в i -й опорной точке;
- w_i – вес i -й опорной точки.

Весовая функция задается формулой:

$$w_i = \frac{1}{d_i^p},$$

где:

d_i – Евклидово расстояние от интерполируемой точки до i -й опорной точки;

p – степенной показатель.

Таблица 5 – Подбор степенного параметра p

p	MAE, °C
1	0.3236
1.5	0.3352
2	0.2836
2.5	0.3287
3	0.3527
4	0.3086

Эксперимент 1. Описание

Проведен следующий эксперимент: удалена и восстановлена температура для 5 квадратных областей размером 50х50 узлов сетки (у. с.). Вычислена средняя квадратичная ошибка(MSE) и средняя абсолютная ошибка (MAE) за 50 экспериментов.

$$MAE = \frac{1}{N} \sum_{i=1}^N |Z^*(x_i) - Z(x_i)|$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (Z^*(x_i) - Z(x_i))^2$$

Эксперимент 1. Результаты

Таблица 6 – Оценка качества общего восстановления температуры для 5 областей при помощи метрик MAE и MSE

Метод интерполяции	MSE, (°C) ²	MAE, °C
Обычный кригинг	0.3050	0.4053
Универсальный кригинг	0.2308	0.3707
IDW	0.1137	0.2339

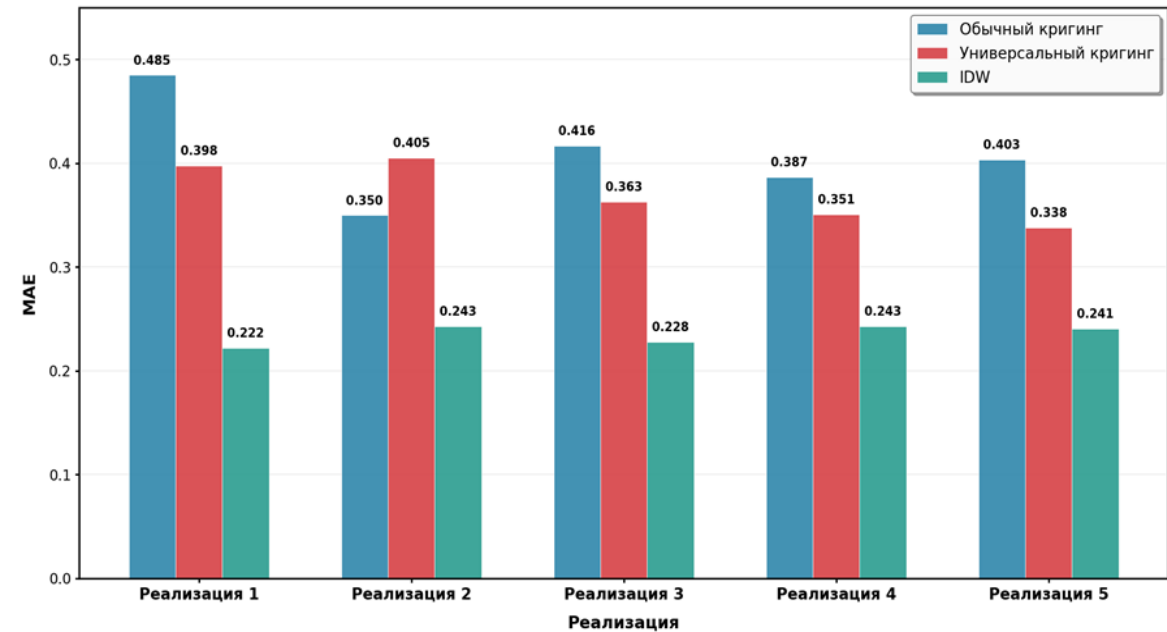


Рис. 6 – Оценка качества восстановления каждой реализации при помощи метрик MSE

Эксперимент 2. Описание.

Для того, чтобы вычислить максимальную по площади квадратную область, в которой температура восстанавливается с относительной погрешностью $\delta \leq 1.5\%$ проведено 20 экспериментов с извлечением областей разных размеров.

$$\delta = \frac{MAE}{AVG(T)}$$

$$MAE = \frac{1}{20} \sum_{i=1}^{20} MAE_i$$

где MAE_i – средняя абсолютная ошибка i -го эксперимента

$$AVG(T) \approx 27,72 \text{ }^{\circ}\text{C}$$

Эксперимент 2. Результаты

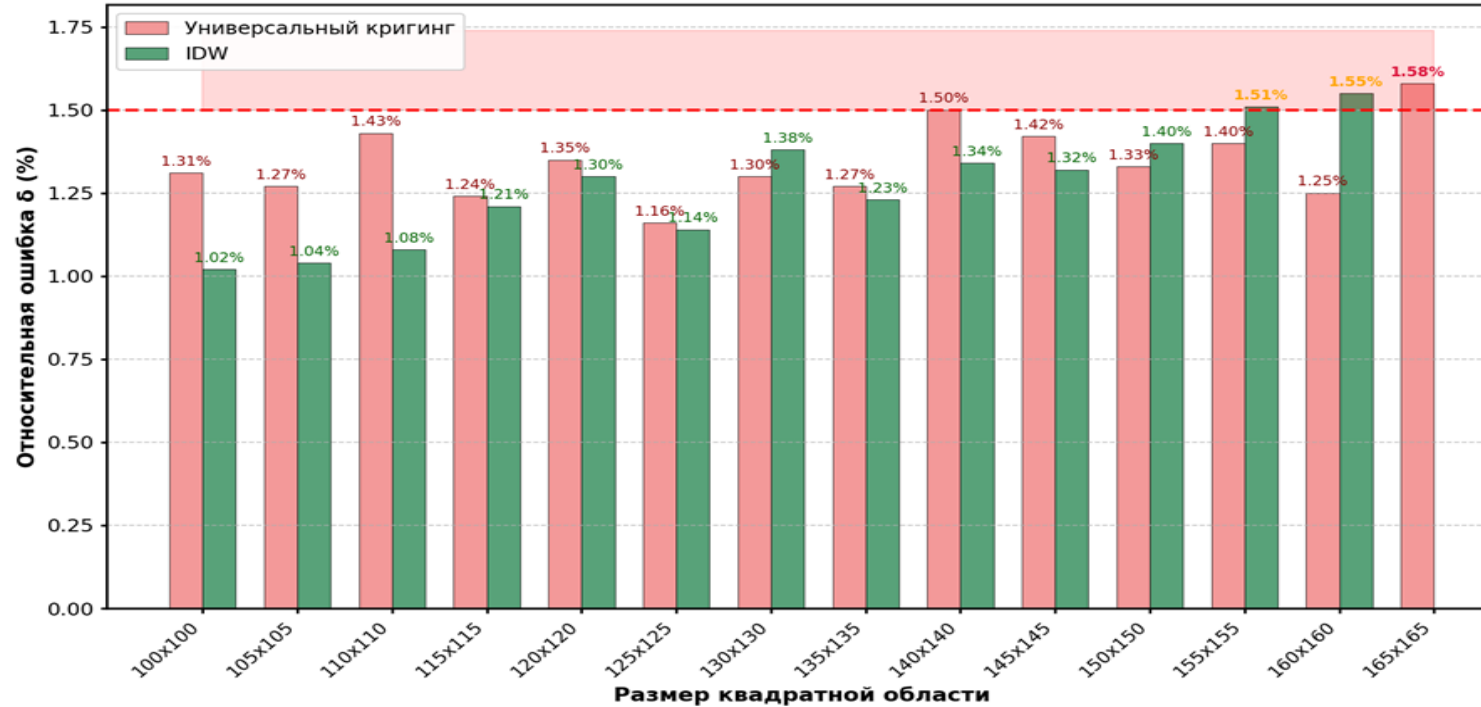
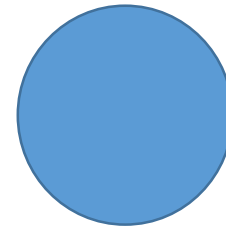


Рис. 7 – Сравнение относительной ошибки UK и IDW при восстановлении температуры для областей разных размеров

Эксперимент 3. Описание

Для того, чтобы выявить зависит ли точность восстановления температуры от формы области, проведено 20 тестов и рассмотрены следующие фигуры, примерно равные по площади: квадрат (50×50 (у. с.)²), прямоугольник (125×20 ((у. с.)²), прямоугольник (100×25 (у. с.)²), прямоугольник (25×100 (у. с.)²), прямоугольник (20×125 (у. с.)²), Прямоугольник (40×62 (у. с.)²), прямоугольник (62×40 (у. с.)²), круг ($r = 28$ (у. с.)).



Эксперимент 3. Результаты

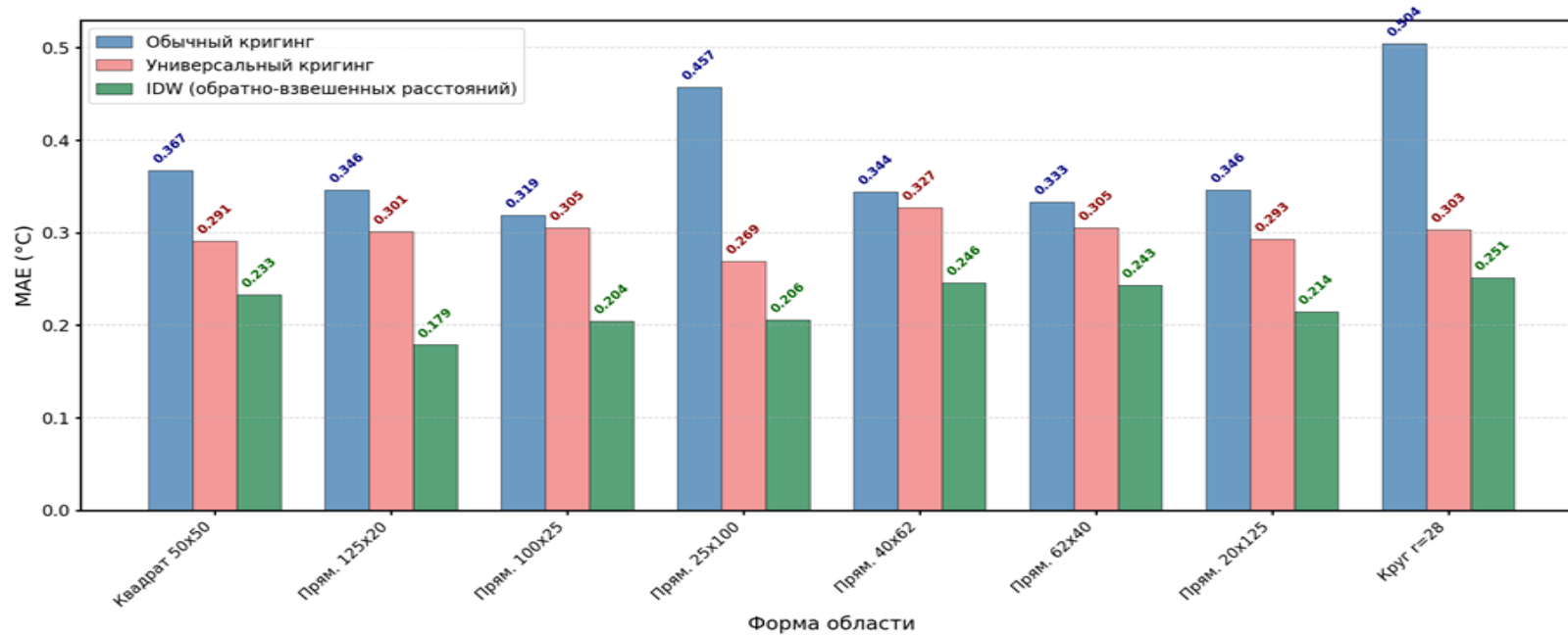


Рис. 8 – Сравнение средней абсолютной ошибки ОК, УК и IDW при восстановлении температуры для областей разных форм.

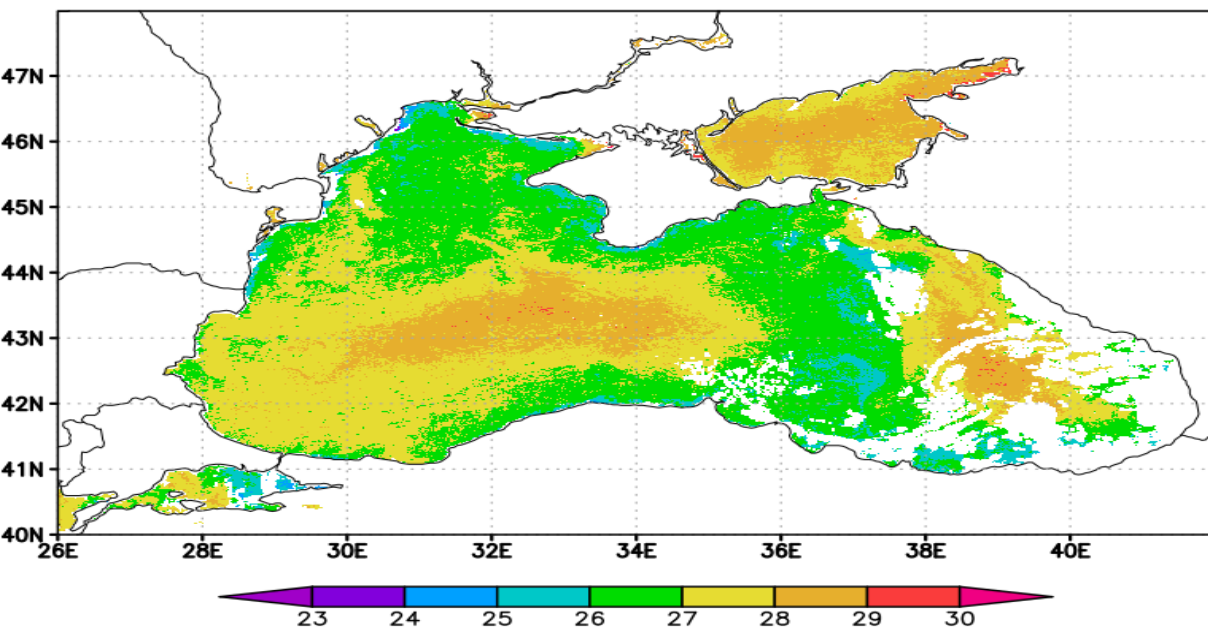


Рисунок 9 – Данные о ТПМ со спутника Terra от 01.08.2021 19:45 UTC

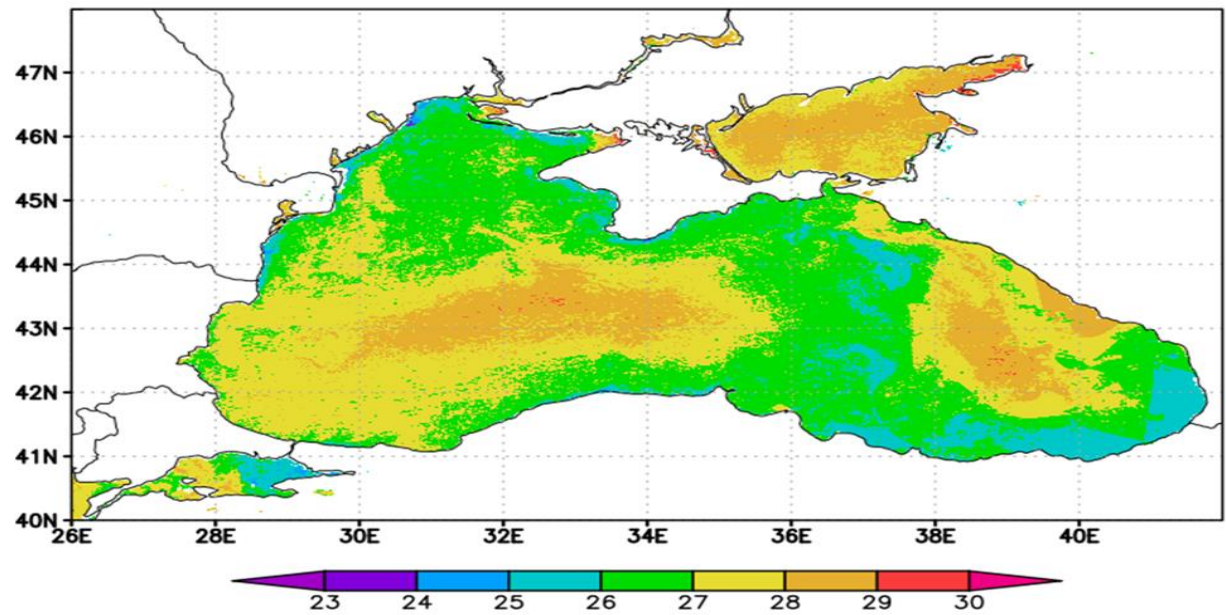


Рисунок 10 – Данные о ТПМ со спутника Terra от 01.08.2021 19:45 UTC с восстановленными пропусками методом UK

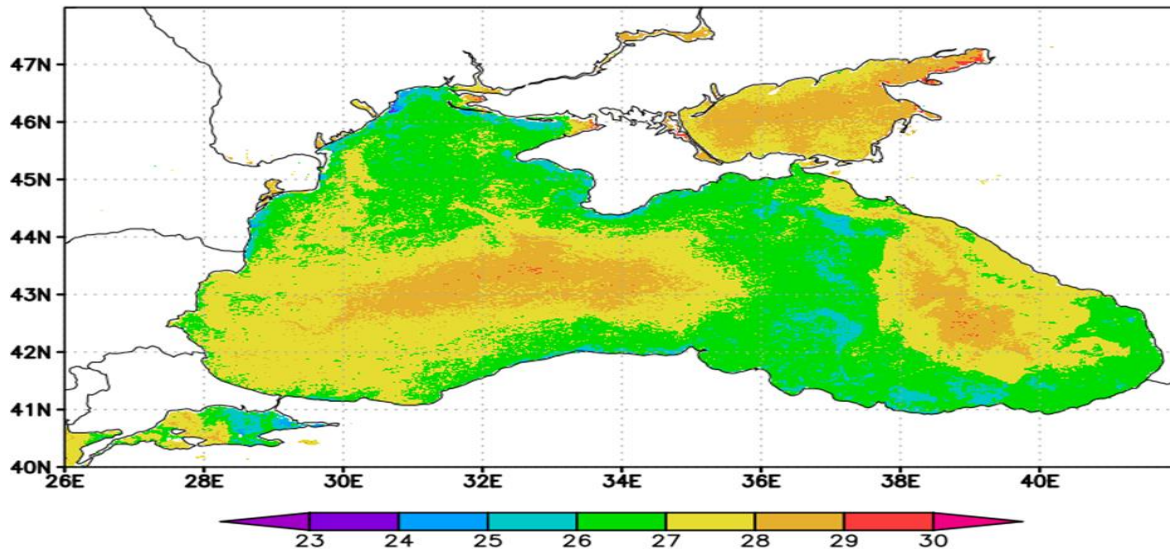


Рисунок 11 – Данные о ТПМ со спутника Terra от 01.08.2021 19:45 UTC с восстановленными пропусками методом IDW

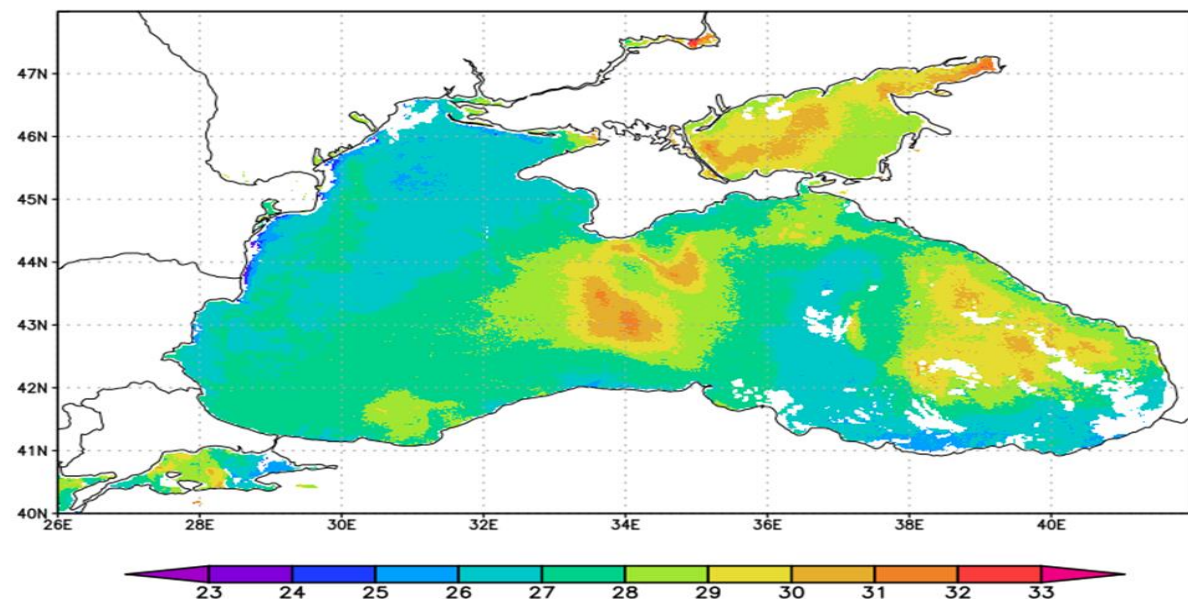


Рисунок 12 – Данные о ТПМ со спутника Aqua от 01.08.2021 11:00 UTC

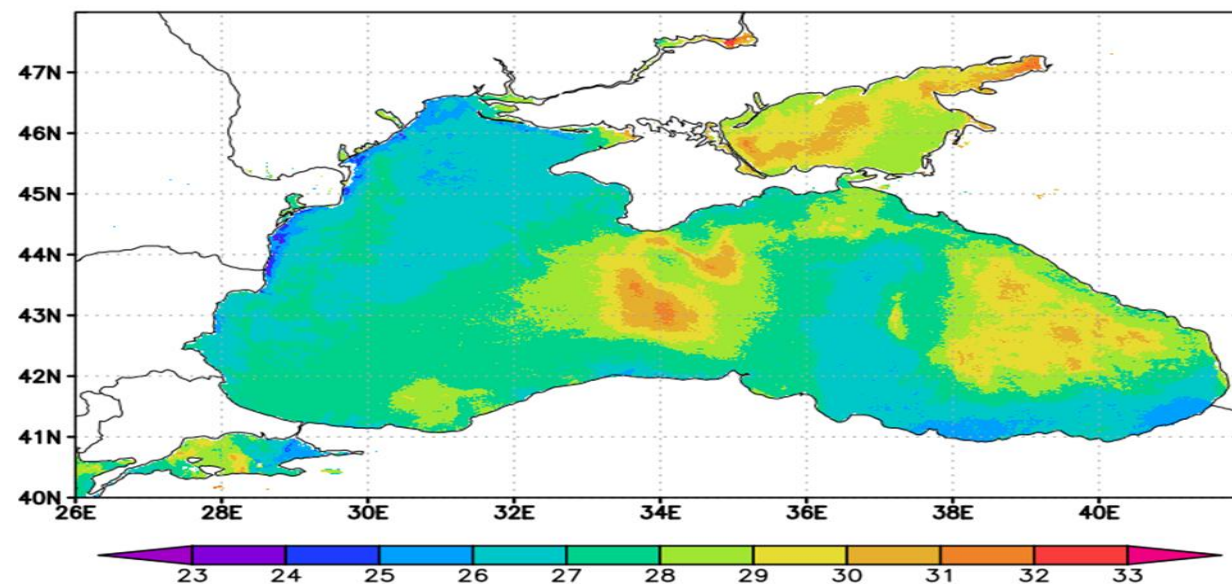


Рисунок 13 – Данные о ТПМ со спутника Aqua от 01.08.2021 11:00 UTC с восстановленными пропусками методом UK

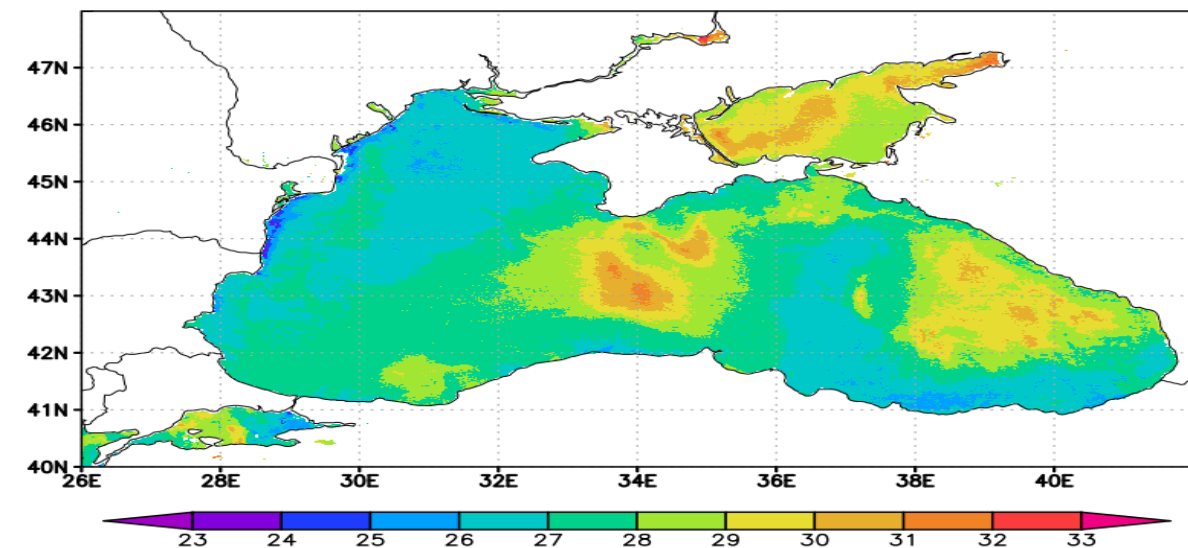


Рисунок 14 – Данные о ТПМ со спутника Aqua от 01.08.2021 11:00 UTC с восстановленными пропусками методом IDW

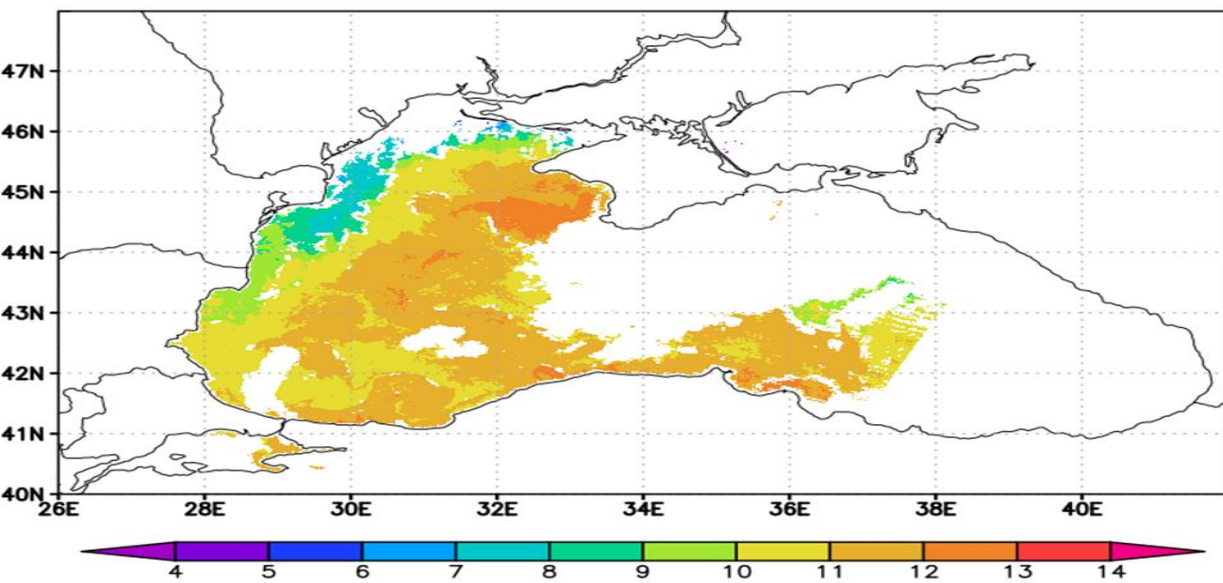


Рисунок 15 – Данные ТПИМ со спутника Aqua от 01.01.2023 00:40 UTC

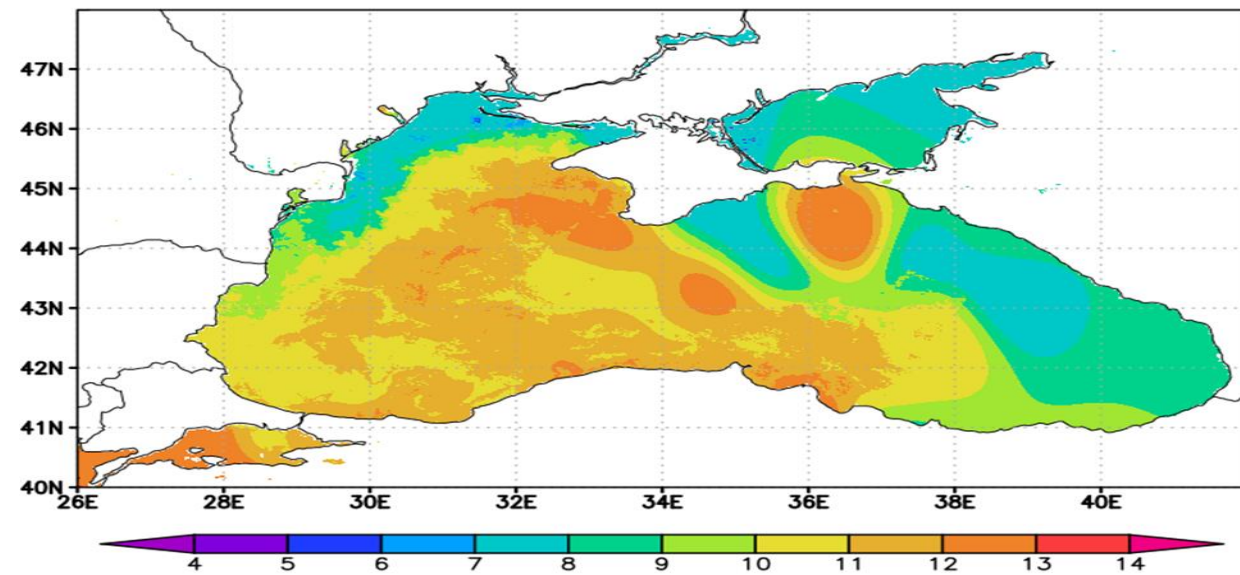


Рисунок 16 – Данные ТПИМ со спутника Aqua от 01.01.2023 00:40 UTC с восстановленными пропусками методом UK

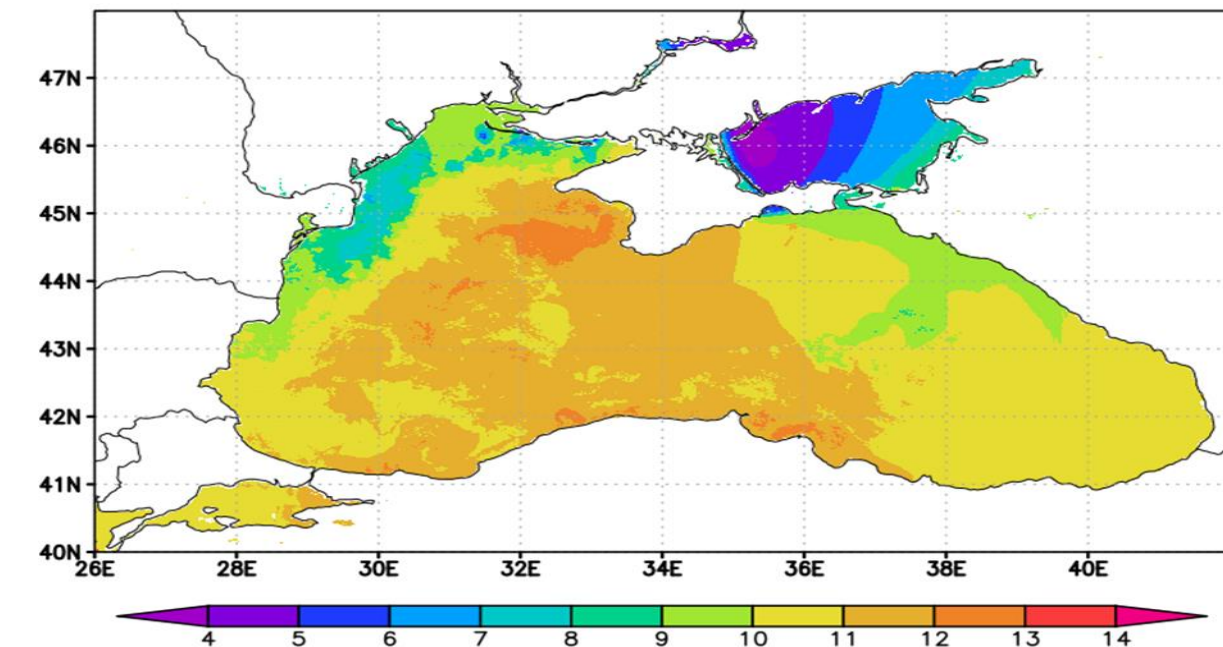


Рисунок 17 – Данные ТПИМ со спутника Aqua от 01.01.2023 00:40 UTC с восстановленными пропусками методом IDW

Эксперимент 4. Результаты

Таблица 7 – Результаты эксперимента 4

Снимок	Е (σ_{UK}^2), (°C) ²	95 % доверительный интервал, °C
Terra от 01.08.2021 19:45 UTC	0.092954	±0.5976
Aqua от 01.08.2021 11:00 UTC	0.130769	±0.7088
Aqua от 01.01.2023 00:40 UTC	0.581218	±1.4943

$$\sigma_{UK}^2(x) = C_R(0) - \sum_{i=1}^{n(x)} \lambda_i C_R(x_i - x) - \sum_{k=0}^K \mu_k f_k(x).$$

Заключение

По результатам работы можно сделать следующие выводы:

- Метод обратных взвешенных расстояний имеет преимущество над универсальным и обычным кригингом для восстановления температуры небольших и геометрически сложных пропусков (Эксперименты 1, 3)
- Универсальный кригинг имеет преимущество над обычным, а также над методом обратных взвешенных расстояний для восстановления крупных областей с температурой (эксперимент 2)
- Универсальный кригинг в отличие от метода обратных взвешенных расстояний позволяет получать статистическую оценку точности восстановления отсутствующих данных (эксперимент 4)