

Г. В. Сутурин¹✉

Автоматизированный выбор метода интерполяции пространственных данных на основе перекрестной проверки

¹ Сибирский государственный университет геосистем и технологий,
г. Новосибирск, Российская Федерация
e-mail: grigroijsuturin@gmail.com

Аннотация. Разработан алгоритм автоматического выбора метода пространственной интерполяции для построения цифровых моделей поверхности по неравномерным измерениям на основе пятикратной перекрестной проверки. Программа автоматически выбирает последовательность интерполяционных алгоритмов и методов сглаживания по значениям средних погрешностей. Перекрестная проверка не является обязательным этапом построения модели поверхности, а используется как инструмент независимого контроля точности. Испытания на синтетической модели чаши водохранилища со сложным извилистым руслом показали преимущество кубического сплайна. При минимальной погрешности общего объема (0,04 %) локальные ошибки высотных отметок в русловой части достигают 1,21 м из-за компенсации разнознаковых отклонений. Для гауссовой модели вариограммы зафиксирована численная неустойчивость вычислений. Разработанный алгоритм может применяться для различных типов пространственных данных и обработки произвольных пространственных выборок.

Ключевые слова: пространственная интерполяция, перекрестная проверка, кригинг, цифровая модель рельефа, вариограмма, синтетическая модель, водохранилище, среднеквадратическая ошибка

G. V. Suturin¹✉

Automated selection of spatial data interpolation method based on cross-validation

¹ Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation
e-mail: grigroijsuturin@gmail.com

Abstract. An algorithm for the automated selection of a spatial interpolation method has been developed for constructing digital surface models from non-uniform measurements based on five-fold cross-validation. The software automatically selects interpolation algorithms and smoothing methods based on mean error values. Cross-validation serves as a tool for independent accuracy control. Tests on a synthetic reservoir model with a complex, winding channel demonstrated the superiority of the cubic spline. Despite a minimal overall volume calculation error (0,04 %), local elevation errors in the channel reach 1,21 m due to the cancellation of multi-directional deviations. Numerical instability was observed for the Gaussian variogram model. The algorithm is applicable to various types of spatial data and arbitrary datasets.

Keywords: spatial interpolation, cross-validation, kriging, digital elevation model, variogram, synthetic model, reservoir, root mean square error

Введение

Переход от нерегулярных пространственных измерений (точечных геодезических пикетов и глубин) к регулярной цифровой модели рельефа фундаментален для геодезии [1-3], гидрографии, горного дела, почвоведения [4] и геофизики [5, 6]. Современное программное обеспечение предлагает широкий спектр детерминированных и геостатистических алгоритмов, классифицируемых по условиям, режимам и виду интерполятора [7]. Однако на практике выбор метода часто субъективен и опирается на настройки по умолчанию, что может искажать расчетные характеристики поверхностей на 3–7 % и более [4]. При этом простое уменьшение шага расчетной сетки не компенсирует методологические ошибки – определяющим фактором остаются качество и плотность исходных данных [8]. В связи с этим автоматизация построения регулярных сеток крайне актуальна [1].

Цель работы – разработка алгоритма автоматического выбора оптимального метода интерполяции на основе статистических критериев. В отличие от традиционных подходов, где параметры вариограммы адаптируются внутри блоков валидации, здесь каждая модификация кригинга рассматривается как независимый интерполятор, что исключает смещение оценок при подгонке модели.

Синтетическая модель рельефа

Для контролируемого эксперимента разработана аналитическая модель чаши водохранилища (500×500 м), позволяющая вычислять истинные ошибки интерполяции в любой точке. Она аналитически воспроизводит речную долину с продольным уклоном, параболическим профилем, меандрирующим руслом, холмами и зоной заиления. Эллиптическая геометрия модели задает выраженную пространственную анизотропию. Максимальная глубина составляет 49,6 м при отметке нормального подпорного уровня 80,6 м.

Из модели извлечена неравномерная выборка из 600 точек: 50 % распределены по спиральным галсам эхолота, 50 % – случайно. Высотный вектор искажен случайным белым шумом со среднеквадратическим отклонением 0,03 м.

Методы интерполяции

Реализованы восемь методов, охватывающих основные классы интерполяционных алгоритмов: детерминированные – линейный, ближайшего соседа, кубический сплайн; геостатистические – обычный кригинг со сферической, экспоненциальной, гауссовой и линейной моделями вариограммы. Методы интерполяции V класса по классификации [7], к которым относится кригинг, используют вариограмму в качестве меры пространственной изменчивости. Выбор модели вариограммы является критически важным этапом геостатистического моделирования [5].

Программная реализация детерминированных алгоритмов выполнена на языке Python с использованием математических модулей библиотеки SciPy [9]. Логика построения вычислительных блоков ядра опирается на принципы создания независимых расчетных модулей с открытым исходным кодом [10].

При реализации геостатистического моделирования в качестве базовой использовалась сферическая модель полувариации $\gamma(h)$, описываемая уравнением:

$$\gamma(h) = \begin{cases} C_0 + C \left(\frac{3h}{2a} - \frac{h^3}{2a^3} \right), & \text{при } h \leq a, \\ C_0 + C, & \text{при } h > a \end{cases}, \quad (1)$$

где C_0 – эффект самородков, C – структурная составляющая, a – радиус влияния, h – расстояние между точками. Параметры модели оценивались непосредственно по экспериментальным данным без привлечения априорной информации.

Перед интерполяцией выполняется оценка пространственной анизотропии по ковариационной матрице координат точек: вычисляются главные направления вариабельности и коэффициент вытянутости, которые передаются в модель кригинга. Учет анизотропии особенно важен для территорий с выраженной пространственной неоднородностью и мозаичной блочной структурой [11].

Процедура перекрестной проверки и верификации

Для сравнения методов применяется 5-кратная перекрестная проверка со случайным перемешиванием. Данные были разделены на обучающую (80 %) и проверочную (20 %) выборки. В качестве метрик точности интерполяции использовались среднеквадратическая ошибка (СКО), вычисляемая по формуле:

$$\text{СКО} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z_{\text{инт}}^{(i)} - Z_{\text{ист}}^{(i)})^2}, \quad (2)$$

и средняя абсолютная ошибка (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Z_{\text{инт}}^{(i)} - Z_{\text{ист}}^{(i)}|. \quad (3)$$

Поскольку универсального метода интерполяции для любых типов поверхностей не существует [7], каждая модификация оценивается как самостоятельный метод без подгонки параметров внутри блоков проверки. Такой подход согласуется с практикой применения перекрестной проверки для выбора лучшего ГИС-интерполятора среди детерминированных альтернатив [4], а также для верификации геостатистических моделей [5].

Для демонстрации практического применения алгоритма выполнен расчет объемов водохранилища. Полный ($V_{\text{полн}}$) и мертвый ($V_{\text{мертв}}$) объемы определялись относительно горизонтальных плоскостей нормального подпорного уровня (НПУ = 80,6 м) и уровня мертвого объема (УМО = 65,6 м) по формуле:

$$V = \sum_{i=1}^n (H - Z_i) \cdot \Delta s, \quad (4)$$

где H – отметка граничной плоскости (НПУ или УМО соответственно), Z_i – расчетная отметка дна в i -м узле регулярной сетки, Δs – площадь элементарной ячейки, а n – количество ячеек, попавших в контур затопления.

Полезный объем определяется как разность полного и мертвого объемов.

Для независимого контроля качества используются пять контрольных точек, соответствующих характерным элементам рельефа: глубоководная зона, донный водовыпуск, буй мониторинга, два береговых створа. Эти точки не участвуют в обучении моделей.

Обсуждение результатов и ограничения метода

Результаты перекрестной проверки приведены в таблице 1. Кубический сплайн показал наилучшие метрики среди всех методов. Из моделей кригинга лучшей оказалась сферическая вариограмма. Гауссова модель продемонстрировала численную неустойчивость – ее ошибки на несколько порядков превысили ошибки остальных методов. Аналогичный эффект отмечается в работе [5], где сферическая модель была признана наиболее устойчивой для интерполяции гравиметрических аномалий.

Таблица 1 – Результаты перекрестной проверки методов интерполяции

Метод	СКО, м	САО, м
Кубический сплайн	0,38	0,19
Сферический кригинг	0,51	0,30
Экспоненциальный кригинг	0,56	0,31
Линейная интерполяция	0,74	0,45
Метод ближайшего соседа	3,22	2,07
Линейный кригинг	3,29	2,49
Гауссов кригинг	*	*

Примечание: гауссова модель продемонстрировала численную неустойчивость (расходимость матрицы весов), СКО превысило $2,9 \cdot 10^3$ м, САО – $2,3 \cdot 10^3$ м.

По выбранному методу построена регулярная сетка из 120×120 узлов с шагом 4,20 м. Перенос данных с нерегулярного массива на регулярную сетку – важнейший этап их подготовки к последующему анализу [6]. Согласно расчетам, относительная ошибка определения суммарного объема чаши водохранилища не

превысила 0,04 %, а баланс полезной и мертвой емкостей сведен с точностью до десятых долей кубического метра.

Таблица 2 – Сравнение расчетных и истинных объемов водохранилища

Параметр	Расчет, тыс. м ³	Истина, тыс. м ³	Ошибка, %
Полный объем	1 943,4	1 943,4	0,001
Полезный объем	808,7	809,0	0,038
Мертвый объем	1 134,7	1 134,4	0,028

Результаты верификации в контрольных точках представлены в таблице 3. Локальные ошибки оказались существенно выше метрик перекрестной проверки: средняя ошибка $\pm 0,73$ м, максимальная $\pm 1,21$ м.

Таблица 3 – Ошибки интерполяции в контрольных точках

Створ	Истинная отметка, м	Расчетная отметка, м	Ошибка, м	Глубина, м
Глубоководный	41,58	42,65	1,07	42,6
Донный водовыпуск	32,38	31,96	0,42	46,5
Буй мониторинга	31,00	30,78	0,22	49,6
Береговой створ №1	89,45	90,18	0,73	18,8
Береговой створ №2	97,93	99,15	1,21	22,1

Сопоставление показателей перекрестной проверки и контрольных створов выявило расхождение погрешностей в 3-4 раза. Средняя составила 0,19 м против 0,73 м, а максимальная – 0,38 м против 1,21 м. Это связано со случайным разбиением геоданных, при котором обучающая и проверочная выборки оказываются близкими, что занижает оценку ошибок. Похожая проблема исследована в работе [4], где рекомендовано учитывать дополнительную топографическую информацию. Кроме того, независимые створы умышленно закладывались на перегибах рельефа с максимальными градиентами глубин.

Контраст между относительной ошибкой объема ($\sim 0,04$ %) и локальными погрешностями (до 1,21 м) объясняется тем, что разнознаковые отклонения ком-

пенсироваться при интегрировании. Такое сглаживание невязок при интегрировании известно в задачах расчета объемов [1, 3]. При этом выбор неверной модели вариограммы может привести к расходимости вычислений. Так, сильный рост погрешностей гауссова кригинга указал на численную неустойчивость весовых матриц. Данная модель требует бесконечно дифференцируемых полей, тогда как реальный ломаный рельеф дна с бровками русла этим свойством не обладает [7]. Этот результат совпадает с выводами работы [5] в пользу сферической модели.

Испытания выявили несколько ограничений алгоритма. Помимо смещенности случайной проверки из-за близости точек, сказывается отсутствие жестких граничных условий. Игнорирование береговой линии и бровок ведет к искусственному сглаживанию изломов дна, поэтому ошибки концентрируются в зонах резких уклонов [4]. Дополнительную погрешность вносит предположение о стационарной анизотропии: единый эллипс поиска не учитывает извилистость русла в пространстве. Для повышения точности нужен переход к локально-анизотропным моделям. Это подтверждается преимуществом байесовского кригинга в работе [5]. Наконец, перенос данных на регулярную сетку требует подбора параметров сглаживания [6], иначе на разреженной периферии возникает неконтролируемая экстраполяция. Перечисленные ограничения будут учтены при дальнейшей доработке вычислительного модуля.

Заключение

В рамках работы выполнено тестирование восьми методов интерполяции. В разработанном алгоритме каждый вариант кригинга оценивался отдельно, без подгонки вариограммы внутри проверочных блоков. При расчете на синтетической модели чаши водохранилища лучшие результаты показал кубический сплайн с СКО 0,38 м. Для сравнения, сферический кригинг выдал СКО 0,51 м. Гауссова модель привела к численной неустойчивости весовых матриц, что совпадает с результатами прошлых исследований [5, 7].

При анализе данных обнаружено расхождение между интегральной оценкой объема и локальными ошибками рельефа. Относительная погрешность общего объема чаши составила всего 0,04 %. Однако в контрольных створах локальные отклонения отметок дна доходят до 1,21 м. Полученные результаты показывают, что одних интегральных метрик недостаточно для оценки качества модели. Так происходит из-за того, что положительные и отрицательные невязки компенсируют друг друга при суммировании по площади. Обычная перекрестная проверка со случайным разбиением также дает искаженную картину. Из-за близкого расположения точек она занижает реальные ошибки моделирования в 3-4 раза по сравнению с независимой верификацией на перегибах рельефа.

Разработанный модуль можно использовать для разных типов пространственных данных. Он подходит для обработки изыскательских материалов в геодезии, маркшейдерии, гидрографии или горном деле. В дальнейшем планируется доработать алгоритм: перейти на пространственную блочную перекрестную проверку, добавить учет тренда для универсального кригинга [4, 5] и автоматизиро-

вать расчет вариограмм. Также нужно внедрить сглаживание [6] и добавить построение карт погрешностей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Дорогова И. Е., Сутурин Г. В. Разработка программного обеспечения для билинейной интерполяции на основе цифровой модели движений земной коры // Интерэкспо Гео-Сибирь. – 2024. – Т. 1. – С. 158–164. – DOI 10.33764/2618-981X-2024-1-158-164.
2. Дорогова И. Е., Демидов К. А. Разработка программной модели движений блоков земной коры для территории Российской Федерации // Регулирование земельно-имущественных отношений в России: сборник материалов VII Национальной научно-практической конференции. – Новосибирск: СГУГиТ, 2024. – Ч. 3. – С. 192–198. – DOI 10.33764/2687-041X-2024-3-192-198.
3. Дорогова И. Е., Духовников К. С. Разработка геодинамического программного модуля для оценивания деформаций земной коры по результатам геодезических измерений // Вестник СГУГиТ. – 2022. – Т. 27. – № 6. – С. 15–27. – DOI 10.33764/2411-1759-2022-27-6-15-27.
4. Мыслыва Т. Н., Куцаева О. А., Подлесный А. А. Сравнение эффективности методов интерполяции на основе ГИС для оценки пространственного распределения гумуса в почве // Вестник Белорусской государственной сельскохозяйственной академии. – 2017. – № 4. – С. 146–152.
5. Kenesbayeva A., Kuldeev E. I., Shalenov E. O., Nurpeissova T. B. Application of the kriging method for gravity data interpolation // Горный журнал Казахстана. – 2025. – № 6. – С. 61–66.
6. Нейман Ю. М., Сугаипова Л. С. О создании регулярной сетки по результатам спутниковой градиентометрии // Известия высших учебных заведений. Геодезия и аэрофотосъемка. – 2017. – № 2. – С. 8–13.
7. Абакушина М. В. Анализ методов интерполяции нерегулярно заданных высот квази-геоида // 70-я научно-техническая конференция студентов, аспирантов и молодых ученых МИИГАиК: сборник тезисов. – М.: МИИГАиК, 2015. – С. 76–81.
8. Свизов Е. М., Ветров А. Л. Численное моделирование сильных летних осадков при различных вариантах шага регулярной сетки // Географический вестник. – 2021. – № 4 (59). – С. 73–83. – DOI 10.17072/2079-7877-2021-4-73-83.
9. Библиотека SciPy для Python // SciPy official site [сайт]. – 2026 – URL: scipy.org (дата обращения: 04.05.2026). – Режим доступа: общ. доступ. – Текст: электронный.
10. Дорогова И. Е., Духовников К. С. Программное обеспечение, сервисы и открытый код для решения задач геодинамики // Интерэкспо Гео-Сибирь. – 2022. – Т. 1. – С. 138–145. – DOI 10.33764/2618-981X-2022-1-138-145.
11. Махмудов М. Д., Фазилова Д. Ш. Построение регулярного поля скоростей Ташкентского региона на основе интерполяции данных ГНСС пунктов // ИнтерКарто. ИнтерГИС. Материалы Междунар. конф. – М.: Географический факультет МГУ, 2023. – Т. 29. – Ч. 1. – С. 535–545. – DOI 10.35595/2414-9179-2023-1-29-535-545.

© Г. В. Сутурин, 2026