

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ
ПРОБЛЕМ БЕЗОПАСНОГО
РАЗВИТИЯ АТОМНОЙ ЭНЕРГЕТИКИ

RUSSIAN ACADEMY OF
SCIENCES
NUCLEAR
SAFETY INSTITUTE

Препринт № NSI-25-94

Preprint NSI-25-94

Р.В. Арутюнян, Л.А. Большов, В.В. Демьянов,
М.Ф. Каневский, И.И. Линге

ПРИМЕР ГЕОСТАТИСТИЧЕСКОГО АНАЛИЗА ЧЕРНОБЫЛЬСКИХ ВЫПАДЕНИЙ

Москва
1994

Moscow
1994

Р.В. Арутюнян, Л.А. Большов, В.В. Демьянов, М.Ф. Каневский, И.И. Линге.
ПРИМЕР ГЕОСТАТИСТИЧЕСКОГО АНАЛИЗА ЧЕРНОБЫЛЬСКИХ ВЫ-
ПАДЕНИЙ. Препринт № NSI-25-94. Москва: Институт проблем безопасного
развития атомной энергетики РАН, 1994. 47 с.

Аннотация

В работе проведен последовательный геостатистический анализ пространственно распределенных данных на примере Чернобыльских выпадений. В представленную методику исследования вошли: ячейковая декластеризация, метод движущегося окна, структурный анализ, кросс-валидация и кригинг. Особое внимание уделено анизотропии и сравнению изотропного и анизотропного моделирования. Результаты пространственной интерполяции для геостатистических подходов сопоставлялись с результатами классического метода обратных квадратов. Все использованные методы подробно описаны, приведены уравнения.

©ИБРАЭ РАН, 1994

R. Arutyunyan, L. Bolshov, V. Demyanov, M. Kanevsky, I. Linge. SAMPLE GEOSTATISTICAL ANALYSIS OF CHERNOBYL FALLOUT (in Russian). Preprint № NSI-25-94. Moscow: Nuclear Safety Institute, 1994. 47 p.

Abstract

One of the approaches to the spatial data analysis is presented in the work. The following methods were sequentially applied to the Chernobyl data: cell declustering, moving window statistics, structural analysis, cross-validation and kriging. Anisotropic case was under special concern, as well as comparison between isotropic and anisotropic modelling. Ordinary kriging results were compared with classical inverse distance squared interpolation. All used methods are fully and clearly described.

©Nuclear Safety Institute, 1994

ПРИМЕР ГЕОСТАТИСТИЧЕСКОГО АНАЛИЗА ЧЕРНОБЫЛЬСКИХ ВЫПАДЕНИЙ

*Р.В. Арутюнян, Л.А. Большов, В.В. Демьянов,
М.Ф. Каневский, И.И. Линге*

ИНСТИТУТ ПРОБЛЕМ БЕЗОПАСНОГО РАЗВИТИЯ АТОМНОЙ ЭНЕРГЕТИКИ

113191 Москва, ул Б.Тулльская, 52

тел.: (095) 952-24-21, факс: (095) 230-20-29, эл. почта: pbl@ibrae.msk.su

Содержание

1 Введение	4
2 Декластеризация	4
3 Метод движущегося окна (Moving Window)	6
4 Структурный анализ	7
5 Построение вариограмм и моделирование	13
6 Кросс-валидация	14
7 Кригинг	17
8 Метод обратных квадратов	20
9 Результаты интерполяции	20
10 Заключение	20
11 Рисунки 1-42	23

1 Введение

Ряд предварительных результатов геостатистического анализа Чернобыльских данных был получен в работах [1, 2, 3]. В данной работе осуществлена попытка провести последовательный анализ пространственно распределенных данных на примере Чернобыльских выпаждений. В рамках данного исследования была сделана глобальная оценка, декластеризация данных, исследована пространственная структура методом движущегося окна, далее был проведен вариограммный анализ с моделированием, и кросс-валидация выбранных моделей. В качестве конечного результата была проведена пространственная интерполяция с привлечением кригинга. Для сравнения использовался широко известный метод обратных квадратов.

Для анализа использовались данные по загрязнению Калужской области цезием 137 (см. рис. 1). В файле насчитывается 227 данных. Для более подробного рассмотрения выбиралась юго-восточная подобласть с наибольшей плотностью измерений. В этой подобласти находится 165 данных.

2 Декластеризация

Большинство пространственно распределенных данных, которые анализируются в геостатистике, имеют кластерную структуру. Кластер образуется, если в какой-либо области проводилось больше измерений чем в другой. В этом случае могут возникнуть существенные отклонения при вычислении, например, среднего значения. Это влечет за собой невозможность получить репрезентативную гистограмму распределения без проведения декластеризации.

Например, в области высоких значений находится в два раза больше точек, чем в области низких значений. Если все значения будут иметь одинаковый вес, то область высоких значений будет иметь слишком большое влияние на распределение. Другими словами, среднее значение будет завышено за счет скопления экспериментальных точек в области высоких абсолютных значений. В этом случае точки из зоны с большими значениями нужно взять с весом в два раза меньшим, чем остальные точки. Этим будет учтена нерегулярность проведенных измерений. Наоборот, при большом скоплении опытных точек в области низких значений, среднее значение, подсчитанное при взятии всех точек с одинаковыми весами, будет занижено по сравнению с действительным. Проблему решает проведение измерений на регулярной сетке, в этом случае наилучшее описание распределения получится при работе с равными весами. Во многих случаях невозможно или нежелательно получить данные на равномерной сетке.

При рассмотрении измерений, проведенных на *нерегулярной* сетке, предполагается существование такого набора весов, при котором может быть получено репрезентативное распределение данных. Из эмпирических соображений можно утверждать, что данные из области с большей плотностью измерений нужно брать с меньшим весом (для уменьшения их влияния на распределение в целом), чем данные из области с меньшей плотностью измерений.

Описанный ниже метод дает представительное распределение всех рассматриваемых данных. Для вычисления весов могут быть использованы разные подходы: метод ячейковой декластеризации, метод ячеек Дирихле, кригинг, для локальной декластеризации следует использовать индикаторный кригинг.

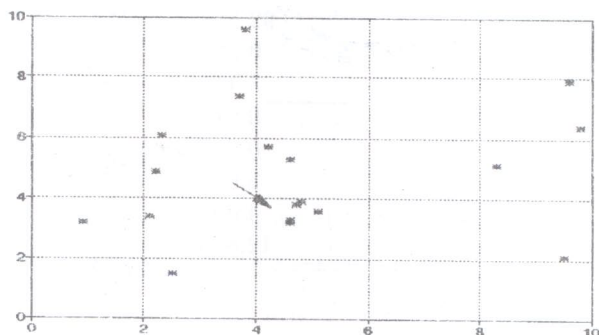
Метод ячейковой декластеризации.

Этот метод (cell-declustering) предложил А.Г.Жорнель (A.G.Journel) в 1983 году [13]. Идея метода заключается в разбиении рассматриваемой области на подобласти кластеризованных данных и определении равных весов для всех точек внутри каждой подобласти в соответствии с количеством точек.

Так, если в ячейку a_i попало n_i точек, то каждое данное будет взято с весом $1/n_i$. Область a_i обычно имеет размерность 3 (время может стать четвертым измерением). Далее, в каждой подобласти a_i данным приписываются равные веса. Для ячейки, не содержащей опытных точек, веса не рассчитываются. Это означает, что область декластеризованных данных состоит

из ячеек, содержащих по крайней мере по одному данному. Это позволяет ограничить влияние граничных данных весом 1. Это может оказаться и недостатком, если границы области четко определены, потому что область декластеризованных значений может не иметь четких границ. Этот недостаток может быть компенсирован тем, что если область четко определена, то в ней должны находиться опытные точки.

На рисунке 1 показано примерное разбиение области на ячейки. При декластеризации значениям попавшим в помеченную стрелкой ячейку будут приписаны веса $1/5$. А точки в ячейке выше будут взяты с весом $1/2$.



Для вычисления весов декластеризации нужно знать два параметра:

1. Размер ячейки (в каждом направлении).
2. Начальная точка сетки (левый нижний угол).

Возможны два предельных случая. Если размер ячейки слишком маленький, то каждая ячейка будет содержать не более одной точки, что приведет к присвоению всем точкам равных весов. В противоположном случае, когда размер ячейки слишком большой, все данные попадут в одну ячейку, что также приведет к получению одинаковых весов для всех данных.

Метод выбора размера ячейки зависит от типа кластеризации. Если данные кластеризованы случайным образом (есть области скопления данных), то размер ячейки выбирается так, чтобы в областях низкой плотности точек приходилось приблизительно по одному данному на ячейку. Если же известно, что есть области высоких или низких значений с большим числом точек, то размер может быть выбран для получения максимального или минимального взвешенного среднего. При декластеризации областей высоких или низких значений нужно брать набор ячеек разного размера. В этом случае строится график зависимости взвешенного среднего значения от размера ячейки и в соответствии с ним выбирается подходящий размер.

Ячейка не обязательно должна быть квадратной. С помощью параметра анизотропии (отношение размеров ячейки) можно построить аналогичные зависимости и на их основе выбрать параметры ячейки соответствующие минимуму или максимуму. Результаты можно также представить в виде контурной карты с размерами ячеек в каждом из направлений в качестве координат. Если при фиксированном размере ячейке перемещать начало сетки, то веса декластеризации могут существенно меняться. Эта проблема решается путем систематического смещения начала сетки. Веса, полученные после каждого смещения, усредняются. Обычно хватает пяти смещений [8].

Рассмотрим файл данных по загрязнению Калужской области цезием 137 (см. рис. 1). Измерения попавшие в этот файл производились пространственно нерегулярно, что привело к образованию кластеров. Всего используется 227 точек. При подсчете кластеризованное среднее (2.7875) оказывается в два раза больше декластеризованного (1.3551). Гистограммы взвешенных и невзвешенных значений представлена рисунке 3. При построении карты распределения весов видно, что наименьшие веса присвоены точкам сконцентрированным в области больших абсолютных значений на юге (см. рис. 2).

Статистика для кластеризованных и декластеризованных данных

Статистика	Кластеризованные данные	Декластеризованные данные
Число данных	227	227
Число опущенных данных	0	0
Число данных меньше нуля	0	0
Среднее значение	2.788	1.3551
Вариация	6.661	3.8109
Стандартное отклонение	2.581	1.9521
Коэффициент вариации %	92.588	144.0539
Коэффициент симметрии	0.965	2.0836
Экспесс	3.431	7.2068
Минимум	0.019	0.019
Первая квартиль	0.545	0.1818
Медиана	2.000	0.5
Вторая квартиль	4.650	1.5052
Максимум	12.810	12.810

На гистограмме весов (рис. 4) построены веса для изотропной декластеризации и анизотропной декластеризации при пяти и десяти смещениях координатной сетки. Из гистограммы видно, что веса накапливаются в области больших значений (вес меньше единицы). При этом веса полученные на анизотропной сетке имеют менее сглаженное распределение. На рисунках 5 и 6 представлены зависимости весов от измеренных значений, которым они приписаны. Заметим, что при декластеризации на анизотропной сетке в области малых значений распределение весов имеет более равномерный характер. Зависимость декластеризованного среднего от размера ячейки имеет минимум при размере ячейки 96 километров. Зависимости построенные с учетом и без учета анизотропии имеют сходный характер, но анизотропная декластеризация более стабильна, то есть кривая испытывает меньше биений (см. рис. 7). Увеличение числа смещений координатной сетки ведет к сглаживанию кривой зависимости от размера ячейки и к некоторому увеличению декластеризованного среднего (см. рис 8).

3 Метод движущегося окна (Moving Window)

При подходе к анализу пространственно распределенных данных важно иметь представление не только о глобальной статистике (однопараметрическая статистика по всей области) но и о распределении статистики в частях области. Для этого используется методика движущегося окна [10].

Статистика движущегося окна (Moving Window Statistics) — подсчет статистики в подобластях (окнах). Этот метод очень полезен при наличии зон различной вариации значений (*heteroscedasticity*) или зон аномальных средних значений. При этом область разбивается на несколько одинаковых обычно прямоугольных окрестностей (окон). Размер окна зависит от среднего расстояния между точками. Хорошим компромиссом между большими и маленькими окнами являются перекрывающиеся окна, при этом два соседних окна имеют несколько общих точек. Это повышает количество окон при достаточно большом размере, дающем достоверную статистику. Таким образом, при смещении окна осуществляется движение. Можно построить карту средних значений и стандартных отклонений в окнах и выделить области локального изменения вариабельности.

Этот же метод позволяет выявить важную связь между пространственными распределениями среднего значения и вариабельности. *Эффект пропорциональности* (*Proportional Effect*) — зависимость между локальными средними значениями и локальными стандартными отклонениями. Можно выделить четыре самых общих случая зависимости:

1. и среднее, и вариабельность постоянны

2. среднее значение имеет локальный тренд, в то время как вариабельность остается постоянной
3. среднее постоянно при изменении вариабельности
4. и среднее, и вариабельность изменяются вместе пропорционально

Для вычисления статистики было выбрано окно размером 20 на 10 км. На рисунке 9 показана диаграмма зависимости между средним значением и стандартным отклонением — это случай 2. Эффект пропорциональности наблюдается только на участке малых средних значений, а далее стандартное отклонение выходит на постоянный уровень (рис. 10). Нулевое значение стандартного отклонения означает, что в окне оказалась только одна точка. На рисунках 11, 12 показаны пространственные распределения локального среднего значения и вариации.

4 Структурный анализ

В рамках структурного анализа проводится построение и моделирование вариограмм пространственно распределенных данных. Вариограмма дает представление о пространственной корреляции данных. В дальнейшем модель может быть использована в кригинге.

Поскольку пространственный анализ имеет дело с взаимным расположением данных, имеет смысл заранее разбить данные на пары — подготовить файл пар. Работа с файлом пар намного ускоряет вычисление вариограмм. Можно оценить количество пар в файле:

$$N_{pair} = \left(\frac{N \cdot (N - 1)}{2} \right)$$

где N — количество данных.

Так, для рассматриваемой части Калужской области включающей 165 точек измерений количество пар составило 13,430.

Напомним некоторые величины, используемые в структурном анализе.

- **Ковариация (Covariance)** — статистическая мера корреляции между двумя значениями.

$$C(h) = E\{(V(x_1) - m(x_1))(V(x_2) - m(x_2))\} \quad (1)$$

Для N экспериментальных точек:

$$C(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} (V(x_i) \cdot V(x_i + h) - m_{-h} \cdot m_{+h}) \quad (2)$$

где:

m_{-h} среднее значение данных, находящихся на расстоянии $-h$ друг от друга:

$$m_{-h} = \frac{1}{N(h)} \sum_{i=1}^{N(h)} V(x_i) \quad (3)$$

m_{+h} среднее значение данных, находящихся на расстоянии $+h$ друг от друга:

$$m_{+h} = \frac{1}{N(h)} \sum_{i=1}^{N(h)} V(x_i + h) \quad (4)$$

- **Момент инерции (Moment of Inertia)** — средний квадрат разницы между x и y координатами точек в каждой паре. Это другая мера размытости облака.

$$moment\ of\ inertia = \frac{1}{2n} \sum_{i=1}^n (x_i - y_i)^2 \quad (5)$$

- **Полувариограмма** (или просто **вариограмма**) (Semivariogram) — вариация разницы значений переменной в двух точках как функция от расстояния между ними и направления:

$$\gamma(x_1, x_2) = \frac{1}{2} \text{Var}\{V(x_1) - V(x_2)\} = E\{(V(x_1) - V(x_2))^2\} \quad (6)$$

Для N экспериментальных точек:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (V(x_i) - V(x_i + h))^2 \quad (7)$$

Можно отметить свойство симметрии вариограммы:

$$\gamma(h) = \gamma(-h) \quad (8)$$

В общем нестационарном случае моменты не существуют. Но при введении некоторых предположений закон распределения случайной величины может быть описан моментами первого и второго порядка. Для существования вариограммы необходимо принять *внутреннюю гипотезу (intrinsic hypothesis)*.

Внутренняя гипотеза (Intrinsic hypothesis).

Случайная функция $V(x)$ называется *внутренней (intrinsic)* если:

- математическое ожидание $m(x)$ существует и не зависит от местоположения x ;
- для любого вектора h разность $(V(x_1) - V(x_2))$ имеет конечную вариацию, независимую от x (стационарность вариограммы):

$$\frac{1}{2} \text{Var}\{V(x+h) - V(x)\} = E\{(V(x+h) - V(x))^2\} = \gamma(h) \quad \forall x \quad (9)$$

Для существования других моментов второго порядка (ковариации, мадограммы и т.д.) необходимо дополнительно ввести *стационарность второго порядка*.

Стационарность второго порядка (Second order stationarity).

Случайная функция $V(x)$ обладает *стационарностью второго порядка*, если:

- математическое ожидание $m(x)$ существует и не зависит от местоположения x :

$$m(x) = E\{V(x)\} \quad \forall x \quad (10)$$

- для каждой пары значений случайной переменной $\{V(x), V(x+h)\}$ ковариация существует и зависит только от расстояния между ними h :

$$C(h) = E\{V(x)V(x+h)\} - m^2 \quad \forall x \quad (11)$$

- **Относительная вариограмма (Relative Variogram)** — вариограмма в которой значение обычной вариограммы для каждого h делится на квадрат среднего значения величин находящихся на расстоянии h .

$$\text{Ral} = \frac{\gamma(h)}{m_h^2} \quad (12)$$

Это бывает полезно, когда присутствует *эффект пропорциональности (proportional effect)* т.е. когда области с большей концентрацией чем средняя также имеют ковариацию большую чем средняя. При использовании моделей относительных вариограмм в кригинге стандартные отклонения кригинга представляют собой десятые доли значений оценок.

- **Обратная ковариация** (Inverted Covariance Variogram) — вариограмма, получаемая при вычитании ковариаций *лагов* (lag) (для каждого заданного набора расстояний h между парами) из экспериментальных вариаций. Такой подход учитывает случаи, когда среднее головных точек отличается от среднего хвостовых точек.

$$S.I.C.(h) = \frac{\sigma^2 - C(h)}{\sigma^2} \quad (13)$$

где *вариация* данных:

$$\sigma^2 = Var\{V(x)\} = E\{(V(x) - m(x))^2\} = E\{V(x)^2\} - m(x)^2 \quad (14)$$

Обратная ковариация может использоваться в кригинге тем же образом, что и обычная полувариограмма.

- **Мадограмма** (Madogram) — среднее значение разницы между измерениями в парах как функция от расстояния между ними и направления:

$$M(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} |u(x_i) - u(x_i + h)| \quad (15)$$

Мадограмма не является действительной вариограммой и обычно не должна быть использована в кригинге.

Вариограмма должна удовлетворять ряду строгих теоретических результатов: положительная определенность, поведение на бесконечности и в нуле. В практическом использовании теоретические моделей и их комбинации должны удовлетворять этим требованиям для описания экспериментальных вариограмм.

Модели вариограмм

Основные типы вариограмм называются соответственно функциям, которые их моделируют.

Модели бывают следующих типов: **Самородок** (Nugget)

$$\gamma(h) = c \quad (16)$$

Наличие у данных такой вариограммы означает отсутствие пространственной корреляции. Данные в этом случае распределены абсолютно случайно (pure nugget).

Сферическая (Spherical)

$$\gamma(h) = \begin{cases} c \cdot \left(1.5 \frac{h}{a} - 0.5 \left(\frac{h}{a}\right)^3\right) & \text{если } h \leq a \\ c & \text{если } h > a \end{cases} \quad (17)$$

где a — действительный радиус корреляции (*range*).

Гауссова (Gaussian)

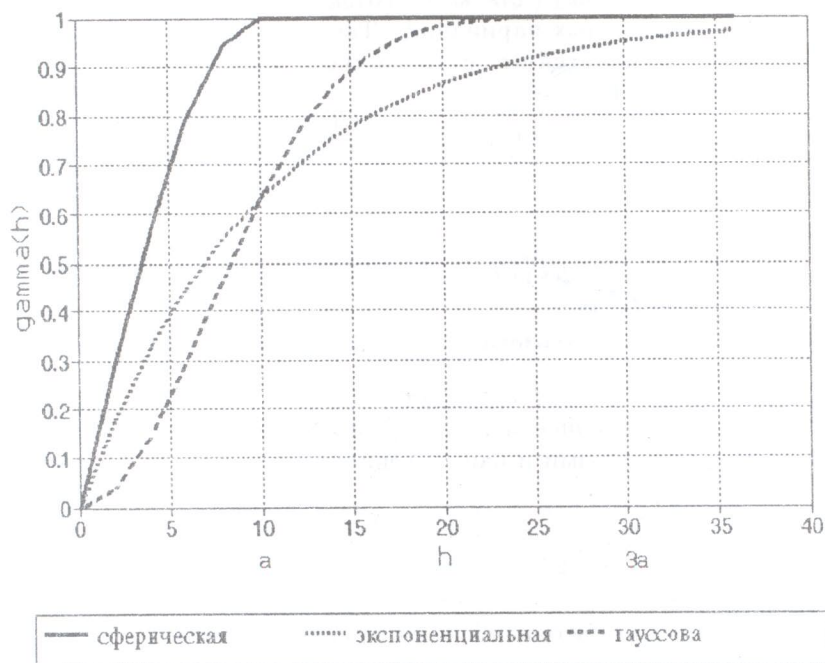
$$\gamma(h) = c \cdot \left(1 - \exp\left(-\frac{3h^2}{a^2}\right)\right) \quad (18)$$

где a — эффективный радиус корреляции (*range*).

Экспоненциальная (Exponential)

$$\gamma(h) = c \cdot \left(1 - \exp\left(-\frac{3h}{a}\right)\right) \quad (19)$$

где a — эффективный радиус корреляции (*range*).



Степенная (Power)

$$\gamma(h) = c \cdot h^\lambda \quad (20)$$

где $0 < \lambda < 2$.

И как ее частный случай — Линейная (Linear)

$$\gamma(h) = c \cdot h \quad (21)$$

Во всех моделях c — сил *sill* (или *slope* в степенной модели) — может быть только положительным.

Типы моделей для $c = 1$, $a = 10$ представлены на рисунке выше.

Во многих случаях используются комбинации моделей различных типов. Их силы и радиусы могут быть различными. Так, наличие суммы моделей с разными радиусами корреляции, говорит о присутствии гнездовой структуры (nested structure) данных.

Поведение вариограмм на больших расстояниях

В соответствии с приведенными выше типами вариограмм можно классифицировать их по поведению на большом расстоянии. При существовании стационарности второго порядка, значение вариограммы “в бесконечности” равно значению ковариации в нуле (исходной ковариации). Это характерно для вариограмм сферического, экспоненциального и гуссового типов. Модель вариограммы сферического типа достигает сила на расстоянии a ; экспоненциального типа — на расстоянии $3a$; а гуссова модель достигает 95% значения сила на расстоянии $a\sqrt{3}$.

Если вариограмма не имеет предела роста на бесконечности, это означает, что ковариация не существует. В этом случае стационарность второго порядка заменяется более слабой внутренней (intrinsic) гипотезой. Бесконечной вариации данных соответствует степенная и линейная модели вариограмм. Это означает, что данные, далеко удаленные друг от друга, продолжают оказывать взаимное влияние.

Поведение вариограмм в нуле

Вариограммы различаются по характеру поведения в нуле. Теоретически $\gamma(0) = 0$ независимо от типа вариограммы. Однако, очень часто вариограмма имеет скачок в нуле, что и называется *наггет (nugget) эффектом* (эффектом самородка, как это было придумано в горном деле). Такой разрыв вариограммы вблизи нуля моделируется включением соответствующей (наггет) составляющей. Эффект происходит в результате ошибок измерений и микровариабельностей. Поскольку структура этих микровариабельностей имеет масштаб меньший, чем масштаб полученных данных, они (микровариабельности) проявляются как белый шум.

Как было уже упомянуто выше, постоянная горизонтальная вариограмма представляет *чистый наггет эффект (pure nugget effect)*. При этом $\gamma(0) = 0$ в некоторой окрестности нуля ϵ , а при $h > \epsilon$ — $\gamma(h) = C_0$. Чистый наггет эффект соответствует полному отсутствию автокорреляций.

При *параболическом* поведении вблизи нуля $\gamma(h) \sim A |h|^2$. В этом случае $\gamma(h)$ дважды дифференцируема в нуле. Такое поведение характеризует высоко регулярную структуру, которая соответствует гауссовой модели вариограммы.

При *линейном* поведении вблизи нуля $\gamma(h) \sim A |h|$. В этом случае $\gamma(h)$ не дифференцируема в нуле, но остается непрерывной при $h = 0$. Этот случай представлен линейной моделью.

Анизотропия вариограмм

Выше рассматривались только *изотропные* вариограммы — те, которые зависят только от расстояния между точками (в евклидовом пространстве). При изотропии изолинии вариограммы будут иметь форму круга. Если имеет место зависимость и от ориентации пары точек в пространстве, то можно говорить о наличии *анизотропии*. Это означает существование различных структур данных в разных направлениях.

Анизотропия вариограмм бывает следующих типов [12]:

- *геометрическая (geometric)* — которую можно представить как изотропную вариограмму, полученную в результате линейной трансформации:

$$\underbrace{\gamma(h_u, h_v, h_w)}_{\text{анизотропная}} = \underbrace{\gamma'(\sqrt{h_u'^2 + h_v'^2 + h_w'^2})}_{\text{изотропная}} \quad (22)$$

$$h_u' = a_{11}h_u + a_{12}h_v + a_{13}h_w$$

$$h_v' = a_{21}h_u + a_{22}h_v + a_{23}h_w$$

$$h_w' = a_{31}h_u + a_{32}h_v + a_{33}h_w$$

Это означает, что изолинии вариограммы будут иметь форму эллипса. В случае комбинации из нескольких структур вариограмма может быть записана следующим образом:

$$\gamma(h) = w_0\gamma_0(h) + w_1\gamma_1(h_1) + w_2\gamma_2(h_2) \quad (23)$$

где: w_0, w_1, w_2 — коэффициенты при вариограммах по направлениям $\gamma_1, \gamma_2, \gamma_3$.

Все вариограммы в этом случае должны иметь одинаковый вид.

Такая анизотропия носит название *анизотропия радиуса (range anisotropy)*.

- *зональная (zonal)* — под ней понимаются все типы негеометрической анизотропии [16]

Негеометрическая анизотропия радиуса (range anisotropy)

Анизотропия радиуса не всегда является геометрической анизотропией. Если радиусы корреляции в вариограммах по направлениям распределены таким образом, что изолинии не имеют форму эллипса, то такая анизотропия радиуса будет *негеометрической*. В таком случае, например для экспоненциальных моделей, вариограммы по четырем направлениям ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) можно записать так:

$$\gamma(h) = w_1(1 - \exp(-a_1h))$$

$$\gamma(h) = w_2(1 - \exp(-a_2h))$$

$$\gamma(h) = w_3(1 - \exp(-a_3h))$$

$$\gamma(h) = w_4(1 - \exp(-a_4h))$$

Нужно отметить, что при стационарности второго порядка w_1, w_2, w_3, w_4 должны быть равны между собой.

Анизотропия сила (sill anisotropy)

Если сила существует, то данные обладают стационарностью второго порядка. При силе, зависящем от направления, вариограмма представляется гнездовой структурой, как и в других случаях зонной анизотропии.

$$\gamma_Z(h) = \sum_{i=1}^n \gamma_i(A_{ih}) \quad (24)$$

где γ_i — изотропные вариограммы, а $A_1, \dots, A_n$ — матрицы, определяющие направления.

Переменный сил свидетельствует о нестационарности, ненулевых пространственных корреляциях, или о коррелированных ошибках измерений с неравными средними значениями. На практике трудно определить какая из трех причин является преобладающей. Если вариограммы в разных направлениях имеют различные силы, то было бы неосмотрительно погоружаться в моделирование вариограмм без предварительного выявления и удаления тренда.

Анизотропия наггета (nugget anisotropy)

При такой анизотропии величина наггета меняется с направлением. Причиной анизотропии сила и наггета являются коррелированные ошибки измерений. Пусть есть пространственно распределенная функция Z . Предположим, что вариограммы в направлениях $45^\circ, 90^\circ$ и 135° имеют близкие радиусы силы и наггеты, а вариограмма в направлении 0° имеет намного меньшие сил и наггет и приблизительно такой же радиус. Предскажем значение функции в точке s_0 . Тогда уравнение для кригинга будет следующим:

$$\Gamma_{ij}\lambda = \gamma$$

где: $\lambda = (\lambda_1, \dots, \lambda_n, \mu)$, μ — множитель Лагранжа.

$$\gamma = (\gamma_Z(s_1 - s_0), \dots, \gamma_Z(s_n - s_0), 1)$$

$$\Gamma_{ij} = \begin{cases} \gamma_Z(s_i - s_j), & \text{если } i = 1, \dots, n; j = 1, \dots, n \\ 1, & \text{если } i = n+1; j = 1, \dots, n \\ 1, & \text{если } i = 1, \dots, n; j = n+1 \\ 0, & \text{если } i = n+1; j = n+1 \end{cases} \quad (25)$$

Очень полезным инструментом в при проведении структурного анализа является *вариограммная поверхность (variogram surface)*. Это поверхность, на которой нанесены абсолютные значения вариограммы в зависимости от разницы координат δx и δy . Таким образом можно сразу получить представление о поле вариограмм в пространстве. Вариограммная поверхность может быть представлена в виде карты изолиний. При этом типы анизотропии просматриваются очень наглядно — например эллипс в случае геометрической анизотропии. Следует отметить, что вариограммное облако обладает точечной симметрией относительно точки $(0,0)$. Это естественно, поскольку каждая пара может браться в двух противоположных направлениях.

Другим используемым инструментом является *вариограммное облако (variogram cloud)*. Это диаграмма разброса значений вариограмм для пар в зависимости от расстояния между точками в паре. Такая диаграмма помогает распознать пары с высоким значением вариограммы поскольку они вносят существенный вклад в значение экспериментальной вариограммы. Вариограммное облако также помогает определить оптимальный лэг (lag — сдвиг расстояния, с которым идет подсчет вариограммы) для вычисления вариограммы. Вариограммное облако может быть построено для любого направления и раствора угла (tolerance angle).

При непосредственном моделировании встает проблема оценки близости модели к экспериментальной вариограмме. Одним из доступных подходов является визуальная похожесть. В этом случае многое зависит от эксперта, проводящего моделирование. Хорошим подспорьем может оказаться специальный *индикатор подгонки* (*IFG — Indicative Goodness of Fit*):

$$IFG = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^{n(k)} \frac{\frac{P(i)}{\sum_{i=1}^{n(k)} P(i)}}{\frac{h(i)}{h_{max}(k)}} \cdot \left(\frac{\gamma(i) - \gamma^*(i)}{\sigma^2} \right)^2 \quad (26)$$

где: N — число значений вариограмм по направлениям, использованных для модели,
 $n(k)$ — номер лага для k -го значения вариограммы,
 $P(i)$ — число пар в i -ом лаге,
 $h(i)$ — среднее расстояния i -го лага,
 $h_{max}(k)$ — максимальное расстояние для i -го значения вариограммы,
 $\gamma(i)$ — значение экспериментальной вариограммы для i -го лага,
 $\gamma^*(i)$ — модельное значение вариограммы для i -го лага,
 σ^2 — вариация данных.

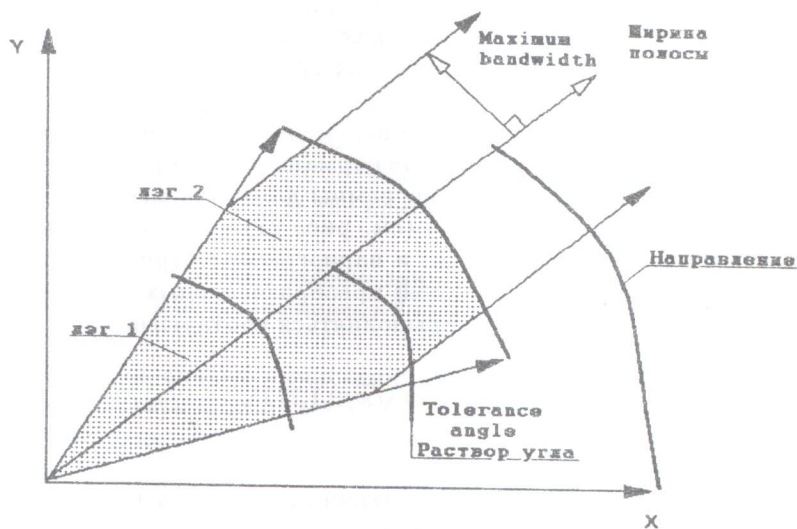
Близкое к нулю показательное согласие указывает на хороший подбор модели. Этот параметр позволяет иметь количественную оценку близости модели к экспериментальной вариограмме.

5 Построение вариограмм и моделирование

Был проведен структурный анализ для выбранной юго-восточной части Калужской области. Работа проводилась с помощью пакета прикладных программ VARIOWIN. Вариограммный анализ проводился в абсолютных и логарифмических переменных.

Прежде всего был построен файл пар (PCF — pair comparison file) для быстрого расчета вариограмм.

На рисунке ниже представлена методика поиска пар в лагах для выбранного направления.



Вариограммные поверхности представлены на рисунке 13. Была замечена разница в структуре в направлениях 90° и 135° . Сразу просматривается анизотропия сила в указанных направлениях. Поверхность для логарифмов имеет более четко выраженную зонную анизотропию.

На рисунках 14, 15, 16 представлены вариограммные облака по всем направлениям и в направлениях 0° и 90° . На них выделены пары, дающие наибольший вклад в экспериментальную вариограмму (находятся внутри эллипса). Местоположение этих пар можно увидеть на карте точек, где пары соединены стрелками.

Из проведенного предварительного анализа можно сделать вывод об имеющейся ограниченной возможности промоделировать вариограмму по всем направлениям, из-за значительного тренда на расстоянии 20–40 км., и о наличии двух совершенно различных структур в условных направлениях “вертикальном” и “горизонтальном”. Указанный тренд проявляется в основном за счет пар, ориентированных в направлении 45° и 90° (см. рис. 17).

На следующей стадии приступим к построению вариограмм для Cs^{137} и $\ln Cs^{137}$.

Вариограмма по всем направлениям для Cs^{137} представлена на рисунке 18. Она имеет тренд на расстоянии 20–30 км., но может быть неплохо промоделирована сферической моделью, поскольку на больших расстояниях (50 км.) она выходит на сил, равный начальной ковариации ($C(0) = 6.4$). При таком подходе мы получаем изотропную модель.

Рассмотрим вариограммы по направлениям. Вариограммы для Cs^{137} в направлении 0° и 135° показывают наличие структуры в этих направлениях с различным радиусом корреляции: 25 км. и 35 км. соответственно (см. рис. 19). Если обратиться к диаграммам разброса лагов (*h-scatter plot*), то можно видеть, какие пары дают одного типа структуру (рис. 20, 21), а какие другого (рис. 22). Эти пары выделены пустыми кружками на диаграммах и соединены стрелками на картах точек. Вариограммы в направлении 0° и 135° очень хорошо моделируются сферической моделью с учетом анизотропии радиуса. Вариограммы в направлении 45° и 90° имеют совершенно отличную структуру (рис. 17). Их сил значительно превосходит значение начальной ковариации, что говорит о большом тренде и наличии корреляций с другой областью данных, к северу от рассматриваемой. Такое поведение можно квалифицировать как негеометрическую анизотропию сила.

В рамках использованного пакета программ удалось промоделировать анизотропию радиуса. Модели вариограмм, параметры которых сведены в таблицу ниже, показаны на рисунках 18, 19.

Аналогично были построены вариограммы для $\ln Cs^{137}$. Вариограмма по всем направлениям, дающая изотропную модель, имеет вид идентичный с соответствующей вариограммой для абсолютных значений (рис. 23). Вариограммы по направлениям отражают те же особенности структуры, что и для абсолютных значений, но имеют более гладкий характер (рис. 24). Из диаграммы разброса лага (*h-scatter plot*) для вариограммы в направлении 135° видно, что пары, нарушающие структуру (пустые кружки), находятся по краям области (рис. 25). Из аналогичной диаграммы видно, что “вертикальная” структуру нарушают пары, находящиеся на западном краю области (рис. 26). На рисунках 23, 24 представлены полученные изотропная и анизотропная модели.

Был проведен вариограммный анализ для переменной, полученной в результате нелинейного преобразования данных в нормированное нормальное распределение $N(0, 1)$. Такое преобразование называется *Normal Score*. Вариограммы для преобразованной переменной имеют гладкий характер и проще моделируются (см. рис. 27). Начальная ковариация преобразованной переменной всегда равна единице, и вариограммы стремятся к этому значению на большом расстоянии. Особенность значений преобразованных переменных состоит в том, что при бинормальности для них должно выполняться соотношение:

$$\frac{\sqrt{\gamma(h)}}{M(h)} = \sqrt{\pi} \approx 1.7725$$

где $M(h)$ — мадограмма.

Это соотношения было рассчитано для различных направлений и результаты в зависимости от номера лага представлены на рисунке 28.

Параметры всех моделей вариограмм сведены в таблицу.

6 Кросс-валидация

Во многих случаях необходимо проверить результаты интерполяции с помощью различных методов и выбрать наилучший. Такой анализ методов может быть проведен путем срав-

Модели вариограмм

Переменная	Направление	Тип модели	Наггет	Сил	Радиус	Аниз.
Cs^{137}	по всем направлениям	Сферическая	0	6.595	23.52	1.0
Cs^{137}	0°	Сферическая	0	6.499	42.47	0.63
$\ln Cs^{137}$	по всем направлениям	Сферическая	0	1.36	27.73	1.0
$\ln Cs^{137}$	0°	Сферическая	0.14	1.199	80.07	0.36
Normal Score	0°	Сферическая	0	0.64	50.74	0.41
	51°	Сферическая	0	0.34	23.37	7.3

нения оцененных значений с действительными. Кросс-валидация позволяет проводить такое сравнение используя только информацию, полученную из измерений. Результаты кросс-валидации обычно используются для сравнения распределений ошибок оценки или разницы (residuals) для различных методов оценки. Это не всегда дает четкий ответ какая оценка лучше. Кросс-валидационные разницы несут в себе важную пространственную информацию. Кросс-валидация является полезной ступенью перед проведением конечных оценок.

Кросс-валидация не является методом проверки гипотез [7]. Метод не может подтвердить правильность какой-то отдельно взятой модели. Поскольку кросс-валидация не зависит от подхода при моделировании (детерминистического или статистического), можно выбрать наиболее подходящие параметры модели. Следует заметить, что если модель хорошо работает с каким-либо набором данных, то это, вообще говоря, не означает, что модель всегда будет работать лучше.

При кросс-валидации вычисляются значения переменной в точках, в которых имеются измерения. Измеренное значение в точке оценивания временно удаляется из набора данных и оценка в точке производится по оставшимся данным. Полученную оценку можно сравнить с изъятым действительным значением. Их разность дает разницу (residual — остаток). Такая процедура повторяется для каждой точки измерений. Окрестность поиска соседних точек выбирается так же, как и в соответствующем методе интерполяции.

Проведем кросс-валидацию для следующих методов пространственной интерполяции: Обычный кригинг (Ordinary kriging) с изотропной и анизотропной вариограммной моделью данных, метод обратных квадратов. В выбранной юго-восточной части Калужской области находится 165 точек измерений, в которых и проведем оценивание. При выборе окрестности поиска для кригинга с изотропной моделью возьмем окружность с радиусом 80 км., а для кригинга с анизотропной моделью выберем эллипс ориентированный в направлении 135° с большой полуосью (80 км.) в два раза превосходящей малую. Для метода обратных квадратов возьмем такую же окрестность, как и для кригинга с изотропной моделью — окружность с радиусом 80 км.

На рисунках 29, 30, 31 построены графики зависимости оцененных значений от действительных для всех трех подходов. Коэффициент корреляции наилучший у кригинга с анизотропной моделью (рис. 31).

Проведем статистический анализ вычисленных разниц. Результаты анализа сведены в таблицу. Видно, что результаты кригинга заметно лучше результатов метода обратных квадратов. Особенно вид распределения характеризуют значения медианы первой и второй квартили. Для распределения оценок кригинга они заметно ближе к значениям для распределения действительных значений. Кригинг с анизотропной моделью не дает значений оценки, меньших нуля, в отличие от кригинга с изотропной моделью.

Полезным инструментом для анализа результатов кросс-валидации является диаграмма зависимости разницы от оцененных значений. По ней можно судить о наличии трендов разницы. Так, для метода обратных квадратов наблюдается смещение разницы в область отрицательных значений с увеличением оценки (рис. 32). Это называется условной смещенностью оценки, в отличие от глобальной смещенности среднего значения разницы. При кригинге этого не наблюдается и можно говорить как о глобальной так и об условной несмещенности

Параметры распределения значений при кросс-валидации.

Статистика	Измеренные значения	Обычный кригинг		Метод обратных квадратов
		изотропная модель	анизотропная модель	
Число данных	165	165	165	165
Число опущенных данных	0	0	0	0
Число данных меньше нуля	0	1	0	0
Среднее значение	3.619	3.639	3.640	3.759
Вариация	6.538	5.393	5.657	2.932
Стандартное отклонение	2.557	2.322	2.378	1.712
Коэффициент вариации	70.644	63.812	65.348	45.549
Коэффициент симметрии	0.634	0.350	0.401	0.164
Экспесс	3.170	2.339	2.421	2.258
Минимум	0.019	-0.028	0.052	0.581
Первая квартиль	1.392	1.573	1.554	2.439
Медиана	3.350	3.765	3.614	3.778
Вторая квартиль	5.202	5.173	5.088	4.832
Максимум	12.810	9.678	9.889	7.807

оценки (рис. 33, 34).

В таблице представлены параметры распределений разниц между оценками и действительными значениями (residuals) для трех методов. Распределения разниц для кригинга отличается большей симметричностью и глобальной несмещенностью. Для кригинга с анизотропной моделью характерна несколько меньшая вариация и более близкое к нулю значение медианы. Графически результаты представлены в виде гистограммы (рис. 35) распределения разниц (residuals) для кросс-валидации метода обратных квадратов, кригинга с изотропной и анизотропной моделями. Видно, что для анизотропной модели вариация меньше, и распределение имеет наиболее симметричный характер.

Пространственное распределение разниц (residuals) представлено на рисунках 36, 37, 38. Для оценки методом обратных квадратов большие положительные разницы (переоценка) расположены в точках по краям области, а отрицательные разницы (недооценка) — во внутренней части области.

Для анализа результатов кросс-валидации можно использовать еще одно свойство — отсутствие корреляции разниц между оцененными и действительными значениями. Для этого нужно построить вариограммы разниц. Если они имеют горизонтальный характер — типа наггет, то разницы не имеют пространственной корреляции. Это значит, что они распределены абсолютно случайно (белый шум). На рисунке 39 представлены вариограммы разниц кросс-валидации для кригинга (верхняя) и для метода обратных квадратов (нижняя). Видно, что вариограмма разницы в случае кригинга точно совпадает с наггет моделью, тогда как в случае метода обратных квадратов разницы имеют структуру, четко выраженную вариограммой по всем направлениям.

В заключении следует отметить, что было бы неправильно автоматически использовать разницы, полученные при кросс-валидации, для улучшения модели. Иногда улучшение модели по результатам кросс-валидации может привести к ухудшению оценки. Если модель требует дополнительной коррекции, то это следует делать с привлечением качественной информации о структуре данных. Более подробно процесс подбора моделей (model fitting) был рассмотрен выше.

Параметры распределения разницы между оценкой и действительными значениями.

Статистика	Обычный кригинг		Метод обратных квадратов
	изотропная модель	анизотропная модель	
Число данных	165	165	165
Число опущенных данных	0	0	0
Число данных меньше нуля	87	80	62
Среднее значение	0.02	0.02	0.140
Вариация	1.344	1.327	1.731
Стандартное отклонение	1.159	1.152	1.316
Коэффициент вариации	5840.437	5725.405	941.756
Коэффициент симметрии	0.160	0.364	-0.692
Эксцесс	4.309	4.363	5.555
Минимум	-3.682	-3.136	-5.728
Первая квартиль	-0.647	-0.666	-0.636
Медиана	-0.028	0.018	0.290
Вторая квартиль	0.674	0.692	0.863
Максимум	4.359	4.365	3.719

7 Кригинг

Одним из часто используемых в геостатистике методов интерполяции является *Обычный кригинг* (*Ordinary kriging*). Этот метод имеет свойство “лучшего линейного несмещенного оценщика” (B.L.U.E. — best linear unbiased estimator). Обычный кригинг “линейный” поскольку его оценки являются взвешенными линейными комбинациями доступных данных. “Несмещенность” достигается нулевым средним значением разницы (m_R) или ошибки. Термин “лучший” характеризует минимизацию вариации ошибок σ_R^2 . Это последнее является отличительной чертой обычного кригинга [10].

Важным моментом метода является то, что m_R неизвестно и поэтому нельзя гарантировать его точное равенство нулю. Так же неизвестно σ_R^2 , поэтому нельзя его минимизировать. Лучшее, что можно сделать в этом случае, это построить модель исследуемых данных и работать со средней ошибкой и вариацией ошибки модели. Для этого используется вероятностная модель, в которой можно подсчитать смещение и ошибку вариации. Затем нужно выбрать веса для всех данных в окрестности так, чтобы средняя ошибка модели ($\bar{m}_R$) равнялась нулю и смоделированная вариация ошибки ($\bar{\sigma}_R^2$) была минимальна.

Оценим значение в каждой непромеренной точке путем взвешенной линейной комбинации доступных значений:

$$\hat{v} = \sum_{j=1}^n w_j \cdot v \quad (27)$$

Набор весов может меняться в зависимости от местоположения оцениваемой точки.

Ошибка каждой оценки определяется как разница между оцененным и действительным значением в точке оценивания:

$$r_i = \hat{v}_i - v_i \quad (28)$$

Тогда средняя ошибка k оценок будет:

$$m_r = \frac{1}{k} \sum_{i=1}^k r_i = \frac{1}{k} \sum_{i=1}^k (\hat{v}_i - v_i) \quad (29)$$

Если представить модель как случайную функцию от случайных переменных, по одной для каждой точки измерений ($V(x_1), \dots, V(x_n)$) и одной для точки оценивания ($V(x_0)$), то оценка также будет случайной переменной:

$$\hat{V}(x_0) = \sum_{i=1}^n w_i \cdot V(x_i) \quad (30)$$

Ошибка оценки будет тоже случайной переменной:

$$R(x_0) = \hat{V}(x_0) - V(x_0) = \sum_{i=1}^n w_i \cdot V(x_i) - V(x_0) \quad (31)$$

Математическое ожидание ошибки оценки в каждой точке должно быть равно нулю:

$$\begin{aligned} E\{R(x_0)\} &= E\{\sum_{i=1}^n w_i \cdot V(x_i) - V(x_0)\} \\ &= \sum_{i=1}^n w_i \cdot E\{V(x_i)\} - E\{V(x_0)\} \end{aligned} \quad (32)$$

$$E\{R(x_0)\} = 0 = E\{V(x_i)\} \cdot E\{\sum_{i=1}^n w_i - E\{V(x_0)\}\}$$

Учитывая стационарность случайной функции:

$$E\{V(x_i)\} = E\{V(x_0)\} = E\{V\}$$

$$\sum_{i=1}^n w_i = 1 \quad (33)$$

Таким образом, сумма весов кригинга должна равняться единице.

Обратимся к вариации ошибки σ_R^2 оценки k данных.

$$\begin{aligned} \sigma_R^2 &= \frac{1}{k} \sum_{i=1}^k (r_i - m_R)^2 \\ &= \frac{1}{k} \sum_{i=1}^k [\hat{v}_i - v_i - \frac{1}{k} \sum_{i=1}^k (\hat{v}_i - v_i)]^2 \end{aligned} \quad (34)$$

Если учесть, что среднее значение ошибки равно нулю, то выражение упростится:

$$\begin{aligned} \sigma_R^2 &= \frac{1}{k} \sum_{i=1}^k (r_i - 0)^2 \\ &= \frac{1}{k} \sum_{i=1}^k [\hat{v}_i - v_i] \end{aligned} \quad (35)$$

Найдем вариацию ошибки:

$$\begin{aligned} Var\{R(x_0)\} &= Cov\{\hat{V}(x_0)\hat{V}(x_0)\} - Cov\{\hat{V}(x_0)V(x_0)\} \\ &\quad - Cov\{V(x_0)\hat{V}(x_0)\} + Cov\{V(x_0)V(x_0)\} \\ &= Cov\{\hat{V}(x_0)\hat{V}(x_0)\} - 2 \cdot Cov\{\hat{V}(x_0)V(x_0)\} + Cov\{V(x_0)V(x_0)\} \end{aligned} \quad (36)$$

$$\begin{aligned} Cov\{\hat{V}(x_0)\hat{V}(x_0)\} &= Var\{\hat{V}(x_0)\hat{V}(x_0)\} \\ &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j \tilde{C}_{ij} \end{aligned} \quad (37)$$

$$Cov\{\hat{V}(x_0)V(x_0)\} = \tilde{\sigma}^2$$

$$\begin{aligned} 2 \cdot Cov\{\hat{V}(x_0)V(x_0)\} &= 2 \cdot \sum_{i=1}^n w_i \cdot E\{V_i, V_0\} - 2 \cdot \sum_{i=1}^n w_i \cdot E\{V_i\} \cdot E\{V_0\} \\ &= 2 \sum_{i=1}^n w_i \tilde{C}_{i0} \end{aligned} \quad (38)$$

В итоге получим:

$$\sigma_R^2 = \sigma^2 + \sum_{i=1}^n \sum_{j=1}^n w_i w_j \tilde{C}_{ij} - 2 \sum_{i=1}^n w_i \tilde{C}_{i0} \quad (39)$$

Далее следует минимизировать вариацию. Для этого вводится параметр Лагранжа μ .

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^n \sum_{j=1}^n w_i w_j \tilde{C}_{ij} - 2 \sum_{i=1}^n w_i \tilde{C}_{i0} - 2\mu \underbrace{\left(\sum_{i=1}^n w_i - 1 \right)}_0 \quad (40)$$

Если взять производные от левой и правой частей по всем весам w_i и приравнять правые части к нулю, то в итоге получится следующая система из n уравнений:

$$\sum_{j=1}^n w_j \tilde{C}_{ij} + \mu = \tilde{C}_{i0}, \quad \forall i = 1, \dots, n \quad (41)$$

Для замыкания системы прибавим уравнение для весов:

$$\sum_{i=1}^n w_i = 1 \quad (42)$$

Для получения вариации умножим n уравнений системы на w_i и сложим их.

$$\sum_{i=1}^n \sum_{j=1}^n w_i w_j \tilde{C}_{ij} = \sum_{i=1}^n w_i \tilde{C}_{i0} - \mu \quad (43)$$

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^n w_i \tilde{C}_{i0} + \mu \quad (44)$$

При переходе к вариограммам система уравнений для обычного кригинга записывается следующим образом:

$$\sum_{j=1}^n w_j \tilde{\gamma}_{ij} - \mu = \tilde{\gamma}_{i0}, \quad \forall i = 1, \dots, n \quad (45)$$

$$\sum_{i=1}^n w_i = 1 \quad (46)$$

а вариация ошибки модели будет равна:

$$\tilde{\sigma}_R^2 = \sum_{i=1}^n w_i \tilde{\gamma}_{i0} + \mu \quad (47)$$

Различают точечные и блочные оценки.

Точечный кригинг

Точечный (point) кригинг — оценка точечного значения из набора окружающих точек. При совпадении оцениваемой точки с экспериментальной точкой значение, полученное из кригинга, совпадает с измеренным значением.

Блочный кригинг

Для более точной оценки можно усреднить несколько точечных оценок, полученных в окрестности оцениваемой точки. Так, вся рассматриваемая область разбивается на блоки, в каждом блоке производится несколько точечных оценок и усредняются. В итоге получается предсказанное локальное значение. Этот метод называется блочным (block) кригингом.

8 Метод обратных квадратов

Для сравнения результатов, полученных из кригинга, был использован известный метод пространственной интерполяции — *метод обратных квадратов* (*Inverse Distance Squared*). В этом методе оценка так же вычисляется как линейная комбинация взвешенных значений. Идея метода состоит в том, чтобы дать наиболее удаленным токам наибольшие веса, а ближайшим — наименьшие. При этом зависимость может быть обратно пропорциональна расстоянию возведенному в степень. В выбранном методе это квадрат расстояния.

Таким образом значение оценки вычисляется следующим образом:

$$\hat{v}_1 = \frac{\sum_{i=1}^n \frac{1}{d_i^2} v_i}{\sum_{i=1}^n \frac{1}{d_i^2}} \quad (48)$$

где: d_i — расстояние от соседней точки со значением v_i до точки оценивания.

9 Результаты интерполяции

Для рассматриваемой юго-восточной части Калужской области был проведен последовательный геостатистический анализ, включая обычный кригинг с изотропной и анизотропной моделями для различных окрестностей поиска. Были использованы полученные выше модели для Cs^{137} . Для кригинга с изотропной моделью окрестностью поиска был круг радиусом 80 км. Для анизотропной модели выбирался эллипс (длины большой и малой полуосей 80 и 40 км.) с различной ориентацией (0° , 135°). Результаты интерполяции представлены на цветной вкладке. Отличия между кригингом с областью поиска разной ориентации незначительны и располагаются в основном в правом нижнем углу. Там образовались зоны с характерной анизотропной ориентацией под 135° . Абсолютные значения разницы между изотропным и анизотропным случаем представлены на рисунке 40. На рисунках в цветной вкладке нанесены значения ошибок кригинга для изотропного и анизотропного моделирования. Во втором случае ошибка значительно меньше на большей части области.

Результаты интерполяции методом обратных квадратов представлены на цветной вкладке. На следующем рисунке цветной вкладки нанесены разницы между оценками кригинга с анизотропной моделью и методом обратных квадратов. Расхождения в оценках особенно существенны в центре области, где метод обратных квадратов дает недооценку.

Результаты проведенной оценки сведены в таблицу.

Был проведен учет тренда при кригинге. В одном случае брался линейный тренд в направлениях x и y , а во втором случае квадратичный тренд только в направлении y . То есть в том направлении, где был выявлен значительный тренд. На это указала вариограмма Гауссова типа в направлении 90° . Результаты представлены в виде карт разниц между оценками кригинга для анизотропной модели без учета и с учетом тренда (рис. 41 и 42 соответственно для первого и второго случаев). Видно, что оценки имеют отличия в основном в углах области, причем при учете только вертикального тренда область большой разницы (правый нижний угол) меньше и вытянута только по вертикали (рис. 42). Это объясняется малым количеством данных в этом углу. Совпадение оценок с учетом и без учета тренда говорит о том, что кригинг с достаточно малым окном (1×1 км.) автоматически учитывает тренд.

10 Заключение

В работе осуществлена одна из методик анализа пространственно распределенных данных. В процессе исследования показано, что различные методы дают разные результаты о пространственном распределении одних и тех же данных. Проведено сравнение между изотропным и анизотропным рассмотрением. Можно сделать вывод о наличии значительной анизотропии в Чернобыльских выпадениях. Учет этой анизотропии дает возможность понизить ошибку интерполяции. Сделан акцент на сравнении геостатистических подходов с классическим методом интерполяции — методом обратных квадратов. Результаты сравнения наглядно представлены на картах и в таблицах пространственного распределения оценок.

Статистика распределений пространственных оценок

Статистика	Обычный кригинг				Метод обратных квадратов
	изотропная мод.		анизотропная мод.		
	Оценка	Вариация	Оценка	Вариация	
Число данных	6000	6000	6000	6000	6000
Число опущенных данных	0	0	0	0	0
Число данных меньше нуля	2	0	36	0	0
Среднее значение	3.085	1.742	3.150	2.367	3.599
Вариация	4.586	3.021	4.372	4.051	2.794
Стандартное отклонение	2.142	1.738	2.091	2.013	1.672
Коэффициент вариации	69.423	99.779	66.368	85.011	46.451
Коэффициент симметрии	.889	2.035	.848	1.408	.402
Экссесс	3.744	6.870	3.776	4.037	3.281
Минимум	-.001	.008	-.039	.013	.064
Первая квартиль	1.410	.679	1.524	.984	2.333
Медиана	2.634	1.026	2.842	1.516	3.621
Вторая квартиль	4.432	2.110	4.379	3.098	4.685
Максимум	12.334	9.196	12.332	8.390	2.156

В работе уделено большое внимание детальному описанию методов структурного анализа, кросс-валидации и теории кригинга.

Список литературы

- [1] M. Kanevsky, R. Arutyunyan, L. Bolshov, I. Linge, E. Savel'eva, T. Haas. Spatial Data Analysis of Chernobyl Fallout. Preliminary Results. Preprint NSI-23-93, Moscow 1993.
- [2] M. Kanevsky, R. Arutyunyan, L. Bolshov, I. Linge, E. Savel'eva, T. Haas. Spatial Data Analysis of Chernobyl Fallout. Monitoring network description. Univariate statistics and declustering. Cross-validation. Preprint NSI-1-94, Moscow 1994.
- [3] Р.В. Арутюнян, Л.А. Большов, В.В. Демьянов, Е.В. Каневская, М.Ф. Каневский, В.П. Киселев, И.И. Линге. Методика Geo-EAS анализа пространственно распределенных данных. Пример исследования: Чернобыльские выпадения. Препринт ИБРАЭ. Москва 1994.
- [4] I. Clark. Practical Geostatistics. Elsevier Applied Science Publishers, London and New York, repr. 1984.
- [5] N. Cressie. Statistics for Spatial Data. John Wiley & Sons, N.Y. 1991.
- [6] M. David. Handbook of Applied Advanced Geostatistical Ore Reserve Estimation. Elsevier Applied Science Publishers, Amsterdam B.V., 1988.
- [7] B. M. Davis. Uses and Abuses of Cross-Validation in Geostatistics. Mathematical Geology, vol. 19, no. 3, 1987.
- [8] C. Deutsch. Declus: A Fortran 77 Program for Determining Optimum Spatial Declustering Weights. Computers & Geosciences, vol. 15, no. 3, 1989.
- [9] Geo-EAS 1.2.1 — Geostatistical Environmental Assessment Software, User's Guide, E. Englund A. Sparks U.S. EPA, Las Vegas 1991.
- [10] Ed.H. Isaaks, R.M. Srivastava, An Introduction to Applied Geostatistics, Oxford, Oxford University Press, 1989.
- [11] C. Deutsch, A.G. Journel. GSLIB, Geostatistical Software Library and User's Guide. Oxford University Press. 1992.
- [12] A.G. Journel, C.J. Huijbregts, Mining Geostatistics, London, Academic Press, 1978.
- [13] A.G. Journel. Nonparametric Estimation of Spatial Distribution. Mathematical Geology, vol. 15, no. 3, 1983.
- [14] M.A. Oliver, R. Webster. Kriging: a method of interpolation for geographical information systems. Int. J. Geographical Information Systems, 1990, v.4, No.3, 313-332.
- [15] B.D. Ripley. Spatial Statistics. John Wiley & Sons, N.Y. 1985.
- [16] D. L. Zimmerman. Another Look at Anisotropy in Geostatistics. Mathematical Geology, vol. 25 no. 4, 1993.

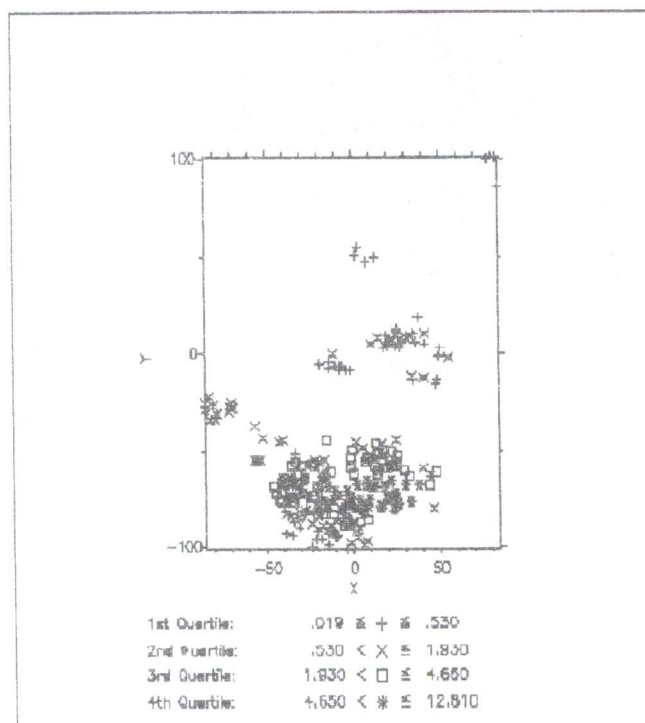


Рис. 1: Карта расположения данных по загрязнению Cs^{137} (Кюри/км.²).

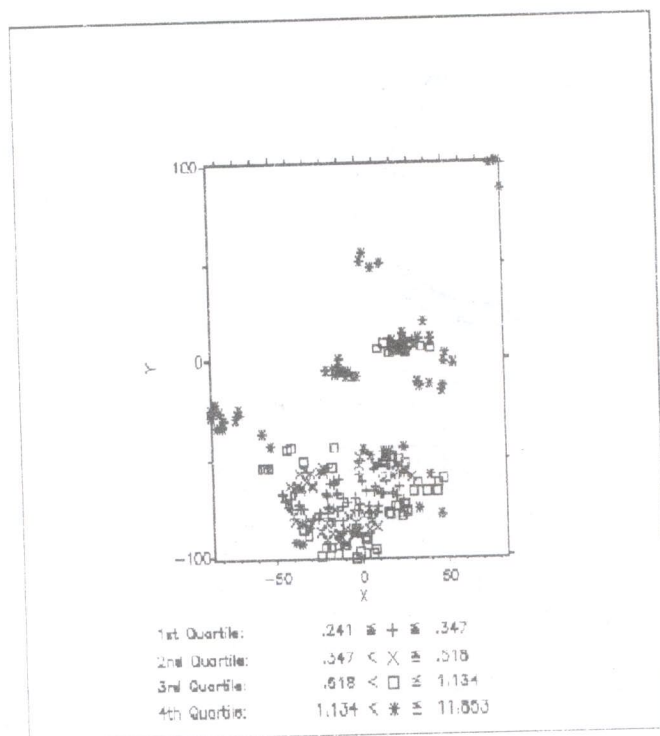


Рис. 2: Карта весов декластеризации по квартилям.

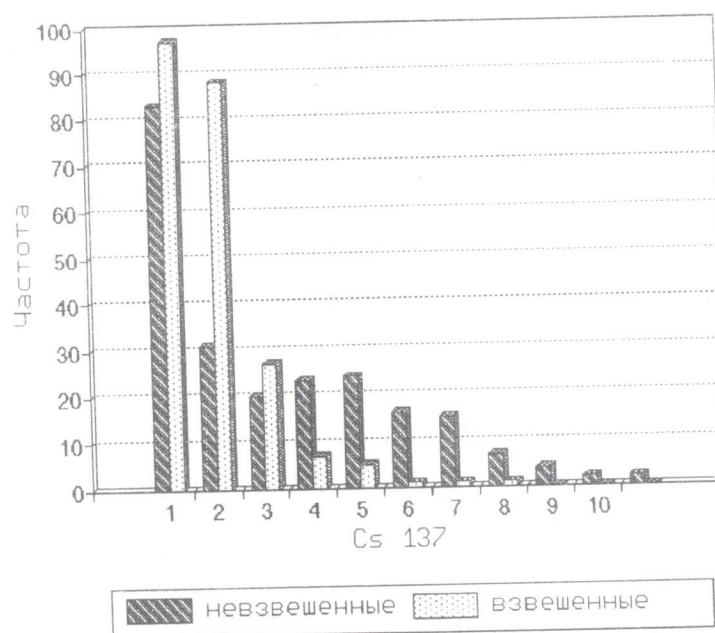


Рис. 3: Гистограмма распределения опытных точек и взвешенных значений.

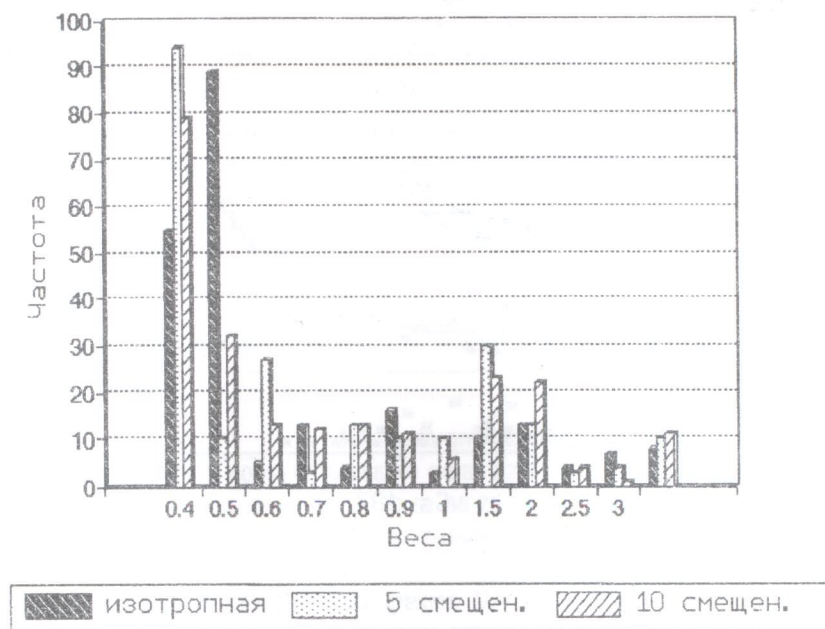


Рис. 4: Гистограмма распределения весов при декластеризации на изотропной сетке и анизотропной сетке при пяти и десяти смещениях начала сетки.

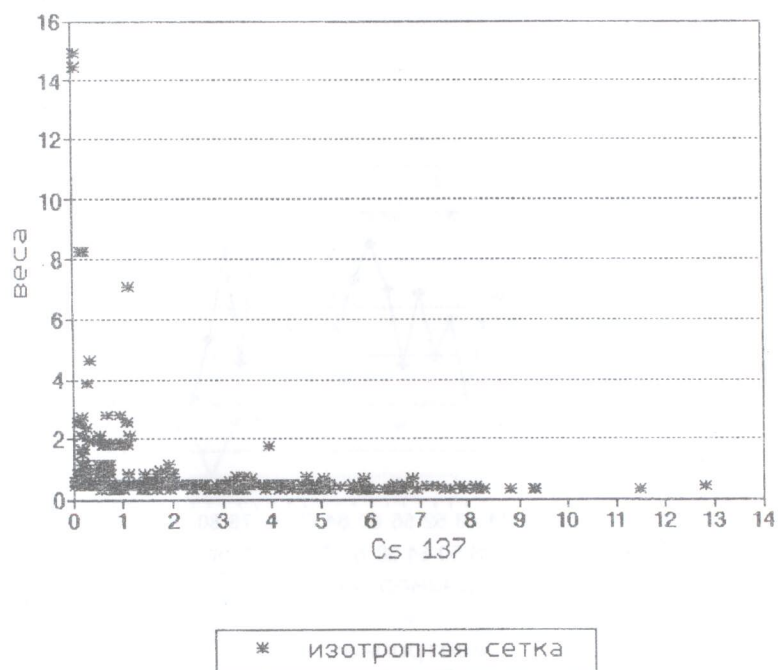


Рис. 5: Диаграмма зависимости весов от измеренных значений при декластеризации на изотропной сетке.

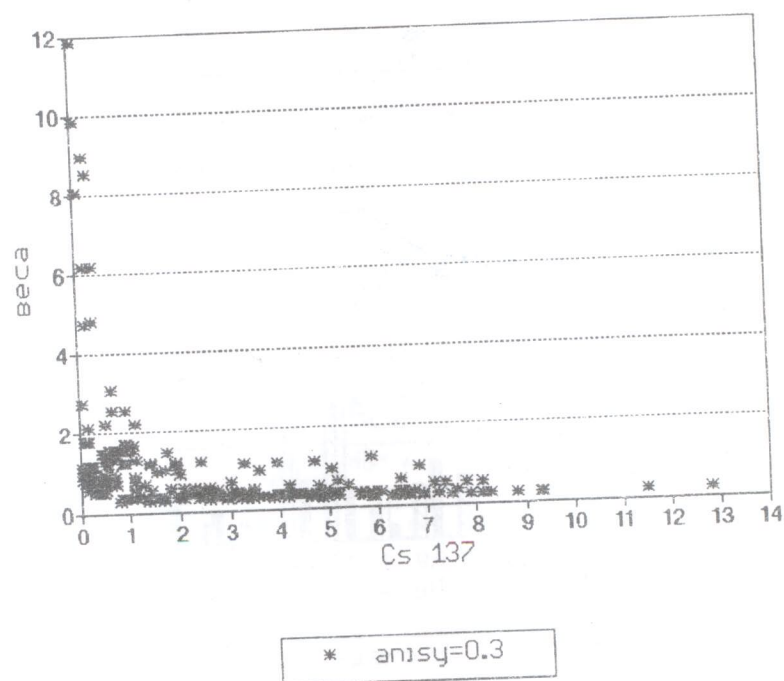


Рис. 6: Диаграмма зависимости весов от измеренных значений при декластеризации на анизотропной сетке.

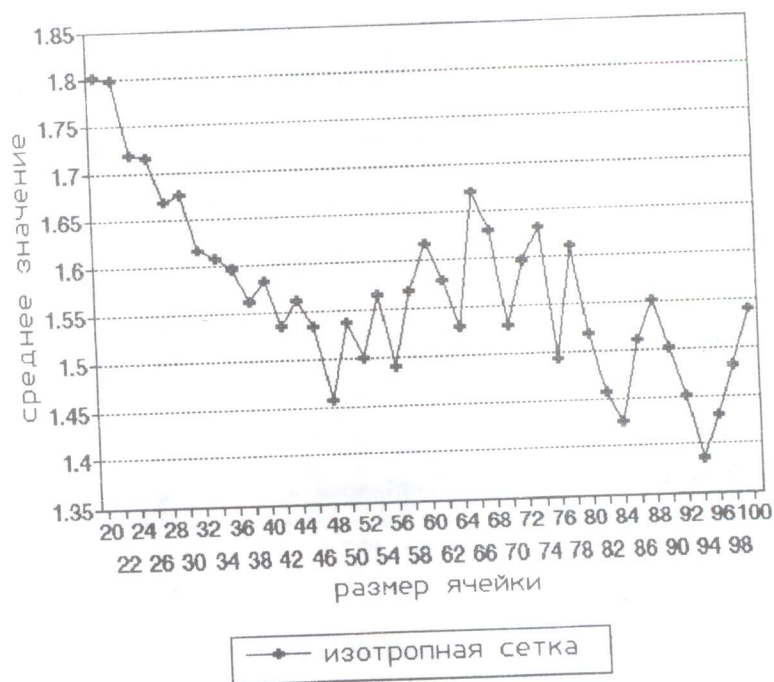


Рис. 7: Диаграмма средних значений в зависимости от размера ячейки при декластеризации на изотропной сетке.

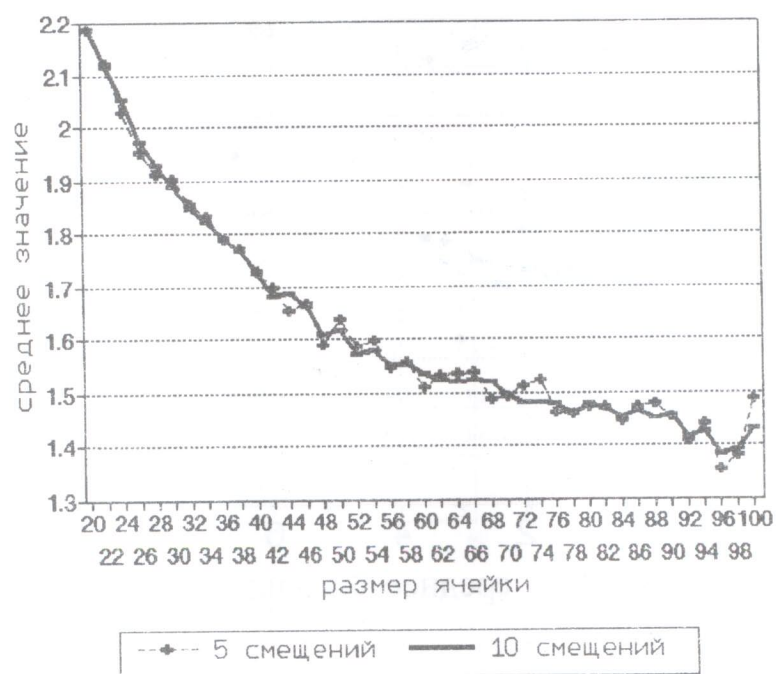


Рис. 8: Диаграмма средних значений в зависимости от размера ячейки при декластеризации на анизотропной сетке с пятью и десятью смещениями.



Рис. 9: Диаграмма зависимости между локальным средним значением и стандартным отклонением.

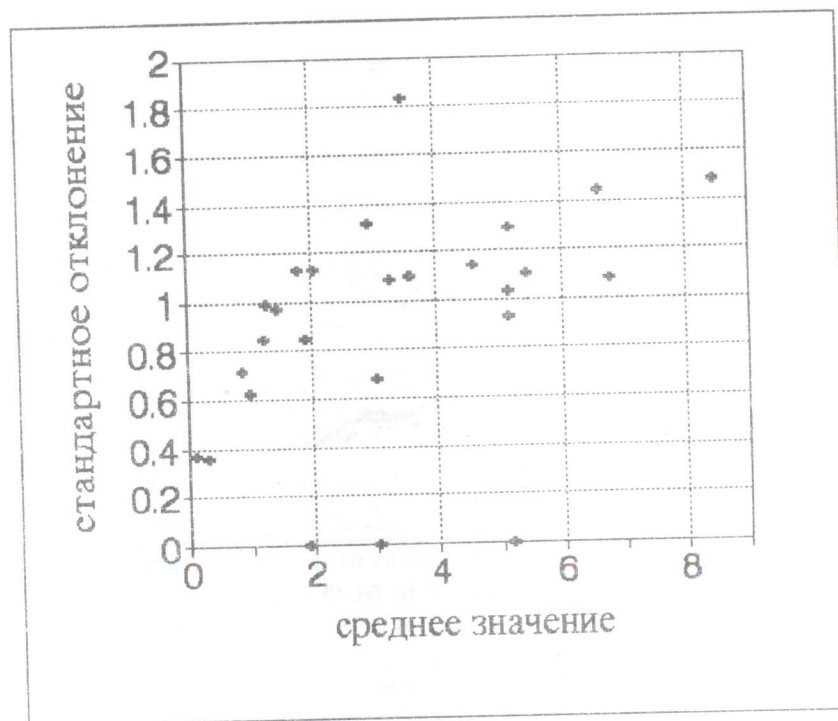


Рис. 10: График зависимости стандартного отклонения от локального среднего значения.

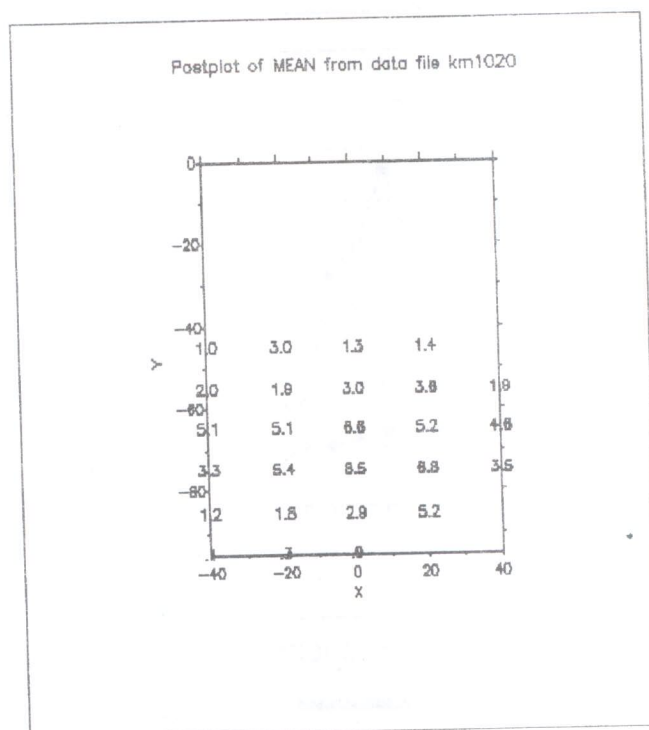


Рис. 11: Пространственное распределение локального среднего значения.

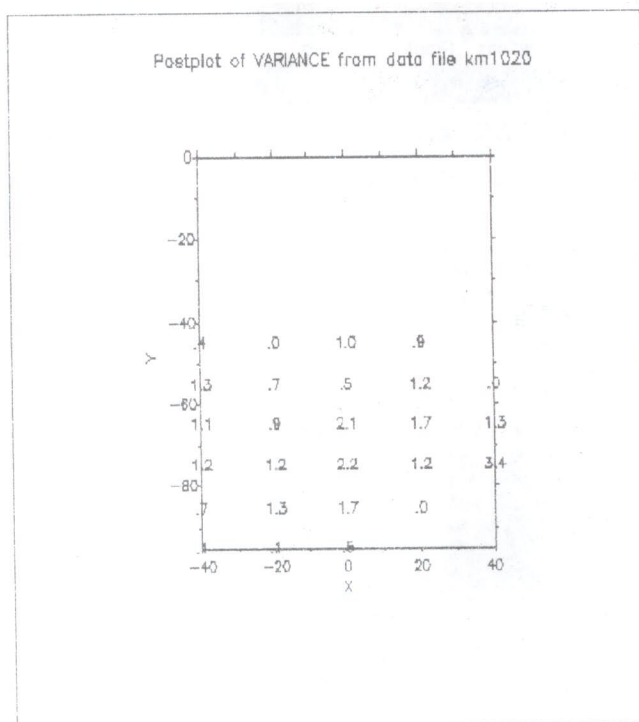


Рис. 12: Пространственное распределение вариации.

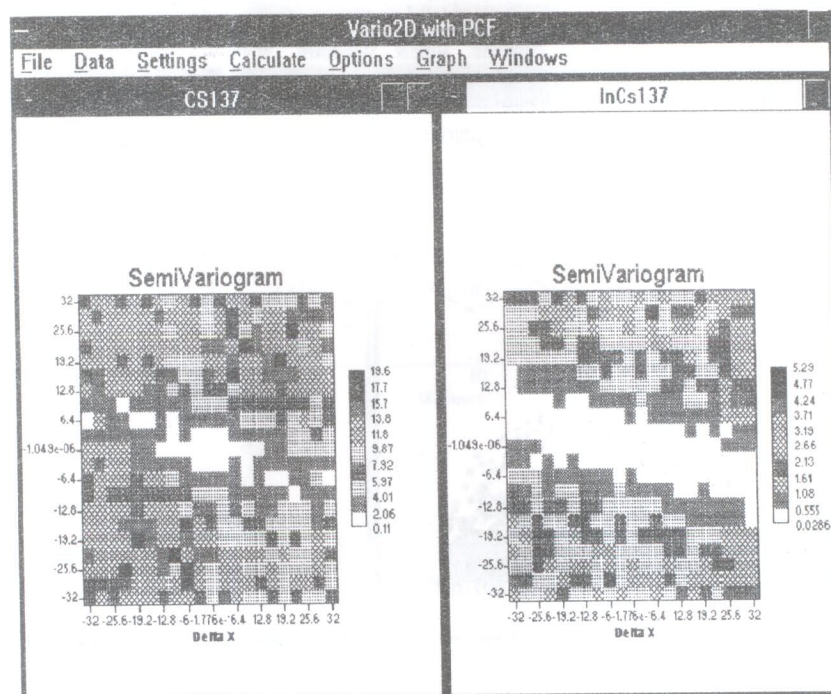


Рис. 13: Вариограммные поверхности для Cs^{137} и $\ln Cs^{137}$.

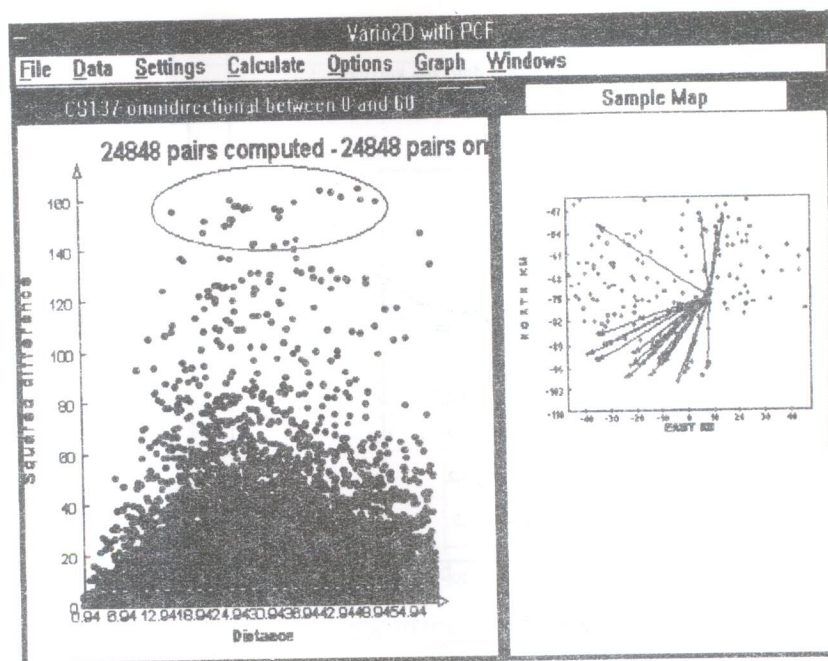


Рис. 14: Вариограммное облако и карта точек для Cs^{137} по всем направлениям.

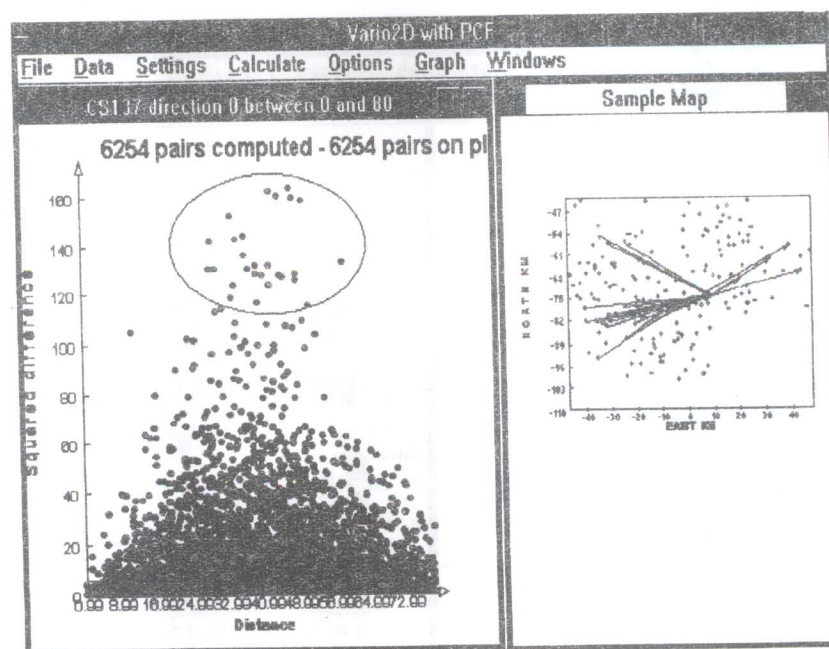


Рис. 15: Вариограммное облако и карта точек для Cs^{137} в направлении 0° .

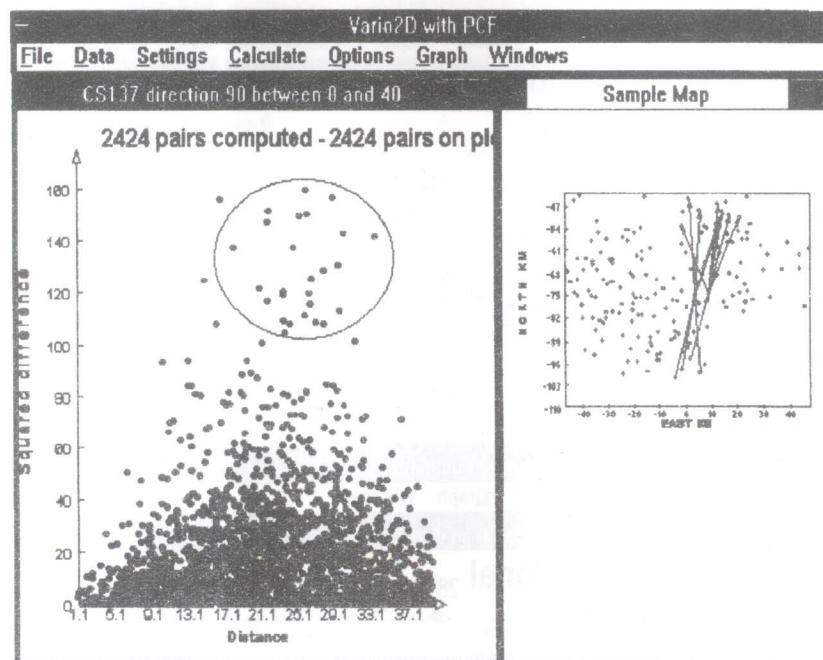


Рис. 16: Вариограммное облако и карта точек для Cs^{137} в направлении 90° .

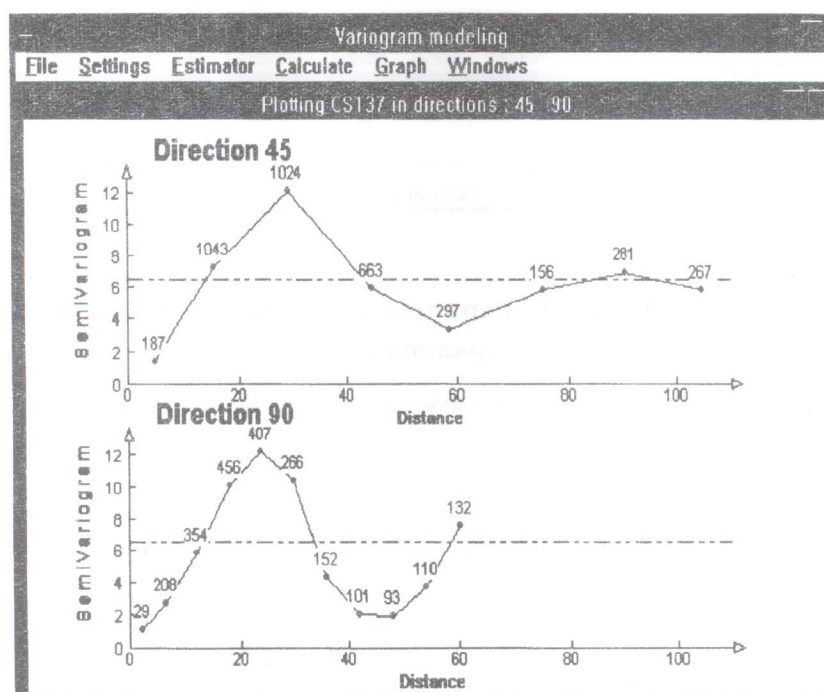


Рис. 17: Экспериментальные вариограммы для Cs^{137} в направлении 45° и в направлении 90° .

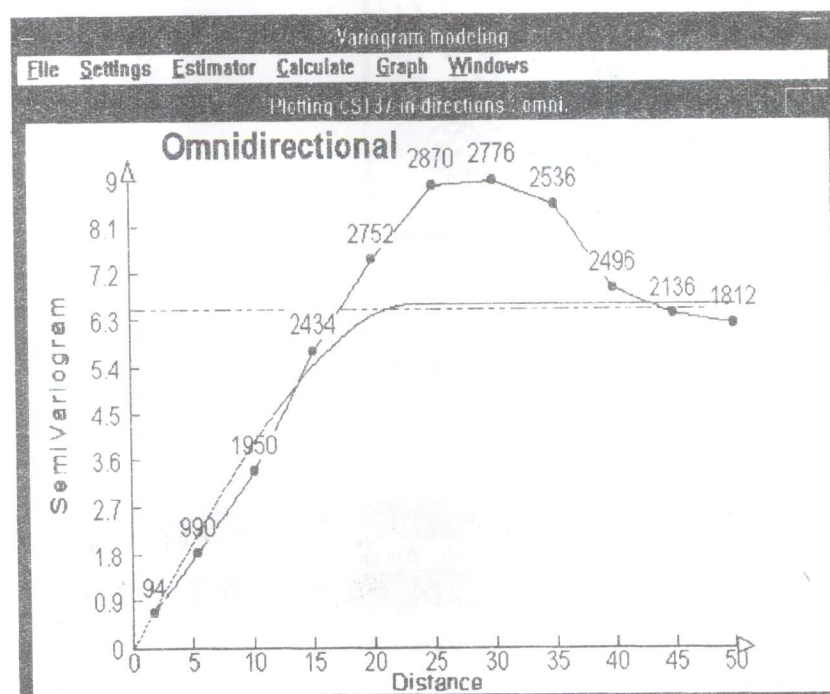


Рис. 18: Экспериментальная вариограмма и модель для Cs^{137} по всем направлениям.

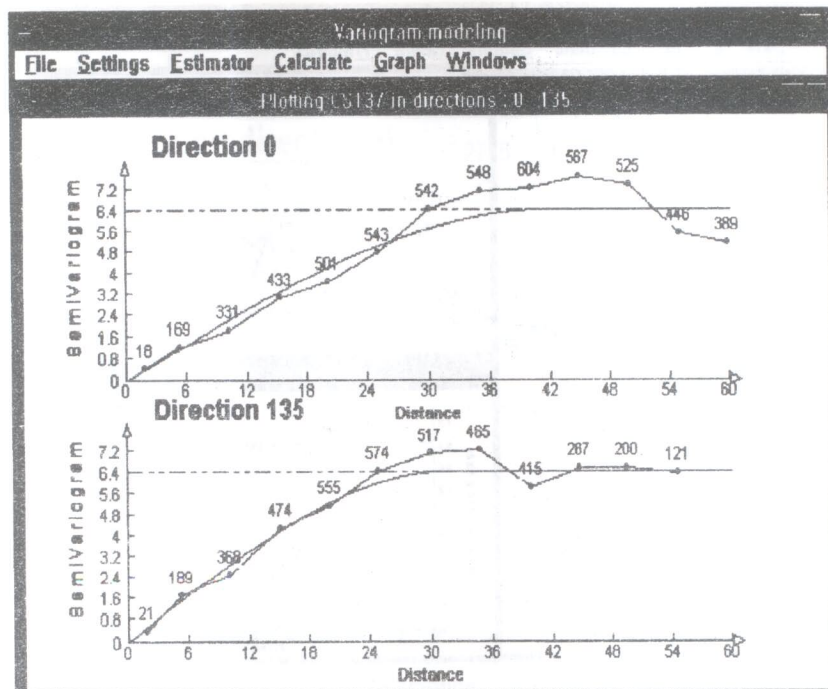


Рис. 19: Экспериментальные вариограммы и модель для Cs^{137} в направлениях 0° и 135°.

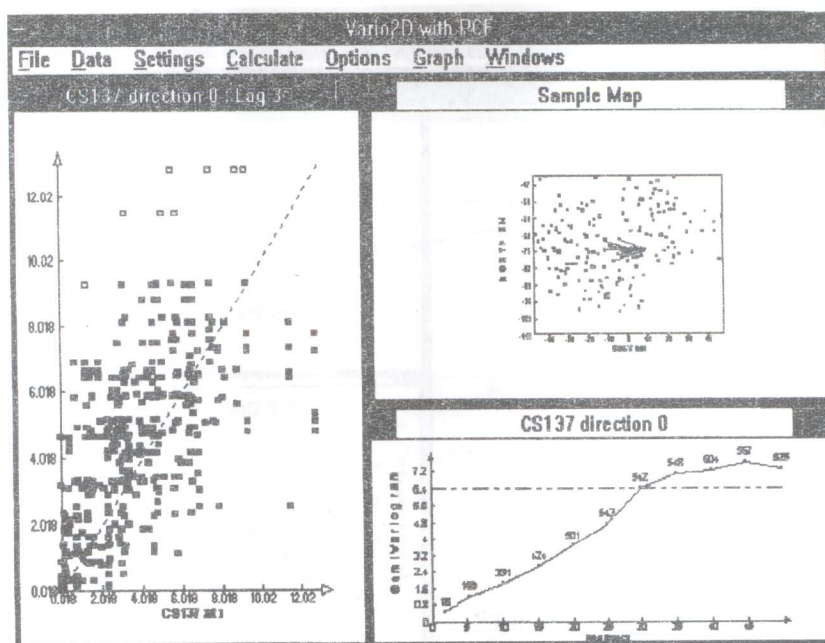


Рис. 20: Диаграмма разброса пар в лаге и экспериментальная вариограмма для Cs^{137} в направлении 0°.

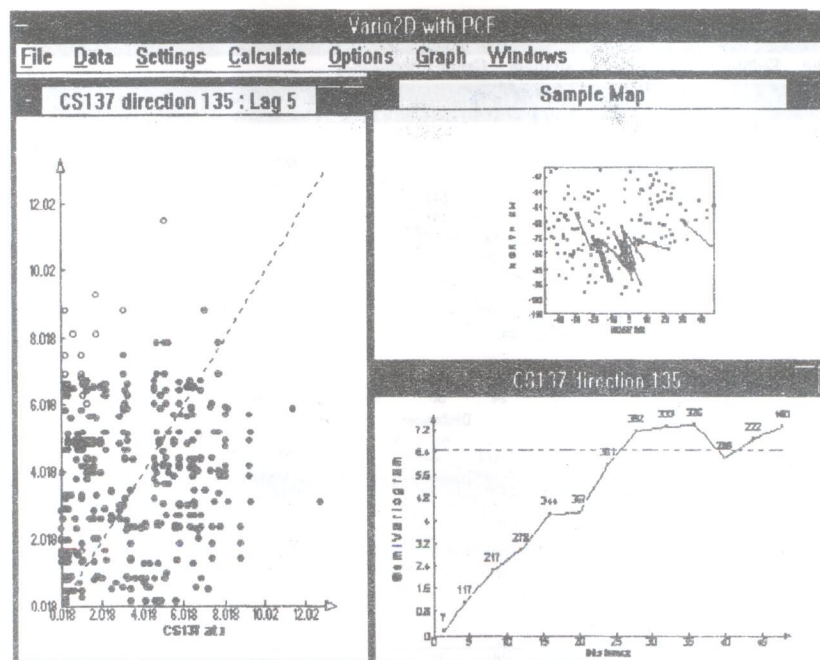


Рис. 21: Диаграмма разброса пар в лаге и экспериментальная вариограмма для Cs^{137} в направлении 135° .

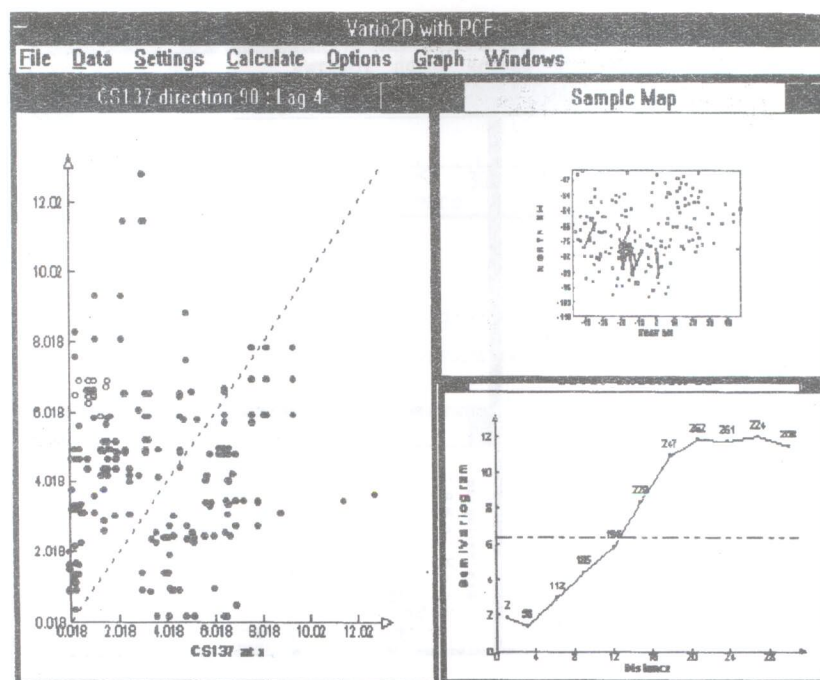


Рис. 22: Диаграмма разброса пар в лаге и экспериментальная вариограмма для Cs^{137} в направлении 90° .

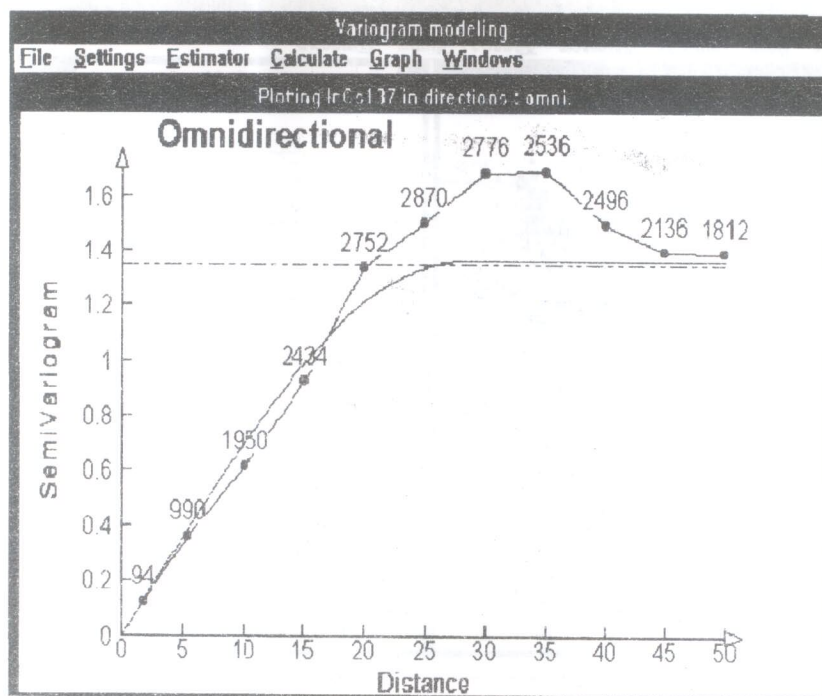


Рис. 23: Экспериментальная вариограмма и модель для $\ln Cs^{137}$ по всем направлениям.

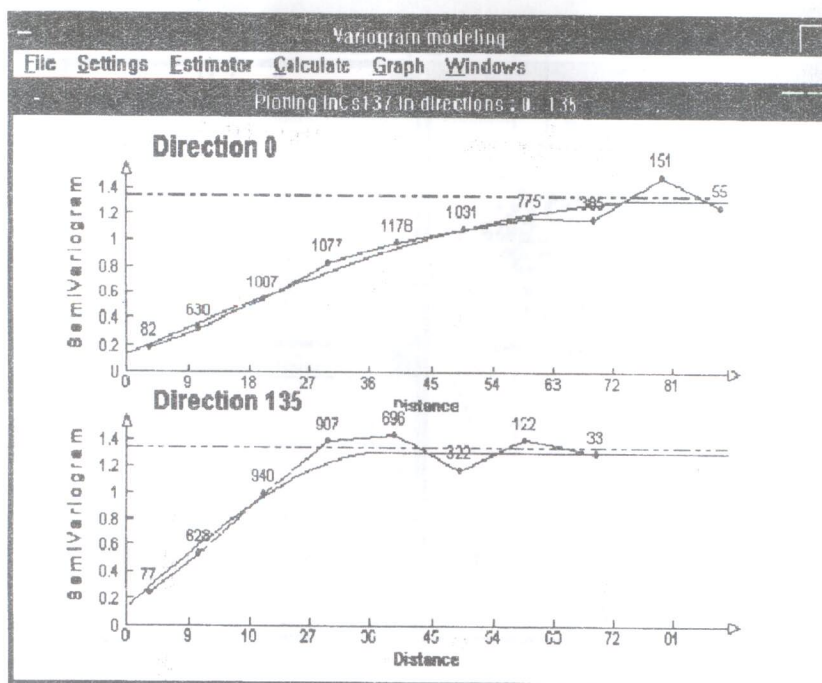


Рис. 24: Экспериментальные вариограммы и модель для $\ln Cs^{137}$ в направлениях 0° и 135° .

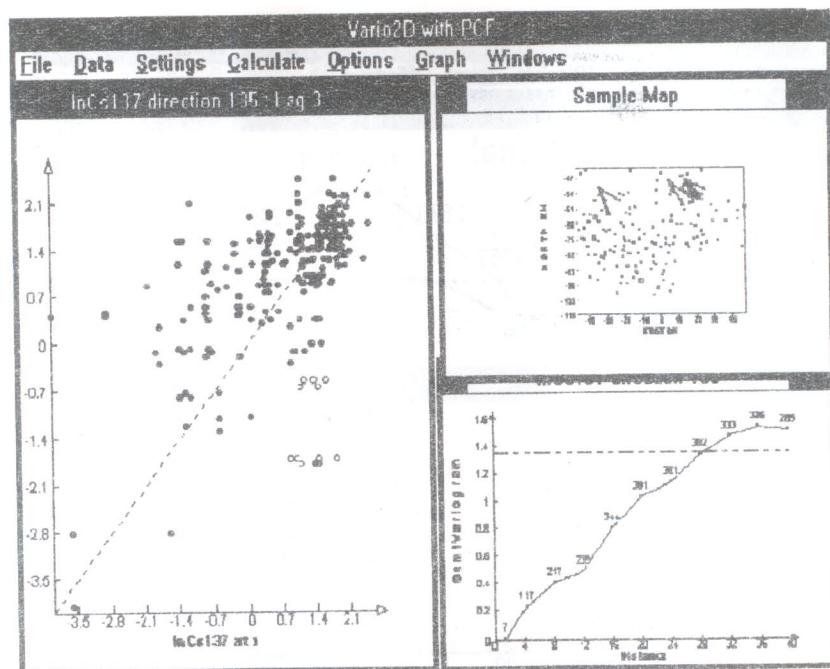


Рис. 25: Диаграмма разброса пар в лаге и экспериментальная вариограмма для $\ln Cs^{137}$ в направлении 135° .

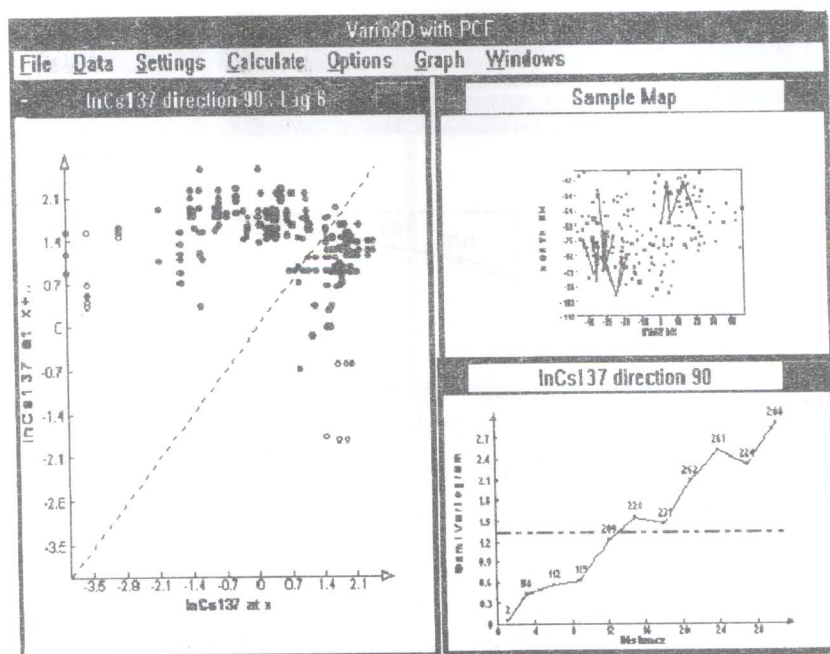


Рис. 26: Диаграмма разброса пар в лаге и экспериментальная вариограмма для $\ln Cs^{137}$ в направлении 90° .

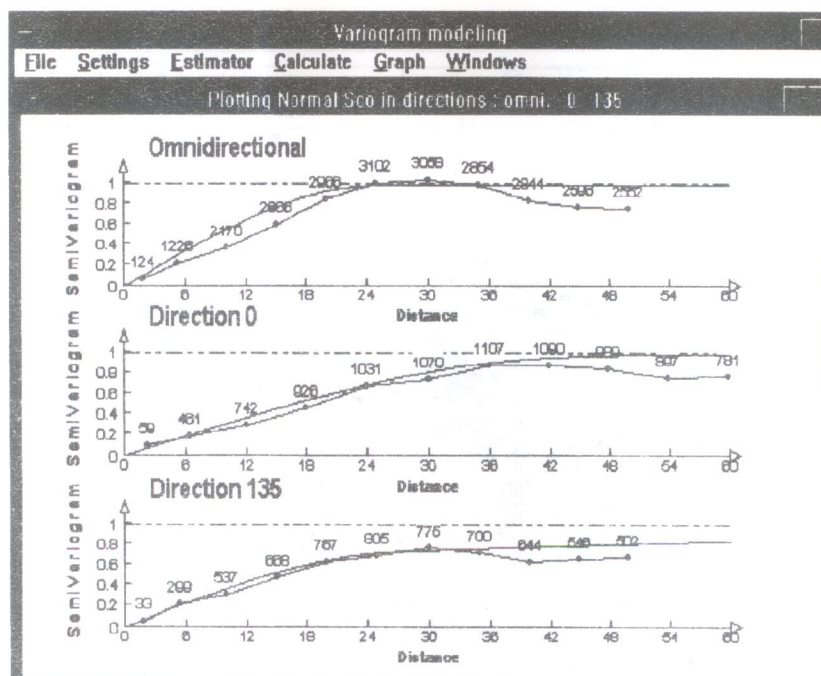


Рис. 27: Экспериментальная вариограмма и модель для преобразованной переменной $Normal\ Score\ Cs^{137}$ по всем направлениям.

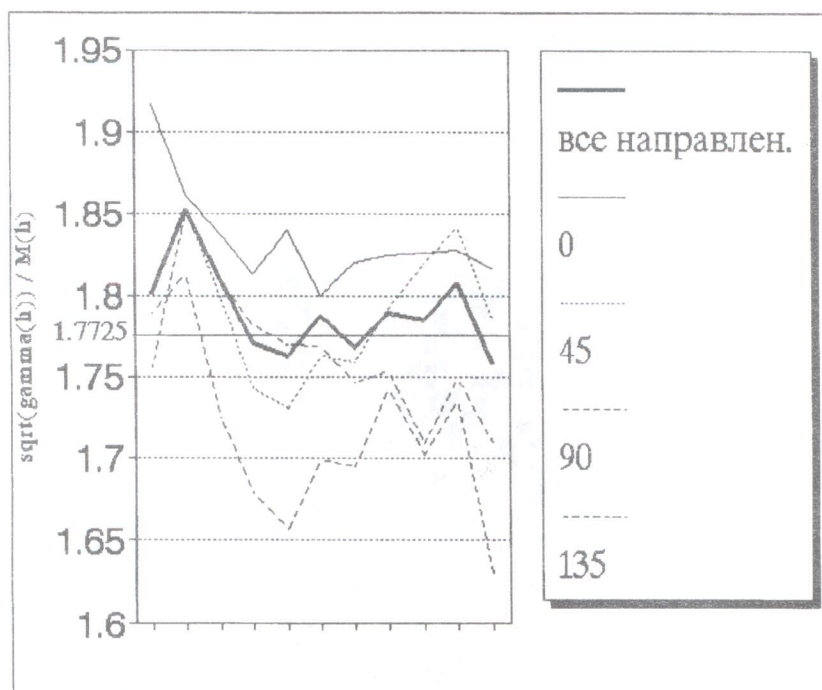


Рис. 28: Диаграмма отношения квадратного корня из вариограммы к мадограмме для преобразованной переменной $Normal\ Score\ Cs^{137}$ в разных направлениях.

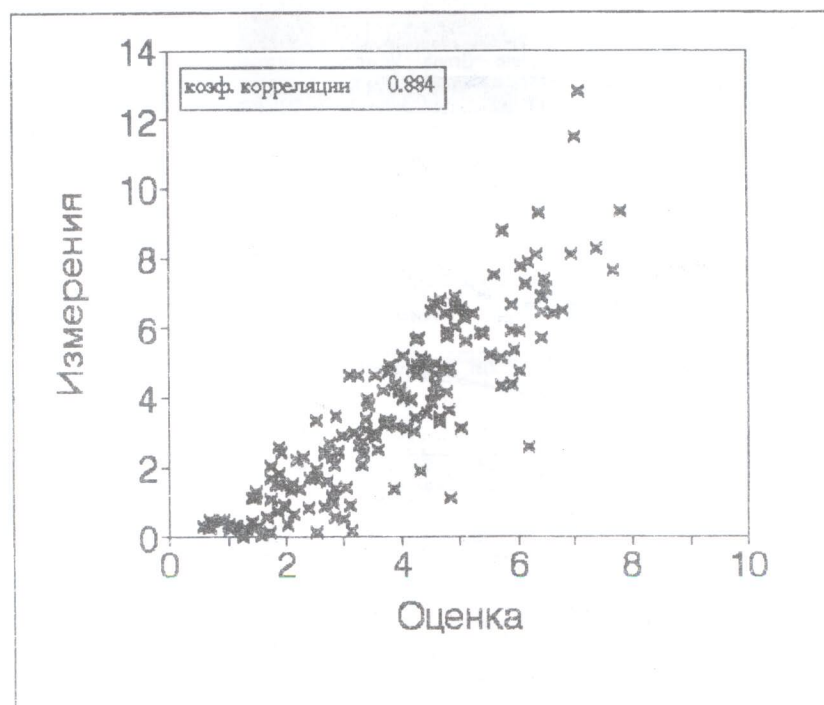


Рис. 29: Диаграмма зависимости измеренных значений от оценок кросс-валидации для метода обратных квадратов.

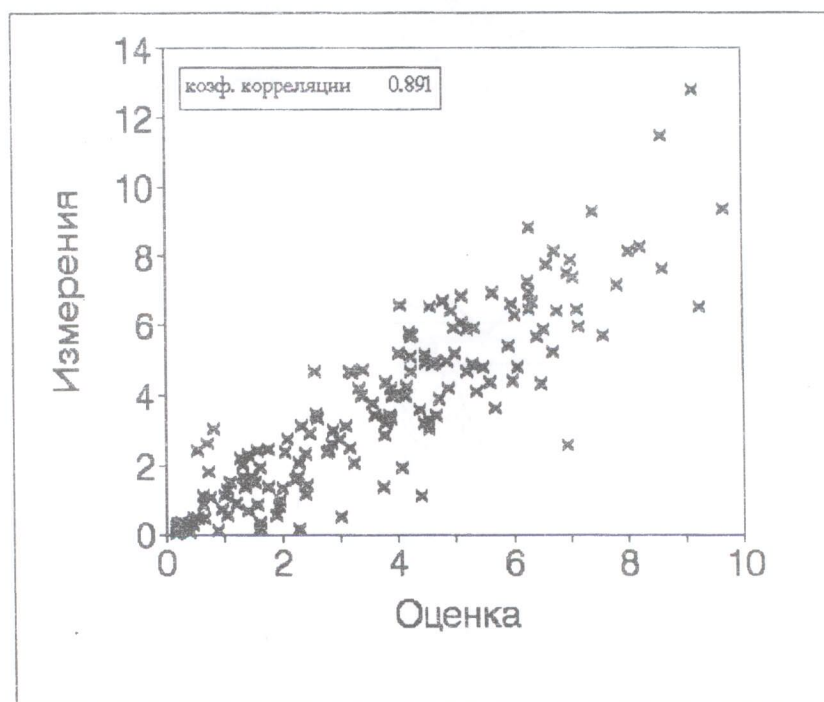


Рис. 30: Диаграмма зависимости измеренных значений от оценок кросс-валидации для изотропной модели.

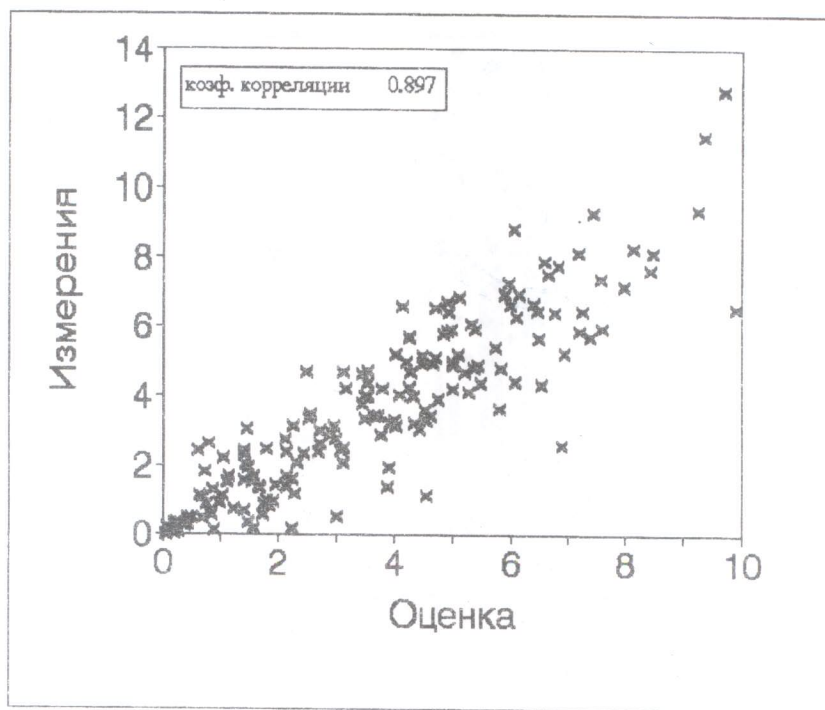


Рис. 31: Диаграмма зависимости измеренных значений от оценок кросс-валидации для анизотропной модели.

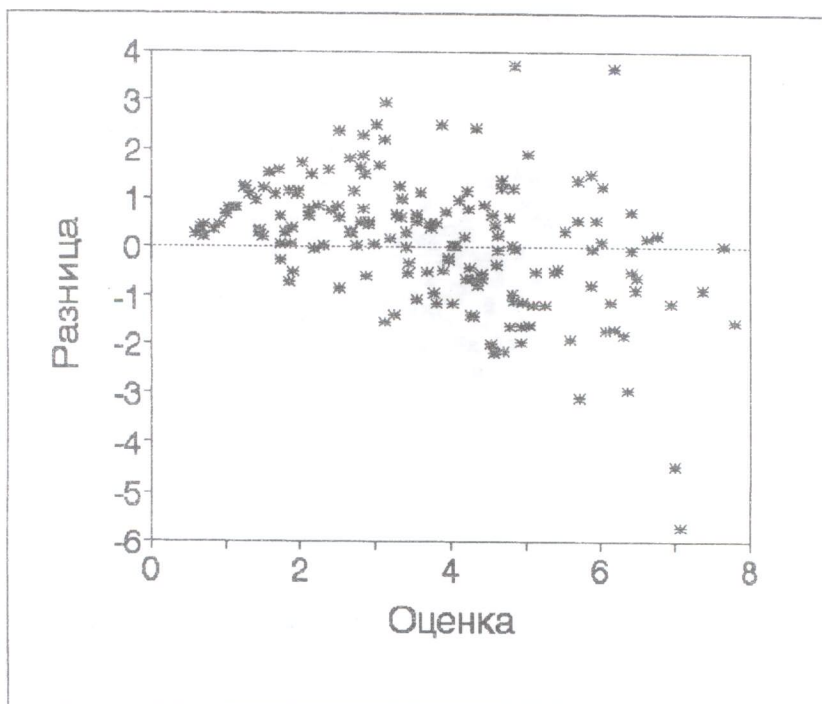


Рис. 32: Диаграмма зависимости разниц (residuals) от оценок кросс-валидации для метода обратных квадратов.

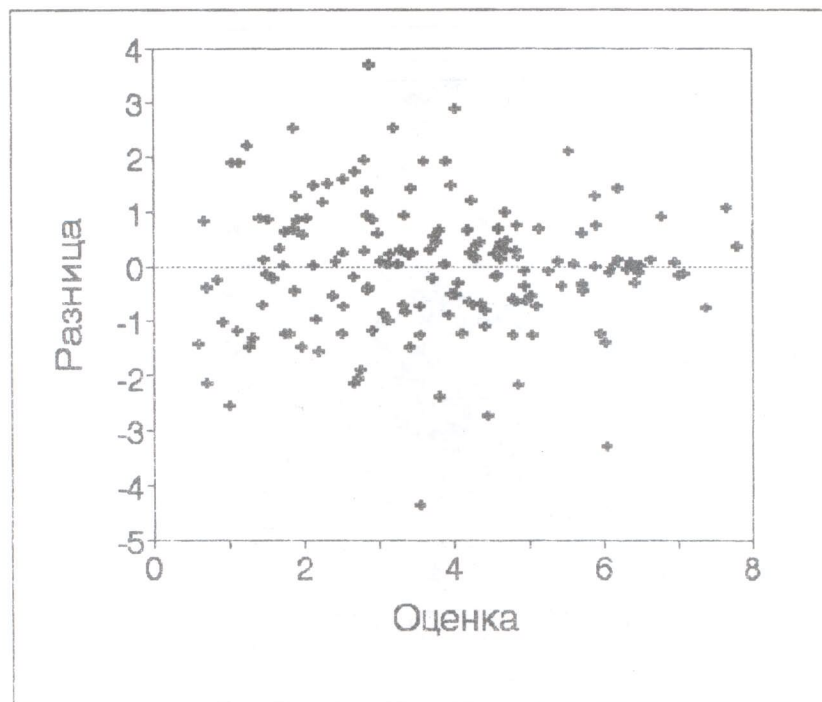


Рис. 33: Диаграмма зависимости разниц (residuals) от оценок кросс-валидации для изотропной модели.

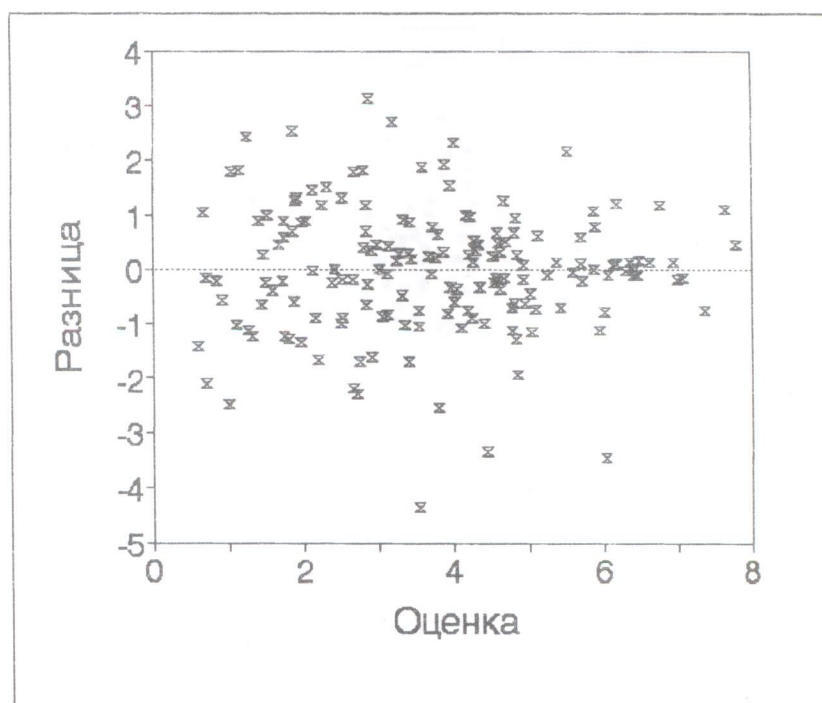


Рис. 34: Диаграмма зависимости разниц (residuals) от оценок кросс-валидации для анизотропной модели.

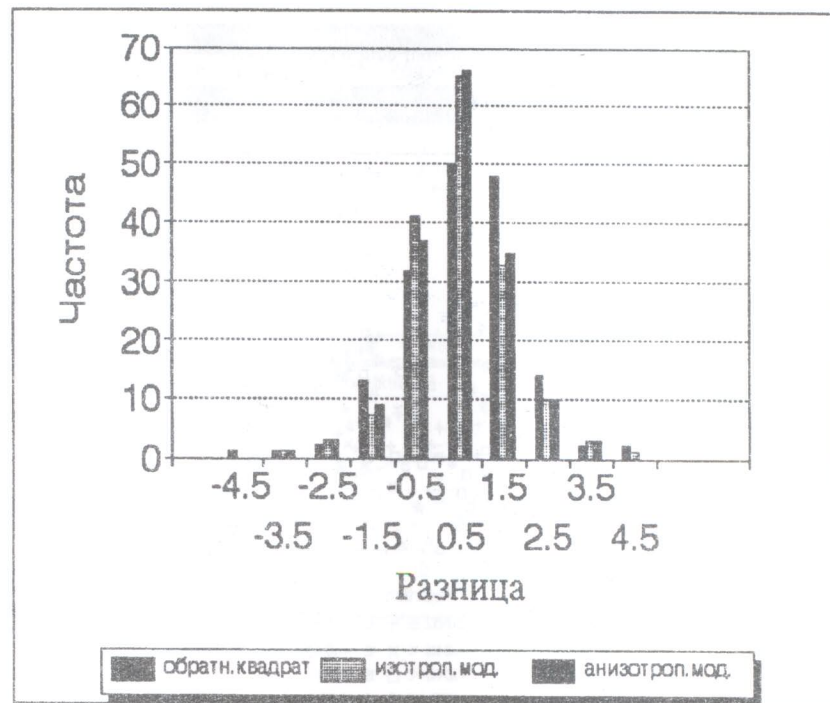


Рис. 35: Гистограмма разниц (residuals) для кросс-валидации трех методов.

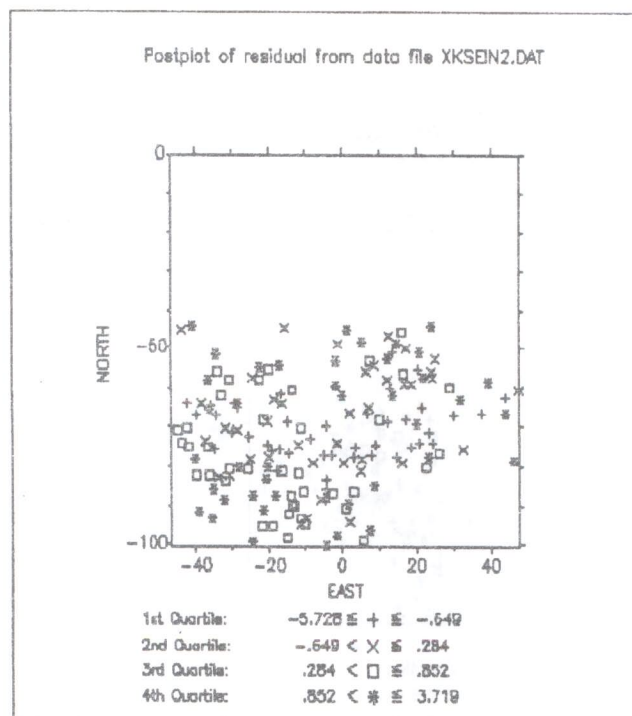


Рис. 36: Пространственное распределение разницы между оценками кросс-валидации для метода обратных квадратов и измеренными значениями.

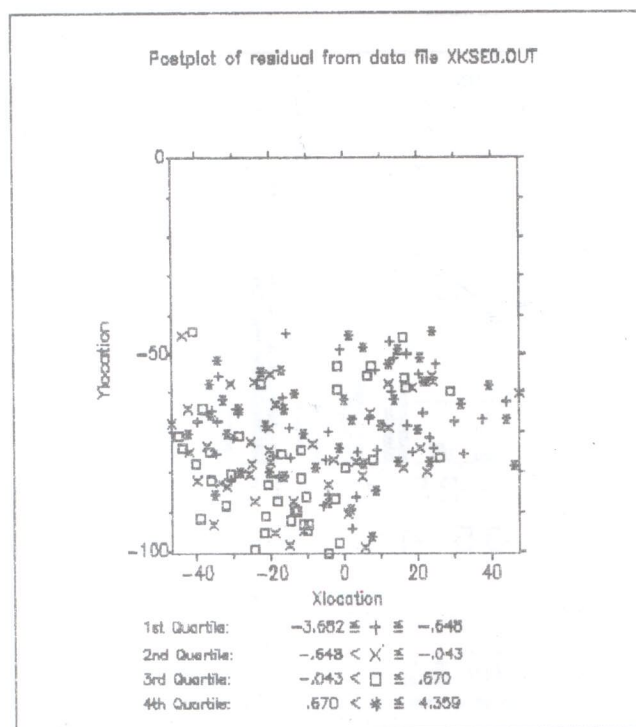


Рис. 37: Пространственное распределения разницы между оценками кросс-валидации для изотропной модели и измеренными значениями.

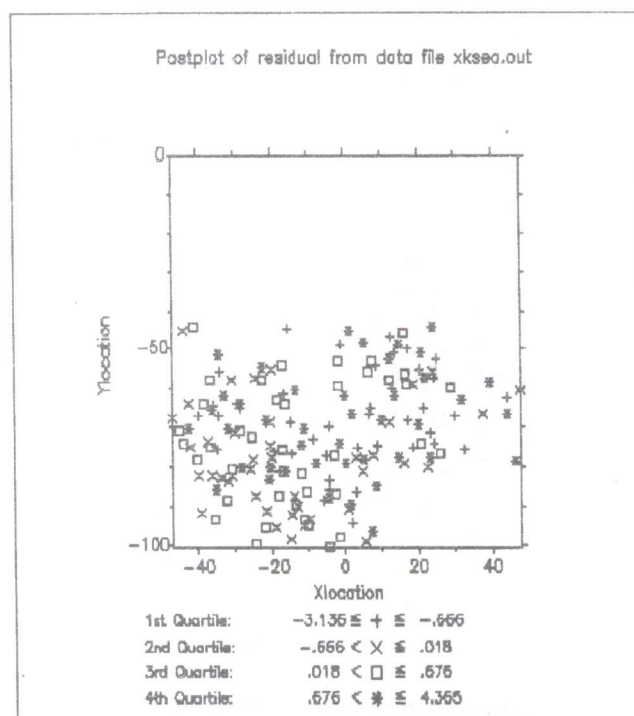


Рис. 38: Пространственное распределения разницы между оценками кросс-валидации для анизотропной модели и измеренными значениями.

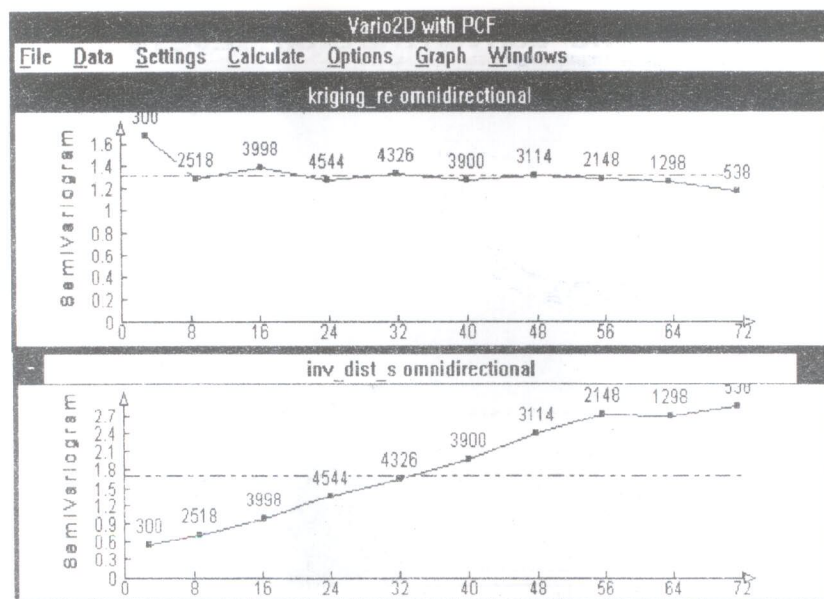


Рис. 39: Вариограммы разниц (residuals) кросс-валидации для метода обратных квадратов (нижняя) и для кригинга (верхняя).

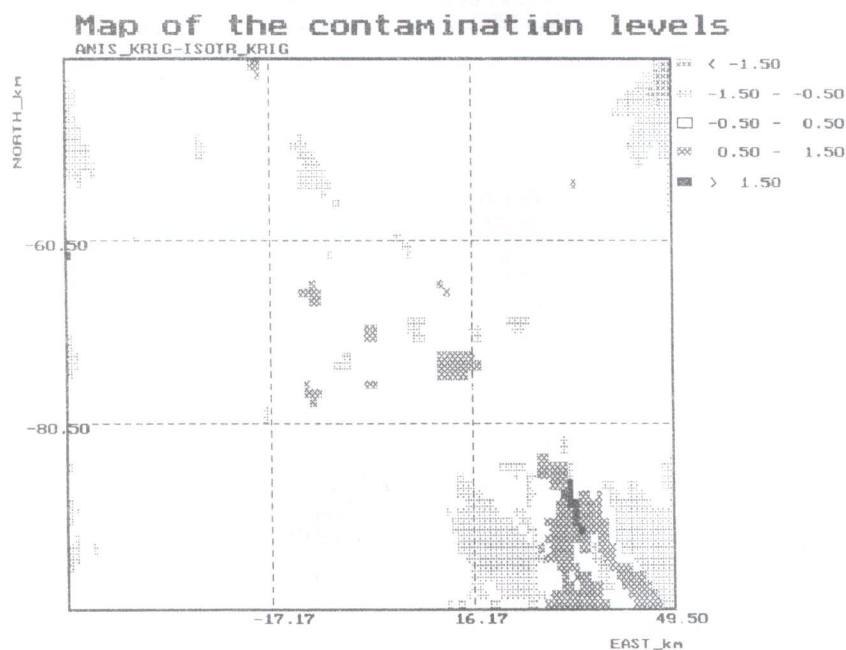


Рис. 40: Карта распределения разницы между оценками кригинга для анизотропной и для изотропной моделей.

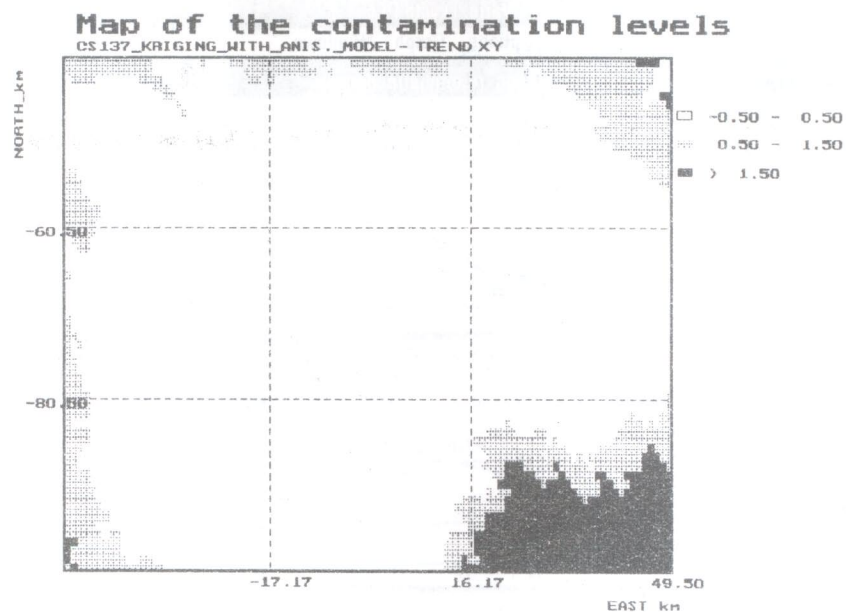


Рис. 41: Карта распределения разницы между оценками кригинга для анизотропной модели без учета тренда и с учетом линейного тренда в направлениях x и y .

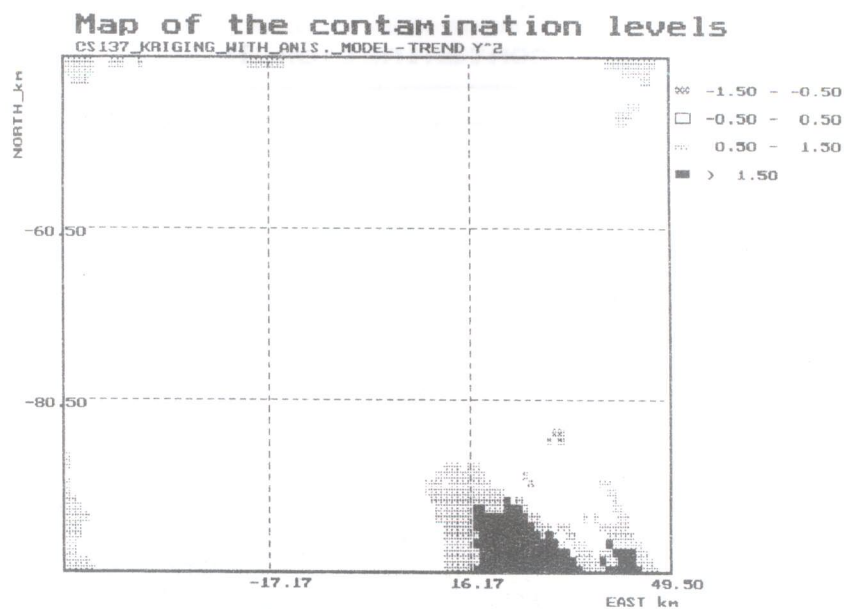


Рис. 42: Карта распределения разницы между оценками кригинга для анизотропной модели без учета тренда и с учетом квадратичного тренда в направлении y .

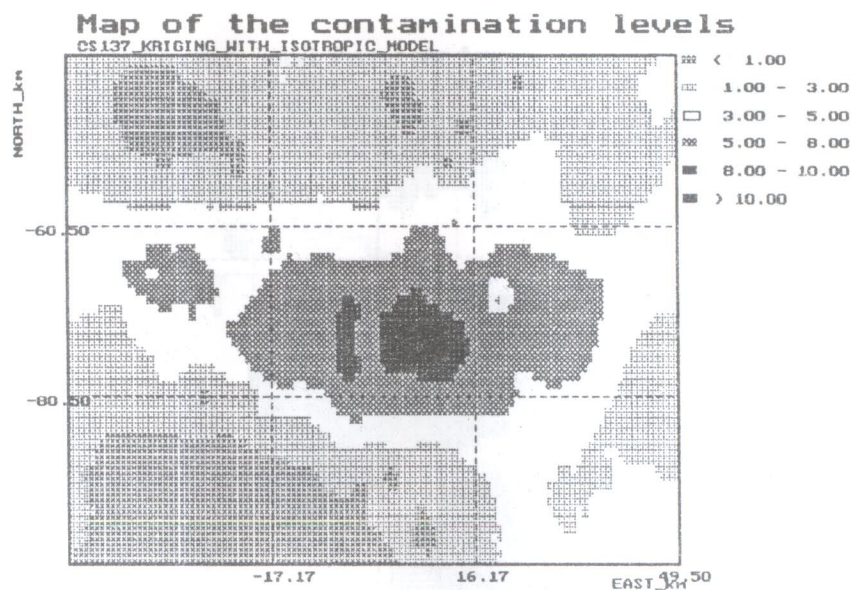


Рис. 43: Карта распределения оценок кригинга для изотропной модели.

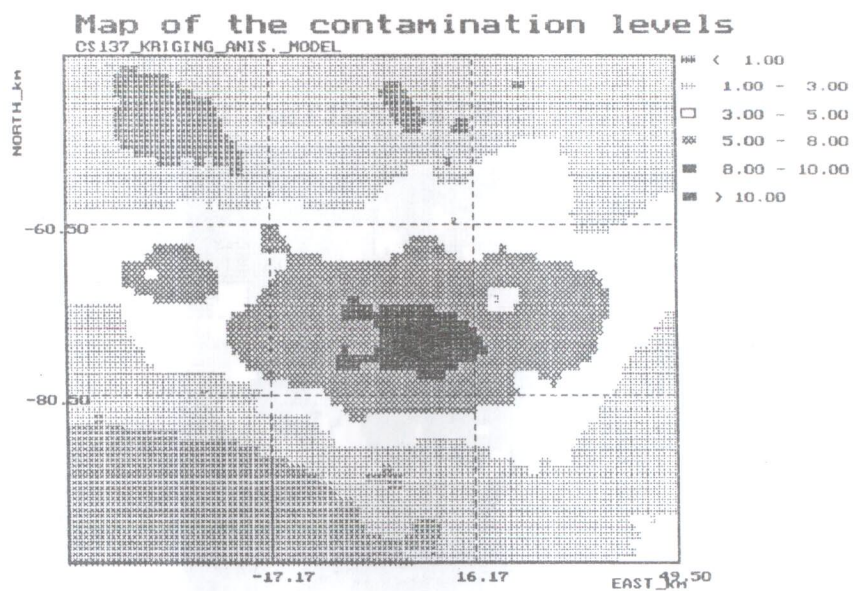


Рис. 44: Карта распределения оценок кригинга для анизотропной модели.

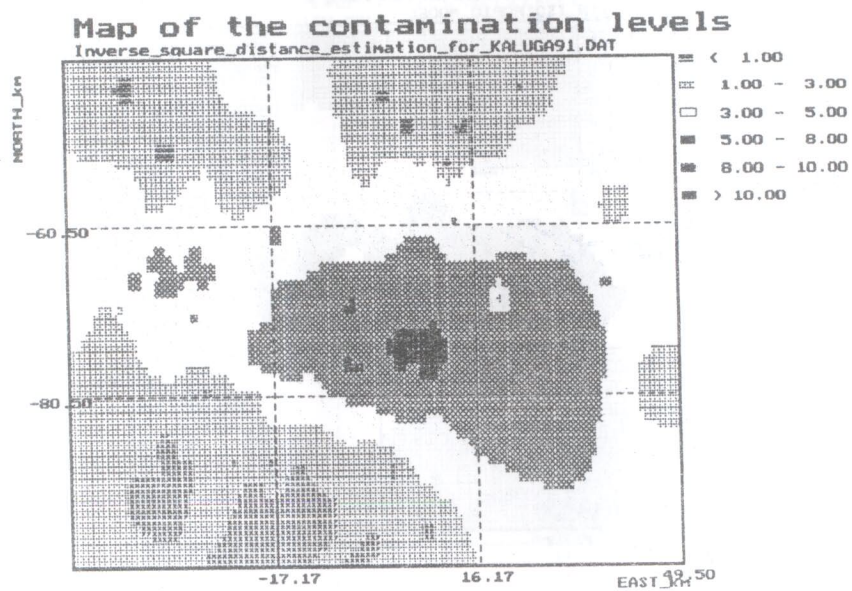


Рис. 45: Карта распределения оценок метода обратных квадратов.

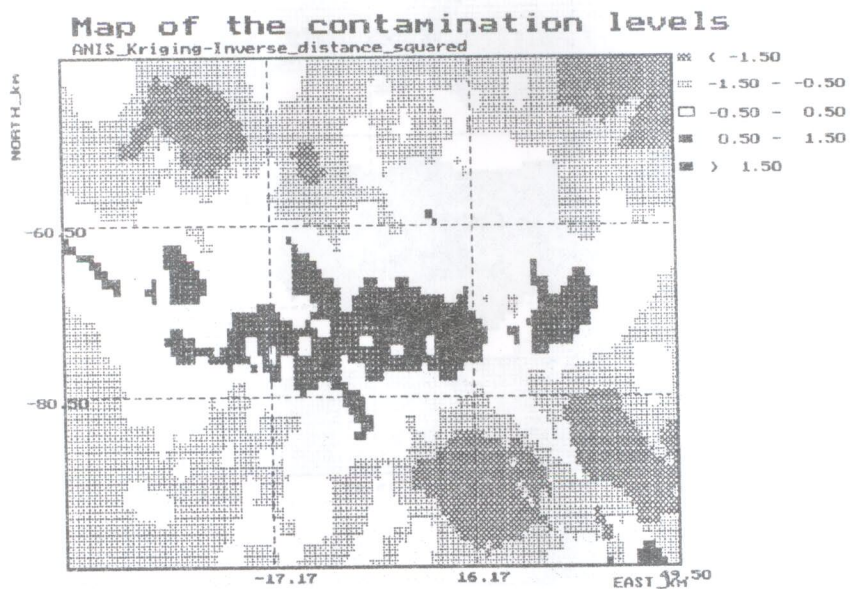


Рис. 46: Карта распределения разницы между оценками кригинга для анизотропной модели и метода обратных квадратов.

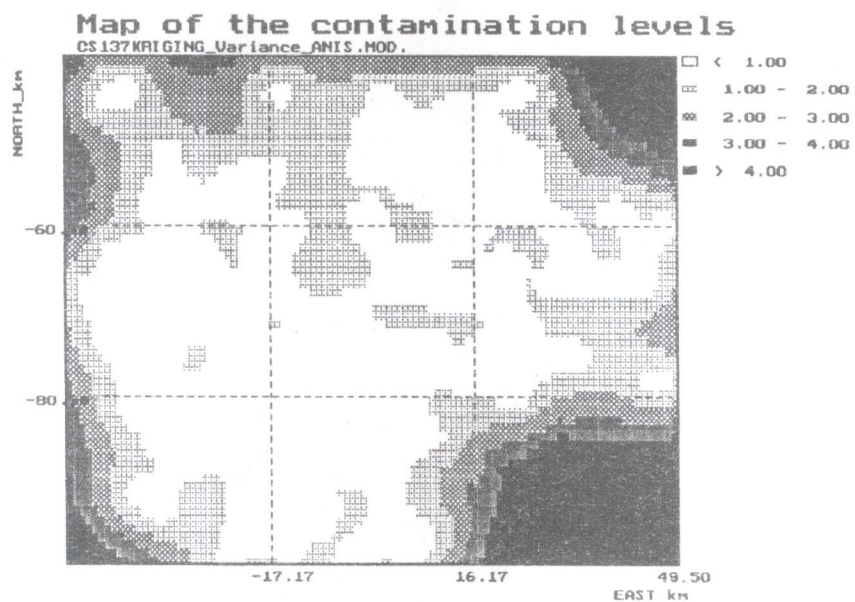


Рис. 47: Карта распределения вариации кригинга для анизотропной модели.

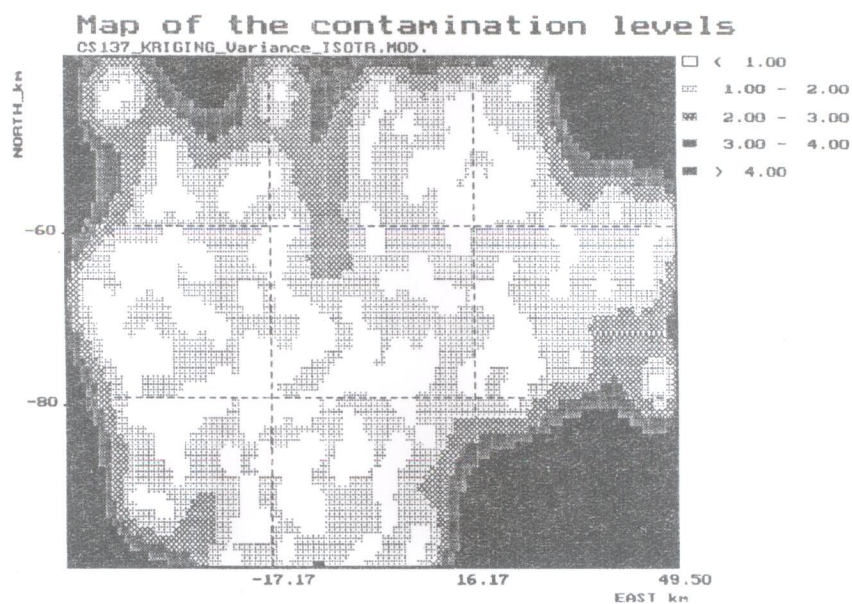


Рис. 48: Карта распределения вариации кригинга для изотропной модели.