

Глава 4. Вариография: исследование и моделирование пространственных корреляционных структур. Меры пространственной непрерывности.

В.В. Демьянов, М.Ф. Каневский, Е.А. Савельева, С.Ю. Чернов

Введение

Одним из важнейших свойств всех природных явлений является пространственная непрерывность, то есть взаимная зависимость (корреляция) для значений в точках расположенных ближе друг к другу больше, чем для разделенных большим расстоянием. Пространственная непрерывность пространственно распределенных данных обычно описывается с помощью корреляционных функций, ковариационных функций, полу-вариограмм и других мер пространственной корреляции. В геостатистике традиционно используются полу-вариограммы, что связано с простотой ее использования в геостатистических интерполяционных моделях. Из-за широкого использования полу-вариограмм для описания пространственной корреляционной структуры данных этот этап анализа принято называть «вариография».

Сущность проведения анализа пространственной корреляционной структуры данных состоит в выявлении наличия у данных корреляционной структуры и определении ее особенностей. Подробнее это означает, например, проверку данных на наличие/отсутствие пространственного тренда (видимой связи значения с реальным местоположением точки), который может быть описан некоторой математической функцией. Проверяется также зависимость корреляционной структуры от взаимной пространственной ориентации точек (наличие/отсутствие пространственной анизотропии). Определяется (если он существует) эффективный радиус корреляции данных — максимальным расстоянием, на котором еще наблюдается зависимость между значениями в различных точках. Конечной целью проведения пространственного корреляционного анализа является построение математической функции, описывающей пространственную корреляционную структуру данных и пригодную для последующего использования в геостатистических оценителях.

Анализ пространственной корреляционной структуры пространственно распределенных данных является начальным и ключевым моментом анализа данных. Качество полученной модели пространственной корреляционной структуры данных определяет качество последующей геостатистической оценки и величину ее ошибки. Полезным бывает предварительный корреляционный анализ и при последующем использовании детерминистических методов, так как позволяет обоснованно выбрать параметры метода. Для любого детерминистического метода разумно использовать параметры, связанные с эффективным радиусом корреляции

Основные меры описывающие пространственную корреляционную структуру пространственно распределенных данных

Для описания пространственной корреляции значений распределений используются различные моменты первого и второго порядка. Они характеризуют похожесть (или не похожесть) данных на расстоянии, тем

самым описывая пространственную непрерывность. Но для их корректного использования при наличии только одной реализации требуется сделать предположений о стационарности данных. Строгая стационарность пространственно распределенных данных означает, что любой набор из конечного числа n точек $x_i (i=1, \dots, n)$ и любого вектора h совместная многомерная функция распределения от $V(x_1), V(x_2), \dots, V(x_n)$ полностью совпадает с многомерной функцией распределения от $V(x_1+h), V(x_2+h), \dots, V(x_n+h)$. На практике такое предположение очень трудно проверить, и так как мы не интересуемся моментами выше второго порядка наличия стационарности второго порядка вполне достаточно. Стационарность второго порядка это стационарность только для моментов первого и второго порядка, то есть $E[V(x)] = m = const$ — среднее постоянно и $Cov(x_1, x_2) = E[V(x_1), V(x_2)] - m^2 = C(h)$ — ковариация существует и зависит только от вектора h .

Очень часто реальные данные не удовлетворяют даже условию стационарности второго порядка. Для использования геостатистических моделей оценивания более слабое условие является достаточным — наличие «внутренней гипотезы» (*Intrinsic Hypothesis*). Это условие можно сформулировать следующим образом: разность значений пространственно разделенных вектором h назовем приращениями, тогда «внутренняя гипотеза» означает стационарность второго порядка для приращений. Среднее для приращений постоянно и равно 0 ($E[V(x+h) - V(x)] = 0$) и вариация приращений (вариограмма) существует и зависит только от h и не зависит от x ($2\gamma(h) = Var[V(x+h) - V(x)]$).

Приняв предположение о стационарности в строгом или более мягком виде (стационарность второго порядка или внутренняя гипотеза), можно заключить, что функции корреляции между данными зависят от их взаимного местоположения, а не от конкретного места в пространстве. Это означает, что основополагающим параметром будет вектор h между двумя в точками в паре, между которыми анализируется корреляция. Для изотропного случая, когда корреляция не зависит от направления, h переходит в расстояние h .

Для проведения пространственного корреляционного анализа используются следующие формулы обеспечивающие различную степень точности:

1. Родограмма (rodogram):

$$R(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{|V(x) - V(x+h)|\}^2$$

2. Мадограмма (madogram) $M(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} |V(x) - V(x+h)|$

3. Ковариация (Covariance) – статистическая мера корреляции между двумя значениями $V(x_1)$ и $V(x_2)$ в точках x_1 и x_2 находящихся на расстоянии $h = x_1 - x_2$.

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

Для $N(h)$ экспериментальных точек находящихся на расстоянии h друг от друга:

$$C(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} V(x)V(x+h) - m_{-h}m_{+h}$$

где: m_{-h} среднее значение данных, находящихся в началах вектора h :

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

m_{+h} среднее значение данных, находящихся в концах вектора \mathbf{h} :

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

Ковариацией можно пользоваться только для данных со стационарностью второго порядка.

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

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

Для $N(h)$ экспериментальных точек находящихся на расстоянии h друг от друга:

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

Для существования вариограммы не обязательна стационарность второго порядка, достаточно наличия *внутренней гипотезы*.

Вариограмма обладает свойством симметрии: $\gamma(h) = \gamma(-h)$.

Вариограмма характеризует различность данных в зависимости от расстояния между ними. Чем больше расстояние, тем выше значение вариограммы. Ковариация характеризует обратное – похожесть данных. Чем больше расстояние между данными тем меньше значение ковариации. Таким образом, вариограмма и ковариация имеют следующую связь:

$$\begin{aligned} \gamma(0) &= C(\infty) \\ \gamma(\infty) &= C(0) \end{aligned}$$

где $C(0)$ – априорная ковариация или экспериментальная вариация $\text{Var}\{V(x)\}$ в точке x .

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

Дрейф (drift) очень важная функция при анализе корреляционной структуре данных $D(h) = E\{V(x) - V(x+h)\}$. Он определяется по следующей формуле:

$$D(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \{V(x) - V(x+h)\}$$

Дрейф может служить указателем подчинения данных внутренней гипотезе. Такой вывод можно сделать если значение $D(h)$ колеблется вокруг 0. Если же $D(h)$ растет (или убывает) с ростом h , то данные не подчиняются даже внутренней гипотезе и тем более, более строгому условию стационарности второго порядка. Это может в частности означать, что у данных имеется систематический тренд, то есть определенная зависимость значения исследуемой функции от пространственного местоположения (координаты). Для таких данных моделирование вариограммы и использование обычных геостатистических оценщиков приведет к необоснованным результатам. Для

таких данных используются специальные методы, о некоторых из которых будет рассказано в последующих главах.

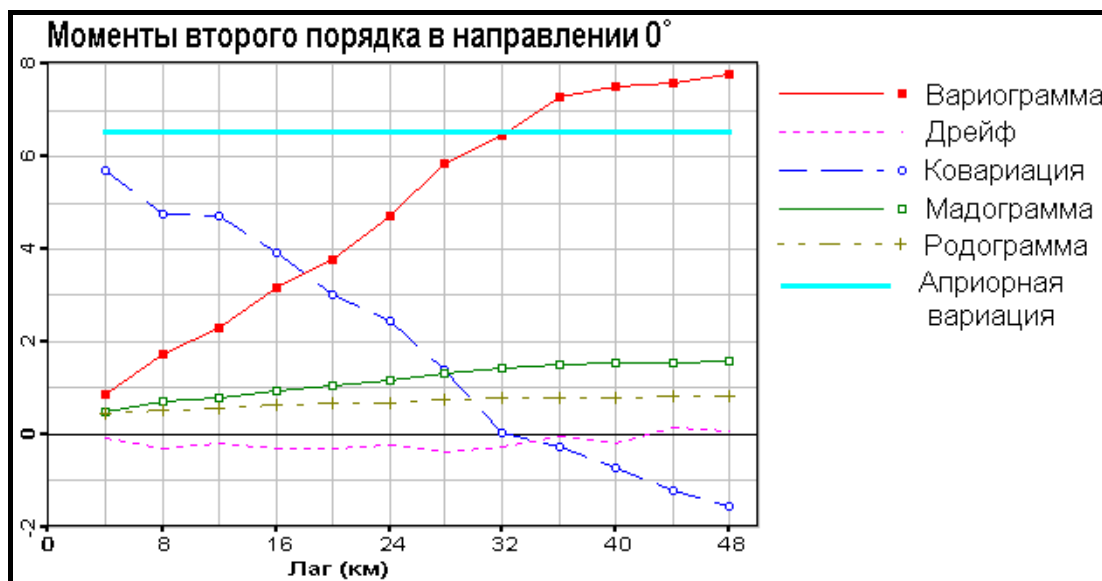


Рис.1 Пример характеристик пространственной корреляционной структуры данных

И для ковариации и для вариограммы существуют стандартизованные варианты, соответственно кореллограмма (correlogram) и стандартизованная вариограмма (standardized variogram). Эти оценки пространственной корреляционной структуры являются более робастными, то есть более устойчивыми к присутствию испорченных данных (например необоснованно высоких или низких значений). Они вычисляются по следующим формулам:

$$\text{Кореллограмма } p(h) = \frac{C(h)}{\sigma_{-h}\sigma_{+h}}$$

$$\text{Стандартизованная вариограмма } \gamma_i(h) = \frac{\gamma(h)}{\sigma_{-h}\sigma_{+h}}$$

где σ_{-h} и σ_{+h} стандартные отклонения для точек находящиеся соответственно в начале и конце вектора \mathbf{h} . Они находятся по формулам

$$\sigma_{-h} = \frac{1}{N(h)-1} \sum_{i=1}^{N(h)} (V(x_i) - m_{-h})^2$$

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

Очень часто в реальных данных наблюдается пропорциональный эффект, это наличие зависимости между средним ($m(h)$) и значением вариограммы ($\gamma(h)$). Пропорциональный эффект сильно усложняет понимание результата анализа пространственной корреляционной структуры, так как искажает результат. Обнаружить пропорциональный эффект можно с помощью вычисления локальных статистических характеристик — среднего и вариации. Для этого исходная исследуемая область разбивается на прямоугольные ячейки

одинакового размера. Размер ячеек определяется количеством точек измерений — требуется, чтобы в большинстве ячеек было достаточно точек измерений для получения обоснованной оценки статистических параметров (среднего и вариации). После проведения такого вычисления результаты представляются на графике, где по одной оси откладывается значение среднего, а по другой вариации. По такому графику легко увидеть наличие или отсутствие зависимости между средним и вариацией, что и определяет наличие или отсутствие пропорционального эффекта.

При наличии пропорционального эффекта предлагается использовать относительные вариограммы, которые уничтожают влияние пропорционального эффекта. Существует две глобальных относительных вариограммы:

Общая относительная вариограмма

$$\gamma_{GR}(h) = \frac{\gamma(h)}{\left(\frac{m_{-h} + m_{+h}}{2}\right)^2}$$

Парная относительная вариограмма

$$\gamma_{PR}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \frac{(V(x_i) - V(x_i + h))^2}{\left(\frac{V(x_i) + V(x_i + h)}{2}\right)^2}$$

В первом случае просто вариограмма нормируется на квадрат среднего для точек разделенных вектором \mathbf{h} . Во втором случае нормировка на квадрат среднего проводится для каждой конкретной пары отдельно.

При рассмотрении пространственных распределений двух случайных величин в одном пространстве используются “инструменты” аналогичные введенным выше. В этом случае к названию добавляется приставка “кросс-”, что означает выявление взаимной пространственной структуры распределений.

При рассмотрении взаимного влияния между двумя распределениями $V(x)$ и $U(x)$ можно ввести моменты второго порядка аналогично случаю с одной переменной.

Кросс-ковариация (Cross-covariance) для $N(h)$ экспериментальных точек находящихся на расстоянии h друг от друга, в которых есть измерения обеих переменных V и U :

$$C(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} V(x)U(x+h) - m_{U-h}m_{V+h}$$

где: m_{-h} среднее значение данных, находящихся в началах вектора \mathbf{h} :

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

m_{+h} среднее значение данных, находящихся в концах вектора \mathbf{h} :

$$m_{U+h} = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} U(x_i + h)$$

Кросс-ковариация не меняется при перемене ориентации пары на противоположную: $C_{uv}(h) = C_{vu}(-h)$. Однако кросс-ковариация зависима от порядка значения переменных в парах: $C_{uv}(h) \neq C_{vu}(-h)$

Кросс-вариограмма (Cross-variogram)

$$\gamma_{uv}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (V(x_i) - V(x_i + h)) \cdot (U(x_i) - U(x_i + h))$$

Кросс-вариограмма обладает симметрией как и обычная вариограмма.

$$\gamma_{uv}(h) = \gamma_{uv}(-h) = \gamma_{vu}(h) = \gamma_{vu}(-h)$$

Особенности вычисления мер пространственной корреляции пространственно распределенных данных на практике

Описанные выше моменты первого и второго порядка служат для анализа пространственной корреляции данных. Для выявления пространственной структуры используются несколько различных инструментов: вариограммы по направлениям, вариограммные поверхности, вариограммные облака. При их построении главный вопрос заключается в выборе точек и образовании пар для вычисления. Согласно пространственной непрерывности два более удаленных друг от друга значения будут менее коррелированы, чем два значения находящихся на близком расстоянии.

Значения экспериментальной вариограммы рассчитываются в зависимости от расстояния между точками в паре (лага) h . Лаг (lag) выбирается в соответствии с размером пространственной корреляции. При таком подходе значение вариограммы не зависит от ориентации пары в пространстве, что означает изотропность структуры. Построенная таким образом вариограмма по всем направлениям (omnidirectional) зависит от всех пар точек в области. Для выявления различий в пространственной структуре в зависимости от направления используются экспериментальные вариограммы в различных направлениях. При этом пары выбираются из соответствующего сектора с заданным направлением, углом раствора и (по усмотрению) шириной полосы, которая сужает область поиска на больших расстояниях. Схема выбора параметров для вариограмм по направлениям представлена на рисунке 2. Для выявления анизотропии можно построить вариограммы в 3-4 направлениях. Когда угол раствора равен 90° вариограмма превращается в омнинаправленную.

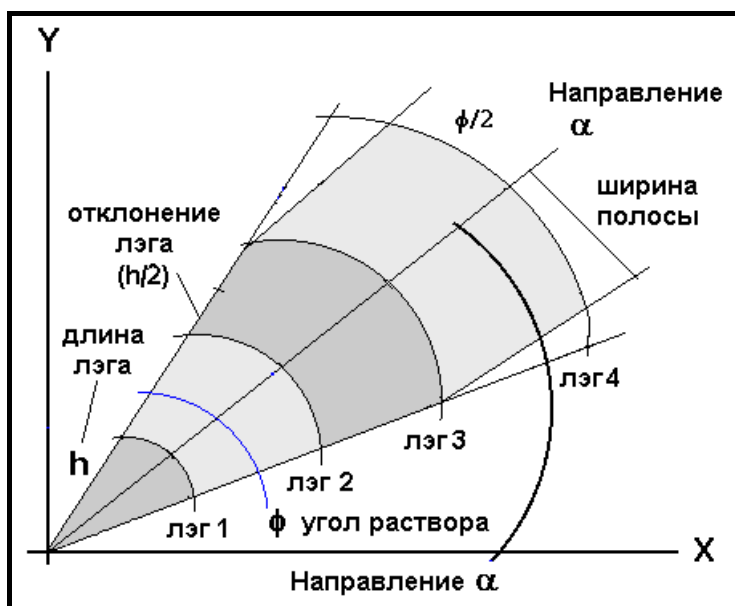


Рисунок 2. Параметры для расчета вариограммы.

Вариограмная роза дает возможность исследовать анизотропию корреляционной структуры. Роза имеет вид лепестков, представляющих вариограммы по направлениям. Вариограмная роза симметрична относительно центра в силу свойств симметрии моментов второго порядка.

Другим инструментом является *вариограмная поверхность*, которая дает представление о поведении пространственной структуры в целом. Вариограмная поверхность это поверхность, на которой нанесены абсолютные значения вариограммы в зависимости от разницы координат Δx и Δy . Вариограмная поверхность позволяет сразу увидеть анизотропию и определить приоритетные направления для построения вариограмм. Значения вариограммы вычисляются в каждой ячейке для соответствующего вектора h . Следует отметить, что вариограмная поверхность обладает точечной симметрией относительно точки $(0,0)$. Это естественно, поскольку каждая пара может браться в двух противоположных направлениях.

Еще одним инструментом пространственного корреляционного анализа 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 лаг для вычисления вариограммы. Вариограммное облако может быть построено для любого направления и раствора угла. Следует отметить, что вариограммное облако обладает точечной симметрией относительно точки $(0,0)$. Это естественно, поскольку каждая пара может браться в двух противоположных направлениях.

Частым явлением при анализе радиологических данных является пространственный *тренд*. Это устойчивая зависимость значений от направления. Тренд может быть связан с различными физическими процессами и условиями при образовании данных, а так же с накладыванием одного

процесса на другой. Визуально можно продемонстрировать тренд на диаграмме разброса значений в зависимости от координат в отдельном направлении. Корреляция между ними означает пространственный тренд в этом направлении. Пространственный тренд может быть обнаружен с помощью вариограммы по направлению. В случае если она имеет значения выше априорной ковариации, это может означать наличие тренда.

Модели вариограмм

Построенные экспериментальные вариограммы, отражающие пространственную структуру моделируются с помощью их теоретических моделей. Вариограмма должна удовлетворять ряду строгих теоретических результатов: положительная определенность, поведение на бесконечности и в нуле. В практическом использовании теоретические модели и их комбинации должны удовлетворять этим требованиям для описания экспериментальных вариограмм. Экспериментальная вариограмма является лишь набором значений для определенных лагов h , тогда как модель вариограммы позволяет получить ее значение для любого расстояния между точками. Это будет необходимо при последующем оценивании на сетке.

Основные типы моделей вариограмм называются соответственно функциям, которые их моделируют. Они представлены графически на рисунке 3. Модели бывают следующих типов:

Самородок (Nugget)

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c, & h \neq 0 \end{cases}$$

Константа $c=C(0)$ носит названия *плато* (sill). Он присутствует и в других стационарных моделях вариограмм. Наличие плато у вариограммы означает стационарность второго порядка. Наличие у данных такой вариограммы означает отсутствие пространственной корреляции. Данные в этом случае распределены абсолютно случайно (pure nugget).

Сферическая

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

где a – действительный *радиус корреляции* (range). На расстоянии a корреляция существует. Для сферической модели $\gamma(a)=C(0)=c_0+c$ – *плато* (sill). Сферическая модель ведет себя линейно вблизи нуля.

Гауссова

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

где a – *эффективный радиус корреляции* (range). На этом расстоянии значение вариограммы достигает 95% плато. Отличительной чертой Гауссовой модели является параболическое поведение вблизи нуля. Гауссова модель достигает

плато асимптотически. Пространственная корреляция промоделированная Гауссовой моделью имеет конечное влияние на бесконечном расстоянии. Это означает, что данные остаются коррелированными на любом расстоянии, но в бесконечности корреляция стремится к нулю.

Экспоненциальная

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

где a – эффективный радиус корреляции (*range*). На этом расстоянии значение вариограммы достигает 95% плато. Экспоненциальная модель достигает плато асимптотически.

Во всех моделях c – плато (*sill*) – может быть только положительным.

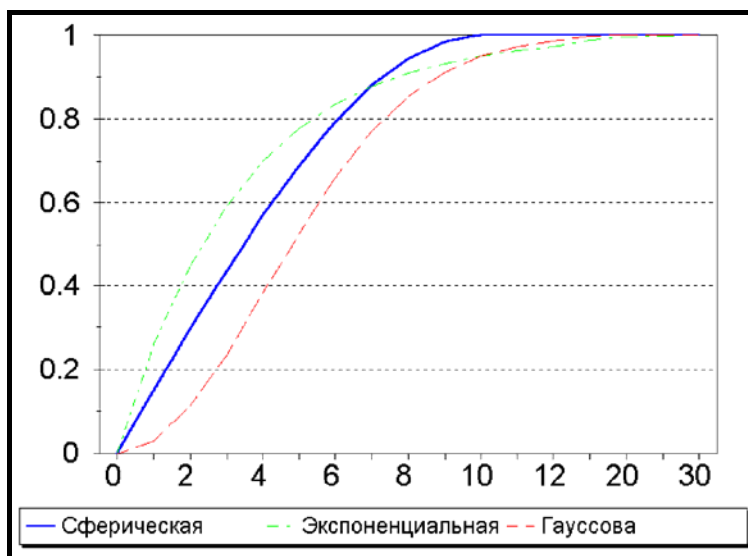


Рисунок 3. Типы моделей вариограмм при $a=10$, $c=1$.

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

При непосредственном моделировании (подборе формы и параметров модели) 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 набор специальных индикаторов подгонки.

$$\text{Индикатор Кресси } I_C = \sum \frac{N(h_i)}{[\gamma^*(h_i, \lambda)]^2} [\gamma^*(h_i, \lambda) - \gamma(h_i)]^2.$$

Индикатор качества подгонки

$$S_0 = \sum (Z(x_i, y_i) - m)^2$$

$$\text{Информационный критерий } I_{AIK} = n \ln \left(\frac{\sum_{i=1}^n (\gamma(h_i) - \gamma^*(h_i, \lambda))^2}{n} \right) + 2p.$$

Здесь $\gamma(h_i)$ — значение экспериментальной вариограммы для i ого лага, $\gamma^*(h_i, \lambda)$ — значение модели вариограммы для лага h_i и набора параметров модели λ , $N(h_i)$ — число пар по которым вычислялось значение экспериментальной вариограммы для лага h_i ,

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 гауссового типов. Модель вариограммы сферического типа достигает плато на расстоянии a ; экспоненциального типа – на расстоянии $3a$; а гауссова модель достигает 95% значения плато 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, данные, далеко удаленные друг от друга, продолжают оказывать взаимное влияние.

Поведение вариограмм в нуле

Вариограммы также различаются и 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 $h > \varepsilon$ — $\gamma(h)=C(0)$. То есть 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]:

1. *геометрическая* (geometric) – при которой полувариограммы (ковариации) по различным направлениям имеют одинаковую формы и плато, но разные радиусы, и изолинии вариограммы на диаграмме имеют форму эллипсов. В этом случае существует положительно определенная матрица \mathbf{B} , такая что, $\gamma(\mathbf{h}') = \gamma(\mathbf{h}^T \mathbf{B} \mathbf{h})$ изотропная вариограмма. Преобразование пространства такое, что геометрически анизотропная вариограмма становится в преобразованном пространстве изотропной определяют следующим образом:

$$\mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \mathbf{h},$$

где a_{max} — длина основной оси эллипса, a_{min} — длина малой оси эллипса, θ — угол, определяющий направление основной оси эллипса (измеряется от направления запад-восток против часовой стрелки). Пример графиков, иллюстрирующих геометрическую анизотропию представлен на рисунке 4.

2. *зонная* – это анизотропия, которая не является геометрической. Например, это может быть случай, когда по различным направлениям различные радиусы, но изолинии вариограммы не образуют эллипс, или когда для различных направлений различаются не только значения радиусов, но и плато. Из-за того, что существуют различные варианты зонной анизотропии удобно провести дальнейшую классификацию.

Случай зонной анизотропии, когда для различных направлений различаются только значения радиуса будем называть *негеометрической анизотропией радиуса* (non-geometric range anisotropy) — пример графиков, иллюстрирующих такой тип анизотропии представлен на рисунке 5. В таком случае для получения вариограммы, которую можно использовать для всех направлений тоже достаточно сделать преобразование пространства:

$$\mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \mathbf{h},$$

где θ — угол, определяющий направление с максимальным значением радиуса (измеряется от направления запад-восток против часовой стрелки).

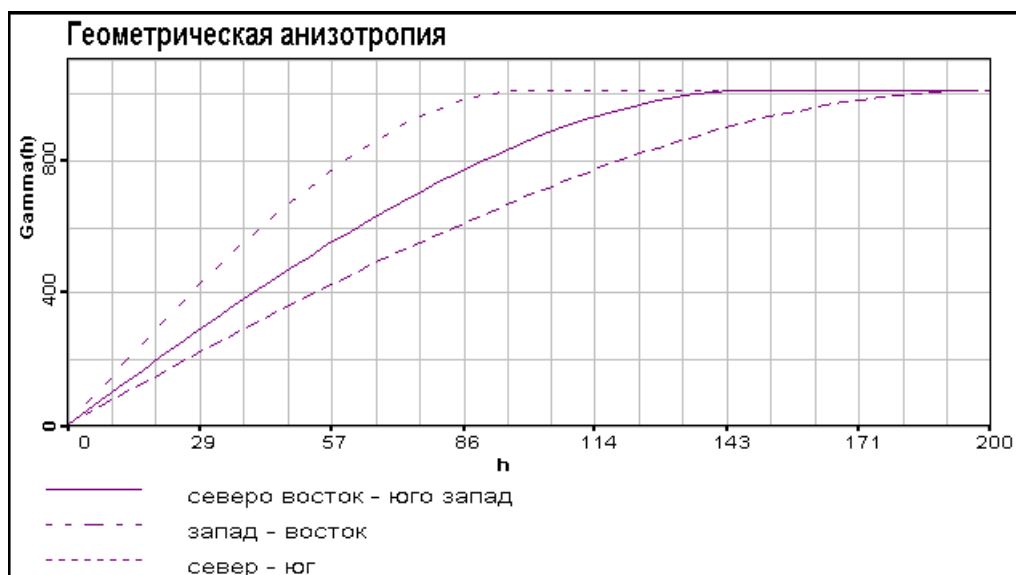


Рисунок 4 Пример модели геометрической анизотропии

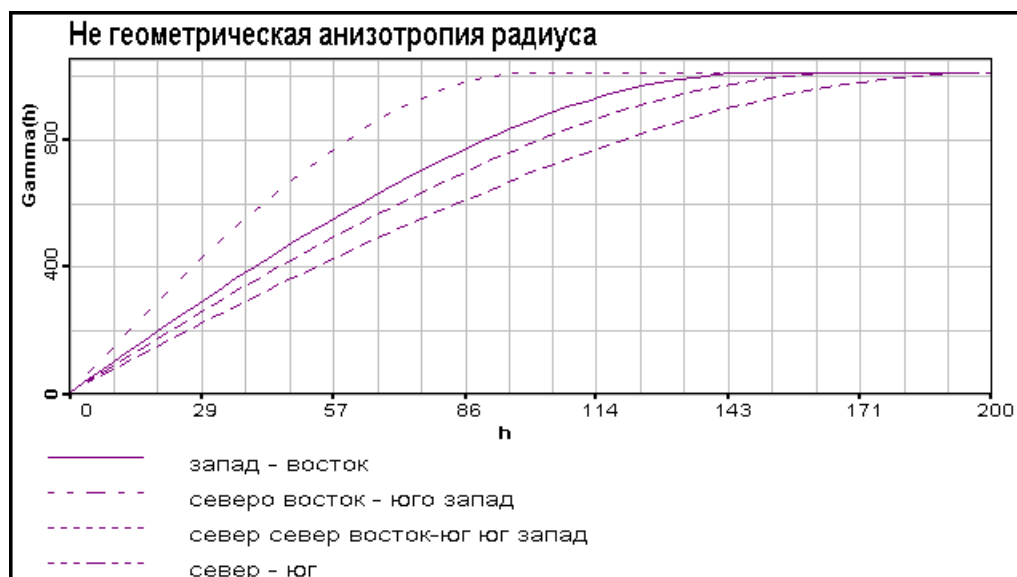


Рисунок 5. Пример негеометрической анизотропии радиуса

Рассмотрим теперь случай зонной анизотропии, когда для различных направлений различаются значения плато (рис. 6) будем называть *анизотропией плато* (sill anisotropy). Наличие плато у вариограммы процесса означает, что процесс не только удовлетворяет внутренней гипотезе, но и стационарности второго порядка. Тогда учитывая поведение вариограммы на больших расстояниях можно написать, что

$$\lim_{\alpha \rightarrow \infty} \gamma(\alpha h) = C(0) - \lim_{\alpha \rightarrow \infty} C(\alpha h)$$

для любого фиксированного h . Очевидно, что такое равенство можно было бы написать для любого направления, где для модели вариограммы обнаружено плато. А так как мы имеем различные значения плато для

различных направлений, следовательно существуют хотя бы два направления (h_1 и h_2) такие, что $\lim_{\alpha \rightarrow \infty} C(\alpha h_1) \neq \lim_{\alpha \rightarrow \infty} C(\alpha h_2)$. А это означает, что по крайней мере в одном направлении $\lim_{\alpha \rightarrow \infty} C(\alpha h) \neq 0$. То есть если плато меняется в зависимости от направления, значит существует хотя бы одно направление, на котором корреляция между значениями не пропадает ни при каких расстояниях.

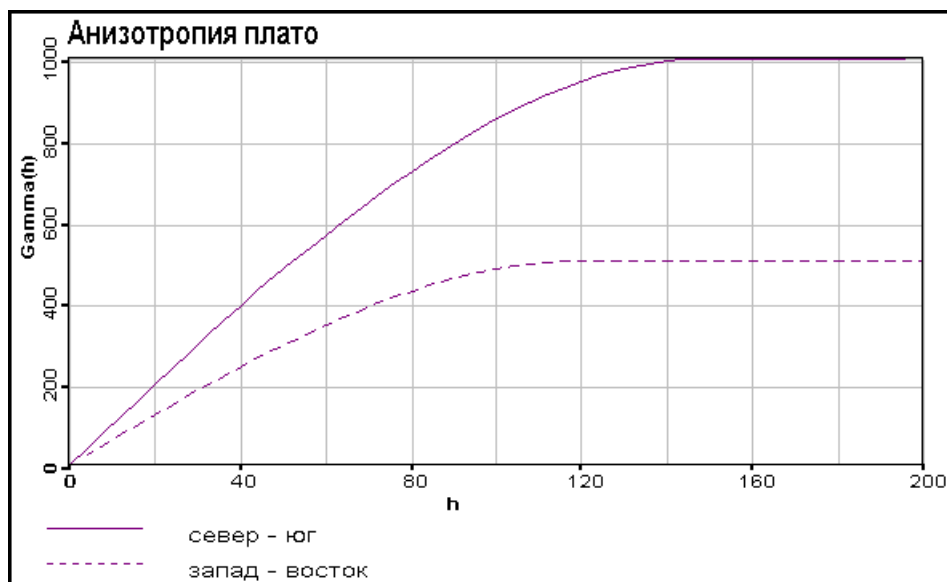


Рисунок 6. Пример анизотропии плато

При обнаружении анизотропии плато можно сделать два предположения о характере процесса: стационарность второго порядка присутствует, но корреляция между значениями не успевает исчезнуть для расстояний на которых проводилось построение модели; стационарности нет из-за присутствия тренда. Хотя второй вариант является более правдоподобным, на практике чаще всего делается предположение о выполнении первого варианта. В этом случае для получения изотропной вариограммы делается преобразование:

$$h' = \begin{bmatrix} 1 & 0 \\ 0 & a_{\max}/a_{\min} \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} h,$$

где a_{\max} задается такое значение, чтобы $a_{\max}/a_{\min} \approx 1000$.

Последнее, что могло бы меняться в зависимости от направления это «самородок», такую анизотропию назовем *анизотропией «самородка»* (nugget anisotropy). Это означает различное поведение вблизи 0. Обычно наличие такого типа анизотропии вызывается коррелированностью неточностей измерений. На практике такая анизотропия не моделируется, подбираются максимально подходящие модели (по индикаторам подгонки), имеющая одинаковое значение «самородка» по всем направлениям.

В случае использования гнездовой структуры для моделирования анизотропной вариограммы необходимо строить ее так, чтобы все i ые элементы (для всех направлений) имели одинаковую модель (сферическую, гауссову и

т.п.) и одинаковое значение плато, но радиусы могут быть любые, в том числе и такие большие, чтобы скрывать анизотропию плато.. В гнездовых структурах по всем направлениям должно быть одинаковое число элементов. Преобразования пространства делаются отдельно для каждого элемента, а в конечном результате опять получается линейная комбинация моделей.

Пример структурного анализа

Проведен структурный анализ загрязнения поверхности Калужской области *CS137*. На рисунке 7 представлено вариограммное облако. Оно показывает, что в данных по загрязнению поверхности Калужской области *CS137* нет ярко выраженных оутлайеров, то есть точек с необоснованно высокими (или низкими) значениями. Такой вывод можно сделать исходя из того, что ширина вариограммного облака растет с ростом расстояния, а значит для близких точек разница значений меньше, чем для более удаленных. Присутствие на вариограммном облаке точек с большим значением квадрата разности значений должно вызывать опасения при построении функций пространственной корреляции, так как они могут вызвать серьезные искажения структуры.

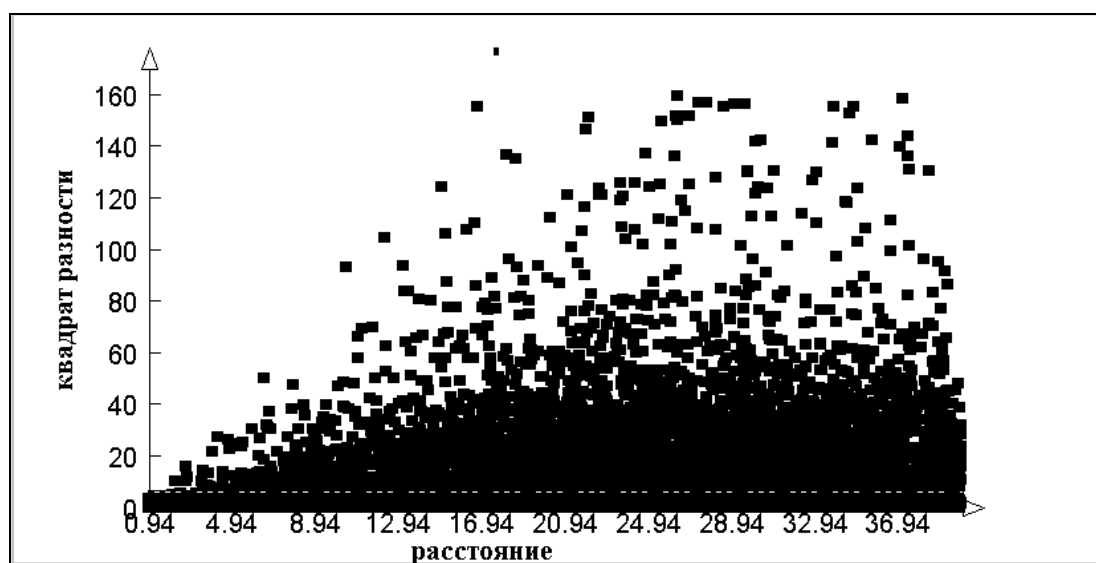


Рисунок 7. Вариограммное облако для *CS137*

Следующим, на что надо обратить внимание является проблема присутствия тренда и анизотропии. Для анализа тренда используется дрейф. На рисунке 8 представлен дрейф, посчитанный для различных направлений (углы указаны в градусах от направления запад - восток против часовой стрелки). По этому рисунку, а также по розе дрейфа (рис. 9 справа) видно, что поведение дрейфа не есть колебание вокруг 0, и оно различается в направлениях с северо-востока на юго-запад и с северо-запада на юго-восток. Это означает, что в данных присутствует тренд, но судя по значениям модуля дрейфа — незначительный. Лучше использовать интерполяционные модели учитывающие тренд, но в данном случае могут сработать и обычные геостатистические методы.

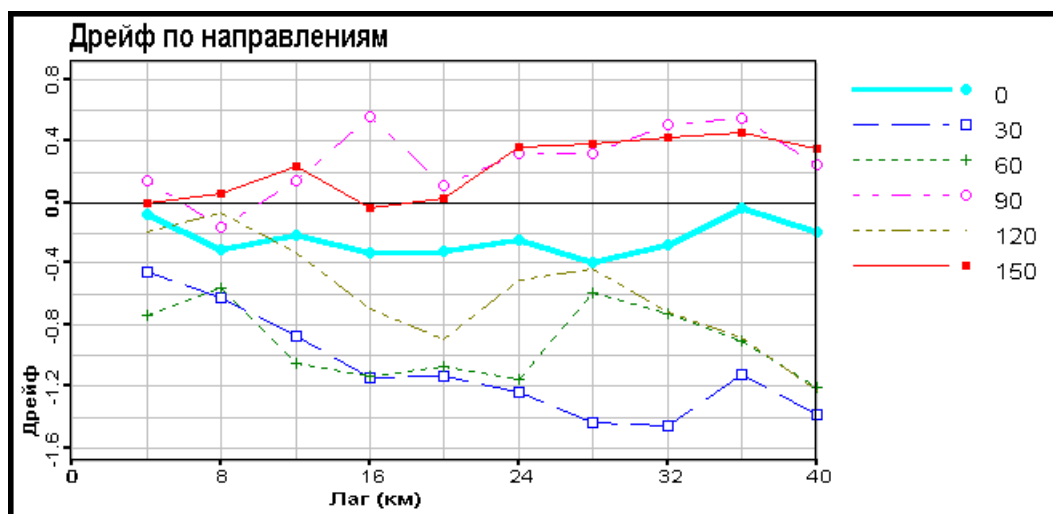


Рисунок 8. Дрейф по направлениям для CS137.

Вариограммная роза (рис. 9 слева) и вариограммная поверхность (рис. 10) демонстрируют анизотропную пространственную структуру. В данном случае, что бывает далеко не всегда довольно четко прочертился эллипс на вариограммной поверхности и вариограммной розе, что дает возможность рассматривать анизотропию, как геометрическую. Основная ось расположена вдоль направления запад-восток, малая соответственно — север-юг.

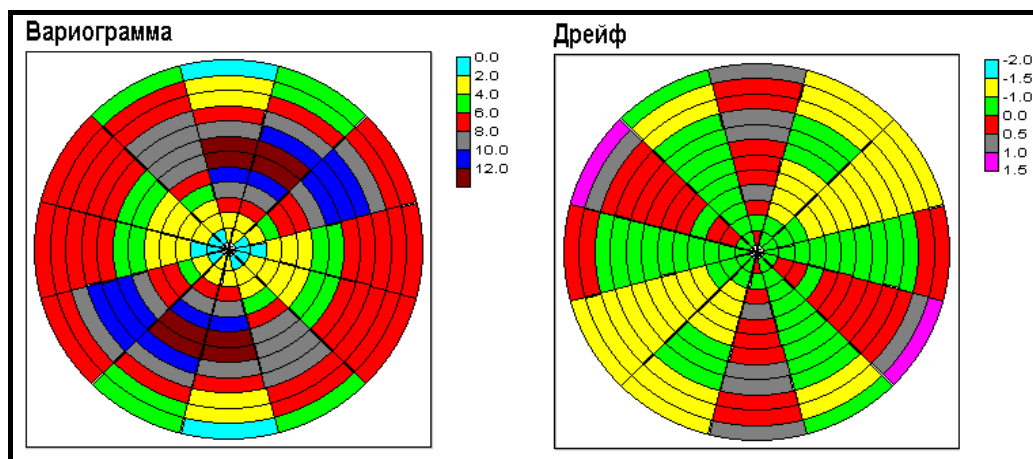


Рисунок 9. Вариограммная роза (слева) и роза дрейфа (справа) для CS137.

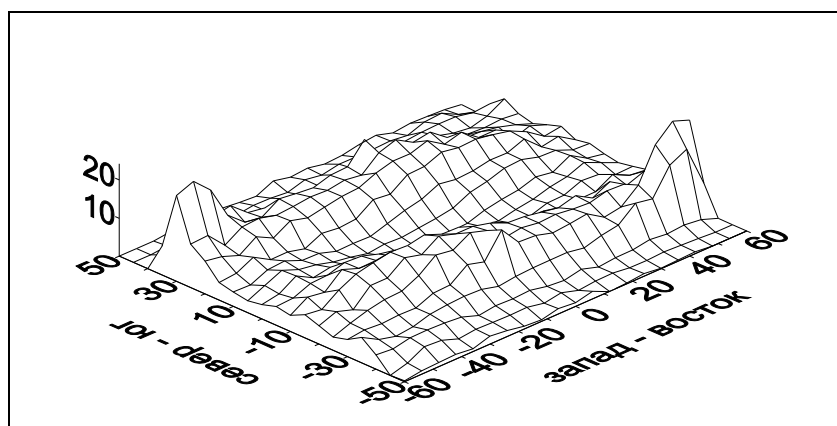


Рисунок 10. Вариограмная поверхность для Cs137

Следующим шагом является моделирование пространственной корреляционной структуры, то есть подбор математической модели, удовлетворяющей всем свойствам вариограммы и позволяющая описать ее для любого лага и направления (экспериментальную вариограмму мы оцениваем для конечного набора лагов и направлений). Вариограмма по всем направлениям (omnidirectional), представленная на рисунке 11 не позволяет провести моделирование из-за влияния анизотропии и трендов в различных направлениях.

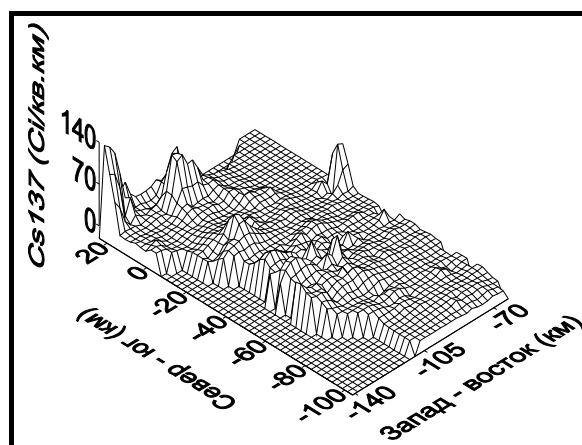


Рисунок 11. Вариограмма по всем направлениям для Cs137.

Из вариограмм по направлениям наиболее стационарными и подходящими для моделирования являются вариограммы в направлениях 0 и 150 градусов. Они отражают устойчивые корреляционные структуры данных в этих направлениях. Дрейф в этих направлениях не существен (см. рис. 8), что позволяет применить модель кригинга с вариограммами.

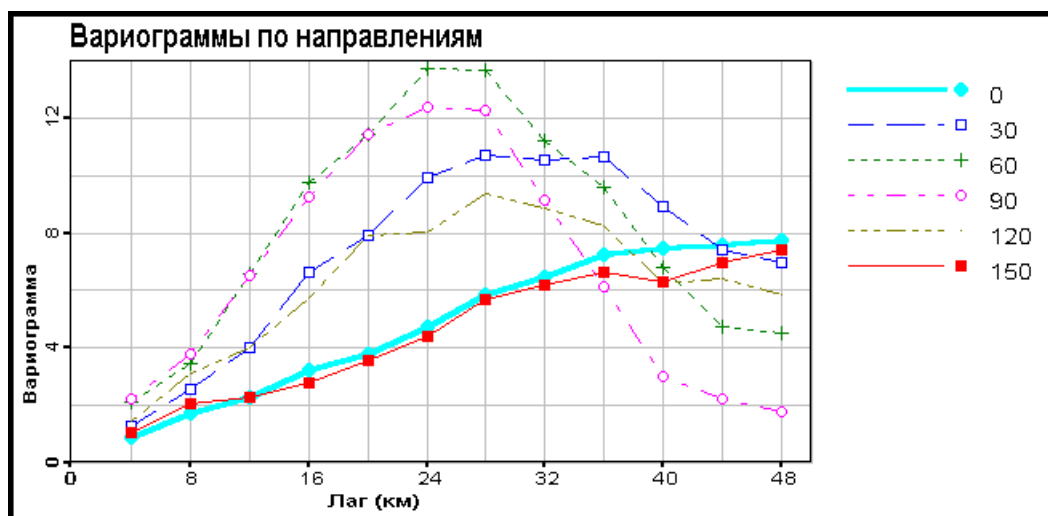


Рисунок 12. Полувариограммы по направлениям для CS137.

Параметры полученных теоретических моделей вариограмм приведены в таблице 1. Выбранная Сферическая анизотропная модель хорошо аппроксимируют точки экспериментальных вариограмм на рисунке 13. Для анизотропной модели был получен радиус корреляции около 40 км, что соответствует примерно половине размера области.

Таблица 1. Параметры моделей вариограмм.

Переменная	Нагге	Модель	Направление	Плато	Радиус	Анизотропия
CS137	0.0	Сферическая	по всем направлениям	6.595	23.52	1.0
CS137	0.0	Сферическая	0°	6.499	42.47	0.63

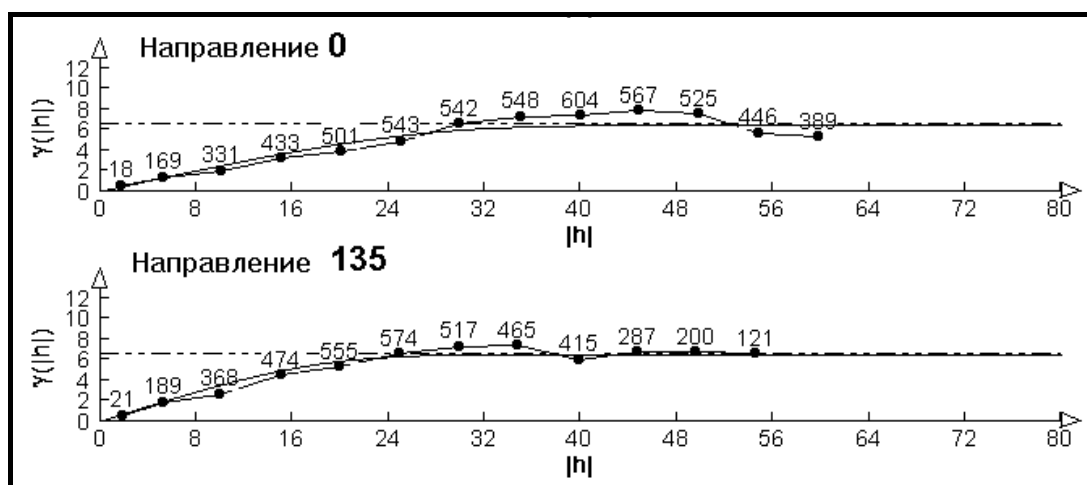


Рисунок 13. Анизотропная модель вариограмм для CS137.

Литература

1. Isaaks E.H., Shrivastava R.M. An Introduction to Applied Geostatistics. Oxford University press, Oxford, 1989.
2. Goovaerts P. Geostatistics for Natural Resources Evaluation. Oxford University Press, 1997.
3. Cressie N. Statistics for spatial data. John Wiley & Sons, New-York, 900p., 1991.
4. Chernov S., V. Demyanov, M. Kanevski, E. Savelieva. VarRose- a Way of Variogram Analysis. Preprint IBRAE-98-03. Mocsow 1998, 27 p.
5. Flamm C., Kanevsky M., Savelieva E. Non-regular variography and multi-method mapping to determination of origin of heavy metals. In 1994 International Association for Mathematical geology Annual Conference. Papers and Extended Abstracts.
6. Pannatier, Y. VARIOWIN Software for Spatial Data Analysis. Springer-Verlag, 1996, New-York.
7. M.David. Handbook of Applied Advanced Geostatistical Ore Reserve Estimation. Elsevier Applied Science Publishers, Amsterdam B.V., 1988.
8. I. Clark. Practical Geostatistics. Elsevier Applied Science Publishers, London and New York, repr. 1984.
9. D. L. Zimmerman. Another Look at Anisotropy in Geostatistics.
10. Mathematical Geology, vol.~25 no.~4, 1993.
11. Dowd P.A., Generalised cross-covariances. Geostatistics, vol. 1, Kluwer Academic Publisher, Dordrecht, 1989, pp. 579-590.
12. Myers D.E. Pseudo-Cross Variograms, Positive-Deginiteness and Cokriging. Mathematical Geology, V.23, No.6, 1991, pp.805-816.
13. A.Papritz, H.R.Kunsch, R.Webster. On the Pseudo Cross Variogram. Mathematical Geology, 1993, V.25, N.8, pp.1015-1026.
14. Armstrong M. Common Problems Seen in Variograms. Mathematical Geology, V.16, No.3, 1984, pp.305-313.
15. Barnes R.J. The Variogram Sill and the Sample Variance. Mathematical Geology, V.23., No.4, 1991, pp.673-678.
16. G.Christakos. On the problem of permissible covariance and semivariogram models. Water Resources Research, 1984, V.20(2), p.251-265.
17. J.Xiaodong, R.A.Olea, Y.-S.Yu. Semivariogram modeling by weighted least squares. Computers & Geosciences, 1996, V.22,N.4,pp.387-397.
18. N.Cressie. Fitting models by weighted least squares. Mathematical Geology, 1985, V.17, N.5, p.563-586.
19. X.F.Zhang, J.C.H.Van Eijkeren, A.W.Heemink. On the weighted least-square method for fitting a semivariogram model. Computers & Geosciences, 1995, V.21, N.4, pp.605-608.
20. K.Levenberg. A method for the solution of certain non-linear problems in least squares. Quart.Applied Mathematics, 1944, V.11, N.2,pp.164-168.
21. R.P.Brent Algorithms for Minimization without Derivatives. 1973, Englewood Cliffs, NJ:Prentice Hall Inc.
22. T.C.Haas. Kriging and automated variogram modeling within a moving window. Atmospheric Environment, 1990, V.24A, p.1759-1769.