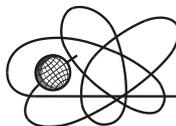




Российская Академия Наук

РОССИЙСКАЯ АКАДЕМИЯ НАУК

**ИНСТИТУТ ПРОБЛЕМ
БЕЗОПАСНОГО РАЗВИТИЯ
АТОМНОЙ ЭНЕРГЕТИКИ**



ИБРАЭ

RUSSIAN ACADEMY OF SCIENCES

**NUCLEAR SAFETY
INSTITUTE**

Препринт ИБРАЭ № IBRAE-2000-05

Preprint IBRAE- 2000-05

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

**СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
ПРОСТРАНСТВЕННО РАСПРЕДЕЛЕННЫХ
ДАННЫХ ПО ОКРУЖАЮЩЕЙ СРЕДЕ**

Москва 2000

Moscow 2000

УДК 502.3

Нехороших Д.С., Демьянов В.В., Каневский М.Ф., Чернов С.Ю., Савельева Е.А.
СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОСТРАНСТВЕННО РАСПРЕДЕЛЕННЫХ
ДАННЫХ ПО ОКРУЖАЮЩЕЙ СРЕДЕ. Препринт № IBRAE-2000-05. Москва: Институт
проблем безопасного развития атомной энергетики РАН. Апрель 2000. 28 с. Библиогр.:
14 назв.

Аннотация

В работе исследуются два метода стохастического моделирования (последовательное Гауссово моделирование и моделирование отжигом) в применении к пространственно распределенным данным по окружающей среде. Приведена теория этих методов и их применение к конкретным данным. Построены карты загрязнений и вероятностные карты — вероятности превышения заданного уровня. Для исследования использовались реальные данные по загрязнению Женевского озера.

©ИБРАЭ РАН, 2000

Nekhoroshih D.S., Demyanov V.V., Kanevski M.F., Chernov S.Y., Savelieva E.A.
STOCHASTIC SIMULATIONS OF SPATIALLY DISTRIBUTED DATA ON ENVIRONMENT
(in Russian). Preprint IBRAE-2000-05. Moscow: Nuclear Safety Institute. April 2000. 28 p. —
Refs.: 14 items.

Abstract

In this work two methods of stochastic simulations (sequential Gaussian simulations and simulated annealing) as applied to spatially distributed data on environment are analyzed. The theory of this methods and its application to the concrete data are adduced. The results are presented as maps of contamination and probability maps. The real data of water pollution of Geneva lake (Lac Lemman) are used for this study.

©Nuclear safety institute, 2000

Стохастическое моделирование пространственно распределенных данных по окружающей среде

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

ИНСТИТУТ ПРОБЛЕМ БЕЗОПАСНОГО РАЗВИТИЯ АТОМНОЙ ЭНЕРГЕТИКИ
113191, Москва, ул. Б. Тульская, 52
тел.: (095) 955-22-31, факс: (095) 958-11-51, e-mail: dasha@ibrae.ac.ru;
<http://www.ibrae.ru/~mkanev/>

Содержание

1	Введение.....	3
2	Стохастическое моделирование.....	4
2.1	Последовательное гауссово моделирование.....	5
2.2	Моделирование отжигом.....	8
3	Пример исследования.....	12
3.1	Описательная статистика и структурный анализ.....	12
3.2	Стохастическое моделирование.....	13
4	Выводы.....	14
5	Благодарности.....	15
6	Литература.....	15
7	Приложение.....	17

1 Введение

В настоящей работе продолжается исследование подходов к анализу пространственно распределенных данных, начатое в работах [4] – [9]. Методы, объединенные общим понятием *геостатистика*, дают возможность эффективно работать с данными и широко применяются при анализе данных в различных областях. В данном случае исследование проводилось на данных по загрязнению водной среды.

Существует два главных подхода к анализу пространственно распределенной информации: детерминистический и статистический. Детерминистический подход предполагает наличие жесткой модели поведения данных, т.е. формульной зависимости, построенной на априорной информации о природе исследуемого феномена. При этом сами данные не говорят о том, какой подход следует применить.

При статистическом подходе (чем и занимается геостатистика) данные рассматриваются как результат случайного процесса. Это, однако, не предполагает действительно случайной природы исследуемого феномена. Основная идея предсказательной статистики — характеризовать любое не измеренное значение z как случайную переменную Z , плотность вероятности которой характеризует неопределенность вокруг z . Такое представление дает возможность получить не только оценку, но и вариацию оценки, которая характеризует пространственную неопределенность предсказанных значений. Заметим, что ни один из традиционных детерминистических методов пространственной интерполяции ошибку оценки получить не позволяет.

Геостатистика сама включает различные подходы к анализу данных: пространственную интерполяцию и стохастическое моделирование. Исследование последнего и является целью данной работы. Стохастическое моделирование, как и остальные методы геостатистики, позволяет оценить пространственную вариабельность и неопределенность данных. В отличие от детерминистических методов и интерполяционных оценщиков, которые усредняют и сглаживают вариабельность исходных данных, стохастическое моделирование позволяет воспроизвести случайный пространственный процесс на основе определенных качественных и количественных критериев исходного распределения. В то время, как интерполяционная оценка в точке единственна для заданных модели и ее параметров,

результатом стохастического моделирования является набор оценок (реализаций) значения функции в точке для заданной модели.

Стохастическое моделирование позволяет сгенерировать множество равновероятных реализаций, обладающих свойствами исходного распределения:

- одинаковой с исходным распределением плотностью распределения;
- точным воспроизведением исходных данных;
- одинаковой с исходным распределением пространственной структурой;

Стохастические методы основываются на предположении, что анализируемые данные являются реализациями случайного процесса. Это позволяет, используя имеющиеся измерения как зафиксированные значения (условное моделирование), получить бесконечно много значений (реализаций) переменной в точке оценивания. Построенные таким образом стохастические реализации обладают той же функцией распределения и такой же пространственной корреляционной структурой, что и исходные данные. Наличие нескольких равновероятных оценок в одной точке позволяет определить неопределенность оценивания и построить вероятностные карты – оценки вероятности превышения заданного уровня значений или оценки, которые могут быть превышены действительными значениями с заданной вероятностью. В настоящей работе использовались методы последовательного Гауссова моделирования и моделирования отжигом (аннилинг) для данных по загрязнению донных отложений Женевского озера (Lac Lemman).

2 Стохастическое моделирование

Чем же стохастическое моделирование отличается от всевозможных геостатистических и детерминистических оценщиков и интерполяторов [1],[4]-[8],[10] ?

Пусть $\{z^*(u_j), u_j \in A, j=1 \dots N\}$ – это набор оценок с помощью кригинга величины z по некоторому изучаемому пространству A . Каждая оценка $z^*(u_j)$, взятая отдельно, т.е. независимо от соседних оценок $z^*(u_j)$, является «наилучшей» в смысле наименьших квадратов, т.к. вариация локальной ошибки $Var\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$ минимальна. (Здесь и далее Z обозначает случайную переменную которая и дает значение z в конкретных точках \mathbf{u} пространства A .) Карты таких наилучших реализаций, однако, не всегда лучшие в целом. Можно показать, что интерполяционные алгоритмы приводят к сглаживанию оценки локальных деталей. Обычно, маленькие значения переоцениваются, в то время как большие – недооцениваются. Такое условное смещение является серьезным препятствием при попытке определить области экстремальных значений величины. Другой недостаток рассматриваемых интерполяционных оценщиков состоит в том, что сглаживание неоднородно – оно зависит от локальной конфигурации данных: сглаживание меньше в областях с большим количеством данных и увеличивается с расстоянием от точки оценивания до ближайших данных.

Сглаженные интерполяционные карты не дадут хороших результатов для приложений чувствительных к присутствию экстремальных значений. Для решения таких задач предпочтительнее использовать альтернативный метод – стохастическое моделирование.

Вместо карты наилучших локальных оценок, стохастическое моделирование генерирует численные модели (реализации) для величины $z: \{z^{(l)}(u), u \in A\}$, где l обозначает номер реализации. Эти карты воспроизводят статистические особенности задачи в целом. При этом модели, как правило, строятся исходя из следующих требований:

1. Точно воспроизводятся исходные данные (если данные измерены точно или ошибки измерений не существенны):

$$z^{(l)}(u) = z(u_\alpha) \quad \forall u = u_\alpha, \alpha=1, \dots, n.$$

Моделирование в этом случае называется условным. В дальнейшем это условие будет кратко обозначаться значком (n) . Это так называемая «жесткая информация». В принципе, простое обобщение моделей позволяет учесть и «мягкую информацию», например, данные с ошибками, неопределенность в исходных данных, качественные и количественные знания об исследуемом явлении и др.

2. Гистограмма смоделированных значений близко воспроизводит репрезентативную гистограмму (декластеризованных, в общем случае) данных.

3. Воспроизводятся вариограмная, ковариационная модели или набор индикаторных вариограм (ковариаций) для различных порогов z_k .

Также могут быть воспроизведены более сложные черты, такие как пространственная корреляция со вторичными данными или многоточечная статистика.

Существует три основных класса алгоритмов стохастического моделирования:

1. Широкий класс алгоритмов, известный под общим названием *последовательного моделирования (sequential simulation algorithms)*: вместо условной функции распределения N -переменных, моделируется ccdf одной переменной и в соответствии с ней присваивается значение каждому из N узлов, перебираемых в случайном порядке. Чтобы воспроизвести вариограмную (ковариационную) модель, каждая ccdf делается условной не только к исходным n данным, но и ко всем значениям, промоделированным ранее.
2. Алгоритм *p-field approach* также использует переход от условной функции распределения N переменных к некоторым ccdf одной переменной. В отличие от подхода последовательного моделирования, все они условны лишь к n исходным данным. Воспроизводство вариограммы (ковариации) здесь аппроксимируется наложением автокорреляционной структуры на вероятностные значения, используемые для выбора этих условных функций распределения.
3. В отличие от двух предыдущих алгоритмов, при *моделировании отжигом (simulated annealing)* создание стохастической реализации формулируется как оптимизационная задача безотносительно к какой бы то ни было модели случайных функций распределения. Основная идея – постепенное возмущение начального образа так, чтобы удовлетворить конечной конструкции, а именно вариограмной (ковариационной) модели. Различные реализации получаются благодаря различным начальным образам и случайности возмущения.

Стохастические реализации моделируются посредством выборки из условной функции распределения (ccdf – conditional cumulative distribution function) в данной точке u :

$$F(u; z|(n)) = Prob\{Z(u) \leq z|(n)\}$$

Карты реализаций генерируются выборкой из ccdf N -переменных, которая моделирует совместную неопределенность в N точках u_i :

$$F(u_1, \dots, u_N; z_1, \dots, z_N |(n)) = Prob\{Z(u_1) \leq z_1, \dots, Z(u_N) \leq z_N |(n)\}$$

Рассмотрим последовательный алгоритм стохастического моделирования, который подразумевает использование ранее промоделированных значений наравне с исходными данными для симуляции в последующих точках.

Идея в том, чтобы перейти от одной условной функции N -переменных к N функциям одной переменной:

$$F(u_1, \dots, u_N; z_1, \dots, z_N |(n)) = F(u_N; z_N |(n+N-1)) \cdot F(u_{N-1}; z_{N-1} |(n+N-2)) \cdot \dots \cdot F(u_1; z_1 |(n)),$$

где $F(u_N; z_N |(n+N-1))$ – это условная функция распределения $Z(u_N)$, учитывающая набор из n исходных данных и $(N-1)$ промоделированных значений $Z(u_i)$.

2.1 Последовательное Гауссово моделирование (sequential Gaussian simulation)

Рассмотрим общий подход последовательного моделирования. Пусть $\{Z(u_j), j=1, \dots, N\}$ – набор случайных переменных, определенных в произвольных N точках u_j , в изучаемом пространстве A . Цель – произвести несколько совместных реализаций этих N случайных переменных:

$$\{z^{(l)}(u_j), j=1, \dots, N\} \quad l = 1, \dots, L$$

при условии набора данных $\{z(u_\alpha), \alpha=1, \dots, n\}$

Возьмем сначала симуляцию величины z только в двух точках u_1 и u_2 . Набор пар $\{z^{(l)}(u_1), z^{(l)}(u_2)\}, l = 1, \dots, L$ может быть произведен с помощью выборки из условной cdf двух переменных:

$$F(u_1, u_2; z_1, z_2 | (n)) = \text{Prob}\{Z(u_1) \leq z_1, Z(u_2) \leq z_2 | (n)\}$$

Другой подход обеспечивается аксиомой Байеса [1], которая позволяет заменять такую cdf произведением двух функций:

$$F(u_1, u_2; z_1, z_2 | (n)) = F(u_2; z_2 | (n+1)) \cdot F(u_1; z_1 | (n)),$$

$(n+1)$ здесь обозначает условие из n данных и первой реализации. Это разложение позволяет генерировать пары $\{z^{(1)}(u_1), z^{(1)}(u_2)\}$ в два шага: значение $z^{(1)}(u_1)$ выбирается из cdf $F(u_1; z_1 | (n))$, потом cdf в точке u_2 принимается условной к реализации $z^{(1)}(u_1)$ вдобавок к исходным данным (n) и выборка из этой функции дает коррелированное значение $z^{(1)}(u_2)$.

Такой принцип последовательности может быть обобщен на большее число точек и разложение будет иметь вид:

$$F(u_1, \dots, u_N; z_1, \dots, z_N | (n)) = F(u_N; z_N | (n+N-1)) \cdot F(u_{N-1}; z_{N-1} | (n+N-2)) \cdot \dots \cdot F(u_1; z_1 | (n)),$$

Теперь мы можем производить реализации случайного вектора $\{Z(u_j), j=1, \dots, N\}$ с помощью следующих шагов:

- Моделируем cdf в первой точке u_1 , условную к n исходным данным $z(u_\omega)$

$$F(u_1; z_1 | (n)) = \text{Prob}\{Z(u_1) \leq z_1 | (n)\}$$

- Выбираем из нее реализацию $z^{(1)}(u_1)$, которая становится условным данным для следующих точек.
- В i -том узле u_i , моделируем условную cdf, при условии n данных и $(i-1)$ значений $z^{(1)}(u_j)$ предыдущих узлов

$$F(u_i; z_i | (n+i-1)) = \text{Prob}\{Z(u_i) \leq z_i | (n+i-1)\}$$

- Выбираем из нее реализацию $z^{(1)}(u_i)$, которая становится условным данным для следующих точек.
- Повторяем два предыдущих шага пока не посетим все N узлов и не присвоим каждому свое значение.

Повторяя все L раз, получим L различных реализаций.

Применение последовательного принципа для мультигауссовой модели случайной функции называется *последовательным Гауссовым моделированием*. Рассмотрим моделирование непрерывной величины z в N узлах сетки (не обязательно регулярной) при условии набора данных $\{z(u_\omega), \alpha=1, \dots, n\}$. Процесс последовательного Гауссового моделирования выглядит так:

1. Сначала нужно проверить применимость мультигауссовой модели, которая требует предварительного преобразования z -данных в y -данные со стандартной нормальной функцией распределения, используя normal score преобразование [1],[10]. Мультинормальность распределения результирующей normal score переменной $Y(u) = \phi(Z(u))$ потом проверяется.
2. Если модель подходит для y -переменной, последовательные симуляции выполняются на y -данных:
 - определяется случайный путь перебора узлов сетки. Чтобы избежать похожих конечных реализаций, которые могут привести к недооценке неопределенности оцениваемой величины, пути следования по узлам для различных реализаций должны отличаться друг от друга.
 - в каждом узле u определяются параметры (среднее и дисперсия) гауссовой cdf $G(u; y | (n))$ с помощью кригинга (простого, обычного и др.) с вариограммой моделью $\gamma(h)$ для нормально преобразованных данных. Условная информация (n) состоит из начальных нормально преобразованных данных $y(u_\omega)$ и значений $y^{(1)}(u_j)$ промоделированных в предыдущих узлах.

Приведем уравнения кригинга [1]. Все интерполяторы семейства кригинга основываются на линейной регрессионной оценке $Z^*(u)$, определяемой так:

$$Z^*(u)-m(u) = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)[Z(u_{\alpha})-m(u_{\alpha})]$$

где $\lambda_{\alpha}(u)$ – веса, приписанные данным $z(u_{\alpha})$, интерпретируемым, как реализации случайной величины $Z(u_{\alpha})$. Величины $m(u)$ и $m(u_{\alpha})$ – средние значения величин $Z(u)$ и $Z(u_{\alpha})$. Все неизвестные величины определяются из решения задачи минимизации вариации невязки

$$\sigma_E^2(u) = \text{Var}\{Z^*(u) - Z(u)\}$$

при условии $E\{Z^*(u) - Z(u)\} = 0$, которое называется условием несмещенности.

Простой кригинг предполагает среднее значение m постоянным во всей области A и известным. Уравнение в этом случае выглядит так:

$$Z^*_{SK}(u) = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)[Z(u_{\alpha}) - m] + m = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)Z(u_{\alpha}) + [1 - \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)]m.$$

Оценка получается сама по себе несмещенной, т.к.

$$E\{Z^*_{SK}(u) - Z(u)\} = m - m = 0.$$

Тогда уравнение можно переписать так:

$$\sum_{\beta=1}^{n(u)} \lambda_{\beta}(u)C(u_{\alpha} - u_{\beta}) = C(u_{\alpha} - u) \quad \alpha = 1, \dots, n(u),$$

а минимизируемую величину так:

$$\sigma^2_{SK}(u) = C(0) - \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)C(u_{\alpha} - u),$$

где $C(h) = E\{[Z(u)-m][Z(u+h)-m]\}$ – ковариация между значениями в точках, разделенных расстоянием (в общем случае вектором) h . Соответственно, $C(u_{\alpha} - u_{\beta}) = E\{[Z(u_{\alpha})-m][Z(u_{\beta})-m]\}$

Для данных, удовлетворяющих внутренней гипотезе (intrinsic hypothesis) [1],[10], верно, что $\gamma(h) = C(0) - C(h)$, где $\gamma(h)$ – *вариограмма*, по определению:

$$2\gamma(h) = \text{Var}\{Z(u+h)-Z(u)\} = E\{(Z(u+h)-Z(u))^2\}.$$

При использовании кригинга предполагается, что данные удовлетворяют внутренней гипотезе, поэтому можно использовать вариограмму $\gamma(h)$, вместо ковариации $C(h)$.

В *Обычном кригинге* среднее m предполагается постоянным во всей области A , но неизвестным. Уравнение выглядит так:

$$Z^*_{OK}(u) = \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)Z(u_{\alpha}) + [1 - \sum_{\alpha=1}^{n(u)} \lambda_{\alpha}(u)]m(u)$$

что в конечном итоге приводит к системе из $(n(u)+1)$ уравнений с $(n(u)+1)$ неизвестными – $n(u)$ весов $\lambda_{\alpha}(u)$ и параметр Лагранжа $\mu_{OK}(u)$:

$$\begin{cases} \sum_{\beta=1}^{n(u)} \lambda_{\beta}(u)\gamma(u_{\alpha} - u_{\beta}) - \mu_{OK}(u) = \gamma(u_{\alpha} - u) \\ \sum_{\beta=1}^{n(u)} \lambda_{\beta}(u) = 1 \end{cases} \quad \alpha=1, \dots, n(u)$$

Последнее уравнение гарантирует несмещенность получаемой оценки.

Возвращаясь к последовательному Гауссовому моделированию, переходим к следующему шагу:

- после получения из уравнений кригинга оценок для среднего и дисперсии Гауссового распределения, разыгрывается симулирующее значение $y^{(l)}(u)$ в соответствии с функцией распределения и добавляется к данным.
 - берется следующий узел вдоль выбранного случайного пути и повторяются два предыдущих шага.
 - продолжаем, пока не переберем все узлы.
3. Последний шаг состоит из обратного преобразования промоделированных normal score значений $\{y^{(l)}(u_j), j=1, \dots, N\}$ в моделируемые значения для исходной переменной:

$$z^{(l)}(u_j) = \Phi^{-1}(y^{(l)}(u_j)), j=1, \dots, N$$

где $\Phi^{-1}(\cdot) = F^{-1}(G(\cdot))$, где $F^{-1}(\cdot)$ – обратная cdf переменной Z и $G(\cdot)$ – стандартная гауссова cdf. Такое преобразование позволяет сохранить исходную гистограмму. Действительно:

$$Prob\{Z^{(l)}(u) \leq z\} = Prob\{\Phi^{-1}(Y^{(l)}(u)) \leq z\} = Prob\{Y^{(l)}(u) \leq \Phi(z)\} = G[\Phi(z)] = F(z),$$

по определению normal score преобразования.

Другие реализации $\{z^{(l')}(u_j), j=1, \dots, N\}$, $l' \neq l$, получаются повторением шагов 2 и 3 с различным случайным путем перебора узлов.

Хотя в теории все ранее промоделированные точки должны быть использованы для моделирования следующих, на практике это является препятствием для воспроизводства вариограммной (ковариационной) модели данных на больших расстояниях. Происходит экранировка удаленных данных большим количеством близких, уже оцененных точек. Мульти-сетевая концепция (multiple grid concept – Gomez-Hernandez, 1991 [11]; Tran, 1994 [12]) позволяет воспроизводить крупномасштабные корреляционные структуры, не принимая в расчет слишком много точек, промоделированных за счет малого количества исходных данных. Идея в том, что сначала используется сетка с крупным шагом, чтобы промоделировать крупномасштабные структуры, а потом уже на мелкой моделируется подробная карта. Процедуру можно обобщить на любое количество промежуточных сеток, но путь следования по узлам для каждой сетки должен быть случайным.

2.2 Моделирование отжигом (simulated annealing)

Simulated annealing [1],[3] – это общее название для семейства оптимизационных алгоритмов, основанных на принципах стохастической релаксации. Заданный образ постепенно возмущается так, чтобы подогнать его под воспроизводство каких-либо целевых структур (гистограмма, вариограмма, ковариация и т.п.), оставляя исходные данные неизменными.

Термин «отжиг» (*annealing*) пришел в математику из металлургии [7]. При высокой температуре металл легко деформируется и меняет свою форму. Понятие температуры связано со скоростью атомов, которые хаотически передвигаются. Чем выше скорость движения атомов, тем выше температура, и тем легче металл поддается деформации. Таким образом, если резко понизить температуру металла, то атомы резко «замрут» и мы получим твердую, но очень неровную поверхность. Если стоит задача получить максимально ровную поверхность, то необходимо сначала сильно разогреть металл, а потом, постепенно очень медленно охлаждать его до полного затвердевания. Таким образом металл будет проходить через множество квазиравновесных состояний, и у всех его атомов будет достаточно времени, чтобы найти свое «лучшее» место среди других атомов, в смысле минимума полной энергии системы, что соответствует идеальной кристаллической решетке. Чем равномернее будут распределены атомы вдоль поверхности, тем ровнее в итоге будет поверхность. С точки зрения математики, задачу можно сформулировать так: минимизировать среднюю квадратичную ошибку отклонения уровня поверхности металла от некоторого постоянного значения. Оптимизируемыми параметрами в этом случае будут координаты положения атомов.

Рассмотрим моделирование непрерывной величины z в N узлах сетки u_j , при заданных условиях $z(u_\alpha)$, $\alpha=1, \dots, n$, таким образом, чтобы вариограмма данных воспроизводилась для первых S лэгов. Аннилинг

требует задания целевой функции (являющейся как раз аналогом энергии), которая измеряет разницу между значениями целевых и текущих статистических параметров на каждом i -м возмущении. Если цель – воспроизвести вариограммную модель, целевая функция может выглядеть так:

$$O(i) = \sum_{s=1}^S [\gamma(h_s) - \gamma_i(h_s)]^2,$$

где $\gamma(h_s)$ – значение требуемой вариограммной модели для лага h_s и $\gamma_i(h_s)$ соответствующее значение вариограммы реализации на i -м возмущении, $\{z_{(i)}^{(l)}(u_j), j=1, \dots, N\}$.

Если целевая функция установлена, процесс моделирования (точнее, процесс оптимизации) включает в себя систематическое модифицирование начальной реализации так, чтобы уменьшить значение целевой функции, делая реализацию приемлемо близкой к целевой статистике. Общий алгоритм аннилинга выглядит так:

1. Создаем начальную реализацию, $\{z_{(0)}(u_j), j=1, \dots, N\}$, которая сохраняет исходные данные и может сразу аппроксимировать какую-нибудь целевую статистику (дисперсию распределения, плато вариограммы или гистограмму).
2. Считаем начальное значение целевой функции, соответствующее этой начальной реализации:

$$O(i) = \sum_{s=1}^S [\gamma(h_s) - \gamma_0(h_s)]^2$$

3. Возмущаем реализацию каким-либо механизмом, например отражением пар z -значений: $z_{(0)}(u_j)$ становится $z_{(0)}(u_i)$ и наоборот. По аналогии с физическими процессами в остывающем металле, механизм возмущения зависит от температуры – чем ниже температура, т.е. чем больше шагов сделано, тем меньше меняется значение в точке при возмущении. Например, в случае отражения пар, уменьшается расстояние между точками, значения которых мы меняем местами.
4. Оцениваем эффект возмущения на воспроизводство целевой статистики, снова вычисляя значение целевой функции, $O_{new}(0)$, учитывая модификацию начальной реализации.
5. Принимаем или не принимаем возмущение на основе какого-либо правила. Обычно, вероятность принятия задается с помощью распределения Больцмана [13]:

$$P\{\text{принять}\} = \begin{cases} 1, & \text{если } O_{new} \leq O_{old} \\ \exp\left(\frac{O_{new} - O_{old}}{T}\right), & \text{в другом случае} \end{cases},$$

где T — температура. Чем выше температура, тем больше вероятность принять неблагоприятное, т.е. не уменьшающее целевую функцию возмущение. Делается это для того, чтобы была возможность избежать локальных минимумов и найти глобальный.

6. Если возмущение принимается, заменяем начальную реализацию на новую $\{z_{(1)}^{(l)}(u_j), j=1, \dots, N\}$, с новой целевой функцией $O(1) = O_{new}(0)$.
7. Повторяем шаги с 3-го по 6-й, пока целевая структура не будет приемлемо достигнута или пока возмущения не перестанут уменьшать целевую функцию. Потом понижаем температуру и проделываем всю вышеуказанную процедуру (шаги 3 – 6) для новой температуры и так поступаем до тех пор, пока не достигнем приемлемого результата (см. ниже критерий остановки),.

Другие реализации $\{z^{(l)}(u_j), j=1, \dots, N\}$, $l' \neq l$, производятся повторением выше описанной процедуры, начиная с другого начального образа. Обычно, число узлов N так велико, а вариограмма накладывает так мало связей, что существует очень много решений оптимизационной задачи. Конечная реализация выбирается из этого набора приблизительных решений.

Существует много способов осуществления алгоритма моделирования отжигом. Варианты отличаются тем, как создавать начальный образ, как его возмущать, компонентами целевой функции и типом критерия, принимать или не принимать эффект после возмущения.

Начальный образ:

- начальный образ должен быть легко производим.
- значения точек начального образа должны сразу подходить какой-нибудь простой целевой структуре (например, воспроизводить гистограмму данных), так чтобы ускорить последующий процесс оптимизации.
- все начальные образы должны быть «равновероятны». Нужно остерегаться использовать один и тот же начальный образ в качестве стартовой точки для нескольких различных реализаций, потому как даже разные дальнейшие пути могут привести к слишком похожим конечным реализациям и вызвать, таким образом, недооценку неопределенности.

Обычно начальный образ производится так: исходные данные «замораживаются» на своих местах, а приписываемое каждому узлу значение z -величины выбирается случайным образом из целевой cdf $F(z)$. Такой подход достаточно быстр и дает набор начальных образов уже удовлетворяющих целевой гистограмме.

Начальный образ также может быть реализацией какого-либо другого алгоритма моделирования, например p -field или последовательного моделирования. В таком случае аннилинг выступает в качестве пост-процессора, с целью улучшить воспроизводство целевой статистики или наложить дополнительную структуру, которая не может быть введена другими алгоритмами.

Механизм возмущения.

Чаще всего используются два следующих механизма:

1. Отражение пар z -величин в случайным образом выбранных точках u_j и u_k , находящихся на расстоянии $d \leq D(T)$, $D(T)$ понижается с понижением T :

$$\begin{cases} z_{(i)}(u_j) = z_{(i-1)}(u_k) \\ z_{(i)}(u_k) = z_{(i-1)}(u_j) \end{cases}$$

Такой механизм возмущения позволяет сохранить неизменной гистограмму начального образа. Значит, нет необходимости включать воспроизводство гистограммы в целевую функцию, если начальный образ уже удовлетворяет ей.

2. Случайным образом выбирается одна точка u_j и модифицируется соответствующее значение $z_{(i-1)}(u_j)$ согласно какому-нибудь механизму, например:

$$z_{(i)}(u_j) = F^{-1}(p_j)$$

где p_j – случайное значение из $[0,1]$, а $F(z)$ – целевая гистограмма. В отличие от механизма отражения, сохраняющего начальную гистограмму, здесь требуется специальная компонента целевой функции.

В обоих случаях условные (т.е. исходные) данные никогда не возмущаются, чтобы конечная реализация сохранила эти значения.

Целевая функция.

Моделирование отжигом позволяет принимать в расчет различные типы информации, вводя ее количественные характеристики в глобальную целевую функцию. Эта функция представляет собой взвешенную сумму C компонент O_c , измеряющих разницу между статистикой текущей реализации (на i -м возмущении) и целевой статистикой:

$$O(i) = \sum_{c=1}^C \alpha_c O_c(i)$$

где веса α_c контролируют относительную важность c -й компоненты целевой функции. Приведем примеры наиболее часто используемых компонент целевой функции:

- (1) **Кумулятивная функция распределения** (учет гистограммы данных)

Типичная целевая статистика – это, по возможности декластеризованная, однопеременная функция распределения z -данных $F(z)$. Если диапазон изменения z описывается серией из K порогов z_k , разница между целевой и текущей cdf может быть измерена так:

$$O_c(i) = \sum_{k=1}^K [F(z_k) - F'_{(i)}(z_k)]^2$$

где $F'_{(i)}(z_k)$ – значение на пороге z_k , вычисленное для реализации на i -м возмущении.

(2) Модель полувариограммы

Воспроизводство вариограммной модели $\gamma(h)$ обычно ограничивается определенным количеством (S) лэгов. Разница между целью и текущим значением в этом случае измеряется так:

$$O_c(i) = \sum_{s=1}^S \frac{[\gamma(h_s) - \hat{\gamma}_{(i)}(h_s)]^2}{[\gamma(h_s)]^2}$$

где $\hat{\gamma}_{(i)}(h_s)$ – значение полувариограммы реализации для лэга h_s на i -м возмущении.

(3) Модели индикаторных полувариограмм

Моделирование отжигом позволяет учитывать специальные пространственные структуры, моделируемые индикаторными полувариограммами $\gamma(h, z_k)$, посчитанными для K различных порогов z_k . Так получаем еще одну компоненту целевой функции:

$$O_c(i) = \sum_{k=1}^K \sum_{s=1}^S \frac{[\gamma(h_s; z_k) - \hat{\gamma}_I^{(i)}(h_s; z_k)]^2}{[\gamma(h_s; z_k)]^2}$$

где $\hat{\gamma}_I^{(i)}(h_s; z_k)$ – значения индикаторной полувариограммы на лаге h_s по порогу z_k для реализации на i -м возмущении.

(4) Многопеременная (многоточечная) статистика

Если для описания структуры данных не достаточно двух-точечной статистики (полувариограммы и ковариации), аннилинг предлагает использовать большее число точек одновременно.

Рассмотрим J -точечную конфигурацию, определенную для J отдельных векторов h_1, \dots, h_J , где $h_1=0$ по определению. Вероятность, что J значений $z(u+h_1), \dots, z(u+h_J)$ вместе не превышают J порогов z_1, \dots, z_J определяется так:

$$\Phi(h_1, \dots, h_J; z_1, \dots, z_J) = E \left\{ \prod_{j=1}^J I(u + h_j; z_j) \right\}$$

Если все J порогов равны одному и тому же значению z_c и все J векторов кратны одному вектору h , то многоточечная статистика – это функция связности, введенная Journel и Alabert (1988) [14]:

$$\Phi(J; z_c) = E \left\{ \prod_{j=1}^J I(u + (j-1)h; z_c) \right\}$$

Эта величина выражает вероятность того, что поток J значений, ориентированных вдоль вектора h , совместно не превышает данного порогового значения z_c .

Статистика такого типа может быть воспроизведена добавлением в целевую функцию компоненты типа:

$$O_c(i) = [\Phi(h_1, \dots, h_J; z_1, \dots, z_J) - \hat{\Phi}_{(i)}(h_1, \dots, h_J; z_1, \dots, z_J)]^2$$

где экспериментальное значение $\hat{\Phi}_{(i)}(h_1, \dots, h_J; z_1, \dots, z_J)$ считается на i -м возмущении.

(5) Коэффициент корреляции

Пусть Y – лучшая выборка или ранее промоделированная величина, вторичная по отношению к величине Z . Если взаимосвязь Z - Y корректно описывается линейным коэффициентом корреляции $\rho_{ZY}(0)$, то его можно ввести в целевую функцию в виде компоненты:

$$O_c(i) = [\rho_{ZY}(0) - \hat{\rho}_{ZY}^{(i)}(0)]^2$$

где $\hat{\rho}_{ZY}^{(i)}(0)$ — коэффициент корреляции, вычисленный по соответствующим парам y -данных и z -значений на i -м возмущении.

(6) Кросс-вариограмма

Пространственная кросс-корреляция между двумя величинами Z и Y , моделируемая с помощью кросс-вариограммы $\gamma_{ZY}(h)$, может быть воспроизведена включением в целевую функцию компоненты типа:

$$O_c(i) = \sum_{s=1}^S \frac{[\gamma_{ZY}(h_s) - \hat{\gamma}_{ZY}^{(i)}(h_s)]^2}{[\gamma_{ZY}(h_s)]^2}$$

где $\hat{\gamma}_{ZY}^{(i)}(h_s)$ — значение кросс-вариограммы между z -значениями и y -данными на i -м возмущении.

Критерий остановки

Как и для любого итерационного алгоритма, для этого оптимизационного процесса должен быть определен критерий остановки. Возможные критерии таковы:

1. Целевая функция достигла достаточно малого значения O_{min} .
2. Максимальное количество возмущений, которые могут быть произведены при одной температуре превысило определенное число.
3. Доля приемлемых возмущений меньше, чем заданное пороговое значение.

3 Пример исследования

3.1 Описательная статистика и структурный анализ

В качестве примера были выбраны данные по загрязнению Женевского озера цинком (Zn), снятые в 1978 г. Использовались измерения в 292 точках (рис.1). Эти данные имеют две точки высоких (выпадающих) значений 320 мкг/г наряду с остальными от 250 мкг/г и ниже, но при обработке они все же принимались в расчет во избежание потери вариабельности модели. Гистограмму распределения данных отображает рис. 21

На рисунке 2 приведена карта изолиний по триангуляции имеющихся данных.

Целью структурного анализа является выявление корреляционной структуры данных с помощью вариографии и построение модели этой структуры. Были построены вариограммы по шести направлениям и вариограммы розы (рис. 9), которые явно демонстрировали анизотропию данных. Структура выявляется на расстоянии 18-20 км, причем в направлениях 90° и -60° наблюдается тренд, на первый взгляд являющийся следствием двух выпадающих значений, но при исследовании данных без их учета, тренд все равно сохранялся.

Для построения модели вариограммы была использована одна сферическая структура, т.к. анизотропию данных можно считать геометрической. О качестве построенной модели можно судить по рис.3.

Для использования последовательного Гауссова моделирования данные были нормализованы. Полученное распределение трансформированных данных (normal score) близко к нормальному. Так же

как и для исходных, была построена анизотропная модель преобразованных данных на основе вариограмм по шести направлениям (рис. 4). Если сравнить вариограммы для нормализованных данных с вариограммами для нетрансформированных, то первые имеют более гладкий характер и в них не наблюдаются значительные тренды, как в нетрансформированных.

Последовательное Гауссово моделирование, являясь параметрическим методом, нуждается в принятии гипотезы о мультинормальности. Проверка на мультинормальность является нетривиальной и практически невозможной в условиях наличия одной реализации. Поскольку параметрические модели как правило используют моменты первого и второго порядка, то достаточным является проверка на нормальность, что обеспечивается нормализованным преобразованием, и бинормальность. Для данных, распределенных бинормально, должно выполняться следующее качественное соотношение:

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

Это соотношение было проверено для вариограмм, использованных при моделировании – по всем направлениям (omnidirectional), рис. 18b и по шести выше указанным направлениям (см. рис. 18a), и, оказалось, что, по крайней мере, на расстоянии 20 км и меньше они удовлетворяют указанному условию. При этом отношение укладывается в десятипроцентную окрестность $\sqrt{\pi}$. По крайней мере, гипотезу о бинормальности нельзя отбросить.

Еще одна проверка бинормальности была произведена с помощью программы BIGAUS (GSLIB) [2]. Эта программа использует модель вариограммы для normal score данных и строит индикаторные полувариограммы для заданных порогов, лэгов и направлений, соответствующие бинормально распределенным данным. Идея в том, чтобы увидеть, насколько экспериментальные индикаторные вариограммы отличаются от тех, что дает двумерное Гауссово распределение. Сравнение индикаторной вариограммы (индикатор вычислялся по порогу 0.5, что соответствует медиане 120 мкг/г) с той, что должна быть у бинормального распределения, приведено на рис. 19a и рис. 19b.

3.2 Стохастическое моделирование

Последовательное Гауссово моделирование проводилось с использованием вариограммных моделей, полученных при структурном анализе преобразованных данных (см. рис. 4). Моделирование проводилось с простым кригингом, количество реализаций в одной серии менялось от 10-ти до 30-ти. Для экстраполяции хвостов использовалась линейная функция, пределы максимума и минимума были установлены 350 мкг/г и 0 мкг/г соответственно; стоит отметить, что алгоритм SGSIM [2] (GSLIB) варьирует минимум и максимум от симуляции к симуляции в диапазоне от значений минимума и максимума исходных данных до тех, что мы устанавливаем сами, в нашем случае от 0 мкг/г до 54 мкг/г и от 320 мкг/г до 350 мкг/г соответственно. Отношение количества исходных данных к количеству уже промоделированных, используемых для дальнейшей оценки, около 1 почти по всей области оценивания, т.к. заранее задавались параметры: максимум 60 ранее полученных узлов, максимум 80 исходных данных (при минимуме 10). Такие параметры позволяют получить достаточно хороший результат внутри области озера, оставляя границу экстраполяции вне ее. Для вычисления задавалась регулярная сетка, достаточно плотная (120×80) и все данные были приписаны к узлам (использовалась процедура GSLIB, сдвигающая данные к ближайшим узлам сетки и оптимизирующая алгоритм поиска). Неприписывание данных к узлам сетки приводило к небольшому понижению вариабельности данных, по сравнению с исходными, и подсчет в этом случае отнимал в несколько раз больше машинного времени. Также использовался мульти-сетевой поиск (multiple grid search), при подключении которого результаты (по визуальной оценке и анализу с помощью вариограмм) получались немного лучше, но при работе с нашими данными, использование этой стратегии не являлось принципиальным.

Стратегия поиска была выбрана в соответствии с вариограммной структурой и геометрической структурой озера: эллипсоид поиска, т.е. область, из которой брались точки для оценивания, задавался с радиусом 20 км в главном направлении (25° от горизонтали к северо-востоку) и 8 км в перпендикулярном, что обеспечивало покрытие от 1/4 до 1/3 исследуемой области для каждого моделируемого узла.

На рисунках 5a, 5b и 5c показаны три различные реализации, наиболее отвечающие действительности по структуре. Видно, что каждая реализация лучше моделирует какую-либо отдельную часть области. На рис. 7 представлены вариограммы по всем направлениям четырех реализаций с различными случайными

зернами (seed). Для сравнения приведена вариограмма преобразованных данных; видно, что вариограммы имеют общую структуру.

Все вычисления производились при помощи программы SGSIM (GSLIB) [2].

В процессе дальнейшей обработки на основе 30-ти реализаций вычислялись оценки E-типа. Поскольку такие оценки являются усреднением результата моделирования по числу реализаций, можно ожидать результаты, схожие с результатами кригинга (рис. 12 — карта оценок E-типа, рис. 20 – оценка обычным кригингом). Объясняется это тем, что именно кригингом определяются исходные среднее и дисперсия распределения и из этого распределения потом выбираются значения для конкретных реализаций. Усреднение по очень большому числу реализаций последовательного Гауссового моделирования должно привести к тому среднему, которое изначально дает решение уравнений кригинга.

Обработка большого числа реализаций позволяет также построить карты риска, т.е. карты вероятности непревышения какого-либо заданного уровня – ведь в каждой точке оценки мы имеем распределение, а, следовательно, знаем вероятность принять то или иное значение. Для иллюстрации построены карты непревышения уровня 100 мкг/г (рис. 13) и 130 мкг/г (рис. 14).

Построение оценок E-типа и карт риска осуществлено с помощью программы POSTSIM (GSLIB) [2].

Моделирование отжигом (*simulated annealing*), в отличие от Гауссового моделирования, базируется на вариограмной модели непреобразованных (raw) данных (рис. 3). При моделировании использовалась двухкомпонентная целевая функция – для гистограммы и вариограммы. Компоненты целевой функции брались в отношении 10% для гистограммы и 90% для вариограммы – такое соотношение было выбрано, как наилучшее, исходя из получаемых результатов. При добавлении веса гистограмме, получалась слишком сглаженная оценка, при уменьшении же ее веса меньше 10%, гистограммы полученных реализаций начинали сильно не соответствовать гистограмме исходного распределения.

Алгоритм SASIM, в отличие от описанного выше SGSIM сохраняет максимальное и минимальное значение, взятое из данных, поэтому во всех реализациях минимумом было значение 54 мкг/г, а максимумом 320 мкг/г. Карты трех различных реализаций представлены на рисунках ба-бс. Вариограммы по двум наиболее хорошо промоделированным направлениям для нескольких реализаций представлены на рис. 8.

Для отжига, также как и для гауссовых симуляций, было проведено усреднение по 30-ти реализациям, т.е. построена карта оценок E-типа и карты непревышения уровней 100 мкг/г и 130 мкг/г. Результаты представлены на рис. 15-17. Теперь, в отличие от Гауссовых симуляций, даже визуально заметно отличие карт оценок E-типа от оценок обычного кригинга, что объясняется тем, что стратегия поиска при моделировании отжигом не использует ни кригинг, ни какие-либо другие интерполяционные методы. Подробная теория моделирования отжигом описана в разделе 2.2.

Все результаты моделирования были сравнены с помощью статистического анализа и вариографии. Суммарная статистика для исходных данных и смоделированных значений для выбранных реализаций представлена в таблице 1. Она показывает, что оба подхода моделирования способны воспроизвести общее распределение. Все статистические характеристики (кроме коэффициента симметрии и эксцесса) воспроизведены в стохастических реализациях с отклонением порядка 5%.

Кроме вариограм, приведенных на рис. 7-8, для исследования результатов применялись вариограммы розы (рис. 9-11). Они показывают воспроизводство анизотропной пространственной корреляционной структуры. Реализации последовательного Гауссова моделирования лучше воспроизводят анизотропию вариограммы, в то время как реализации моделирования отжигом лучше воспроизводят пространственную вариабельность, характеризующуюся плато вариограммы.

4 Выводы

Проведено стохастическое моделирование загрязнения донных отложений Женевского озера с помощью

- последовательного Гауссового моделирования
- моделирования отжигом

Проведено исследование модельных параметров и анализ чувствительности:

- 1) Методы стохастического моделирования позволили

- промоделировать пространственную вариабельность
 - построить карты вероятности
- 2) Оба метода дают сравнимые по качеству результаты. Отличия:
- моделирование отжигом дает более устойчивое воспроизводство вариограмм и более высокую вариабельность результатов
 - последовательное Гауссово моделирование точно воспроизводит анизотропию пространственной корреляционной структуры.

Перспективы продолжения данной работы заключаются в многопеременном стохастическом моделировании на этих же данных, т.к. использование большого количества вторичной информации может улучшить результаты.

Таблица 1. Итоговая статистика для исходных данных и реализаций Гауссового моделирования и аннилинга.

	Zn	SASIM 1	SASIM 2	E-type	SGSIM 1	SGSIM 2	E-type
Среднее значение	114.64	113.28	113.74	114.56	113.69	112.22	111.97
Стандартное отклонение	30.83	29.62	29.25	25.72	28.91	28.85	24.62
Минимум	54	54	54	54	28.42	14.19	54
Первая квартиль	96.5	89.09	93	90.81	97	95	95.96
Медиана	120	120	120	120	120	110	113.43
Вторая квартиль	130	130	133.8	134.33	130	130	124.05
Максимум	320	320	320	320	344.27	332.57	320
Козф. симметрии	1.04	0.26	0.20	0.20	1.63	1.33	0.74
Экссесс	7.28	0.45	0.60	1.16	1.11	7.84	3.36

5 Благодарности

Работа проводилась при частичной поддержке гранта ИНТАС 97-31726.

6 Литература

1. Pierre Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford University Press. 1997.
2. Clayton V. Deutsch, Andre G. Journel. *GSLIB Geostatistical Software Library and User's Guide*. Oxford University Press. 1998.
3. Oliver Perrin, Serge Iovleff. *Estimation a non-stationary spatial structure using simulated annealing*. GeoComputation 99, http://www.geovista.psu.edu/geocomp/geocomp99/Gc99/028/gc_028.htm

4. М.Ф. Каневский, В.В. Демьянов. *Непараметрическая геостатистика, стохастическое моделирование и анализ радиоэкологических данных*. Препринт ИБРАЭ, IBRAE-95-10, Москва 1994.
5. М.Ф. Каневский, В.В. Демьянов, С.Ю. Чернов. *Совместный пространственный анализ ^{137}Cs и ^{90}Sr черновыльских выпадений*. Препринт ИБРАЭ, IBRAE-96-04, Москва 1996.
6. В.В. Демьянов, М.Ф. Каневский, Е.А. Савельева, В.А. Тимонин, С.Ю. Чернов. *Совместное стохастическое моделирование пространственной вариабельности и неопределенности ^{90}Sr и ^{137}Cs* . Препринт ИБРАЭ, IBRAE-98-01, Москва 1998.
7. М. Kanevsky. *Chernobyl fallout: stochastic simulations of the spatial variability and probabilistic mapping*. Препринт ИБРАЭ, NSI-23-94, Москва 1994.
8. М. Kanevsky. *Artificial networks and spatial interpolations. Case study: chernobyl fallout*. Препринт ИБРАЭ, IBRAE-95-07, Москва 1994.
9. В.А. Тимонин, В.В. Демьянов, М.Ф. Каневский. *Обзор и применение методов нелинейной оптимизации для обучения ИНС многослойный перцептрон*. Препринт ИБРАЭ, IBRAE-99-09, Москва 1999.
10. Каневский М.Ф., Демьянов В.В., Савельева Е.А., Чернов С.Ю., Тимонин В.А. *Общие проблемы охраны окружающей среды*. ВИНТИ, 1999.
11. Gomez-Hernandez, J. *A Stochastic Approach to the Simulation of Block Conductivity Fields Conditioned upon Data Measured at a Smaller Scale*. Doctoral dissertation, Stanford University, Stanford, CA. 1991.
12. Tran, T. *Improving variogram reproduction on dense simulation grids*. Computers & Geosciences, 20(7):1161-1168. 1994.
13. E. Aarts and J. Korst. *Simulated Annealing and Boltzmann Machines*. John Wiley & Sons, New York, 1989.
14. Journel, A.G. and F.G. Alabert. *Focusing on spatial connectivity of extreme valued attributes: Stochastic indicator models of reservoir heterogeneities*. SPE paper # 18324. 1988.

7 Приложение

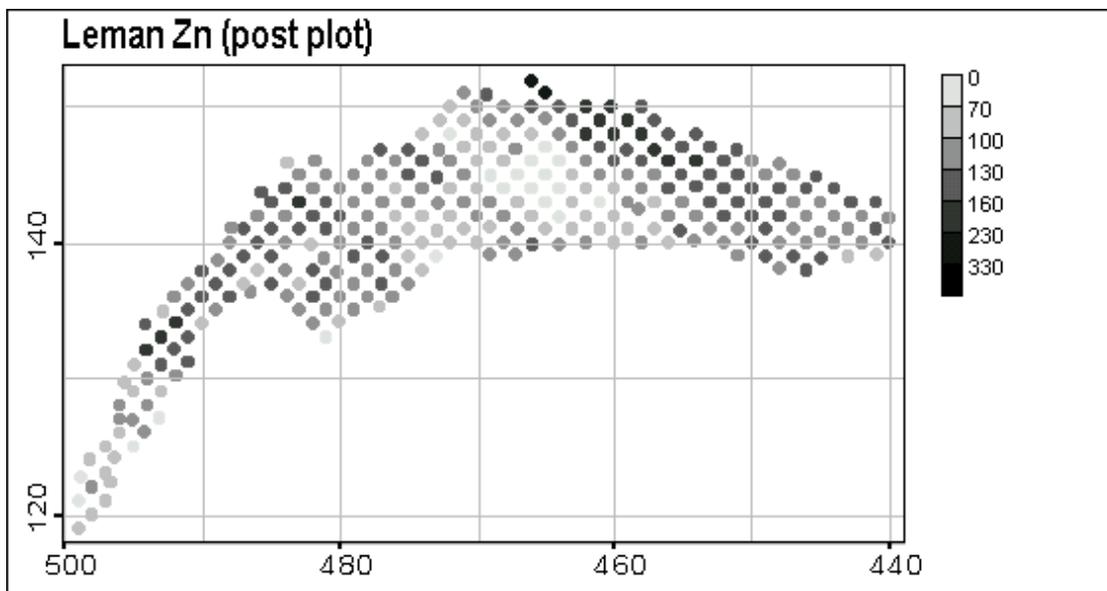


рис.1 Данные по загрязнению Женевского озера (Zn)

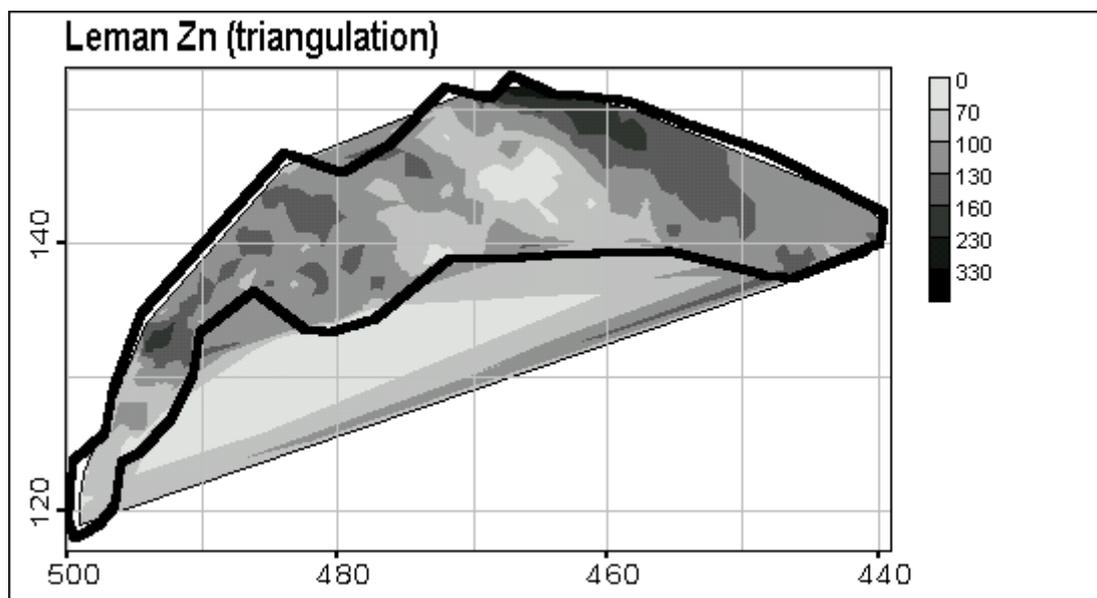


рис.2 Женевское озеро. Изолинии по триангуляция (Zn)

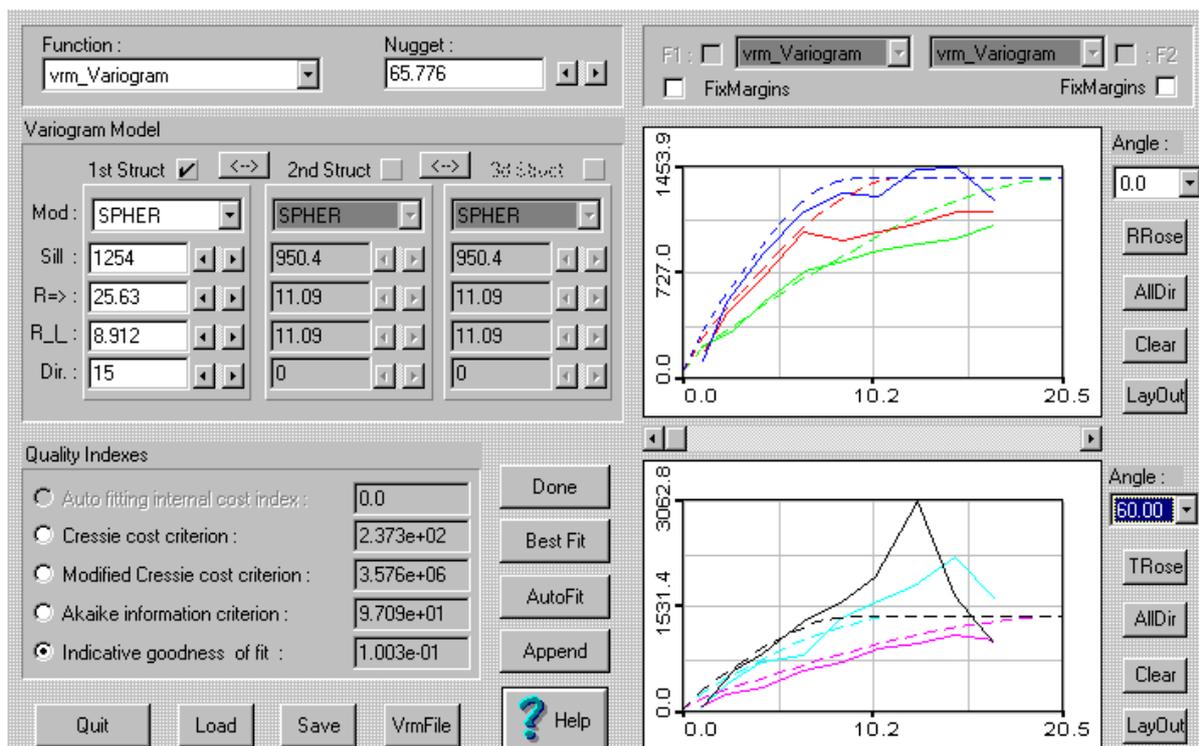


рис.3 Фиттинг вариограммной модели для исходных данных

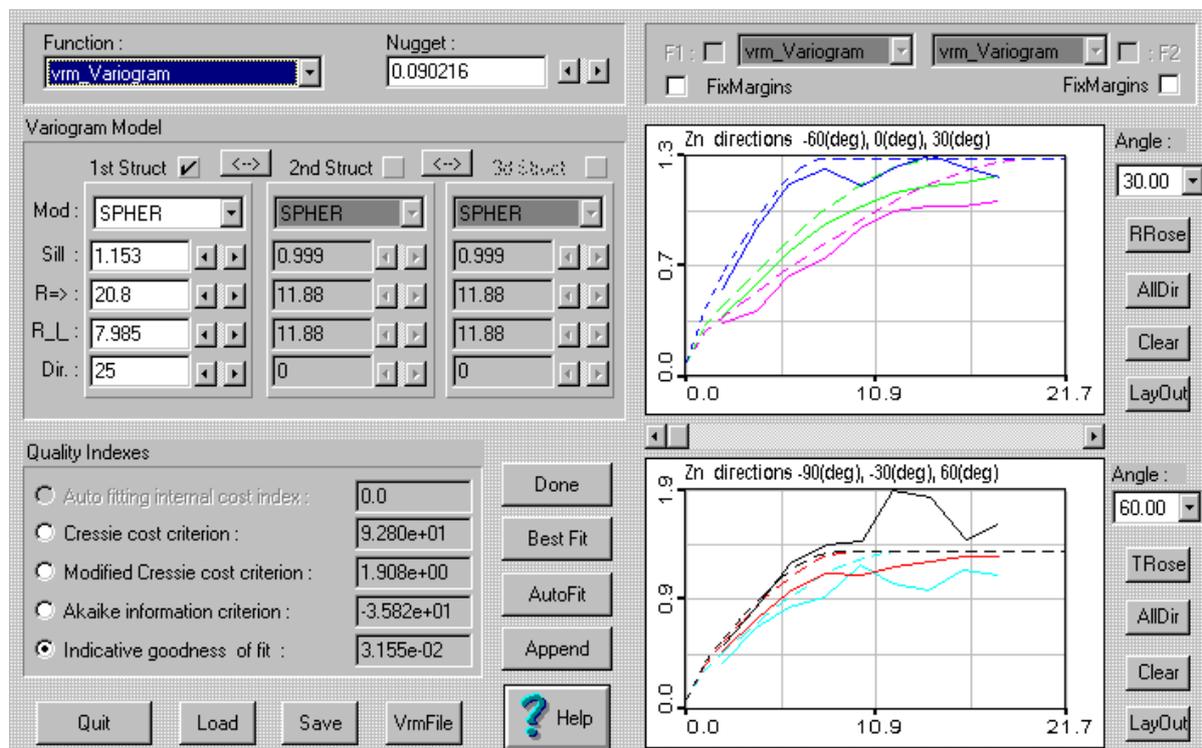


рис.4 Фиттинг вариограммной модели для нормально-преобразованных данных

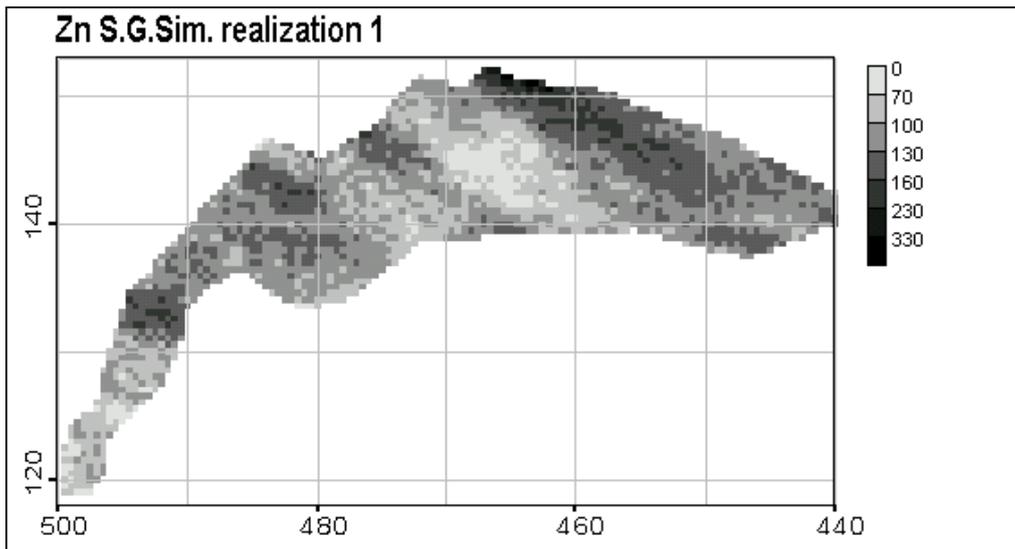


рис.5a Женевское озеро. Реализация последовательного Гауссового моделирования

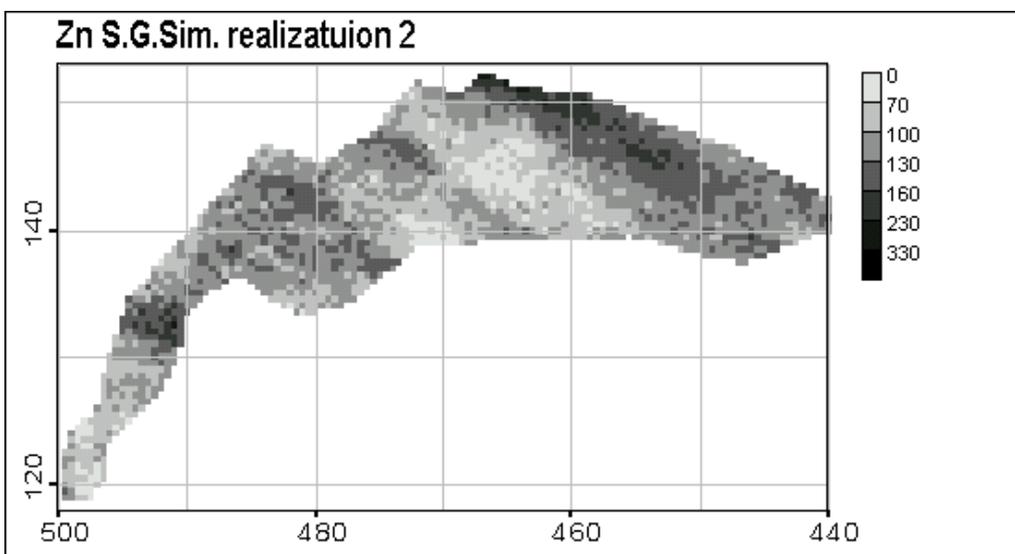


рис.5b Женевское озеро. Реализация последовательного Гауссового моделирования

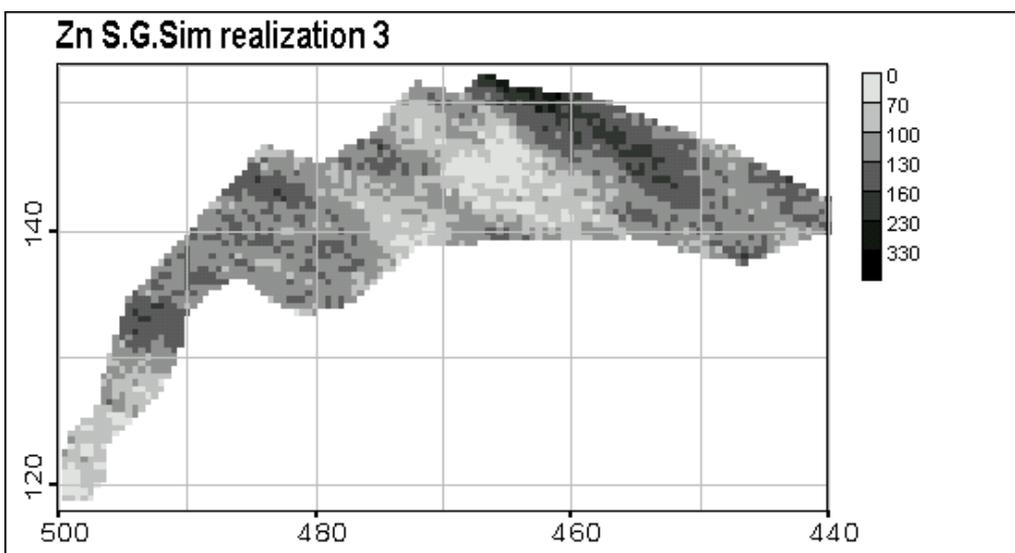


рис.5c Женевское озеро. Реализация последовательного Гауссового моделирования

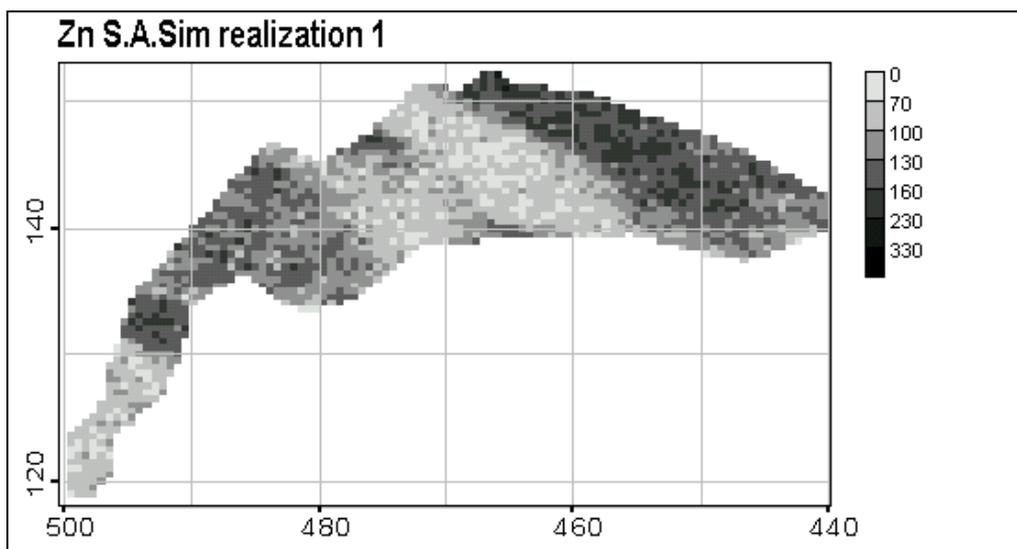


рис.6а Женевское озеро. Реализация моделирования отжигом

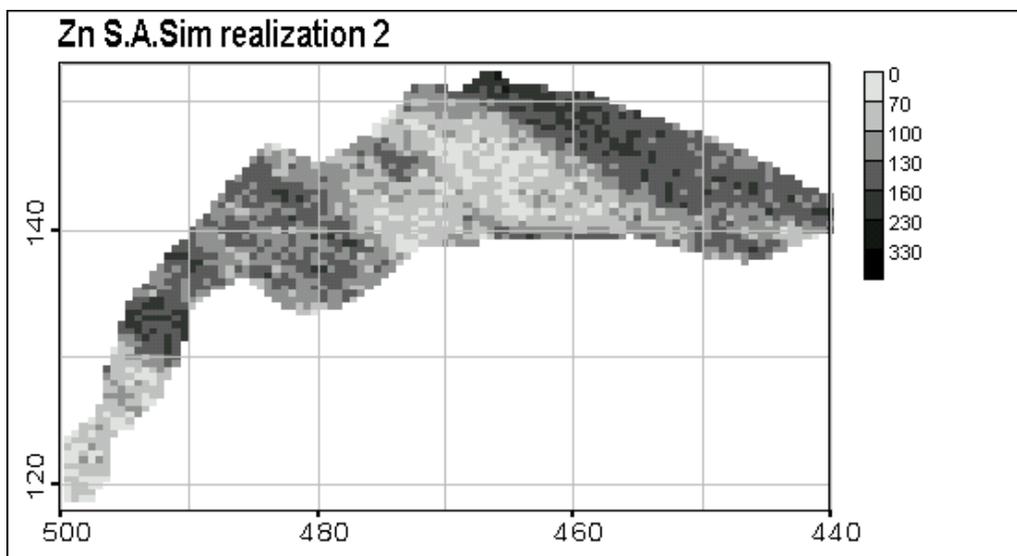


рис.6б Женевское озеро. Реализация моделирования отжигом

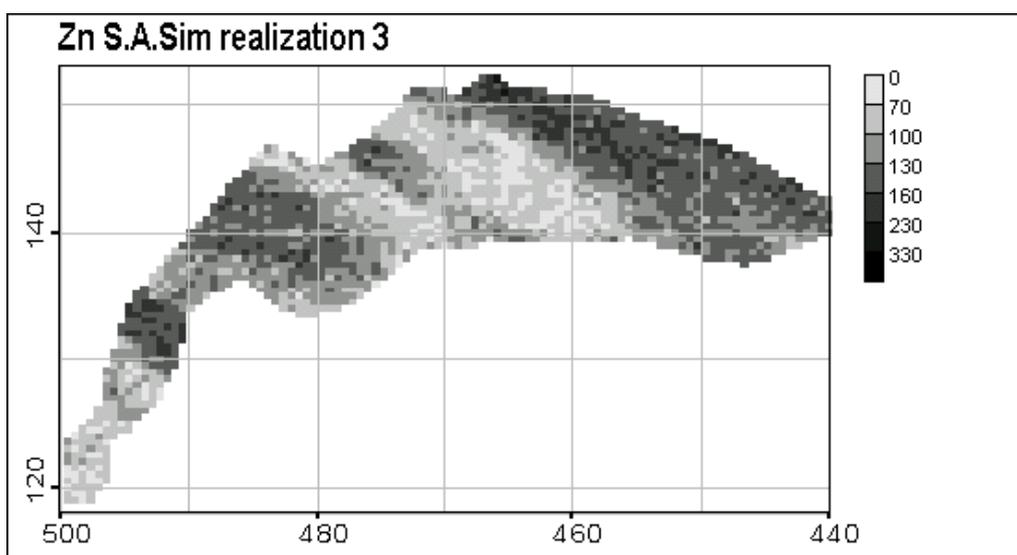


рис.6с Женевское озеро. Реализация моделирования отжигом

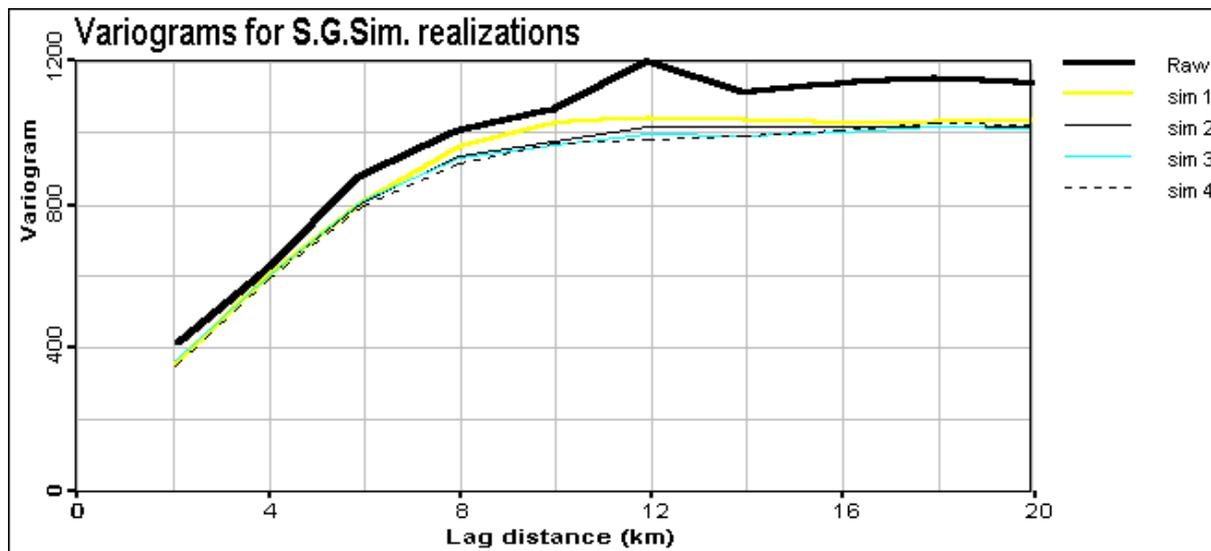


рис.7 Женевское озеро. Вариограммы по всем направлениям для исходных данных и SGSIM реализаций

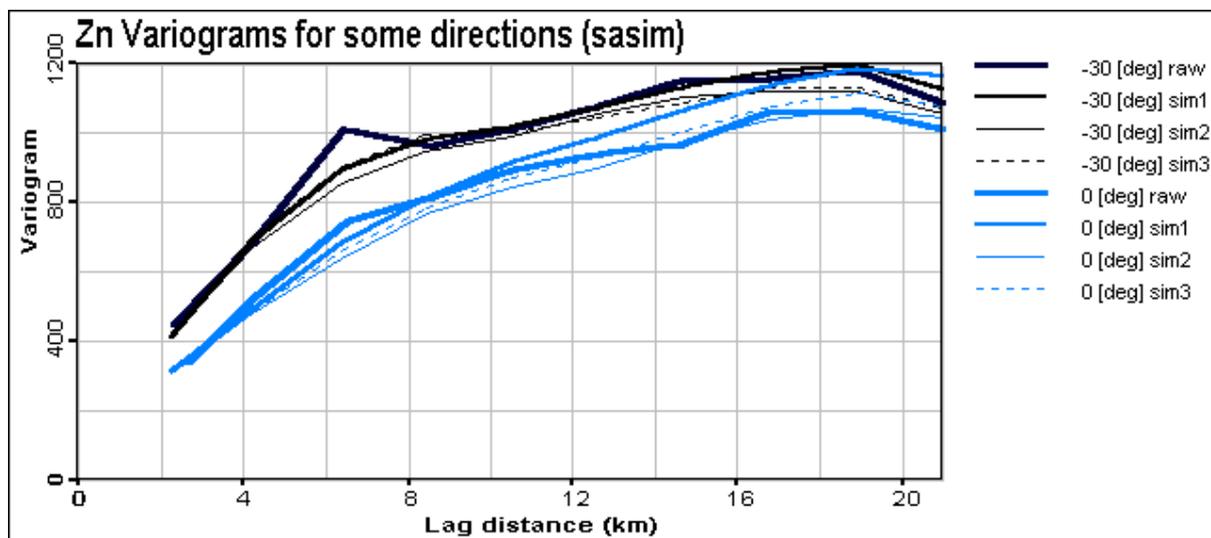


рис.8 Женевское озеро. Вариограммы по двум направлениям для исходных данных и SASIM реализаций

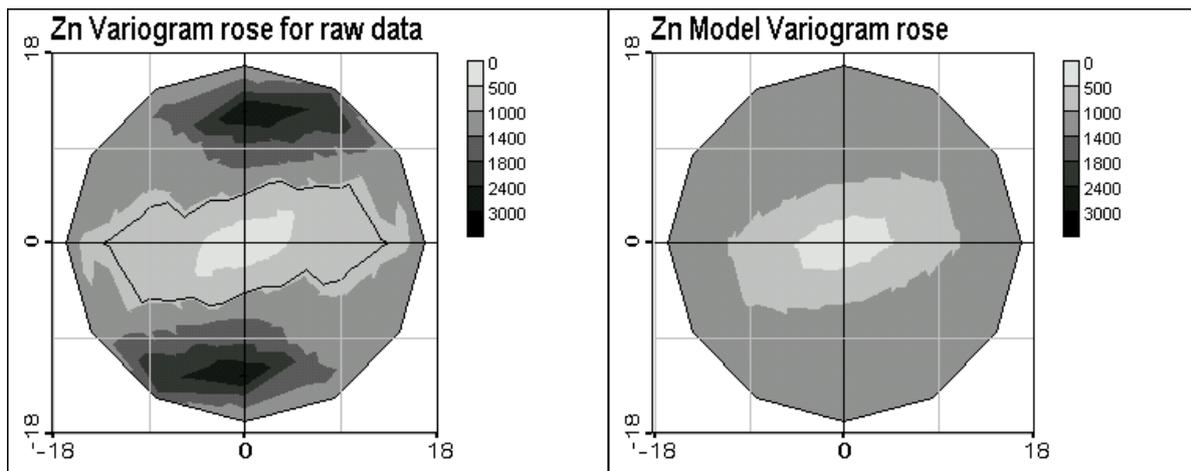


рис.9 Женевское озеро. Вариограммы розы для исходных данных и модели

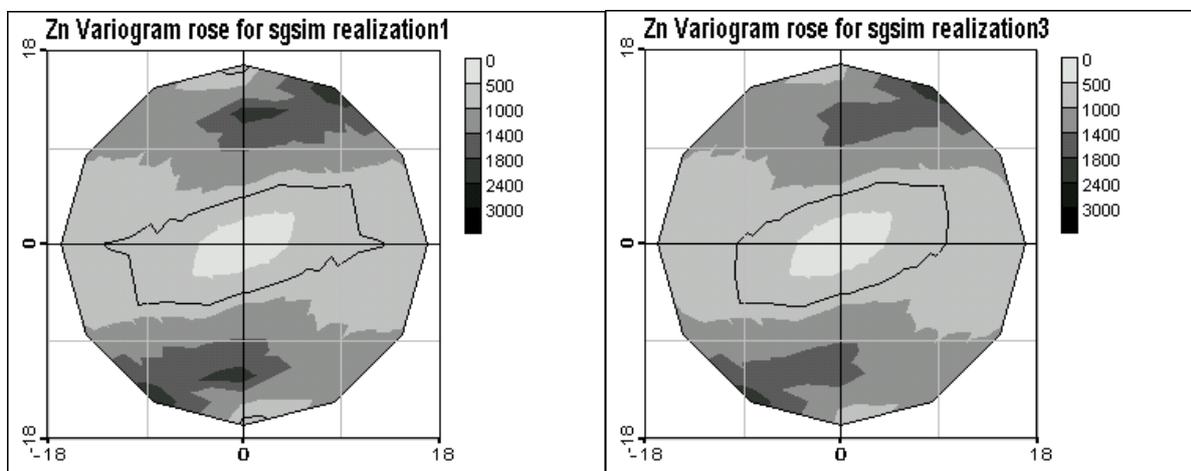


рис.11 Женевское озеро. Вариограммы розы для реализаций последовательного Гауссового моделирования

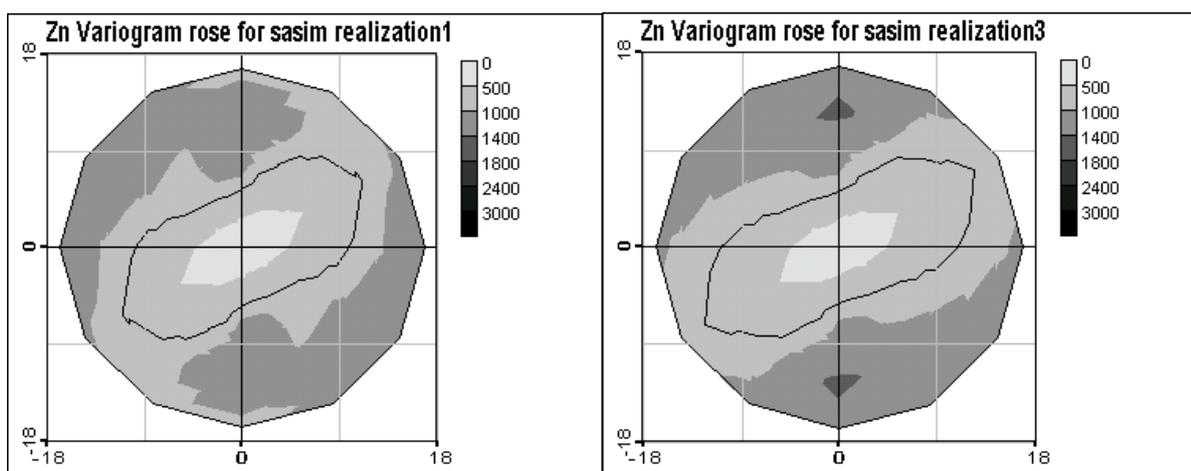


рис.10 Женевское озеро. Вариограммы розы для реализаций моделирования отжигом

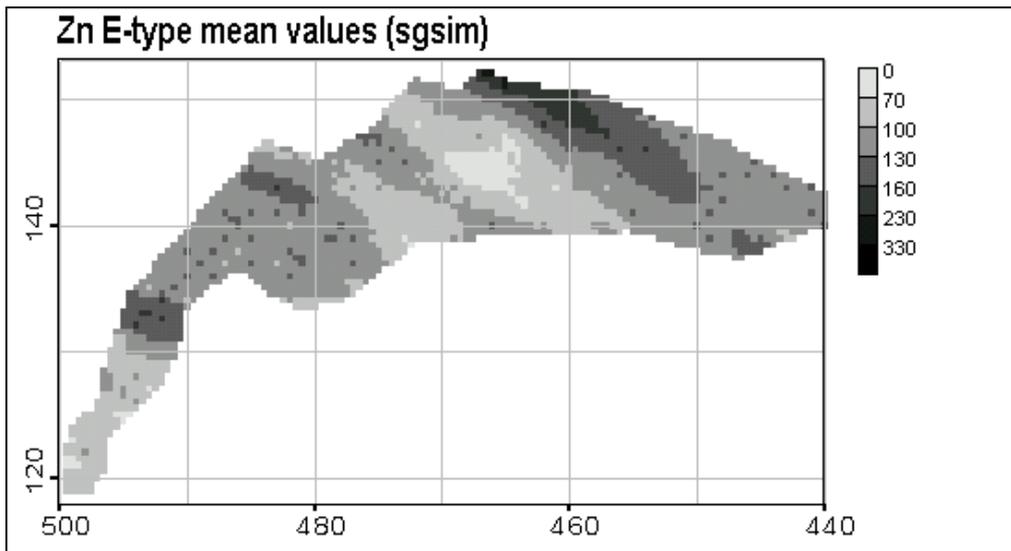


рис.12 Оценка E-типа для 30-ти реализаций SGSIM

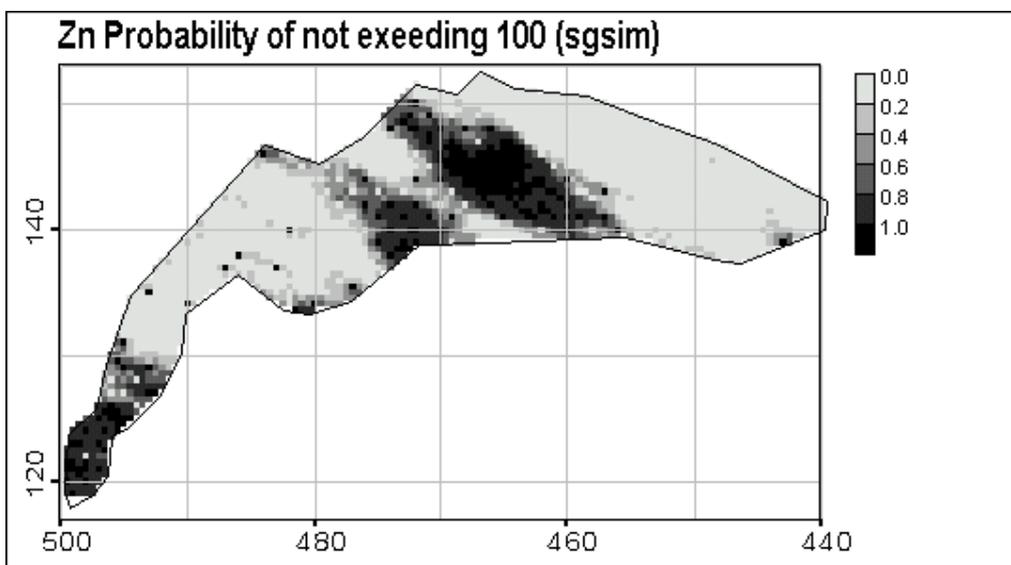


рис.13 Женевское озеро. Вероятностная карта SGSIM реализаций

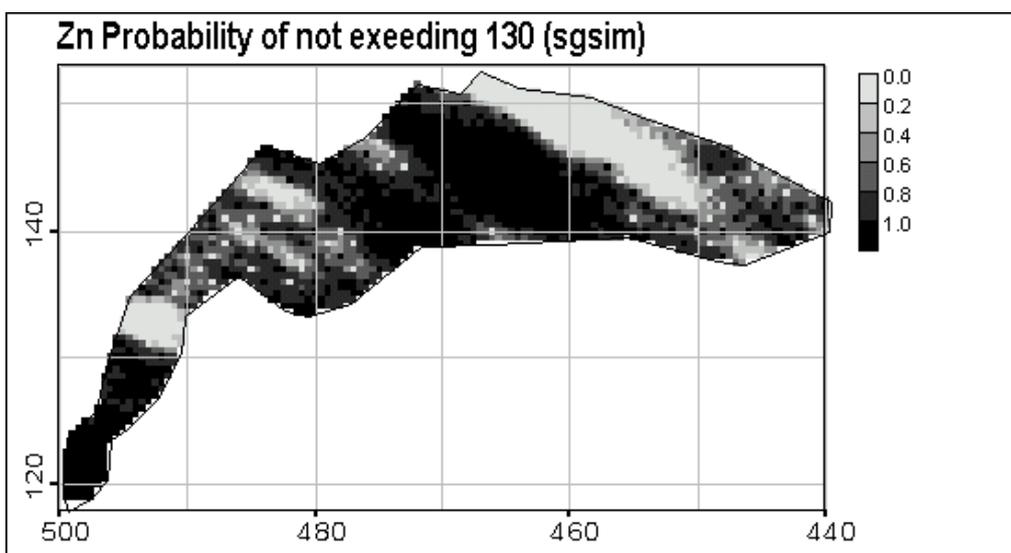


рис.14 Женевское озеро. Вероятностная карта SGSIM реализаций

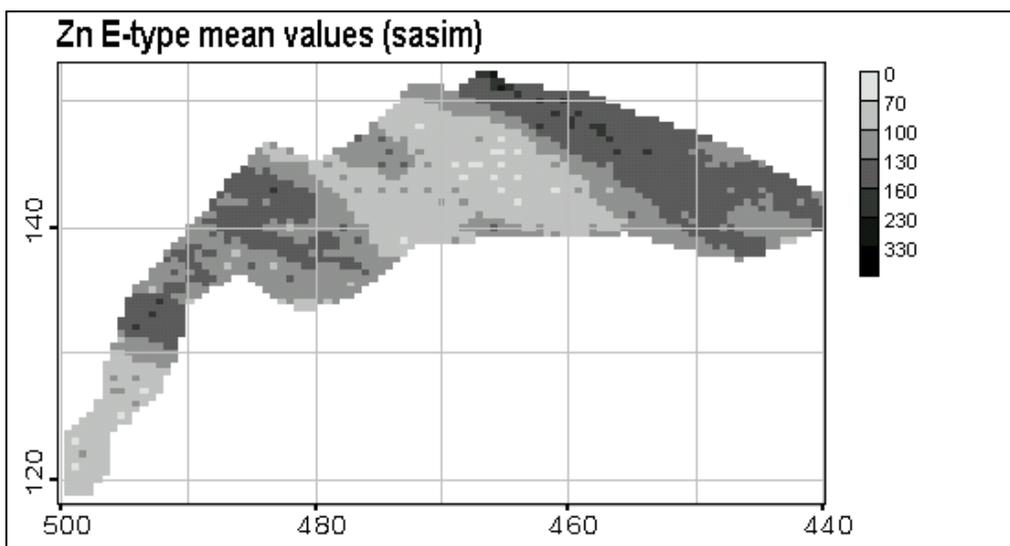


рис.15 Оценка E-типа для 30-ти реализаций SASIM

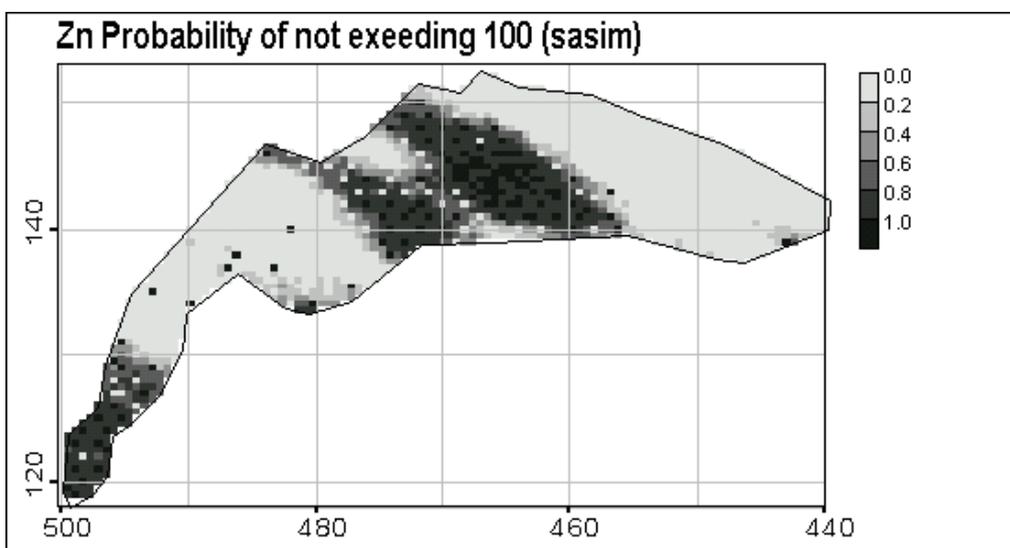


рис.16 Женевское озеро. Вероятностная карта SASIM реализаций

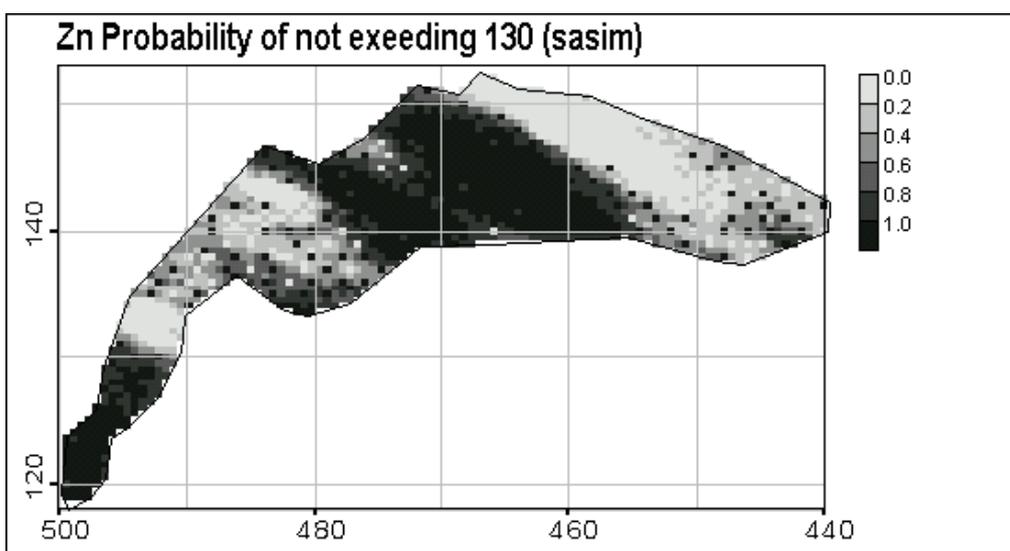


рис.17 Женевское озеро. Вероятностная карта SASIM реализаций

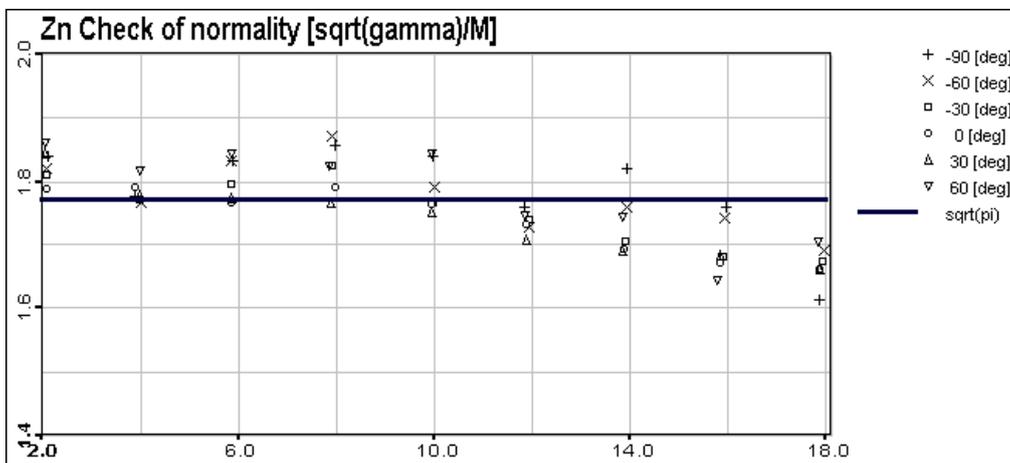


рис.18а Женевское озеро. Проверка нормальности распределения данных

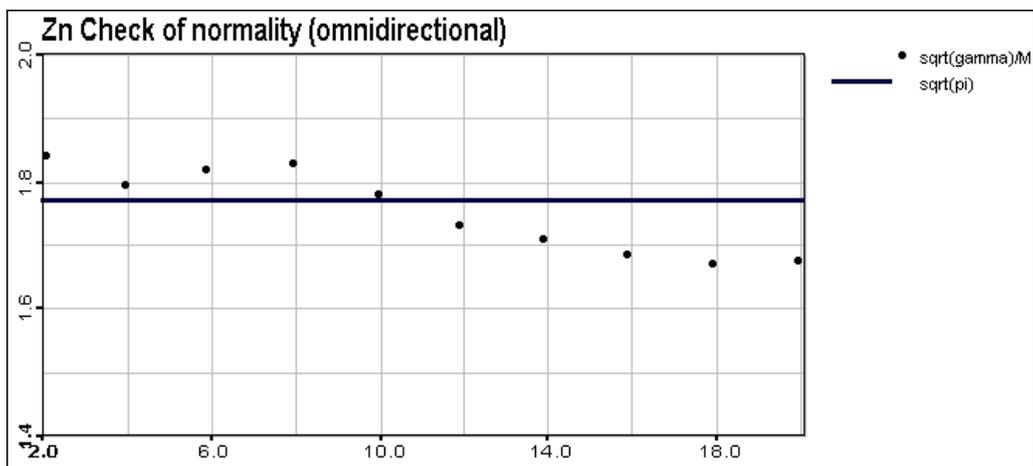


рис.18b Женевское озеро. Проверка нормальности распределения данных

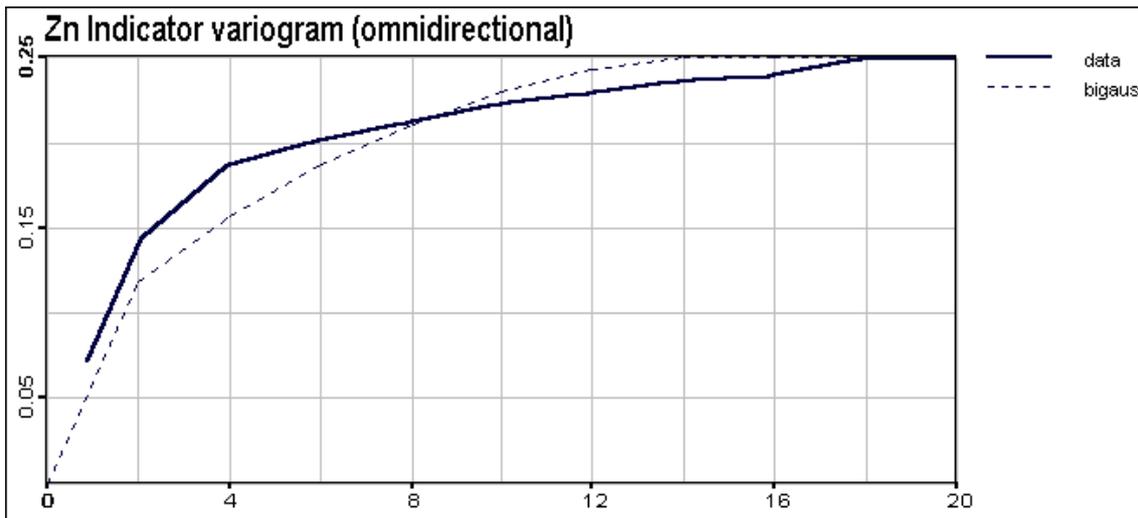


рис.19а Женевское озеро. Проверка на бинормальность. Индикаторные вариограммы по порогу 0.5

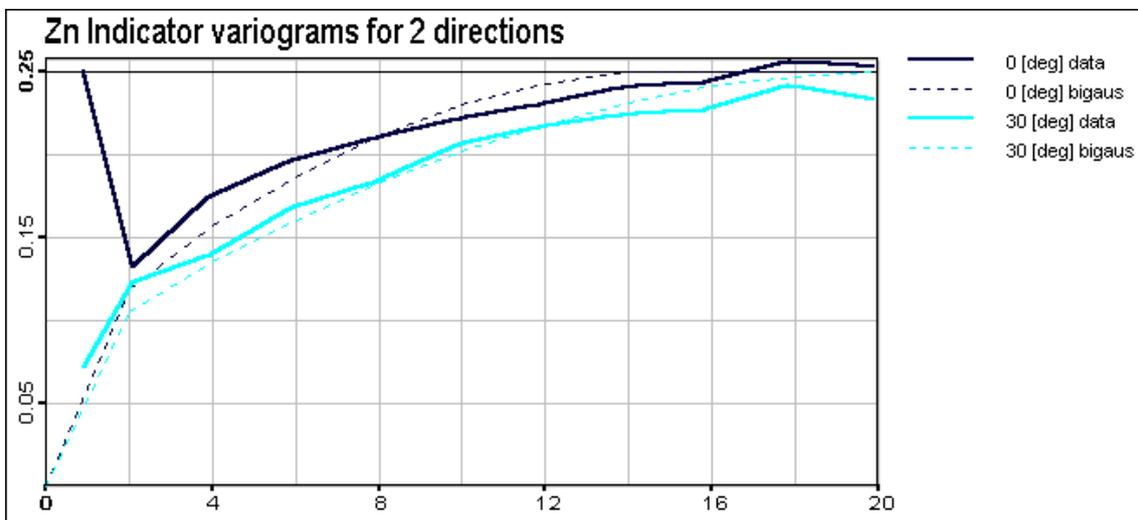


рис.19b Женевское озеро. Проверка на бинормальность. Индикаторные вариограммы по порогу 0.5

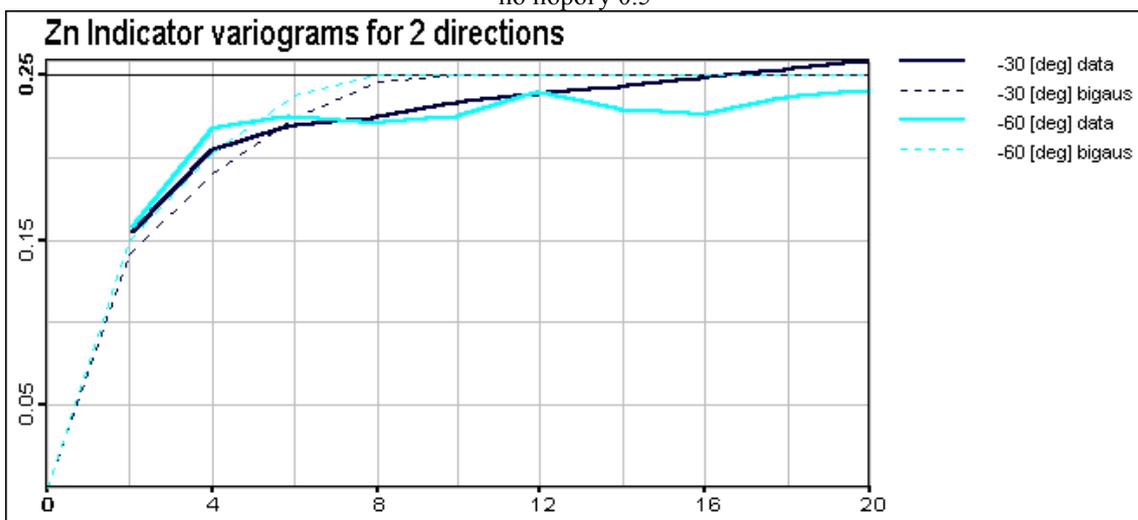


рис.19с Женевское озеро. Проверка на бинормальность. Индикаторные вариограммы по порогу 0.5

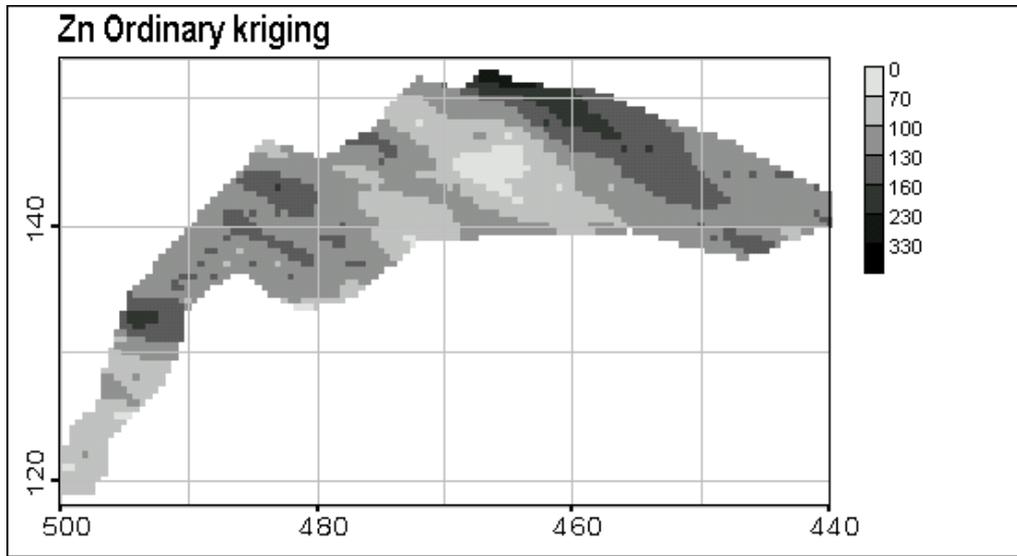


рис.20 Карта оценки обычного кригинга

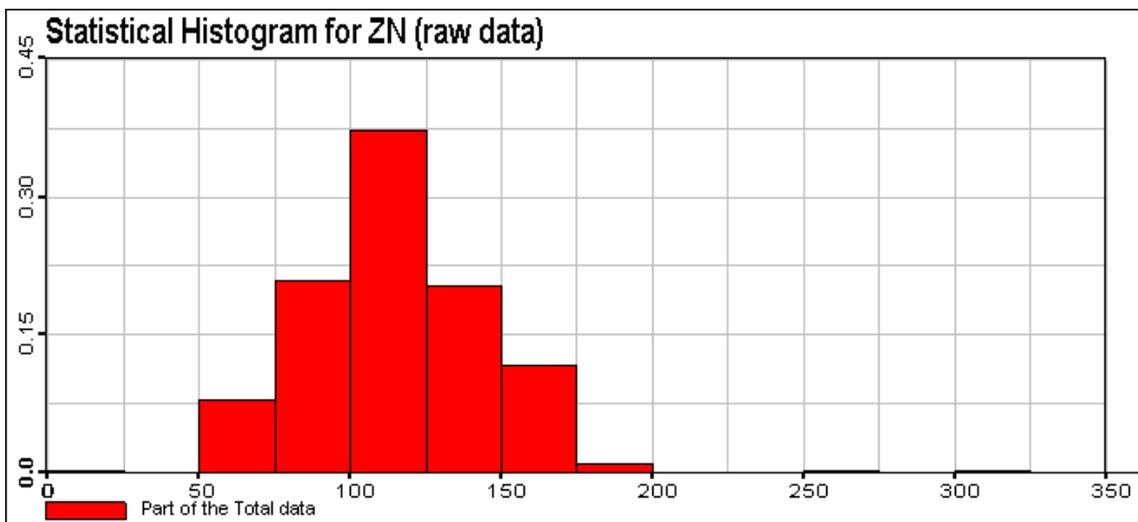


рис. 21 Гистограмма распределения данных по загрязнению Женевского озера цинком (Zn)

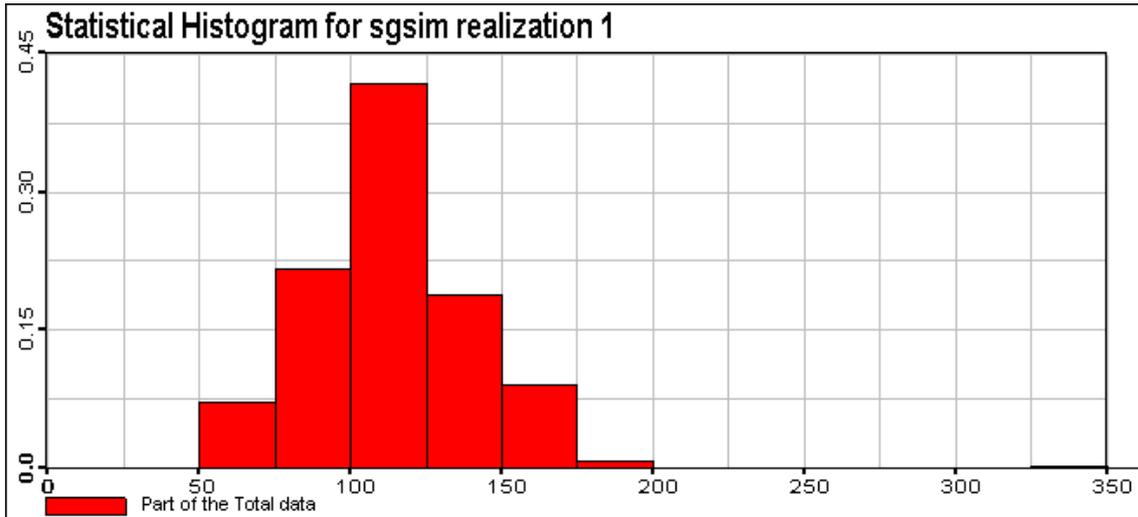


рис. 22а Женевское озеро. Гистограмма распределения значений для sgsim реализации.

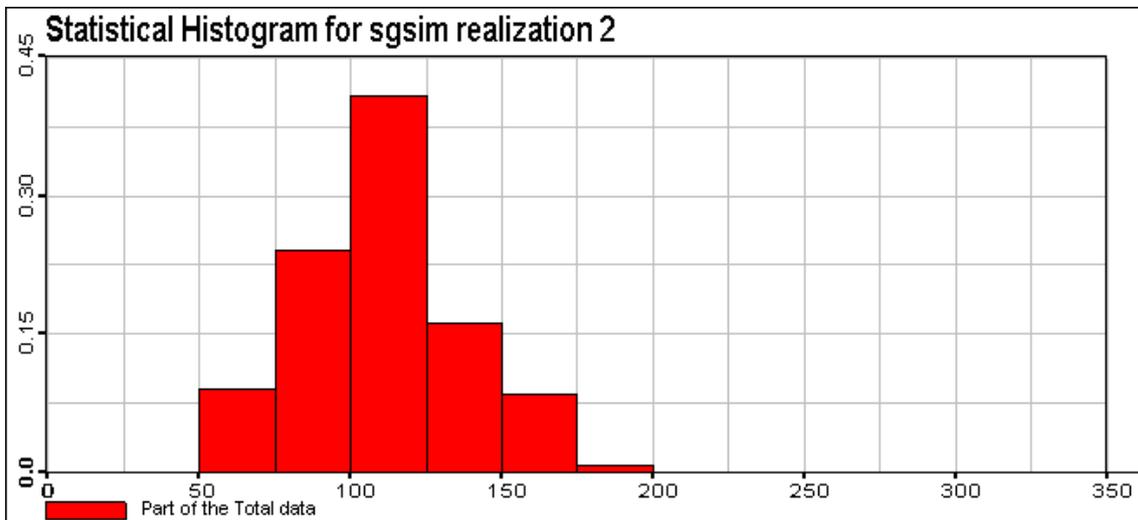


рис. 22b Женевское озеро. Гистограмма распределения значений для sgsim реализации.

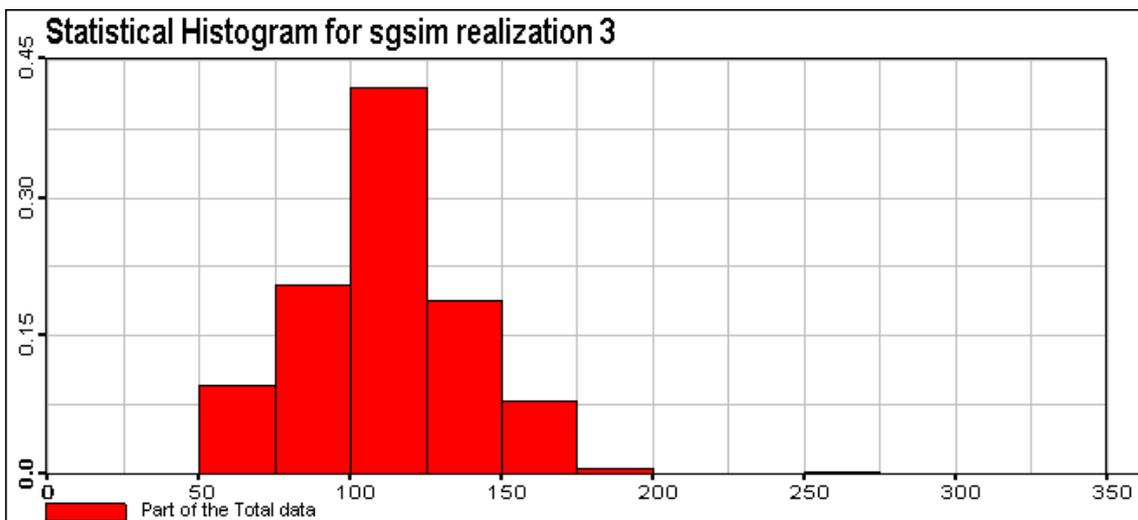


рис. 22с Женевское озеро. Гистограмма распределения значений для sgsim реализации.