

**Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего образования
Санкт-Петербургский горный университет**

Кафедра геоэкологии

ДИСТАНЦИОННЫЕ И ГИС-ТЕХНОЛОГИИ В ГЕОЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

*Методические указания к самостоятельным работам
для студентов бакалавриата направления 05.03.06*

**САНКТ-ПЕТЕРБУРГ
2021**

УДК 504.064 (073)

ДИСТАНЦИОННЫЕ И ГИС-ТЕХНОЛОГИИ В ГЕОЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ: Методические указания к самостоятельным работам / Санкт-Петербургский горный университет. Сост. *А.В. Стриженок*. СПб, 2021. 37 с.

Самостоятельная работа по дисциплине «Дистанционные и ГИС-технологии в геоэкологических исследованиях» направлена на формирование у студентов навыков обобщения знаний, полученных на лекциях и лабораторных занятиях, и применения этих знаний для самостоятельного решения практических задач.

В результате освоения материала, предназначенного для самостоятельной проработки, студенты должны сформировать и продемонстрировать следующие профессиональные компетенции: готовность использовать научные законы и методы при оценке состояния окружающей среды в сфере функционирования производств по эксплуатационной разведке, добыче и переработке твердых полезных ископаемых, а также при строительстве и эксплуатации подземных объектов; пользование компьютером как средством управления и обработки информационных массивов; владение методами геохимических и геофизических исследований, общего и геоэкологического картографирования, обработки, анализа и синтеза полевой и лабораторной геоэкологической информации.

В методических указаниях представлены основы создания моделей пространственных переменных средствами программного продукта *Golden Software Surfer*, а также приведено подробное описание методов математической аппроксимации, которые используются в указанной программной оболочке при построении тематических поверхностей.

Предназначены для студентов направления подготовки 05.03.06 «Экология и природопользование», профиль подготовки «Природопользование».

Научный редактор проф. *М.А. Пашкевич*

Рецензент канд. техн. наук *Н.И. Горошкова* (Государственный гидрологический институт)

ВВЕДЕНИЕ

Целью освоения дисциплины «Дистанционные и ГИС-технологии в геоэкологических исследованиях» является получение студентами знаний о методах и средствах дистанционного зондирования земли (ДЗЗ), методиках дешифрирования и обработки данных дистанционного зондирования природно-территориальных комплексов, в том числе с использованием современных ГИС-оболочек.

Целью самостоятельной работы является систематизация, закрепление и расширение теоретических знаний в области использования ГИС в геоэкологических исследованиях, приобретение навыков практического применения этих знаний для решения конкретных научных, инженерных и производственных задач.

Материал для самостоятельной проработки, приведенный в настоящих методических указаниях, представляет собой теоретические основы численного описания пространственно-распределенных явлений на основе наблюдений в отдельных точках средствами программного продукта *Golden Software Surfer*, описание основных методов математической аппроксимации, использующихся программой для построения моделей пространственных переменных, а также алгоритм создания картографических тематических поверхностей в программе *Surfer*.

Самостоятельное освоение материала, представленного в настоящих методических указаниях, осуществляется студентом на протяжении всего семестра, в котором изучается дисциплина, а проверка степени освоения материала проводится в рамках сдачи дифференцированного зачета по дисциплине «Дистанционные и ГИС-технологии в геоэкологических исследованиях» и при защите курсового проекта по данной дисциплине.

1. ПОСТРОЕНИЕ МОДЕЛЕЙ ПРОСТРАНСТВЕННЫХ ПЕРЕМЕННЫХ

Для описания поведения пространственно-распределенных явлений используется понятие пространственная переменная – это числовая переменная $Z = Z(X, Y)$, которая принимает значение в произвольной точке области исследования, где X, Y - координаты точки в пространстве (на плоскости). Например, содержание тяжелых металлов в верхнем слое почвы, высота рельефа местности, мощность геологического пласта могут рассматривать как пространственные переменные в двумерном пространстве.

Значения этих переменных, в силу различных причин, обычно известны лишь в некоторых точках исследуемой территории (рис. 1а). Эти точки называют точками выборки (точками измерений, точками наблюдений, постами, выборочными точками). Обозначим $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ – координаты точек выборки, n – объем выборки, а Z_1, Z_2, \dots, Z_n - значения изучаемого параметра в этих точках (выборочные данные, исходные данные, экспериментальные данные).

Задача пространственной аппроксимации – получить (с минимально возможной погрешностью) значения изучаемой пространственной переменной Z в произвольных точках области исследования на основе обработки и анализа ее значений, измеренных в ограниченном числе выборочных точек:

$$Z(X_0, Y_0) = Z_0 \Rightarrow \check{Z}_0 = F(Z_1, Z_2, \dots, Z_n),$$

где (X_0, Y_0) – координаты произвольной точки в области исследования, Z_0 – неизвестное значение изучаемой переменной Z в этой точке, \check{Z}_0 – интерполяционная оценка неизвестного значения Z_0 , получаемая на основе выборочных данных (Z_1, Z_2, \dots, Z_n) .

Для оценки значения переменной Z в точке, где она не измерена, применяют различные методы пространственной аппроксимации. Принято различать два основных подхода к аппроксимации: детерминированный и геостатистический.

Методы детерминированной аппроксимации определяют неизвестную переменную параметрической функцией, чья форма задается либо явно (например, полином), либо неявно (условие минимальной кривизны). Параметры выбираются так, чтобы оптимизи-

ровать некоторый критерий наилучшего приближения в точках выборки (например, наименьшие квадраты, точное совпадение).

Геостатистические методы (кригинг) используют статистические свойства измеренных данных, оценивая пространственную автокорреляцию и учитывая ее при аппроксимации.

Для представления результатов пространственной аппроксимации используется прямоугольная регулярная сетка, в узлах которой вычисляется значение изучаемой переменной. То есть в качестве точки (X_0, Y_0) последовательно рассматриваются узлы этой сетки. Такое представление называют моделью пространственной переменной или пространственной моделью (рис. 1б). Расстояние между узлами сетки задают существенно меньше характерного размера изучаемого пространственного параметра, а сама сетка должна полностью покрывать территорию исследования.

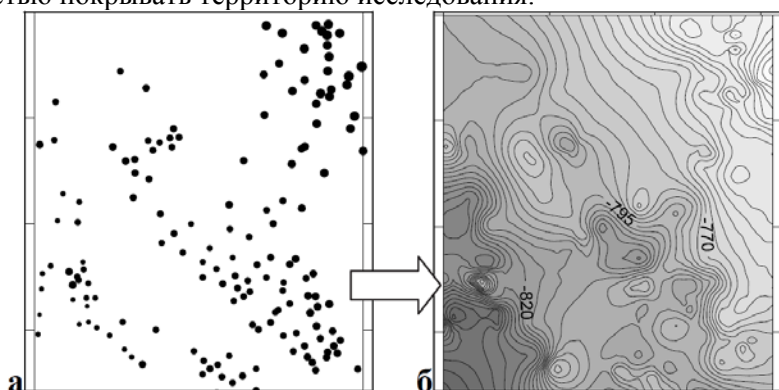


Рис. 1. Выборочные данные (а) и пространственная модель (б)

2. СОЗДАНИЕ ПРОСТРАНСТВЕННОЙ МОДЕЛИ В *SURFER*

2.1 ВВОД ИСХОДНЫХ ДАННЫХ

В программе *Surfer* для построения модели пространственной переменной нужно выбрать пункт меню «*Grid*», подпункт «*Data*», после чего появляется диалоговое окно, предлагающее выбрать файл с исходными данными.

Исходными данными для построения модели являются измеренные значения переменной Z в точках с известными координатами – точках выборки. Эти данные должны быть записаны в табличной

форме и храниться в текстовом файле в кодировке *ASCII*. Таблица с данными должна иметь три или более колонок. В двух колонках записываются координаты X и Y выборочных точек: (X_1, Y_1) , (X_2, Y_2) , ..., (X_n, Y_n) , остальные колонки содержат измеренные (наблюдаемые) в этих точках значения одной или более пространственных переменных. Одна строка таблицы соответствует одной точке наблюдений. Первая строка может являться заголовком и содержать имена колонок.

После выбора файла с данными появляется диалоговое окно, которое имеет несколько интерфейсных элементов управления: кнопок, переключателей и строк ввода, объединенных в различные блоки (рис. 2). В первом блоке «*Data Columns*» в выпадающих списках « X », « Y », « Z » необходимо указать, в каких колонках входного файла находятся значения координат X , Y точек выборки и в какой колонке значения изучаемой переменной Z , измеренные в этих точках.

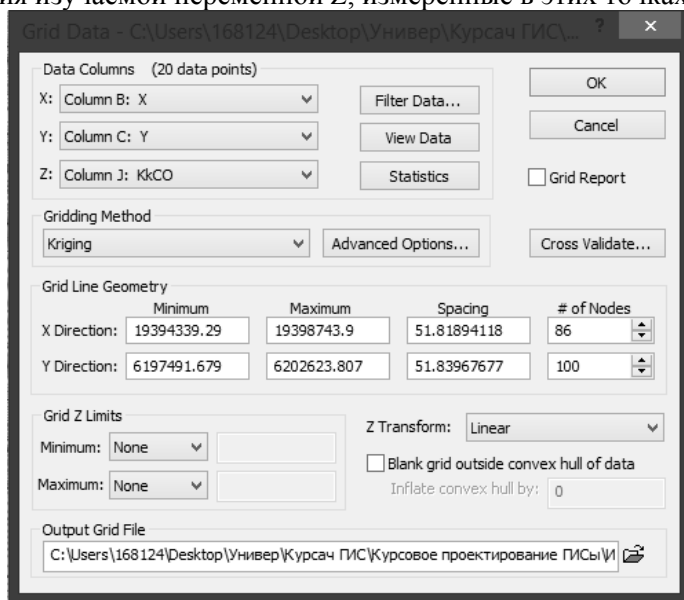


Рис. 2. Окно для задания входных данных, метода аппроксимации и параметров выходной сетки

В этом же блоке расположены еще три кнопки: «*Filter Data...*», «*View Data*», «*Statistics*». При нажатии на кнопку «*Filter Data...*» появляется диалоговое окно, изображенное на рис. 4.

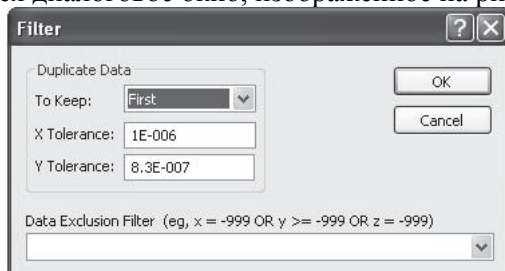


Рис. 3. Окно для задания условий исключения данных и правила обработки дублирующихся данных

Здесь в строке «*Data Exclusion Filter*» можно задать условие исключения части данных из расчетов (исключены будут данные, удовлетворяющие условию).

Кроме этого, блок элементов управления «*Duplicate Data*» позволяет задать правило обработки дублирующихся в пространстве данных. Например, если в выпадающем списке «*To Keep*» выбран пункт (по умолчанию) «*First*», то при вычислении оценок будет использоваться значение Z первой из дублирующихся точек выборки, а если выбрать опцию «*Average*», то будет использоваться среднее значений Z всех дублирующихся точек. Строки ввода «*X Tolerance*», «*Y Tolerance*» позволяют задать минимальное расстояние (по осям координат) между точками. Точки, расположенные на меньшем расстоянии, будут считаться дублирующимися.

При нажатии на кнопку «*View Data*» появляется возможность просмотра таблицы исходных данных, а при нажатии на кнопку «*Statistics*» генерируется окно с отчетом, в котором отображаются статистические показатели, рассчитанные по выборкам-столбцам входного файла.

2.2 ЗАДАНИЕ ГЕОМЕТРИЧЕСКИХ ПАРАМЕТРОВ ВЫХОДНОЙ СЕТКИ

Как уже говорилось, для представления результатов пространственной аппроксимации используется прямоугольная регу-

лярная сетка, полностью покрывающая исследуемую территорию. Блок «*Grid Line Geometry*» (рис. 2) позволяет задать геометрические параметры сетки, в узлах которой будут рассчитываться значения переменной Z :

- координаты X и Y левого нижнего угла (в столбце «Minimum»);

- координаты X и Y правого верхнего угла (в столбце «Maximum»);

- шаг сетки по X и по Y – на каком расстоянии будут отстоять друг от друга узлы регулярной сетки (в столбце «*Spacing*»).

Здесь же в столбце «*# of Lines*» вычисляется количество узлов регулярной сетки.

В блоке «*Output Grid File*» отображается имя файла формата *.*grd*, в котором будут сохранены результаты расчетов в узлах сетки - растровая модель. Имя файла и путь можно сменить, нажав на кнопку.

2.3 ВЫБОР МЕТОДА АППРОКСИМАЦИИ

В программе *Surfer* представлено 12 методов пространственной аппроксимации. Эти методы и их параметры, а также рекомендации по использованию будут описаны ниже. Выбор метода производится в блоке «*Gridding Method*» (рис. 2), который состоит из выпадающего списка с названиями методов аппроксимации и кнопки «*Advance Options*» для задания параметров выбранного метода.

Выбранный алгоритм пространственной аппроксимации запускается после нажатия на кнопку «*OK*». В правой части основного диалогового окна (рис. 2) расположен переключатель «*Grid Report*», который дает возможность сформировать текстовый отчет о проделанной процедуре интерполяционного оценивания.

2.4 ПЕРЕКРЕСТНАЯ ПРОВЕРКА

Кнопка с названием «*Cross Validate...*» вызывает диалоговое окно, которое позволяет задать параметры, так называемой, «перекрестной проверки» (рис. 4).

Суть данной проверки заключается в том, что из исходного набора данных случайным образом удаляется одно наблюдение Z_i , и затем, используя оставшиеся данные и выбранный алгоритм аппрок-

симации, рассчитывается значение \check{Z}_i в точке этого наблюдения. Ошибка аппроксимации вычисляется как разность между интерполированным значением и наблюдаемым значением: $\check{Z}_i - Z_i$. Этот процесс повторяется заданное число m раз и генерирует m ошибок аппроксимации. На основе анализа этих ошибок можно сделать выводы о точности аппроксимации. Перекрестную проверку можно считать объективным способом оценки качества методов аппроксимации, а также использовать для сравнения относительного качества двух или более методов.

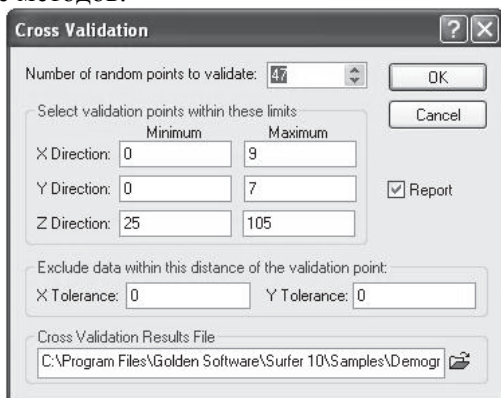


Рис. 4. Окно для задания условий перекрестной проверки

Диалоговое окно «*Cross Validation*» (рис. 4) имеет четыре блока элементов управления. Первая строка ввода «*Number of random points to validate*» задает число m случайных точек из исходного набора, на которых будет проверена модель. Второй блок строк ввода «*Select validation points within these limits*» определяет условия отбора точек проверки: точки за указанными пределами не будут участвовать в перекрестной проверке. Блок «*Exclude data within this distance of the validation point*» необходим для задания буферной зоны вокруг точек проверки в том случае, если исходные данные группируются в кластеры. Точки выборки, попадающие внутрь этих буферных зон, не участвуют при вычислении значений (аппроксимации) в проверочных точках, тем самым удается избежать влияния кластеризации на результат. В строке ввода «*Cross validation Results*

File» задается имя файла, в который будет помещен результат перекрестной проверки.

3. МЕТОДЫ ПРОСТРАНСТВЕННОЙ АППРОКСИМАЦИИ

Каждый метод аппроксимации имеет свой собственный набор параметров. Некоторые из параметров являются одинаковыми для различных методов. Прежде чем перейти к описанию методов, рассмотрим некоторые общие для них понятия.

Точные и сглаживающие интерполяторы. Все методы аппроксимации можно разделить на точные и сглаживающие интерполяторы.

Точные интерполяторы в точке выборки дают оценку, равную измеренному в ней значению, $\check{Z}_i = Z_i$. При этом нужно понимать, что даже при использовании точных интерполяторов, вполне возможно, что построенная на растровой сетке модель не будет воспроизводить значения в выборочных точках, если эти точки не точно совпадают с узлами сетки. Борьба с этим можно лишь путем уменьшения шага сетки (увеличения числа узлов).

Оценка методами сглаживающей аппроксимации, вычисленная в точке выборки, не совпадает с измеренным в ней значением, $\check{Z}_i \neq Z_i$. Сглаживающие интерполяторы уменьшают эффект локальной изменчивости. Этот тип аппроксимации используется в том случае, когда нет уверенности в точности измерения данных - изучаемая величина измерена в выборочных точках с некоторой погрешностью.

Некоторые точные интерполяторы имеют дополнительный параметр, называемый параметром сглаживания, использование которого превращает их в сглаживающие интерполяторы.

Взвешенное усреднение. Методы, реализованные в программе Surfer, в основном используют алгоритмы средневзвешенной аппроксимации. В общем случае оценка вычисляется как:

$$\check{Z}_i = \sum_{j=1}^n w_{j0} \cdot Z_j, \text{ где}$$

w_{j0} – вес i -ой точки данных при оценке значения Z в точке (X_0, Y_0) .

Это означает, что, при прочих равных условиях, чем ближе выборочная точка к точке, где делается оценка, тем больше ее вклад

в эту оценку Z . Разница между методами заключается в том, как весовые коэффициенты вычисляются и применяются к точкам данных.

Окрестность поиска. Методы аппроксимации могут работать как глобальные и как локальные интерполяторы. Глобальные интерполяторы вычисляют оцениваемые значения на основе всего набора данных. Локальные интерполяторы используют исходные данные лишь из некой локальной окрестности интерполируемой точки, которая является меньшей пространственной областью внутри большей изучаемой территории. Глобальная интерполяция делает модельную поверхность более гладкой, локальная позволяет выявлять локальные особенности.

В программе *Surfer* задание локальной окрестности (или окрестности поиска) для методов, поддерживающих локальную интерполяцию, осуществляется в окне, которое появляется при нажатии на кнопку «*Advance Options*», во вкладке «*Search*» (рис. 5). Эта вкладка позволяет задать правила, определяющие, какие выборочные точки будут участвовать в вычислении оценок.

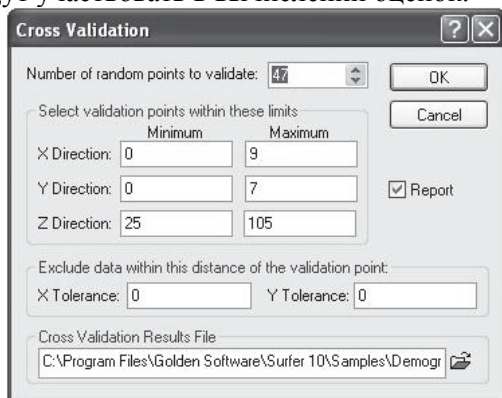


Рис. 5. Окно для определения окрестности поиска

По умолчанию задана опция «*No Search (use all of the data)*» - то есть для вычисления оценок используются все исходные данные, и метод будет работать как глобальный интерpolator.

Если опция «*No Search*» не выбрана, то становятся доступными элементы вкладки, задающие правила поиска данных для локальной аппроксимации.

Прежде всего, нужно задать параметры поискового эллипса. Центр поискового эллипса будет находиться в текущей точке, где делается оценка. Размер и ориентация этого эллипса накладывает ограничения на дальность и место поиска ближайших точек данных, используемых при вычислении оценки в текущей точке. Параметры поискового эллипса задаются в блоке «*Search Ellipse*». Это значения большой и малой полуосей эллипса - «*Radius 1*» и «*Radius 2*», и угол, определяющий его ориентацию в пространстве - «*Angle*».

В случае если в заданную окрестность поиска (эллипс) во-круг интерполируемой точки не попало ни одной выборочной точки, ей присваивается пустое значение.

Кроме этого, для определения используемых данных, можно поделить окрестность поиска на сектора - строка ввода «*Number of sectors to search*» (можно указать от 1 до 32 секторов), и сформулировать следующие четыре дополнительных правила поиска:

- в строке «*Maximum number of data to use from ALL sectors*» задается максимальное количество выборочных точек, используемых при расчетах;

- в строке «*Maximum number of data to use from each sectors*» задается максимальное количество выборочных точек из каждого сектора, значения в которых будут использоваться при расчетах;

- в строке «*Minimum number of data in all sectors (node is blanked if fewer)*» задает минимально необходимое количество точек выборки во всей окрестности поиска (если точек меньше, то интерполируемой точке присваивается пустое значение);

- в строке «*Blank node if more than this many sectors are empty*» задается минимально допустимое количество пустых секторов (куда не попали точки выборки).

Эти правила поиска полезны в том случае, если исходные данные показывают наличие кластеров (отдельных скоплений выборочных точек). Деление окрестности поиска на сектора позволяют выполнить декластеризацию данных (рис. 6). Также эти виды поиска уместны, когда данные собраны вдоль разнесенных траверсов. В этом случае односекторный поиск может привести к оценке на основе точек данных с одного направления, что может проявиться на модельной поверхности, например, как нереалистичные склоны ме-

жду траверсами. Использование 4-8 секторов должно устранить или уменьшить этот эффект.

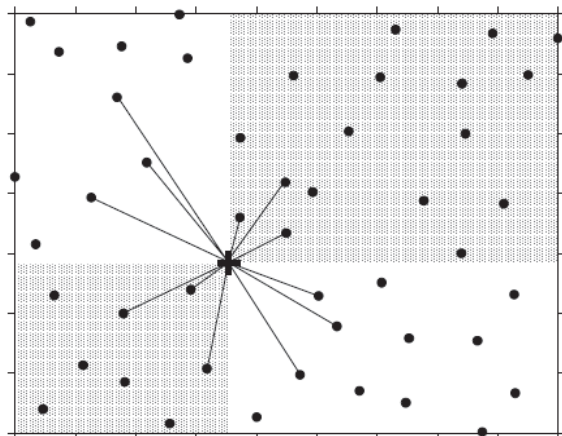


Рис. 6. Пример использования 4-х секторов поиска: *Number of sectors to search = 4, Maximum number of data to use from each sectors = 3*

Для аппроксимации в точке, отмеченной крестиком, будут использоваться значения в 3-х ближайших к ней выборочных точках из 4-х секторов (всего 12 выборочных значений).

Линии нарушения пространственной непрерывности и линии разломов. При оценке значений пространственной переменной могут приниматься в расчет известные линии нарушения пространственной непрерывности и/или линии разломов в области исследования. Примерами таких структурных нарушений являются тектонические разломы, геохимические барьеры, линейные антропогенные сооружения (дамбы, дороги) и пр. Для методов, позволяющих учитывать подобные особенности, такие линии нарушения пространственной непрерывности и/или линии разломов задаются в окне, которое появляется при нажатии на кнопку «*Advance Options*», во вкладке «*Breaklines and Faults*».

На этой вкладке в строке ввода «*File Containing Breaklines*» указывается имя файла, содержащего нетопологическое векторное представление линий нарушения пространственной непрерывности - *Breaklines*. Каждая полилиния, представляющая линию нарушения,

задана перечислением X , Y координат ее вершин и значений переменной Z в вершинах.

Эти линии предоставляют дополнительную информацию, которая может быть задействована при моделировании. Для точки аппроксимации находят ближайшую точку, расположенную на линии нарушений, и с помощью линейной аппроксимации между двумя вершинами вычисляют в ней значение Z (рис. 7). Это значение затем используется, наряду со значениями в выборочных точках, для оценки Z в точке аппроксимации.

Линии типа *Breaklines* можно использовать, например, при моделировании рельефа для учета дополнительной информации о руслах рек с известными отметками уровня воды, о линиях горных хребтов и водоразделов с известными отметками высот.



Рис. 7. Линии типа Breaklines

В строке ввода «*File Containing Fault Traces*» может быть указано имя файла, содержащего векторное представление линий разломов - «*Faults*». Полилинии, представляющие линии разломов, задаются перечислением X , Y координат их вершин.

При учете этого типа структурных нарушений (разломов, барьеров), расстояние от интерполируемой точки до точки выборки вычисляется в обход линии разлома (рис. 8).

Заметим, что файлы с линиями нарушений и разломов можно подготовить непосредственно в программе *Surfer*, используя пункт главного меню «*Map*», подпункт «*Digitize*».



Рис. 8. Линии типа *Faults*

Анизотропия. Изучаемые пространственно-распределенные явления часто создаются физическими процессами, которые имеют определенную ориентацию в пространстве. Например, в устье реки крупные взвешенные вещества оседают быстро, а мелкие - медленнее. Отсюда, чем ближе к береговой линии, тем отложения мощнее, а чем дальше, тем тоньше. Если точка наблюдения находится на расстоянии 100 метров от точки аппроксимации в направлении, параллельном береговой линии, то значение в ней скорее будет похоже на значение в точке аппроксимации, чем значение в точке наблюдения, расположенной на таком же расстоянии, но в направлении, перпендикулярном к береговой линии. Такую зависимость свойств изучаемых явлений от направлений в пространстве принято называть анизотропией.

Многие методы аппроксимации позволяют учитывать анизотропию при построении пространственных моделей. Если точки в одном направлении показывают большее сходство значений, чем точки в другом направлении, то предпочтительнее, чтобы выборочные точки первого направления получили больший вес при оценке значения Z . Такой относительный вес будет определяться отношением анизотропии (*Ratio*) и ее направлением (*Angle*).

Если нет веских оснований считать изучаемое явление анизотропным, то параметр *Ratio* надо оставить равным 1 (по умолчанию). Если значение *Ratio* установлено больше 3 или меньше 1/3, то

анизотропный эффект очевидным образом визуализируется при картографическом представлении построенной модели. Параметр *Angle* задается в градусах.

Заметим, что в большинстве случаев рекомендуется устанавливать параметры поискового эллипса («*Search Ellipse*») таким образом, чтобы его направление и отношение большой и малой полуосей (*Radius1/Radius2*) совпадали с задаваемыми направлением анизотропии и отношением анизотропии, соответственно.

3.1 МЕТОД ОБРАТНЫХ ВЗВЕШЕННЫХ РАССТОЯНИЙ (*INVERSE DISTANCE TO A POWER*)

В качестве оценки переменной *Z* в некоторой точке исследуемой области этот метод использует среднее взвешенное значений переменной *Z* в выборочных точках:

$$\hat{z}_0 = \frac{\sum_{i=1}^n w_{i0} \cdot Z_i}{\sum_{i=1}^n w_{i0}}$$

Метод обратных расстояний работает в предположении, что экспериментальные значения, измеренные близко от интерполируемого местоположения, оказывают большее влияние на результат оценки, чем удаленные от него на значительное расстояние. Следовательно, точки выборки, расположенным ближе всего к точке, где делается оценка, должны иметь больший вес; вес выборочных точек должен уменьшаться с увеличением расстояния от интерполируемого местоположения. Такие веса для данного метода задаются следующим образом:

$$w_{i0} = \frac{1}{(\sqrt{d_{i0}^2 + \delta^2})^\beta}, \text{ где } \delta$$

где d_{i0} – расстояние между точкой (X_0, Y_0) , где вычисляется оценка (например, узлом регулярной сетки), и i -ой точкой измерения; δ – параметр сглаживания; β – степенной параметр.

Степенной параметр β определяет, как быстро будет уменьшаться вес с увеличением расстояния. Если значение β слишком высоко, то на интерполяцию окажут влияние только выборочные точки, расположенные в непосредственной близости. По мере увеличения β оценка будет приближаться к оценке, даваемой методом бли-

жайшего соседа. Для меньших β веса будут распределены более равномерно между соседними точками данных. Если значение β приближается к нулю, то веса всех точек выборки становятся одинаковыми (равными 1), а оценка вырождается в простое среднее значение. На практике, как правило, используют значения β от 1 до 3.

Параметр сглаживания δ позволяет учитывать фактор «неточности» данных. Если $\delta = 0$, то метод работает как точный интерполятор. То есть в любой выборочной точке вычисленная оценка будет равна измеренному значению. Если $\delta > 0$, то метод работает как сглаживающий интерполятор, и чем больше параметр сглаживания, тем меньше подавляющее влияние любого конкретного наблюдения при вычислении оценки в произвольной точке. Заметим, что на изолинейных картах, построенных по рассчитанной этим методом растровой модели, часто наблюдается эффект, так называемых, «блуждающих глаз» - рисунка из концентрических контурных линий вокруг точек выборки. Уменьшить этот эффект можно увеличив значение параметра сглаживания δ .

Параметры метода обратных расстояний задаются в диалоговом окне, которое появляется после нажатия на кнопку «*Advance Options*». Для задания параметра β служит строка ввода «*Power*» на вкладке «*General*», а для задания параметра δ - строка ввода «*Smoothing*».

Хотя этот простой и быстрый метод может работать и как глобальный интерполятор, его имеет смысл применять как локальный интерполятор. То есть в тех случаях, когда изучаемое пространственное явление приводится в действие локальными изменениями, которые могут быть смоделированы с помощью определения подходящей окрестности поиска (вкладка «*Search*»), накладывающей ограничения на дальность, место и условия поиска измеренных значений, используемых для аппроксимации.

Для метода могут быть заданы линии нарушения пространственной непрерывности и линии разломов (вкладка «*Breaklines and Faults*»), а также задана анизотропия. Метод не экстраполирует значения Z за пределы диапазона выборочных данных.

3.2 МОДИФИЦИРОВАННЫЙ МЕТОД ШЕПАРДА (MODIFIED SHEPARD'S METHOD)

Этот метод работает в тех же предположениях, что и метод обратных взвешенных расстояний, но позволяет уменьшить эффект «бычьих глаз». В качестве оценки переменной Z в произвольной точке (X_0, Y_0) области исследования используется среднее взвешенное значений аппроксимирующих функций в этой точке:

$$\hat{Z}_0 = \frac{\sum_{i \in N_0} w_{i0} \cdot Q_i(X_0, Y_0)}{\sum_{i \in N_0} w_{i0}}$$

В общем случае, суммирование ведется не по всем n выборочным точкам, а лишь по n_w ближайшими соседями той точки, где делается оценка. Здесь N_0 – набор номеров таких ближайших выборочных точек ($|N_0| = n_w$). Число n_w является параметром метода, контролирующим его локальность.

Веса выборочных точек, также как и в методе обратных расстояний, обратно пропорциональны расстоянию до точки оценки, но задаются другим способом:

$$w_{i0} = \left(\frac{d_{\max 0} - d_{i0}}{d_{\max 0} \cdot d_{i0}} \right)^2, \text{ где}$$

где d_{i0} – расстояние от точки аппроксимации до i -ой точки измерения; $d_{\max 0}$ – расстояние до максимально удаленной выборочной точки, среди тех, которые учитываются при вычислении оценки.

В отличие от метода обратных расстояний вместо значений самой переменной Z , измеренных в выборочных точках, усредняются значения n_w аппроксимирующих функций, построенных вокруг каждого из n_w наблюдений. В качестве таких функций программа *Surfer* использует квадратичные функции:

$$Q_i(X, Y) = a + b \cdot X + c \cdot Y + d \cdot X \cdot Y + e \cdot X^2 + f \cdot Y^2$$

Коэффициенты i -ой функции определяются методом взвешенных наименьших квадратов по n_q выборочным точкам – ближайшим соседям i -ой точки выборки, с ограничением $Q_i(X_i, Y_i) = Z_i$; веса определяются в соответствие с вышеприведенной формулой. Параметр n_q отвечает за другой аспект локальности метода. Применение такого подхода позволяет избавляться от появления эффекта «бычьих глаз». Подходящее значение n_q в 1.5-2 раза больше, чем

количество свободных параметров у аппроксимирующей функции; у квадратичной функции их шесть.

Модифицированный метод Шепарда по определению является точным интерполятором. В программе *Surfer* реализация данного метода позволяет использовать еще один параметр - параметр сглаживания δ , действующий так же, как и в методе обратных расстояний, и заставляющий метод работать как сглаживающий интерполятор. Наиболее приемлемые значения этого параметра – от 0 до 1. Метод может экстраполировать значения Z за пределы диапазона выборочных данных.

Параметры метода задаются в диалоговом окне, которое появляется после нажатия на кнопку «*Advance Options*»: параметр n_w задается в строке ввода «*Weighting Neighbors*», параметр n_q - в строке ввода «*Quadratic Neighbors*», для задания параметра δ служит строка ввода «*Smoothing Factor*». В блоке «*Search*» можно задать параметры поискового эллипса, накладывающего ограничения на дальность и место поиска ближайших выборочных точек, используемых при оценках.

3.3 МЕТОД БЛИЖАЙШЕГО СОСЕДА (*NEAREST NEIGHBOR*)

В качестве оценки переменной Z в некоторой точке исследуемой области берется значение, которое имеет ближайшая (по Евклидовому расстоянию) выборочная точка:

$$\check{Z}_0 = Z_i, \text{ где}$$

Z_i – значение в выборочной точке, расположенной ближе других к местоположению (X_0, Y_0) .

Этот метод рекомендуется использовать в том случае, если исходные выборочные точки расположены в пространстве регулярно (в узлах некоторой регулярной сетки, покрывающей исследуемую область) или почти регулярно (незначительно смещены относительно узлов сетки). Выходная растровая модель должна иметь шаг равный или близкий к шагу выборочной сетки. Метод полезен, когда данные, расположенные в узлах регулярной сетки, записаны в текстовом *ASCII*-файле, но должны быть преобразованы в формат сетки (*.*grd*) программы *Surfer*.

Иногда в некоторых узлах выборочной сетки данные отсутствуют и в них изучаемая переменная не должна интерполироваться

(эти узлы попали в такие подобласти, где переменная Z не распределена). В этом случае выходная растровая модель должна иметь пустое значения в тех местах, где нет данных наблюдения. Для этого в соответствующем диалоговом окне «*Advance Options*» необходимо задать радиусы поискового эллипса, установив их меньше, чем шаг выборочной сетки.

Для метода ближайшего соседа могут быть заданы линии нарушения пространственной непрерывности и линии разломов (вкладка «*Breaklines and Faults*»). Метод не экстраполирует значения Z за пределы диапазона выборочных данных.

3.4 МЕТОД СКОЛЬЗЯЩЕГО СРЕДНЕГО (*MOVING AVERAGE*)

Оценкой переменной Z в некоторой точке исследуемой области, полученной по этому методу, будет среднее арифметическое значений в выборочных точках, попавших в поисковый эллипс с центром в этой точке:

$$Z_0 = \frac{\sum_{i \in N_0} Z_i}{n_0}$$

Суммирование ведется по n_0 ближайшим соседям точки аппроксимации, попавшим в окрестность поиска. Здесь N_0 – набор номеров таких выборочных точек ($|N_0| = n_0$). Параметры поискового эллипса задаются в диалоге «*Moving Average Advance Options*» на вкладке «*General*» в блоке «*Search Ellipse*». Это значения большой и малой полуосей эллипса и угол, определяющий его ориентацию в пространстве.

В строке «*Minimum number of data (node is blanked if fewer)*» этого же диалогового окна задается минимальное количество n_0 точек выборки, необходимых для вычисления среднего. В случае если количество выборочных точек, попавших в поисковый эллипс, окажется меньше заданного, вычисления не производятся и точке присваивается пустое значение.

Метод скользящего среднего не рекомендуется использовать, если объем выборочных данных небольшой или средний. Однако в случае больших и очень больших наборов пространственных данных, он является полезным инструментом исследования. Этот метод позволяет выявлять разномасштабные тренды в данных и быстро

работает даже на очень больших выборках. Он также может рассматриваться как альтернатива методу ближайшего соседа для получения растровых моделей по большому объему исходных данных, расположенных на регулярной выборочной сетке.

Метод позволяет учитывать линии нарушения пространственной непрерывности, но не разломы (вкладка «*Breaklines*»).

3.5 ПОЛИНОМИАЛЬНАЯ РЕГРЕССИЯ (*POLYNOMIAL REGRESSION*)

Метод используется для оценки пространственного тренда в данных, то есть зависимости изучаемой пространственной переменной Z от координат X , Y . Этот метод не является интерполяционным в прямом смысле: он не пытается предсказать неизвестные значения Z . Он предназначен для выявления глобальных пространственных трендов и теряет детальную локальную информацию, содержащуюся в данных.

В диалоговом окне, которое появляется после нажатия на кнопку «*Advance Options*», задается вид полинома, с использованием которого строится полиномиальная модель тренда. Для этого в блоке «*Surface Definition*», используя переключатель, выбирается предопределенный вид полинома: линейный, билинейный, квадратичный или кубический. Либо можно, выбрав пункт «*User defined polynomial*», вручную задать степень полинома отдельно для переменной X (в поле «*Max X Order*») и переменной Y (в поле «*Max Y Order*»). В поле «*Max Total Order*» определяется максимум суммы степеней X и Y . Полином будет включать все комбинации X и Y , сумма степеней которых не превосходит значения в поле «*Max Total Order*».

Уравнение полиномиальной регрессии Z на X , Y строится с использованием метода наименьших квадратов на основе всех входных данных (окрестность поиска не применяется), и метод можно считать глобальным и сглаживающим интерполятором. Оценкой переменной Z в произвольной точке исследуемой области будет значение, вычисленное по полученному уравнению регрессии:

$$\hat{Z}_0 = Q(X_0, Y_0), \text{ где}$$

$Q(X, Y)$ – построенный полином.

3.6 МЕТОД ЛОКАЛЬНЫХ ПОЛИНОМОВ (LOCAL POLINOMIAL)

Этот метод также оценивает зависимость пространственной переменной Z от координат X, Y , но в отличие от предыдущего использует кусочно-полиномиальное аппроксимирование. Он строит полиномиальную регрессию Z на X, Y по данным, попавшим в поисковый эллипс с центром в заданной точке области исследования, получая, таким образом, полиномиальное приближение в локальной окрестности этой точки. Реализация метода в *Surfer* использует локальные полиномы степени 1, 2 или 3:

$$a+b\cdot X+c\cdot Y$$
$$a+b\cdot X+c\cdot Y+d\cdot X\cdot Y+e\cdot X^2+f\cdot Y^2$$
$$a+b\cdot X+c\cdot Y+d\cdot X\cdot Y+e\cdot X^2+f\cdot Y^2+g\cdot X^2\cdot Y+h\cdot X\cdot Y^2+i\cdot X^3+j\cdot Y^3$$

Для определения коэффициентов полиномов применяется метод взвешенных наименьших квадратов, где вес каждой выборочной точки зависит от ее расстояния от точки, где делается оценка, параметров поискового эллипса и задаваемого пользователем параметра *Power*. Этот параметр задается в диалоге «*Local Polynomial Advance Options*» на вкладке «*General*» в строке «*Power*». Чем больше его значение, тем больший вес будут иметь близлежащие точки выборки, и меньший – удаленные (то есть, тем «локальнее» будет полином). Степень аппроксимирующего полинома задается в блоке «*Polynomial Order*».

Локальная окрестность для поиска данных определяется на вкладке «*Search*». Чтобы избежать резких изменений поведения модельной поверхности (разрывности производных), необходимо обеспечить плавность соединения соседних полиномов. Для этого размеры локальной окрестности задают таким образом, чтобы обеспечить частичное перекрытие наборов данных, используемых для аппроксимации в соседних узлах сетки.

Оценкой переменной Z в точке аппроксимации будет являться значение полиномиального уравнения:

$$\check{Z}_0 = Q_0(X_0, Y_0), \text{ где}$$

$Q_0(X, Y)$ – локальный полином, построенный по данным из локальной окрестности точки (X_0, Y_0) .

При построении локальных полиномов можно учитывать известные линии нарушения пространственной непрерывности на территории исследования (вкладка «*Breaklines*»).

Метод локальных полиномов наиболее подходит для моделирования локально гладких поверхностей. Глобальный полином первого порядка соответствует плоскости (рис. 9а); глобальный полином второго порядка соответствует поверхности с изгибом, позволяя отображать долины; глобальный полином третьего порядка может характеризовать два изгиба. Но для поверхности с меняющейся формой, например ландшафта с чередованием склонов и ровных поверхностей, глобальный полином не подходит. Локальные полиномы представляют такую поверхность более точно (рис. 9б).

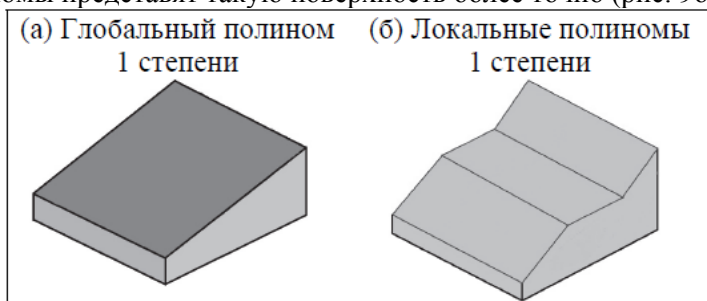


Рис. 9. Иллюстрация поверхностей глобального и локальных полиномов

3.7 ТРИАНГУЛЯЦИЯ С ЛИНЕЙНОЙ ИНТЕРПОЛЯЦИЕЙ (*TRIANGULATION WITH LINEAR INTERPOLATION*)

Этот метод использует алгоритм построения триангуляции Делоне - разбиения области исследования на совокупность треугольников. Вершинами треугольников являются соседние точки выборки; точки соединяются таким образом, что ребра треугольников не пересекаются. Внутри окружности, описанной вокруг любого построенного треугольника, не попадает ни одна из точек триангуляции (рис. 10). Для оценки переменной Z в некоторой точке (X_0, Y_0) области исследования сначала определяется, в какой треугольник она попала.

Пусть интерполируемая точка попала в треугольник, образованный выборочными точками с координатами (X_1, Y_1) , (X_2, Y_2) и (X_3, Y_3) . Тогда в трехмерном (X, Y, Z) пространстве строится плоскость,

проходящая через точки с координатами (X_1, Y_1, Z_1) , (X_2, Y_2, Z_2) и (X_3, Y_3, Z_3) . Уравнение плоскости:

$$a \cdot X + b \cdot Y + c \cdot Z + d = 0, \text{ где}$$

коэффициенты могут быть найдены по известным формулам:

$$a = Y_1(Z_2 - Z_3) + Y_2(Z_3 - Z_1) + Y_3(Z_1 - Z_2)$$

$$b = Z_1(X_2 - X_3) + Z_2(X_3 - X_1) + Z_3(X_1 - X_2)$$

$$c = X_1(Y_2 - Y_3) + X_2(Y_3 - Y_1) + X_3(Y_1 - Y_2)$$

$$d = X_1(Y_2Z_3 - Y_3Z_2) + X_2(Y_3Z_1 - Y_1Z_3) + X_3(Y_1Z_2 - Y_2Z_1)$$

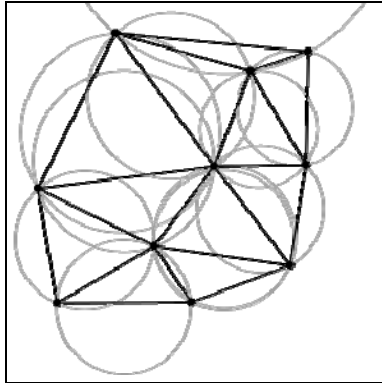


Рис. 10. Триангуляция Делоне

Оценкой переменной X в точке (X_0, Y_0) будет соответствующее значение на этой плоскости (рис. 11):

$$\hat{Z}_0 = \frac{-a \cdot X_0 - b \cdot Y_0 - d}{c}$$

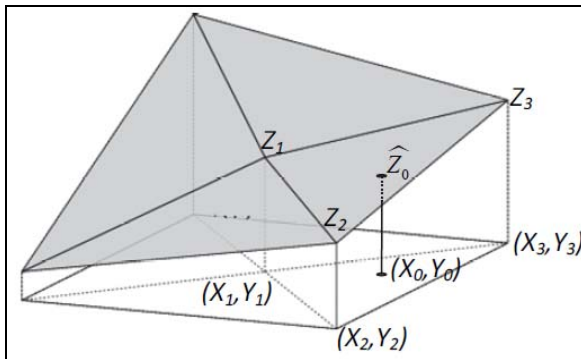


Рис. 11. Триангуляция с линейной интерполяцией

Использование этого метода при небольшом числе точек выборки приводит к появлению явных треугольных граней на модельной растровой поверхности, или больших прямолинейных сегментов на карте изолиний. Триангуляция с линейной аппроксимацией работает лучше, если данные равномерно распределены по исследуемой территории. Метод эффективен, если требуется сохранить линии разрывов поверхности.

Метод точно воспроизводит значения в выборочных точках, то есть является точным интерполятором. Метод может учитывать анизотропию изучаемой переменной. Полученную схему триангуляции можно сохранить в файл векторного формата, задав ему имя в соответствующем диалоге «*Advance Options*» (рис.20).

3.8 МЕТОД ЕСТЕСТВЕННОЙ ОКРЕСТНОСТИ (*NATURAL NEIGHBOR*)

В этом методе оценка переменной X в некоторой точке области исследования определяется как среднее взвешенное значений этой переменной в выборочных точках (фактически в ближайших выборочных точках):

$$\hat{Z}_i = \sum_{j=1}^n w_{ij} \cdot Z_j, \text{ где}$$

веса w_{ij} определяются с использованием диаграмм Вороного.

Диаграмма Вороного для заданного множества точек - это такое разбиение плоскости на совокупность областей, при котором любая точка в отдельной области разбиения ближе к одной из точек множества, чем к любой другой точке множества (рис. 12а). Диаграммы Вороного также называют разбиением Тиссена или ячейками Дирихле. Одним из главных свойств диаграммы Вороного является её взаимно-однозначное соответствие с триангуляцией Делоне. А именно, если соединить рёбрами исходные точки, чьи многоугольники (ячейки) Вороного граничат друг с другом, то получим триангуляцию Делоне (рис. 12б).

В методе естественной окрестности строятся две диаграммы Вороного. Сначала строится диаграмма Вороного для множества точек выборки (рис. 13а). Затем в это множество встраивается точка (X_0, Y_0) , где делается оценка, и строится диаграмма Вороного для новой конфигурации точек (рис. 13б). Диаграммы накладываются одна на другую. Ячейка новой точки (заполненная серой заливкой

на рис. 13в) покрывает части ячеек некоторых точек первой диаграммы. Именно эти точки выборки (естественные соседи) и будут учитываться при оценке X в точке (X_0, Y_0) . Веса w_i0 определяются как доля площади ячейки интерполируемой точки (X_0, Y_0) , принадлежащей ячейке i -ой выборочной точки на первой диаграмме.

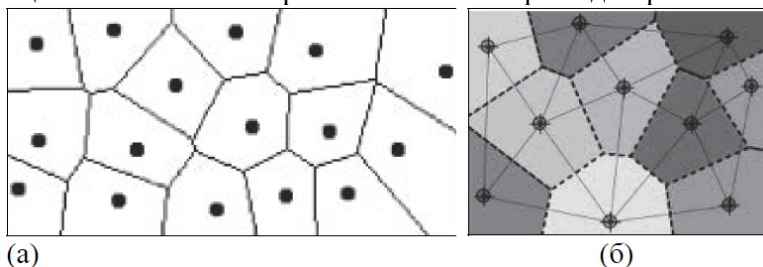


Рис. 12. а – диаграмма Вороного, б – диаграмма Вороного и триангуляция Делоне

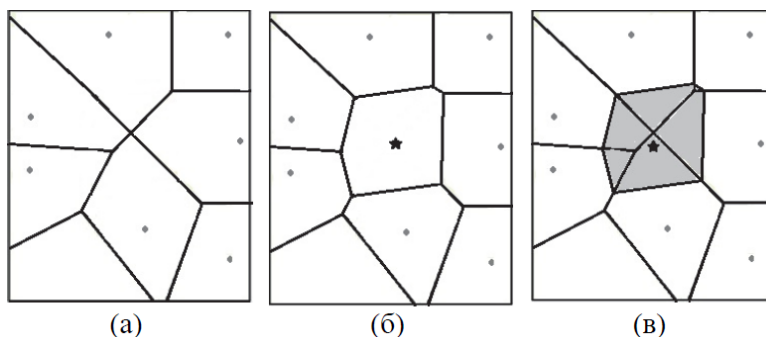


Рис. 12. а – диаграмма Вороного для точек выборки, б – диаграмма Вороного для точек выборки и точки аппроксимации, в – наложение двух диаграмм

Метод позволяет получать хорошие изолинейные карты по наборам данных, содержащим скопления выборочных точек в одних подобластях исследуемой территории и разреженные выборочные точки в других подобластях. Получаемая в произвольной точке оценка X всегда будет заключаться в пределах диапазона значений в ближайших точках. В построенной модели не будет «пиков», «ям», «хребтов», «долин», которые не были бы представлены исходными данными. Метод является точным интерполятором.

В диалоге «*Natural Neighbor Advance Options*» предоставляется возможность задать параметры анизотропии, а также сохранить в файл векторного формата схему триангуляции Делоне, соответствующую первой диаграмме Вороного.

3.9 МЕТОД МИНИМАЛЬНОЙ КРИВИЗНЫ (*MINIMUM CURVATURE*)

Под этим названием в *Surfer* используется двухмерный сплайн с натяжением (с минимизацией кривизны). Поверхность, построенная с помощью этого метода, подобна тонкой эластичной пластине, которая стремится пройти через все экспериментальные точки с минимальным числом изгибов. Метод не является точным интерполятором, он генерирует наиболее гладкую поверхность, которая проходит настолько близко к экспериментальным точкам, насколько это возможно.

Предварительно с использованием метода наименьших квадратов и всех выборочных данных строится аппроксимирующая поверхность на основе простого полинома $Z(X, Y) = a + b \cdot X + c \cdot Y$, и в точках выборки вычисляются остатки – разность между результатами аппроксимации и наблюдаемыми значениями:

$$e_i = (a + b \cdot X_i + c \cdot Y_i) - Z_i$$

Далее, используя алгоритм минимальной кривизны, остатки интерполируются в узлах сетки. Алгоритм решает модифицированное дифференциальное уравнение:

$$(1 - T_i) \nabla^2 (\nabla^2 e) - (T_i) \nabla^2 e = 0$$

с граничными условиями:

$$(1 - T_b) \frac{\partial^2 e}{\partial n^2} + T_b \frac{\partial e}{\partial n} = 0, \frac{\partial (\nabla^2 e)}{\partial n} = 0, \frac{\partial^2 e}{\partial X \partial Y} = 0, \text{ где}$$

∇^2 – оператор Лапласа, n – нормаль к границе, T_i – параметр «внутреннее натяжение», T_b – параметр «натяжение на границе». Алгоритм работает итерационно, перевычисляя оценки в узлах модельной сетки, пока не будет достигнута либо заданная невязка, либо заданное число итераций.

Результирующую оценку X получают путем сложения аппроксимирующей полиномиальной поверхности и проинтерполированных остатков:

$$\check{Z}(X, Y) = a + b \cdot X + c \cdot Y + \check{e}(X, Y)$$

В диалоговом окне, которое появляется после нажатия на кнопку «*Advance Options*», задаются параметры, управляющие работой метода минимальной кривизны, в том числе критерий сходимости.

В строке ввода «*Maximum Residual*» задается максимальная невязка (измеряется в тех же единицах, что и экспериментальные данные). Наиболее подходящее значение этого параметра равно, примерно, 10% от точности исходных данных. Например, если исходные данные измерены с точностью до единицы, то рекомендуемое значение $Maximum\ Residual = 0,1$. По умолчанию значение этого параметра устанавливается равным $0,001 \cdot (X_{max} - X_{min})$. Итерации продолжаются до тех пор, пока максимальная невязка для всей итерации не станет меньше значения *Maximum Residual*.

В строке ввода «*Maximum Iteration*» необходимо задать максимальное количество итераций. Значение этого параметра должно устанавливаться в зависимости от числа узлов растровой сетки, в которых вычисляется оценка. Если M - число узлов сетки, то значение этого параметра рекомендуется выбирать в интервале от M до $2M$. Например, если используется сетка 50 на 50, то максимальное количество итераций должно быть установлено между 2500 и 5000.

Если при работе алгоритма количество итераций превысит заданное значение *Maximum Iteration*, и при этом расчетная невязка окажется выше заданной *Maximum Residual*, то модель не будет создана (появится сообщение, что итерационный процесс не сходится).

В строке ввода «*Relaxation Factor*» задается, так называемый, фактор расслабления. Его значение может изменяться от 0 до 2, и по умолчанию, равно 1, что хорошо работает в большинстве случаев. Чем меньше значение этого параметра (ближе к 0), тем медленнее работает алгоритм, но тем вероятнее схождение (достижение нужной невязки за заданное число итераций); и наоборот, чем больше это число (ближе к 2), тем быстрее работает алгоритм, но вероятность схождения ниже. Оптимальное значение фактора расслабления определяется путем проб и ошибок.

В строках ввода «*Internal Tension*» и «*Boundary Tension*» можно задать параметры T_i и T_b , управляющие «внутренним натяже-

нием» сплайна и «натяжением на границе», соответственно. Оба эти параметра управляют степенью кривизны (натяжением) поверхности и изменяются от 0 до 1.

Если взять число близкое к нулю, то натяжение будет минимальным, и наоборот, если взять число близкое к 1, то это будет приводить к максимальному натяжению (меньшей кривизне) моделируемой поверхности. Например, высокое натяжение создаст области между наблюдениями, которые будут выглядеть как грани драгоценного камня.

При построении поверхности методом минимальной кривизны могут быть учтены линии нарушения пространственной непрерывности и линии разломов (вкладка «*Breaklines and Faults*»), а также задана анизотропия.

Этот метод лучше всего подходит для медленно меняющихся поверхностей, например, высот грунтовых вод, рельефа (без резко меняющихся элементов рельефа), концентрации загрязняющих веществ в воздухе. Важным свойством метода является непрерывность поверхности производной второго порядка.

3.10 МЕТОД РАДИАЛЬНЫХ БАЗИСНЫХ ФУНКЦИЙ (*RADIAL BASIC FUNCTION*)

В этом методе оценка переменной X в произвольной точке области исследования находится как линейная комбинация значений радиальных базисных функций:

$$\hat{z}_0 = \sum_{i=1}^n \lambda_i B(d_{i0}), \text{ где}$$

d_{i0} – расстояние между точкой, где вычисляется оценка, и i -ой точкой измерения; λ_i – коэффициент i -ой выборочной точки; $B()$ – радиальная базисная функция, аргументом которой является расстояние d_{i0} .

Радиальная базисная функция является ядерной функцией, которая определяет оптимальные веса, применяемые к точкам данных во время аппроксимации. Существует множество разновидностей радиальной базисной функции, из них в Surfer доступны пять типов:

1. Мультикватричная функция (*Multiquadric*):

$$B(d) = \sqrt{d^2 + R^2}$$

2. Обратная мультикватричная функция (*Inverse Multiquadric*):

$$B(d) = \frac{1}{\sqrt{d^2 + R^2}}$$

3. Мультилогарифмическая функция (*Multilog*):

$$B(d) = \log(d^2 + R^2)$$

4. Плоский сплайн (*Thin-Plate Spline*):

$$B(d) = (d^2 + R^2)\log(d^2 + R^2)$$

5. Кубический сплайн (*Natural Cubic Spline*):

$$B(d) = (d^2 + R^2)^{3/2}$$

где d – расстояние (с учетом анизотропии) от точки аппроксимации до выборочной точки, R^2 – параметр сглаживания.

Каждая базисная функция имеет различную форму и позволяет получать разные интерполированные поверхности. Наиболее часто используется мультикватричная функция, которая, по мнению многих, является наилучшей с точки зрения построения гладкой поверхности, проходящей через экспериментальные точки. Для параметра сглаживания R^2 нет общих рекомендаций по выбору оптимального значения; разумные значения этого параметра находятся в интервале от среднего расстояния между точками выборки до половины этого среднего.

Коэффициенты λ_i получают решением системы n уравнений, которые составляются из условия точной аппроксимации - прохождения выходной поверхности через значения X в n выборочных точках:

$$\sum_{i=1}^n \lambda_i B(d_{ij}) = Z_j, j = 1, \dots, n$$

Неизвестными в уравнениях являются n коэффициентов λ_i . И последовательность действий для определения λ_i включает: вычисление расстояний между всеми выборочными точками (d_{ij}), вычисление по ним значений радиальной базисной функции ($B(d_{ij})$), решение системы уравнений.

Метод радиальных базисных функций является точным интерполятором. Параметр сглаживания позволяет регулировать гладкость получаемой поверхности. С помощью радиальных базисных функций можно успешно создавать модели слабо изменяющихся

поверхностей, например поверхности высот. Однако этот метод плохо подходит, если возможны значительные изменения изучаемой пространственной переменной в пределах коротких расстояний, и/или если существует подозрение, что исходные данные могут содержать ошибки.

Для вычислений с использованием радиальных базисных функций в *Surfer* необходимо в диалоговом окне на вкладке «*General*» в строке «*Basic Function*» выбрать вид радиальной базисной функции, и в строке «*R² Parameter*» задать значение параметра сглаживания R^2 . Метод позволяет ограничивать окрестность поиска выборочных данных (вкладка «*Search*»), учитывать линии нарушения пространственной непрерывности (вкладка «*Breaklines*») и анизотропию.

3.11 МЕТОД КРИГИНГА (KRIGING)

В отличие от всех описанных выше методов, которые относятся к классу детерминированных, кригинг является геостатистическим методом. Кригинг строит скорее статистическую модель реальности, чем модель интерполяционной функции. Геостатистические методы основываются на вероятностной модели, рассматривающей изучаемую пространственную переменную $Z(X, Y)$ как реализацию случайной функции $Z(X, Y)$. Такой подход позволяет учитывать пространственную корреляцию данных и дает возможность не только создавать модели поверхностей, но также получать оценку точности этих моделей.

Данный метод основан на использовании методов математической статистики. В его реализации применяется идея регионализованной переменной, то есть переменной, которая изменяется от места к месту с некоторой видимой непрерывностью, поэтому не может моделироваться только одним математическим уравнением. Поверхность рассматривается в виде трех независимых величин. Первая - тренд, характеризующий изменение поверхности в определенном направлении. Вторая - небольшие отклонения от общей тенденции, вроде маленьких пиков и впадин, которые являются случайными, но все же связанными друг с другом пространственно. Третья - случайный шум, выбивающиеся из преобладающей тенденции значения. С каждой из трех переменных надо оперировать в отдельно-

сти. Тренд оценивается с использованием математического уравнения, которое наиболее близко представляет общее изменение поверхности и во многом подобно поверхности тренда.

Реализованные в различных приложениях методы кригинга решают задачу аппроксимации с применением линейных оценок. Это три основные формы кригинга: простой кригинг (используется, если известно математическое ожидание случайной функции Z), ординарный кригинг (математическое ожидание случайной функции Z неизвестно, но постоянно) и универсальный кригинг (математическое ожидание случайной функции Z неизвестно и непостоянно).

В простейшем случае, кригинг-оценка значения пространственной переменной Z в некоторой точке, где не было измерений, определяется как:

$$\hat{Z}_0 = \sum_{t=1}^n \lambda_t Z_t, \text{ где}$$

n весов λ_t являются решением системы кригинга:

$$\sum_{t=1}^n \lambda_t C_{tj} = C_{0j}, j = 1, \dots, n$$

а ошибка оценки кригинга определяется как квадратный корень из дисперсии оценки:

$$\begin{aligned} \sigma^2 &= \text{Var}\{Z(X_0, Y_0) | Z_1, \dots, Z_n\} = E \left[\left(\hat{Z}_0 - Z(X_0, Y_0) \right)^2 \right] = \\ &= C(0) - \sum_{t=1}^n \lambda_t C_{0t} \end{aligned}$$

В этих уравнениях $C_{ij} = C((X_i, Y_i); (X_j, Y_j))$ – значение функции ковариации, характеризующее связь между случайной величиной $Z(X_i, Y_i)$ в i -ой точке выборки и случайной величиной $Z(X_j, Y_j)$ в j -ой точке выборки. Соответственно, $C_{0i} = C((X_0, Y_0); (X_i, Y_i))$ – значение функции ковариации между случайной величиной $Z(X_0, Y_0)$ в интерполируемой точке и случайной величиной $Z(X_i, Y_i)$ в i -ой точке выборки.

Функция $C()$ не известна. В предположении стационарности изучаемого явления, эта функция может быть оценена по выборочным данным, и по результатам оценки построена ее модель $\hat{C}()$. При этом оценивается и моделируется не сама ковариация, а другая мера пространственной корреляции – вариограмма, зная которую можно

достаточно просто получить функцию ковариации. Этот процесс носит название вариограммного анализа данных. Полученная модель ковариации $\hat{C}()$ затем подставляется в систему кригинга, которая решается для определения весов λ_i .

Таким образом, для применения метода кригинга предварительно должна быть получена модель вариограммы – меры пространственной корреляции данных, которая и является основным параметром метода.

Основные этапы создания геостатистической модели включают:

1. анализ и предварительную обработку данных (декластеризация, выявление трендов и областей пространственной неоднородности, анализ распределения, выпадающих значений, анизотропии);
2. расчет значений эмпирической вариограммы или ковариации;
3. построение модели вариограммы или ковариации;
4. решение системы уравнений кригинга для определения весов;
5. получение прогнозного значения и ошибки (неопределенности) оценки в произвольной точке области исследования (например, в узлах регулярной сетки).

В *Surfer* параметры кригинга задаются в диалоговом окне «*Kriging Advanced Options*» во вкладке «*General*», которая разделена на блоки.

В блоке «*Variogram Model*» предоставляется возможность добавить (кнопка «*Add*»), удалить (кнопка «*Remove*»), отредактировать (кнопка «*Edit*») модель вариограммы, подобранную в ходе вариограммного анализа (пункт главного меню «*Variogram*» → «*New variogram*»).

В блоке «*Output Grid of Kriging Standard Deviation*» можно задать имя файла, в котором будет записана сетка стандартных отклонений – ошибок кригинга.

В выпадающем списке «*Kriging Type*» можно выбрать либо «*Point*», либо «*Block*». В случае выбора «*Point*» (по умолчанию), метод оценивает значения в узлах сетки. В случае выбора «*Block*», оцениваются средние значения в прямоугольных блоках с центром в

узлах сетки. Блоки размером и формой соответствуют ячейкам сетки. Поскольку «блочный» кригинг оценивает среднее в «блоке», он генерирует гладкую поверхность. То есть, даже, если точка наблюдения расположена точно в узле сетки, оценка «блочного» кригинга в этом узле не воспроизведет наблюдаемое значение.

Выпадающий список «*Drift Type*» содержит три значения: «*None*» (по умолчанию), «*Linear*», «*Quadratic*». Он предназначен для извещения о наличии в данных пространственного тренда (дрифта) и выбора формы кригинга. Если выбрана опция «*None*», то считается, что исходные данные не содержат пространственного тренда. В этом случае оценка будет выполняться методом ординального кригинга. Если выбрана опция «*Linear*» или «*Quadratic*», то считается, что в данных присутствует линейный или квадратичный тренд, и оценка выполняется методом универсального кригинга.

Для кригинга может задаваться окрестность поиска выборочных точек (вкладка «*Search*»), а также могут учитываться известные линии нарушения пространственной непрерывности (вкладка «*Breaklines*»). Анизотропное поведение изучаемой переменной учитывается на этапе построения модели вариограммы.

Кригинг является очень гибким методом, он может быть либо точным, либо сглаживающим интерполятором в зависимости от значений параметров. Он позволяет эффективным и естественным образом включать в модель анизотропию и тренды. Если рассматривать только методы аппроксимации, которые работают на основе взвешенного усреднения, то кригинг является наилучшим несмещенным предиктором. Если данные имеют нормальное распределение, то кригинг становится наилучшим предиктором среди всех несмещенных предикторов (не только взвешенных средних). Метод кригинга дает хорошие результаты для большинства наборов данных, в силу чего является рекомендуемым и выбранным по умолчанию в Surfer.

3.12 РАСЧЕТ СТАТИСТИК (*DATA METRICS*)

Выбор этого метода позволяет рассчитывать в узлах заданной сетки различные, в основном статистические, показатели на основе значений в выборочных точках, попавших в поисковый эллипс. Данный метод, в общем случае, не является интерполяционным. Он

предназначен для предварительного анализа исходных данных. Рассчитываемый показатель выбирается в списке «*Data Metric to Grid*» на вкладке «*General*» диалога «*Data Metrics Advance Options*». Например, если выбрать пункт «*Mean*» в блоке «*X Moment Statistics*», а на вкладке «*Search*», задавая поисковый эллипс, выбрать пункт «*Use all data*» (использовать все исходные данные), то на результирующей сетке в каждом узле будет находиться одно и то же число - среднее значение, подсчитанное по всем выборочным данным.

При подсчете статистик может быть задана окрестность поиска выборочных данных (вкладка «*Search*»), а также заданы и учтены известные линии нарушения пространственной непрерывности и линии разломов (вкладка «*Breaklines and Faults*»).

РЕКОМЕНДУЕМЫЙ БИБЛИОГРАФИЧЕСКИЙ СПИСОК

а) основная литература

1. *Блиновская Я.Ю.* Введение в геоинформационные системы: Учебное пособие./ Я.Ю. Блиновская, Д.С. Задоя. - М.: НИЦ ИНФРА-М, 2014. - 112 с.
2. *Добрякова В.А.* Основы Mapinfo: Учебное пособие. - Тюмень: ТГУ, 2006. - 81 с.
3. *Капралов Е.Г.* Геоинформатика: Учебное пособие / Е.Г. Капралов, А.В. Кошкарёв, В.С. Тикунов. Т.1. - М.: Академия, 2008. – 373 с.
4. *Капралов Е.Г.* Геоинформатика: Учебное пособие / Е.Г. Капралов, А.В. Кошкарёв, В.С. Тикунов. Т.2. - М.: Академия, 2008. – 391 с.
5. *Мальцев К.А.* Построение моделей пространственных переменных (с применением пакета Surfer): Учебное пособие / К.А. Мальцев, С.С. Мухарамова. – Казань: Казанский университет, 2014. – 103 с.
6. *Мурзинцев П.П.* Инженерное управление территориями (использование ГИС MapInfo в зонировании городов): Учебное пособие. - Новосибирск: СГГА, 2008. - 105 с.
7. *Овчинников В.А.* Программирование для MapInfo на примерах: Учебник. – Москва, 2011. - 181 с.
8. *Савельев А.А.* Пространственный анализ в растровых геоинформационных системах: Учебно-методическое пособие / В.А. Овчинников, С.С. Мухарамова, А.Г. Пилюгин. – Казань: Казанский государственный университет, 2007. – 28 с.

б) дополнительная литература

1. *Гараевская Л.С.* Картография: Учебник. - М.: Недра, 1971. – 360 с.
2. *Дензин П.В.* Геодезия: Учебник. М.: Издательство Московского университета, 1953. – 376 с.
3. *Кравцова В.И.* Космические методы картографирования: Учебное пособие. - М.: Издательство Московского университета, 1995. - 236 с.
4. *Тикунов В.С.* Моделирование в картографии: Учебник. - М.: Издательство Московского университета, 1997. – 405 с.

СОДЕРЖАНИЕ

Введение	3
1. Построение моделей пространственных переменных	4
2. Создание пространственной модели в surfer	5
2.1 Ввод исходных данных	5
2.2 Задание геометрических параметров выходной сетки.....	7
2.3 Выбор метода аппроксимации	8
2.4 Перекрестная проверка	8
3. Методы пространственной аппроксимации.....	10
3.1 Метод обратных взвешенных расстояний (<i>inverse distance to a power</i>).....	16
3.2 Модифицированный метод шепарда (<i>modified shepard's method</i>).....	18
3.3 Метод ближайшего соседа (<i>nearest neighbor</i>).....	19
3.4 Метод скользящего среднего (<i>moving average</i>).....	20
3.5 Полиномиальная регрессия (<i>polynomial regression</i>).....	21
3.6 Метод локальных полиномов (<i>local polinomial</i>).....	22
3.7 Триангуляция с линейной интерполяцией (<i>triangulation with linear interpolation</i>).....	23
3.8 Метод естественной окрестности (<i>natural neighbor</i>).....	25
3.9 Метод минимальной кривизны (<i>minimum curvature</i>).....	27
3.10 Метод радиальных базисных функций (<i>radial basic function</i>).....	29
3.11 Метод кригинга (<i>kriging</i>).....	31
3.12 Расчет статистик (<i>data metrics</i>).....	34
Рекомендуемый библиографический список.....	36

ДИСТАНЦИОННЫЕ И ГИС-ТЕХНОЛОГИИ В ГЕОЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

*Методические указания к самостоятельным работам
для студентов бакалавриата направления 05.03.06*

Сост. *А.В. Стриженок*

Печатается с оригинал-макета, подготовленного кафедрой
геоэкологии

Ответственный за выпуск *А.В. Стриженок*
Лицензия ИД № 06517 от 09.01.2002

Подписано к печати 04.06.2021 Формат 60×84/16.
Усл. печ. л. 2,2. Усл.кр.-отг. 2,2. Уч.-изд.л. 2,0. Тираж 75 экз. Заказ 540.

Санкт-Петербургский горный университет
РИЦ Санкт-Петербургского горного университета
Адрес университета и РИЦ: 199106 Санкт-Петербург, 21-я линия, 2