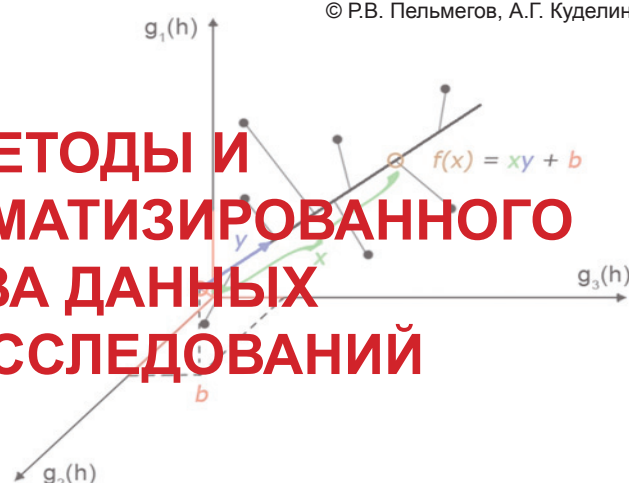


УДК 550.832:681.5.03

© Р.В. Пельмегов, А.Г. Куделин

Р.В. Пельмегов, А.Г. Куделин

ЭВРИСТИЧЕСКИЕ МЕТОДЫ И ТЕХНОЛОГИИ АВТОМАТИЗИРОВАННОГО КОНТРОЛЯ КАЧЕСТВА ДАННЫХ ГЕОФИЗИЧЕСКИХ ИССЛЕДОВАНИЙ СКВАЖИН



Моделирование геообъектов и геопроцессов

В настоящее время бурение любой скважины обязательно сопровождается комплексом геофизических исследований скважин (ГИС) [1]. Во время проведения или непосредственно по завершении исследования необходимо провести оперативную верификацию каротажных диаграмм и при необходимости принять решение о повторном замере. Однако в большинстве случаев эта процедура переносится на этап обработки и интерпретации, тем самым увеличивая затраты времени на выдачу заключений. В связи с вышесказанным представляется актуальной разработка автоматизированной экспертной системы оценки качества (АЭСОК) первичных материалов ГИС, обеспечивающей возможность оперативно оценивать качество диаграммного материала перед его отправкой на интерпретацию.

Экспертиза и экспертные оценки при всех очевидных недостатках до сегодняшних дней остается единственным способом анализа и решения плохо формализуемых задач. Наиболее эффективно применение экспертных методов в сочетании с другими применяемыми ныне методами оценки качества данных, в основном со статистическими методами. Широко используются воспроизводимость данных на перекрывающихся интервалах, повторных и контрольных измерений. Однако важно помнить, что для оценки качества материала можно рассматривать данные, записанные только во время одного и того же каротажа, т.е. с минимальным временным интервалом, поскольку сложность применения статистических методов связана с большим количеством возможных преобразований регистрируемых параметров.

Недостатки проверки качества кривых ГИС статистическими методами частично могут быть устранены применением способа многоскважинного статистического контроля [2], основанного на предположении о том, что выбранный геофизический параметр имеет постоянное распределение по площади для достаточно мощной толщи пластов [3].

Проблема метода состоит в получении эталонного распределения изучаемого параметра. Идеальным является случай, когда можно построить распределение параметра по представительной выборке керна. В случае когда имеются лишь единичные образцы керна, строить по ним распределение какого-либо параметра не имеет смысла.

Классической задачей статистической обработки геофизических материалов является спектральный анализ или более конкретно – оценивание значений частот, амплитуд, фаз и периодических составляющих различных геофизических полей [4]. Спектральный анализ позволяет судить о наличии в данном процессе тех или иных ритмов, частота которых соответствует имеющимся в анализаторе фильтрам, с помощью которых выделяются все колебания данного периода. Недостатком метода является отсутствие информации о локализации частоты в пространстве (или во времени). Таким образом, при применении в качестве метода контроля качества данных, спектральный анализ не позволяет локализовать искаженный участок, а лишь предупреждает о наличии возможных ошибок регистрации.

Распространенной практикой при поиске и разведке месторождений нефти и газа является оценка достоверности результатов геофизических методов исследования на основе их сопоставления друг с другом. При заключении о достоверности результатов всех сопоставляемых геофизических методов, внимание обращается на участки их «рассогласованности», которые могут быть связаны либо с аномальным объектом, либо с наличием помех, либо с технологическим браком, допущенным при выполнении полевых работ или обработке полученных материалов.

Основная идея предлагаемого метода базируется на представлении о том, что ряды измеряемых геофизических величин формируют дискретное изображение исследуемого объекта. В случае исследования схожих объектов даже существенно

отличающиеся (ввиду многих причин) по амплитуде и форме диаграммы-изображения имеют и сходное содержание. Сравнивая тестируемое изображение с подобранными по содержанию изображениями из имеющейся базы эталонов, являющейся аккумулятором опыта эксперта, можно судить о качестве изучаемого материала.

Для определения подобия сигналов обычно применяется некоторая мера расстояния, с помощью которой можно получить численную оценку их сходства. Два сигнала также могут называться подобными, если возможно такое преобразование определяющих параметров, после которых сигналы «выглядят» одинаковыми. Понятие меры сходства широко используется в теории подобия [4, 5, 6], более того, эта задача является одной из основных. Однако проблема формализации этого понятия при решении конкретных задач, как правило, не является тривиальной.

В контексте вышесказанного предлагается следующий критерий эквивалентности изображений геофизических сред с соответствующим алгоритмом вычисления оценки подобия сигналов. В качестве основной характеристики среды предложено считать морфологический градиент, отображающий перепады сигнала [7]. Для числовой оценки подобия двух сигналов, в этом случае, достаточно вычислить площадь фигуры, образуемой взаимными уклонениями кривых градиентов. Для корректного сравнения тестируемого и эталонного сигналов их уровни должны быть нормированы. Это необходимо, поскольку результаты измерений, даже проведенные в одной скважине, могут отличаться по мощности вплоть до порядков. И это возможно, поскольку понятие мощности сигнала не относится к характеристикам каких-либо измеряемых геофизических величин. Таким образом, разработанный алгоритм вычисления оценки подобия двух сигналов равной длины по критерию изображаемой среды имеет следующий вид:

1. Нормирование сигналов по ограничиваемой их кривыми площади.
2. Нахождение градиентов нормированных кривых.
3. Вычисление интеграла модуля разности градиентов.

Для решения поставленной задачи предлагается использовать метод локального усреднения показателя, т.н. метод скользящего окна, который позволяет в определенной мере снять ограничения на протяженность исследуемых объектов и ослабить влияние случайных явлений. Критерий подобия рассчитывается в пределах окна, полученное значение присваивается средней точке. Затем в заданном

направлении окно перемещается на одну точку и операция повторяется. Размер окна зависит от характера распределения и обычно включает от 10 до 50 ближайших точек. Результат подбора участков по разработанному алгоритму представлен на рисунке (рис. 1).

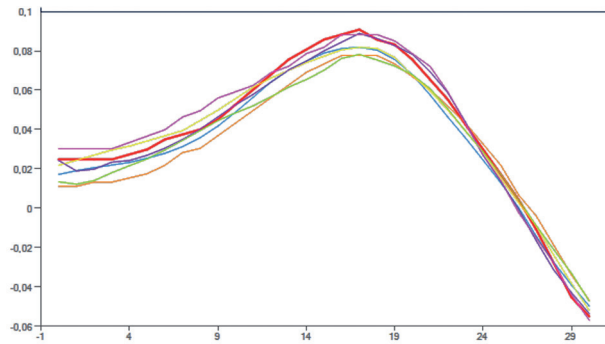


Рис. 1. Результат подбора похожих участков

Следующим этапом предлагаемого метода является исследование тестируемой кривой на предмет выявления особенностей, не присущих эталонным кривым. Для нахождения таких особенностей предлагается применение метода итерационного моделирования неполных данных с помощью многообразий малой размерности [8]. Все версии метода трактуются как построение нейросетевого конвейера, решающего следующие задачи:

- заполнение пробелов в данных;
- ремонт данных, корректировка значений исходных данных так, чтобы наилучшим образом работали построенные модели;
- построение вычислителя, заполняющего пробелы в поступающей на вход строке данных (в предположении, что данные о новых объектах связаны теми же самыми отношениями, что и в исходной таблице).

Рассматриваются три версии метода:

- 1) линейный – с моделированием данных последовательностью линейных многообразий малой размерности;
- 2) квазилинейный – с построением «главных кривых» (или «главных поверхностей»), однозначно проектируемых на линейные главные компоненты;
- 3) существенно нелинейный – основанный на построении «главных кривых» с использованием вариационного принципа; итерационная реализация этого метода близка методу самоорганизующихся карт Кохонена.

Столбец матрицы A есть вектор a с k пробелами, который представляется как k -мерное линейное многообразие L_a , параллельное k координатным

осям, которые соответствуют удаленным данным. При наличии априорных ограничений на пропущенные значения место L_a занимает прямоугольный параллелепипед $P_a \subset L_a$. Построим моделирующее эти данные линейное многообразие малой размерности следующим образом.

За основу возьмем прямую $f(x) = xy + b$, которая задается направляющим вектором y и проходит через точку, определяемую вектором b . Причем, задавая ограничения на значения свободного члена b , мы можем требовать, чтобы прямая проходила или не проходила через начало координат. Далее, расположим эту прямую так, чтобы она наилучшим (в некотором точном смысле) образом приближала исходные данные. Если взять в качестве проектора данных на эту прямую ортогональный проектор, то исходный вектор данных a ортогонально проецируется таким образом в вектор $x = Pr_f(a)$ на полученной прямой (рис. 2).

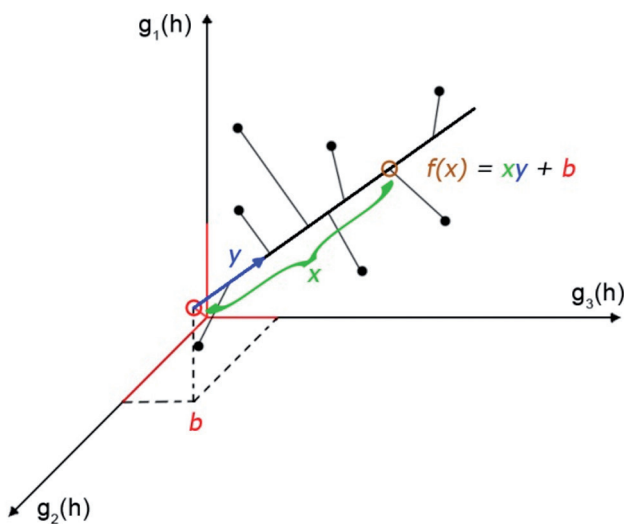


Рис. 2. Моделирующее многообразие малой размерности

Для исходных данных можно посчитать их отклонения от линейной модели, которые находятся из разницы между исходными данными и их проекциями на полученную прямую. Для полученных отклонений также можно построить приближающую наилучшим образом (в определенном точном смысле) прямую, для которой тоже можно рассчитать отклонения. В результате получается итерационный процесс моделирования данных, который заключается в том, что для исходных данных строится наилучшая (в определенном точном смысле) модель – линейное многообразие M малой размерности. Далее из данных A вычитаются проекции $x = Pr_M(a)$. Получаем отклонения от первой модели. Для этого множества отклонений снова строится простая

модель и т.д., пока все отклонения не станут достаточно близки к нулю.

Пусть задана прямоугольная матрица $A = (a_{ij})$, клетки которой заполнены действительными числами или значком @, означающим отсутствие данных. Требуется представить исходную матрицу A в виде суммы одноранговых матриц P_q : $A = \sum_q P_q$, где каждая P_q имеет вид $x_j y_j + b_j$. Следовательно, ставится задача поиска наилучшего приближения A матрицей вида $x_j y_j + b_j$ методом наименьших квадратов:

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq @}} (a_{ij} - x_i y_j - b_j)^2 \rightarrow \min. \quad (1)$$

Решая эту задачу последовательными итерациями по явным формулам, мы получим линию, на которую не накладывается ограничение обязательного прохождения через начало координат. При фиксированных векторах y_j и b_j значения x_i , доставляющие минимум форме (1), однозначно и просто определяются из равенств $\partial\Phi/\partial x_i = 0$:

$$\frac{\partial\Phi}{\partial x_i} = -2 \sum_{\substack{i,j \\ a_{ij} \neq @}} (a_{ij} - x_i y_j - b_j) y_j = 0, \quad (2)$$

$$x_i = \left(\sum_{\substack{j \\ a_{ij} \neq @}} (a_{ij} - b_j) y_j \right) / \left(\sum_{\substack{j \\ a_{ij} \neq @}} (y_j)^2 \right). \quad (3)$$

Аналогично и при фиксированном векторе x_i значения y_j и b_j , доставляющие минимум форме (1), определяются явно из двух равенств $\partial\Phi/\partial y_j = 0$ и $\partial\Phi/\partial b_j = 0$:

$$\frac{\partial\Phi}{\partial y_j} = -2 \sum_{\substack{i \\ a_{ij} \neq @}} (a_{ij} - x_i y_j - b_j) x_i = 0, \quad (4)$$

$$\frac{\partial\Phi}{\partial b_j} = -2 \sum_{\substack{i \\ a_{ij} \neq @}} (a_{ij} - x_i y_j - b_j) = 0. \quad (5)$$

Представляя полученные значения в виде системы, получим:

$$\begin{cases} y_j \sum_{\substack{i \\ a_{ij} \neq @}} x_i^2 + b_j \sum_{\substack{i \\ a_{ij} \neq @}} x_i = \sum_{\substack{i \\ a_{ij} \neq @}} a_{ij} x_i \\ y_j \sum_{\substack{i \\ a_{ij} \neq @}} x_i^2 + b_j \sum_{\substack{i \\ a_{ij} \neq @}} 1 = \sum_{\substack{i \\ a_{ij} \neq @}} a_{ij} \end{cases} \quad (6)$$

Решение полученной системы может быть найдено при помощи метода Крамера или метода квадратного корня [9]. Поскольку процедура является итерационной, то в качестве начального приближения вектора y возьмем случайное значение, но нормированное на 1, а в качестве b возьмем средние значения в соответствующих столбцах исходной матрицы A : y – случайный, нормирован на 1 (т.е. $\sum_j y_j^2 = 1$).

$$b_j = \frac{1}{n_j} \sum_{\substack{i \\ a_{ij} \neq @}} a_{ij}, \text{ где } n_j = \sum_{\substack{i \\ a_{ij} \neq @}} 1. \quad (7)$$

Задавая практически произвольными начальными приближениями для y_j и b_j , ищем значение x_i , далее объявляем неизвестными y_j и b_j , находим их значения при фиксированном x_i и т.д. Эти простые итерации сходятся, так как на каждой итерации происходит уменьшение функционала (1).

Несмотря на уменьшение функционала на каждой итерации, возможна ситуация, когда процесс идет так медленно, что относительное уменьшение значения на каждой итерации близко к нулю. В этом случае необходимо ввести дополнительный критерий остановки – малость относительного улучшения $\Delta\Phi/\Phi$, где $\Delta\Phi$ – полученное за цикл уменьшение значения Φ , а Φ – само текущее значение. Таким образом, процедура останавливается, если $\Delta\Phi/\Phi < \varepsilon$ или $\Phi < \delta$ для некоторых $\varepsilon, \delta < 0$.

Решая задачу (1), для данной матрицы A находим наилучшее приближение матрицей P_1 вида $x_i y_j + b_j$. Далее, из матрицы A вычитаем полученную матрицу P_1 и для полученной матрицы уклонений $A - P_1$ вновь ищем наилучшее приближение P_2 этого же вида и т.д. Контроль ведется по остаточной дисперсии столбцов.

В результате исходная матрица данных A представляется в виде суммы матриц P_q , т.е. $A = P_1 + P_2 + \dots + P_q$. Если пробелы отсутствуют, т.е. все значения a_{ij} известны, то описанный метод приводит к методу главных компонент – сингулярному разложению центрированной исходной таблицы данных. В этом случае начиная с $q=2$, $P_q = x^q y_j^q (b=0)$. В общем случае это не так. *Следует обратить особое внимание на то, что центрирование (переход к нулевым средним) к данным с пробелами неприменимо.*

С использованием Q полученных факторов можно решать задачи заполнения пропусков в таблице и ремонта искаженных значений:

Q-факторное заполнение пропусков: пропущенные значения в исходной матрице A определяются из суммы Q полученных матриц вида $x_i y_j + b_j$;

Q-факторный «ремонт» таблицы: значения в исходной матрице заменяются на сумму Q полученных матриц вида $x_i y_j + b_j$.

При отсутствии пробелов полученные прямые будут ортогональны, и мы получим ортогональную систему факторов. Для неполных данных это не так, но возможен процесс ортогонализации полученной системы факторов, который, к примеру, заключается в том, что исходная таблица восстанавливается при помощи полученной системы факторов, после чего эта система пересчитывается заново, но уже на дополненных данных.

Когда для исходной матрицы A имеется набор исчерпывающих ее с заданной точностью матриц P_q ,

то, как это было показано в предыдущем параграфе, операции заполнения пропусков и «ремонта» данных не представляют особого труда. Однако часто возникает необходимость обработать отдельно взятый вектор данных, при условии, что данные в нем связаны теми же соотношениями, что и в исходной матрице.

На основании этого опишем операцию восстановления данных в поступающей на обработку строке a_j с пробелами (некоторые $a_j = @$). Пусть построена последовательность матриц P_q вида $x_i y_j + b_j$ ($P_q = x_i^q y_j^q + b_j^q$), исчерпывающая исходную матрицу A с заданной точностью. Для каждого q по заданной строке определим число $x^q(a)$ и вектор a^q :

$$x^q(a) = \frac{\sum_{j, a_j \neq @} (a_j^{q-1} - b_j^q) y_j^q}{\sum_{j, a_j \neq @} (y_j^q)^2}; \tag{8}$$

$$a_j^q = a_j^{q-1} - x^q(a) y_j^q - b_j^q, (a_j \neq @).$$

Здесь многообразие M – прямая, координаты точек на M задаются параметрическим уравнением $z_j = t y_j + b_j$, а проекция $Pr_M(a)$ определяется согласно (8):

$$Pr(a) = t(a) y_j + b_j, \tag{9}$$

$$t(a) = \frac{\sum_{j, a_j \neq @} (a_j - b_j) y_j}{\sum_{j, a_j \neq @} (y_j)^2}.$$

То же самое в векторной форме:

$$t(a) = \frac{(a - b, y)_a}{\|y\|_a^2}. \tag{10}$$

Для Q -факторного восстановления данных полагаем:

$$\bar{a}_j = \sum_{q=1}^Q x^q(a) y_j^q + b_j^q, (a_j \neq @). \tag{11}$$

Геометрическая интерпретация этой процедуры состоит в том, что для заданного входного вектора a находим его проекцию a_1 на прямую $f_1(t) = t y_1 + b_1$. Далее, вычитая из вектора полученную проекцию, находим его уклонение от прямой f_1 . Для этого уклонения снова ищем проекцию a_2 , но уже на прямую $f_2(t) = t y_2 + b_2$ и т.д. Таким образом, исходный вектор представляется в виде суммы Q векторов:

$$a = \sum_{q=1}^Q a^q.$$

В итоге, используя это разложение, восстанавливаются пропущенные значения в поступающем на обработку векторе a , а также находятся «исправленные» оценки уже известных его значений.

Для построения линейных многообразий размерности больше 1 используются совершенно

аналогичные формулы. Но так как в составе одного многообразия размерности больше 1 задающие его векторы могут быть не ортогональными, то для облегчения интерпретации полученных факторов рекомендуется проводить соответствующую ортогонализацию.

Остановимся на двумерных и трехмерных линейных многообразиях, для которых опишем процедуры построения и ортогонализации.

Однако ко всему сказанному хочется заметить, что использование многомерных линейных моделей не приносит ничего существенно нового. Увеличение размерности линейных моделей может дать эффект только при использовании их соответствующих квазилинейных форм.

Пусть дано подпространство размерности n и в нем система линейно независимых векторов $\{y_k\}$, $k = 1 \dots n$. По определению, они образуют базис этого подпространства. Требуется провести процедуру ортогонализации этого базиса, т.е. построить такой новый базис этого подпространства $\{\bar{y}_k\}$ ($k = 1 \dots n$), чтобы $(\bar{y}_n) = 0$ для любых $i \neq j$.

Пусть набор векторов $\{y_k\}$ ($k = 1 \dots n-1$) уже ортогонализирован, а y_n – еще нет. Тогда построим новый вектор и потребуем, чтобы $(\bar{y}_n, y_k) = 0$ для всех $k=1 \dots n-1$. В результате имеем систему из $n-1$ уравнения с $n-1$ неизвестными:

$$\left(\sum_{i=1}^n \alpha_i y_i + y_n, y_k \right) = 0, \quad k=1 \dots n-1.$$

Учитывая, что $(y_i, y_k) = 0$ при $i \neq k$, то легко заметить, что полученная система будет диагональной. Отсюда неизвестные легко находятся по формулам:

$$\alpha_k = -\frac{(y_n, y_k)}{(y_k, y_k)}.$$

Следовательно, если задана система из n векторов, в которой имеется ортогональная подсистема из $n-1$ вектора, то оставшийся вектор «ортогонализируется» по формуле:

$$\bar{y}_n = y_n - \sum_{i=1}^{n-1} \frac{(y_n, y_i)}{(y_i, y_i)} y_i. \quad (12)$$

Тогда процедура ортогонализации системы из n базисных векторов выглядит как последовательная ортогонализация систем из 2, 3, ..., n векторов с использованием (12).

Для построения двумерного линейного многообразия минимизируем квадратичную форму:

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq @}} (a_{ij} - x_{1j} y_{1j} - x_{2i} y_{2j} - b_j)^2 \rightarrow \min. \quad (13)$$

Решение дается последовательными итерациями по явным формулам. При фиксированных y_{1j}, y_{2j} и b_j значения x_{1i} и x_{2i} однозначно находятся из системы равенств $\partial\Phi/\partial x_{1i} = 0$ и $\partial\Phi/\partial x_{2i} = 0$:

$$\begin{cases} \frac{\partial\Phi}{\partial x_{1i}} = -2 \sum_{\substack{j \\ a_{ij} \neq @}} (a_{ij} - x_{1i} y_{1j} - x_{2i} y_{2j} - b_j) y_{1j} = 0, \\ \frac{\partial\Phi}{\partial x_{2i}} = -2 \sum_{\substack{j \\ a_{ij} \neq @}} (a_{ij} - x_{1i} y_{1j} - x_{2i} y_{2j} - b_j) y_{2j} = 0. \end{cases}$$

То же самое в векторном виде:

$$\begin{cases} x_{1i} (y_{1, y_1})_{ai} + x_{2i} (y_{1, y_2})_{ai} = (a_i - b, y_1)_{ai}, \\ x_{1i} (y_{1, y_2})_{ai} + x_{2i} (y_{2, y_2})_{ai} = (a_i - b, y_2)_{ai}. \end{cases}$$

Аналогично при фиксированных x_{1i} и x_{2i} значения, доставляющие минимум квадратичной форме (19), однозначно находятся из равенств $\partial\Phi/\partial y_{1j} = 0$, $\partial\Phi/\partial y_{2j} = 0$ и $\partial\Phi/\partial b_j = 0$:

$$\begin{cases} \frac{\partial\Phi}{\partial y_{1j}} = -2 \sum_{\substack{i \\ a_{ij} \neq @}} (a_{ij} - x_{1i} y_{1j} - x_{2i} y_{2j} - b_j) x_{1i} = 0, \\ \frac{\partial\Phi}{\partial y_{2j}} = -2 \sum_{\substack{i \\ a_{ij} \neq @}} (a_{ij} - x_{1i} y_{1j} - x_{2i} y_{2j} - b_j) x_{2i} = 0, \\ \frac{\partial\Phi}{\partial b_j} = -2 \sum_{\substack{i \\ a_{ij} \neq @}} (a_{ij} - x_{1i} y_{1j} - x_{2i} y_{2j} - b_j) = 0. \end{cases}$$

То же самое в векторной форме:

$$\begin{cases} y_{1j} (x_1, x_1)_{aj} + y_{2j} (x_1, x_2)_{aj} + b_j (x_1, 1)_{aj} = (a_j, x_1)_{aj}, \\ y_{1j} (x_1, x_2)_{aj} + y_{2j} (x_2, x_2)_{aj} + b_j (x_2, 1)_{aj} = (a_j, x_2)_{aj}, \\ y_{1j} (x_1, 1)_{aj} + y_{2j} (x_2, 1)_{aj} + b_j (1, 1)_{aj} = (a_j, 1)_{aj}. \end{cases}$$

Полученная система решается любым численным способом, например, методом квадратного корня [9] (т.к. полученная матрица является симметричной).

Учитывая, что полученные векторы y_1 и y_2 могут быть не ортогональны, в некоторых случаях их необходимо ортогонализировать. А так как они образуют базис подпространства размерности 2, то эта задача легко решается с использованием (12).

Для построения трехмерного линейного многообразия минимизируем квадратичную форму:

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq @}} (a_{ij} - x_{1j} y_{1j} - x_{2i} y_{2j} - x_{3i} y_{3j} - b_j)^2 \rightarrow \min. \quad (14)$$

Решение дается последовательными итерациями по явным формулам. При фиксированных y_{1j}, y_{2j}, y_{3j} и b_j значения x_{1i}, x_{2i} и x_{3i} однозначно находятся из системы равенств $\partial\Phi/\partial x_{1i} = 0$, $\partial\Phi/\partial x_{2i} = 0$ и $\partial\Phi/\partial x_{3i} = 0$:

$$\begin{cases} \frac{\partial \Phi}{\partial x_{1i}} = -2 \sum_{\substack{j \\ a_{ij} \neq @}} (a_{ij} - x_{1i}y_{1j} - x_{2i}y_{2j} - x_{3i}y_{3j} - b_j)y_{1j} = 0, \\ \frac{\partial \Phi}{\partial x_{2i}} = -2 \sum_{\substack{j \\ a_{ij} \neq @}} (a_{ij} - x_{1i}y_{1j} - x_{2i}y_{2j} - x_{3i}y_{3j} - b_j)y_{2j} = 0, \\ \frac{\partial \Phi}{\partial x_{3i}} = -2 \sum_{\substack{j \\ a_{ij} \neq @}} (a_{ij} - x_{1i}y_{1j} - x_{2i}y_{2j} - x_{3i}y_{3j} - b_j)y_{3j} = 0. \end{cases}$$

То же самое в векторном виде:

$$\begin{cases} x_{1i}(y_1, y_1)_{a_i} + x_{2i}(y_1, y_2)_{a_i} + x_{3i}(y_1, y_3)_{a_i} = (a_i - b, y_1)_{a_i}, \\ x_{1i}(y_1, y_2)_{a_i} + x_{2i}(y_2, y_2)_{a_i} + x_{3i}(y_2, y_3)_{a_i} = (a_i - b, y_2)_{a_i}, \\ x_{1i}(y_1, y_3)_{a_i} + x_{2i}(y_2, y_3)_{a_i} + x_{3i}(y_3, y_3)_{a_i} = (a_i - b, y_3)_{a_i}. \end{cases}$$

Для лучшего приближения исходных данных можно подобрать такую гладкую вектор-функцию, значения переменных которой определяются через проекции данных на уже построенное многообразие, что суммарное значение квадратов уклонений будет минимальным среди всех возможных функций данного класса (при одинаковых ограничениях на гладкость). Такой тип линий называется квазилинейным.

Построение квазилинейных моделей, наилучшим (в определенном точном смысле) образом приближающих данные, предлагается проводить в несколько этапов. Пусть, как и в случае линейных моделей, задана таблица с пропусками $A = (a_{ij})$, т.е. некоторые $a_{ij} = @$.

1. *Построение линейной модели*: решение задачи (1). Для определенности полагаем, что $(y, b) = 0$, $(y, y) = 1$.

2. *Интерполяция (сглаживание)*: строится вектор-функция $f(t)$, минимизирующая функционал:

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq @}} (a_{ij} - f_j(a_{ik}y_k))^2 + \alpha \int_{-\infty}^{+\infty} (f''(t))^2 dt, \quad (15)$$

где $\alpha > 0$ – параметр сглаживания.

Для решение этой задачи могут быть применены полиномы небольшой степени, кубические сплайны или функция Карлемана [10].

3. *Экстраполяция*: самая простая экстраполяция полученной вектор-функции $f(t)$ может быть получена при использовании касательных к полученной функции на концах интервала. Намного более интересным представляется использование для экстраполяции формул Карлемана.

Таким образом, сглаженная вектор-функция $f(t)$ экстраполируется с некоторого конечного множества $\{t_k\}$ (которое не обязательно связано с проекциями на прямую $z_j = ty_j + b_j$ исходных строк данных) на всю вещественную прямую с использованием формул Карлемана (с помощью формул Карлемана экстраполируется отклонение кривой $f(t)$ от прямой $ty + b$).

$$f(t) \approx ty + b + \sum_{k=1}^m (f(t_k) - t_k y - b) \times \frac{2(e^{\lambda t} - e^{\lambda t_k})}{\lambda(e^{\lambda t} + e^{\lambda t_k})(t - t_k)} \prod_{\substack{j=1 \\ j \neq k}}^m \frac{(e^{\lambda t} + e^{\lambda t_k})(e^{\lambda t} + e^{\lambda t_k})}{(e^{\lambda t} + e^{\lambda t_k})(e^{\lambda t} + e^{\lambda t_k})}, \quad (16)$$

где λ – параметр метода, характеризующий, насколько широка полоса на плоскости комплексных чисел, в которой гарантированно голоморфна экстраполируемая функция (эта ширина равна π/λ).

Процедура использования квазилинейных моделей несколько отличается от аналогичной процедуры в линейном случае, хоть и базируется на ее основе. Точка на построенной кривой $f(t)$, соответствующая полному («комплектному») вектору данных a строится как $f(a, y)$. В этом и заключается квазилинейность метода: сначала ищется проекция вектора данных на прямую $Pr(a) = ty + b$, $t = (a, y)$, а затем строится точка на кривой $f(t)$. Также и для неполных векторов данных – сначала ищется на прямой ближайшая точка $t(a)$ (проекция неполного вектора a), а затем – соответствующая точка на кривой $f(t)$, при $t = t(a)$.

После построения кривой $f(t)$ из данных вычитаются их проекции, то есть матрица данных заменяется на матрицу уклонений от модели. Далее снова ищется наилучшее линейное приближение (к примеру, вида $x_j y_j + b_j$) для матрицы уклонений, вновь строится сглаживание, потом – экстраполяция по Карлеману и т.д., пока уклонения не приблизятся в достаточной степени к нулю. Критерием остановки опять же могут выступать остаточные дисперсии. Таким образом, процедура моделирования данных квазилинейными многообразиями малой размерности тоже является итерационной, как и в линейном случае. В результате исходная таблица предстает в виде Q -факторной модели:

$$a_{ij} \cong \sum_q f_j^q(t_i^q), \quad (17)$$

где аргументом функции служит специальным образом нормированное скалярное произведение исходного данного на линейную основу квазилинейного многообразия или, другими словами, значение проекции данного на прямую.

Заметим также, что если $a_{ij} \neq @$, то эта формула аппроксимирует исходные данные, иначе она дает способ восстановления пропущенных значений.

Самоорганизующиеся карты Кохонена (Self-Organizing Map – SOM) [12] – это модифицированный алгоритм линейного векторного квантования данных, т.е. представления N точек данных с помощью меньшего числа точек-ядер. Каждое из ядер заменяет собой локальное сгущение данных (рис. 3), назовем это сгущение таксоном. В результате такой замены данные представляются с определенной

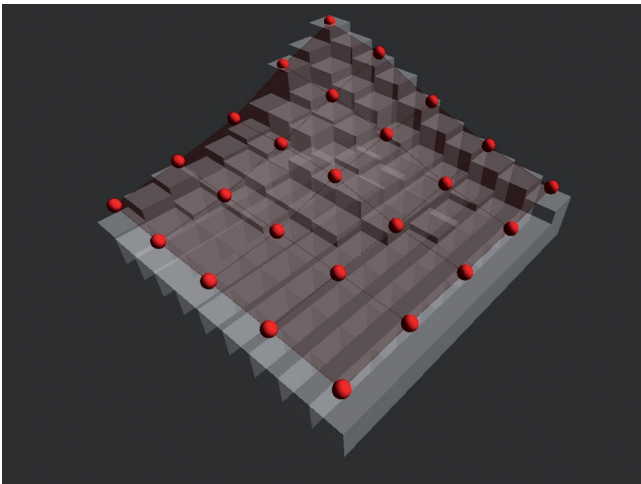


Рис. 3. Аппроксимация SOM набора данных

ошибкой аппроксимации – среднеквадратичного расстояния от точки данных до ближайшего к ней образца.

Пусть SOM определяется набором точек (ядер) $Y = \{y_{ij}\}$ ($i, j = 1 \dots m$). Будем считать, что сетка квадратная, но для прямоугольной все описанные формулы аналогичны), последовательно расположенных на квадратной сетке и требуется отобразить на ней набор точек данных $X = \{x_i\}$. Введем преобразование Π , которое каждому вектору $x \in X$ сопоставляет ближайшую к нему точку из Y :

$$x \xrightarrow{\Pi} y_{ij}, |y_j - x|^2 \rightarrow \min, \quad (18)$$

каждому ядру y_{ij} сопоставляется его таксон

$$K_{ij} = \{x \in X \mid x \xrightarrow{\Pi} y_{ij}\}. \quad (19)$$

Минимизируемый функционал будет состоять из следующих слагаемых:

$$D_1 = \sum_{ij} \sum_{x \in K_{ij}} \|x - y_{ij}\|^2; \quad (20)$$

$$D_2 = \sum_{ij} \|y_{ij} - y_{i+1,j}\|^2 + \sum_{ij} \|y_{ij} - y_{i,j+1}\|^2; \quad (21)$$

$$D_3 = \sum_{ij} \|2y_{ij} - y_{i-1,j} - y_{i+1,j}\|^2 + \sum_{ij} \|2y_{ij} - y_{i,j-1} - y_{i,j+1}\|^2. \quad (22)$$

Таким образом, для построения SOM требуется минимизировать функционал:

$$D = \frac{D_1}{|X|} + \lambda \frac{D_2}{m^2} + \mu \frac{D_3}{m^2}, \quad (23)$$

где λ, μ – параметры связности и нелинейности – «модули упругости» (деление на число точек $|X|$ и число ядер m означает нормировку «на одно слагаемое» и позволяет для выборок разной мощности использовать одинаковые способы изменения λ и μ).

Пусть метрика является евклидовой. В этом случае функционал D является квадратичным по положениям узлов y^{ij} , это значит, что при заданном разбиении множества точек данных на таксоны для его минимизации потребуется решить систему линейных

уравнений размерами $pq \times pq$. Следовательно, эффективным методом минимизации функционала D окажется такой алгоритм:

1. Узлы сетки так или иначе располагаются в пространстве данных.
2. При заданных положениях узлов сетки производится разбиение множества точек данных на таксоны – подмножества K_{ij} .
3. При заданном разбиении множества точек данных на таксоны производится минимизация функционала D .

Шаги 2 и 3 повторяются до тех пор, пока функционал D не перестанет изменяться (в пределах заданной точности). Процесс сходится, поскольку на каждом этапе минимизации величина D , будет уменьшаться, вместе с тем она ограничена снизу нулем (величина D неотрицательна). Более того, процесс сходится за конечное число шагов, поскольку число вариантов разбиения точек данных на таксоны конечно.

Выпишем явно коэффициенты матрицы системы линейных уравнений, которую необходимо решать на каждой итерации алгоритма минимизации.

Непосредственное дифференцирование дает:

$$\frac{1}{2} \frac{\partial D}{\partial y_{kl}} = a_{kl}^{(-2)} y^{k-2,l} + a_{kl}^{(-1)} y^{k-1,l} + a_{kl} y^{kl} + a_{kl}^{(+1)} y^{k+1,l} + a_{kl}^{(+2)} y^{k+2,l} + b_{kl}^{(-2)} y^{k,l-2} + b_{kl}^{(-1)} y^{k,l-1} + b_{kl}^{(+1)} y^{k,l+1} + b_{kl}^{(+2)} y^{k,l+2} - \sum_{x \in K_{kl}} x, \quad (24)$$

где

$$a_{kl}^{(-2)} = \frac{\mu}{pq} (1 - \delta_{l,2})(1 - \delta_{l,1}),$$

$$a_{kl}^{(-1)} = \frac{\lambda}{pq} (\delta_{l,1} - 1) + \frac{2\mu}{pq} (\delta_{l,2} - 1)(1 - \delta_{l,1}),$$

$$a_{kl} = \frac{n_{kl}}{|X|} + \frac{\lambda}{pq} [4 - \delta_{k,1} - \delta_{k,p} - \delta_{l,1} - \delta_{l,q}] +$$

$$+ \frac{\mu}{pq} \left[(1 - \delta_{k,2})(1 - \delta_{k,1}) + (1 - \delta_{k,p-1})(1 - \delta_{k,p}) - 2(1 - \delta_{k,p})(1 - \delta_{k,1}) + (1 - \delta_{l,2})(1 - \delta_{l,1}) + (1 - \delta_{l,q-1})(1 - \delta_{l,q}) - 2(1 - \delta_{l,q})(1 - \delta_{l,1}) \right].$$

$$a_{kl}^{(+1)} = \frac{\lambda}{pq} (\delta_{l,q} - 1) + \frac{2\mu}{pq} (\delta_{l,q-1} - 1)(1 - \delta_{l,q}),$$

$$a_{kl}^{(+2)} = \frac{\mu}{pq} (1 - \delta_{l,q-1})(1 - \delta_{l,q}),$$

$$b_{kl}^{(-2)} = \frac{\mu}{pq} (1 - \delta_{k,2})(1 - \delta_{k,1}),$$

$$b_{kl}^{(-1)} = \frac{\lambda}{pq} (\delta_{k,1} - 1) + \frac{2\mu}{pq} (\delta_{k,2} - 1)(1 - \delta_{k,1}),$$

$$b_{kl}^{(+1)} = \frac{\lambda}{pq} (\delta_{k,p} - 1) + \frac{2\mu}{pq} (\delta_{k,p-1} - 1)(1 - \delta_{k,p}),$$

$$b_{kl}^{(+2)} = \frac{\mu}{pq} (1 - \delta_{k,p-1})(1 - \delta_{k,p}),$$

где n_{kl} – число элементов в таксоне K_{kl} , δ_{ij} – символ Кронекера, множители вида $(1 - d_{ij})$ введены для того, чтобы учесть «краевые эффекты». Если индексы k, l при y^{kl} не соответствуют никакому узлу сетки, то этот множитель автоматически обратит это слагаемое в ноль.

Уравнения $\frac{\partial D}{\partial y^{kl}} = 0, k = 1 \dots p, l = 1 \dots q$, дают

m систем линейных уравнений (по одной на каждую из m компонент векторов y^{kl}).

«Вытянем» набор ядер y^{kl} в один столбец.

В результате вектор неизвестных будет

$$x = (y_{11} \dots y_{1q}, y_{21} \dots y_{2q}, \dots, y_{(p-1)1} \dots y_{(p-1)q}, y_{p1} \dots y_{pq}).$$

Система имеет вид $Ax = b$, где s -я компонента вектора свободных членов равна

$$b_s = \frac{\sum_{x \in K_{ij}} x}{|X|}, i = \left[\frac{s-1}{q} \right] + 1, j = \left[\frac{s-1}{q} \right] q,$$

где $[\dots]$ – операция взятия целой части числа.

$$A_{st} = \begin{cases} a_{kl}, i=k, j=l \\ a_{kl}^{(-1)}, i=k-1, j=l \\ a_{kl}^{(+1)}, i=k+1, j=l \\ a_{kl}^{(-2)}, i=k-2, j=l \\ a_{kl}^{(+2)}, i=k+2, j=l \\ b_{kl}^{(-2)}, i=k, j=l-2 \\ b_{kl}^{(-1)}, i=k, j=l-1 \\ b_{kl}^{(+1)}, i=k, j=l+1 \\ b_{kl}^{(+2)}, i=k, j=l+2 \\ 0, else \end{cases}, \text{ где } \begin{cases} i = \left[\frac{s-1}{q} \right] + 1 \\ j = s \left[\frac{s-1}{q} \right] q \\ k = \left[\frac{t-1}{q} \right] + 1 \\ l = t - \left[\frac{t-1}{q} \right] q \end{cases}$$

Алгоритм выявления локальных особенностей кривых на основе методов заполнения пробелов был построен нами следующим образом:

1. Нормирование подобранных на предыдущем этапе кривых по мощности.
2. Удаление участка тестируемой кривой, качество которого подвергается сомнению.
3. Формирование из кривых матрицы (поверхности) с пробелами.
4. Восстановление участка по имеющимся избыточным данным.
5. Вычисление интеграла квадрата разности исходной и восстановленной кривой.

Оценка отклонения предстанет числом прямо пропорциональным величине неуверенности системы в качестве данного участка (рис. 5).

С целью улучшения работы используемых методов восстановления данных формирующие матрицу кривые ранжируются по критерию образования максимально «гладкой» поверхности (рис. 4). Подобранные кривые уже обладают достаточным сходством, поэтому разработанный алгоритм

построения «гладкой» поверхности представлен следующей последовательностью операций:

1. Нормирование подобранных кривых по мощности.
2. Нахождение градиентов нормированных кривых.
3. Вычисление среднего квадратического отклонения кривой градиента.
4. Ранжирование кривых по полученной величине.
5. Выравнивание центров масс по вертикальной оси.

Результаты вычислительных экспериментов позволяют утверждать о применимости и конкурентоспособности предлагаемого метода. При наличии слабых гармонических помех в наблюдениях он, по-видимому, является наиболее предпочтительным. Участок с помехами явно выделялся аномально высоким уровнем среднеквадратического отклонения (рис. 5). Реализация подобного подхода к решению задачи автоматизированного контроля качества записи данных способна в значительной мере снизить вероятность получения недостоверных результатов ГИС. Дальнейшее развитие предложенного решения с включением моделей сред в качестве контролируемых данных позволит строить новые интерпретационные схемы в скважинной геофизике.

Ключевые слова: геофизические исследования, контроль качества, итерационное моделирование, критерии подобия, автоматизация.

ЛИТЕРАТУРА

1. ГОСТ Р 54362-2011. Геофизические исследования скважин. Термины и определения. – Введ. 2012-07-26. – URL: <http://standartgost.ru/ГОСТ%20Р%2054362-2011> (дата обращения: 09.02.2014).
2. Губерман Ш.А., Овчинникова М.И. Некоторые возможности использования статистических характеристик геологических разрезов // Изв. АН СССР. Сер. Геофиз. – 1964. – № 7. – С. 87-94.
3. Никитин А.А. Статистические методы выделения геофизических аномалий. – М. : Недра, 1979. – 280 с.
4. Веников В.А. Теория подобия и моделирование. – М. : Высшая школа, 1968.
5. Седов Л.И. Методы подобия и размерности в механике. – 10-е изд., доп. – М. : Наука. Гл. ред. физ.-мат. лит., 1987. – 432 с.
6. Герасимато Ф.Г. Теория подобия [Электронный ресурс] // Академия Тринитаризма. – М., Эл № 77-6567. – Публ. № 17850, 21.01.2013. – URL: <http://www.trinitas.ru/rus/doc/0016/001d/00162064.htm> (дата обращения: 12.03.2014).
7. Сергиенко А.Б. Цифровая обработка сигналов : учебник для вузов. – СПб. : Питер, 2002. – 606 с.

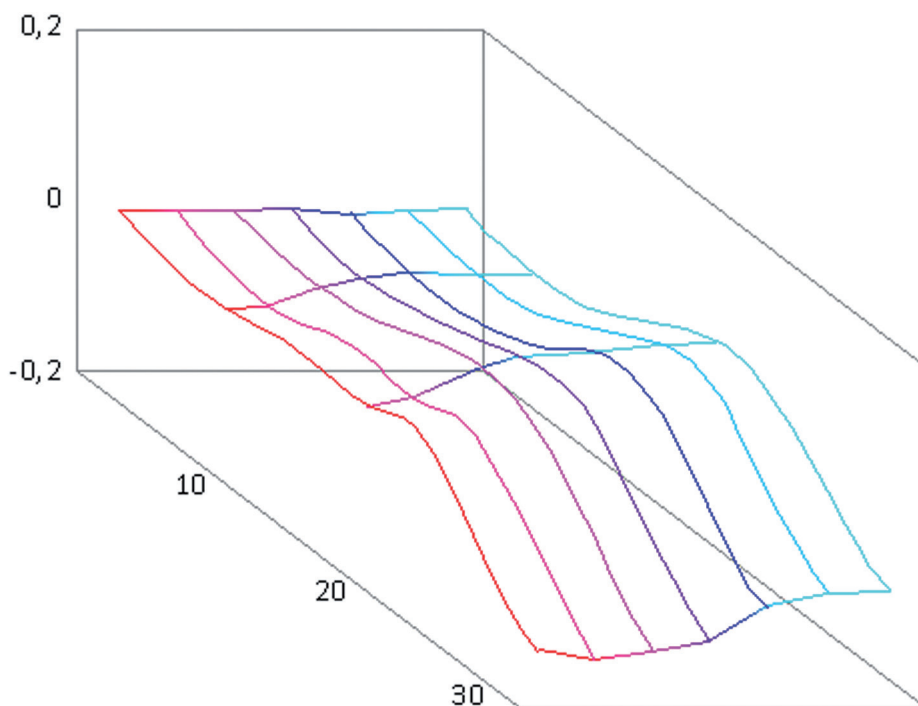


Рис. 4. «Гладкая» поверхность, формируемая ранжированными кривыми

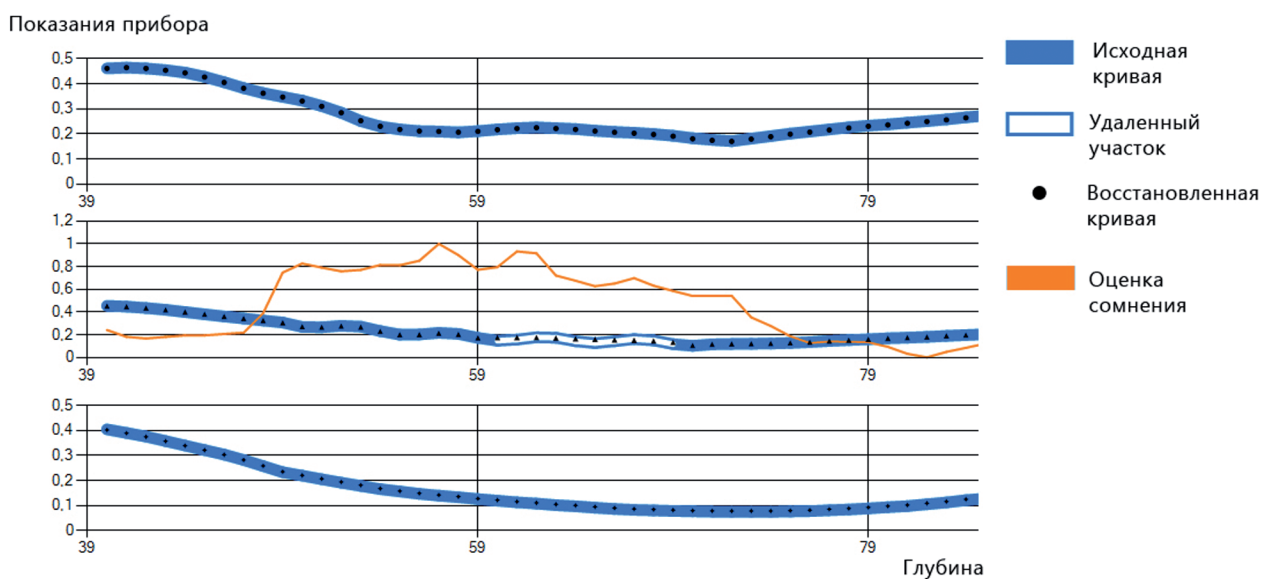


Рис. 5. Оценка сомнения в качестве участка, искаженного гармоническими колебаниями

8. Россиев А.А. Итерационное моделирование неполных данных с помощью многообразий малой размерности : дис. ... канд. физ.-мат. наук : 05.13.16. – Красноярск, 2000. – 121 с.
 9. Самарский А.А. Введение в численные методы. – М. : Наука. Гл. ред. физ.-мат. лит., 1982. – 272 с.

10. Айзенберг Л.А. Формулы Карлемана в комплексном анализе. Первые приложения. – Новосибирск : Наука, 1990. – 248 с.
 11. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. – Новосибирск : Наука, 1996. – 276 с.
 12. Зиновьев А.Ю. Визуализация многомерных данных. – Красноярск : Изд-во КГТУ, 2000. – 168 с.