

**К вопросу автоматизированного контроля качества данных геофизических исследований скважин\*****Р. В. Пельмегов, А. Г. Куделин**

*Работа посвящена развитию научных основ, методов и технологий первичного автоматизированного контроля регистрируемых данных при проведении геофизических исследований в скважинах. Решение вопроса о прекращении каротажа на определённой скважине зачастую связано с субъективными оценками экспертов. Авторы аргументируют необходимость использования автоматизированных средств обнаружения ошибок и контроля качества регистрируемых данных в процессе или непосредственно по завершению исследований. Приводится краткий анализ симптомов ненадёжности фрагментов записи. Предлагается алгоритм, использующий методы итерационного моделирования неполных данных с помощью многообразий малой размерности, для решения задачи оценки качества записи данных геофизических исследований скважин. Приводятся примеры, демонстрирующие высокую эффективность методов итерационного моделирования неполных данных с помощью многообразий малой размерности в задаче заполнения пробелов данных геофизических исследований, а также результаты численного эксперимента по решению задачи контроля качества данных электрического каротажа.*

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

**Введение.** Одной из важнейших задач нефтепромысловой геофизики является повышение точности и достоверности количественной интерпретации промыслово-геофизических данных. Решение этой задачи невозможно без достоверных данных геофизических исследований скважин (далее — ГИС) [1].

Первичная обработка и оценка качества полевых данных, оценка материалов производится оператором по исследованию скважин визуально, непосредственно на объекте. Такая оценка требует от оператора значительных познаний в теории геофизических методов исследования скважин, навыков интерпретации ГИС, а также большого опыта работы. Однако большинство операторов, по тем или иным причинам, либо не обладают требуемой квалификацией, либо просто не имеют возможности оценить качество полученных данных вследствие большого объёма проводимых исследований, сложных условий исследования [2], сложных и незнакомых геофизических разрезов скважин и т. д. При выявлении «брака» материала на этапе интерпретации геофизических исследований, требуется повторное проведение исследований, что связано с весьма значительными финансовыми расходами.

В статье «A Hierarchical Approach to Improving Data Quality» [3] присутствует во многом схожая проблематика задачи контроля качества данных. Предлагаемый авторами подход к решению задачи автоматизированного контроля качества геоинформационных данных основан на сопоставлении изображений одной местности, полученных из разных источников. Критерием качества считаются величины взаимных уклонений пар матриц смежности графов, представляющих наборы данных. Подбор и сопоставление изображений выполняется экспертами вручную, являясь при этом актом формализации их опыта, поскольку основная проблема инженерии знаний — это процесс извлечения знаний.

\* Работа выполнена по тематическому плану Министерства образования и науки РФ № 2.3.13 «Метод сбалансированной дискретизации для задач имитационного моделирования динамических процессов в распределённых объектах».

**Постановка задачи.** Существует проблема «ранней» оценки качества зарегистрированного материала ГИС, то есть оценки качества в процессе записи и непосредственно после окончания записи. Таким образом, использование автоматических средств обнаружения ошибок и контроля качества регистрируемых данных становится весьма актуальным.

Работа посвящена развитию научных основ, методов и технологий первичного автоматизированного контроля регистрируемых данных при проведении геофизических исследованиях в скважинах. Симптомами ненадёжности фрагментов записи выступают:

1) значения показаний приборов, не имеющие смысла с геолого-геофизической точки зрения, которые могут быть связаны со срывами, выбросами, затяжками, утечками в кабеле, ошибками аналого-цифрового преобразования [4, 5] и т. д.;

2) противоречивые показания приборов, проявляющиеся в расхождениях показаний, полученных различными методами исследования, либо при разных условиях работы одного метода. Например, показания электрических методов на постоянном и переменном токе, либо несоответствие показаний разных зондов при исследовании на постоянном токе;

3) несоответствие показаний приборов априорной информации, имеющейся об объекте исследований. Например, отсутствие артефактов на записи электромагнитных методов при прохождении интервалов перфорации колонны.

Целью автоматизированного контроля качества является выявление вышеуказанных проблемных ситуаций, которые определяют информационную нерегулярность данных и привлечение к ним внимания оператора. Достижение цели обеспечивается обнаружением с некоторой, заранее заданной вероятностью, симптомов проблемных ситуаций в процессе записи данных ГИС, а также непосредственно по завершению исследования.

Скважинные геофизические исследования с математической точки зрения описываются, как функции  $g = f(h)$ , где  $g$  — измеряемый геофизический параметр,  $h$  — глубина. Симптомы нарушения информационной регулярности записей каротажных данных  $g = f(h)$  могут быть определены с помощью методов итерационного моделирования неполных данных с использованием многообразий малой размерности [6] для автоматизированного контроля качества данных геофизических исследований.

**Алгоритм оценки качества материалов с применением метода моделирования неполных данных.** Рассмотрим алгоритм решения задачи оценки качества записи данных ГИС с применением метода восстановления на примере данных метода кажущихся сопротивлений (КС). Скважинные исследования методом КС основаны на расчленении пород, окружающих скважину, по их удельному электрическому сопротивлению (УЭС). При исследованиях методом КС может регистрироваться либо сила тока (токовый каротаж), либо разность потенциалов. В результате каротажа получают токовые диаграммы, характеризующие изменение силы тока по стволу скважины. Исходя из предположения, что большая часть данных не искажена, алгоритм может иметь следующий вид:

1. Из исходного набора данных выбирается окно — окрестность исследуемой точки;
2. Часть данных строки содержащей исследуемую точку удаляется;
3. Проводится восстановление удалённого участка;
4. Восстановленная кривая сравнивается с исходной по критерию минимума среднеквадратических отклонений.

Шаги повторяются для каждой точки исходного набора данных (рис. 1).

Пусть имеется двумерный прямоугольный числовой массив  $G$ , строки которого представляют собой функции  $g = f(h)$  КС пород скважины, заданные в дискретной форме с одинаковым интервалом и нормированные на диапазон  $[0...1]$  независимо друг от друга. Таким образом, каж-

дый столбец этой матрицы есть вектор данных  $a$  — «изображение» исследуемой среды различными зондами [7–9]. Достоверность каждой точки матрицы  $G$  подвергнута сомнению и может быть оценена за счёт скрытых зависимостей между точками её окрестности.

Представим  $G$  как одномерный массив векторов  $a$  и выберем интервал — окрестность вектора, содержащего исследуемую точку. Таким образом получаем «окно» — матрицу  $A$  размера  $m \times n$ , где  $m$  (количество строк) — размерность вектора  $a$ ,  $n$  (количество столбцов) — длина интервала, содержащего окрестность точки. Слева и справа от исследуемой точки, в содержащей её строке, удалим участки данных одинаковой длины, получив, таким образом, версию матрицы  $A$  с «пробелами». Применив к матрице с «пробелами» один из методов итерационного моделирования неполных данных восстановим удалённый участок. Контролируемая точка должна находиться строго в центре отбрасываемого участка. Контролируемый участок должен находиться в центре окна, если это возможно. Допускается перестановка местами строк матрицы.

Необходимо также отметить, что хотя количество строк матрицы  $A$  и может быть произвольным, для удовлетворительной работы методов восстановления данных требуется не менее трёх строк. Длина интервала подбирается экспериментально для каждого метода ГИС. Эффективный размер интервала для КС составляет от 30 до 50 точек. Длина восстанавливаемого участка должна находиться в пределах 20–50 % от ширины матрицы.

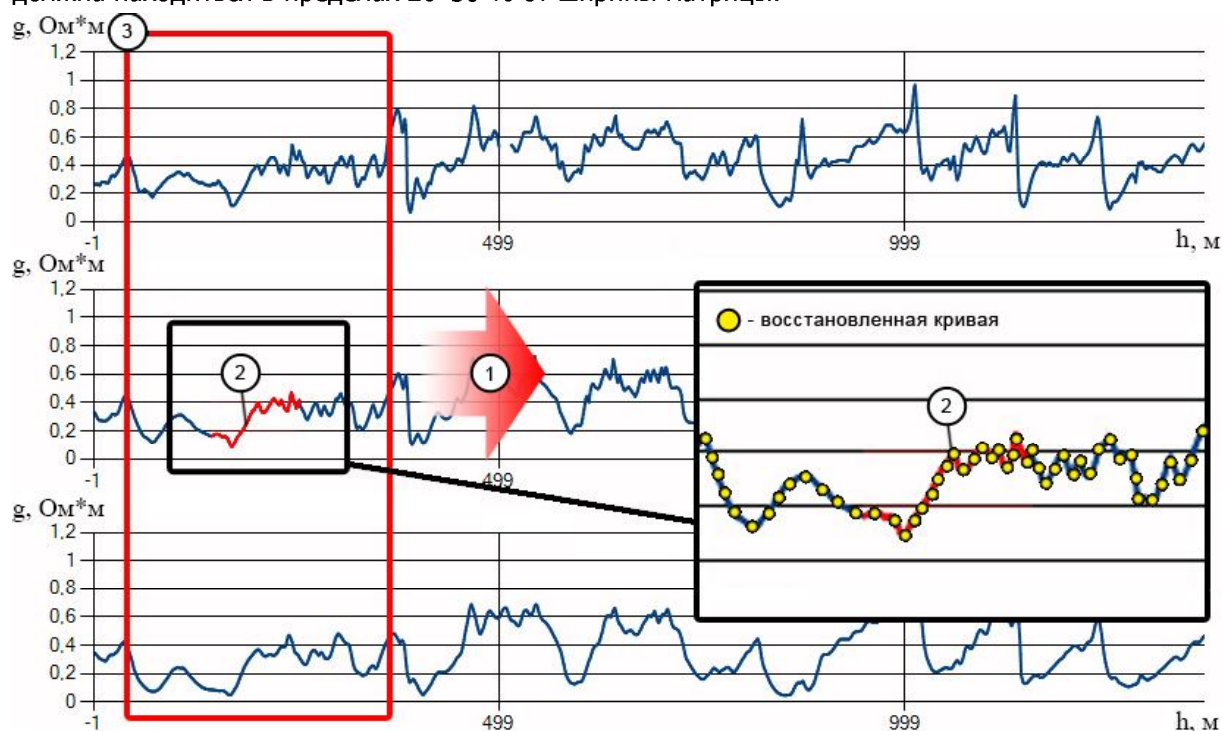


Рис. 1. Иллюстрация алгоритма оценки достоверности данных: 1 — движение скользящего окна; 2 — удалённый участок данных; 3 — скользящее окно

**Итерационное моделирование неполных данных с помощью многообразий малой размерности.** Рассмотрим три версии метода итерационного моделирования неполных данных с помощью многообразий малой размерности:

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.

Все версии метода могут трактоваться как построение нейросетевого конвейера, решающего следующие задачи:

1. заполнение пробелов в данных;
2. ремонт данных, корректировка значений исходных данных так, чтобы наилучшим образом работали построенные модели;
3. 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).

Столбец матрицы  $A$  есть вектор  $a$  с  $k$  пробелами, который представляется как  $k$ -мерное линейное многообразие  $L_a$ , параллельное  $k$  координатным осям, которые соответствуют удалённым данным. При наличии априорных ограничений на пропущенные значения место  $L_a$  занимает параллелепипед  $P_a \subset L_a$ .

Построим моделирующее эти данные линейное многообразие малой размерности следующим образом: за основу возьмём прямую  $f(x)=xy+b$ , которая задаётся направляющим вектором  $y$  и проходит через точку, определяемую вектором  $b$ . 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 в вектор  $x=Pr_{f(x)}(a)$  на полученной прямой (рис. 2).

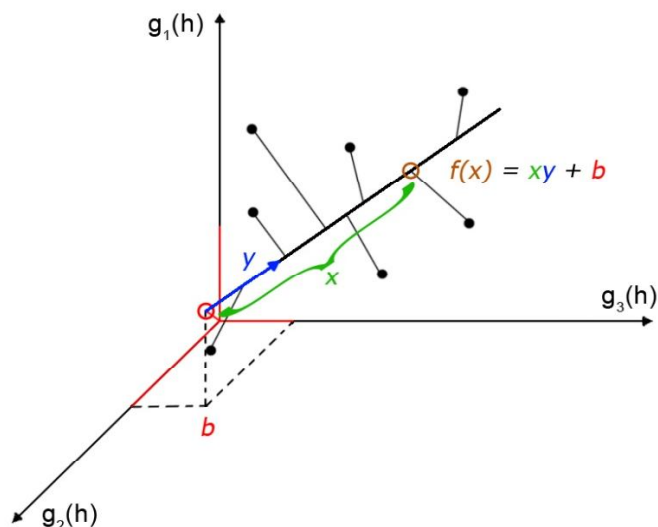


Рис. 2. Моделирующее многообразие малой размерности

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 (в определённом 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 определённом 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  $x_i y_j + b_j$ . Следовательно, ставится задача поиска наилучшего приближения  $A$  матрицей вида  $x_i y_j + b_j$  методом наименьших квадратов:

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

На первой итерации допускается принять значения вектора  $y$  случайным, но нормированным на 1, значения вектора  $b$  вычисляются по формуле:

$$b_j = \frac{1}{n_j} \sum_{\substack{i \\ a_{ij} \neq @}} a_{ij}, \text{ где } n_j = \sum_{\substack{i \\ a_{ij} \neq @}} 1. \quad (2)$$

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 ведётся по остаточной дисперсии столбцов. В результате исходная матрица данных  $A$  представляется в виде суммы матриц  $P_q$ , т. е.  $A=P_1+P_2+\dots+P_q$ . 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 искажённых 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.

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. Такой тип линий называется квазилинейным

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

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

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

$$\Phi = \sum_{\substack{i,j \\ a_{ij} \neq @}} (a_{ij} - f_j(a_{ik} y_k))^2 + \alpha \int_{-\infty}^{+\infty} (f''(t))^2 dt, \quad (3)$$

где  $\alpha > 0$  — параметр сглаживания.

Для решения 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 или формул Карлемана.

Таким образом, сглаженная вектор-функция  $f(t)$  экстраполируется с некоторого конечного множества  $\{t_k\}$  на всю вещественную прямую с использованием формул Карлемана:

$$f(t) \approx ty + b + \sum_{k=1}^m (f(t_k) - t_k y - b) \frac{2(e^{\lambda t} - e^{\lambda t_k})}{\lambda(e^{\lambda t} + e^{\lambda t_k})(t - t_k)} \prod_{\substack{j=1 \\ j \neq k}}^m \frac{(e^{\lambda t} + e^{\lambda t_k})(e^{\lambda t} + e^{\lambda t_k})}{(e^{\lambda t} + e^{\lambda t_k})(e^{\lambda t} + e^{\lambda t_k})}, \quad (4)$$

где  $\lambda$  — параметр, характеризующий ширину полосы на плоскости комплексных чисел, где экстраполируемая функция гарантированно голоморфна.

Процедура использования квазилинейных моделей несколько отличается от аналогичной процедуры в линейном случае. Точка на построенной кривой  $f(t)$ , соответствующая полному 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 на прямой ищется ближайшая точка  $t(a)$ , а затем — соответствующая точка на кривой  $f(t)$ , при  $t = t(a)$ . После построения кривой  $f(t)$  из данных вычитаются их проекции, т. е. матрица данных заменяется на матрицу уклонений. Далее снова ищется наилучшее линейное приближение для матрицы уклонений, вновь строится сглаживание, экстраполяция и т. д., пока уклонения не приблизятся в достаточной степени к нулю. Критерием остановки могут выступать остаточные дисперсии.

В результате исходная таблица предстаёт в виде  $Q$ -факторной модели:

$$a_{ij} \cong \sum_q f_j^q(t_i^q), \quad (5)$$

где аргументом функции служит нормированное скалярное произведение исходного данного на линейную основу квазилинейного многообразия.

Самоорганизующиеся карты Кохонена (Self-Organizing map — SOM) [11] — 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 — таксон. В результате такой замены данные представляются с ошибкой аппроксимации — среднеквадратичного расстояния от точки до ближайшего к ней образца.

Пусть SOM определяется набором точек (ядер)  $Y = \{y_{ij}\}$  ( $i, j = 1..m$ ), последовательно расположенных на квадратной сетке. Требуется отобразить на ней набор точек данных  $X = \{x_i\}$ . Введём преобразование  $\Pi$ , которое каждому вектору  $x \in X$  сопоставляет ближайшую к нему точку из  $Y$ :

$$x \xrightarrow{\Pi} y_{ij}, |y_{ij} - x|^2 \rightarrow \min, \quad (6)$$

а каждому ядру  $y_{ij}$  сопоставляется его таксон

$$K_{ij} = \{x \in X \mid x \xrightarrow{\Pi} y_{ij}\}. \quad (7)$$

Минимизируемый функционал, таким образом, будет состоять из следующих слагаемых:

$$D_1 = \sum_{ij} \sum_{x \in K_{ij}} \|x - y_{ij}\|^2; \quad (8)$$

$$D_2 = \sum_{ij} \|y_{ij} - y_{i+1,j}\|^2 + \sum_{ij} \|y_{ij} - y_{i,j+1}\|^2; \quad (9)$$

$$D_3 = \sum_{ij} \|2y_{ij} - y_{i-1,j} - y_{i+1,j}\|^2 + \sum_{ij} \|2y_{ij} - y_{i,j-1} - y_{i,j+1}\|^2. \quad (10)$$

Для построения SOM требуется минимизировать функционал:

$$D = \frac{D_1}{|X|} + \lambda \frac{D_2}{m^2} + \mu \frac{D_3}{m^2}, \quad (11)$$

где  $\lambda, \mu$  — параметры связности и нелинейности, «модули упругости».

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 нулём. 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 даёт следующие результаты:

$$\begin{aligned} \frac{1}{2} \frac{\partial D}{\partial y^{kl}} = & a_{kl}^{(-2)} y^{k-2,l} + a_{kl}^{(-1)} y^{k-1,l} + a_{kl} y^{kl} + a_{kl}^{(+1)} y^{k+1,l} + a_{kl}^{(+2)} y^{k+2,l} + \\ & + b_{kl}^{(-2)} y^{k,l-2} + b_{kl}^{(-1)} y^{k,l-1} + b_{kl}^{(+1)} y^{k,l+1} + b_{kl}^{(+2)} y^{k,l+2} - \sum_{x \in K_{kl}} x. \end{aligned} \quad (12)$$

Из уравнения  $\frac{\partial D}{\partial y^{kl}} = 0$ ,  $k=1 \dots p$ ,  $l=1 \dots q$  получим  $m$  систем линейных уравнений.

«Вытянем» набор ядер  $y_{kl}$  в один столбец. В результате вектор неизвестных примет вид:

$$x = (y_{11}, \dots, y_{1q}, y_{21}, \dots, y_{2q}, \dots, y_{(p-1)1}, \dots, y_{(p-1)q}, y_{p1}, \dots, y_{pq}). \quad (13)$$

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

$$b_s = \frac{\sum_{x \in K_{ij}} x}{|X|}, \quad i = \left[ \frac{s-1}{q} \right] + 1, \quad j = \left[ \frac{s-1}{q} \right] q, \quad (14)$$

где [...] — операция взятия целой части числа.

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

Матрица, таким образом, имеет девятидиагональный вид.

На протяжении всей работы алгоритма значения элементов матрицы остаются неизменными, изменяются лишь компоненты вектора  $b$ .

Численные эксперименты показывают высокую эффективность методов итерационного моделирования неполных данных с помощью многообразий малой размерности в задаче заполнения пробелов данных геофизических исследований (рис. 3). Для данных, качество которых не подвергается сомнению, протяжённость удалённого контролируемого участка, успешно восстанавливаемого методом самоорганизующихся карт, составляет до 50 % ширины окна и до 20 % при восстановлении линейным методом.

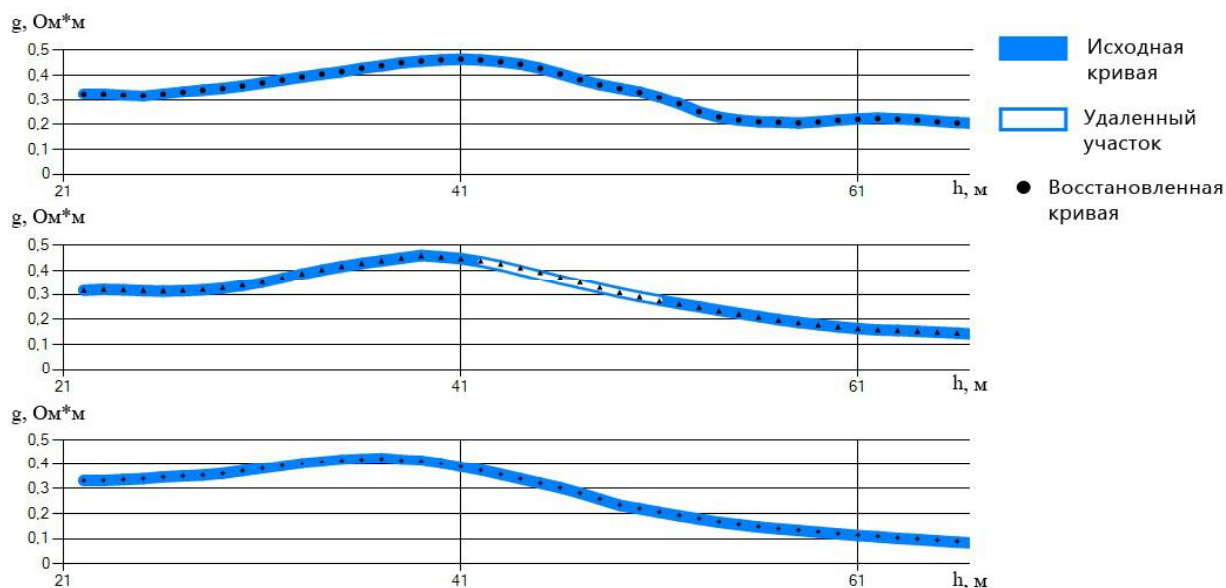


Рис. 3. Графическое представление содержимого LAS-файла с удалённым контролируемым участком

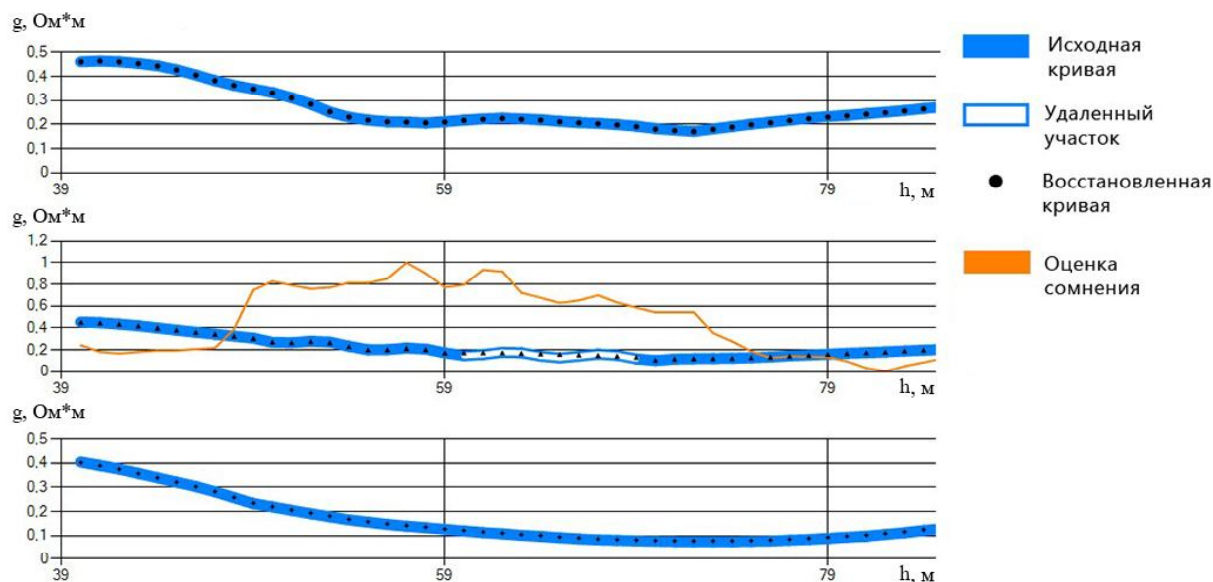


Рис. 4. Оценка сомнения для участка, искажённого гармоникой

Точки — данные, восстановленные методом SOM.

Далее из исходных данных отброшенного участка вычитаем данные полученные методом восстановления, получая, таким образом, дискретно заданную случайную величину — кривую



уклонений  $e$ . Величина среднеквадратического отклонения  $e$  и есть оценка достоверности для тестируемой точки, расположенной в центре отброшенного участка. Все операции повторяются для каждой точки, формируя кривую оценки достоверности.

Приведём результаты расчётов для рассматриваемой задачи контроля качества записи данных каротажа КС с использованием методов итерационного моделирования неполных данных самоорганизующимися многообразиями.

С целью проверки эффективности предложенного алгоритма на участке от 50 до 70 метров на испытываемой кривой были искусственно внесены помехи, представляющие собой гармонические колебания постоянной частоты и постоянной малой амплитуды. Участок с помехами явно выделялся аномально высоким уровнем среднеквадратического отклонения (рис. 4, 5).

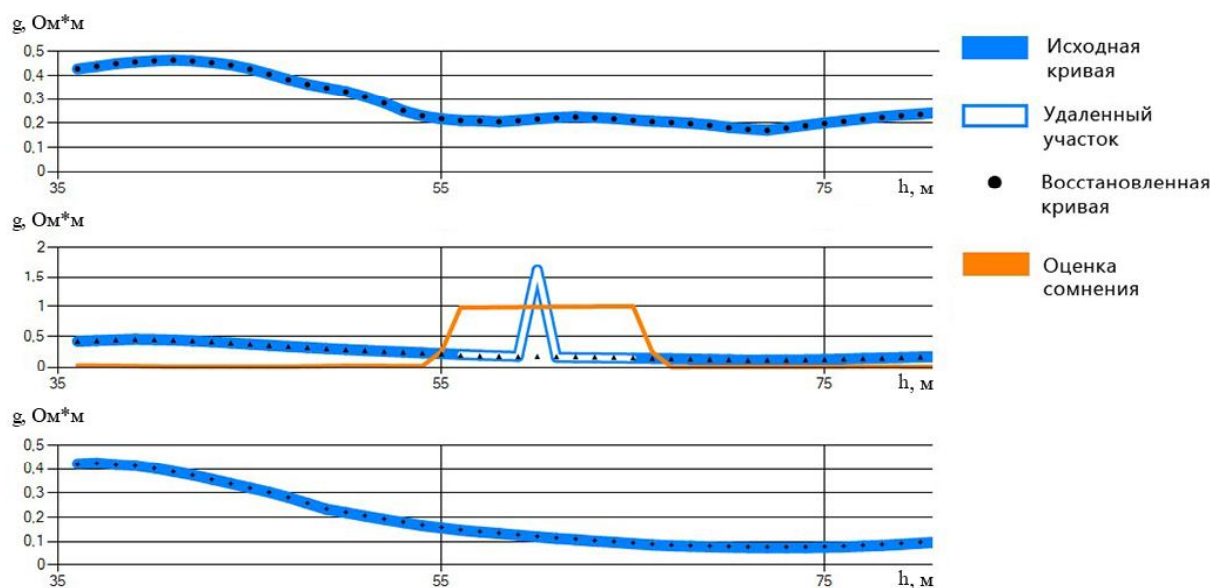


Рис. 5. Оценка сомнения для участка, искажённого выбросом

**Заключение.** Сформулирована задача автоматизированного контроля качества ГИС с использованием методов восстановления данных. Построена вычислительная схема, алгоритмы и программа вычислительной машины для решения задачи контроля качества ГИС. Доказана возможность применения методов восстановления данных в задачах оценки качества. Методика основана на использовании методов теории подобия и восстановления данных с пропусками. Автоматизация первичного контроля каротажных материалов позволит избежать дополнительных выездов на скважины с целью повторного проведения ГИС.

### Библиографический список

1. ГОСТ Р 54362-2011. Геофизические исследования скважин. Термины и определения. — Режим доступа : [http://www.sanse.ru/text/GOST\\_2008.pdf](http://www.sanse.ru/text/GOST_2008.pdf) (дата обращения : 09.02.2014).
2. Нейдорф, Р. А. Осесимметричная задача о распространении фронта фазового превращения в гетерогенной твёрдой пористой среде в условиях граничного теплообмена и наличия центрального теплозащитного кольцевого слоя / Р. А. Нейдорф [и др.] // Вестн. Дон. гос. техн. ун-та. — 2011. — Т. 11, № 7 (58). — С. 979–986.
3. Abate, M. L. Hierarchical Approach to Improving Data Quality / Marcey L. Abate, Kathleen V. Diegert, Heather W. Allen // Data Quality Journal. — 1998. — Vol. 4, no. 1. — Pp. 365–369.

4. Нейдорф, Р. А. Аппроксимационное построение математических моделей по точечным экспериментальным данным методом «cut-glue» / Р. А. Нейдорф // Вестн. Дон. гос. техн. ун-та. — 2014. — № 1. — С. 45–58.
5. Сергиенко, А. Б. Цифровая обработка сигналов / А. Б. Сергиенко. — Санкт-Петербург : Питер, 2003. — 584 с.
6. Россиев, А. А. Итерационное моделирование неполных данных с помощью многообразий малой размерности / А. А. Россиев. — Красноярск, 2000. — 83 с.
7. Веников, В. А. Теория подобия и моделирование / В. А. Веников. — Москва : Высш. шк., 1976. — 479 с.
8. Седов, Л. И. Методы подобия и размерности в механике / Л. И. Седов. — Москва : Наука, 1987. — 432 с.
9. Герасимато, Ф. Г. Теория подобия / Ф. Г. Герасимато. — Режим доступа : <http://www.trinitas.ru/rus/doc/0016/001d/00162064.htm> (дата обращения : 09.02.2014).
10. Айзенберг, Л. А. Формулы Карлемана в комплексном анализе. Первые приложения / Л. А. Айзенберг. — Новосибирск : Наука, 1990. — 248 с.
11. Зиновьев, А. Ю. Визуализация многомерных данных / А. Ю. Зиновьев. — Изд-во КГТУ, 2000. — 180 с.

Материал поступил в редакцию 01.04.2014.

## References

1. GOST R 54362-2011. Geofizicheskiye issledovaniya skvazhin. Terminy i opredeleniya. [GOST R 54362-2011. Well logging. Terms and definitions.] Available at: [http://www.sanse.ru/text/GOST\\_2008.pdf](http://www.sanse.ru/text/GOST_2008.pdf) (accessed: 09.02.2014) (in Russian).
2. Neydorf, R. A., et al. Osesimmetrichnaya zadacha o rasprostraneni fronta fazovogo prevrashcheniya v geterogennoy tverдой poristoy srede v usloviyakh granichnogo teploobmena i nalichiya tsentralnogo teplozashchitnogo koltsevoogo sloya. [Axisymmetric problem on propagation of phase transformation front in heterogeneous solid porous medium in terms of boundary heat transfer and presence of central heatproof ring layer.] Vestnik of DSTU, 2011, vol., 11, no. 7 (58), pp. 979–986 (in Russian).
3. Abate, Marcey L., Diegert, Kathleen V., and Allen, Heather W. Hierarchical Approach to Improving Data Quality. Data Quality Journal, 1998, vol. 4, no. 1, pp. 365–369.
4. Neydorf, R. A. Approksimatsionnoye postroyeniye matematicheskikh modeley po tochechnym eksperimentalnym dannym metodom «cut-glue». [Approximating mathematical model development according to point experimental data through “cut-glue” method.] Vestnik of DSTU, 2014, no. 1, pp. 45–58 (in Russian).
5. Sergiyenko, A. B. Tsifrovaya obrabotka signalov. [Digital signal processing.] Sankt-Peterburg : Piter, 2003, 584 p. (in Russian).
6. Rossiyeв, A. A. Iteratsionnoye modelirovaniye nepolnykh dannyykh s pomoshchyu mnogo-obraziy maloy razmernosti. [Iterative modeling of incomplete data using low-dimensional manifolds.] Krasnoyarsk, 2000, 83 p. (in Russian).
7. Venikov, V. A. Teoriya podobiya i modelirovaniye. [Similarity theory and modeling.] Moscow : Vysshaya shkola, 1976, 479 p. (in Russian).
8. Sedov, L. I. Metody podobiya i razmernosti v mekhanike. [Similarity and dimensional methods in mechanics.] Moscow : Nauka, 1987, 432 p. (in Russian).
9. Gerasimato, F. G. Teoriya podobiya. [Similarity theory.] Available at: <http://www.trinitas.ru/rus/doc/0016/001d/00162064.htm> (accessed: 09.02.2014) (in Russian).

10. Aizenberg, L. A. Formuly Karlemana v kompleksnom analize. Pervyye prilozheniya. [Carleman formula in complex analysis. First applications.] Novosibirsk: Nauka, 1990, 248 p. (in Russian).
11. Zinovyev, A. Y. Vizualizatsiya mnogomernykh dannykh. [Multivariate data visualization.] Izdatelstvo KGTU, 2000, 180 p. (in Russian).

## **ON AUTOMATED QUALITY CONTROL OF GEOPHYSICAL WELL LOGGING DATA\***

**R. V. Pelmegov, A. G. Kudelin**

*The development of scientific principles, methods, and techniques of the automated control of the primary data recorded in the geophysical well logging is considered. The issue of termination of logging on a particular well is often associated with subjective expert assessments. The need for automated detection of errors and quality control of the recorded data during or immediately upon completion of studies is discussed. A brief analysis of the unreliability symptoms of record fragments is provided. An algorithm using iterative modeling techniques of incomplete data with low-dimensional manifolds for solving the problem of assessing the quality of the well logging data is proposed. Examples demonstrating the high efficiency of the iterative modeling methods for incomplete data using low-dimensional manifolds in the problem of space filling of geophysical study data are given; and the results of the numerical experiments on solving the problem of data quality control of electric logging are provided.*

**Keywords:** geophysical research, quality control, iterative modeling, automation.

---

\* The research is done on theme no. 2.3.13 "Balanced sampling method for the problems of dynamic process simulation in distributed objects" on the thematic research plan of the Ministry of Education and Science of the Russian Federation.