

УДК 004.89

В. П. ЖИТНИКОВ, Н. А. ЗАЙКО

КОМПЛЕКСНЫЙ ПОДХОД К ОЦЕНКЕ ПОГРЕШНОСТЕЙ В ЗАДАЧЕ ЧИСЛЕННОГО АНАЛИЗА ДАНЫХ НАТУРНОГО ЭКСПЕРИМЕНТА

Для анализа вероятностных характеристик результатов измерений параметров физических процессов построена математическая модель процесса обработки данных, учитывающая дискретность измерений и вычислений, имеющих место при использовании цифровой техники. Погрешность результата обработки данных состоит из четырех составляющих: внешние помехи, погрешность измерительного прибора, погрешность численного метода обработки и накопленная погрешность округления, вызванная ограниченной разрядностью процессора. Получены характеристики случайных процессов и оценки их погрешностей. *Случайный процесс; вероятностные характеристики; математическая модель; численная фильтрация*

ВВЕДЕНИЕ

Информационные процессы, протекающие в комплексе измерений и обработки их результатов, заключаются в получении информации в виде совокупности измеряемых величин, воспринимаемых с помощью датчиков, которые поступают в измерительный канал в виде некоторых сигналов, преобразуемых здесь в цифровые отсчеты. Далее эти отсчеты поступают в систему обработки, где с помощью алгоритмов обработки информации, а также правил их выбора рассчитываются характеристики процесса с оценкой их погрешностей, что в конечном итоге и предоставляется потребителю информации.

В последнее время требования к системе обработки ужесточаются. Например, при наклонном бурении скважин требуемая точность данных о траектории движения бура после обработки сигнала может достигать до 3–4-х значащих цифр. Реальный датчик, предназначенный для работы, дает около 2-х цифр. Помеха, которая может иметь и случайную и детерминированную составляющие, может быть на порядок больше. Отсюда возникает задача: построить такую систему, которая позволила бы уменьшить погрешность в 100–1000 раз.

Реальный процесс измерения может быть подвержен различным внешним помехам. В случае бурения это может быть влияние буровой установки, при измерении параметров двигателя – вращение турбины и т. п. В любом случае исследуемый сигнал может содержать

как случайную, так и детерминированную составляющие. В качестве примера можно рассмотреть процесс вида

$$x(t) = b + (a + \alpha)\sin(\omega t),$$

где a , b , ω – константы, α – случайная величина, t – время, $(a + \alpha)\sin(\omega t)$ – помеха.

Теоретически уменьшить погрешность, вызванную помехами и неточностью измерений можно статистическими методами. Но уменьшение погрешности при применении статистических методов происходит примерно как $1/\sqrt{N}$ (N – число измерений). Тем самым, для уменьшения погрешности на 2–3 порядка нужно использовать 104–106 слагаемых. В этих условиях (высокой точности и большого числа слагаемых) проявляются погрешности, которые ранее были несущественными: погрешности метода численной обработки и накопленная погрешность округления.

Таким образом, новыми объектами исследования являются эти два вида погрешностей при применении статистических методов анализа.

Поскольку при исследовании результатов натурального эксперимента постулирование каких-либо свойств измеряемого приборами сигнала, несущего информацию об исследуемом процессе (стационарность, эргодичность и т. п.) может привести к непредсказуемым искажениям оценок, предлагается подход, заключающийся в использовании априорной информации в виде определенных математических моделей, базирующихся на опыте исследований в других областях, с обязательной последующей проверкой, уточнением и построением апостериорных мо-

делей. Для этого используются специальные средства, позволяющие провести анализ составляющих сигнала, имеющих различный характер (случайный или детерминированный в виде функциональных зависимостей различного вида) и представить результаты пользователю в наглядном виде.

В соответствии с этим считается известной последовательность результатов измерений $\langle x(t_n) \rangle$, $n = 1, \dots, N$ некоторого параметра $x(t)$ и характеристика системы измерений (СИ) в виде плотности вероятности $\omega(\Delta)$ погрешности измерения $\Delta = \langle x(t) \rangle - x(t)$. Эта характеристика должна быть получена при тестировании СИ на стенде с помощью более точных приборов.

Вначале по последовательности $\langle x(t_n) \rangle$ и зависимости $\omega(\Delta)$ строятся оценки одномерной

$$\langle \omega(X) \rangle = \langle \omega(x(t) = X) \rangle$$

и двумерной

$$\langle \omega(X_1, X_2, \tau) \rangle = \langle \omega(x(t) = X_1, x(t + \tau) = X_2) \rangle$$

плотностей вероятности, и с их помощью вычисляются требуемые характеристики сигнала.

Затем по предложенной ниже методике находят оценки погрешностей, проводится анализ адекватности априорных и построение апостериорных моделей. Альтернативой является установление факта о невозможности применения разработанной методики к данной последовательности x_j .

Ранее были изучены процессы непрерывного измерения [1], которые характерны для аналоговых устройств. В связи с широким использованием цифровой техники и компьютеров для анализа результатов измерений, возникает необходимость исследования влияния дискретизации на результат измерений.

В работе [2] была предложена дискретная модель для исследования характеристик на основе оценки одномерной плотности вероятности. Оценка двумерной плотности вероятности позволяет находить и исследовать статистические связи между двумя случайными величинами. Они широко используются в различных технических приложениях.

Для анализа результатов решения задач вычислительной математики (численного дифференцирования, интегрирования, решения задач для дифференциальных уравнений) были разработаны методы фильтрации, основанные на априорном представлении математической модели погрешности численного метода в виде суммы нескольких степенных функций параметра дискретизации (числа узлов или шага сетки) и неизвестной нерегулярной части погрешности [3–

6]. Представляет интерес экспериментальная проверка применимости этих методов к задачам статистической обработки данных, имеющих наряду с детерминированной и случайную составляющую, существенно меняющую математические свойства функций.

1. ПОСТАНОВКА ЗАДАЧИ

Концептуальная постановка задачи моделирования выглядит следующим образом.

Дано:

- дискретная последовательность данных, представленных в цифровом формате $\langle x(t_n) \rangle$;
- точностная характеристика СИ в виде плотности вероятности погрешности $\omega[\Delta]$, моментов: математического ожидания m_Δ , дисперсии D_Δ , и др.

Определить:

- плотность распределения вероятности входного сигнала;
- с его помощью найти вероятностные характеристики сигнала $x(t)$;
- оценки погрешностей вычисленных значений вероятностных характеристик.

Поскольку закон распределения можно определить только приближенно, то необходимо найти оценки погрешностей характеристик, вызванные различными источниками (внешними помехами, измерительным прибором, методом вычисления характеристик и ограниченностью разрядности представления чисел в ЭВМ).

Расширить возможности методов обработки экспериментальной информации можно за счет комплексного учета основных факторов, влияющих на точность статистических измерений.

Идея *комплексного подхода* заключается в использовании априорной математической модели общей погрешности в виде:

$$\Delta_{\text{общ}} = \Delta_{\text{возм}} + \Delta_{\text{СИ}} + \Delta_{\text{метод}} + \Delta_{\text{окр}}. \quad (1)$$

Представление погрешности в виде суммы подразумевает линеаризацию некоторой функциональной зависимости более общего вида, что обосновано при достаточно малых относительных значениях погрешностей. Это имеет место для измерений высокой точности. Кроме того, при таком представлении не предполагается отсутствие зависимости составляющих погрешности, например от амплитуды измеряемой величины. Конкретизация этих зависимостей представляет одну из задач методики исследования результатов измерений.

Внешние воздействия возмущающих факторов $\Delta_{\text{возм}}$ описываются случайной моделью с

априори неизвестной плотностью распределения вероятности $\omega_1[X]$.

Погрешность СИ $\Delta_{СИ}$ считается случайной величиной с известной плотностью распределения вероятности $\omega[\Delta]$.

Погрешность метода $\Delta_{метод}(N, B) = y_{N, B} - y$ (где $y_{N, B}$ и y – вычисленное и точное значения) обусловлена дискретизацией входного сигнала, ограниченностью диапазона измерения и ее модель представляется определенными функциональными зависимостями от количества слагаемых суммы (4) и величины диапазона измерения B

$$y_{N, B} - y_B = c_1(B)N^{-k_1} + c_2(B)N^{-k_2} + \dots, \quad (2)$$

$$y_B - y = C_1B^{-l_1} + C_2B^{-l_2} + \dots, \quad (3)$$

где y – точное значение; $y_{N, B}$ – приближенный результат, полученный при числе слагаемых суммы, равном N , и длине диапазона измерения, равном B ; y_B – предельное при $N \rightarrow \infty$ значение $y_{N, B}$ при данном B ; $k_1, k_2, \dots, l_1, l_2, \dots$ – действительные числа ($0 < k_1 < k_2 < \dots, 0 < l_1 < l_2 < \dots$), которые могут быть априори известны или неизвестны в зависимости от конкретной ситуации. То есть, модель этих составляющих погрешности является детерминированной. Представление (2), (3) является априорным и изначально выбирается на основе результатов исследования решения задач вычислительной математики [3–7].

Погрешность округления $\Delta_{окр}$ считается неизвестной величиной, диапазон изменения которой определяется экспериментально и может зависеть как от метода вычисления, так и от длины мантииссы машинного слова и способа округления. Эта погрешность и ее оценочная зависимость от N может существенно отличаться при применении различных численных методов.

До обработки измерительной информации погрешность определяется первыми двумя составляющими (1). В результате статистической обработки первая и вторая составляющая уменьшаются, но появляются третья и четвертая. Дальнейшая обработка уменьшает третью составляющую.

2. МЕТОДИКА ОБРАБОТКИ ИЗМЕРИТЕЛЬНОЙ ИНФОРМАЦИИ

Комплексный подход предполагает использование определенной методики обработки данных с целью выделения полезной информации путем уменьшения или подавления составляющих погрешности. Для этого в определенной последовательности используется набор известных, адаптированных к данной задаче алгорит-

мов оценки и уменьшения разных видов погрешности.

1. Поскольку закон распределения случайной величины неизвестен, то определяется приближенная зависимость плотности вероятности на основе дискретных эмпирических данных $\langle x_n \rangle = \langle x(t_n) \rangle$:

- оценка одномерной плотности вероятности [2]

$$\langle \omega_1(X) \rangle = \frac{1}{N} \sum_{n=1}^N \omega[\langle x_n \rangle - X],$$

- оценка двумерной плотности вероятности

$$\begin{aligned} \langle \omega_2(X_1, X_2, k) \rangle &= \\ &= \frac{1}{N-k} \sum_{n=1}^{N-k} \omega[\langle x_n \rangle - X_1] \omega[\langle x_{n+k} \rangle - X_2] \end{aligned}$$

2. Алгоритмы вычисления оценок моментов случайного сигнала сводятся на основе полученных оценок плотностей вероятности к вычислению средних арифметических результатов измерений или их комбинаций.

Например, оценка корреляционного момента определяется по формуле

$$\begin{aligned} \langle R(k) \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [X_1 - \langle m_X \rangle] [X_2 - \langle m_X \rangle] \times \\ &\times \langle \omega_2(X_1, X_2, k) \rangle dX_1 dX_2 = \\ &= \frac{1}{N-k} \sum_{n=1}^{N-k} \left[\langle x_n \rangle - \frac{1}{N} \sum_{n=1}^N \langle x_n \rangle \right] \times \\ &\times \left[\langle x_{n+k} \rangle - \frac{1}{N} \sum_{n=1}^N \langle x_n \rangle \right], \end{aligned} \quad (4)$$

где

$$\begin{aligned} \langle m_X \rangle &= \int_{-\infty}^{\infty} X \langle \omega_1(X) \rangle dX = \frac{1}{N} \sum_{n=1}^N \int_{-\infty}^{\infty} X \omega[\langle x_n \rangle - X] dX = \\ &= \frac{1}{N} \sum_{n=1}^N \langle x_n \rangle - m_{\Delta} \end{aligned}$$

– оценка математического ожидания параметра X .

3. Оценка и уменьшение погрешностей дискретизации и ограниченности диапазона измерения проводится с помощью фильтрации результатов расчета характеристик сигнала. Для этого разработаны алгоритмы, приведенные ниже и использующие методы экстраполяции Ричардсона, Ромберга, Нэвилла, встроенные в алгоритмы фильтрации, а в случае неизвестных показателей в (2), (3) используются δ^2 -, ϵ -, θ -алгоритмы [3]. При этом сначала проводится фильтрация по N в соответствии с (2), а затем отфильтрованные по N значения фильтруются по B согласно (3).

4. Применение методов фильтрации позволяет подавить составляющие погрешностей (2),

(3), что дает возможность обнаружить скрытую за погрешностями метода накопившуюся погрешность округления.

5. На основании изучения результатов расчетов строится апостериорная модель сигнала, которая может быть использована для оценки погрешностей при изменении характеристик сигнала.

6. Анализ результатов расчетов на этом этапе позволяет подобрать оптимальные параметры численных методов (длины кванта дискретизации и диапазона измерения, тип переменных). Далее эти параметры могут оставаться неизменными, или периодически корректироваться, если характер сигнала претерпевает значительные изменения.

7. Проводится комплекс вычислений, оценок погрешностей и проверка достоверности этих оценок с помощью адаптированного программно-алгоритмического обеспечения.

3. АЛГОРИТМЫ ФИЛЬТРАЦИИ

Рассмотрим алгоритмы фильтрации математических моделей (2) и (3).

В качестве фильтруемых последовательностей $z_i^{(0)}$ могут использоваться как вычисленные результаты вероятностных характеристик $y_{N,B}$, так и полученные в результате первого этапа фильтрации значения y_B .

Суть алгоритмов заключается в том, что на начальном этапе делается попытка определить коэффициенты модели k_1, k_2, \dots . Обычно это возможно, когда они являются целыми числами.

В этом случае для фильтрации в алгоритме 1 применяется формула Ричардсона. В случае невозможности определения коэффициентов модели в алгоритме 1 используется δ^2 -алгоритм или ε -алгоритм [3], не требующие знания порядков точности (рис. 1).

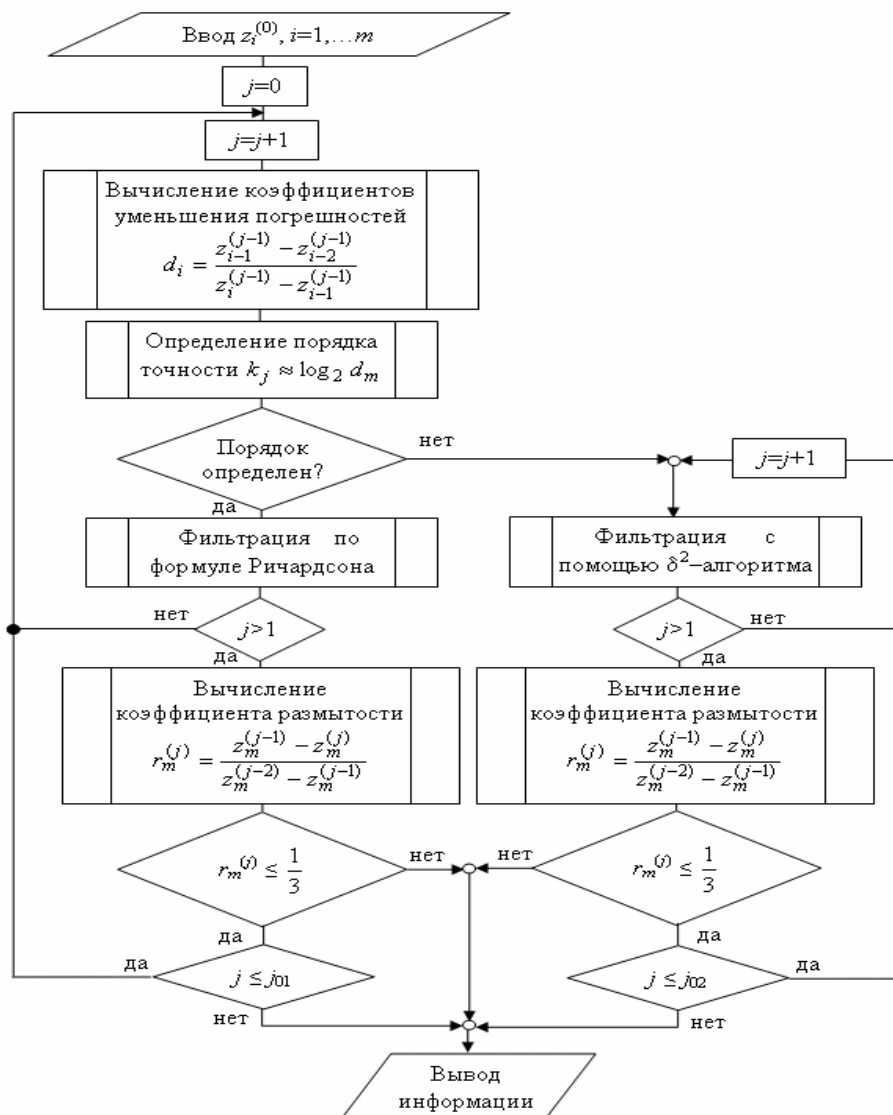


Рис. 1. Алгоритм 1 идентификации математических моделей

Далее проводится повторная фильтрация, на j -м шаге которой по вычисленной ранее последовательности $z_i^{(j-1)}$ применением одного из перечисленных в п. 2 методики алгоритмов экстраполяции определяется отфильтрованная j раз последовательность $z_i^{(j)}$. Количество шагов фильтрации либо ограничено заранее числом j_{01} , либо процесс прерывается, если не выполняется определенное условие $r_m^{(j)} \leq \frac{1}{3}$, где

$$r_m^{(j)} = \frac{z_m^{(j-1)} - z_m^{(j)}}{z_m^{(j-2)} - z_m^{(j-1)}}, \quad z_m^{(j)} - \text{значения, полученные на } j\text{-м шаге фильтрации.}$$

Смысл этого ограничения заключается в том, что значительного уменьшения оценки погрешности после очередной фильтрации не происходит. Критическое значение $1/3$ эмпирически определено в работе [8].

Алгоритм 1 обычно используется для оценки погрешности дискретизации по числу отсчетов N . Он требует экспоненциального роста числа отсчетов, например, $N_i = 2N_{i-1}$. Алгоритм 2 более удобен для оценки погрешности, вызванной ограниченностью диапазона измерения V , поскольку он требует линейного роста числа отсчетов.

В алгоритме 2 вместо формулы Ричардсона используется формула Нэвилла, а вместо δ^2 -алгоритма – θ -алгоритм [3].

В отличие от первого, в алгоритме 2 формула Нэвилла может быть использована, только если показатели являются частью натурального ряда чисел. В остальном алгоритм 2 аналогичен алгоритму 1.

Результаты применения этих алгоритмов приведены на рис. 2–5.

4. РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ МЕТОДИКИ АНАЛИЗА ПОГРЕШНОСТЕЙ

Рассмотрим зависимости относительной погрешности коэффициента корреляции $r(k)$ от количества отсчетов N (рис. 2, 3). Данные зависимости удобнее всего проиллюстрировать в логарифмическом масштабе, поэтому по оси абсцисс будет откладываться десятичный логарифм от количества отсчетов N , а по оси ординат – десятичный логарифм относительной погрешности δ вычисления величины коэффициента корреляции $r(k)$. Величину $-\lg \delta$ назовем точностью, выраженной в количестве точных десятичных значащих цифр.

С помощью алгоритма 1 проведена идентификация математической модели (2) по результатам численных экспериментов, что позволило

путем фильтрации получить оценки погрешности и уточнить эти результаты. На рис. 2, 3 цифрой 0 обозначена зависимость погрешности исходных результатов вычисления коэффициента корреляции, цифрами 1–3 показаны результаты однократной, двукратной и т. д. фильтрации численных данных. Буквой a обозначены прямые $y = 2 + \frac{1}{2} \lg N$ и $y = 12 + \frac{1}{2} \lg N$, оценивающие уровень случайной составляющей погрешности результата, буквой b – прямые $y = 6,5 - \frac{1}{2} \lg N$ и $y = 16,5 - \frac{1}{2} \lg N$, оценивающие уровень накопленной погрешности округления.

Положительным результатом каждого шага фильтрации является увеличение ординаты и изменение (увеличение) углового коэффициента линий с возрастанием их номера. При попытке фильтрации несуществующей составляющей (при ошибке в указании показателя степенной функции в (2)) уточнения не происходит, возможно даже ухудшение точности.

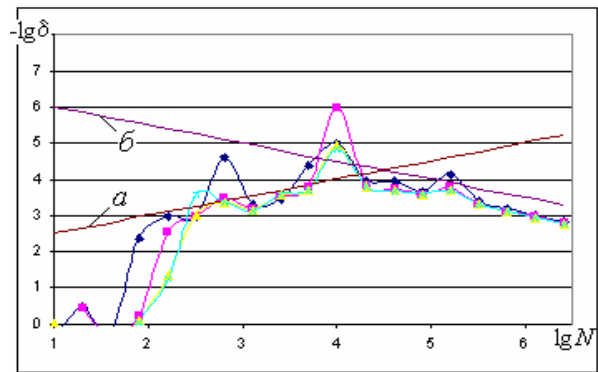


Рис. 2. Зависимость погрешности вычисления коэффициента корреляции $r(k)$ от количества отсчетов N . Переменные типа single. Амплитуда случайной составляющей составляет 10^{-2}

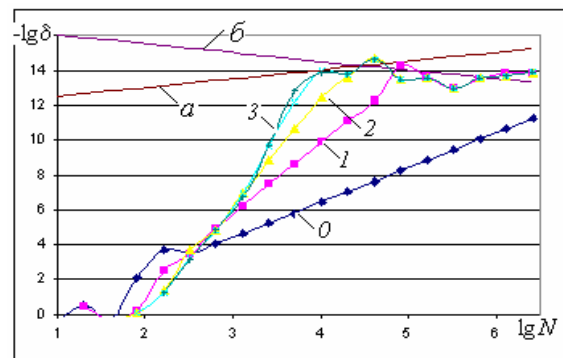


Рис. 3. Зависимость погрешности вычисления коэффициента корреляции $r(k)$ от количества отсчетов N . Переменные типа double. Амплитуда случайной составляющей составляет 10^{-12}

При небольшой ошибке в показателе происходит неполное подавление составляющей, что может обнаружиться после нескольких после-

дующих фильтраций при появлении линии с угловым коэффициентом, соответствующим этой составляющей, сдвинутой вверх по графику.

Таким образом, сравнение несколько раз отфильтрованных результатов позволяет проверить непротиворечивость априорных математических моделей.

Фильтрация позволяет достаточно надежно оценить погрешность численного метода, увеличить точность результата. Путем идентификации и подавления составляющих погрешности численного метода удается визуально оценить зависимость погрешности округления от N . Как видно из рис. 2, 3, при суммировании погрешность округления накапливается по статистическому закону. Из-за накопления погрешности округления наблюдается уменьшение точности при $\lg N > 4$. Рис. 3 показывает, что фильтрация позволяет уменьшить погрешность метода на несколько порядков, но только при $\lg N > 3$.

Результаты показывают, что накопление погрешности округления происходит по статистическому закону. Оценка относительной погрешности округления выглядит следующим образом: $\delta_{окр} \approx 10^{-M} \sqrt{N}$, где M – число разрядов мантиссы.

Для изучения характера изменения коэффициента корреляции при расширении диапазона были проведены расчеты с помощью алгоритмов 1 и 2, результаты которых представлены на рис. 4 и 5.

По оси абсцисс откладывается $\lg(B / 1000)$, где B – величина диапазона измерения, а по оси ординат – $\lg \delta$, где δ – относительное отличие текущего вычисленного значения от эталонного значения. Каждое использованное значение исследуемой зависимости для каждого B получалось в результате фильтрации по N . Как видно из рис. 4 и 5, зависимость коэффициента корреляции от величины диапазона B приближенно можно представить функцией $r_{\infty} - \text{const}/B$. Фильтрация позволяет уточнить предельные значения коэффициента корреляции на несколько значащих цифр. При этом сами расчетные значения могут оказаться недостаточно точными. Наличие предельного значения, закономерностей изменения погрешностей при увеличении B или их отсутствие представляет собой важнейшую информацию не только для оценки точности результатов обработки, но и для принятия решения об адекватности модели случайного процесса и реального сигнала. Как видно из сравнения двух рисунков, алгоритм 2 позволяет добиться большей точности при меньшем обрабатываемом диапазоне измерения.

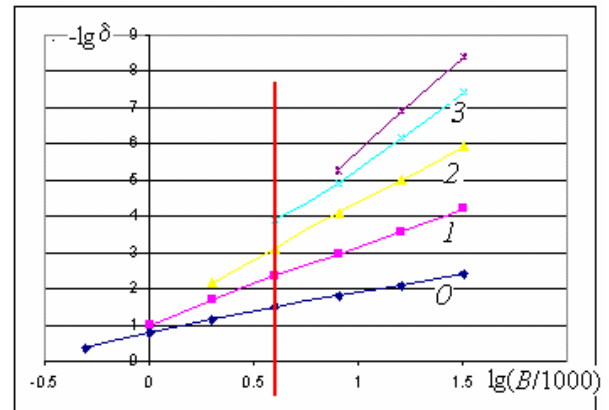


Рис. 4. Результаты фильтрации зависимости значений коэффициента корреляции от величины диапазона измерения с помощью алгоритма 1

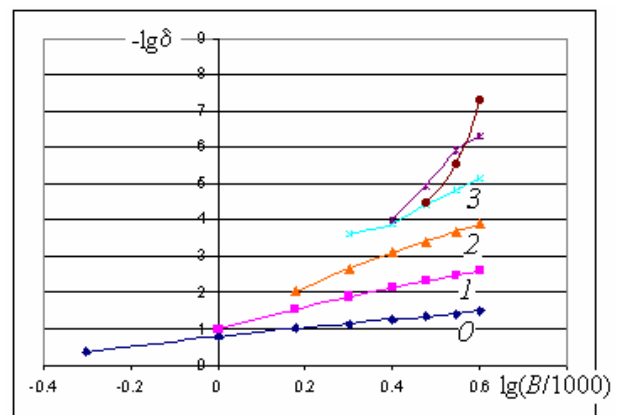


Рис. 5. Результаты фильтрации зависимости значений коэффициента корреляции от величины диапазона измерения с помощью алгоритма 2

ЗАКЛЮЧЕНИЕ

Разработанная дискретная модель процесса обработки результатов измерения характеристик случайного сигнала, в отличие от известных, позволяет учесть влияние на результаты обработки всех видов источников погрешности в комплексе.

На основе этой модели разработана методика комплексного подхода к оценке погрешностей обработки результатов измерений, которая не требует знания закона распределения измеряемой составляющей и сочетает статистические и детерминированные способы уменьшения погрешностей. Это предоставляет возможность оценить все их составляющие. Исследование показало, что этот подход позволяет уточнить результаты и увеличить надежность полученных оценок, проверить адекватность моделей, применяемых для описания случайных процессов.

Разработанные алгоритмы позволяют пользователю в наглядном виде получать информацию об исследуемых параметрах и оценки их

погрешностей, что дает возможность оценки всех составляющих погрешностей, имеющих разный порядок. Это обеспечивает, в отличие от известных средств, принятие обоснованных решений о достоверности полученных результатов обработки данных натурального эксперимента.

Анализ результатов внедрения в нефтегазовую промышленность позволяет утверждать, что применение разработанной модели процесса идентификации характеристик случайных процессов и методов обработки результатов измерений на основе этой модели позволяет при обработке сигналов из забоя скважины в процессе наклонно-направленного бурения на глубине 2–5 км попасть в нефтеносный пласт толщиной 1,5–3 м и пробурить там канал сбора нефти длиной 700–1500 м.

СПИСОК ЛИТЕРАТУРЫ

1. **Заико, А. И.** Определения и алгоритмы измерения характеристик эргодических процессов // Метрология. 2003. № 4. С. 3–15.
2. **Заико Н. А.** Дискретная модель измерения эргодических случайных процессов Вестник УГАТУ (сер. «Управление, вычислительная техника и информатика»). 2007. Т. 9, № 5(23). С. 101–107.
3. **Шерыхалина Н. М.** Методы обработки результатов численного эксперимента для увеличения их точности и надежности // Вестник УГАТУ (сер. «Управление, вычислительная техника и информатика»). 2007. Т. 9, № 2(20). С. 127–137.
4. **Шерыхалина Н. М.** Применение фильтрации для обработки результатов численного эксперимента // Вестник УГАТУ (сер. «Управление, вычислительная техника и информатика»). 2007. Т. 9, № 7(25). С. 90–96.
5. **Житников В. П., Шерыхалина Н. М.** Обоснование методов фильтрации результатов численного эксперимента // Вестник УГАТУ (сер. «Фунда-

ментальная и прикладная математика»). 2007. Т. 9, № 3(21). С. 71–79.

6. **Шерыхалина Н. М.** Применение фильтрации численных результатов для увеличения надежности САПР // Информационные технологии. 2008. № 9. С. 16–22.

7. **Шерыхалина Н. М., Поречный С. С.** Моделирование погрешности и численная фильтрация при решении смешанных задач для дифференциальных уравнений // Вестник УГАТУ (сер. «Управление, вычислительная техника и информатика»). 2008. Т. 11, № 1(28). С. 181–188.

8. **Житников В. П., Шерыхалина Н. М., Ураков А. Р.** Линейные некорректные задачи. Верификация численных результатов. Уфа: УГАТУ, 2008. 100 с.

ОБ АВТОРАХ



Житников Владимир Павлович, проф., зав. каф. компьют. матем. Дипл. инж.-физ. (МФТИ, 1973). Д-р физ.-мат. наук по механ. жидкости, газа и плазмы (Казанск. гос. ун-т, 1993). Иссл. в обл. волновых течений жидкости, электрохим. формообразования, числ.-аналит. методов.



Заико Наталья Александровна, ст. преп. той же каф. Дипл. инж.-сист. (УГАТУ, 2004). Канд. техн. наук (УГАТУ, 2008). Иссл. в обл. повыш. точности и уменьш. длительности измерений интеллект. систем обработки информации.