ЕЖЕГОДНИК-2010, Тр. ИГГ УрО РАН, вып. 158, 2011, с. 205–208

МЕТОДОЛОГИЯ И МЕТОДЫ ИССЛЕДОВАНИЙ

ПОМЕХОУСТОЙЧИВЫЙ И ДОСТОВЕРНЫЙ МЕТОД ОБРАБОТКИ МНОГОКРАТНЫХ НАБЛЮДЕНИЙ

Н. С. Иванов

В ряде областей науки и практики иногда приходится обрабатывать малые наборы повторяющихся наблюдений (выборки из нескольких десятков отсчётов), где доля аномально сильно искажённых наблюдений велика и составляет до половины и более половины всех наблюдений. С подобной ситуацией приходится сталкиваться, например, при геофизических измерениях в промышленно развитых районах; при нарушениях радиосвязи, когда части установки разнесены в пространстве и связаны по радиоканалу; в радиотелеметрии, когда принимаемый сигнал мал. Но задача подавления аномально больших, обычно импульсных, помех, искажающих постоянный или периодический сигнал более половины времени измерения, ранее была не решена.

В статье описывается разработанный автором [2] помехоустойчивый метод получения результата измерений при многократных наблюдениях, названный методом достоверного сравнения. Метод сохраняет работоспособность, когда засорение – доля выбросов в выборке наблюдений – доходит до 60%, а остальные наблюдения искажены флуктуационной помехой. Метод работоспособен, даже если все выбросы имеют одинаковую полярность и стремятся исказить результат в одном направлении. Точность метода, когда засорение превышает 5–10%, значительно выше, чем у медианы и очень немного уступает последней при меньших засорениях. Метод достоверного сравнения по алгоритму получения результата измерения идентичен ранее разработанному нами методу сравнения [3, 4], но значительно превосходит последний в точности и помехоустойчивости при определении величины доверительного интервала результата измерения.

Для пояснения работы методов достоверного сравнения и сравнения рассмотрим влияние аномально больших помех на величины известных оценок. Пусть наблюдения периодического или постоянного сигнала искажены флуктуационной помехой с нормальным распределением $N(0, \sigma_1^2)$. Сравним, как влияет на результат обработки многократных наблюдений определяемый по методу среднего арифметического и по методу медианы появление аномально искаженного наблюдения. Оценить воздействие на оценку (на результат измерений) по реализациям с известным распределением наблюдений от появления аномально испорченного на величину x наблюдения можно с помощью функции влияния, впервые введённой Хампелем. Хампелем доказано [6], что при нормированном нормальном распределении помехи $N(0, 1)$ функция влияния для среднего арифметического $IF(x) = x$, а для медианы $IF(x) = (\pi/2)^{1/2} \operatorname{sign}(x)$. Следовательно, наблюдение, искажение которого выходит за пределы $\pm(\pi/2)^{1/2}$, а для ненормированного распределения флуктуационных помех $N(0, \sigma_1^2)$ – за пределы $\pm\sigma_1(\pi/2)^{1/2}$, отклоняет результат измерений по среднему арифметическому в соответствующую сторону больше, чем по медиане.

Указанное расхождение результатов измерений используем для отбраковки наблюдений-выбросов, искаженных аномальной помехой. По выборке наблюдений найдем медиану и среднее арифметическое значение. Если среднее арифметическое больше медианы, то отбракуем максимальное наблюдение, а если меньше – минимальное. По оставшимся реализациям наблюдений повторно найдем медиану и среднее арифметическое и повторим отбраковку. Процедуру будем повторять, пока очередные медиана и среднее арифметическое не совпадут с нужной точностью или их разность не изменит свой знак, или пока не будет отбракована определенная часть (например, 2/3) выборки. Если разность изменила знак, то за результат измерений примем полусумму последних медиан. В других случаях – последнюю медиану.

Какой же точности совпадения медианы и среднего арифметического значения нужно добиваться? Такой, чтобы отбраковать как можно больше аномально испорченных наблюдений-выбросов, но избежать отбраковки наблюдений, искаженных только флуктуационной помехой. Требуемая точность совпадения зависит от величины флуктуационной помехи и размера выборки. Но обычно величина флуктуационной помехи неизвестна, а ее определение затруднительно, поскольку неизвестна доля наблюдений-выбросов, искаженных аномально. Поэтому разумно связать требуемую точность совпадения с параметром выборки наблюдений, в основном определяемым величиной флуктуационной помехи и как можно слабее зависящим от доли наблюдений-выбросов, осложненных аномальной помехой, и от величины этой помехи. В наибольшей степени указанным пожеланиям соответствует размах небольшой заданной части выборки, например ее четверти, имеющей наиболее плотное распределение [3]. Моделирование показало, что хорошие результаты у предложенного метода получаются при точности совпадения медианы и среднего арифметического значения равной kR/N, где R –

размах наиболее плотно расположенной четверти ранжированных наблюдений выборки, N – первоначальное число наблюдений в выборке, k – коэффициент, выбираемый в интервале от 14 до 30. Причем, от увеличения k, при отсутствии в выборке выбросов, получаемый результат измерений асимптотически приближается к результату измерений по методу медианы, а при уменьшении k улучшается результат измерений при большом числе выбросов. Дальнейшее улучшение результатов измерений можно получить, задавая точность совпадения медианы и среднего арифметического значения переменной, уменьшающейся после отбраковки каждого очередного наблюдения, например по формуле $Rk[(N - 1.5j)/N]^2/N$, где j – количество уже отбракованных наблюдений.

Помехоустойчивость предлагаемого способа получения результата измерений при многократных наблюдениях была проверена и подтверждена моделированием его работы на ЭВМ сравнительно с наиболее сильными методами применяемыми на практике. С медианой как наиболее устойчивой оценкой. Со сравнительно устойчивыми и эффективными: медианой Ходжеса–Лемана, как наиболее распространённой ранговой оценкой; М-оценкой Тьюки, как наилучшей М-оценкой. Со средним арифметическим – наиболее эффективным, хотя и не устойчивым.

Моделирование проводилось согласно рекомендации Госстандарта [5], обработкой выборок засоренных нормальных распределений, но жестче, чем требует Госстандарт. Для засорения поочередно использовались не только рекомендуемые Госстандартом симметричные распределения, но и “трудные” для всех оценок односторонние и двумодальное – “трудное” для всех оценок, основанных на медиане. Нами ранее [3] опубликовано 35 рисунков с графиками зависимости среднего квадратического отклонения (СКО) результатов измерений от принятого математического ожидания по сравниваемым методам при увеличении засорения с 0 до 80% для различных распределений аномальной помехи и для выборок различных размеров (4, 8, 16, 32 и 96 наблюдений). Выигрыш в помехоустойчивости у предлагаемых методов наблюдался при размере выборки в 8 и более наблюдений и был особенно велик в случае наиболее “опасных” для всех ранее известных методов односторонних выбросов. Наиболее “опасных” так как каждый выброс стремится сместить результат измерений в одну и ту же сторону. Последнее имеет важное значение, так как иногда на практике вблизи сильного техногенного источника помех встречаются именно односторонние помехи. Такие помехи вызываются, например, сменой режимов работы электродвигателей поездов при ускорении или торможении, когда период полезного сигнала гораздо меньше длительности одного переходного режима электропоезда.

Помехоустойчивость получения результата измерений по предлагаемому методу подтверждена практикой геофизических работ ряда геологоразведочных организаций [3]. У автора имеются акты производственных организаций, подтверждающие, что использование предлагаемого метода позволяло вести геофизические измерения на участках с очень высоким уровнем промышленных помех от ЛЭП, газопроводов, электропоездов и тому подобное, где известной аппаратурой вести измерения невозможно. А так же подтверждающие, что использование предлагаемого метода, сравнительно со средним арифметическим и медианой, обеспечивало увеличение глубинности исследований в несколько раз.

Изложенное выше относительно получения результата измерения справедливо как для метода достоверного сравнения так и для метода сравнения, поскольку по алгоритму получения результата измерения они идентичны. Однако при измерениях недостаточно знать значение собственно результата, не менее важно знать его точность и достоверность, что оценивают с помощью доверительных интервалов. Для определения величины 95% доверительного интервала результата измерений Δ (0.95) в методе сравнения нами ранее [3] было предложено выражение

$$\Delta(0.95) = \frac{4\, med\,\{|x(i) - T|\}}{\sqrt{N - j}}$$

где T – результат измерения, N – первоначальное число наблюдений, j – количество отбракованных наблюдений, $x(i)$ – последовательность наблюдений оставшихся после отбраковки. Однако там же [3] на страницах 70–71 нами оговаривалось, что “доверительные интервалы, определяемые по предложенной формуле, не являются строго 95%. В зависимости от уровня засорения и вида аномальной помехи частота попаданий оценок в определяемые интервалы заметно отклоняется от 95%. Однако ориентировочное представление о точности оценки в пределах работоспособности метода вычисляемые доверительные интервалы дают, и с этой точки зрения их вычисление полезно. При большом уровне засорения значительная часть получаемых оценок начинает сильно выходить за пределы определяемых интервалов и даже за пределы двух доверительных интервалов”. Как в ранее известных методах, так и в методе сравнения наибольшие погрешности при определении доверительных интервалов возникают при наиболее неблагоприятном распределении аномально испорченных наблюдений – антисимметричном с ненулевым математическим ожиданием. Это не позволяло всегда точно определять достоверность результата измерений, и уверенно судить о применимости соответствующего метода при существующем в момент измерения уровне помех.

В последующем вопрос получения правильного доверительного интервала результата измерения был рассмотрен более подробно. Были проведены многочисленные пробы различных способов определения величины 95% доверительного интервала результата измерений моделированием на ЭВМ при увеличении засорения с 0 до 80% для различных распределений аномальной помехи и для выборок наблюдений различных размеров. При моделировании определялся процент d попаданий результата измерения в определяемый для соответствующего опыта 95% двусторонний доверительный интервал Δ (0.95). По результатам моделирования для определения величины 95% доверительного интервала результата измерений в методе достоверного сравнения было предложено преобразованное выражение

$$\Delta(0.95) = \frac{4\, med\,\{|x(i) - T|\}}{\sqrt{N - j}} \times \frac{N + 2.5j}{N}$$

более хорошо отображающее реальность.

Типичные результаты моделирования представлены на рис. 1 и 2. На рисунках по оси ординат отложен процент d попаданий результата измерения T в определяемый по предложенной для соответствующего метода методике 95% двусторонний доверительный интервал Δ (0.95) соответствующего опыта. По оси абсцисс отложено ε – доля наблюдений, искаженных аномально сильной помехой. Кривая 1 отображает получение 95% доверительного интервала в методе достоверного сравнения, кривая 2 – в методе сравнения, кривая 3 – в методе среднего арифметического. Для метода среднего арифметического доверительный интервал вычислялся по обычно применяемой методике предполагающей нормальность распределения [1]. Прямой 4 показан уровень 95%. Рис. 1 показывают воздействие односторонних аномальных помех, полученных с помощью взятых по модулю случайных чисел нормального распределения, рис. 2 – симметричных аномальных помех, полученных с помощью случайных чисел равномерного распределения. На приведённых рисунках доверительный интервал определялся для 32 наблюдений сигнала в одном измерении.

На рис. 1, для одностороннего распределения аномальных помех, видно что процент d попаданий результата измерения в соответствующий 95% двусторонний доверительный интервал, определяемый по методу достоверного сравнения (кривая 1), очень близок к 95% (к прямой 4) пока ε – доля наблюдений, искаженных аномальной помехой, меньше 0.65, процент попаданий по методу сравнения (кривая 2) заметно отличается от 95%, а процент попаданий по методу среднего арифметического (кривая 3) даже при сравнительно небольшом увеличении ε начинает очень сильно отличатся от 95%. Следовательно, при антисимметричном распределении аномальных помех, помехоустойчивость определения 95% двустороннего доверительного интервала по методу достоверного сравнения превосходит помехоустойчивость его определения по методу сравнения и даже стала несколько превосходить помехоустойчивость определения величины самого результата измерения. При симметричном распределении аномальных помех (см. рис. 2) помехоустойчивость определения 95% двустороннего доверительного интервала по всем упомянутым методам практически совпадает.

Правильное помехоустойчивое определение доверительного интервала в методе достоверного

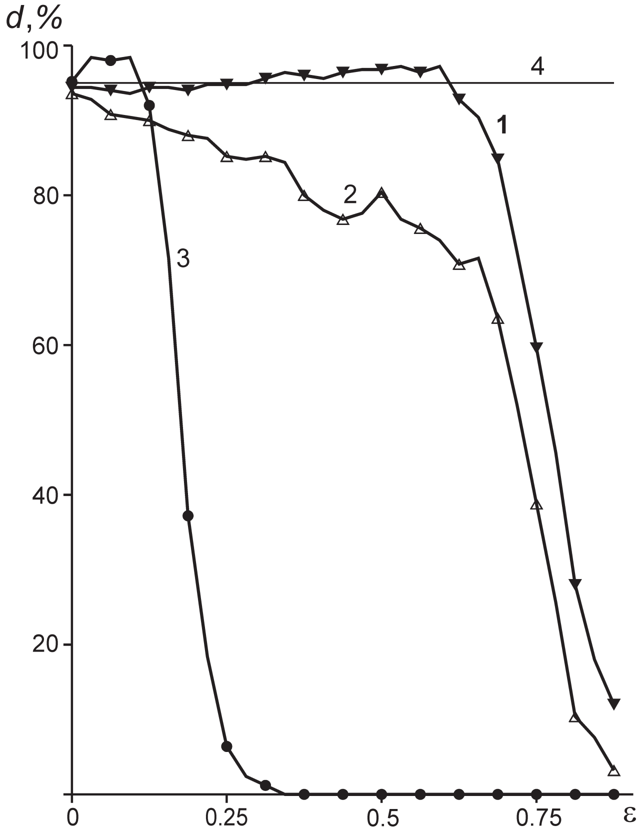

Рис. 1. Процент d попаданий результатов измерений в 95% доверительный интервал в зависимости от доли наблюдений ε, искаженных аномальной односторонней помехой.

Кривая 1 соответствует методу достоверного сравнения, кривая 2 – методу сравнения, кривая 3 – методу среднего арифметического. Прямой 4 показан уровень 95%.

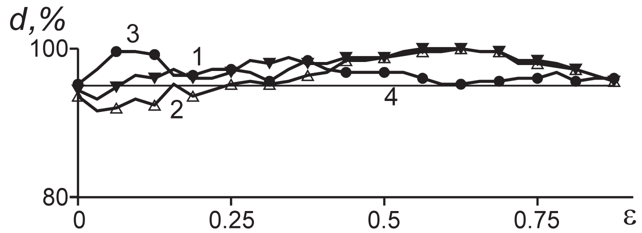

Рис. 2. Аналогично рис. 1, но аномальная помеха симметричная.

сравнения позволяет уверенно определять точность и достоверность значений результатов измерений и применимость метода при существующем в момент измерения уровне помех. По применению метода опубликованы подробные методические указания [2], зарегистрированные Центром метрологии и сертификации "Сертимет" УрО РАН под шифром МУ 88–16360–10–2010 и включённые в Реестр методических указаний УрО РАН.

Исследования проведены в рамках проекта 09-М-2345-2001 "Освоение недр Земли…".

СПИСОК ЛИТЕРАТУРЫ

1. ГОСТ Р 50779. 21-2004. Правила определения и методы расчёта статистических характеристик по выборочным данным. Часть 1. Нормальное распределение.
2. *Иванов Н.С.* Робастный метод обработки результатов измерений засоренных до 60% выбросами с неизвестным законом распределения (метод достоверного сравнения). Методические указания. Екатеринбург: ИГГ УрО РАН, Центр "Сертимет" УрО РАН, 2010. 29 с.
3. *Иванов Н.С.* Новые методы цифровой нелинейной фильтрации аномальных помех с неизвестным законом распределения. Екатеринбург: УрО РАН, 2002. 132 с.
4. *Иванов Н.С., Человечков А.И., Байдиков С.В.* Способ преобразования периодического электрического сигнала в код и устройство для его осуществления. А.с. 1800927 (Россия). 1989.
5. МИ 2174-91. Аттестация алгоритмов и программ обработки данных при измерениях. Основные положения. Госстандарт России. Рекомендация. Государственная система обеспечения единства измерений. СПб.: ВНИИМ, 1993. 27 с.
6. *Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В.* Робастность в статистике. Подход на основе функции влияния. М.: Мир, 1989. 512 с.