

МОДИФИКАЦИЯ МЕТОДА ИРВИНА ДЛЯ ВЫЯВЛЕНИЯ АНОМАЛЬНЫХ УРОВНЕЙ ВРЕМЕННЫХ РЯДОВ: МЕТОДИКА И ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

^{1,2}Трофименко С.В., ¹Маршалов А.Я., ¹Гриб Н.Н., ³Колодезников И.И.

¹Технический институт (филиал) ФГАОУ ВПО "Северо-Восточный федеральный университет имени М.К. Аммосова", г. Нерюнгри, Россия (678960, Республика Саха (Якутия), г. Нерюнгри, ул. Южно – Якутская 23), e-mail: trofimenko_sergei@mail.ru;

²Институт тектоники и геофизики им. Ю.А. Косыгина ДВО РАН;

³Академия наук республики Саха (Якутия), anrsya@mail.ru

Приведены результаты численного моделирования выявления аномальных (пиковых) уровней положительно определенного временного ряда методом Ирвина. Для моделирования использован ряд числовых данных в виде отдельных пиковых значений, осложненных случайной помехой и синусоидальным сигналом. В рамках аддитивной модели исходного ряда задача разделения нормального и аномального уровней методом Ирвина успешно решается в случаях отдельных импульсов. Однако, даже в этом случае, следующий за пиковым значением член ряда с фоновым уровнем интерпретируется как аномальный. В задачах автоматизированной обработки данных ряды пиков осложняются пропорциональным количеством значений фоновых компонент. Для тех случаев, когда количество пиков превышает 50% от общего количества значений ряда и тем более для подряд стоящих пиков сформированный ряд фоновых данных осложняется пиковой компонентой за счет увеличения дисперсии. В результате численного моделирования была разработана методика в виде модифицированного метода Ирвина для адекватного выделения аномальных уровней ряда. Данная задача актуальна для систем электромагнитного мониторинга грозových разрядов и поиска различных источников электромагнитного излучения тектонической природы.

Ключевые слова: временные ряды, аномальные уровни ряда, метод Ирвина, модификация метода Ирвина, разделение уровней ряда, электромагнитное излучение

MODIFICATION OF THE METHOD FOR IRWIN DETECT ABNORMAL LEVELS TIME SERIES: METHOD AND NUMERICAL EXPERIMENTS

^{1,2}Trofimenko S.V., ¹Marshalov A.J., ¹Grib N.N., ³Kolodeznikov I.I.

¹Technical Institute (branch) "North-Eastern Federal University of MK Ammosov" Neryungri, Russia (678960,

²Republic of Sakha (Yakutia), Neryungri Street. South - Yakut 23), e-mail: trofimenko_sergei@mail.ru;

³Institute of Tectonics and Geophysics FEB RAS,

³Academy of Sciences of the Republic of Sakha (Yakutia), anrsya@mail.ru

The results of numerical modeling to detect abnormal (peak) levels positively certain time series by Irwin . To simulate a number of numerical data used as separate peaks, complicated by the random noise and a sinusoidal signal. Under the additive model of the original series attempts to separate normal and abnormal levels by Irwin successfully solved in the cases of individual pulses. However, even in this case, following the peak of the term of the background level is interpreted as abnormal. In the problems of automated data processing complicated series of peaks proportional to the values of background components. For those cases where the number of peaks higher than 50 % of the total values of the series and even more so for a row facing the peaks formed by a number of background data is complicated by the peak component by increasing the dispersion. The numerical simulation method was developed as a modification of the method of Irwin for adequate allocation of abnormal levels of the series. This problem is important for systems of lightning electromagnetic monitoring and research of various sources of electromagnetic radiation of the tectonic nature.

Keywords: time series, abnormal levels of a number, the method of Irvine, Irvine modification of the method, the division series levels, the electromagnetic radiation

Геофизический мониторинг атмосферных источников электромагнитного излучения, разработанный для регистрации и контроля движения грозových разрядов, может быть использован для дистанционного отслеживания геодинамических процессов в земной коре [3, 4], связанных с формирующимися очагами землетрясений и активными разломами

земной коры [6-9]. Важным вопросом при интерпретации аномалий электромагнитного излучения (ЭМИ) является разделение аномалий по типу источника излучения, что детально исследовано в работе [5].

Суммарный сигнал ЭМИ, можно представить в виде:

$$E_t = f(E_t^{im}, E_t^f, \varepsilon_t), \quad (1)$$

где E_t^{im} и E_t^f импульсная аномальная (пики) и фоновые составляющие сигнала, ε_t – случайный компонент. Причем, априори, функция источника сигнала не известна, т.е. для неё не известен тип модели, определяемой по (1).

В линейной теории электромагнитных волн принята аддитивная модель исходного ряда наблюдений ЭМИ на основе принципа суперпозиции. Если временной ряд представляется в виде суммы соответствующих компонент, то полученная модель носит название аддитивной и имеет вид (без учета тренда и сезонной компоненты):

$$Y_t = E_t^{im} + E_t^f + \varepsilon_t, \quad (2)$$

где Y_t – уровни временного ряда.

В данной работе представлены результаты имитационного моделирования разделения полей ЭМИ на основе статистического анализа временного ряда наблюдений, отражающих реальное состояние геофизической среды.

На рис. 1,2 показаны результаты наблюдений сигналов ЭМИ в периоды слабо возмущенного (рис. 1) и возмущенного (рис. 2) состояния геофизической среды. Качественно выделить фоновую составляющую E_t^f в возмущенный день весьма проблематично.

Для отработки технологии разделения полей ЭМИ на фоновую E_t^f и импульсную E_t^{im} составляющие в модели (2) был создан имитационный файл в среде электронных таблиц Excel. Совокупный ряд исходных данных (2) был составлен из суммы рядов пиковых значений E_t^{im} , количество импульсов которых можно задавать произвольно и фоновой компоненты ряда в виде суммы периодической синусоидальной и случайной компонент ряда $E_t^f = \varepsilon_t + f(\sin(\omega t))$ (рис.3). Суммарный ряд подвергался алгоритму выделения импульсов по методу Ирвина [2] (метод пиков), который используется в практике статистического анализа временных рядов с целью выделения аномальных уровней ряда.

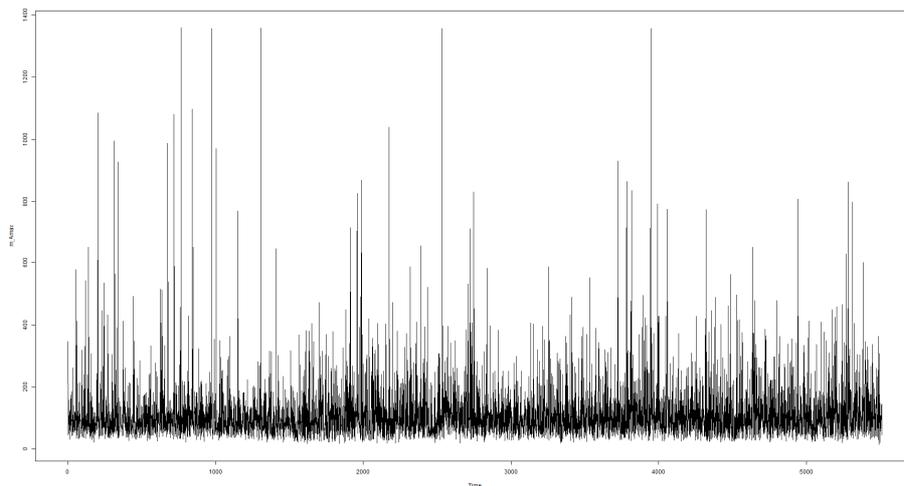


Рис. 1. Исходный ряд значений амплитуд сигналов ЭМИ в слабо возмущенный день 04.01.2013.

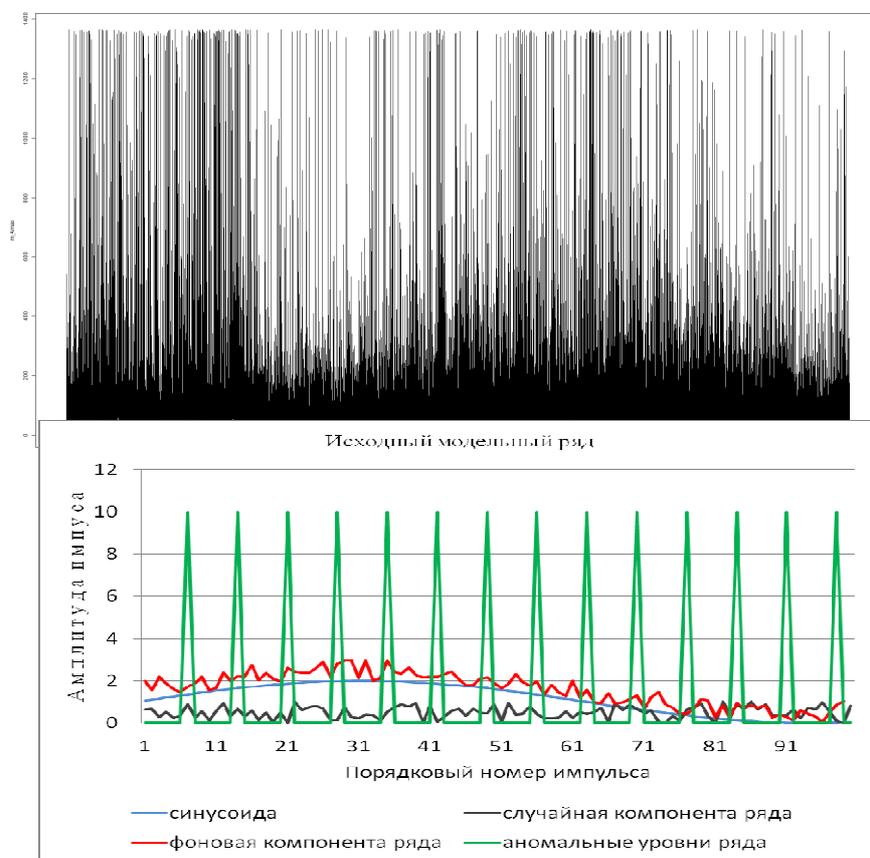


Рис. 2. Исходный ряд значений амплитуд сигналов ЭМИ в сильно возмущенный день 13.01.2013.

Рис. 3. Компоненты модельного ряда для имитации сигналов ЭМИ

Пусть имеется временной ряд $\{y_t | t = 1, 2, \dots, n\}$. Метод Ирвина предполагает использование соотношения

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y}, \quad (3)$$

где $\sigma_y = \sqrt{\frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n-1}}$ – стандартное отклонение, $\bar{y} = \frac{\sum_{t=1}^n y_t}{n}$ – среднее значение амплитуд,

для оценки аномального приращения амплитуды последующего члена ряда $y_t - y_{t-1}$ в

предположении, что предыдущий член ряда относится к фоновой составляющей сигнала E_t^f . Расчетные значения $\lambda_t : \lambda_2, \lambda_3, \dots, \lambda_n$ сравниваются с табличными значениями критерия Ирвина λ_α ; если какое-либо из них λ_t оказывается больше табличного λ_α , то соответствующее значение y_t уровня t ряда считается аномальным. Значения критерия Ирвина для уровня значимости $\alpha = 0,05$ приведены в таблице 1.

Таблица 1

Значения критических значений λ_α для уровня значимости $\alpha = 0,05$

| | | | | | | | |
|------------------|-----|-----|-----|-----|-----|-----|-----|
| n | 2 | 3 | 10 | 20 | 30 | 50 | 100 |
| λ_α | 2.8 | 2.8 | 1.5 | 1.3 | 1.2 | 1.1 | 1 |

Для практических расчетов при известной длине реализации n_i можно использовать аналитическое выражение для определения критических значений λ_α^i , для которой погрешность аппроксимации меньше 1%:

$$\lambda_\alpha^i = 1.478 \cdot x_i^{-0.1767}, \quad (4)$$

где $x_i = n_i / 10$.

Численные эксперименты показали следующее: во-первых, для одиночного пика (аномального уровня) Y_t относительно фонового уровня Y_{t-1} следующий член ряда Y_{t+1} также интерпретируется как аномальный, хотя амплитуда его импульса сравнима с фоновым значением. Во-вторых, эта же погрешность будет присутствовать при наличии в исходном ряде нескольких аномальных уровней $Y_t, Y_{t+1}, \dots, Y_{t+k}$. В данном случае, за аномальный уровень ряда будет принят первый аномальный Y_t и следующий за серией аномалий фоновый уровень Y_{t+k+1} . Все промежуточные аномалии Y_{t+1}, \dots, Y_{t+k} будут отнесены к фоновой составляющей ряда. Это приведет к искажению (завышению) огибающей фоновой компоненты (рис. 4).

При интерактивном способе разделения уровней ряда фиктивные аномальные уровни отбраковываются оператором. В задачах автоматизированной обработки данных ряды пиков осложняются пропорциональным количеством значений фоновых компонент.

Для тех случаев, когда количество пиков превышает 50% от общего количества значений ряда и тем более для подряд стоящих пиков сформированный ряд фоновых данных осложняется пиковой компонентой (рис.5). Видимо эти погрешности до сих пор не позволили методу Ирвина найти достаточно широкое применение в практике статистических исследований временных рядов.

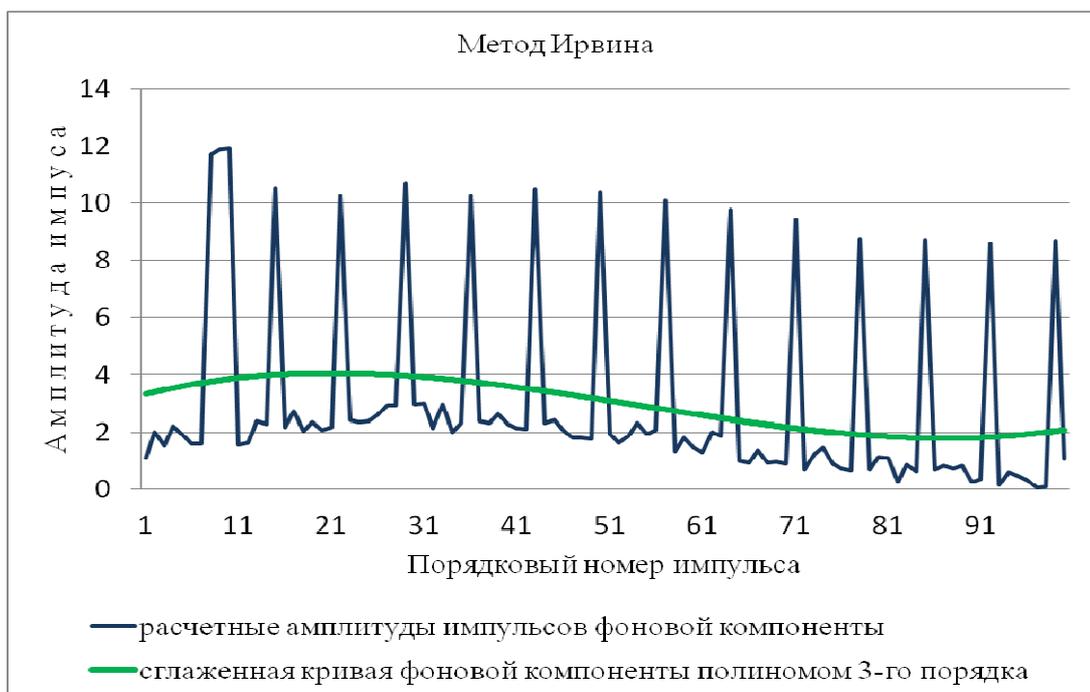


Рис. 4. Иллюстрация осложнения фоновой компоненты ряда аномальными уровнями при использовании метода Ирвина

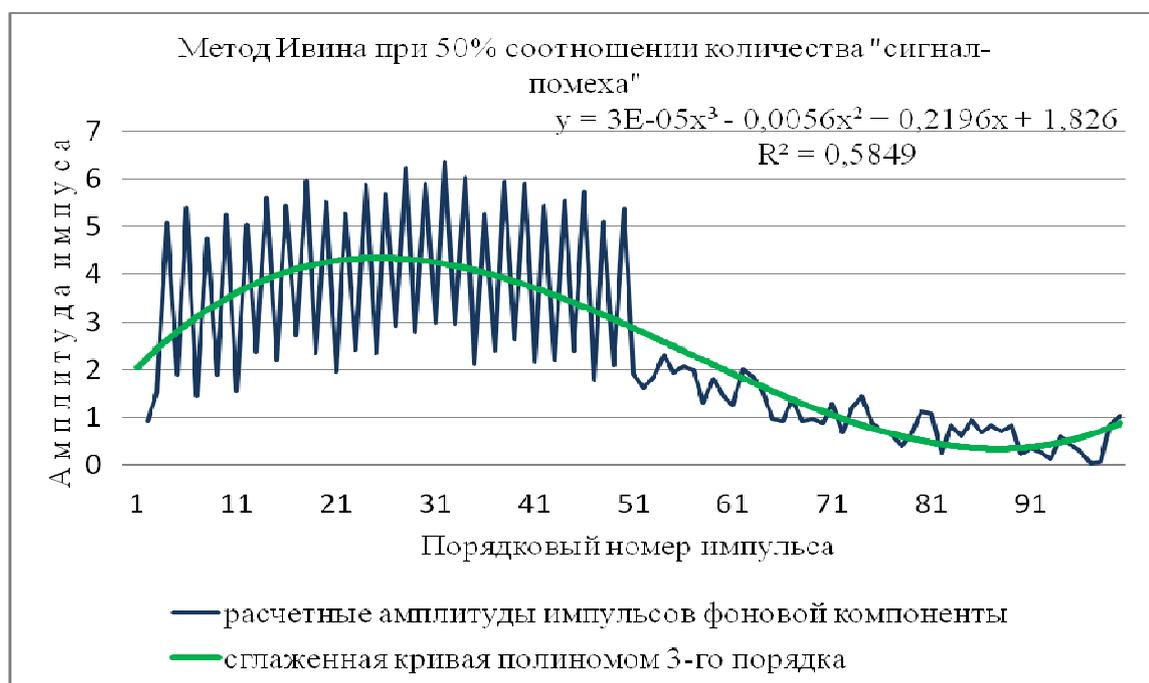


Рис. 5. Осложнение фоновой составляющей имитационного ряда при 50% соотношении пиковых и фоновых компонент

В результате численного моделирования различных соотношений «сигнал - помеха» была разработана методика в виде модифицированного метода Ирвина для адекватного разделения аномальных (пиковых) и фоновых уровней ряда. Данная задача, как было указано

выше, актуальна для систем электромагнитного мониторинга грозовых разрядов и поиска различных источников электромагнитного излучения тектонической природы.

Для исключения указанных погрешностей метода Ирвина было разработано три варианта расчетов для разделения уровней ряда, применение которых обосновано сложностью соотношений «сигнал-помеха» в реальных экспериментальных данных, например, как это показано на рис. 1, 2.

В простейшем случае, для выделения одиночных пиков положительно определенного ряда к условию Ирвина (3) добавляется условие не отрицательности разности $y_t - y_{t-1}$, т.е. $y_t - y_{t-1} > 0$. Все аномальные уровни в фоновой компоненте ряда заменяются значениями по формуле параболического интерполирования:

$$\bar{y}_n = \frac{-y_{n-2} + 2y_{n-1} + 5y_n}{6}, \quad (5)$$

где y_{n-2}, y_{n-1}, y_n члены ряда фоновой компоненты, предшествующие аномальному уровню y_{n+1} . Так как первый член ряда может быть с аномальной амплитудой, то в начале ряда проставляются три дополнительных члена ряда с минимальной амплитудой. Результаты численного моделирования по данному алгоритму показаны на рис. 6.



Рис. 6. Выделение фоновой составляющей имитационного ряда при наличии одиночных пиковых компонент с дополнительным условием не отрицательности разности сравниваемых амплитуд последовательных членов ряда

Сравнение с результатами расчетов с применением простого алгоритма Ирвина (рис. 4) наглядно показывает необходимость введения в алгоритм дополнительного условия не

отрицательности разности $y_t - y_{t-1} > 0$.

При увеличении количества аномальных уровней до 50% замена пиковых значений по формуле (5) приводит к осложнению фоновой компоненты усредненными значениями амплитуд аномальных уровней (рис. 5). В данном случае применяется алгоритм последовательного исключения аномальных уровней.

Амплитуды сигнала аномальных уровней в фоновой компоненте заменяются по формуле (5) с возвратом на начало цикла расчетов. При этом значение стандартного отклонения σ_y рассчитывается для первоначального ряда и при повторных расчетах принимается за константу.

Данный алгоритм позволяет адекватно разделять фоновую и импульсную компоненты, однако требует большего времени для расчетов, так как в массиве данных за сутки накапливается до 80000 импульсов и более, что приведет к задержке отображения результатов в реальном времени в системах автоматизированного мониторинга.

Для случаев возмущенного состояния геофизической среды (см. рис. 2) разработан алгоритм с построением вариационного ряда с возрастающими амплитудами сигнала. Для $\alpha = (20-40)\%$ от общего количества членов вариационного ряда рассчитывается стандартное отклонение σ_y и применяется метод Ирвина до появления первого аномального уровня Y_k . Максимальное значение из первых $(k-1)$ уровней вариационного ряда принимается за граничное значение фоновой компоненты, по которому производится разделение компонент исходного ряда.

Численное моделирование с использованием данных натурального эксперимента показало, что, если в исходном ряду присутствует только две компоненты, то фоновая компонента выделяется из исходного ряда без искажений. При наличии нескольких источников ЭМИ метод чувствителен к выбору количества членов вариационного ряда для расчета σ_y . Это может привести к пропуску аномальных источников сигнала либо к выделению мнимых источников (ошибки первого и второго родов). Для исключения данных погрешностей в системе автоматического мониторинга предполагается производить расчеты с использованием различных процентных выборок α из вариационного ряда для расчета σ_y и применения метода Ирвина. Адекватность модели во всех случаях проверяется по статистическим характеристикам остаточного ряда [1]. Проведенные расчеты для двух дней с различным состоянием возмущенности геофизической среды (рис.1, 2) показали сходимость моделей фоновых компонент, выделенных по методам исключения пиков и построения вариационного ряда при $\alpha = (35 - 45)\%$.

Список литературы

1. Калинина В. Н., Панкин В. Ф. Математическая статистика. – М.: Высшая школа, 2001. – 336 с.
2. Трофименко С.В. Методы и примеры статистических оценок временных рядов //Международный журнал экспериментального образования", №9, 2013.-С.41-42.
3. Трофименко С.В. Проявление землетрясений на фоне стационарного сейсмического процесса Олекмо-Становой зоны (ОСЗ) Горный информационно-аналитический бюллетень (научно-технический журнал). 2007. Т. 17. №1. С. 208-213.
4. Трофименко С.В. Геофизические поля и сейсмичность Южной Якутии //Горный информационно-аналитический бюллетень (научно-технический журнал). 2007. Т. 17. №1. С. 188-196.
5. Трофименко С.В. Детальные геолого-геофизические исследования зон активных разломов и сейсмическая опасность Южно-Якутского региона /А.Н. Овсюченко, Трофименко С.В., Мараханов А.В., П.С., Карасев, Е.А. Рогожин, В.С. Имаев, В. М. Никитин, Н.Н.Гриб // Тихоокеанская геология, 2009. – том 28. – №4. – С. 55-74.
6. Трофименко С.В. Очаговые зоны сильных землетрясений Южной Якутии /Овсюченко А.Н., Трофименко С.В., Мараханов А.В., П.С., Карасев, Е.А. Рогожин, В.С. //Физика Земли. 2009. №2. С. 15-33.
7. Трофименко С.В. Сейсмоструктура переходной области от Байкальской рифтовой зоны к орогенному поднятию Станового хребта /А.Н.Овсюченко, Трофименко С.В., Мараханов А.В., П.С., Карасев, Е.А. Рогожин, В.С. //Геотектоника, 2010, № 1, с. 29-51.
8. Трофименко С.В. Тектоническая интерпретация статистической модели распределений азимутов аномалий гравимагнитных полей Алданского щита. – Тихоокеанская геология. – 2010. – Том 29. – №3. – С. 64-77.
9. Трофименко С.В. Активные нектонические нарушения участка Алдан-Нагорный нефтепроводной системы Восточная Сибирь – Тихий океан /Карасев П.С., Овсюченко А.Н., Мараханов А.В., Трофименко С.В. // Нефтяное хозяйство. – 2008. – № 9. –С. 80-84.

Рецензенты:

Омельяненко А.В., д.т.н., профессор, главный научный сотрудник лаборатории инженерной геокриологии Института мерзлотоведения им. П.И.Мельникова СО РАН, г. Якутск;
Имаев В.С., д.г.-м.н., профессор, гл. научный сотрудник Института земной коры СО РАН, г.Иркутск.