

УДК: 550.343.3+550.34.016

А.С. Черепанцев

**МЕТОДИЧЕСКИЕ АСПЕКТЫ ВЫДЕЛЕНИЯ ВНЕШНИХ ПРИРОДНЫХ
ФАКТОРОВ В ЗАПИСЯХ ВАРИАЦИЙ ОБЪЕМНОЙ ДЕФОРМАЦИИ**

Рассмотрены возможности улучшения выделения внешних факторов в вариациях геофизических полей на примере поля объемной деформации. Описана методика робастной оценки параметров факторной модели на основе цензурирования выборки. Получены оценки искажения параметров факторной модели в случае применения последовательных моделей снятия связанных факторов. На примере снятия внешних факторов барического воздействия и приливных деформаций в вариациях объемной деформации показана эффективность рассмотренных методик.

Геофизические наблюдения; многофакторный анализ; робастное оценивание.

A.S. Cherepantsev

**METHODOLOGICAL ASPECTS OF THE EXTERNAL FACTORS
ALLOCATION IN VOLUMETRIC STRAIN VARIATIONS**

Improving the allocation of external factors in the variations of geophysical fields on the example of the volumetric strain variations is studied. A technique of parameters of the factor model robust estimation, based on the censored sample is shown. Score values of the parameter distortion in the factor model in the case of sequential patterns of removed factors is obtained. On the example of removing the atmospheric pressure and the tidal strain variations in volumetric strain shows the efficiency of the considered techniques.

Geophysical observations; multivariate analysis; robust estimation.

Одним из направлений по исследованию и контролю состояния верхних слоев литосферы является сбор и последующий анализ комплекса геофизических полей, чувствительных к напряженно деформированному состоянию геофизической среды в точке наблюдения.

Особенностью исследуемых временных рядов наблюдаемых полей является присутствие в них наряду с компонентой, источником которой являются деформационные тектонические процессы, компонент, имеющих иную природу. К основным внешним источниками деформации приповерхностных слоев литосферы относятся барическая атмосферная нагрузка земной поверхности, лунно-солнечные приливные изменения формы Земли, антропогенное воздействие, гидрологические процессы как на поверхности Земли (выпадение осадков, снегонакопление), так и в гидрологически активных слоях. Задача динамического анализа выделяемой эндогенной компоненты накладывает дополнительное требование максимальной эффективности снятия внешних факторов, амплитудный вклад которых в наблюдаемое поле может превышать искомую составляющую. Ряд методов улучшения эффективности такого выделения рассмотрены в работе.

Выбор в качестве исходных данных временных рядов вариаций геофизических полей, полученных американскими учеными в ходе многолетних наблюдений на геофизическом полигоне Северной Калифорнии (сегмент разлома Сан-Андреас Паркфилд), обусловлен, прежде всего, накопленным большим объемом информации (длительность 16 лет, интервал дискретизации 10 минут) о характере тектонических процессов в данном регионе, широкой сети долговременных режимных наблюдений комплексов геофизических полей и сопутствующих факторов, открытостью источников данных (www.ncedc.org/ul/) [1]. Для анализа на каждом полигоне взяты временные ряды скважинных наблюдений поля объемной деформации.

Наряду с исследуемыми полями в точках наблюдения проводилась регистрация сопутствующих метеорологических полей (атмосферное давление, температура на поверхности, выпадение осадков).

Предварительная обработка включала в себя наряду с выделением и изъятием искаженных значений, также компенсацию дефектов временных рядов, таких как ступенчатые скачки нулевого отсчета, меандровые скачки среднего значения.

Традиционный подход к разделению различных компонент предполагает отсутствие взаимодействия влияющих процессов. В этом случае систему можно представить отдельными динамическими системами. Результатом суммирования данных компонент и является наблюдаемое поле:

$$y_{dilatation}(t) = \sum_i x_i(t) + \varepsilon(t), \quad (1)$$

где $\varepsilon(t)$ – случайная компонента.

Рассмотрим особенности применения линейного регрессионного анализа к задаче разделения компонент различных динамических систем. Линейная регрессионная модель динамической системы может быть представлена в виде

$$y(t) + \sum_{i=1}^s \alpha_i y^{(i)}(t) = d + \sum_{j=0}^m \mathbf{x}^{(j)}(t) \boldsymbol{\beta}_j + u(t) + \delta(t), \quad (2)$$

где $y(t)$, $y^{(i)}(t)$ – анализируемый временной ряд и его i -я производная по времени, $\mathbf{x}^{(j)}(t)$ – j -я производная по времени временного ряда вектора воздействующих факторов размерности $m \times n$ (m – число факторов, n – длина временного ряда). Нулевая производная соответствует непосредственно вектору воздействующих факторов, d – статическое смещение, $u(t)$ – неизвестная компонента, $\delta(t)$ – случайный процесс с законом распределения $N(0, \sigma^2)$.

Для решения задачи оценки неизвестных параметров $\{\alpha_i, \boldsymbol{\beta}_j, d\}$ традиционно используется метод наименьших квадратов или максимального правдоподобия.

Принципиальные особенности использования метода наименьших квадратов применительно к анализу геофизических рядов включают в себя:

1. Остаточные слагаемые в (2) включают как чисто случайный процесс $\delta(t)$, так и компоненту динамической системы $u(t)$, поведение которой может существенно отличаться от нормального случайного процесса. С другой стороны, использование метода наименьших квадратов для оценки параметров регрессионной модели необходимо требует нормальности остатка. Данное противоречие можно избежать, если учесть свойство нелинейных динамических систем с хаотическим поведением, заключающееся в существовании инвариантной конечной величины – времени предсказания поведения системы T . При $t > T$ значение компоненты системы может быть предсказано только статистически и отсчеты компоненты, выполненные с дискретизацией $\Delta t > T$ независимыми. В [2] показана возможность использования этого свойства в задаче выделения компоненты динамической системы при наложении медленно меняющейся компоненты тренда или сезонной составляющей, характерных для наблюдаемых геофизических полей.

2. Чувствительность метода к присутствию выбросов и смещений «0» в анализируемой выборке, обусловлена невыполнением необходимого требования независимости и нормальности распределения остатка $u(t) + \delta(t)$. Построение робастной оценки параметров модели может заключаться в цензурировании выборки.

Предлагаемая итерационная схема включает в себя: расчет модели $\{\alpha_i, \beta_j, d\}$ по имеющимся данным, исключение из ряда отсчетов, имеющих невязки $\varepsilon_k = \left| y_k - \sum_{i=1}^s \alpha_i y_k^{(i)} - \sum_{j=0}^m \mathbf{x}_k^{(j)} \beta_j \right|$, превышающие заданную величину a , расчет модели $\{\tilde{\alpha}_i, \tilde{\beta}_j, \tilde{d}\}$ по оставшимся данным. Окончание итерационного процесса определяется насыщением дисперсии остатка, т.е. для заданной величины h выполнение на p и $p+1$ итерации неравенства $|\sigma_\varepsilon^p - \sigma_\varepsilon^{p+1}| \leq h$.

Данный метод жесткой фильтрации можно считать частным случаем общего метода робастной непараметрической регрессии – локально взвешенного сглаживания, которое представляет собой итерационный процесс, в котором объекты с большими отклонениями получают меньший вес, после чего регрессия строится заново. Процесс повторяется до тех пор, пока веса не стабилизируются. Так, в качестве весовой функции можно использовать квадратичное ядро $H(r) = (1 - r^2)^2$,

$$(|r| < 1), \text{ где } r = \frac{\varepsilon_i}{6 \cdot \text{median}(\varepsilon_i)}.$$

3. Особенностью применения регрессионного моделирования при снятии компонент внешних воздействующих факторов является их структурная организация. Как правило, геофизические факторы различной природы имеют компоненты, не являющиеся независимыми между собой (для исследуемого поля объемной деформации это, прежде всего, барическое давление и приливные деформации). Присутствие суточной и полусуточной компонент характерно для обоих полей. В этом случае принципиально важно построение модели с учетом комплекса связанных факторов. Встречающееся в литературе указание на компенсацию отдельного фактора (для объемной деформации – барического давления) и приводимые последующие заключения о вариации во времени коэффициент передачи давление-деформации, может приводить к существенному искажению значений коэффициента связи и соответственно выводов о его поведении. Несложно получить соотношения для оценки искажений модели в этом случае.

Пусть $\mathbf{F} = \{f_n^m\}$ – матрица m факторов, каждый длиной по n отсчетов. Решение метода наименьших квадратов задачи построения линейной регрессионной модели данных \mathbf{y} (n -мерный вектор) по m данным $\mathbf{f}^i: \mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$, определяется произведением $\boldsymbol{\alpha} = \mathbf{F}^+ \mathbf{y} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{y}$.

Погрешность в оценке параметров модели $\boldsymbol{\alpha}$ в случае неполного учета факторов рассмотрим на примере модели последовательного учета первых k факторов ($1 \leq k < m$) в исходных данных \mathbf{y} , а далее остальных $(m-k)$ факторов в данных остатка $\boldsymbol{\varepsilon}$ первой модели. Для сравнения параметров полной и последовательной моделей линейной регрессии, представим матрицу факторов \mathbf{F} в виде объединения матриц двух групп факторов: $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2)$, тогда обобщенная обратная матрица представляется в виде

$$(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T = \left((\mathbf{F}_1, \mathbf{F}_2)^T (\mathbf{F}_1, \mathbf{F}_2) \right)^{-1} (\mathbf{F}_1, \mathbf{F}_2)^T = \begin{pmatrix} \mathbf{F}_1^T \mathbf{F}_1 & \mathbf{F}_1^T \mathbf{F}_2 \\ \mathbf{F}_2^T \mathbf{F}_1 & \mathbf{F}_2^T \mathbf{F}_2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{F}_1^T \\ \mathbf{F}_2^T \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{F}_1^T \\ \mathbf{F}_2^T \end{pmatrix},$$

где Σ_{11}, Σ_{22} – автоковариационные матрицы первой и второй группы факторов соответственно, а Σ_{12}, Σ_{21} – их взаимноквариационные матрицы, причем $\Sigma_{12} = \Sigma_{21}^T$.

Для представления обратной матрицы справедливо

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} = \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix}.$$

Тогда окончательное соотношение для параметров полной модели регрессии

$$\alpha = \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \mathbf{F}_1^T - \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \mathbf{F}_2^T \\ -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} \mathbf{F}_1^T + (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \mathbf{F}_2^T \end{pmatrix} \mathbf{y}. \quad (3)$$

При последовательном учете факторов, параметры модели определяется:

$$\alpha^* = \begin{pmatrix} (\mathbf{F}_1^T \mathbf{F}_1)^{-1} \mathbf{F}_1^T \mathbf{y} \\ (\mathbf{F}_2^T \mathbf{F}_2)^{-1} \mathbf{F}_2^T (\mathbf{y} - (\mathbf{F}_1^T \mathbf{F}_1)^{-1} \mathbf{F}_1^T \mathbf{y}) \end{pmatrix} = \begin{pmatrix} \Sigma_{11}^{-1} \mathbf{F}_1^T \\ -\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \mathbf{F}_1^T + \Sigma_{22}^{-1} \mathbf{F}_2^T \end{pmatrix} \mathbf{y}. \quad (4)$$

Формулы (3), (4) для повышения наглядности удобнее представить через множественный коэффициент корреляции между факторами \mathbf{F}_1 и \mathbf{F}_2 [4]:

$$\rho_{\mathbf{F}_i \bullet \mathbf{F}_j} = \sqrt{\Sigma_{ij} \Sigma_{jj}^{-1} \Sigma_{ji}} / \sqrt{\Sigma_{ii}}, \quad i, j = 1, 2.$$

Квадрат множественного коэффициента корреляции $\rho_{\mathbf{F}_1 \bullet \mathbf{F}_2}$ характеризует величину линейной связи между факторами \mathbf{F}_1 , \mathbf{F}_2 и равен доле дисперсии \mathbf{F}_1 , описанной линейной регрессионной моделью $\mathbf{F}_1 = \mathbf{F}_2 \beta_{1 \bullet 2} + \xi$: $\rho_{\mathbf{F}_1 \bullet \mathbf{F}_2}^2 = \frac{\sigma_{\mathbf{F}_1}^2 - \sigma_{\xi}^2}{\sigma_{\mathbf{F}_1}^2}$ ($\beta_{1 \bullet 2}$ – вектор параметров регрессионной модели, имеющий размерность $(m-k)$).

С учетом того, что $\beta_{i \bullet j} = \Sigma_{ij}^{-1} \Sigma_{ji}$, для параметров моделей справедливо

$$\alpha = \begin{pmatrix} \frac{\Sigma_{11}^{-1} \mathbf{F}_1^T - \beta_{2 \bullet 1} \frac{\Sigma_{22}^{-1} \mathbf{F}_2^T}{1 - \rho_{\mathbf{F}_2 \bullet \mathbf{F}_1}^2}}{1 - \rho_{\mathbf{F}_1 \bullet \mathbf{F}_2}^2} \\ -\beta_{1 \bullet 2} \frac{\Sigma_{11}^{-1} \mathbf{F}_1^T}{1 - \rho_{\mathbf{F}_1 \bullet \mathbf{F}_2}^2} + \frac{\Sigma_{22}^{-1} \mathbf{F}_2^T}{1 - \rho_{\mathbf{F}_2 \bullet \mathbf{F}_1}^2} \end{pmatrix} \mathbf{y}, \quad (5)$$

$$\alpha^* = \begin{pmatrix} \Sigma_{11}^{-1} \mathbf{F}_1^T \\ -\beta_{1 \bullet 2} \Sigma_{11}^{-1} \mathbf{F}_1^T + \frac{\Sigma_{22}^{-1} \mathbf{F}_2^T}{1 - \rho_{\mathbf{F}_2 \bullet \mathbf{F}_1}^2} \end{pmatrix} \mathbf{y}. \quad (6)$$

Как следует из полученных соотношений, в случае отсутствия линейной связи факторов \mathbf{F}_1 и \mathbf{F}_2 , обе модели имеют одинаковые параметры ($\rho_{\mathbf{F}_1 \bullet \mathbf{F}_2} = \rho_{\mathbf{F}_2 \bullet \mathbf{F}_1} = 1$, $\beta_{1 \bullet 2} = \beta_{2 \bullet 1} = 0$). В случае же связи факторов модельные параметры α^* могут существенно отличаться от параметров α полной модели. Неполный учет связанных факторов и оценка влияния лишь части факторов \mathbf{F}_1 приводит к искажению истинных значений параметров. Величина таких искажений определяется значениями множественных коэффициентов корреляции ρ между первой и второй груп-

пой факторов и коэффициентами связи $\beta_{2\bullet 1}$ второй группы факторов по первой (верхняя строка в (5) и (6)).

В качестве основных воздействующих факторов рассмотрены основные компоненты приливных объемных деформаций (11 компонент), полученные путем непосредственного расчета приливного потенциала в заданной точке поверхности Земли с помощью программного пакета Egtab [3]. Рассмотренные приливные компоненты группируются около суточного и полусуточного периодов.

В качестве иллюстрации необходимости учета взаимной связи внешних факторов при анализе, на рис. 1 представлены результаты компенсации давления и приливов, полученные при последовательном и одновременном снятии факторов. Как следует из полученных спектральных закономерностей, последовательное снятие, в отличие от одновременного, не обеспечивает компенсации суточной и полусуточной компонент в спектре остатка.

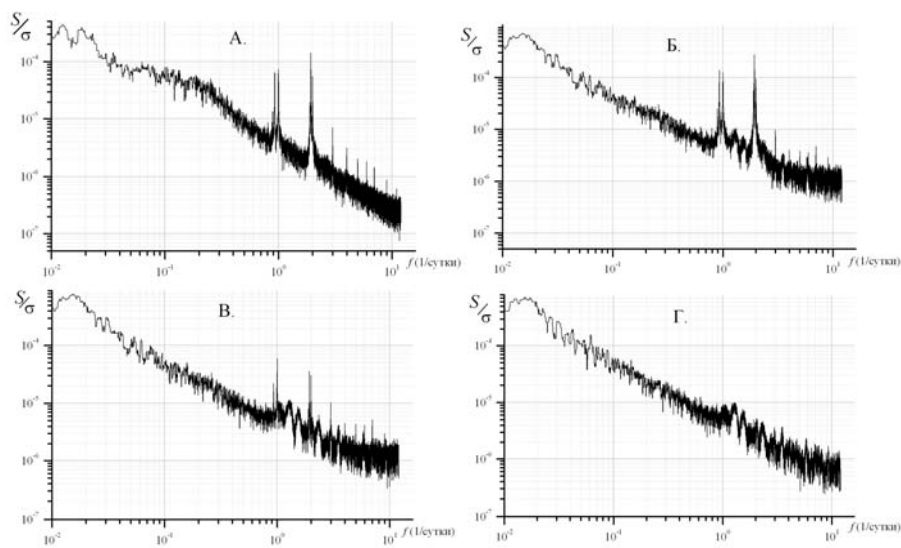


Рис. 1. Нормированные амплитудные спектры объемной деформации в точке наблюдения PKDLT: А – спектр исходного сигнала ($\sigma = 57 \text{ nstrn}$); Б – спектр ряда после компенсации барического давления ($\sigma = 37 \text{ nstrn}$); В – спектр ряда после последовательной компенсации барического давления и приливных компонент ($\sigma = 36 \text{ nstrn}$); Г – спектр ряда после совместной компенсации барического давления и приливных компонент ($\sigma = 35 \text{ nstrn}$)

С целью учета в регрессионной модели неизвестных факторов отклонения реальных значений амплитуды и фазы от нормальных величин, рассмотрены отдельные нормальные компоненты, каждая из которых может иметь свои собственные факторы отклонения. Амплитудный фактор учитывается непосредственно в передаточном коэффициенте модели. Для учета фазового фактора, в предположении его малости рассмотрен фактор производной каждой отдельной компоненты приливных деформаций, исходя из соотношения для гармонического воздействия.

Пусть $x(t)$ – расчетное нормальное значение компоненты приливной объемной деформации, $y(t)$ – реальное значение приливной деформации, имеющее фазовый сдвиг относительно $x(t)$. Тогда

$$y(t) = A \cdot \sin(\omega t + \varphi) = A \cos \varphi \sin \omega t + A \sin \varphi \cos \omega t = \alpha_1 x(t) + \alpha_2 \dot{x}(t).$$

Коэффициенты α_1 , α_2 оцениваются по регрессионной модели. Тогда оценка фазового фактора в случае гармонического приливного воздействия

$$\varphi = \arctg\left(\frac{\omega \cdot \alpha_2}{\alpha_1}\right), \text{ а амплитудного фактора } A = \sqrt{\alpha_1^2 + \omega^2 \alpha_2^2}.$$

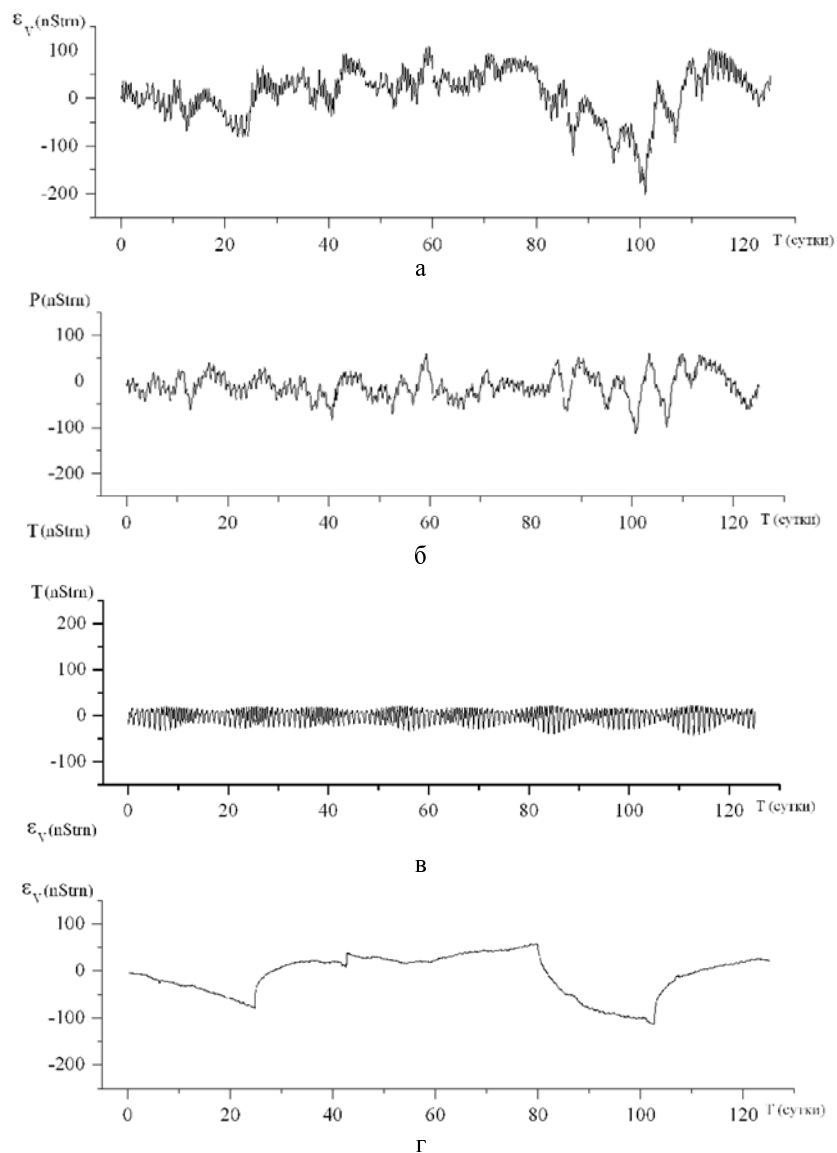


Рис. 2. Фрагмент компенсации аддитивных факторов барического давления и приливного воздействия в вариации объемной деформации. Точка наблюдения PKDLT, начальное время – 7h 12.7.1990. Положительное приращение деформации соответствует сжатию; а – исходные вариации объемной деформации; б – вариации атмосферного давления, приведенные к эквивалентной деформации; в – вариации приливных деформаций; г – вариации объемной деформации с компенсированными внешними факторами

В общем случае для малых сдвигов τ оценку можно также получить исходя из приближения

$$y(t) = A \cdot x(t - \tau) \approx A \cdot x(t) - A\tau \cdot \dot{x}(t) = \alpha_1 x(t) + \alpha_2 \dot{x}(t);$$

$$\tau = -\frac{\alpha_2}{\alpha_1}.$$

Предполагая малость временных задержек τ , справедливо следующее приближенное модельное соотношение:

$$d(t) + \dot{d}(t)\tau_d = \beta_1 p(t) + \beta_1 \tau_p \dot{p}(t) + \sum_{k=1}^{11} \lambda_k x_k(t) + \sum_{k=1}^{11} \lambda_k \tau_{x_k} \dot{x}_k(t) + \varepsilon_i.$$

Таким образом, в общем виде модель представляет собой простейшую динамическую систему, описываемую системой обыкновенных дифференциальных уравнений первого порядка:

$$d_i = \gamma + \alpha \dot{d}_i + \beta_1 p_i + \beta_2 \dot{p}_i + \sum_{k=1}^{11} (\lambda_1^k x_i^k + \lambda_2^k \dot{x}_i^k) + \varepsilon_i,$$

где d – временной ряд деформаций, p – временной ряд барического давления, x^k – временные ряды компонент теоретических значений приливных деформаций.

На рис. 2 представлены исходные ряды наблюдений объемной деформации и внешних воздействующих факторов, а также результат компенсации внешних факторов в вариациях объемной деформации.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Prescott W., Hodgkinson K., etc. Access to Strain and Other Low Frequency Geophysical Observations. EarthScope Workshop: Making and Breaking a Continent. 2001, Snowbird, Utah. – P. 5.
2. Черепанцев А.С. Выделение динамической составляющей в вариациях геофизических полей на основе сходимости выборочного среднего // Физика Земли. – 2008. – № 11. – С. 31-46.
3. Wenzel, H.-G., 1996. The nanogal software: Earth tide processing package ETERNA. Bull. Inf. Marées Terrestres, [124]:9425-9439.
4. Афуфи А., Эйзен С. Статический анализ. Подход с использованием ЭВМ. – М., 1982. – 488 с.

Статью рекомендовал к опубликованию д.ф.-м.н. Г.В. Куповых.

Черепанцев Александр Сергеевич

Технологический институт федерального государственного автономного образовательного учреждения высшего профессионального образования «Южный федеральный университет» в г. Таганроге.

E-mail: s6319a@mail.ru.

347928, г. Таганрог, пер. Некрасовский, 44.

Тел.: 88634371663.

К.ф.-м.н.; доцент.

Cherepantsev Alexandr Sergeevich

Taganrog Institute of Technology – Federal State-Owned Autonomous Educational Establishment of Higher Vocational Education “Southern Federal University”.

E-mail: s6319a@mail.ru.

44, Nekrasovsky, Taganrog, 347928, Russia.

Phone: +78634371663.

Cand. of Eng. Sc.; Associate Professor.