

Анализ временных рядов на примере регистрации вариаций гравитационного поля

*Р.З. Буртиев¹, Ю.В. Семенова³, В.Т. Кирияк², Е.В. Сидоренко²,
С.В. Тройан¹, В.Ю. Карданец¹, Д.В. Нука², 2021*

¹Институт геологии и сейсмологии АН Молдовы, Кишинев, Молдова

²Технический университет Молдовы, Кишинев, Молдова

³Институт геофизики им. С.И. Субботина НАН Украины, Киев, Украина

Поступила 1 апреля 2021 г.

В статье использована модель временного ряда для исследования структуры рядов гравиметрических данных. Наблюдения за активностью геофизических процессов показали, что периоды вариаций геофизических процессов хаотичны на оси времени. По их графику нельзя определенно говорить о закономерности в продолжительности периодов вариаций, чередовании периодов сейсмического затишья и высокой сейсмической активности. Структура ряда проанализирована формальными методами для поиска статистической закономерности в вариациях геофизических параметров во времени. Для исследования динамики геофизических событий были задействованы модели временных рядов. Прогнозирование осуществлено с помощью пакета SPSS20 и программы EXCEL 2016. На точность прогноза указывает сравнение прогнозируемого ряда с фактическими данными. Предсказанные значения гравиметрических данных не выходят за пределы доверительных интервалов. Если начать прогнозирование со слишком ранней даты, то построенный прогноз может отличаться от составленного на основе всех статистических данных. Если в данных прослеживаются сезонные тенденции, то рекомендуется начинать прогнозирование с даты, предшествующей последней точке статистических данных. Для исследования динамики геофизических событий можно применить пространственные модели и модели временных рядов. Пространственная модель описывает совокупность геофизических параметров в данный момент времени. Временной ряд — это ряд регулярных наблюдений над некоторым параметром в последовательные моменты времени или на промежутках времени. Модель временного ряда применяется: для выявления статистической связи между частотой и глубиной возникновения землетрясений, а также статистической зависимости этих данных от гравиметрических вариаций; определения закономерностей в изменении уровней ряда и построения его модели в целях прогнозирования и исследования взаимосвязей между геофизическими явлениями.

Ключевые слова: SPSS, временной ряд, гравиметрические данные.

Введение. Наблюдения за гравиметрическими полями показывают, что периоды вариаций гравиметрических полей разбросаны хаотично на оси времени. По графику гравиметрических данных нельзя определенно говорить о закономерности в продолжительности периодов вариаций. Цель данного исследования — анализ структуры ряда формальными методами для по-

иска статистической закономерности в вариациях гравиметрических данных во времени. Для исследования динамики гравиметрических данных можно применить пространственные модели и модели временных рядов. Пространственная модель описывает совокупность гравиметрических данных в данный момент времени. Временной ряд — это ряд регулярных наблю-

дений над некоторым параметром в последовательные моменты времени или на промежутках времени. В настоящей статье модель временного ряда применяется для исследования структуры гравиметрических данных — выявления статистической связи между частотой и глубиной возникновения землетрясений и гравиметрическими данными, а также закономерностей в изменении уровней ряда и построения его модели для прогнозирования и исследования взаимосвязей между гравиметрическими данными. Теория временных рядов позволяет решать следующие основные задачи: определение природы ряда и его основных параметров; предсказание будущих значений временного ряда по имеющимся данным. Составляющими временных рядов являются: тренд, циклическая компонента, сезонная компонента и случайная составляющая. Временной ряд состоит из детерминированной и случайной частей. Тренд, циклическая и сезонная компоненты образуют детерминированную часть, которая используется для предсказания будущих значений ряда. Сезонная компонента используется для обозначения неслучайной функции, которая формируется на основе периодически повторяющихся в определенное время года колебаний исследуемого ряда. В нашем случае предполагается, что форма взаимодействия перечисленных компонент может быть мультипликативной (1), смешанной (2, 3) и аддитивной (4) [Андерсон, 1971; Хеннан, 1974; Бриллинджер, 1980; Кендалл, 1981]:

$$x_t = m_t c_t s_t \varepsilon_t, \quad (1)$$

$$x_t = m_t c_t + s_t + \varepsilon_t, \quad (2)$$

$$x_t = m_t c_t s_t + \varepsilon_t, \quad (3)$$

$$x_t = m_t + c_t + s_t + \varepsilon_t, \quad (4)$$

где m_t — тренд (устойчивая долговременная тенденция); c_t — циклическая составляющая (достаточно длинные нерегулярные колебания с периодом больше одного года); s_t — сезонная составляющая (достаточно регулярные периодические колебания с периодом, не превышающим одного года); ε_t —

случайная составляющая (погрешность). Ряд $g_t = m_t + c_t$, состоящий из суммы тренда и циклической компоненты, называется циклической компонентой (конъюнктурным циклом).

Наличие тренда означает присутствие долговременной систематической составляющей во временном ряде, описывающей основную тенденцию динамики временного ряда. Тренд как гладкая кривая отражает влияние долговременных непериодических факторов. Формально под трендом понимается последовательность условных средних значений, обусловленных временной переменной. Периодические колебания можно подразделять на сезонные колебания, вызванные климатическими условиями, у которых период не превышает одного года, и циклические с периодом колебаний несколько лет. Периодичность означает, что $x_{t+kp} = x_t$; $k = 1, 2$, и является сезонной при $p = 1$ год, циклической при $p = 2, 3$ года. Случайный шум затрудняет обнаружение регулярных компонент, однако методы исследования временных рядов включают различные алгоритмы фильтрации шума, позволяющие более четко определить регулярную составляющую. Теория временных рядов может быть полезной в исследовании временного режима геофизических событий, указывать на периоды затишья, также на периоды умеренной и повышенной активности возникновения сейсмических событий. Трендом называют неслучайную функцию, формируемую под действием общих либо долговременных тенденций. Циклическая компонента — это тоже неслучайная функция. Анализу подлежат данные, зарегистрированные на гравиметрах 115 и 139. Предварительным этапом статистической обработки должен быть этап проверки однородности выборки, т. е. неизменности вероятностных характеристик. Мерой однородности статистической совокупности данных является коэффициент вариации [Крамер, 1975]:

$$V = \frac{\sigma}{\bar{a}} 100 \%, = 4,89 \%,$$

где $\sigma = 0,0139254$ — среднеквадратическое

отклонение, $\bar{a} = -0,284456$ — среднее арифметическое значение выборки (табл. 1). Выборка считается однородной, если коэффициент вариации не превышает 33 %. Выборочное значение ряда гравиметрических данных $V_q = 5 \%$ для гравиметра 115, $V_q = 4 \%$ для гравиметра 139, что значительно меньше порога 33 % и, соответственно, подтверждает однородность временного ряда гравиметрических данных.

Гравиметрические данные используются во многих сферах деятельности человека как научных, так и практических. Например, на связи формы поверхности Земли с элементами гравитационного поля основан физический метод определения фигуры нашей планеты. Решение редуцированной задачи высшей геодезии, определение геодезических координат, высот практически невозможно без знания величин уклонения отвеса и аномалий высот, которые определяются по гравиметрическим данным. Результаты измерения силы тяжести на поверхность Земли, фиксация ее изменения во времени позволяют изучать внутреннее строение Земли, определять ее физические параметры. Все явления, происходящие на поверхности Земли, связаны с перемещением масс внутри нее. Их можно фиксировать при постоянном мониторинге гравитационного поля в тектонически активных областях и тем самым прогнозировать возможные землетрясения

или извержения вулканов. Повышение точности расчета орбит искусственных спутников Земли увеличивает значимость гравиметрических данных, которые используются для целей автономной навигации. Незнание особенностей гравитационного поля порождает погрешности в показаниях инерциальных навигационных систем и определении координат точек земной поверхности. Строительство уникальных объектов народного хозяйства, таких как ускорители частиц, антенны радиотелескопов, телебашни, прокладка протяженных подземных коммуникаций, требующих выполнения монтажных работ, не возможны без учета неоднородностей гравитационного поля, с которыми связано положение ответных линий.

Гравиметр является многопрофильным прибором, который применяется во многих областях. Однако на практике большинство из них используется в разведочных работах. Из-за этого сами гравиметры стали прочно ассоциироваться именно с гравиразведкой. Изначально в гравиразведке и гравиметрии для точных измерений поля силы тяжести применялись различные градиентометры и вариометры. Данные приборы позволяют измерить полные значения вторых производных потенциала, что достаточно информативно, но они имеют крайне низкую производительность — на одно измерение силы тяжести может ухо-

Т а б л и ц а 1 . Описательные статистики

Гравиметр	Число наблюдений	Минимум	Максимум	Среднее значение	Стандартное отклонение	Асимметрия		Экссесс	
						Статистика	Стандартная погрешность	Статистика	Стандартная погрешность
Гравиметр 115	130	-0,331	-0,232	-0,28285	0,011485	-0,513	0,212	5,609	0,422
Гравиметр 139	130	-0,319	0,281	-0,26877	0,069453	7,758	0,212	60,240	0,422
Допустимое число (по списку)	130	—	—	—	—	—	—	—	—

дуть до 40 мин и более. Гравиметр представляет собой тонкий измерительный прибор, который зависит от ряда мешающих факторов: температуры, давления, вибраций (всевозможные микросейсмы и тряска). Поэтому измерения проводят исключительно в покое, устанавливая гравиметр по очереди на каждом пикете заранее подготовленной сети. Чувствительная часть гравиметра помещена в защитный корпус, в котором поддерживаются постоянные температура и давление. Гравиметры могут быть установлены на автомобилях, судах, самолетах, космических кораблях, но более распространены конструкции, переносимые вручную.

Измеренные показания могут быть абсолютными (измеряется сама величина силы тяжести) и относительными (в этом случае измеряется разность силы тяжести в двух соседних точках). Также измерения могут проводиться в движении, но чаще гравиметры неподвижно устанавливают на пункте измерения и лишь после окончания измерений переносят на новый пункт. При выполнении полевой гравиметрической съемки создается сеть гравиметрических пунктов. При необходимости приведения наблюдений к единому абсолютному уровню значений силы тяжести полевая сеть привязывается к государственной гравиметрической сети, которая является основой для выполнения гравиметрических исследований, с целью изучения гравитационного поля и фигуры Земли, их изменений во времени, а также метрологического обеспечения гравиметрической съемки. Государственная гравиметрическая сеть представляет собой совокупность закрепленных на местности и гравиметрически связанных между собой пунктов с известными значениями силы тяжести, координат и высоты. Измерение абсолютного значения силы тяжести основаны на том, что период колебаний T у маятника зависит от величины поля, в котором совершаются эти колебания. Основная научная задача, которая решается с помощью фундаментальной сети, — изучение изменений гравитационного поля во времени. Для этого на фундамен-

тальных пунктах, размещаемых в различных геолого-тектонических регионах, систематически выполняются абсолютные и относительные определения силы тяжести с максимально возможной на данный момент точностью. Гравиметрические съемки являются одним из важных видов гравиметрических исследований.

В зависимости от поставленной задачи относительная точность измерений силы тяжести при выполнении различных видов гравиметрических съемок может изменяться в пределах от 10^{-5} до 10^{-9} мГал. Например, для общей характеристики гравитационного поля обширных территорий земной поверхности (несколько миллионов км²) измерения силы тяжести достаточно выполнять с точностью до 5^{-10} мГал. Наиболее точные измерения (до 1^{-2} мкГал) проводятся при изучении суточных флуктуаций силы тяжести. Для решения большинства практических задач гравиметрии и геодезии необходимо измерять силу тяжести с точностью 1^{-5} мкГал. В то же время точность определения опорных пунктов гравиметрической съемки характеризуется ошибкой порядка 0,1 мГал.

Потенциал силы тяжести — это сумма потенциала тяготения всех масс Земли и потенциала центробежной силы, вызванного суточным вращением Земли. Если известен потенциал, то можно найти любой другой элемент гравитационного поля. Дифференцирование потенциала по направлению позволяет найти проекцию силы тяжести на это направление. Дифференцирование по нормали к земной поверхности потенциала дает полную величину силы тяжести. Вторые производные потенциала — градиенты силы тяжести и градиенты кривизны земной поверхности. Когда геодезист перед измерениями на пункте устанавливает инструмент по уровням, он неизбежно использует натуральную (естественную) систему координат: вертикальная ось его инструмента совмещена с отвесной линией — единственной доступной координатной линией, а ось уровня является касательной к земной поверхности потенциала силы тяжести. Направление от-

весной линии можно найти из астрономических определений — его задают астрономические широта и долгота. Геометрические задачи (вычисление расстояний, азимутов) в натуральной системе координат не решают, поскольку направление отвесной линии от пункта к пункту изменяется нерегулярно. Поле силы тяжести сложное, потому земные поверхности имеют сложную форму, а силовые линии — двоякую кривизну. Важно, что превышение в поле силы тяжести выражается через разность потенциалов.

По графику рядов (рис. 1) видно, что данные гравиметра 139 более компактны, наблюдается стационарность. Описательные статистики (см. табл. 1) содержат ряд описательных статистических процедур. Частота является средством детального описания данных. С этой процедуры начинается первичный анализ временного ряда. Полученные первичные распределения дают представление о частоте встречаемости анализируемых переменных. Таблицы частот пригодны для суммирования и отражения данных. Описательные статистики дают описание средних, квадратичного отклонения, дисперсии и других параметров статистик для нормального распределения, а также минимального значения, размаха и суммы для асимметричного распределения количественной переменной.

Первичный анализ дает возможность описания подмножеств наблюдений с помощью разнообразных статистик (подсчет частот, процентов, средних и др.) и графи-

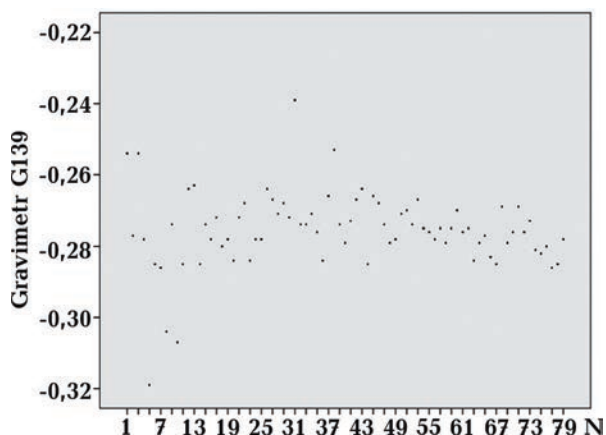


Рис. 1. График рядов.

Fig. 1. Time series plot.

ков. Между гравиметрическими данными имеет место статистически корреляционная зависимость $r_{115, 139} = -0,356$ (табл. 2), что означает правдоподобие гипотезы H_0 об отсутствии корреляционной зависимости между гравиметрическими данными на уровне вероятности значимости 0,356.

Между гравиметрическими данными и временем имеет место статистически значимая корреляционная зависимость $r_{115, n} = 0,054$; $r_{139, n} = 0,903$. Вероятность значимости 0,327 превышает уровень значимости 0,1.

Экссесс (Kurtosis) является мерой «сглаженности» распределения. В данном случае значение эксцесса составляет 3,912 для гравиметра 115 и 4,761 для гравиметра 139, асимметрия в ряде гравиметрических данных для гравиметра 115 составляет $-0,055$ и

Т а б л и ц а 2 . Корреляция

		Глубина	Магнитуда
		Корреляция Пирсона	1
Глубина	Число (2-стороннее)	0,701	0,408
	№	130	119
Магнитуда	Корреляция Пирсона	0,077*	1
	Число (2-стороннее)	0,408	—
	№	119	186

* Корреляция значима на 1 %-ном уровне.

-0,551 для гравиметра 139 (см. табл. 2). Положительный эксцесс указывает на «плосковершинное» распределение, у которого максимум вероятности выражен не столь ярко, как у нормального. Значения эксцесса, превышающие 5,0, свидетельствуют о том, что по краям распределения находится больше значений, чем вокруг среднего. Отрицательный эксцесс, напротив, характеризует «островершинное» распределение, график которого более вытянут по вертикальной оси, чем график нормального распределения. Считается, что распределение с эксцессом в диапазоне от -1 до +1 примерно соответствует нормальному виду. В большинстве случаев вполне допустимо считать нормальным распределение с эксцессом по модулю, не превосходящим 2. В большинстве случаев за нормальное принимается распределение с асимметрией, лежащей в пределах от -1 до +1.

Асимметрия указывает, в какую сторону относительно среднего сдвинуто большинство значений распределения. Нулевое значение асимметрии означает симметричность распределения относительно среднего значения. Положительная асимметрия указывает на сдвиг распределения в сторону меньших значений, а отрицательная асимметрия — в сторону больших. В ис-

следованиях, не требующих высокой точности результатов, нормальным считают распределение с асимметрией по модулю, не превышающей 2.

В SPSS (Statistical Package for the Social Sciences — «статистический пакет для общественных наук») заложена возможность вычисления множества существующих коэффициентов корреляции. Из них самый простой и один из часто применяемых — это коэффициент линейной корреляции Пирсона, а также разновидности коэффициентов ранговой корреляции Спирмена и Кендалла.

Автокорреляция. Если верно предположение о случайном характере колебаний уровней в исследуемом ряду, то не должно быть связи между уровнями. Альтернативное предположение допускает наличие зависимости между последовательными уровнями, т. е. в этом случае временной ряд не случаен. Для оценки связи между последовательными уровнями ряда может быть использована автокорреляционная функция ряда, например гравиметры 115 и 139, для которых значения автокорреляций не выходят за пределы 95 % доверительных интервалов (рис. 2). Следовательно, гравиметрические данные гравиметров 115 и 139 являются «белым шумом».

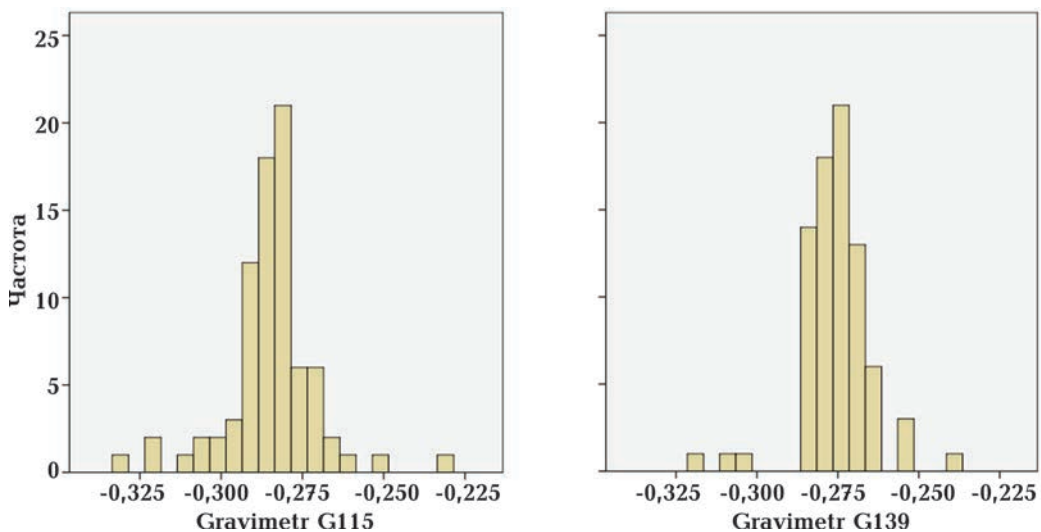


Рис. 2. Гистограмма гравиметрических данных.

Fig. 2. Histogram of gravity data.

Цель анализа временных рядов — определение модели реализации ряда. Корреляционный анализ позволяет выявить структуру ряда, т. е. определить присутствие в ряде той или иной периодической компоненты заранее неизвестной частоты.

Автокорреляционная функция определяется формулой [Бриллинджер, 1980]

$$ACF(\tau) = \frac{\sum_{k=\tau+1}^n (y_k - \bar{y})(y_{k-\tau} - \bar{y})}{\sum_{k=1}^n (y_k - \bar{y})^2}, \quad (5)$$

где $\tau = 1, 2, n - 1$ — величина сдвига, называемая лагом, определяет порядок коэффициента корреляции. Хотя автокорреляционная функция определена только для стационарных процессов, вычислить ее можно для любых рядов и далее анализировать природу ряда. Так, если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию. Если в ряду ни

один из коэффициентов автокорреляции не является значимым (рис. 3), то это может означать: ряд не содержит тенденции и циклических колебаний и имеет случайную структуру — «белый шум»; ряд содержит сильную нейтральную тенденцию, для выявления которой необходимо провести дополнительный анализ. Для определения вида и порядка процессов, порождающих стационарный временной ряд, используется аппарат автокорреляционных функций: обычной (АКФ) и частной (ЧАКФ).

Если временной ряд близок к «белому шуму», то коррелограмма осциллирует близко к горизонтальной оси (см. рис. 3), а ее значения близки к 0. При больших значениях τ оценка $ACF(\tau)$ для коэффициента автокорреляции содержит ошибки. Это связано с частичным суммированием, так как из суммирования отбрасываются наблюдения. Поэтому коррелограмма при больших значениях τ не отражает истинную структуру ряда. Для стационарного ряда $ACF(\tau)$ быстро убывает с увеличением τ . При наличии тренда автокорреляционная функция приобретает вид медленно спадающей кривой. В случае сезонной периодичности в графике АКФ присутствуют пики для лагов, кратных периоду сезонно-

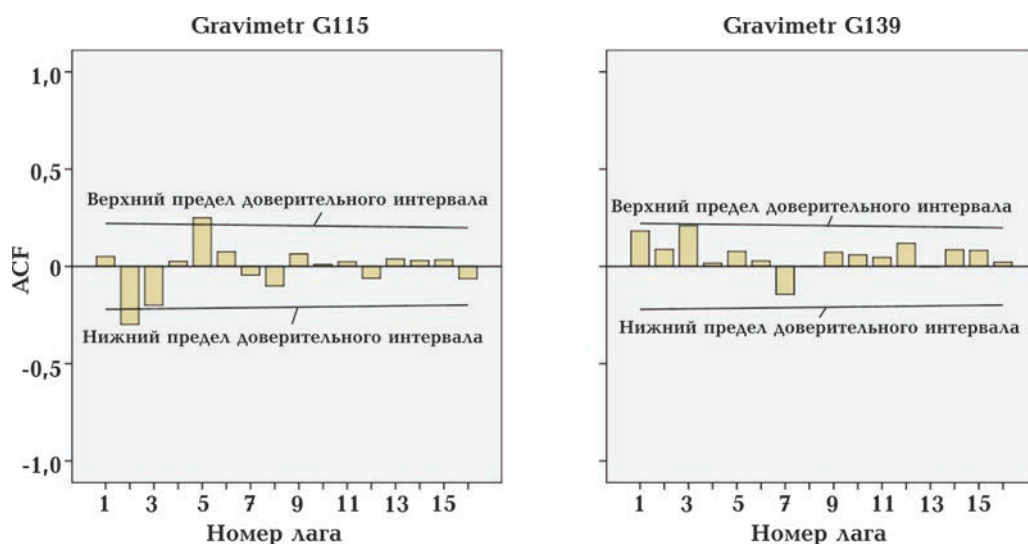


Рис. 3. Автокорреляция ряда и частная автокорреляция.

Fig. 3. Autocorrelation of a series and partial autocorrelation.

сти. Однако эти пики могут быть скрыты присутствием тренда или большой дисперсией случайной компоненты. Для заданного уровня значимости α (по умолчанию $\alpha = 5\%$) можно вычислить границы доверительного интервала, в котором нахождение значения функции автокорреляции при заданном лаге τ с вероятностью $1 - \alpha$ не противоречит предположению об отсутствии корреляции сечений с этим лагом (см. рис. 3). При графическом изображении функции автокорреляции или ее оценок эти интервалы дают две граничные кривые (выше и ниже основного графика). Выход за эти граничные кривые рассматривается как указание на значимость корреляции с соответствующим лагом. В данном случае за граничные кривые в АКФ выходят только 2-й и 4-й уровни, а в ЧАКФ — только 2-й уровень. Для ряда, состоящего из уровней гравиметра 139, значения автокорреляции и частной автокорреляции не выходят за пределы доверительного интервала.

Порядок *AR* (авторегрессии) определяется по поведению ЧАКФ, порядок *MA* (Метод скользящего среднего) — по поведению АКФ (см. рис. 3). При наличии закономерности в АКФ в модель включается только *AR*, а при наличии закономерности в ЧАКФ — только *MA*. Что касается порядков, то они определяются номером лага, на котором происходит выход за границу доверительного интервала. Кроме того, наличие статистически значимого коэффициента автокорреляции указывает, что ряд не случаен и между последовательными наблюдениями существует определенная связь. Значимый отрицательный коэффициент автокорреляции свидетельствует о высокочастотных осцилляциях уровней. Линейные коэффициенты автокорреляции характеризуют тесноту только линейной связи текущего и предыдущих уровней ряда. Поэтому по коэффициентам автокорреляции можно судить только о наличии или отсутствии линейной зависимости между уровнями ряда. Для проверки ряда на наличие нелинейной тенденции вычисляются линейные коэффициенты автокорреляции для временного ряда, состоящего из логариф-

мов исходных уровней. Отличные от нуля значения коэффициентов автокорреляции свидетельствуют о наличии нелинейной тенденции.

Другой полезный метод исследования периодичности состоит в исследовании ЧАКФ (см. рис. 3), представляющей собой углубление понятия обычной автокорреляционной функции. В ЧАКФ устраняется зависимость между промежуточными наблюдениями. Другими словами, частная автокорреляция на данном лаге аналогична обычной автокорреляции, за исключением того, что при вычислении из нее удаляется влияние автокорреляций с меньшими лагами [Андерсон, 1971; Бриллинджер, 1980]. На лаге 1 (когда нет промежуточных лагов) частная автокорреляция равна обычной автокорреляции. Как отмечалось, периодическая составляющая для данного лага k может быть удалена взятием разности соответствующего порядка. Это означает, что из каждого i -го элемента ряда вычитается $(i - k)$ -й элемент. Имеются два аргумента в пользу рассмотренных преобразований:

- во-первых, таким образом можно оп-ределить скрытые периодические составляющие ряда. Напомним, что автокорреляции на последовательных лагах зависимы. Поэтому удаление некоторых автокорреляций изменит другие автокорреляции, которые, возможно, подавляли их, и сделает некоторые другие сезонные составляющие более заметными;

- во-вторых, удаление сезонных составляющих преобразовывает временной ряд в стационарный, что необходимо для применения *ARIMA* (Autoregressive Integrated Moving Average) и других методов, например спектрального анализа. Тестовым критерием проверки значимости коэффициентов автокорреляции является критерий Бокса—Пирса [Шанченко, 2008]:

$$Q = n \sum_{k=1}^m r_k ,$$

где r_k — коэффициент автокорреляции с лагом k ; m — наибольший лаг; n — длина

ряда. Выборочная статистика критерия Бокса—Пирса является χ^2 распределенной случайной величиной с m степенями свободы.

Критерием адекватности модели временному ряду является неразличимость ряда остатков от процесса «белого шума» [Андерсон, 1971; Бокс, Дженкинс, 1974; Бриллинджер, 1980; Вербик, 2008]. Статистика Льюнга—Бокса, вычисленная для проверки значимости автокорреляций, меньше критического значения, соответствующего вероятности значимости $\alpha = 0,05$. Более глубокий анализ периодических колебаний с целью обнаружения скрытых периодичностей проводится с помощью спектрального анализа временного ряда.

Коэффициент Дарбина—Уотсона показывает автокорреляцию, которая принимает значения из интервала (0,4) [Сеньо, 2007]. Значения, близкие к 0, указывают на сильную положительную, близкие к 4 — на сильную отрицательную автокорреляцию, а близкие к 2 — на ее отсутствие. В данной задаче значение коэффициента Дарбина—Уотсона составляет $d = 1,876$, т. е. выполняется условие $1,5 < d < 2,5$ и, следовательно, автокорреляция отсутствует [Сиденко и др., 2011]. Согласно критерию Фишера, регрессионная модель оказалась незначимой для всех включенных в пять моделей регрессоров, так как вероятность значимости выборочной статистики F -критерия, примененной для проверки нулевой гипотезы H_0 о незначимости уравнения регрессии, равна 0,562. Для

прогнозирования временного ряда применен EXCEL 2016 (рис. 4). Если начать прогнозирование перед последней точкой, можно получить оценку точности прогноза, сравнив прогнозируемый ряд с фактическими данными. Но если начать прогнозирование со слишком ранней даты, построенный прогноз может отличаться от составленного прогноза на основе всех статистических данных. При этом прогноз будет более точным. Если в данных прослеживаются сезонные тенденции, то рекомендуется начинать прогнозирование с даты, предшествующей последней точке статистических данных.

Тест Колмогорова—Смирнова. Основой теста является расчет максимальной разности между кумулятивными частотами выборки и нормальной функцией распределения. Эта разность обозначается величиной z , на основании которой вычисляется вероятность значимости. По критерию Колмогорова—Смирнова нулевая гипотеза H_0 о соответствии нормальному распределению гравиметрических данных подтверждается [SPSS Trends 14.0, 2006; Сеньо, 2007; Bernhardt, 2007]. Одним из показателей нормальности распределения выборки является эксцесс и асимметрия.

Сводка модели и оценки параметров. Рассматриваются регрессии: линейная, логарифмическая, обратная, квадратическая, кубическая, экспоненциальная. Выбор оптимальных моделей временных рядов: R-квадрат для стационарной части [Spezialvorlesung ..., 2006; SPSS Trends 4.0, 2006];

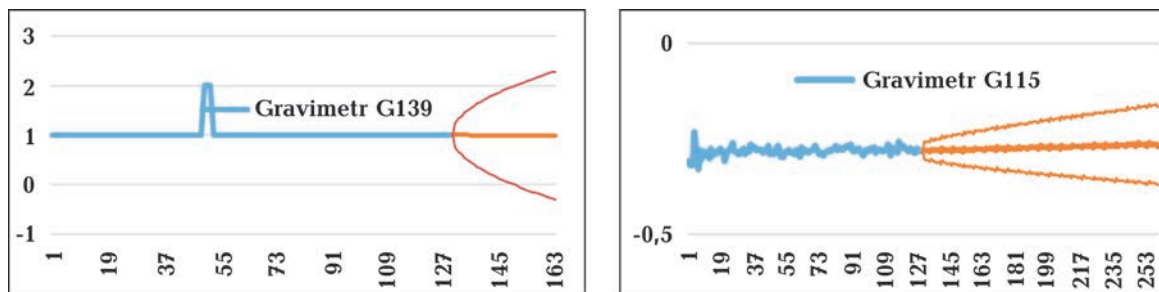


Рис. 4. Прогноз временного ряда.

Fig. 4. Time series forecast.

показатель сравнения стационарной части модели (без тренда) с простой моделью среднего значения. Эта мера предпочтительнее обычного R-квадрата при наличии тренда или сезонной компоненты. Критерий R-квадрат для стационарной части может принимать и отрицательные значения, которые означают, что рассматриваемая модель хуже, чем простая модель, а положительные значения означают обратное, т. е. рассматриваемая модель лучше, чем простая. Отрицательное значение критерия R-квадрат можно получить и для уравнений, которые не содержат свободный член. В этом случае значение критерия R-квадрат не может быть истолковано как квадрат корреляции. Такие ситуации показывают, что постоянный член должен присутствовать в модели [Володин, 2006].

Метод экспоненциального сглаживания и прогнозирования временных рядов. Теория временных рядов располагает множеством методов предсказания значений уровня ряда, которые реализуют схему экстраполяции, т. е. исследуется ряд и предполагается, что свойства его в дальнейшем не будут изменяться.

Одним из распространенных методов — метод простого экспоненциального сглаживания. Несмотря на простоту использованного математического аппарата, прогнозный потенциал метода не уступает методам, где применяются более глубокие математические методы экстраполяции. Метод экспоненциального сглаживания относится к непараметрическим методам анализа временных рядов, так как его применение не зависит от вида распределения случайной компоненты. Метод экспоненциального сглаживания дает возможность получить оценку параметров тренда [Spezial vorlesung ..., 2006; Backhaus et al., 2011], характеризующих не средний уровень ряда, а тенденцию, сложившуюся к моменту последнего наблюдения. Часто для прогнозирования нестационарных временных рядов применяется самая распространенная модель простого экспоненциального сглаживания:

$$S_t = \alpha x_t + \beta S_{t-1},$$

где S_t — значения сглаженного x_t исходного ряда в момент t ; α — параметр сглаживания, $0 < \alpha < 1$; $\beta = 1 - \alpha$. Эта формула применяется рекурсивно — каждое новое значение вычисляется как среднее значение данного наблюдения (которое является также прогнозом) и сглаженного ряда.

Спектральный анализ временного ряда. Спектральный анализ применяется для определения периодической компоненты при известной длине периода. По сути — это линейная регрессия, где зависимой переменной являются уровни ряда, а регрессорами — функции синусов и косинусов. Спектральный анализ определяет корреляцию регрессоров различной частоты с наблюдаемыми данными. Известна теорема [Бокс, Дженкинс, 1974; Колмогоров, Фомин, 1976], согласно которой среди всех тригонометрических многочленов порядка n наименьшее среднеквадратичное отклонение имеет многочлен, искомыми коэффициентами которого являются коэффициенты Фурье. На применении одномерных рядов Фурье основан один из методов моделирования сезонных и циклических колебаний. Ряды Фурье — эвристический алгоритм, является одной из разновидностей спектрального анализа. С помощью спектрального анализа в структуре временного ряда определяется пик отклонений от тренда, что позволяет рассчитать длительность периодической компоненты ряда (рис. 5). При применении спектрального анализа случайный стационарный процесс представляется в виде суммы гармонических колебаний различных частот, называемых гармониками. Спектр описывает распределение амплитуд случайного стационарного процесса по различным частотам. Исследование частотной структуры ряда выполняется процедурой «Спектральный анализ» пакета SPSS. Как известно, почти любую периодическую функцию можно аппроксимировать рядом Фурье, суммой синусов и косинусов [Пискунов, 1964; Смирнов, Дунин-Барковский, 1969; Бокс, Дженкинс, 1974; Spezial vorlesung ..., 2006; Backhaus et al., 2011]: $x_t = a \cos(\omega t) + b \sin(\omega t)$, $0 \leq \omega \leq \pi$.

При моделировании временного ряда

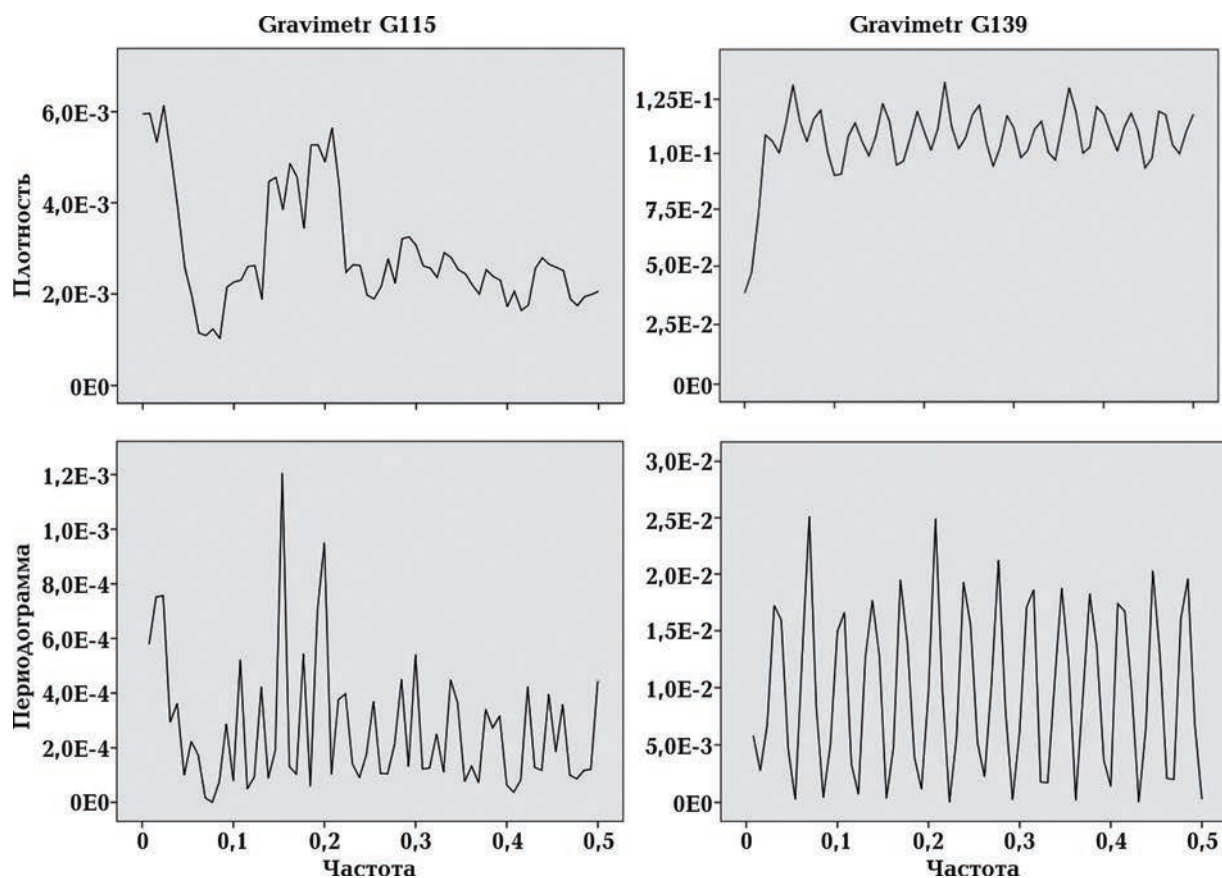


Рис. 5. Периодограмма и спектральная плотность временных рядов.

Fig. 5. Periodogram and spectral density of time series.

суммой синусов и косинусов синусоидальные периодические компоненты появляются на периодограмме в форме отдельных вершин, а несинусоидальные — в форме ряда равноотстоящих вершин разной высоты. Вершина, соответствующая наименьшей частоте, указывает на частоту периодической компоненты во временном ряду [Spezial vorlesung ..., 2006]. Возникает вопрос, влияют ли эти различия на значения функции спектральной плотности? Особый интерес представляют низкочастотные колебания, так как долговременные тенденции в динамике ряда содержатся в области низких частот. Чем лучше модель трендовой составляющей, тем больше колебаний он содержит. Согласно теореме Дирихле [Смирнов, Дунин-Барковский, 1969; Цапаева, 2011], непрерывная функция с периодом 2π однозначно определяется своим рядом Фурье:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nt) + b_n \sin(nt).$$

Применение данного способа требует знание частотной структуры ряда. Для моделирования нестационарных временных рядов применяется модель авторегрессии проинтегрированного скользящего среднего (ARIMA). Данная модель является одним из методов оценки неизвестных параметров и прогнозирования временных рядов. При моделировании временного ряда суммой синусов и косинусов синусоидальные периодические компоненты появляются на периодограмме в форме отдельных вершин (см. рис. 3), а несинусоидальные — в форме ряда равноотстоящих вершин разной высоты. Вершина, соответствующая наименьшей частоте, указывает на частоту пе-

риодической компоненты во временном ряду [Spezial vorlesung ..., 2006].

Описательные статистики. Характеристики центральной тенденции, которые описывают положение распределения и разброса, выводятся по умолчанию. Они отражают степень различия значений исследуемых данных. Характеристики центральной тенденции включают среднее значение, медиану и 5 %-ное усеченное среднее, стандартную ошибку, дисперсию, стандартное отклонение, минимальное и максимальное значения переменных, размах и межквартильный размах. Описательные статистики включают также характеристики формы распределения, такие как асимметрия и эксцесс, которые выводятся вместе со своими стандартными ошибками. Выводится также 95 %-ный доверительный интервал для среднего, можно задать иное значение доверительного уровня.

Выводы. В настоящей статье модель временного ряда применяется для исследования структуры геофизических событий: выявление статистической связи между частотой и глубиной возникновения землетрясений и гравиметрическими данными; закономерностей в изменении уровней ряда и построения его модели в целях прогнозирования и исследования взаимосвязей между явлениями.

Активность геофизических процессов указывает, что периоды вариаций геофизических процессов разбросаны хаотично на оси времени. По их графику нельзя оп-

ределенно утверждать о закономерности в продолжительности периодов вариаций и в чередовании периодов сейсмического затишья с периодом высокой сейсмической активности. Стимулом для проведения данного исследования было желание анализировать структуру ряда формальными методами для поиска статистической закономерности в вариациях геофизических параметров во времени. Для исследования динамики геофизических событий были задействованы модели временных рядов. Прогнозирование осуществлено с помощью EXCEL 2016 (см. рис. 5). На точность прогноза указывает сравнение прогнозируемого ряда с фактическими данными. Предсказанные значения гравиметрических данных не выходят за пределы доверительных интервалов. Если начать прогнозирование со слишком ранней даты, построенный прогноз может отличаться от составленного прогноза на основе всех статистических данных. При этом прогноз будет более точным. Если в данных прослеживаются сезонные тенденции, то рекомендуется начинать прогнозирование с даты, предшествующей последней точке статистических данных.

Благодарность. Авторы выражают искреннюю благодарность за редактирование и полезные замечания Петру Павловичу Павлову, научному сотруднику Лаборатории сейсмологии Института геологии и сейсмологии.

Список литературы

- Андерсон Т. Статистический анализ временных рядов. Мир: Москва, 1971. 746 с.
- Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Москва: Мир, 1974. 406 с.
- Бриллинджер Д. Временные ряды. Обработка данных и теория. Москва: Мир, 1980. 532 с.
- Вербик М. Путеводитель по современной эконометрике. Москва: Научная книга, 2008, 616 с.
- Володин И.Н. Лекции по теории вероятностей и математической статистике. Казань, 2006. 270 с.
- Кендалл М. Временные ряды. Москва: Финансы и статистика, 1981. 198 с.
- Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. Москва: Наука, 1976. 542 с.

- Крамер Г. Математические методы статистики. Москва: Мир, 1975. 648 с.
- Пискунов Н.С. Дифференциальное и интегральное исчисление. Москва: Наука, 1964. 312 с.
- Сеньо П.С. Теорія ймовірностей та математична статистика. Київ: Знання, 2007. 558 с.
- Сиденко А.В., Вишняков В.В., Исаев С.М. Теория статистики: учебник. Москва: МАКС Пресс, 2011. 376 с.
- Смирнов И.В., Дунин-Барковский Н.В. Курс теории вероятностей и математической статистики. Москва: Наука, 1969. 511 с.
- Хеннан Э. Многомерные временные ряды. Москва: Мир, 1974. 575 с.
- Цапаева С.А. Ряды Фурье. Великий Новгород, 2011.
- Шанченко Н.И. Лекции по эконометрике. Ульяновск: Изд. УЛГТУ, 2008. 139 с.
- Backhaus, K., Erichson, B., Plinke, W., Schuchard-Ficher, C., & Weiber, R. (2011). *Multivariate Analyse methoden* (pp. 120—154). Berlin-Heidelberg: Springer Verlag.
- Bernhardt, C. (2007). *Modellierung von Elektrizitätspreisen durch lineare Zeitreihenmodelle und Value-at-Risk-Schätzung mittels Methode-naus der Extremwerttheorie Technische Universität München*. Zentrum Mathematik. München, 97 p.
- Spezial vorlesung Zeitreihenanalyse-MitBeispielen in Mathematica Institut fur Stochastik, Johannes Kepler Universität Linz. (2006). Linz, 277 p.
- SPSS Trends 14.0. (2006). Chicago, 165 p.

Analysis of time series by the example of registration of variations in the gravitational field

R.Z. Burtiev¹, Yu. V. Semenova³, V.T. Kiriya², E.V. Sidorenko², S.V. Troyan¹, V.Yu. Kardanets¹, D.I. Nuka², 2021

¹Institute of Geology and Seismology of the Academy of Sciences of Moldova, Chisinau, Moldova

²Technical University of Moldova, Chisinau, Moldova

³S.I. Subbotin Institute of Geophysics of the National Academy of Sciences of Ukraine, Kiev, Ukraine

In this work, a time series model is used to study the structure of gravimetric data series to identify patterns in the change in the levels of the series and model in order to predict and study the relationships between the levels of gravimetric data. Observations of the activity of geophysical processes showed that the periods of variations in geophysical processes are scattered chaotically on the time axis. According to their schedule, it is impossible to definitely speak about the regularity in the duration of the periods of variations, and in the alternation of periods of seismic calm with a period of high seismic activity. The impetus for this study was the desire to analyze the structure of a number of formal methods to search for statistical patterns in the variations of geophysical parameters over time. Time series models were used to study the dynamics of geophysical events. Forecasting was carried out using the SPSS 20 package and EXCEL 2016. The accuracy of the forecast is indicated by the comparison of the forecast series with the actual data. The predicted values of the gravimetric data are within the confidence intervals. If you start forecasting too early, the forecast may differ from the forecast based on all statistical data. If the data show seasonal trends, it is recommended to start forecasting from the date before the last point of the statistical data. Spatial and time series models can be used to

study the dynamics of geophysical events. A spatial model describes a set of geophysical parameters at a given point in time. A time series is a series of regular observations of a certain parameter at successive points in time or at intervals of time. In this work, the time series model is used: to identify the statistical relationship between the frequency and depth of occurrence of earthquakes, as well as to identify the statistical dependence of these data on gravimetric variations; determination of patterns in the change in the levels of the series and the construction of its model in order to predict and study the relationships between geophysical phenomena.

Key words: SPSS, time series, gravity data.

References

- Anderson, T. (1971). *Statistical Analysis of Time Series*. Mir: Moscow, 746 p. (in Russian).
- Box, J., & Jenkins, G. (1974). *Time Series Analysis. Forecast and management*. Moscow: Mir, 406 (in Russian).
- Brillinger, D. (1980). *Time series. Data processing and theory*. Moscow: Mir, 532 p. (in Russian).
- Verbik, M. (2008). *A guide to modern econometrics*. Moscow: Nauchnaya kniga, 616 p. (in Russian).
- Volodin, I.N. (2006). *Lectures on probability theory and mathematical statistics*. Kazan, 270 p. (in Russian).
- Kendall, M. (1981). *Time series*. Moscow: Finansy i statistika, 198 p. (in Russian).
- Kolmogorov, A.N., & Fomin, S.V. (1976). *Elements of the theory of functions and functional analysis*. Moscow: Nauka, 542 p. (in Russian).
- Kramer, G. (1975). *Mathematical Methods of Statistics*. Moscow: Mir, 648 p. (in Russian).
- Piskunov, N.S. (1964). *Differential and integral calculus*. Moscow: Nauka, 312 p. (in Russian).
- Senyo, P.S. (2007). *Probability theory and mathematical statistics*. Kyiv: Znannya, 558 p. (in Ukrainian).
- Sidenko, A.V., Vishnyakov, V.V., & Isaev, S.M. (2011). *The theory of statistics: a textbook*. Moscow: MAKS Press, 376 p. (in Russian).
- Smirnov, I.V. & Dunin-Barkovskiy, N.V. (1969). *Course in Probability Theory and Mathematical Statistics*. Moscow: Nauka, 511 p. (in Russian).
- Hennan, E. (1974). *Multidimensional time series*. Moscow: Mir, 575 p. (in Russian).
- Tsapaeva, S.A. (2011). *Fourier series*. Veliky Novgorod (in Russian).
- Shanchenko, N.I. (2008). *Lectures on econometrics*. Ulyanovsk: Ed. UISTU, 139 p. (in Russian).
- Backhaus, K., Erichson, B., Plinke, W., Schuchard-Fischer, C., & Weiber, R. (2011). *Multivariate Analyse methoden* (pp. 120—154). Berlin-Heidelberg: Springer Verlag.
- Bernhardt, C. (2007). *Modellierung von Elektrizitätspreisen durch lineare Zeitreihenmodelle und Value-at-Risk-Schätzung mittels Methode-naus der Extremwerttheorie Technische Universität München*. Zentrum Mathematik. München, 97 p.
- Spezial vorlesung Zeitreihenanalyse-MitBeispielen in Mathematica Institut fur Stochastik, Johannes Kepler Universität Linz. (2006). Linz, 277 p.
- SPSS Trends 14.0. (2006). Chicago, 165 p.

Аналіз часових рядів на прикладі реєстрації варіацій гравітаційного поля

***Р.З. Буртієв¹, Ю.В. Семенова³, В.Т. Кіріяк², О.В. Сидоренко²,
С.В. Тройан¹, В.Ю. Карданець¹, Д.В. Нука², 2021***

¹Інститут геології і сейсмології АН Молдови, Кишинів, Молдова

²Технічний університет Молдови, Кишинів, Молдова

³Інститут геофізики ім. С.І. Субботіна НАН України, Київ, Україна

У статті використано модель тимчасового ряду для дослідження структури рядів гравіметричних даних. Спостереження за активністю геофізичних процесів показали, що періоди варіацій геофізичних процесів розкидані хаотично на осі часу. За їх графіком можна впевнено говорити про закономірності в тривалості періодів варіацій і в чергуванні періодів сейсмічного затишшя та високої сейсмічної активності. Структуру ряду проаналізовано формальними методами для пошуку статистичної закономірності у варіаціях геофізичних параметрів у часі. Для дослідження динаміки геофізичних подій задіяно моделі часових рядів. Прогнозування здійснено за допомогою пакета SPSS 20 і програми EXCEL 2016. На точність прогнозу вказує порівняння прогнозованого ряду з фактичними даними. Передбачені значення гравіметричних даних не виходять за межі довірчих інтервалів. Якщо почати прогнозування із занадто ранньої дати, то побудований прогноз може відрізнятись від складеного на основі всіх статистичних даних. Якщо в даних простежуються сезонні тенденції, то рекомендується починати прогнозування з дати, що передує останній точці статистичних даних. Для дослідження динаміки геофізичних подій можна застосувати просторові моделі та моделі часових рядів. Просторова модель описує сукупність геофізичних параметрів у певний момент часу. Часовий ряд — це ряд регулярних спостережень над деяким параметром у послідовні моменти часу або на проміжках часу. У статті модель тимчасового ряду застосовано: для виявлення статистичного зв'язку між частотою і глибиною виникнення землетрусів, а також статистичної залежності цих даних від гравіметричних варіацій; визначення закономірностей у зміні рівнів ряду і побудови його моделі з метою прогнозування та дослідження взаємозв'язків між геофізичними явищами.

Ключові слова: SPSS, часовий ряд, гравіметричні дані.