УДК 550.348.098.42 DOI: https://doi.org/10.24028/gzh.0203-3100.v42i5.2020.215080

Модель главных компонент в макросейсмике

Р.З. Буртиев, В.Ю. Карданец, 2020

Институт геологии и сейсмологии АН Республики Молдова, Кишинев, Молдова
Поступила 2 марта 2020 г.

Формирование сейсмических процессов обусловлено сложными, многообразными геолого-геофизическими процессами, происходящими в недрах Земли. Сейсмические процессы характеризуются множеством параметров, а результаты наблюдений над ними представляют в виде многомерных случайных величин. При исследовании таких многопараметрических процессов встает вопрос: нельзя ли отбросить часть параметров или заменить их меньшим числом каких-то функций от них, сохранив при этом всю информацию? Для решения данной задачи служит факторный анализ, который основан на определении минимального числа факторов, составляющих наибольшую долю в дисперсии данных. В исследовании сложной природы сейсмичности факторный анализ помогает глубже понять сущность сейсмических процессов, поскольку взаимозависимость между сейсмическими параметрами должна быть обусловлена связями между параметрами, выявление которых является задачей факторного анализа.

Для того чтобы регрессионный анализ, основанный на обычном методе наименьших квадратов, давал наилучшие результаты, случайная ошибка должна удовлетворять условиям Гаусса—Маркова: математическое ожидание случайной ошибки в любом наблюдении должно быть равно нулю, т.е. она не должна иметь систематического смещения. Обычно если уравнение регрессии включает свободный член, то это значит, что условие выполнено автоматически, так как роль константы состоит в определении любой систематической тенденции объясняемой переменной, включенной в уравнение регрессии. Мультиколлинеарность означает высокую взаимную коррелированность объясняющих переменных регрессии. Отсутствие высокой коллинеарности регрессоров — одно из условий применения метода наименьших квадратов для оценки параметров многомерной линейной регрессии. Для оценки значений коэффициентов функции затухания, при наличии мультиколлинеарности, применяется регрессионный анализ на главных компонентах, где сильно коррелированные регрессоры заменяются компонентами F_1, F_2, F_3, F_4, выявленными моделью главных компонент факторного анализа, между которыми корреляционная связь отсутствует.

Ключевые слова: модель главных компонент, вероятностный анализ сейсмической опасности (ВАСО), SPSS-статистический пакет для социальных наук.

Введение. Изучение сейсмичности связано с выявлением причинно-следственных связей в эндогенных процессах. Для исследования силы связей между причинными и следственными процессами может быть полезен регрессионный (факторный) анализ, предполагающий выполнение условий теоремы Гаусса—Маркова и позволяющий выявить тенденции в массовых явлениях.

В исследовании сложной природы сейсмичности факторный анализ оказался полезным для понимания сущности сейсмических процессов. Между параметрами, отражающими механизм и геометрию очага Вранча, наблюдается статистически значимая корреляционная зависимость. Анализ позволяет выявить латентные факторы, обусловливающие связи между этими параметрами. Факторный анализ приводит к потере исходных данных информации, однако значительное уменьшение количества параметров оправдывает его применение и помогает выявить закономерности в сейсмических процессах, которые не поддаются непосредственному наблюдению.

В данной задаче значения параметров заданы в разных единицах измерения, поэтому для оценки степени влияния параметров на интенсивность сейсмических толчков используются стандартизованные коэффициенты в уравнении регрессии, т. е. стандартизованный вариант уравнения регрессии (вместо значений всех параметров используются их стандартизованные значения). При этом изменяется только форма записи уравнения регрессии, а коэффициенты корреляции между всеми параметрами, коэффициент детерминации, значимость коэффициентов регрессии не изменяются. По уровню значимой вероятности — величины, применяемой при статистической проверке гипотез (в данном случае используется критерий *t*-статистики Стьюдента), проверяется нулевая гипотеза о незначимости коэффициента регрессии.

Многомерная линейная регрессия и мультиколлинеарность. Мультиколлинеарность означает высокую взаимную коррелированность объясняющих переменных регрессии. Отсутствие высокой коллинеарности регрессоров является одним из условий применения метода наименьших квадратов для оценки параметров многомерной линейной регрессии.

Считается, что два регрессора x и z коллинеарны, если парный коэффициент корреляции r_{xz}>0,7 [Габриелян, 2008; Мхитарян и др., 2008]. Отсутствие мультиколлинеарности и гомоскедастичности, постоянство дисперсии остатков регрессии — необходимые условия для получения несмещенных, эффективных и состоятельных оценок коэффициентов регрессии. Проверка модели на коллинеарность регрессоров определяется по значениям фактора инфляции дисперсии VIF (Variance Inflation Factor):

$$\mathrm{VIF}_j = \frac{1}{1-R_j^2}, \qquad (1)$$

где R_j^2 — коэффициент множественной детерминации:

$$R_j^2 = \frac{\sum_{i=1}^{n}\left(\hat{x}_{ji}-\bar{x}_j\right)^2}{\sum_{i=1}^{n}\left(x_{ji}-\bar{x}_j\right)^2}. \qquad (2)$$

Здесь n — число наблюдений; x_{ji} — наблюденные значения, $\hat{x}_{ji}$ — расчетные значения, $\bar{x}_j$ — среднее арифметическое значение зависимой переменной в регрессии переменной x_j на остальные переменные.

В практических приложениях рекомендуется, чтобы число наблюдений n превышало число независимых переменных m не менее, чем в три раза [Лешинский и др., 2003; Габриелян, 2008]. Высокие значения величины VIF_j (1), превышающие 4 хотя бы для одного регрессора X, означают мультиколлинеарность. Наиболее детальным показателем проблем, связанных с мультиколлинеарностью, является коэффициент увеличения дисперсии, определяемый для каждой переменной, как в формуле (4).

Общность параметра i равна сумме квадратов ее нагрузок по всем факторам:

$$\mathrm{Com}_i^2 = \sum_{i=1}^{M} a_{ik}^2, \qquad (3)$$

где a_{ik} — нагрузка i-го параметра на k-й фактор, Com_i^2 — коэффициент общности множественной детерминации в регрессии x_j на прочие x, т. е.

$$x_j = c_0 + c_1x_1 + \ldots + c_{j-1}x_{j-1} + \\ + c_{j+1}x_{j+1} + \ldots + c_kx_k + \varepsilon,\ j = 1,\ldots,m. \qquad (4)$$

Общность означает долю дисперсии в данном параметре, которая обусловлена совокупным воздействием всех факторов. Это значение должно превышать 0,5, иначе соответствующие параметры будут удалены. Другими словами, общность указывает на надежность данного параметра. Например, значение 1 означает, что дисперсия параметра полностью определяется выделяемым фактором. Извлечение — доля дисперсии, которая принимает значения из интервала 0—1 и объясняется всеми параметрами, оставшимися после извлечения факторов. Общность является квадратом множественной корреляции параметра, которая использует факторы в качестве независимых параметров. Извлеченные факторы объясняют свыше 97,6 % дисперсии параметра P_{pl} (табл. 1). Общность определяет для каких параметров факторный анализ работает лучше или хуже. Если параметр обладает низкой общностью (например, Azim 13,9 %), то для этого параметра применение факторной модели не имеет смысла и его следует удалить из модели.

Т а б л и ц а 1. Общность

Параметр	Начальное	Извлеченное
M_w	1	0,805
ln$R$	1	0,858
Azim	1	0,139
$NP1_{stk}$	1	0,932
$NP1_{dp}$	1	0,965
$NP1_{stip}$	1	0,897
$NP2_{stk}$	1	0,931
$NP2_{dp}$	1	0,967
$NP2_{stip}$	1	0,919
P_{az}	1	0,951
P_{pl}	1	0,976
B_{az}	1	0,789
B_{pl}	1	0,827
T_{az}	1	0,871
T_{pl}	1	0,923

Мера выборочной адекватности Кайзера—Мейера—Олкина (КМО) используется для проверки гипотезы о том, что частные корреляции между переменными малы. Выборочное значение индекса КМО 0,57 указывает на значимую корреляционную связь между всеми переменными. Критерий Бартлетта применяется для проверки гипотезы H_0 о том, что в выборке отсутствует зависимость между переменными, при альтернативной гипотезе H_1 о наличии взаимосвязи. Для исследуемой выборки

Т а б л и ц а 2. КМО и критерий сферичности Бартлетта

Мера выборочной адекватности Кайзера—Мейера—Олкина		0,570
Критерий сферичности Бартлетта	Приблизительный	277018,98
	Степень свободы	120
	Значимость	0,000

Т а б л и ц а 3. Матрица компонентов

Параметры	Компоненты				Повернутая матрица компонентов			
	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
P_{az}	,951	–,138	–,004	–,163	,910	,150	,237	–,208
$NP2_{stk}$	–,894	,047	,356	,043	–,932	–,093	–,157	–,166
$NP2_{dp}$	–,859	,436	,087	,184	–,934	,125	–,081	,271
$NP1_{dp}$	,858	–,427	–,173	–,139	,951	–,156	,075	–,180
B_{az}	–,795	–,217	–,131	–,289	–,548	–,234	–,651	–,042
M_w	,709	,222	,274	,438	,392	,178	,796	,002
$NP1_{stk}$	,706	,376	–,535	,013	,711	,352	,127	,529
T_{az}	–,541	,486	–,402	,429	–,549	,026	,014	,756
P_{pl}	–,043	–,918	,159	,308	,072	–,895	,095	–,387
T_{pl}	,214	,752	,221	–,511	,032	,955	–,034	–,086
$NP1_{slip}$	,610	,708	,119	–,099	,345	,787	,381	,115
$NP2_{slip}$	–,095	,403	,775	–,374	–,371	,639	,129	–,591
B_{pl}	–,161	,513	–,650	,333	–,116	,124	–,013	,891
MSK	,257	–,052	,471	,412	–,002	–,095	,626	–,243
ln$R$	–,285	–,250	–,456	–,745	,114	–,009	–,944	–,037
Azim	,176	,069	,069	,206	,074	,012	,274	,047

Таблица 4. Таблица факторов

F_1	F_2	F_3	F_4
P_{az}	P_{pl}	$NP1_{stk}$	$\ln R$
$NP2_{stk}$	T_{pl}	$NP2_{slip}$	Azim
$NP2_{dp}$	$NP2_{slip}$	MSK	
$NP1_{dp}$		B_{az}	
B_{az}			
M_w			
$NP1_{stk}$			
T_{az}			

рассчитанная статистика при 120 степенях свободы χ^2=277018,989 (табл. 2). Это позволяет принять альтернативную гипотезу о значимости всей корреляционной матрицы на 5 %-ном уровне значимости, что также подтверждает целесообразность проведения факторного анализа.

О мультиколлинеарности свидетельствует значение VIF от 4 и выше хотя бы для одного индекса j. Характеристикой качества регрессионной модели служит коэффициент детерминации R-квадрат (2), который является отношением дисперсии расчетной части к дисперсии случайной части. Регрессионная модель тем лучше, чем большая часть рассеивания объясняется изменением объясняющей составляющей регрессора (4).

Для оценки значений коэффициентов функции затухания при наличии мультиколлинеарности применяется регрессионный анализ на главных компонентах, где сильно коррелированные регрессоры заменяются компонентами F_1, F_2, F_3, F_4, выявленными моделью главных компонент факторного анализа, между которыми корреляционная связь отсутствует (табл. 3, 4).

Если значения параметров измерены в несопоставимых величинах, их следует стандартизовать.

Регрессионная модель на главных компонентах имеет вид

$$\begin{cases} z_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1\mathrm{p}}x_\mathrm{p}, \\ z_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2\mathrm{p}}x_\mathrm{p}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ z_p = a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pp}x_p, \end{cases} \qquad (5)$$

где z_i — i-я главная компонента, a_{ij} — j-я факторная нагрузка. Необходимо построить регрессионную модель интенсивности сейсмических толчков на параметрах сейсмичности.

Обозначим параметры как x_1, x_2, …, x_{16}, где x_1=M_w, $x_2 = \ln\sqrt{d^2 + h^2}$, x_3=$NP1_{stk}$ и т. д., а компоненты — как z. В методе главных компонент первая главная компонента является линейной комбинацией параметров, обладающих наибольшей дисперсией. Далее для каждой следующей компоненты дисперсия убывает [Сиденко и др., 2011].

Метод множественной регрессии применяется для построения модели главных компонент на основе их значений:

$$x_{ji} = a_{1i}z_{j1} + a_{2i}z_{j2} + a_{3i}z_{j3} + \ldots + a_{ki}z_{jk}\,. \quad (6)$$

Чтобы получить уравнение регрессии при исходных параметрах, в уравнение подставляются значения главных компонент $z_{1,2}, \ldots, z_k$, выраженные через стандартизованные параметры x_1, x_2,…,x_k:

$$I = c_0 + c_1x_1 + c_2x_2 + c_3x_3 + \ldots + c_nx_n + \varepsilon\,. \quad (7)$$

С помощью последнего уравнения (7) проводилось определение коэффициентов затухания интенсивности сейсмических толчков по мере удаления от источника землетрясения. Однако выборочные статистики тестовых критериев проверки качества регрессии свидетельствуют о неудовлетворительной регрессионной модели на главных компонентах. Регрессия на выделенные главные факторы оказалась слабой, низкое значение коэффициента детерминации R^2=0,018 (2), хотя по критерию Фишера гипотеза о незначимости регрессии отвергается согласно значимой регрессионной зависимости значения интенсивности от главных компонент. Таким образом, модель затухания строилась на параметрах землетрясений.

Как было отмечено, аппроксимация эллипсом макросейсмического поля является упрощенной моделью реальной картины, которая мозаична, а линии, разделяющие зоны одинаковой балльности, —

извилисты. Поэтому оценка региональных коэффициентов затухания проводилась по схеме, основанной на макросейсмических наблюдениях румынских землетрясений (10.11.1940, 4.03.1977, 31.08.1986, 30.05.1990, 31.05.1990) [Kronrod et al., 2012]:

- начало координат совмещается с эпицентром исследуемых землетрясений в каждом скользящем азимутальном створе шириной 40° с шагом дискретности 20°;
- определяются значения коэффициентов уравнения затухания с помощью процедуры множественного регрессионного анализа из пакета SPSS-20;
- методом наименьших квадратов подбирается линия регрессии, при которой общая сумма квадратов отклонений (Residuals) зависимой переменной от наблюденных была бы наименьшей;
- при проведении регрессионного анализа переменные, которые объявляются независимыми, могут быть коррелированы между собой, т. е. имеет место мультиколлинеарность, что приводит к некорректности регрессионного анализа.

В множественном регрессионном анализе существует три способа включения независимых переменных в уравнение регрессии:

1) стандартный — включаются все переменные;

2) обратная пошаговая регрессия — из рассмотрения последовательно исключаются наименее значимые переменные, если статистический показатель значимости переменной F-критерия меньше вероятности значимости α=0,05 (либо α=0,01), и строится регрессия по оставшимся переменным. На начальном этапе в регрессию включаются все переменные;

3) прямая пошаговая регрессия действует в противоположном направлении. Вначале в уравнение включается регрессор, который имеет наибольшее значение коэффициента корреляции с зависимой переменной, и с помощью тестовых критериев проверяется адекватность регрессии. При значимости модели включается следующий регрессор и т. д. Если значимость какого-то регрессора меньше критической величины, то регрессор исключается, если больше, то сохраняются, т. е. регрессоры,

Т а б л и ц а 5. Корреляционная матрица

Параметры	MSK	M_w	ln$R$	Azim	$NP1_{stk}$	$NP1_{dp}$	$NP1_{slip}$	$NP2_{stk}$	$NP2_{dp}$	$NP2_{slip}$	P_{az}	P_{pl}	B_{az}	B_{pl}	T_{az}	T_{pl}
MSK	1,000	,366	–,458	,010	–,016	,114	,114	–,089	–,137	,093	,163	,146	–,261	–,191	–,186	–,037
M_w	,366	1,000	–,704	,192	,430	,367	,559	–,513	–,370	,066	,582	–,043	–,713	–,003	–,270	,147
ln$R$	–,458	–,704	1,000	–,194	–,066	,029	–,334	,046	–,030	–,175	–,118	–,070	,554	–,016	–,116	,047
Azim	,010	,192	–,194	1,000	,082	,082	,103	–,108	–,083	–,005	,120	–,008	–,174	,013	–,041	,035
$NP1_{stk}$	–,016	,430	–,066	,082	1,000	,543	,665	–,820	–,488	–,329	,580	–,469	–,506	,379	,034	,278
$NP1_{dp}$	,114	,367	,029	,082	,543	1,000	,223	–,830	–,993	–,336	,884	,277	–,535	–,315	–,623	–,101
$NP1_{slip}$	,114	,559	–,334	,103	,665	,223	1,000	–,447	–,229	,357	,477	–,672	–,620	,092	–,040	,674
$NP2_{stk}$	–,089	–,513	,046	–,108	–,820	–,830	–,447	1,000	,796	,323	–,862	,065	,599	–,098	,432	–,082
$NP2_{dp}$	–,137	–,370	–,030	–,083	–,488	–,993	–,229	,796	1,000	,268	–,895	–,283	,545	,397	,677	,056
$NP2_{slip}$	,093	,066	–,175	–,005	–,329	–,336	,357	,323	,268	1,000	–,093	–,327	–,006	–,354	–,250	,595
P_{az}	,163	,582	–,118	,120	,580	,884	,477	–,862	–,895	–,093	1,000	,036	–,688	–,232	–,672	,196
P_{pl}	,146	–,043	–,070	–,008	–,469	,277	–,672	,065	–,283	–,327	,036	1,000	,079	–,457	–,356	–,845
B_{az}	–,261	–,713	,554	–,174	–,506	–,535	–,620	,599	,545	–,006	–,688	,079	1,000	,032	,169	–,242
B_{pl}	–,191	–,003	–,016	,013	,379	–,315	,092	–,098	,397	–,354	–,232	–,457	,032	1,000	,625	,048
T_{az}	–,186	–,270	–,116	–,041	,034	–,623	–,040	,432	,677	–,250	–,672	–,356	,169	,625	1,000	–,053
T_{pl}	–,037	,147	,047	,035	,278	–,101	,674	–,082	,056	,595	,196	–,845	–,242	,048	–,053	1,000

которые связаны с зависимой переменной с наибольшей частной корреляцией, включаются в уравнение регрессии.

В данном случае комбинируются прямая и обратная пошаговые регрессии, где регрессоры последовательно включаются в уравнение и выключаются из него. Метод множественной регрессии позволяет определить значимость линейной связи между параметрами и интенсивностью сейсмических толчков, качество аппроксимации данных уравнением регрессии, пригодность вычисленных значений коэффициентов уравнения в целях наилучшего прогноза, а также значимость параметров для предсказания силы сейсмических толчков. Из корреляционной матрицы всех выборочных данных видно (табл. 5), что заметная корреляционная зависимость прослеживается между параметрами $NP1_{stk}$ и $NP2_{stk}$ (r=–0,820), между интенсивностью MSK и магнитудой (r=0,366) и между интенсивностью MSK и гипоцентральным расстоянием (r=–0,458).

Первое представление о зависимых параметрах можно получить по корреляционной матрице, которая характеризует степень корреляционной связи между параметрами исходного массива. Чем выше доля высоких корреляций, тем лучше данные подходят для факторного анализа.

Также наблюдается функциональная и сильная зависимость между некоторыми физическими параметрами очагов, вероятно обусловленными эндогенными процессами, природу которых предстоит выяснять (табл. 5). Чтобы охарактеризовать степень корреляции в словесной форме, используется шкала классификации тесноты связи Чеддока (табл. 6).

Из табл. 7 видно, что методом пошаговой регрессии в регрессионное уравнение поочередно включались параметры Constant, ln$R$, M_w, $NP2_{slip}$, T_{pl}, Azim, начиная с константы и логарифма гипоцентрального расстояния. Приведены значения коэффициента детерминации для каждой модели, который принимает наибольшее значение R^2=0,823. Для модели, где в уравнение включены регрессоры Constant, ln$R$, M_w, $NP2_{slip}$, T_{pl}, Azim и, это означает, что уравнением регрессии объясняется около 82,3 % вариации интенсивности сейсмических воздействий. Критерий Дарбина—Уотсона проверяет автокорреляцию, которая принимает значения из интервала (0,4) [Сеньо, 2007]. Значения, близкие к 0, указывают на сильную положительную автокорреляцию, близкие к 4 — на сильную отрицательную, а близкие к 2 — на отсутствие автокорреляции. На практике, если для критерий Дарбина—Уотсона выполняется условие 1,5<d<2,5, автокорреляция отсутствует [Сиденко и др., 2011].

Таблица 6. Шкала классификации тесноты связи Чеддока

Количественная мера	0—0,1	0,11—0,3	0,31—0,5	0,51—0,7	0,71—0,9	0,91—0,99	0,991—1
Характер связи	Отсутствует	Слабая	Умеренная	Заметная	Тесная	Сильная	Функциональная

Таблица 7. Сводная таблица модели

Модель	R	Коэффициент детерминации R-квадрат	Скорректированный R-квадрат	Стандартная ошибка оценки	Измененная статистика					Критерий Дарбина—Уотсона
					R-квадрат измененный	F измененное	df_1	df_2	Значение F измененное	
1	.836	.699	.699	.715	.699	2421,724	1	1043	.000	
2	.886	.784	.784	.606	.085	411,105	1	1042	.000	
3	.897	.804	.804	.577	.020	106,931	1	1041	.000	
4	.903	.816	.815	.560	.012	66,228	1	1040	.000	
5	.907	.823	.822	.549	.007	41,741	1	1039	.000	1,158

Наблюдаемое значение составляет 1,158. Следовательно, можно говорить о наличии положительной автокорреляции. Согласно критерию Фишера, регрессионная модель оказалась значимой 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. Наилучшей по статистическим характеристикам оказалась модель, в которую включены пять параметров: Constant, $\ln R$, M_w, $NP2_{slip}$, T_{pl}, Azim. В этом случае коэффициент детерминации является наибольшим: R^2=0,823. Выборочная функция распределения остатков регрессии (рис. 1, *а*) почти совпадает с теоретической функцией распределения, что свидетельствует в пользу модели с пятью регрессорами.

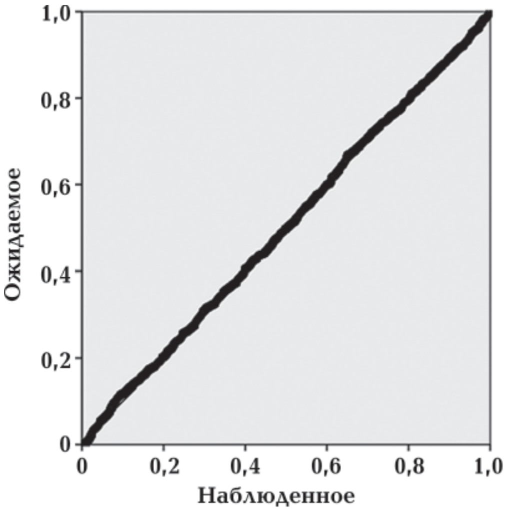
Однако тестовые критерии, характеризующие адекватность регрессионной модели, указывают на достаточную статистическую значимость модели затухания, где включены Constant, магнитуда M_w и натуральный логарифм гипоцентрального расстояния $\ln R$. Включение в модель других параметров не привело к существенному улучшению качества модели затухания, и такие модели тяжело поддаются статистической обработке. Поэтому было принято решение включить в регрессионное уравнение затухания магнитуду и гипоцентральное расстояние, как это принято в сейсмологии.

В данной задаче значения параметров заданы в разных единицах измерения, поэтому для оценки степени влияния параметров на интенсивность сейсмических толчков используются стандартизованные коэффициенты в уравнении регрессии, т. е. стандартизованный вариант уравнения регрессии (вместо значений всех параметров — их стандартизованные значения). При этом изменяется только форма записи уравнения регрессии, а коэффициенты корреляции между всеми параметрами, коэффициент детерминации, значимость коэффициентов регрессии не изменяются. По величине вероятности значимости t-статистики проверяется нулевая гипотеза о незначимости коэффициента регрессии. Из сводной табл. 7 видно, что в третьей, четвертой и пятой моделях коэффициент Constant статистически незначим — вероятность значимости превышает 5 %,уровень значимости α=0,05. Значения коэффициента VIF меньше 4 (табл. 8). Следовательно, мультиколлинеарность отсутствует. Один из показателей коллинеарности параметра — его толерантность (допуск) когда выборочные значения параметра значительно превышают пороговую величину 0,2. Таким образом, соответствующий параметр не является линейной комбинацией других регрессоров. Одним из показателей, определяющих адекватность регрессионной модели, служит нормальность распределения остатков, на которую указывают незначительно уклоняющиеся от функции плотности распределения и функции распределения нормального закона (рис. 1, *а*, *б*) гистограмма и выборочная функция распределения остатков.

Выше было, аппроксимация эллипсом макросейсмического поля упрощенная модель реальной картины, которая является мозаичной, а линии, разделяющие зоны равной балльности, — извилистыми. Для построения более реальной модели затухания интенсивности сейсмических толчков по мере удаления от источника землетрясения предлагается следующая схема:

- начало координат совмещается с эпицентром исследуемых землетрясений;
- в каждом скользящем азимутальном створе шириной 40° с шагом дискретности 20° определяются значения коэффициентов уравнения макросейсмического поля с помощью регрессионного анализа.

Регрессионная модель интенсивности сейсмических воздействий строилась в пунктах-баллах: на магнитуду M_w и логарифм гипоцентрального расстояния $\ln R$ и на все параметры (M_w, $\ln R$, Azim, $NP1_{stk}$, $NP1_{dp}$, $NP1_{slip}$, $NP2_{stk}$, $NP2_{dp}$, $NP2_{slip}$, P_{az}, P_{pl}, B_{az}, B_{pl}, T_{az}, T_{pl}). Тестовые критерии проверки значимости регрессии указывают на



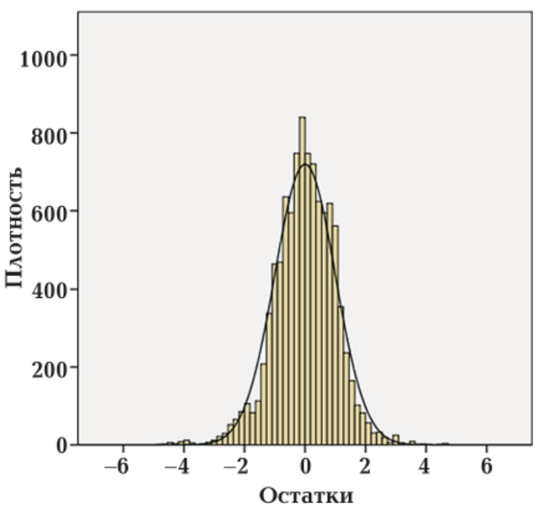


Рис. 1. Выборочная функция распределения остатков регрессии (*а*), функция плотности распределения остатков и теоретическая функция плотности распределения нормального закона (*б*).

Fig. 1. The sample distribution function of the regression residues (*a*), the density distribution function of the residues, and the theoretical distribution density function of the normal law (*б*).

Таблица 8. Регрессионная модель интенсивности сейсмических воздействий в пунктах-баллов

Створ	Константа	Регрессоры		Коэффициент детерминации	Вероятность значимости F-статистики	Значение VIF	Критерий Дарбина—Уотсона
		M_w (*a*)	ln$R$ (*b*)				
0—40°	9,56	0,96	–4,40	0,784	0,0	1,011	1,007
20—60°	8,84	1,06	–4,36	0,805	0,0	1,003	1,014
40—80°	9,01	1,05	–4,42	0,798	0,0	1,026	0,719
60—100°	11,20	0,94	–5,10	0,767	0,0	1,005	0,691
80—120°	10,27	1,13	–5,30	0,743	0,0	1,002	0,346
100—140°	6,28	1,29	–4,01	0,430	0,0	1,37	0,379
120—160°	6,03	1,37	–4,05	0,501	0,0	1,394	0,314
140—180°	5,91	1,53	–4,39	0,774	0,0	1,748	0,403
160—200°	7,19	1,48	–4,73	0,476	0,0	1,215	0,608
180—220°	9,33	1,27	–4,94	0,533	0,0	1,228	0,652
200—240°	9,33	1,27	–5,02	0,546	0,0	1,159	0,666
220—260°	4,94	1,64	–4,60	0,495	0,0	1,124	0,659
240—280°	3,30	2,04	–5,40	0,466	0,0	1,311	0,726
260—300°	7,10	1,61	–5,88	0,618	0,0	1,321	0,897
280—320°	9,78	1,05	–5,16	0,612	0,0	1,162	0,595
300—340°	9,17	1,04	–4,72	0,605	0,0	1,016	0,596
320—360°	7,86	1,091	–1,977	0,566	0,0	1,02	0,710
340—20°	9,14	0,914	–1,742	0,697	0,0	1,022	1,014

достаточно качественную регрессию значений интенсивности сейсмических толчков на два параметра — M_w, ln$R$. Регрессия значения интенсивности на магнитуду и гипоцентральное расстояние характеризуется выборочными значениями: R^2=0,784 и

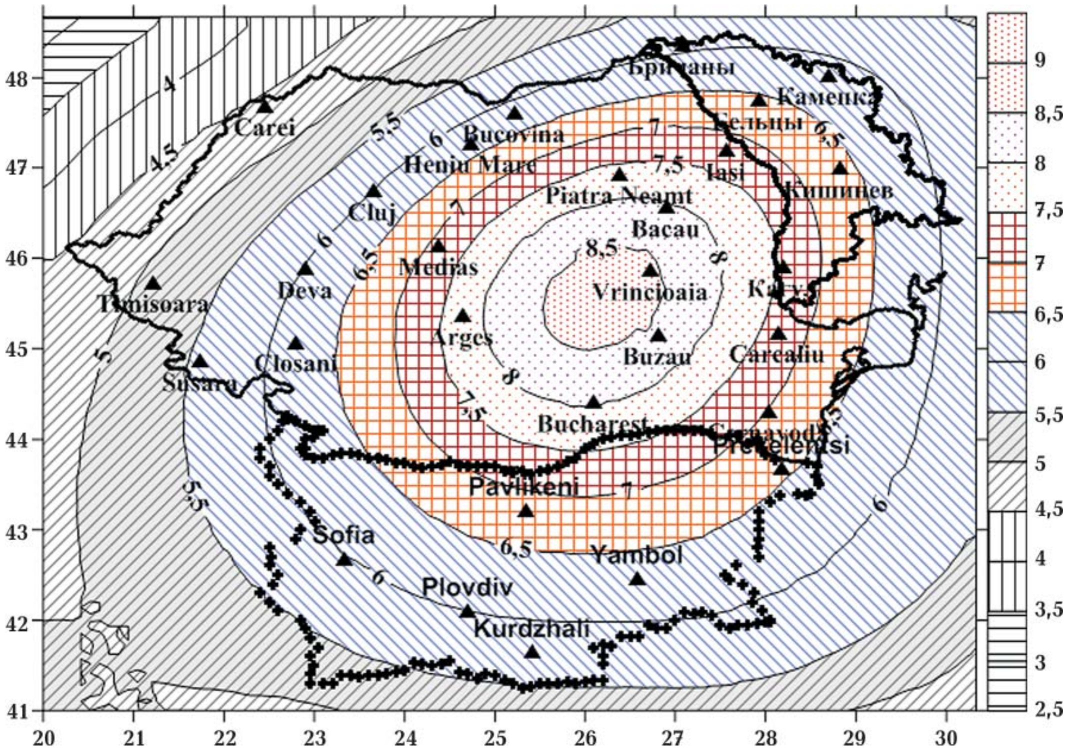


Рис. 2. Карта сейсмической опасности территорий Молдовы, Румынии и Болгарии, подверженных воздействию очага Вранча, рассчитана по значениям коэффициентов, полученных методом регрессионного анализа.

Fig. 2. A map of the seismic hazard of the territory of Moldova, Romania and Bulgaria exposed to the Vrancea focus, calculated from the values of the coefficients obtained by the method of regression analysis.

высоким значением тестового критерия Фишера, который отвергает гипотезу о незначимости регрессии в пользу ее значимости:

$$I = aM + b\sqrt{h^2 + R^2} + c\,, \qquad (8)$$

где M — магнитуда, h — глубина, R — эпицентральное расстояние. Наибольшее наблюдаемое значение критерия Дарбина—Уотсона составляет 1,014. Следовательно, можно говорить о слабой положительной автокорреляции. Значения коэффициента VIF меньше 4, т. е. мультиколлинеарность отсутствует (см. табл. 8).

Была рассчитана сейсмическая опасность по формуле (8) и построена карта сейсмического районирования территорий Молдовы, Румынии и Болгарии, подверженная воздействию очага Вранча, а по значениям коэффициентов, полученных методом регрессионного анализа, где вычислялась сейсмическая опасность (рис. 2).

Применялась разработанная в Институте геологии и сейсмологии АН Республики Молдова методология ВАСО (вероятностный анализ сейсмической опасности) для рас чета вероятности возникновения толчков определенной интенсивности. Картируемая территория покрывалась географической сеткой с шагом дискретности 0,2°. В узлах сетки рассчитывалась сейсмическая опасность по формуле [Burtiev, 2017]

$$p_{\phi,\psi}\left(t,n,m_k,I^k\right) =$$

$$= p\left(t,n\right)\frac{n!}{m_k!\left(n-m_k\right)!}\omega_k^{m_k}\left(1-\omega_k\right)^{n-m_k}\,, \qquad (9)$$

$$\tilde{\omega} = \left(\omega_1, \omega_{12}\right), \qquad (10)$$

где ω_k — вероятность возникновения в узлах сетки сейсмического толчка с интенсивностью I^k, которая состоит из вектора распределения вероятностей возникновения в узлах сети сейсмических толчков с интенсивностью I^k=1, 2, …,12 баллов шкалы MSK-64.

Выводы. В исследовании сложной природы сейсмичности факторный анализ оказался полезным для понимания сущности сейсмических процессов. Между параметрами, отражающими механизм и геометрию очага Вранча, отмечена статистически значимая корреляционная зависимость. С помощью факторного анализа выявлены латентные факторы, обусловливающие связи между этими параметрами. Факторный анализ приводит к потере в исходных данных информации, однако значительное уменьшение количества параметров оправдывает его применение и помогает выявить закономерности в сейсмических процессах, которые не поддаются непосредственному наблюдению.

Список литературы

Алказ В.Г. Основы прогноза сейсмической опасности и сейсмического риска территории Республики Молдова. Кишинев, 2007, 229 с.

Габриелян Р.Г. Эконометрика: методическое пособие. Видное, 2008. 85 с.

Лещинский О.Л., Рязанцева В.В., Юнькова О.О. Економетрія: навч. посіб. для студ. вищ. навч. закл. Київ: ДП «Вид. дім «Персонал», 2003, 208 с.

Мхитарян В.С., Архипова М.Ю., Сиротин В.П. Эконометрика: учебно-методический комплекс. Москва: Изд. центр ЕАОИ, 2008. 144 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.

Burtiev, R. (2017). Seismic hazard assessment method based on the stochastic models of seismicity. *Bulletin of the International Institute of Seismology and Earthquake Engineering, 51*, 22—38.

Principal component model in macroseismicity

R.Z. Burtiev, V.Yu. Cardanets, 2020

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

Seismic processes are complex and diverse, since their formation is caused by complex, diverse geological and geophysical processes occurring in the Earth's interior, and are characterized by a set of parameters, and the results of observations over them are presented as multidimensional random variables. When studying such multiparametric processes, the question arises: can we discard some of the parameters, or replace them with a smaller number of some functions from them, while preserving all the information? To solve this problem, we use factor analysis, which is based on determining the minimum number of factors that make up the largest share in the data variance. In the study of the complex nature of seismicity, factor analysis helps to understand better the essence of seismic processes, since the interdependence between seismic parameters must be due to the relationships between parameters, the identification of which is the task of factor analysis.

In order for the regression analysis based on the usual least squares method to give the best results, the random error must satisfy the Gauss-Markov conditions: the mathematical expectation of the random error in any observation must be zero, which means it should not have a systematic bias. Usually, if the regression equation includes a free term, then this means that the condition is satisfied automatically, since the role of the constant is to determine any systematic tendency of the explained variable included in the regression equation. Multicollinearity means a high cross-correlation of the explanatory regression variables. The lack of high collinearity of the regressors is one of the conditions for applying the least squares method to estimate the parameters of multidimensional linear regression. To assess the values of the coefficients of the attenuation function, in the presence of multicollinearity, we use regression analysis on the main components, where the strongly correlated regressors are replaced by components F_1, F_2, F_3, F_4, identified by the model of the main components of factor analysis, between which there is no correlation.

Key words: principal component model, probabilistic of seismic hazard analysis (PSHA), SPSS-statistical package for social sciences.

References

Alkaz, V.G. (2007). *Fundamentals of seismic hazard and seismic risk prediction for the territory of the Republic of Moldova*. Chisinau, 229 p. (in Russian).

Gabrielyan, R.G. (2008). *Econometrics: Methodological Gu*ide. Vidnoe, 85 p. (in Russian).

Leshchinskyy, O.L., Ryazantseva, V.V., & Yunkova, O.O. (2003). *Econometrics: a textbook for students of higher educational institutions*. Kyiv: Publishing House «Personnel», 208 p. (in Ukrainian).

Mkhitaryan, V.S., Arkhipova, M.Yu., & Sirotin, V.P. (2008). *Econometrics: educational and methodological complex*. Moscow: EAOI Publishing Center, 144 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). *Theory of statistics: textbook*. Moscow: MAKS Press, 376 p. (in Russian).

Burtiev, R. (2017). Seismic hazard assessment method based on the stochastic models of seismicity. *Bulletin of the International Institute of Seismology and Earthquake Engineering, 51*, 22—38.

Модель головных компонент у макросейсміці

Р.З. Буртієв, В.Ю. Карданець, 2020

Інституту геології і сейсмології АН Республики Молдова, Кишинів, Молдова

Формування сейсмічних процесів зумовлене складними, різноманітними геолого-геофізичними процесами, що відбуваються в надрах Землі. Сейсмічні процеси характеризуються безліччю параметрів, а результати спостережень над ними подають у вигляді багатовимірних випадкових величин. При дослідженні таких багатопараметричних процесів постає питання: Чи не можна відкинути частину параметрів або замінити їх меншим числом якихось функцій від них, зберігши при цьому всю інформацію? Для розв'язання цієї задачі слугує факторний аналіз, який заснований на визначенні мінімального числа факторів, які становлять найбільшу частку в дисперсії даних. У дослідженні складної природи сейсмічності факторний аналіз дає змогу глибше зрозуміти сутність сейсмічних процесів, оскільки взаємозалежність між сейсмічними параметрами має бути зумовленою зв'язками між параметрами, виявлення яких є задачею факторного аналізу.

Для того щоб регресійний аналіз, заснований на звичайному методі найменших квадратів, давав найліпші результати, випадкова помилка має задовольняти умови Гаусса—Маркова: математичне сподівання випадкової помилки у будь-якому спостереженні має дорівнювати нулю, тобто не має бути систематичного зсуву. Зазвичай, якщо рівняння регресії включає вільний член, то це означає, що умова виконана автоматично, оскільки роль константи полягає у визначенні будь-якої систематичної тенденції у поясненні змінної, яку включено у рівняння регресії. Мультиколінеарность означає високу взаємну корельованість у поясненні змінних регресії. Відсутність високої колінеарності регресорів є однією з умов застосування методу найменших квадратів для оцінювання параметрів багатовимірної лінійної регресії. Для оцінки значень коефіцієнтів функції згасання, при наявності мультиколінеарності, застосовано регресійний аналіз на головних компонентах, де сильно корельовані регресори замінюються компонентами F_1, F_2, F_3, F_4, виявленими моделлю головних компонент факторного аналізу, між якими кореляційний зв'язок відсутній.

Ключові слова: модель головних компонент, імовірнісний аналіз сейсмічної небезпеки (ІАСН), SPSS-статистичний пакет для соціальних наук.