

УДК 519.876:55.435.62(477.75)

В.М. Таран

Європейський університет, Ялтинська філія, Ялта

КЛАСТЕРИЗАЦІЯ ДАНИХ ДЛЯ ФОРМУВАННЯ ВИСНОВКУ ПРИ ПРОГНОЗУВАННІ ЗСУВНИХ ПРОЦЕСІВ ПІВДЕННОГО БЕРЕГА КРИМУ

Запропоновано при аналізі зсувних процесів Південного берега Криму розглядати результати спостережень в стандартному вигляді з метою привести їх до однієї розмірності, за допомогою вагових коефіцієнтів об'єднати всі вхідні фактори в одну змінну – середнє зважене, та запропоновано модель РОСЗ – регресію однофакторну середньо зважену. Встановлено, що при впорядкуванні незалежного фактору за зростанням також впорядковується і результуючий показник. Упорядковані дані розбиваються на три великих кластери, встановлено межі для вхідного фактору, при яких результуючий показник набуває одного з трьох значень: маленька кількість зсувів, середня та велика. Визначено причини виникнення розбіжностей між прогнозними і реальними даними для моделі РОСЗ – це рівень вкладених коштів в укріплення схилів. Для інших моделей наявність цієї закономірності не спостерігається. За допомогою критеріїв оцінки моделей та оцінки прогнозу проаналізовано п'ять моделей, запропонованих раніше, і модель РОСЗ, зроблено висновок, що всі моделі придатні для прогнозування.

Ключові слова: прогнозування зсувних процесів Південного берега Криму, регресійний аналіз, оцінка моделі і оцінка прогнозу, кластеризація.

Вступ

Постановка проблеми. Південний берег Криму є унікальним регіоном України не лише завдяки географічному розташуванню і клімату, а ще й тому, що на цій достатньо вузькій полосі узбережжя пролягає більше 50% всіх доріг півострова. Таким чином, зсувні процеси становлять значну загрозу, а іноді наносять велику шкоду їх експлуатації, руйнують не тільки дороги, а й будівлі і всілякі споруди.

Аналіз та прогнозування зсувних процесів Південного берега Криму – задача, яка на фоні катастрофічних природних явищ останніх років набуває все більшої актуальності, а також є складною задачею аналізу динамічних процесів довільної природи, що ускладнюється присутністю в даній системі чинника ризику непередбачених ситуацій [1]. Таким чином, постає проблема побудови моделей та оцінювання прогнозів за цими моделями задля подальшого розподілу коштів на усунення наслідків цих процесів або для завчасного укріплення зсувонебезпечних ділянок.

Аналіз літератури. У наукових дослідженнях [2 – 7], що присвячені цій проблемі, основна увага приділяється картуванню територій, геологічному та геоморфологічному дослідженню схилів та ділянок, на яких відбулися руйнівні зсуви, а прогнози базуються на інтуїції фахівця-геолога та на математичних моделях, які відтворюють схили, ковзні поверхні ґрунтової маси та залежать від суто геоморфологічних показників. Отже, в цих дослідженнях обмеженням є використання сучасних інформаційних технологій в нестабільних умовах невизначеності в еколого-економічному середовищі протікання цих процесів.

На Південному березі Криму накопичено різноманітні дані багаторічних кліматичних, геологічних, екологічних, астрономічних спостережень, але науковці пев-

ної галузі при моделюванні використовують лише «свої» дані і «свої» моделі, не пов'язуючи між собою всі спостереження. Таким чином, накопичена інформація використовується не в повному обсязі, не системно.

Мета статті – системний аналіз динаміки зсувних процесів та розробка моделей і методів, які спрощуватимуть процес вироблення і прийняття управлінських рішень. Отже, актуальною стає проблема створення системи підтримки прийняття рішень (СППР), яка б допомагала приймати ефективні управлінські рішення та використовувала не один, а декілька методів прогнозування і на їх основі будувала би узагальнений прогноз.

Моделі, що використовуються в інтелектуальній системі “Forecasting of Landslide for Southern Coast of Crimea” («FLSSCC») при прогнозуванні зсувних процесів Південного берега Криму

Фактори, які сприяють активізації зсувних процесів, – це опади, сейсмічна активність і сонячна активність [10, 11]. Опади враховувалися багатьма іншими дослідниками, але такі прогнози мали невелику достовірність. Отже, для їх покращення будемо накопичувати кількість опадів за два роки (O). В якості фактору сейсмічної активності візьмемо кількість поштовхів впродовж року з магнітудою, що перевищує 8,5К, які відбулися на відстані до 200 км від Південного берега Криму (CeA). Сонячну активність будемо вимірювати за середньою річною кількістю чисел Вольфа (CoA). Зсувний процес опишемо за допомогою кількості активних зсувів за рік (A3) [11].

Вище перелічені чинники сприяють виникненню нових зсувів та активізації старих. Таким чином, модель має містити незалежні фактори: опади, сей-

смічна активність і сонячна активність – і результуючий показник: кількість активних зсувів (на рис. 1 показано нормовані значення цих змінних).

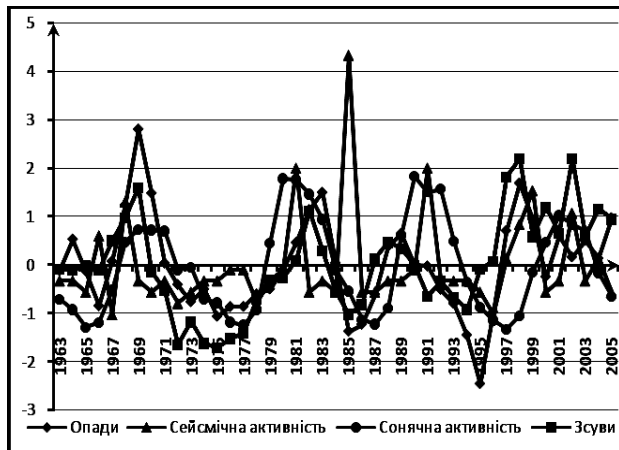


Рис. 1. Незалежні фактори і залежний показник для прогнозування зсувних процесів Південного берега Криму

Щоб показати всі дані на одному графіку, було проведено нормування всіх показників за формулою (дані спостережень за 43 роки):

$$X_{ij_норм} = x_{ij} - \bar{x}_i / \sigma_i, \quad i = \overline{1, 4}, \quad j = \overline{1, 43},$$

де x_{ij} – j -те спостереження i -го фактору; \bar{x}_i – його математичне сподівання; σ_i – його середнє квадратичне відхилення.

Наведений рисунок свідчить, що дійсно існує залежність між незалежними факторами і залежним показником. Значне підвищення активізації зсувних процесів в 1968-1969, 1982, 1988, 1997-1998, 2000-2005 роках супроводжується підвищенням майже всіх факторів, а зниження в 1972-1977, 1985-1986, 1991, 1993-1994 роках – зниженням. Не зважаючи на те, що в 1991 році спостерігався підйом сейсмічної і сонячної активності, підйому зсувної активності не відбулося. Це, мабуть, пов'язано з масовим проведенням завчасного укріплення небезпечних ділянок після активізації екзогенних процесів в 1982 і 1988 роках.

Введемо позначення: O – кількість опадів за два гідрогеологічних роки, SeA – сейсмічна активність, SoA – сонячна активність, $A3$ – кількість активних зсувів. Наведемо моделі, що використовуються в інтелектуальній системі “Forecasting of Landslide for Southern Coast of Crimea” («FLSSCC») і описують зсувні процеси Південного берега Криму: регресійна однофакторна модель відносно опадів (PO), авторегресійна (AP), комплексна множинна лагова авторегресійна модель (ЛАР), метод групового врахування аргументів (МГУА) та метод комплексування аналогів (КА) [10, 11].

Розглянемо нову модель, яка використовує, як і ЛАР, чотири вхідних даних (опадів, сейсмічну і сонячну активність та кількість зсувів за попередній рік). Проте, по-перше, всі ці фактори спочатку нормуються за наведеною вище формулою, щоб дані для роз-

рахунків були порівнянні між собою та належали проміжку $[-2; 2]$, крім декількох значень. По-друге, за допомогою обраних ваг всі фактори зводяться до однієї змінної $x_{сер_зв}$ (середнього зваженого), а потім будеться однофакторна регресія (РОЗС – регресія однофакторна середньо зважена). В табл. 1 наведено ваги для факторів, які визначаються за формулою:

$$w_{i_норм} = R_i^2 / \sum_{k=1}^4 R_k^2, \quad i = \overline{1, 4},$$

де R_i^2 – коефіцієнт детермінації i -го фактору та результуючого показнику.

Таблиця 1

Ваги вхідних факторів для моделі РОЗС

	Опади	СеА	СоА	A3(-1)	Сума
R^2	0,4219	0,0059	0,0285	0,5116	0,9679
Ваги (w_i)	0,4359	0,0061	0,0294	0,5286	1,0000

Аналогічним способом пронормуємо результуючий показник і помістимо обидві змінних на один графік, враховуючи, що вхідний фактор беремо на рік раніше, щоб можливо було отримати прогноз на наступний рік (рис. 2).



Рис. 2. Незалежний фактор, розрахований як середнє зважене, і залежний показник для прогнозування зсувних процесів Південного берега Криму для моделі РОЗС

Аналізуючи рис. 2, можна зробити висновок, що залежність існує, проте лаг, рівний одиниці, погіршує якість моделі. Такий крок виправдовує себе тим, що ми можемо моделювати наступний стан процесу, спираючись на поточні дані, не прогнозуючи їх для наступного спостереження.

Таким чином, до п'яти моделей, розглянутих в попередніх дослідженнях (РО, АР, ЛАР, КА, МГУА), слід додати модель РОЗС – регресію однофакторну середньо зважену.

Наведемо критерії вибору кращих моделей з множини оцінюваних кандидатів [9 – 11]:

- коефіцієнт множинної детермінації (R^2);
- середній квадрат похибки ($e^2_{сер}$);
- критерій Дарбіна-Уотсона для перевірки на корельованість похибок моделі (DW).

В практиці частіше використовують такі формальні статистики перевірки якості прогнозу похибки точності [9 – 11]:

- середньоквадратична похибка RMSE (root

mean squared error);

– середня абсолютна похибка MAE (mean absolute error);

– середня абсолютна відсоткова похибка MAPE (mean absolute percentage error);

– коефіцієнт нерівності Тейла U.

Таким чином, для порівняння наведемо в табл. 2 результати оцінювання всіх моделей за запропонованими критеріями. Розглядатимемо шість моделей (з урахуванням РОСЗ) та сім критеріїв.

Таблиця 2
Оцінювання запропонованих моделей

Моделі	Якість моделі			Якість прогнозу			
	R2	$\Sigma e^2 / n$	DW	RMSE	MAE	MAPE	K_{-T} Тейла U
AP(1)	0,501	1625	2,02	40,32	31,39	28,8%	0,139
PO(O)	0,422	1926	0,94	43,88	32,60	34,8%	0,151
ЛАР	0,643	1190	1,84	34,50	28,37	27,9%	0,118
КА	0,502	2144	1,53	46,89	32,26	32,9%	1,169
МГУА	0,805	648	2,30	25,46	18,98	18,9%	0,087
РОСЗ	0,536	1545	1,69	39,13	30,0	29,0%	0,135

Наведена таблиця свідчить, що незважаючи на те, що нова модель РОСЗ не є найкращою, оцінки якості моделі та її прогнозу посідають лише третє місце, тобто поступаються моделям МГУА та ЛАР; а також є кращими за такі моделі: AP, PO, КА – ця модель відповідає умовам моделювання.

Повернемося знову до моделі РОСЗ. Результати спостережень за 1962 – 2005 роки проранжуємо за вхідним фактором $x_{сер-зв}$, тобто впорядкуємо вхідний фактор за зростанням. Всі значення результуючого показника, крім декількох значень, при цьому також набувають певного порядку (фактично майже всі значення кількості зсувів також зростають). Графік цієї залежності, прогнозних даних та відхилень (MAE і MAPE) наведено на рис. 3.

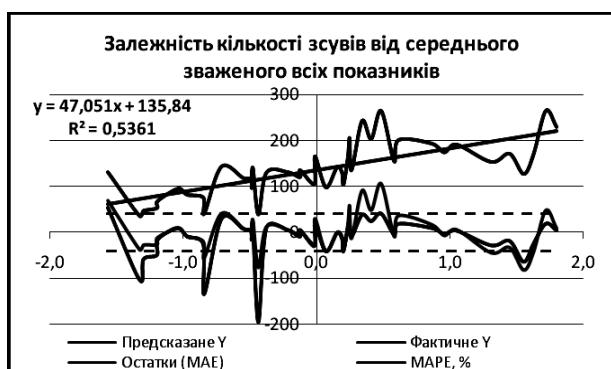


Рис. 3. Залежність результуючого показника від середнього зваженого факторів і залишки (MAE і MAPE)

При порівнянні років, в які відбулися значні відхилення від загальної тенденції, помічаємо, що відхилення в меншу сторону відбулися в 70-80 роки, а в більшу – в 90-2000 роки. Це можна пояснити тим, що наприкінці шістдесятих років Південний берег

Криму набуває статусу міжнародної здравниці СРСР, а тому на будівельні роботи виділяються в цей період значні кошти, які вкладаються в тому числі і в споруди, що укріплюють схили.

З початку дев'яностих років Україна входить в складний економічний стан, в зв'язку з чим коштів ледве вистачає на усунення катастрофічних наслідків зсувних процесів (на завчасне укріплення зсуво-небезпечних ділянок кошти практично не витрачаються), як наслідок – різке зростання кількості зсувів. Значне абсолютне (MAE) і відносне (MAPE) відхилення фактичних значень від прогнозних упорядкуємо за роками і помістимо в табл. 3.

Червоним виділено роки, коли фактичне значення перевищує прогнозне. З наведеної таблиці слідує, що всі вище перелічені припущення щодо впливу вкладених коштів на кількість зсувних процесів підтверджуються. Серед останніх років виділяється 1987, це пов'язано з сильними землетрусами на Кавказі, а саме в Нагорному Карабаху, які нанесли великі руйнування цьому регіону і спровокували людські жертви. Відповідно і Південний берег Криму відчув на собі ці поштовхи, – усталеність схилів було порушено, відбулася активізація зсувів, проте були виконані роботи по укріпленню схилів і далі спостерігався спад активності. Починаючи з 1995 року вкладання коштів значно зменшується, що відображається на зсувній активності, тобто вона зростає, відхилення набувають великих значень.

Таблиця 3

Залишки для моделі РОСЗ, які набувають занадто великого значення за модулем (абсолютні та відносні)

Рік	Залишки (MAE)	Залишки, % (MAPE)
1970	-82	-64%
1972	-76	-195%
1974	-55	-134%
1975	-37	-103%
1976	-28	-59%
1977	-26	-50%
1984	-41	-40%
1987	41	29%
1991	-41	-42%
1995	69	53%
1997	90	37%
1998	47	18%
2000	58	28%
2002	106	40%

На рис. 4 показано діаграму будівельно-відновлювальних робіт за 1968 – 2000 роки [2], а також розраховано в середньому за указаний період вартість робіт на один зсув за один рік.

Таким чином, ми бачимо, що кошти вкладаються нерівномірно, це впливає на результуючий показник, а завчасно вкладені кошти значною мірою стримують руйнівні процеси, які відбулися внаслідок

док активізації зсувів.

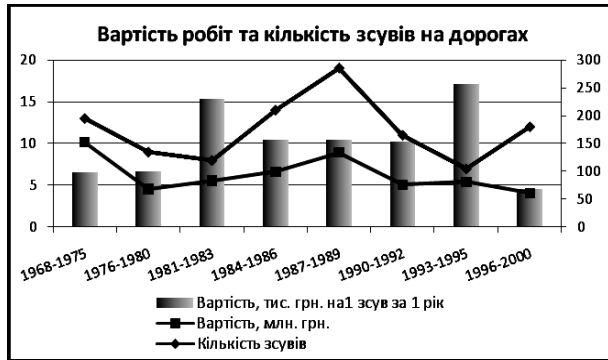


Рис. 4. Вартість робіт для запобігання зсувним процесам Південного берега Криму

На рис. 5 зображено будівництво буронабивних паль (стовпів) 2008 року для укріплення траси Сімферополь – Ялта – Севастополь.



Рис. 5. Укріплення дорожнього полотна буронабивними пальями для стримування зсувних процесів Південного берега Криму

На цій трасі 92 зсувонебезпечних ділянки, які закріплені за допомогою палих утримуючих споруд, з них лише шість зазнали деформації та ушкодження в період сплеску активізації зсувних процесів [2]. Це свідчить про високу надійність таких споруд.

Модель РОСЗ висвітлює ще один цікавий момент. В процесі упорядкування за вхідним фактором $X_{\text{ср.зв.}}$ ми отримали зростаючий ряд незалежної змінної, якому відповідає майже всюди зростаючий ряд залежної змінної.

Розіб'ємо ряд спостережень на три частини (кластери): 1-й відповідає незначній активності зсувних процесів; 2-й – активізація є, але незначна і катастрофічних наслідків за собою не несе; 3-й – кількість зсувів велика і руйнація важливих об'єктів значна. Оскільки ряд нараховує 43 спостереження, то в кожний кластер попаде 14 спостережень: 1-й – значення незалежного фактора лежать в інтервалі $[-1,5; -0,5]$; 2-й – в інтервалі $[-0,4; 0,2]$ і 3-й – $[0,3; 1,8]$. При цьому всі значення залежного показника відповідають своєму кластеру, крім певних

значень, виділених раніше і показаних в табл. 3, які суттєво залежать від вкладених коштів. Це означає, що перевищення середнього зваженого всіх нормованих вхідних факторів на 30 відсотків тягне за собою збільшення активності зсувних процесів, а також збільшення витрат на усунення катастрофічних наслідків цих процесів. А також завчасне вкладення коштів і проведення відповідних робіт зменшують кількість катастрофічних зсувів.

Жодна модель із запропонованих раніше не дала такої чіткої кластеризації за вхідними факторами і не пояснювала певні відхилення від розрахункових значень результуючого показника рівнем вкладених коштів на закріплення зсувонебезпечних ділянок.

Таким чином, наведена модель РОСЗ – регресія однофакторна середньо зважена належним чином описує зсувні процеси Південного берега Криму, і з врахуванням рівня проведених робіт по укріпленню зсувонебезпечних ділянок можна достовірно оцінювати прогноз на наступний рік, а також рекомендувати заходи по попередженню катастрофічних наслідків цих процесів.

Висновки

В даній статті було проаналізовано необхідність створення моделей, які пов'язують спостереження різних незалежних організацій за кліматичними, сейсмічними та сонячними даними в сучасних умовах моделювання і прогнозування зсувних процесів Південного берега Криму, які використовують попередні відомі дані для отримання оцінок прогнозу в майбутньому, що є новим в моделюванні цих процесів.

Запропоновано розглядати результати спостережень не в абсолютному значенні, а пронормувати різні фактори з метою привести їх до однієї розмірності, а також за допомогою вагових коефіцієнтів об'єднати всі вхідні фактори в одну змінну – середньо зважене. Таким чином, отримано нову регресійну модель – РОСЗ, яка пов'язує нову змінну з результуючим показником.

За допомогою критеріїв оцінки моделей та оцінки прогнозу проаналізовано п'ять моделей, запропонованих раніше, і модель РОСЗ. Результати розрахунків наведено у вигляді таблиці, з якої слідує, що нова модель придатна для прогнозування.

Встановлено, що при впорядкуванні середньо зваженого фактору за зростанням також впорядковується і результуючий показник, крім деяких значень, які реагують на інші фактори впливу. Упорядковані дані розбиваються на три великих кластери, які відображають зв'язок між незалежними даними спостережень та кількістю активних зсувів. Встановлено межі для вхідного фактору, при яких результуючий показник набуває одного з трьох значень: маленька кількість зсувів, середня (незначні наслідки і незначні витрати) та велика – катастрофічні на-

слідки, великі кошти на відновлення ушкоджених ділянок.

Визначено причини виникнення розбіжностей між прогнозними і реальними даними для моделі РОСЗ – це рівень вкладених коштів в укріплення схилів. Для інших моделей наявність цієї закономірності не спостерігається.

В подальшому передбачається для виводу результату в інтелектуальній системі поєднувати всі отримані прогнози, показувати їх ефективність, а також додати рекомендації щодо вкладення коштів в наступному і поточному році у протизсувні заходи з урахуванням моделі РОСЗ.

Список літератури

1. Селін Ю.М. Системний аналіз екологічно небезпечних процесів різної природи / Ю.М. Селін // Системні дослідження та інформаційні технології. – 2007. – № 2. – С. 22-32.
2. Круцик М.Д. Захист гірських автомобільних доріг від зсувів / М.Д. Круцик. – Коломия, 2003. – 425 с.
3. Логвинова К.Т. Климат и опасные гидро- метеорологические явления Крыма / К.Т. Логвинова, М.Б. Барабаш. – Л.: Гидрометеоиздат, 1982. – 320 с.
4. Безсмертний А.Ф. Геофізична інтерпретація схилових деформацій в межах розвитку зсувів на Південному березі Криму / А.Ф. Безсмертний // Вісник Київського національного університету ім. Т.Г. Шевченка; серія «Геологія». – К.: Видавничо-поліграфічний центр «Київський університет», 2001. – Вип. №19. – С. 37-41.

5. Шеко А.И. Прогноз экзогенных геологических процессов на Черноморском побережье СССР / А.И. Шеко. – М.: Недра, 1979. – 242 с.

6. Оползни Крыма. Часть 1. История отечественного оползневедения / И.Ф. Ерыш, В.Н. Саломатин. – Симферополь: Апостроф, 1999. – 249 с.

7. Оползни Крыма. Часть 2. Методы изучения оползней / А.Г. Бакиров, А.Ф. Бессмертный, И.Ф. Ерыш, В.Ф. Котлов, А.В. Макушин, В.Н. Саломатин, В.Н. Сальников. – Симферополь: Апостроф, 1999. – 175 с.

8. Бідюк П.І. Аналіз часових рядів / П.І. Бідюк. – К.: ННКА ІПСА, 2006. – 188 с.

9. Бідюк П.І. Системний підхід до прогнозування на основі моделей часових рядів / П.І. Бідюк // Системні дослідження та інформаційні технології. – 2003. – №3. – С. 88-110.

10. Таран В.М. Методика оцінювання регресійних моделей, побудованих за даними спостережень, що описують зсувні процеси Південного берега Криму / В.М. Таран // Матеріали XII Міжнародної науково-технічної конференції «Системний аналіз та інформаційні технології САІТ – 2010». – К., 2010. – С. 159.

11. Таран В.М. Практичне впровадження розроблених методів прогнозування зсувних процесів Південного берега Криму / В.М. Таран // Вісник НТУ «ХПИ» MicroCAD – 2010. – С. 162-172.

12. Свідоцтво про реєстрацію авторського права на твір №31281. Комп'ютерна програма "Forecasting of Landslide for Southern Coast of Crimea" («FLSSCC»). / В.М. Таран // 08.12.2009.

Надійшла до редколегії 13.08.2010

Рецензент: д-р техн. наук, проф. П.І. Бідюк, Національний технічний університет України "КПІ", Київ.

КЛАСТЕРИЗАЦИЯ ДАННЫХ ДЛЯ ФОРМИРОВАНИЯ ВЫВОДА ПРИ ПРОГНОЗИРОВАНИИ ОПОЛЗНЕВЫХ ПРОЦЕССОВ ЮЖНОГО БЕРЕГА КРЫМА

В.Н. Таран

Предложено при анализе оползневых процессов Южного берега Крыма рассматривать результаты наблюдений в стандартном виде с целью привести их к одной размерности, с помощью весовых коэффициентов объединить все входные факторы в одну переменную – среднюю взвешенную, предложена модель РОСЗ – регрессия однофакторная средняя взвешенная. Установлено, что при упорядочении независимого фактора по возрастанию также упорядочивается и результирующий показатель. Упорядоченные данные разбиваются на три больших кластера, установлены пределы для входного фактора, при которых результирующий показатель приобретает одно из трех значений: маленькое количество оползней, среднее и большое. Определены причины возникновения расхождений между прогнозными и реальными данными для модели РОСЗ – это уровень вложенных средств в укрепление склонов. Для других моделей наличие этой закономерности не наблюдается. С помощью критериев оценки моделей и оценки прогноза проанализировано пять моделей, предложенных раньше, и модель РОСЗ, сделан вывод, что все модели пригодны для прогнозирования.

Ключевые слова: прогнозирование оползневых процессов Южного берега Крыма, регрессионный анализ, оценка модели и оценка прогноза, кластеризация.

CLUSTERIZATION OF DATA FOR FORECASTING FOR LANDSLIDE PROCESSES OF THE SOUTHERN COAST OF CRIMEA

V.N. Taran

It is suggested at the analysis of landslide processes of the Southern Coast of Crimea to examine the results of supervisions in a standard kind on purpose to bring them over to one dimension, by means of gravimetric coefficients to unite all entrance factors in one variable: middle weighted. There is offered ROMW-model (regression onefactor middle weighted). It is set that at the ascending ordering of independent factor a resulting index is also put in order. The ranked data are broken up on three large clusters, limits for an entrance factor, at which a resulting index acquires one of three values, are set: little amount of landslides, AV and large. Reasons of origin of divergences are certain between prognosis and real data for the model of ROMW is the level of the inlaid facilities in strengthening of slopes. For other models the presence of this conformity to law is not observed. By means of criteria of estimation of models and estimation of prognosis five models, offered before, are analyzed, and model of ROMW, drawn conclusion, that all models are suitable for forecasting. Figs: 5. Tabl.: 4. Refs: 12 titles.

Keywords: forecasting of landslide processes of the Southern Coast of Crimea, regressive analysis, model estimation and prognosis estimation, clusterization.