СПИСАНИЕ НА БЪЛГАРСКОТО ГЕОЛОГИЧЕСКО ДРУЖЕСТВО, год. XLIII, кн. 2, 1982
REVIEW OF THE BULGARIAN GEOLOGICAL SOCIETY, vol. XLIII, part 2, 1982

Стохастични структурни модели в инженерната геология

С. Маматарков

Научноизследователски институт по полезни изкопаеми, 1505 София

S. Mamatarkov — Stochastic Structural Models in Engineering Geology. The paper presents an analysis of the implementation of the criteria for applicability of statistical methods in data processing and generalization of results from measurements of the indices of engineering geological properties of rocks. To describe their spatial variability random variable and regression (trend) structural models are discussed.

The advantages of regression structural models are analysed and explained, namely: 1) higher accuracy even when sufficiently close correlations between the index and the coordinates of measurement points is lacking; 2) plurality in determination of the calculated (generalized) values of the indices and of the interval estimates with a certain probability.

Attention is drawn to recent methods for overcoming computation difficulties and to the proposed program realizations.

Изучаването и обобщеното (моделно) представяне на пространствената изменчивост на физичните и механичните свойства на скалите, т. е. създаването на структурни модели, е основна задача на инженерната геология, особено при проучването на находищата на полезни изкопаеми. За целта могат да се използуват различни математически методи, но най-широко приложение имат статистическите методи. Обосновката за тяхното прилагане може да се търси или в същността на свойствата на скалите (Вистелиус, 1980), или в начина на тяхното измерване (Рац, 1973).

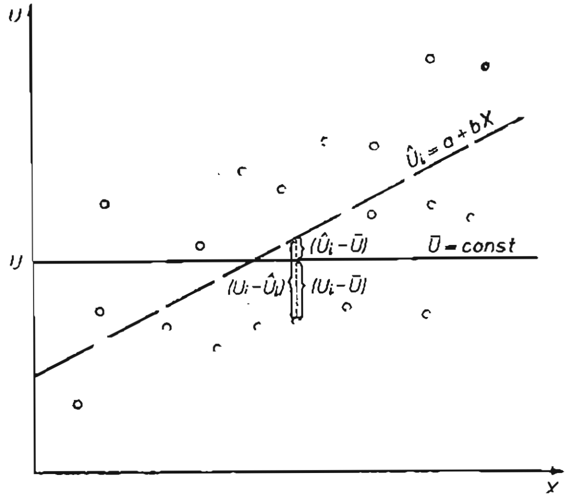
В аспекта на същността на свойствата на скалите съществуват две противоположни мнения: а) формирането на свойствата на скалите във всяка точка на скалния масив е резултат от съвместното действие на такова разнообразие от фактори, всеки от които действува детерминирано, че да се опише и оцени това действие поотделно е невъзможно и е необходимо усредняване. В този случай „статистическата природа" на свойствата на скалите има епистемологичен характер и прилагането на статистическите методи е чисто прагматично; б) контратезата е, че действието на литогенетичните фактори е стохастично по своята природа. Следователно свойствата на скалите онтологично са стохастични и не нашата ограничена възможност да вникнем в детайлите води до неопределеност, невъзпроизводимост и оттам до случайност на измерваните стойности на свойствата на скалите. Независимо кое от двете положения е вярно, а засега все още няма достатъчно данни, за да се приеме едно от тях, статистическите моделни представи имат законно място в инженерногеоложката практика.

Аргументите, свързани с начина на получаване и използуване на инженерногеоложката информация, се свеждат до: 1) показателите на инженерногеоложките свойства на скалите се измерват с помощта на експерименти, провеждани в различни физически точки на скалния масив. Резултатът от измерването представлява някакво усредняване по някаква област, наричана определяща област на експеримента. Удобно е този резултат да се представя като някаква статистика от полето на това свойство в границите на определящата област на експеримента. Тъй като нито размерите, нито положението на тази област могат да бъдат фиксирани съвършено точно, измерените показатели се оказват случайни величини. Това обстоятелство е в сила независимо от степента на детайлност и точност, с които е определена структурата на полето на дадения показател. В това се състои „статистическата природа“ на свойствата на скалите или по-точно на постулираната моделна представа; 2) обикновено размерите на определящите области на експериментите са малки в сравнение с размерите на областите на въздействие на съоръженията или мероприятията на инженерно-стопанската дейност на човека. Получаваните в експериментите усреднявания в сравнително малки обеми от геоложкото пространство са подложени на нерегулярна изменчивост. Това се дължи на съществуващите различни нива на нееднородност в тези обеми. От двата съществуващи класа методи за описание на тази нерегулярна изменчивост — детерминистичния и статистическия, последният е по-достъпен, по-удобен и по-икономичен; 3) дискретният характер на инженерногеоложките измервания (в отделни физически точки) прави детерминираното описание на обекта много чувствително към пространственото разположение и броя на точките на измерване. Статистическото описание е относително устойчиво и в по-малка степен зависимо от проучвателната мрежа.

Математическият апарат на теорията на вероятностите, който се използува от статистическите методи, е построен за работа с явления, на които наред с неопределеност е присъща още и статистическа устойчивост (еднородност). Това означава, че на всяка стойност на случайната величина съответствува, макар и неизвестна, но напълно определена вероятност. Общо взето, отсъствието на статистическа устойчивост не изключва използуването на статистиките, но при нейно отсъствие са неприложими методите за оценка достоверността на получаваните резултати. И тъй като в инженерногеоложката и инженерната практика са важни не само самите стойности на показателите на свойствата на скалите, но и тяхната точност и надеждност, то изискването за статистическа устойчивост придобива принципно значение.

В инженерногеоложката практика за оценяване и описание на показателите на свойствата на скалите са приложими и се използуват няколко структурни модела: детерминирани, случайна функция, случайна величина и регресионни (трендови). Детерминираните структурни модели са приложими в много редки случаи и почти нямат практическо значение. Моделите, използуващи апарата на теорията на случайните функции за оценка и описание на нееднородността и изменчивостта на инженерногеоложките свойства на скалите, изискват значително по-голямо количество информация от това, с което обикновено се разполага. Тяхното приложение се ограничава при много отговорни съоръжения и при провеждането на специални проучвания. Най-широко разпространение имат моделите, използуващи теорията на случайните величини, а значително по-малко — регресионните модели.

При модела на случайната величина априори се допуска и приема: а) че резултатите от измерванията са независими, т. е. разстоянието между точките на измерване е по-голямо от радиуса на автокорелация; б) че те имат нормално или незначимо отличаващо се от него разпределение; в) че генералната съвкупност, от която е направена извадката с измерванията, т. е. отделеното инженерногеоложко тяло, е статистически еднородна; г) че съществува стационарен режим на изменчивост на свойствата на скалите по целия обем на инженерногеоложкото тяло, т. е. математическите очаквания на показателя и неговата дисперсия са постоянни, а корелационната функция е нулева.



Фиг. 1. Съотношение на разликите при модела на независимата случайна величина и при регресионния структурен модел

Някои изследвания (Бабенышев, 1974; Комаров, 1972) показват, че радиусът на автокорелация в статистически еднородна скала не превишава 1—2 m. Следователно стойностите от измервания във физически точки на по-голямо разстояние ще бъдат статистически независими. При проучването на находища на полезни изкопаеми параметрите на проучвателната мрежа в плоскостта (X, Y) значително превишават това разстояние. По дълбочина (Z) разстоянието между точките на измерване също най-често е по-голямо.

Нашият опит показва, че показателите на инженерногеоложките свойства на скалите по изключение имат нормално или незначимо отличаващо се от него разпределение. Въпреки твърденията на редица авторитетни автори в реалните съвкупности, с които има работа инженер-геологът, е налице много често съществено отклонение от нормалното разпределение. Причините могат да бъдат много, но според нас най-важни са две. Първата е, че отделянето на статистически еднородни (устойчиви) инженерногеоложки тела е трудна за решаване проблема. За прилагането на формални критерии обикновено не достига експерименталният материал и това налага отделянето да става най-вече по геоложки признаци. Липсата на бимодално и силно асиметрично разпределение на практика се приема като достатъчно условие за статистическа еднородност, което най-общо невинаги е вярно. Следствието е статистически недостатъчно еднородни инженерногеоложки тела. Втората причина е свързана с много строгите критерии, на които отговаря нормалното разпределение — дефиниране в точка ($\beta_1=0$; $\beta_2=3$). Съществуват различни по мощност критерии за оценяване на значимостта на отклонението на едно емпирично разпределение от нормалното. Тяхното приложение обаче е чисто констативно и на практика не води до използуването на адекватни методи за обработка на измерванията и получаване на обобщени и разчетни показатели.

Структурният модел на случайната величина при описаните ограничения е приложим при измервания на една и съща величина или в тези случаи, когато може да се приеме, че инженерногеоложките измервания могат да се моделират като многократни измервания на една и съща величина. В геоложки аспект изложеното изглежда така. Известно е, че свойствата на скалите във всяка точка на геоложкото пространство се формират под действието на множество от литогенетични фактори $\{F_i\}$. Също така е известно, че последователността, времетраенето и интензивността на действието на тези фактори са определими само в най-общ вид, и то качествено. Засега те не могат да бъдат измерени или оценени както за дадена физическа точка, така и за даден обем от геоложкото пространство. Моделът на случайната величина предполага постоянство на литогенетичните фактори по целия обем на инженерногеоложкото тяло $\{F_i=const\}$. И това постоянство обхваща както последователността, така и времетраенето и интензивността на действие. Само в такъв случай може да се приеме, че всички проби, взети от инженерногеоложкото тяло, представляват една-единствена проба и всички измервания на даден показател са многократни измервания на една и съща величина.

За всеки е ясно, че такова постоянство реално не съществува и не може да съществува. Следователно структурният модел на случайната величина е само най-грубо приближение към действителността, тъй като не отговаря на геоложката реалност $\{F_i(X, Y, Z)\neq const\}$. Усредняването по целия обем на инженерногеоложкото тяло е неправомерно или най-малкото свързано със значителна дисперсия и грешки. Тъй като функциите на изменение на литогенетичните фактори в пространство-времето са практически неопределяеми, алтернативата е свойствата на скалите да се поставят в съответствие с пространствените координати $(t=const)$. Така се стига до регресионните структурни модели, в които независими променливи (регресори) са пространствените координати.

Съдържателната същност на регресионните структурни модели се състои в постулирането, че измененията на стойностите на даден показател U в определена площ или обем на геоложкото пространство са някаква неизвестна нам функция от координатите

$$M(U(X, Y, Z))=f(X, Y, Z)+\varepsilon. \tag{1}$$

Ние постулираме, че измененията на математическото очакване на дадения показател се състоят от две части: функцията $f(X, Y, Z)$, която е някаква апроксимация и отразява детерминираната (систематичната, тренд) компонента, и сумарната грешка ε, която е обусловена от случайните вариации на U и от неточностите и непълната възпроизводимост при измерванията на U. Случайните вариации могат да имат отчасти закономерен характер. Теоретично те биха могли да се изразят чрез някаква бързо флуктуираща функция $\varphi(X, Y, Z)$. В такъв случай сумарната грешка ε би била сума от тази функция и грешката от неточностите и невъзпроизводимостта на измервателните есперименти:

$$\varepsilon(X, Y, Z)=\varphi(X, Y, Z)+\delta. \tag{2}$$

За съжаление поради дискретния характер на получаването на инженерногеоложката информация и нейното недостатъчно количество практически е невъзможно да се определят видът и параметрите на функцията $\varphi(X, Y, Z)$. Следва да се задоволим с отчитането на сумарната грешка ε, което също носи значителни преимущества.

При създаването на регресионните структурни модели не съществуват или много рядко съществуват априорни съображения за вида на функцията $f(X, Y, Z)$. Освен това колкото по-сложен е математическият модел, толкова по-трудно той се поддава на съдържателна интерпретация. Не са за пренебрегване и изчислителните трудности, които, макар и несъществени при съвременните бързодействуващи ЕИМ, имат определен икономически ефект. Всичко това естествено води до линейните (по коефициентите) полиномиални модели. Полиномиалните представяния, както е известно, позволяват колкото искаме точно да апроксимираме всякакви непрекъснати функции. Освен това те дават възможност за лесна линеаризация на нелинейните по регресорите полиноми.

Полиномиалните модели имат вида

$$M(U(X, Y, Z))=\widehat{U}=\beta_0+\beta_1 X+\beta_2 Y+\beta_3 Z+\cdots+\beta_p Z^k+\varepsilon \tag{3}$$

или

$$\widehat{U}=exp\,[\beta_0+\beta_1 X+\beta_2 Y+\beta_3 Z+\cdots+\beta_p Z^k]+\varepsilon. \tag{4}$$

Експоненциалната функция (4) има редица предимства (Романова, 1968): 1) експонентата не може да приема отрицателни стойности; 2) логаритмуването на показателите U с голяма положителна асиметрия на разпределението изглажда стойностите, силно отклоняващи се от средната, а това има положителен ефект за апроксимиращата функция; 3) експоненциалната функция леко се линеаризира.

Върху математическата страна на построяването на линейни регресионни модели и по-специално на полиномиалните модели съществува обширна математическа и дори геоложка и инженерногеоложка литература (Комаров и др., 1976). Необходимо е да се разгледат обаче някои въпроси, на които в геоложката и инженерногеоложка литература не се обръща достатъчно внимание: основни допущания и свойства на получаваните оценки.

Относно случайната компонента ε, наричана сумарна или привидна грешка, се правят следните допущания: 1. Грешката ε е случайна величина. 2. Математичното очакване на ε е равно на нула — $M(\varepsilon)=0$. 3. Дисперсията на ε е постоянна — $D(\varepsilon)=const$. 4. Последователните стойности на ε са независими една от друга — $cov(\varepsilon)=0$.

Така при построяването на регресионния структурен модел се приема, че за всяко измерване на показателя U е справедлива връзката

$$\widehat{U}_i=\beta_0+\beta_1 X_i+\beta_2 Y_i+\beta_3 Z_i+\cdots+\beta_p Z_i^k+\varepsilon_i. \tag{5}$$

В резултат на инженерногеоложките проучвания се получава отбрана съвкупност (извадка) от множеството стойности $\{U_i, X_i, Y_i, Z_i\}$ от изучаваната генерална съвкупност — отделеното инженерногеоложко тяло. Задачата се състои в определянето на параметрите β_p. Истинските стойности на тези параметри е невъзможно да се определят, тъй като въпросът опира до ограничения обем на разполагаемата информация — извадка с ограничен обем. Ние можем да получим само статистически оценки на истинските параметри b_p и тогава (3) става

$$\widehat{U}=b_0+b_1 X+b_2 Y+b_3 Z+\cdots+b_p Z^k+\varepsilon. \tag{6}$$

Приемайки някаква хипотеза за формата на връзката между U и (X, Y, Z), ние не можем еднозначно да подберем параметрите b_p, тъй като през съответната област на пространството (U, X, Y, Z) могат да се прокарат множество полиномиални повърхнини. Необходим е някакъв критерий. В качеството на такъв критерий е естествено да се приеме изискването за съотношението на стойностите от наблюденията U_i с изчислените $\widehat{U}_i$, и то такова, че полиномиалната повърхнина да мине възможно най-близко към данните от измерванията. Различните методи за оценяване на параметрите β_p, т. е. намирането на оценките b_p, се опират на различни критерии, измерващи степента на близост на изчислените с фактическите данни, и, разбира се, дават различни оценки b_p за една и съща съвкупност $\{U_i, X_i, Y_i, Z_i\}$. При това получаваните оценки притежават различни статистически свойства.

По силата на своята простота и широка област на приложение най-разпространен е методът на най-малките квадрати (МНМК). Същността на МНМК се изразява в това да се намерят такива оценки b_p, че

$$\sum_i \varepsilon_i^2 = \sum_i (U_i - \widehat{U}_i)^2 = min. \qquad (7)$$

Не е маловажно и това, че получаваните по МНМК оценки b_p при условие, че направените предположения за $\varepsilon(1 \div 4)$ са справедливи, притежават редица важни свойства:

1. Оценките b_p са неизместени, т. е. $M(b_p) = \beta_p$. Това свойство е логическо следствие от второто предположение за характера на грешката ε. Неизместеността означава, че оценките b_p по извадката се концентрират около неизвестните истински параметри β_p.

2. Оценките са състоятелни, т. е. $\lim\limits_{N \to \infty} D(b_p) = 0$.

3. Оценките са ефективни в този смисъл, че те имат минимална дисперсия в сравнение с всякакви други оценки.

Ако предположението 3 $(D(\varepsilon) = const)$ или 4 $(cov(\varepsilon) = 0)$ е нарушено, то неизместеността и състоятелността на оценките се запазва, но те се оказват по-слабо ефективни, отколкото в случаите, когато тези допускания са справедливи.

Съвършено очевидно е, че за прогнозирането на стойностите на показателите на свойствата на скалите в различни точки на геоложкото пространство съвсем не е безразлично какви стойности ще притежават оценките. Що се отнася до свойството неизместеност, то е необходимо. Наистина изместените оценки априори дават неверни сведения за положението на $f(X, Y, Z)$ в признаковото пространство. Свойството състоятелност означава, че при увеличаване броя на измерванията N оценките на параметрите стават по-надеждни във вероятностен смисъл, т. е. с нарастването на N оценките по-плътно се концентрират около β_p. Свойството ефективност е най-важно, доколкото то определя степента на възможната грешка на прогнозата.

Полиномът (3) в матрична форма е

$$U = V\beta + \varepsilon, \qquad (8)$$

където $U = (U_i)$ — вектор на измервания показател; $\beta = (\beta_p)$ — вектор на неизвестните регресионни коефициенти; $V = (v_{ij})$ — матрица на независимите променливи, чийто размер се определя от броя на измерванията $i = \overline{1, N}$ и от броя на променливите $j = \overline{1, K+1}$; $\varepsilon = (\varepsilon_i)$ — вектор на грешките.

При множествената линейна регресия към описаните четири допускания за грешката ε се добавя и

5. Матрицата V се състои от линейно независими вектори-стълбове, т. е. между векторите $v_{i1}, v_{i2}, v_{i3}, \ldots, v_{ik}$ няма линейни зависимости.

Последното е равносилно на това, че рангът на матрицата V е равен на $K+1$, а това на свой ред означава, че $|V^TV|=0$, т. е. матрицата V^TV е обратима.

Условието 5 се изпълнява, когато като регресори се използуват координатите на точките на измерване. Тук обаче възникват изчислителни трудности, водещи до нереални резултати и свързани с необходимостта от обръщане на матрицата V^TV, нейната голяма размерност и лоша обусловеност. Някои програми за обръщане на матрици имат тенденцията да губят устойчивост при увеличение на размерността (Д е в и с, 1977). Това може да се дължи и на натрупване на грешката от закръглявания, особено при недостатъчно значещи цифри. В последно време се появиха математически методи (Ф о р с а й т и др., 1980), които дават възможност за решаване на тази проблема в рамките на МНМК дори когато условие 5 не се удовлетворява. Това е методът на сингулярното разложение.[1]

Структурният модел на случайната величина дава обобщена характеристика за показателите на свойствата на скалите за целия обем на отделеното инженерногеоложко тяло. Регресионният структурен модел в сравнение със структурния модел на случайната величина дава следните преимущества:

1. Условно разделя изменчивостта на инженерногеоложкия показател във фиксираното статистически устойчиво (еднородно) инженерногеоложко тяло от геоложкото пространство на две компоненти: детерминирана — $f(X, Y, Z)$, и случайна — ε. При това функцията $f(X, Y, Z)$ представлява най-добрата в определен смисъл апроксимация на фактическите данни. В нашия случай всеки полином е най-добрата апроксимация в смисъла на най-малките квадрати за дадената степен. Това означава, че даден полином от трета степен е най-добрата апроксимация от трета степен по смисъла на МНМК, но полином от втора, четвърта или друга степен най-общо може да бъде по-добра апроксимация. Подчертаваме това обстоятелство, тъй като определянето на степента на полинома и (или) вида на функцията $f(X, Y, Z)$, която ще бъде най-добрата апроксимация измежду всички възможни, е по принцип трудна, макар и решима задача. Решаването на тази задача става или чрез сравнение на различни апроксимации по условието

$$\sum_i \varepsilon_i=0 \quad \text{или} \quad 0\leq\sum_i \varepsilon_i\leq\zeta,$$

като ζ е някакво предварително зададено малко число, или чрез определяне степента на полинома и (или) вида на функцията по съществуваща съдържателна инженерногеоложка информация.

Апроксимацията (3) дава възможност да бъде определена интервална оценка на U във всяка точка на интересуващото ни геоложко пространство, т. е. безкрайно много интервални оценки за разлика от единствената при случайната величина.

2. Самата същност на регресията, т. е. поставянето на U в зависимост от регресорите (X, Y, Z), води до по-голяма точност на определяне на оценката $\widehat{U}$. Това нагледно много добре се вижда на фиг. 1. Съвсем ясно се виждат двата вида разлики $(U_i-\bar{U})$ и $(U_i-\widehat{U}_i)$. Средните суми от квадратите на тези разлики дават съотношенията

$$\Sigma(U_i-\bar{U})^2=\Sigma(\widehat{U}_i-\bar{U})^2+\Sigma(U_i-\widehat{U}_i)^2 \qquad (10)$$

[1] Под ръководството на автора в НИПИ от н. с. Кр. Калинов е създаден комплекс програми на *FORTRAN(H)* за *IBM-360/145* за създаване на регресионни модели по МНМК и чрез сингулярни разложения.

и

$$S_u^2=\delta^2_{u|x,y,z}+S^2_{u|x,y,z}, \tag{11}$$

където S_u^2 — общата дисперсия на U; $\delta^2_{u|x,y,z}$ — дисперсия на вариациите на U, възникващи под влияние на (X, Y, Z) — дисперсия на регресията, или обяснена дисперсия; $S^2_{u|x,y,z}$ — остатъчна (необяснена) дисперсия.

И при най-слаба връзка на изменението на U в зависимост от координатите X, Y, Z следва, че

$$\Sigma(U_i-\bar{U})^2\geq\Sigma(U_i-\hat{U}_i)^2 \tag{12}$$

или

$$S_u^2\geq S^2_{u|x,y,z}. \tag{13}$$

Степента на увеличение на точността на оценката на инженерногеоложкия показател в дадена точка на пространството може непосредствено да се оцени по съотношението

$$S_{u|x,y,z}=S_u\sqrt{1-R^2_{u|x,y,z}},$$

където $R^2_{u|x,y,z}$ е коефициентът на множествена детерминация, или квадратът на индекса на корелация.

Посочените две преимущества на регресионните структурни модели пред структурните модели на случайната величина имат много голямо практическо значение. При проучването на находищата на полезни изкопаеми регресионните структурни модели са единствените най-достъпни и удобни модели за обработка и обобщение на инженерногеоложката информация.

Литература

Бабенышев, А. П. 1974. Взаимосвязь между показателями свойств в соседних точках и ее зависимость от степени статистической однородности. — *Инженерные изыскания в строительстве*, 2, 2, 7—13.

Вистелиус, А. Б. 1980. *Основы математической геологии*. М., Недра. 216 с.

Девис, Дж. 1977. *Статистика и анализ геологических данных*. М., Мир. 572 с.

Комаров, И. С. 1972. *Накопление и обработка информации при инженерно-геологических исследованиях*. М., Недра. 296 с.

Комаров, И. С., Н. М. Хайме, А. П. Бабенышев. 1976. *Многомерный статистический анализ в инженерной геологии*. М., Недра. 199 с.

Рац, М. В. 1973. *Структурные модели в инженерной геологии*. М., Недра. 216 с.

Романова, М. А. 1968. Сортировка обломочного материала эоловых отложений централных каракумов. — В: *Вопросы математической геологии*. Л., Наука, с. 207—224.

Форсайт, Дж., М. Малькольм, К. Моулер. 1980. *Машинные методы математических вычислений*. М., Мир. 280 с.

(Постъпила на 15. VI. 1981 г.)