V. БИАВТОМАТИКА И БИОИНФОРМАТИКА

МЕТОДОЛОГИЧЕН ПРОБЛЕМ НА ИМИТАЦИОННОТО МОДЕЛИРАНЕ НА АГРОЕКОСИСТЕМИТЕ

Александър Н. Садовски

METHODOLOGICAL PROBLEMS IN SIMULATION OF THE AGRO-ECOSYSTEMS

Alexander N. Sadovski

Summary: Simulation models describe the functioning of agro-ecosystems and the flow of processes in them using computer programs. With their help can be made computer experiments and to predict the behavior of agro-ecosystems. One of the main problems in the creation of simulation models is imitation of uncontrollable soil and weather factors. A number of studies of soil properties and meteorological characteristics indicate that these stochastic variables do not follow the normal (Gaussian) distribution. In the selection of statistical models that approximate the distribution of such stochastic variables, the most appropriate is a family of Johnson distributions. The article examines ways to generate pseudorandom numbers from Johnson and Burr distributions families for simulating soil and weather factors determining agro-ecosystems.

Keywords: Simulation modeling, agro-ecosystems, empirical distributions, pseudorandom numbers.

УВОД

Най-просто растителната екосистема може да бъде определена като растителен организъм плюс средата, в която той живее. Въз основа на тази идея ние можем да определим агроекосистемите като управлявани и експлоатирани от човека растителни екосистеми, които се обуславят от четири групи фактори - биологични, метеорологични, почвени и антропогенни. В агроекосистемите посевът и неговото абиотично обкръжение са свързани помежду си и се намират в постоянно взаимодействие.

Съществуват два пътя на математическо изследване [Л. Илиев, 1980]. Първият път започва с изследване на свойствата и отношенията на реалните системи и процеси, и завършва с построяване на абстрактни теории за тях, в които се отразяват най-съществените свойства и отношения на изучаваните системи. Вторият път тръгва от готови математически теории, структури, формули и уравнения и завършва с намиране на такива обекти, за които моделите са конкретни интерпретации. Намирането на математически модели в различните науки и практиката се нарича математическо моделиране.

Понякога се срещат така наречените „съмнителни изследователски практики“ (Questionable Research Practices). Умишлено или не,

такива практики обикновено имат влошаващ ефект върху качеството и надеждността на резултатите от научните изследвания. Целта на настоящата статия е да се покаже, че предположението за нормално разпределение на множеството случайни променливи в научните изследвания е несъстоятелно. Като следствие от това е необходимо да се прилагат широко други статистически разпределения. Това е валидно и при имитационното моделиране на явления и процеси, в частност на агроекосистемите.

МАТЕРИАЛИ И МЕТОД

Имитационните модели описват функционирането на обекта и протичането на процесите с помощта на компютърни програми. С тяхна помощ могат да се извършват машинни експерименти и да се прогнозира поведението на системата. По правило те се използват, когато не е възможно да се определи аналитичната форма на зависимостите между величините, когато аналитичният модел трудно може да се изследва и интерпретира или когато експериментът с реалния обект е труден, скъп или практически невъзможен. Това обуславя широкото приложение на имитационното моделиране при изследване на такива големи, сложни и динамични системи, каквито са агроекосистемите.

Един от основните проблеми при създаването на имитационни модели на агроeko-

системите е имитиране на неуправляемите почвени и метеорологични фактори.

Изследванията върху различни такива фактори (например, почвените свойства съдържание на хумус, механичен състав, почвена реакция, съдържание на общ азот, общ фосфор, подвижни форми на фосфор, калий и др. или метеорологичните характеристики температура на въздуха, относителна влажност, количество на валежите, скорост на вятъра, слънчева радиация и т.н.) показват, че тези стохастични величини обикновено не следват нормалното (Гаусово) разпределение.

В изследване върху данни от Института по почвознание "Н. Пушкиров" е направен статистически анализ на редица количествени свойства на сивите горски почви от планинските и

предпланински райони на Северна България [А. Петрова, А. Садовски, 1984]. Анализирани са съдържание на хумус (%), механичен състав ($< 0.01 \text{ mm}$) в %, почвена реакция pH в KCl, общ азот (%), общ фосфор (%), подвижни форми на фосфор (ppm P) и калий (mg/100 g). Резултатите от анализа на показателите - съдържание на хумус и подвижен калий са представени в Табл. 1 и 2. Всички данни се отнасят за горния почвен слой (орния хоризонт) на сиви горски почви.

В Табл. 3 са дадени получените стойности на параметрите на разпределенията за всички изследвани свойства на сивите горски почви.

На Фиг. 1 и 2 са изобразени емпиричните разпределения на разглежданите свойства хумус подвижен калий. Вижда се, че те имат ясно изразена асиметрия.

Табл. 1. Съдържание на хумус в Сиви горски почви

Окръг	Площ dka	Хмин	Хмакс	Хср	Медиана	Мода	Ср.кв. откл.	Ср.абс. откл.	Асим.	Ексцес	Тип разпр.
Варна	142194	0,74	3,67	1,781	1,565	1,585	0,8893	0,670	1,632	3,419	SB
В.Търново	230559	1,03	3,04	1,887	1,752	1,725	0,4533	0,365	0,574	2,845	SB
Видин	171209	0,48	4,38	1,722	1,646	1,535	0,5211	0,402	0,534	4,863	SU
Враца	201271	0,62	5,18	2,314	2,145	2,008	0,7833	0,568	0,941	4,518	SB
Габрово	65318	0,19	3,53	1,754	1,604	1,337	0,5760	0,466	0,832	4,235	SB
Ловеч	438418	0,54	5,11	2,289	2,042	1,680	0,9184	0,707	1,540	4,153	SB
Монтана	131781	0,62	4,22	2,345	2,299	2,325	0,6643	0,488	0,850	4,417	SB
Търговище	280847	0,68	4,23	1,855	1,828	1,317	0,5123	0,379	0,707	4,732	SU
Шумен	69425	1,30	3,58	1,855	1,710	1,599	0,5410	0,432	1,678	4,050	SB
ОБЩО	1735343	0,19	5,18	2,036	1,896	1,865	0,7542	0,554	1,796	5,420	SB

Табл. 2. Съдържание на подвижен калий в Сиви горски почви

Окръг	Площ dka	Хмин	Хмакс	Хср	Медиана	Мода	Ср.кв. откл.	Ср.абс. откл.	Асим.	Ексцес	Тип разпр.
Варна	142609	9,0	43,0	26,60	27,51	29,32	6,490	4,988	0,050	3,187	SU
В.Търново	234975	7,0	55,0	26,18	25,78	25,84	7,136	5,093	1,159	5,429	SU
Видин	173779	3,8	52,0	20,34	20,92	20,95	6,926	5,399	0,002	3,822	SU
Враца	201243	4,3	47,0	21,59	20,28	19,68	8,023	6,208	0,490	3,328	SB
Габрово	64680	2,8	32,3	17,29	16,70	16,14	5,687	4,590	0,169	2,953	SB
Ловеч	438414	6,0	56,0	20,61	21,16	21,89	5,308	4,168	0,011	3,533	SU
Монтана	131967	4,0	43,8	22,89	23,61	25,11	7,661	6,151	0,001	2,718	SB
Търговище	286547	7,8	53,0	21,69	20,63	15,49	7,476	5,873	1,079	4,856	SB
Шумен	69465	15,0	58,3	29,12	27,74	28,38	8,252	5,838	1,754	5,053	SB
ОБЩО	1747301	2,8	58,3	22,49	22,06	21,15	7,440	5,721	0,417	4,413	SU

Табл. 3. Сиви горски почви - разпределения на почвените свойства

Почвени свойства	Тип разпр.	П а р а м е т р и			
		ξ	λ	γ	δ
Хумус %	SB	2.566	1.942	8.700	0.100
Мех. състав %	SB	-0.449	1.245	68.000	7.000
pH в KCl	SB	0.534	1.158	4.800	3.200
Общ азот %	SB	4.635	2.148	1.000	0.000
Общ фосфор %	SU	-2.392	1.776	0.021	0.050
Подв. фосфор	SB	2.065	0.888	64.600	0.400
Подв. калий	SU	-1.256	2.485	14.919	13.964

Фиг. 1. Хумус

Фиг.2. Подвижен калий

Друг пример е анализът на разпределението на валежите (в мм) измерени в тр. Плевен през 1971 г. [Л. Дъбов (ред.), 1976]. Те се характеризират със следните параметри:

Брой на наблюденията - 365
 Минимална стойност - 0,0
 Максимална стойност - 74,0
 Средна стойност - 1,6595
 Стандартно отклонение - 5,3465
 Асиметрия - 8,1735
 Ексцес - 95,821

Направен и анализ на разпределението на облачността (в баловете) за същия пункт и година. Неговите параметри са:

Брой на наблюденията - 365
 Минимална стойност - 0
 Максимална стойност - 10
 Средна стойност - 5,4548
 Стандартно отклонение - 3,6134
 Асиметрия - 0,00289
 Ексцес - -1,433

На Фиг. 3 и 4 са представени хистограмите на тези разпределения, които очевидно не са нормални.

Фиг. 3. Хистограма на валежите

Фиг. 4. Хистограма на облачността

Получени са следните стойности на параметрите на S_B разпределенията:

Валежи $\gamma = 3.0697$, Облачност

$\gamma = 0.00234$,

$\delta = 0.51387$, $\delta = 0.45181$,

$\lambda = 132.92$, $\lambda = 11.004$,

$\xi = -0.13654$. $\xi = -0.03921$.

Получените резултати потвърждават тезата, че много от количествените свойства на почвите и метеорологични характеристики имат разпределения, които съществено се отличават от нормалното разпределение. При изследване размера на частиците, получени в процеса на раздробяване на естествени твърди тела /камъни и др./, Колмогоров [А. Колмогоров, 1941] е установил, че същият се описва с логаритмически нормалното разпределение (S_L).

При избора на статистически модели за разпределенията, които апроксимират разпределението на емпиричните честоти на извадките от такива стохастични величини, най-подходяща се явява фамилията от разпределения на Джонсън [N. Johnson, 1949], която включва като частни случаи нормалното и логаритмически нормалното разпределения.

Емпиричните разпределения на Джонсън се представят от преобразуването

$$\frac{z - \gamma}{\delta} = g\left(\frac{x - \mu}{\lambda}\right) \quad (1)$$

където μ и γ са параметри на положението, λ и δ - параметри на мащаба, z е нормално разпределена случайна величина, а g е подходящо избрана функция. Джонсън е предложил следните три фамилии функции, които обхващат цялата област от равнината (β_1, β_2) на Фиг. 1, аналогично на седемте типа разпределения на Пирсън:

а) логаритмически нормално S_L

$$g\left(\frac{x - \xi}{\lambda}\right) = \ln\left(\frac{x - \xi}{\lambda}\right), \quad x \geq \xi; \quad (2)$$

Плътноста на това разпределение е

$$f_1(x) = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda}{x - \xi} \exp\left\{-\frac{1}{2}\left[\gamma + \delta \ln\left(\frac{x - \xi}{\lambda}\right)\right]^2\right\},$$

където

$x \geq \xi$, $\delta > 0$, $-\infty < \gamma < \infty$, $-\infty < \xi < \infty$.

б) система S_B

$$g\left(\frac{x - \xi}{\lambda}\right) = \ln\left(\frac{x - \xi}{\lambda + \xi - x}\right), \quad \xi \leq x \leq \xi + \lambda; \quad (3)$$

Плътноста на това разпределение е

$$f_2 = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda^2}{(x - \xi)(\lambda - x - \xi)} \exp\left\{-\frac{1}{2}\left[\gamma + \delta \ln\left(\frac{x - \xi}{\lambda - x + \xi}\right)\right]^2\right\},$$

където

$\xi \leq x \leq \xi + \lambda$, $\delta > 0$, $-\infty < \gamma < \infty$, $\lambda > 0$, $-\infty < \xi < \infty$

в) система S_U

$$g\left(\frac{x - \xi}{\lambda}\right) = \operatorname{arsh}\left(\frac{x - \xi}{\lambda}\right), \quad -\infty < x < +\infty. \quad (4)$$

Плътноста на това разпределение е

$$f_3 = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda}{\sqrt{(x - \xi)^2 + \lambda^2}} \exp\left\{-\frac{1}{2}\left[\gamma + \delta \ln\left(\frac{x - \xi}{\lambda} + \sqrt{\left(\frac{x - \xi}{\lambda}\right)^2 + 1}\right)\right]^2\right\},$$

където

$-\infty < x < \infty$, $\delta > 0$, $-\infty < \gamma < \infty$, $\lambda > 0$, $-\infty < \xi < \infty$.

Определянето на стойностите на параметрите се извършва по метода на моментите [I. Hill et al., 1976] или както е показано от Джордж и Рамачандран [F. George, K.M. Ramachandran, 2011]. Оценка на параметрите по метода на най-малките квадрати се дава в работата [J.J. Swain, S. Venkatraman, J.R. Wilson, 1988].

За генериране на псевдослучайни числа от фамилията разпределения на Джонсън е необходимо да се следва следния алгоритъм:

1. Генерират се равномерно разпределени случайни числа $U(0,1)$ например с подпрограмата URAND, виж [Дж. Форсайт и др., 1986, стр. 245].

2. Генерират се нормално разпределени случайни числа $N(0,1)$ например с Алгоритъм 267, виж [M. Pike, 1965, стр. 606]. Може да се използва и следната процедура [Н. Хастингс, Дж. Пикок, 1980]:

Нека R_i са независими и равномерно разпределени случайни числа $U(0,1)$. Тогава е валидна следната апроксимация за нормално разпределени случайни числа $N(0,1)$

$$N(0,1) \approx \frac{\sum_{i=1}^k R_i - k/2}{(k/12)^{1/2}},$$

където точността на апроксимацията расте с нарастването на k . За практически цели е достатъчно да се приеме $k = 12$, тъй че

$$N(0,1) \approx \sum_{i=1}^{12} R_i - 6.$$

3. Правим трансформацията $Z = (N - \gamma) / \delta$.

4. Определят се случайни числа X от разпределението на Джонсън в зависимост от неговия тип по формулите:

$$S_L: X = \xi + \lambda \exp(Z),$$

$$S_B: X = \xi + \lambda / (1 + \exp(-Z)),$$

$$S_U: X = \xi + \lambda \operatorname{sh}(Z).$$

Друга алтернатива е да бъдат използвани емпирични разпределения от фамилията на Пирсън [М. Кендалл, А. Стюарт, 1966]. Последните имат неудобството, че фамилията се състои от 7 типа функции, което затруднява

практическото ѝ приложение при определяне параметрите на разпределенията и генериране на псевдослучайни числа, докато фамилията разпределения на Джонсън си състои само от три функции, обхващащи цялата равнина (β_1, β_2) , виж Фиг. 5, където

$$\text{- асиметрия} \quad \beta_1 = \frac{\mu_3^2}{\mu_2^3};$$

$$\text{- ексцес} \quad \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

Фиг. 5. Разпределения в равнината (β_1, β_2)

Друг подход към проблема за използване на разпределения, отличаващи се от нормалното е предложен от Бърр [I.W. Burr, 1942]. Възможно е да се определи функцията на разпределението, а след това да се получи плътността на разпределението чрез диференциране.

Разпределенията на Бърр могат да съответствуват приблизително на много други добре известни вероятностни разпределения като нормално, гама, логаритмично-нормално, експоненциално, както и Пирсъновите семейства от тип I, II, V, VII, IX, X, XII. Като се има предвид широк спектър от форма и мащаб на параметрите на разпределението на Бърр, то може да има важна роля в моделиране на

надеждността, анализ на риска и контрол на статистическия процес.

Предлага се функцията

$$F(x) = 1 - \frac{1}{(1 + x^\alpha)^\beta}, \quad 0 \leq x \leq \infty, 0 < \alpha, \beta. \quad (5)$$

Изчисленията параметри на разпределенията на Бърр е сложна задача и използването на конвенционални методи, като например оценки по Метода на Максималното Правдоподобие (MLE) и метода на моментите (MM) не става директно. Определянето им се извършва с помощта на числени методи и компютърни програми. За целта може да се използва популярната програма MatLab.

Генерирането на псевдослучайни числа от фамилията разпределения на Бърр се извършва с помощта на формулата

$$X = (U^{-1/\beta} - 1)^{1/\alpha},$$

където U са равномерно разпределени случайни числа $U(0,1)$.

ЗАКЛЮЧЕНИЕ

Често изследователите в областта на екологията и опазването на околната среда се срещат със случайни величини, които не следват нормалното /Гаусово/ разпределение, което по-скоро се явява изключение, отколкото правило. Тяхното апроксимиране или провеждането на числени експерименти с тях изисква използване на други фамилии от статистически разпределения, каквито са разпределенията на Джонсън, Пирсън и Бърр. Настоящата статия представя начините за генериране на псевдослучайни числа от тези разпределения.

Тези, които се интересуват от по-задълбочено запознаване с проблемите за генериране на случайни числа и прилагане на метода на имитационното моделиране /наричан още метод Монте Карло/, могат да използват книгата на Джентъл [J.E. Gentle, 2005], в която има обширна библиография по тази тема. Книгата е предназначена да бъде едновременно справочник и учебник. Може да се използва като основен текст или допълнителен текст за различни специалисти и изследователи..

ЛИТЕРАТУРА

- [1] Илиев Л. (1980). Математика и развитие - методологични проблеми. Наука и изкуство, София.
- [2] Кендалл М., А. Стюарт (1966). Теория разпределений. Наука, Москва.

[3] Колмогоров А.Н. (1941). О логарифмически нормальном законе распределения размеров частиц при дроблении. Доклады АН СССР, т. 31, 99-101.

[4] Дъбов Л. (1976). Метеорологичен годишник 1971. Управление Хидрология и Метеорология, София.

[5] Петрова А., А. Садовски (1984). Статистически анализ на количествени свойства на почви от планинските райони в България. Internat. Meeting on Classification and Management of Soils in mountain regions, 25.- 29 Sept. Sofia, 254-264.

[6] Форсайт Дж., М. Малкълм, К. Молър (1986). Компютърни методи за математически пресмятания. Наука и изкуство, София.

[7] Хастингс Н., Дж. Пикок (1980). Справочник по статистическим распределениям. Статистика, Москва.

[8] Burr I.W. (1942). Cumulative frequency functions. The Annals of Mathematical Statistics, Vol. 13, Number 2, 215-232.

[9] Gentle J.E. (2005). Random Number Generation and Monte Carlo Methods. Springer.

[10] George F., K.M. Ramachandran (2011). Estimation of Parameters of Johnson's System of Distributions. Journal of Modern Applied Statistical Methods: Vol. 10: Iss. 2, Article 9.

[11] Hill, I. D., R. Hill, R.L. Holder (1976). Fitting Johnson curves by moments. Applied Statistics, 25 180-189.

[12] Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. Biometrika, 36, 149-176.

[13] Pike M.C. (1965). Algorithm 267: Random normal deviate. Comm. ACM, v. 8, 10, p. 606.

[14] Swain, J.J., S. Venkatraman, J.R. Wilson (1988). Least-squares estimation of distribution functions in Johnson's translation system. Journal of Statistical Computation and Simulation, 29, 271-297.

Проф. Александър Н. Садовски

Академик на МАНЕ

Тел. 02-981-5586

e-mail: alexsad@hotmail.com

Prof. Alexander N. Sadovski

Academician of the IEAS

Phone: 02-981-5586

e-mail: alexsad@hotmail.com