DYNAMICS OF THE RODENT COMMUNITY IN THE CHIHUAHUAN DESERT OF NORTH AMERICA III. FREQUENCY DISTRIBUTIONS AND PREDICTION OF VARIABLES*

ZENG Zongyong YANG Yuemin SONG Zhiming LUO Mingshu

(*Department of Biology, Sichuan University, Chengdu, 610064*)

Abstract

The frequency distributions of the 6 ecological variables of the rodent community in the Chihuahuan Desert of southwestern North America were verified using the chi-square test and Kormorgorov test. And moving averages and exponential smoothing methods were employed to predict dynamics of the 6 community variables. The results reveal that: 1) joint population density, biomass, species evenness and biomassevenness had a normal distribution; 2) frequency distribution of number and diversity of species was skewed to the right (skewness < 0), no common theoretical probability distributions could describe these two variables; 3) single moving averages was a better model for predicting fluctuations of species diversity and double moving averages for that of species evenness; 4) Single exponential smoothing fitted dynamics of the other variables well.

Key words Chihuahuan desert; Chi-square test; Exponential smoothing; Moving averages; Frequency distribution; Rodent community

Community variables are often used to characterize a community. In this sense, the values of the community variables in a sampling are regarded as the estimates of the corresponding parameters of the community. But responding to fluctuations of its biotic and abiotic environment, population densities of the species in a community vary along the time dimension. The variations of population densities are integrated with to form the fluctuations of parameters of the community which is consisted of the species. So the values of the community variables in a sampling are just the values of random variables. We should be careful when we study a community statistically using ecological variables, e. g. species composition, diversity, or space patterns, just by a sampling from the community (Zeng, 1994). Even though a sample which includes only a unit can be used to draw the estimates of the community parameters, the estimates are biased (Sokal et al., 1981). But if we identify the probability distributions of the community variables, we can obtain not only the expectations of the community parameters, but also

---

the standard deviations of the corresponding variables. Furthermore, if a variable is distributed normally, we can obtain their means, which are the unbiased estimates of the corresponding parameters, and confidence intervals from a sample which contain more than two sample-units theoretically . In this paper the chi-square test and Kormorgorov test of hypotheses of frequency distribution are applied to identify frequency distribution of the 6 community variables, including number of species, joint population density, biomass, Shannon index of species diversity, and evenness of species diversity and biomass-allocation evenness. After that moving average and exponential smoothing models (Sullivan et al. , 1977, Legendre et al. , 1983, Jassby et al. , 1990) is used to predict dynamics of the 6 variables.

METHODS

1. The data base

The field data were collected by mark-recapture, for details of the study site and rodents see Brown et al. (1985) and Zeng (1987a, b). In this paper we considered 6 main community variables, i. e. number of species, joint population density, biomass, Shannon index of species diversity, and evenness of species diversity and biomass-allocation evenness among the 17 species (Zeng et al. , 1994). The 6 variables are defined as the follows (in this section, i represents the ith month):

1) Number of species (S_i): number of species of rodents (unit: no. of species/month).

2) Joint population density (N_i): the sum of population densities of the 17 species of rodents (unit: individuals/ha). Note: in Zeng et al. , (1994) this variable was named "number of individuals".

3) Biomass (B_i): the sum of biomass of the 17 species of rodents (unit: g/ha).

4) Shannon index of species diversity (H_i).

5) Species evenness (E_i).

6) Biomass-allocation evenness (BE_i): was calculated by

$$BE_i = HB_i / (\log S_i) \qquad (1)$$

here, HB was Shannon index of biomass-allocation among 17 species in the ith month:

$$HB_i = -\sum_{j=1}^{17} p_{ij} \log p_{ij} \qquad (2)$$

in which p_{ij} is the proportion of biomass of the jth species in the community biomass B_i in the ith month.

For more details of the 6 community variables see Zeng (1994).

2. The chi-square test and kormorgorov test

Because all of the 6 variables fluctuated, it is not appropriate to determine parameters of a rodent community by a unique census. We should know at first frequency distributions of the 6 community variables. Here the chi-square test and Kormorgorov test are employed to identify frequency distributions of the 6 variables.

The Chi-square test is a common method to verify if a random variable has a certain theoretical distribution or not. The null hypothesis for testing is that a random variable has a normal distribution. First, the data of each variable were pooled to obtain the observed frequency. Here, O_i represents observed frequency of the ith group of the pooled data. Second, the mean and corrected standard deviation of each variable were calculated and regarded as the two parameters of a normal distribution, by which the corresponding theoretical probability of the ith group can be drawn and multiplied then by the total sampling number of months to give the theoretical frequency T_i of each group. Third, the Chi-square test was applied to test the null hypothesis that frequency of a variable had a normal distribution. The statistics being calculated was $X^2 = \sum_{i=1}^{K} (O_i - T_i)^2 / T_i$, here $\underline{k}$ is the number of groups. Comparison between and the critical value $X_{0.05}^2$ gave if the null hypothesis could be accepted, 0.05 is the level of significance. If accepted, frequency of the variable had a normal distribution at the 0.05 level of significance (Sokal et al., 1981).

The Kormorgorov test is another method to verify frequency distribution of a random variable. If a random variable is approaching a certain theoretical distribution, the difference between its observed and theoretical frequency must be less enough to accept the null hypothesis. Here, the null hypothesis is still that the frequency distribution of a certain community variable was normal. The statistics for testing is $D_n = \sup [F_n(x) - F(x)]$, in which $F_n(x)$ is the cumulative frequency of the sample and obtained by cumulatively adding O_i, $F(x)$ the theoretical cumulative frequency and obtained by cumulatively adding T_i. The comparison of $F_n(x)$ and $F(x)$ of each pooled group gives D_n, which is the maximum one of all $[F_n(x) - F(x)]$, $\underline{n}$ is the number of raw data points and x is the median value of each pooled group. $F_n(x) - F(x)$ reflects the difference between the observed frequency and the theoretical frequency. We have the statistics $\lambda = \sqrt{n} \, D_n$, which distribution is near $Q(\lambda)$. The density function of $Q(\lambda)$ is $\sum_{k=-\infty}^{\infty} (-1) e^{-2k^2\lambda^2}$ (when $\lambda > 0$) or 0 (when $\lambda \leqslant 0$). If $1 - Q(\lambda)$ has a probability which is less than 0.05, the null hypothesis that the variable has a normal distribution should be rejected. Because $F_n(x) - F(x)$ is too large. The larger D_n leads to larger $\sqrt{n} \, D_n$, which leads to a larger $Q(\lambda)$ and a smaller $1 - Q(\lambda)$. Otherwise the null hypotheses should be accepted. The Kormorgorov test compares the difference between each observed frequency of the pooled groups and the corresponding theoretical frequency, so it is more accurate than the chi-square test (Sokal et al., 1981).

3. Prediction of the 6 community variables

In time series data there is generally series correlation between successive data points. In this case the greater errors resulting from regression analysis make regression equations unbelievable. So the moving average or exponential smoothing, which are two methods of the time series analysis, are better choices for fitting the time series data

(Sullivan et al. , 1977).

Here, 6 mathematical models, i. e. simple linear regression, single- and double moving averages, and single-, double- and triple- exponential smoothing method, were applied to predict the temporal variation of the 6 variables.

Suppose X_i is the value of a time series at time i, N is the number of moving data points each time, then the predicted value of the time series at time t is

$$Y_t = \frac{1}{N} \sum_{i=t-N}^{T-1} X_i \qquad (3)$$

So we must have at least N raw data points for predicting the value of the time series at time t. This is the single moving average method. It is suited to those time series in which there is no trends.

If there are trends in a time series, the double moving averages are suitable. Suppose

$$M_t^{(1)} = \frac{1}{N} \sum_{i=t-N+1}^{t} X_i \qquad (4)$$

$M_t^{(1)}$ is the first moving average based on which the second moving average is

$$M_t^{(2)} = \frac{1}{N} \sum_{i=t-N+1}^{t} M_i \qquad (5)$$

Then we get the two parameters of the function for predicting:

$$a_t = 2 M_t^{(1)} - M_t^{(2)} \qquad (6)$$

$$b_t = \frac{2}{N-1} [M_t^{(1)} - M_t^{(2)}] \qquad (7)$$

The function for predicting is

$$Y_{t+1} = a_t + b_t T \qquad (8)$$

here, $\underline{T}$ is the predicting time from t, Y_{t+T} is the predicted value at time $\underline{t+T}$ based on the 2N raw data points. We must have at least 2N raw data points for predicting the value at time $\underline{t+T}$. This is the double moving average method which is suited to those time series in which there is a linear trend. The double moving averages need more raw data than the single moving averages.

The basic model of the exponential smoothing is

$$Y_{t+1} = \alpha X_t + (1-\alpha) Y_{t-1} \qquad (9)$$

here, Y_{t-1} is the last predicted value and when $t=2$, $Y_2 = X_1$, α is the smoothing constant which is between 1 and 0. Y_{t+1} the predicted value at time $\underline{t+1}$. This is the single exponential smoothing. From the above equation we know α is an empirical constant which tells that we put how much weight on the historical data and new data.

From each method we get a predicted series Y_i, then the residual square is

$$RS = \sum (X_i - Y_i)^2 \qquad (10)$$

We prefer to the method which gives the minimum residual squares.

According to the criterion that the residual square was minimum, different methods, different number of data points was used to calculated moving averages, and differ-

ent exponentials was subjected to computation [for the details of the 6 mathematical models see Sullivan et al. (1977) or Legendre et al. (1983)].

RESULTS

1. Frequency distributions of the 6 variables

Figure 1 shows the frequency distribution of the 6 variables. The abscissa represents pooled data groups and the ordinate the frequency (unit: number of month) of the 6 variables.
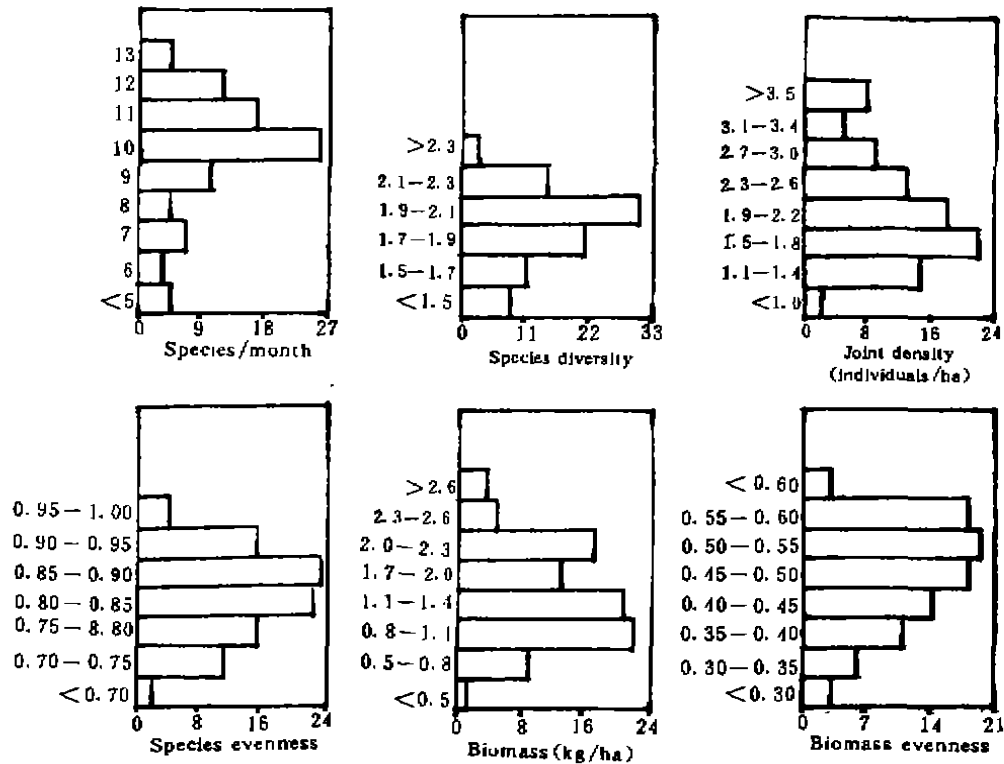

Fig. 1 The frequency distributions of the 6 community variables

The basic statistics reveal the characteristics of the 6 variables clearly (Table 1). The mode of number of species was 10 which means that there were 10 or more species in the community in most of 92 months. Joint population density were skewed to the left and the number of species and species diversity to the right, which implicates that number of individuals of rodents was less than 20 while the number of species greater than 8 and species diversity greater than 1.5 in most of months. Biomass and species evenness were symmetrically distributed with the least skewness. Species diversity had the greatest kurtosis 2.005, which suggests that its most values were between a narrow interval. In fact the Shannon indexes of species diversity in 53 months were between 1.7—2.1, even though the minimum and maximum of this variable were respectively 0.79 and 2.70, which is a quite wide interval.

Table 1　The basic statistics and results of the statistical hypotheses testing
of frequency distribution of the 6 community variables

Community variable	No. of species	Joint density (no. /ha)	Biomass (g/ha)	Species diversity	Species evenness	Biomass evenness
Maximum *	13	38.06	2564	2.71	1.00	0.63
Minimum *	3	6.42	402	0.79	0.63	0.27
Mean *	9.67	21.04	1368	1.88	0.84	0.48
Standard deviation *	2.19	7.5	475	0.32	0.08	0.08
Mode	10	17	1200	2.00	0.85	0.55
Median	10	19	1200	1.90	0.84	0.47
Skewness	−1.029	0.538	0.385	−0.880	−0.320	−0.453
Kurtosis	0.992	−0.496	−0.617	2.005	−0.103	−0.527
	23.30#	5.389	9.600	9.805#	4.169	9.635
Critical value	12.59	7.815	11.07	7.815	9.488	11.07
Degree of freedom	6	5	5	3	4	5

* , From Zeng 1994.

, The null hypothesis that frequency of the variable had a normal distribution was rejected at a 0.05 level of significance.

The results of the chi-square test suggest that frequency of join population density, biomass, species evenness and biomass-allocation evenness had a normal distribution (Table 1). Their smaller values were originating from the insignificant difference between the observed and theoretical frequency and resulted in accepting the null hypotheses. Frequency of the number of species and species diversity was not normally distributed. The Kormorgorov test gives the same results. Both their skewness and kurtosis were greater than those of the other 4 variables. We also tested the frequency distributions of the two variables when the null hypothesis is that they had a Poisson distribution. The null hypothesis was also rejected. Maybe it is impossible to describe the two variables by common theoretical probability distributions.

2. Prediction of the 6 variables

The results suggest that prediction of species diversity by single moving averages and species evenness by double moving averages gave the minimum residual square. The other 4 variables were fitted appropriated by single exponential smoothing (Table 2).

In Fig. 2, dotted lines are predicted values of the 6 variables, while the solid lines are the actual values. We can make comparison of residual square between simple regression and moving averages or exponential smoothing to draw a conclusion that time series method fitted the real data perfectly.

DISCUSSION

Dynamics of the rodent community in the Chihuahuan Desert tell us that it is not appropriate to determine community parameters by sampling a community just once because the parameters fluctuated greatly along the time dimension. We should sample a community randomly and repeatedly to give both expectations of the community vari-

ables and their confidence intervals.

Table 2   The number of data points being used to calculate moving averages and exponentials

when the residual square is minimum in simple regression, moving

averages and exponential smoothing of time series of the 6 variables

A. Moving averages and simple regression

Community variable	Single		Double		Simple regression residual square
	No. of moving data points	Residual square	No. of moving data points	Residual square	
No. of species	20	2.968	24	3.112	3.377
Joint density	5	26.93	6	44.81	45.87
Biomass	6	170636	18	250838	219579
Species diversity	30 * *	0.0495 *	24	0.0599	0.0797
Species evenness	48	0.0422	24 * *	0.0042 *	0.0059
Biomass evenness	2	0.0030	24	0.0038	0.0033

B. Exponential smoothing

Community Variable	Single		Double		Triple	
	α	Residual square	α	Residual square	α	Residual square
No. of species	0.5 #	2.912 *	0.2	3.261	0.15	3.427
Joint density	0.75 #	21.33 *	0.3	24.03	0.2	24.92
Biomass	0.7 #	129169 *	0.2	150065	0.2	158684
Species diversity	0.6	0.0526	0.3	0.0599	0.2	0.0797
Species evenness	0.5	0.0049	0.2	0.0055	0.2	0.0058
Biomass Evenness	0.7 #	0.0027 *	0.3	0.0030	0.1	0.0033

* ， The minimum residual square

* * ， The number of data points being used to calculate moving averages when residual square was the minimum

， The exponential smoothing constant when residual square was the minimum

The premise of parameter estimates of small samples is that the random variable in our question has a normal distribution. In this study we make it sure that joint population density, biomass, evenness of species and biomass-interspecific allocation were distributed normally, which indicating that we can obtain the estimates of means and their confidence intervals of the 4 parameters by a small sample (theoretically, the number of samples unit$>$2, see Sokal et al. , 1981).

However, we cannot do it with number of species and species diversity in the same way because of their skewed frequency distributions. Their skewed frequency distribution can be found from the absolute values of both their skewness and kurtosis which are greater than those of other variables and 0 (both skewness and kurtosis of a random variable that has a normal distribution equal to 0). The negative skewness means that the frequency of both variables was skewed to the right. It is difficult to identify their frequency theoretically. Of course we can do it with a large sample if only the number of sample units is greater than 50. In this study, estimates of the 6 variables are respectively (p=0.05): 9.67$\pm$0.47 (species), 21.04$\pm$1.59 (individuals/ha), 1.37$\pm$0.1 (kg/ha), 1.88 $\pm$ 0.07, 0.84 $\pm$ 0.02 and 0.48 $\pm$ 0.02. Biomass evenness is much less than species evenness which means uneven biomass allocation among the 17 species populations and furthermore leads to the conclusion that there is a monopolizing energy allocation patterns among the species in this desert rodent community (Zeng et al. ,1994). We can also just calculate their mathematical characteristics, e. g. mean, variance or standard deviations, mode median, skewness and kurtosis from a small sample if it is hard to get a large sample. From these mathematical characteristics we can get some information on central tendency, dispersion or symmetry of frequency distribution of the vari-
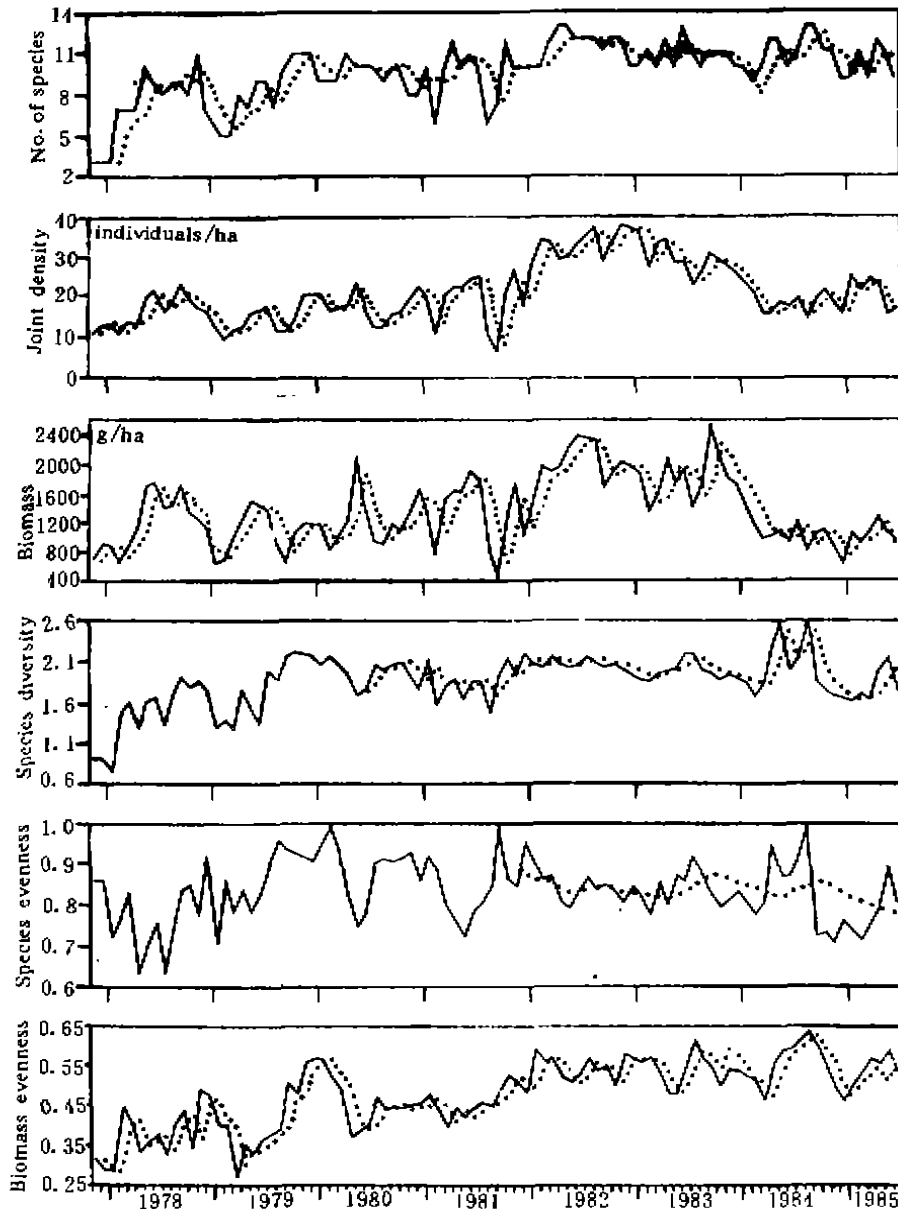
ables being studied.

Fig. 2 The time series (——) of the 6 variables and their predicted
series (...) by moving averages or single exponential smoothing

Moving averages or exponential smoothing are better models for fitting the ecological variables of this community because the parameters are changing in predicting every coming point. They can give much better results for prediction of a quite short period than regression method. Thirty and forty-eight data points, respectively, are needed at least for predicting species diversity and species evenness. It shows that the historical data points are more important in predicting than new ones. The smoothing constants of

joint density, biomass and biomass-allocation evenness are 0.7 or more, that of number of species is 0.5. It reveals that current data points of the number of species have been given less weight in predicting than the other 3 variable.

REFERENCE

Brown J H, Munger J C. 1985. Experimental manipulation of a desert rodent Community; food addition and species removal. *Ecology*, 66; 1545—1563.

Jassby A D, Powell T M. 1990. Detecting changes in ecological time series. *Ecology*, 71; 2044—2052.

Legendre L, Legendre P. 1983. Numerical ecology. Elsevier Scientific Publishing Company, NY.

Sokal R R, Rohlf F J. 1981. Biometry. W. H. Freeman. San Francisco, California, USA.

Sullivan W G, Claycombe W W. 1977. Fundamentals of forecasting. Reston Publishing Company, VA.

Zeng Z, Brown J H. 1987a. Population ecology of a desert rodent; *Dipodomys merriami* in the Chihuahuan Desert. *Ecology*, 68; 1328—1340.

Zeng Z, Brown J H. 1987b. A method for distinguishing dispersal from death in mark—recapture studies. *Journal of Mammalogy*, 68; 656—665.

Zeng Z. 1994. Dynamics of the desert rodent community in the Chihuahuan Desert of North America I. Interannual fluctuations and trends. *Acta Theriologica Sinica*, 14 (1); 24—34.

Zeng Z, Yang Y, Song Z. 1994. A comparative study of biomass dynamics of species populations of the desert community in the Chihuahuan Desert. *Zoological Research*, 15; 32—41.

中文摘要

北美 CHIHUAHUAN 荒漠啮齿动物群落动态
III. 变量的频次分布和预测

曾宗永　杨跃敏　宋志明　罗明澍
（四川大学生物系，成都，610064）

本文用 χ^2 方和柯尔莫哥洛夫检验分析了北美 Chihuahuan 荒漠啮齿动物群落6个生态学变量的频次分布，并用移动平均和指数平滑方法拟合了这些变量的动态变化。结果显示：1）结合种群密度、生物量、物种均匀性和生物量均匀性服从正态分布；2）物种数和物种多样性的频次为向右偏斜的分布，无法用常见的理论分布来近似表示；3）单移动平均和双移动平均分别较好地描述了物种多样性和物种均匀性；4）其余4个变量可用单指数平滑来较好的描述。

关键词　Chihuahuan 荒漠；χ^2 方检验；指数平滑；移动平均；频次分布；啮齿动物群落