Fitting competing models and evaluation of model parameters of the abundance distribution of the land snail *Vallonia pulchella* (Pulmonata, Valloniidae)


*Oles Honchar Dnipro National University, Dnipro, Ukraine
**Mykolaiv National Agrarian University, Mykolaiv, Ukraine
***Bohdan Khmelnitsky Melitopol State Pedagogical University, Melitopol, Ukraine

**Received 15.03.2018**
**Accepted 25.04.2018**

This paper summarizes the mechanisms behind the patterning of the intra-population abundance distribution of the land snail *Vallonia pulchella* (Müller, 1774). The molluscs were collected in recultivated soil formed on red-brown clays (Pokrov, Ukraine). Data obtained in this study reveal that *V. pulchella* population abundance ranges from 1 to 13 individuals per 100 g of soil sample. To obtain estimates of the mean, three models were used: the model of the arithmetic mean, the Poisson model and a log-normal model. The arithmetic mean of the occurrence of this species during the study period was 1.84 individuals/sample. Estimation of the average number of molluscs in one sample calculated using the Poisson model is lower and equals 1.40 individuals/sample. The distribution of the number of individuals in a population was described by the graphics “rank – abundance”. The individual sample plot sites with molluscs may be regarded as equivalents of individual species in the community. For the analysis, the following models were used: broken sticks model, niche preemption model, log-normal model, Zipf model, and Zipf-Mandelbrot model. Applying the log-normal distribution gives a lower estimate of the mean density at 1.28 individuals/sample. Median value and mode is estimated at 1.00 individuals/sample. The Zipf-Mandelbrot model was shown as the most adequate to describe distribution of the *V. pulchella* population within the study area. The Zipf-Mandelbrot model belongs to the family of so-called non-Gaussian distributions. This means that the sample statistics do not possess asymptotic properties and by increasing the sample size, they tend to infinity, and are not close to the values of the general population. Therefore, the average value of the random variable that describes the non-Gaussian distribution has no statistical meaning. From an environmental point of view, this means that within the study area the capacity of the habitat is large, and for some combination of environmental conditions the rapid growth of the abundance of a given species is possible.

**Keywords**: abundance; intra-population structure; land snails

**Introduction**

There are plenty of publications devoted to habitat selection by land snails based on the investigation of mollusc communities from spatially different biotopes which differ in vegetation cover, soil type, moisture level (Millar & Waite, 1999; Martin & Sommer, 2004; Müller et al., 2005; Weaver et al., 2006). The number of invertebrates on a particular plot changes randomly. It is impossible to predict in advance for any moment of time at a given plot which of the factors will operate within the optimum range and which at an extreme range for any given species of invertebrate (Brygadyrenko, 2015a). The calcium content in the soil, pH, the soil texture (Ondina et al., 2004), and the content of exchangeable cations and aluminum (Ondina et al., 1998) are the most important among the edaphic factors which affect the molluscs. The soil moisture plays a significant role in forming ecological niches of molluscs (Nekola, 2003). However, P. Ondina et al. (2004) noted the restrictions of the importance of soil moisture data at any given moment because of the considerable changeability of this parameter. To help address this situation phytointicication data application is an appropriate method for assessment of land snails’ ecological characteristics and community properties (Horsák et al., 2007; Dvořáková & Horsák, 2012). Ellenberg phytointicication scales have been effectively used to explain the habitat preferences of the land snail *Vertigo geyeri* in Poland and Slovakia (Schenkovi et al., 2012). Interaction of litter macrofauna with herbaceous plants has been investigated in natural forests of Ukraine’s steppe zone (Brygadyrenko, 2015b) and in poplar plantations within city parks (Faly & Brygadyrenko, 2014). The importance of soil factors in the spatial distribution, abundance and diversity of mollusc communities has been revealed on a large-scale level of research (Nekola & Smith, 1999; Juřičková et al., 2008; Szybiak et al., 2009). The issue of the spatial scale and the hierarchy of factors acting on molluscs are of special interest (Nekola & Smith, 1999; Bohan et al., 2000; McClain & Nekola, 2008; Myšák et al., 2013).

Habitat is attributed by the presence of a certain range of resources and ecological conditions which make it possible for a given species to occupy, survive and reproduce in a given area (Hall et al., 1997). For identification of the environmental characteristics which make a territory appropriate for the existence of a species, it is important to investigate the selection of habitat types (Calenge & Basille, 2008).

The mollusc *Vallonia pulchella* (Muller, 1774) is Holarctic species distributed around the world at high latitudes. In Europe it prefers wetter habitats such as wet grasslands and swamps, as well as dry dunes and meadows (Kerney & Cameron, 1979). In Ukraine it is found in moderately dry and wet grasslands habitats (Gural-Sverlova & Gural, 2012).
There have been few publications on the results of aut- and demeological research on the land snail *V. pulchella*. In the 1930s, research was carried out on the reproduction of this mollusc (Whitney, 1937, 1941), which was continued in this century (Kuznik-Kowalska & Proków, 2016). It was revealed that the density of *V. pulchella* is affected by many soil physico-chemical properties, especially calcium and magnesium concentration and pH of water extract (Hermida et al., 1993, 2000). Furthermore, interrelation was found between the abundance of *V. pulchella* and pH and electrical conductivity of the water in marshy lowland central Scandinavia (Schenkova et al., 2015). Local trends, as well as the mosaic nature of the organization of the soil body determine the structure of the vegetation cover. This explains the role of indicators in the organization of ecological niche of mollusc *V. pulchella*, the dimension of which exceeds the size of the ecological space of an individual of this species. Soil mechanical impedance, as well as phytodegradation indicators of vegetation, determines the peculiarities of marginality and specialization of the ecological niche of *V. pulchella*. It is the measurement of these indicators that makes it possible to draw a map of habitat preferences for *V. pulchella*. Thus, ecological niche optima may be presented by integral variables such as marginality and specialization axes and may be plotted in geographic space by means of habitat suitability index reproduction (Yorkina et al., 2018).

Thus, the main aim of our work is to select the optimal model and to evaluate model parameters of the abundance distribution of the land snail *V. pulchella* (Müller, 1774) (Gastropoda, Pulmonata, Valloniidae).

**Material and methods**

The studies were conducted at the research station of Dnipro Agrarian and Economic University in Prikov in June 2012. Sampling was carried out on a type of technogenic soil formed on red-brown clays (geographic coordinates of the southwest corner of the polygon are 47°38'55" N, 34°08'33" E) (Zhukov et al., 2013; Zverkovskiy et al., 2017; Kharytonov et al., 2018; Yorkina et al., 2018). According to WRB 2007 (IUSS Working group WRB, 2007), the examined soil belongs to the RSG Technosols. A perennial legume-grass agrophytocoenosis was cultivated at this research station between 1995 and 2003, following which the process of naturalization of vegetation began (Zhukov et al., 2017).

The area within which the samples were taken consisted of seven transects of 15 samples each. The sampling points formed a regular grid of a mesh size of 3 m. From the center of each sampling site a soil sample weighing 100 g was collected. Each sample was examined in the laboratory with a *10* binocular microscope MBS-9 and the number of live specimens of *V. pulchella* was noted.

To obtain estimates of the mean, the three models were used: the model of the arithmetic mean, the Poisson model and a log-normal model. Estimation of the mean from assumptions about the distribution of the random variable according to Poisson was obtained by the formula (Shebanin et al., 2008):

\[
D = - \ln \frac{n_0}{n}
\]

(1)

where D is the mean estimation; n0 is number of samples without the presence of the species; n is total number of samples.

Statistical calculations were performed using Statistica 7.0 program (Khalafayan, 2007) and a 95% confidence interval for the mean estimation obtained using different models has been calculated on the basis of the bootstrap approach using bootES package (Kirby & Gerlanc, 2013). The distribution of the number of individuals in a population may also be described by the graphics "rank – abundance", which are often used in community ecology (Whittaker, 1965). In this case, the individual sample plot sites with molluscs may be regarded as equivalents of individual species in the community. For the analysis in this case, the following model can be used:

- broken sticks model (MacArthur, 1957):
  \[
  \hat{a}_b = \frac{N}{S} \frac{1}{\gamma}
  \]
  (2)

- Motomura model (the Whittaker niche preemption model) (Motomura, 1932):
  \[
  \hat{a}_m = N \alpha (1 - \alpha)^{\gamma}
  \]
  (3)

- log-normal model (Preston, 1948, 1962):
  \[
  \hat{a}_l = \exp[\log(\eta) + \log(\sigma)\Phi]
  \]
  (4)

- Zipf model (Zipf, 1949):
  \[
  \hat{a}_z = Np r^{\gamma}
  \]
  (5)

- Zipf-Mandelbrot model (Mandelbrot, 1983):
  \[
  \hat{a}_z = Nc (r + \beta)^{\gamma}
  \]
  (6)

where \( \hat{a} \) is the expected abundance of species of rank \( r \); \( S \) is the number of species; \( N \) is the number of individuals; \( \Phi \) is a standard normal distribution function; \( p \) is the estimated proportion of the most abundant species; \( \mu, \sigma, \gamma, \beta \) and \( c \) are the parameters in each model.

The degree of adequacy of the model was evaluated using Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). The best model has the lowest AIC and BIC.

Plant and soil data used for distance matrix calculation were discussed previously in our publication (Yorkina et al., 2018). Statistical calculations were performed with the help of the Statistica 7.0 program and the project for statistical computations R (www.r-project.org) using vegan (Oksanen et al., 2017).

**Results**

In examining 105 samples, 193 *V. pulchella* were detected. Thus, the arithmetic mean of the occurrence of this species in sod lithogenic soils on red-brown clays during the study period was 1.84 individuals per 100 g soil sample (Table 1). Thus, the observed distribution of *V. pulchella* abundance in one sample is adequately described by the Poisson law (Kolmogorov-Smirnov test: \( d_{KS} = 0.089, P > 0.05 \)) (Fig. 1).

This type of discrete distribution is described by a random flow (Puza-chenko, 2004; Shebanin et al., 2008). Estimation of the average number of molluscs in one sample calculated using the Poisson model is lower and equals 1.40 individuals/sample (Table 1).

Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>D</th>
<th>D_{boot} ± SD</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arithmetic mean</strong></td>
<td>1.84</td>
<td>1.70 ± 0.23</td>
<td>1.35 - 2.27</td>
</tr>
<tr>
<td><strong>Poisson distribution</strong></td>
<td>1.40</td>
<td>1.43 ± 0.20</td>
<td>1.08 - 1.85</td>
</tr>
<tr>
<td><strong>Log-normal distribution</strong></td>
<td>1.28</td>
<td>1.26 ± 0.14</td>
<td>1.03 - 1.52</td>
</tr>
</tbody>
</table>

Notes: D – density estimates of the population; D_{boot} – bootstrap-based density estimates of the population; SD – standard deviation; CI – confidence interval.

The pronounced asymmetric character of the distribution makes the arithmetic mean an insignificant value of the general population mean. It is quite logical that applying the log-normal distribution gives a lower estimate of the mean density at 1.28 individuals/sample (Table 1). This value is closest to the median value and mode, which is estimated at 1.00 individuals/sample.

![Fig. 1. Histogram of the empirical distribution of V. pulchella per 100 g of soil sample (dashed line marks the theoretical Poisson distribution for \( \lambda = 1.40 \)](image-url)

The use of curves "ranked-abundance" shows that the Zipf-Mandelbrot model is the most adequate to describe the abundance of the

population of *V. pulchella* within the study area, as the Akaike information criterion for this model is the smallest among all least studied (Table 2).

### Table 2

Statistics for the various models of distribution “rank-abundance” experimental data

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviation Information criterion</th>
<th>Akaike (AIC)</th>
<th>Bayesian (BIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broken sticks</td>
<td>r = 0.15, p = 0.014</td>
<td>232.29</td>
<td>232.29</td>
</tr>
<tr>
<td>Niche preception</td>
<td>r = 0.15, p = 0.016</td>
<td>220.82</td>
<td>223.19</td>
</tr>
<tr>
<td>Log-normal</td>
<td>r = 0.09, p = 0.79</td>
<td>214.77</td>
<td>219.51</td>
</tr>
<tr>
<td>Zipf</td>
<td>r = 0.09, p = 0.03</td>
<td>213.23</td>
<td>217.96</td>
</tr>
<tr>
<td>Zipf-Mandelbrot</td>
<td>r = 0.93, p = 0.120</td>
<td>211.36</td>
<td>218.47</td>
</tr>
</tbody>
</table>

It should be noted that the Zipf-Mandelbrot model belongs to the family of so-called non-Gaussian distributions (Khaytun, 2005). This means that the sample statistics do not possess asymptotic properties and by increasing the sample size, they tend to infinity, and are not close to the values of the general population. Therefore, the average value of the random variable that describes the non-Gaussian distribution has no statistical meaning. From an environmental point of view, this means that within the study area the capacity of the habitat is large, and for some combination of environmental conditions the rapid growth of the abundance of a given species is possible.

The Mantel test allows us to compare the correlation matrices among themselves (Table 3). The Mantel test for matrices of distance measures based on the number of molluscs (Euclidean distance) and measured ecological characteristics (Euclidean distance) indicates a reliable correlation of these matrices (r = 0.15, P = 0.02). Mantel tests indicate that in a complex of environmental indicators that affect molluscs, a key role is played by edaphic factors (positive correlation) and phytoindication indicators (negative correlation). The space (the matrix of geographical distances) does not play an important role in the variability of the molluscs' density.

### Table 3

Correlations of distance matrices (Mantel test)

<table>
<thead>
<tr>
<th>Factor</th>
<th>All environmental indicators</th>
<th>Edaphic indicators</th>
<th>Phytoindication scales</th>
<th>Spatial coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r = 0.15, p = 0.014</td>
<td>r = 0.15, p = 0.016</td>
<td>r = 0.11, p = 0.05</td>
<td>r = 0.08, p = 0.04</td>
</tr>
<tr>
<td></td>
<td>r = 0.09, p = 0.79</td>
<td>r = 0.09, p = 0.03</td>
<td>r = 0.04</td>
<td></td>
</tr>
<tr>
<td>A partial Mantel test with control variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edaphic indicators</td>
<td>r = 0.13, p = 0.097</td>
<td>r = 0.06, p = 0.016</td>
<td>r = 0.09, p = 0.02</td>
<td></td>
</tr>
<tr>
<td>Phytoindication scales</td>
<td>r = 0.16, p = 0.009</td>
<td>r = 0.16, p = 0.009</td>
<td>r = 0.07</td>
<td></td>
</tr>
<tr>
<td>Climate</td>
<td>r = 0.10, p = 0.015</td>
<td>r = 0.08, p = 0.04</td>
<td>r = 0.08</td>
<td></td>
</tr>
<tr>
<td>Edaphic</td>
<td>r = 0.08, p = 0.004</td>
<td>r = 0.05, p = 0.012</td>
<td>r = 0.05</td>
<td></td>
</tr>
<tr>
<td>Spatial coordinates</td>
<td>r = 0.12, p = 0.014</td>
<td>r = 0.05, p = 0.09</td>
<td>r = 0.09</td>
<td></td>
</tr>
</tbody>
</table>

It should be noted that the information value for describing the variation in the number of molluscs has a complex of phytoindication indicators, and does not separate the matrices of climatic and edaphic indicator values. The use of edaphic indicators as a control variable in the partial Mantel test leads to a change in the correlation coefficient of the distance matrices in terms of the number of molluscs and the matrix of ecological indicators (r = 0.15, P = 0.02). This confirms the result that edaphic factors and phytoindication indicators mark opposite tendencies of influence of environmental factors on the mollusc population, while the Mantel test between the matrices of edaphic and phytoindicating distances is positive (r = 0.11, P = 0.013).

Phytoindication scales as doers in the partial Mantel test do not affect the correlation of molluscs and both environmental indicators in general and edaphic indicators.

Thus, the analysis of the results of Mantel tests indicates a specific characteristic of the influence of edaphic factors and vegetation on the population of the molluscs *V. pulchella* and the absence of influence of spatial factors on the chosen scale. It is likely that the selected set of indicators can fully describe the spatial variation of the population of *V. pulchella* within the studied test site.

**Discussion**

Distribution of the species in the loci can be described by a number of models. Abundance curves are widely used to describe dominance structures in communities of living organisms. This tool can also be used to describe the statistical distribution of the individuals in a population. The difference consists in the fact that the x-axis will not be delayed in the rank of species abundance and species abundance levels in the locus of a given population. Ideally, the extent to which the actual distribution of a particular model makes it possible to make a choice in favour of one of the theories that explain the structuring of the population (or community) (Pielou, 1975). Geometric series (Motomura, 1932) was proposed to describe the benthic communities of a lake. The parameter k of the distribution is a measure of the complexity of the species composition of the community. According to Tokeshi (1993), Motomura applied a geometric series as the simplest mathematical form for the description of an ecological community, although later this model was interpreted as a reflection of the distribution of resources among species (Ferreira & Petrere-Jr., 2008). From this viewpoint, the distribution should be described by a situation in which the dominant species uses a proportion k of the whole initially available resource, leaving accessible proportion 1-k. The next dominant species uses the same fraction k of the remaining available resource. This goes for all species of the community. May (1975) believes that the geometric series is a situation in which all species of a community are energy equivalent and value of ecological interactions is proportional to the abundance of species – more abundant species require more energy from the system. Geometric series take into account only the number of species and neglect the influence of the size of the animals on the energy needs of a species (Ferreira, Petrere-Jr., 2008).

Logarithmic series distribution proposed by Fischer et al. (1943) slightly resembles a hyperbola decreasing with increasing number of species and can be predicted from the expression:

\[ S/N = (1-x)/x \times [-\ln (1-x)] \]

where \( ax^n \) represents the number of species with \( n \) number of individuals. The number \( x \) can be obtained from the equation

\[ S/N = \frac{1}{1-x} \times \ln (1-x) \]

and in practice is in the range 0.9–1.0 (Magurn, 2004). The constant \( a \) is independent of the sample size and may be viewed as an index of diversity, while remaining a robust assessment even in the case where the data is not very well matched logarithmic series (Ferreira & Petrere-Jr., 2008).

Truncated lognormal distribution was proposed by Preston (1948). With the species abundances in the histogram in logarithmic scale, the author obtained a curve that is a good description of a large number of community data. The R classes resulting logarithms are called octaves, wherein each class represented twice the number of species of the previous class (1, 2, 4, 8, 16 ...). The lognormal model has two parameters: \( \mu \) – mean value of the logarithmic score abundance and \( \sigma \) – standard rating average logarithmic abundance deviation (Ferreira & Petrere-Jr., 2008). For real samples distribution is truncated on the left.

The area behind this point represents species that is not included in the sample, and it decreases with increasing sample size. MacArthur (1957) suggested that the ecological niches in a community can be compared with a stick of unit length, at which n–1 are randomly generated points form n segments of length proportional to the abundance of each species in the community. MacArthur’s broken stick model does not contain parameters for evaluation.

Model Zipf (Zipf, 1949) has two parameters – abundance estimation dominant form γ and the parameter β, which indicates the probability of occurrence of the species in the community. The Model Zipf-Mandelbrot (Mandelbrot, 1983) has a third parameter α, which may be regarded as ecological variety of potential environmental or ecological niche diversity.

The geometrical model of the distribution of species abundance is consistent with the hypothesis interception niches Whittaker (1972). It assumes that every species of community sites in decreasing order of abundance (increase in rank) uses a constant (K-th) of the remaining resources of the community. For example, if the strongest competitor (dominant) occupies 70% of the niche space using the appropriate community resources share, the second most important species of able to take a similar share of the remainder of the first niche space, the third type – the same piece of space left over from the first and the second kind, etc. According to this model, the proportion of community resources used by the most abundant species (dominant) is not a special case (the result of the biological and ecological characteristics or accidental circumstances), and reflects the general character of the distribution of niche space between species in different conditions. Hence the community with a higher dominance of the most competitive species must be characterized as having not only fewer resources available for related species, but also a more "rigid" means of distribution of resources among them (i.e. higher K values) that ceteris paribus conditions may affect their species richness.

There are other types (models) of species abundance patterns, a review of which is available in many publications (Whittaker, 1972; Megarran, 2004; Ferreira & Petere-Jr., 2008). They assume a more or less uniform, compared to the geometrical model, allocation of resources and thus more or less marked predominance of one species. The former include, for example, log-normal model and the "broken stick" MacArthur; to the second – hyperbolic model.

According to the log-normal model (for the analysis of biological communities, this was first used by Preston), K values in the species of the first rank higher than the species next few ranks, and in accordance with the hypothesis of random boundaries between niches MacArthur vice versa. At the same time as the lognormal model and MacArthur model typical for communities with a relatively low level of dominance. Therefore, the allocation of resources between the associated views in the three models (hyperbolic lognormal and MacArthur) is related to the level less than the first domination rank species than in geometric.

Particular attention is paid to the types of distribution in the community ecology and population ecology due to the need to open or close a picture of the mechanisms of the relationship of the community or population with the resources offered by the habitat. It is expected that the community or the population is subject to the theoretical concepts, on the basis of which are derived the corresponding types of distributions. The fact that the population size distribution of V. pulchella is not subject to a fully stochastic processes (Poisson distribution model of the broken sticks) indicates that the selected site is not entirely uniform for the species under study. It should be noted that an alternative statistical broken stick model is a little more realistic, which can only be the basis for the assumption of the presence of any processes that may underlie the alternative models.

The best of the studied models to describe the number of mollusks is the model of the Zipf-Mandelbrot. This type of distribution is often found in the description of complex organized systems (Khaytun, 2005).The non-Gaussian nature of this distribution makes it difficult to use traditional statistics to describe the corresponding processes. With respect to the type of population, Mandelbrot-Zipf distribution systems can arise as the result of significant increase in population numbers in the formation of optimal environmental conditions. In other words, the number is not in this situation, a marker of environmental conditions in the sense that the relationship is not proportional. Then a great informative value is simply the fact of the presence (or absence) of the form.

Studies of terrestrial micromolluscs has always been accompanied by significant methodological difficulties related to their small size. For example, the diameter of the shell of many species of *Vallonia Risso*, 1926 does not exceed 2–3 mm (Gerber, 1996). However, the procedure for assessing the abundance of species, using a widespread approach based on test plots, characterized by significant errors and obtained on the basis of such an assessment approach, cannot always be trusted unconditionally. In addition, these estimates may differ even by two-four orders of magnitude and cannot always be clear, these differences are fundamental and are the consequences of violations of population estimation procedure. For example, the population density of *V. pulchella* in alder and oak forests of Belarus was 4–8 individuals/m² (Zemoglyadnukh, 2005), on artificial *Robinia* stands in a recultivated area near the town of Zhovti Vodi it was 5.6 individuals/m² (Kal’bachko & Unkov’ska, 2008), whereas in asl-elm forests in Poland it did not exceed the average of 0.13 individuals/m² (Konarelska-Batara & Błoszyk, 2007), and in floodplain forests of Slovakia it was 0.07 individuals/m² (Čejka & Harnerlík, 2009). On the other hand, Hermida et al. (1993) estimated the average density for three studied populations of *V. pulchella* in Spain (meadow and forest habitats, as well as near the river bank) to be 5.9–10.1 individuals/m², reaching on separate test plots values of the order of 200 individuals/m². In bush willow depressions in Kazakhstan Uvalieva (1990) a density of 224 individuals/m² was estimated. Therefore, in principle other methods have been proposed for abundance estimates for litter micromolluscs, for example, collecting a significant amount of topsoil, moss, lichen, leaf litter (volume 10–12 liters), followed by hand sorting of these samples in a laboratory. In addition, for complete extraction of the shells (if possible), a set of sieves with decreasing mesh sizes may be used (Horsak, 2003).

For typical soil species it has been proposed to conduct sampling of soil (specific weight), followed by its analysis under a microscope in the laboratory. This method originated in paleozoology (Evans, 1972). Using this approach, Davies et al. (1996) found the abundance of *V. pulchella* to be 0.4–40.4 individuals per 100 g sample of soil on chalky soils in UK, and for the Jurassic limestone in the center of Krakow (Poland), its abundance was 1–22 individuals per 100 g of soil sample (Gózka-Szarszewska, 2013). The results thus obtained are entirely consistent with the data obtained in this study – 1–13 individuals per 100 g of soil sample, which may indicate the higher accuracy of this method.

References


