Lubos Matejicek

Development of Software Tools for Ecological Field Studies Using ArcPad

Integration of data collection, statistical analysis and dynamic modeling in ecology requires new hardware and software tools that offer mobile data management together with implementation of advanced calculation methods. Recent advances in mobile computing representing by mobile devices and ESRI's ArcPad make possible to automate a number of procedures. In addition to mapping of spatial and temporal attributes of ecological systems, the statistical methods and dynamic models can be used to estimate new features and impacts of man-made activities. The software tools include estimates of population, community and ecosystem parameters, which are extended by numerical models based on population dynamics and energy transport.

Introduction

Progress in applied ecology depends on a number of components. In addition to ecological theory and experimental design, new technological advances in collection and management of spatial and temporal data arise in the form of the GIS supported by remote sensing and GPS. GISs represented traditionally by computer hardware and software on WindowsNT or UNIX platforms are also transformed to the mobile computing equipment that includes notebook or pocket PCs. Especially, mobile integrated GISs enforce revisions of traditional ecological field estimates. Moreover entering, storing, retrieving, transforming and displaying spatial data, the aerial and satellite images together with determining of the location by the GPS can be directly used for the estimates of ecological parameters. The measurements of the locations on Earth complemented by data from observations are directly used to estimate the areas or compute the density. Simultaneously, GPSs also assist for navigation of observers. Aerial and satellite data support classification of the surface and the estimates of habitats of the plants and species. Also, new advances in radiometric methods, images from digital cameras and network connections through the mobile phone devices represent potential extensions of the mobile GISs. So, the ecological estimates can become more precise and efficient. Applications of the GISs in ecology are in-depth discussed in the book written by Johnston (1998).

Methodology

Ecology by itself is empirical science that requires data from the real world. But data are not all there is to ecology. Ecological hypotheses or ideas are the other part. Ecological field measurements depend on good sampling devices, which vary greatly for different groups of animals and plants specific to a given field of study, and appropriate statistical and modeling design. In spite of that statistical methods and modeling tools are widely used in other scientific disciplines, now they really get going in the ecological practice. It is caused by the great progress in designing of sampling devices in nearly all areas of ecology.

Ecological methodology

In this paper, the methods of parameter estimates are divided into a few levels that are formed by ecology of

individuals, populations, communities and ecosystems. Data about individuals, which are represented by their parameters (weight or size, length or height, location, habitat and other attributes), are collected by observers and consequently recorded directly into the digital media. Population parameters are represented mainly by the abundance or density, growth ratio, capacity and migration. Mostly, the whole population cannot be observed by itself. So, the various sampling techniques and statistical methods are needed (mark-recapture or removal methods, line intercept techniques, distance or transect methods). In addition to data analysis, dynamic modeling represents more general and complex methodology for population estimates (the models of exponential or logistic growth extended by other factors). In ecology, the next hierarchical level is represented by the community, which is described with another attributes (similarity, diversity, interactions, niche). Again, probability and statistical techniques are used to estimates parameters of communities. In this case, dynamic models can simulate spatial and temporal interaction among population or communities. More complex ecological units are represented by ecosystems, which are mostly studied from a specific point of view (matter or energy flows, food chains, interactions between biotic and abiotic parts). Much of the discussion about ecological methodology is described in the book written by Krebs (1999). Statistical theory and applications implemented in the software tools come out from the books written by Legendre (1998), Zar (1996), Sokal (1995), Heath (1995) and Kanji (1993). Examples of ecological dynamic models are derived from Pielou (1977) and Krebs (1994). Some dynamic models are extended by spatial aspects, which are discussed by Okubo (1980) and Shigesada (1997).

Computing environment

The software tools for ecological analysis are developed with the ArcPad Studio, which assists in the design of user data entry forms and scripts. All the applications run in the ArcPad that can be also installed in the Windows CE environment. The ArcPad is a GIS's product for mobile computing that supports a multi-layer environment with industry-standard vector map and raster image themes. Its ability to work directly in shape formats with symbology exported from ArcView or ArcGIS, and the interface with the GPS makes it useful in most any field ecological data collections. The software tools are developed with VBScript language that also supports Windows CE environment. The ArcPad user interface is customized by files based on the ArcPad's XML format. All these tools are sufficient for calculations of estimated ecological parameters and simulation of smaller spatial dynamic models. The spatial data exchange in the frame of the GIS’s data formats enables to move some larger time consuming calculations to the ArcGIS environment. Consequently, the computations can be carried out by extensions developed with the use of the ArcObjects technology. The numeric algorithms included in the multidimensional data analysis and dynamic modeling are mostly discussed in the book written by Flowers (1995).

Results

Software Tools for Ecological Field Studies-EcoPad

The software tools include a number of methods focused on spatial data management, multidimensional data analysis and dynamic modeling. The basic schema of the system (EcoPad) is in figure 1, which illustrates its basic interconnection and operability.

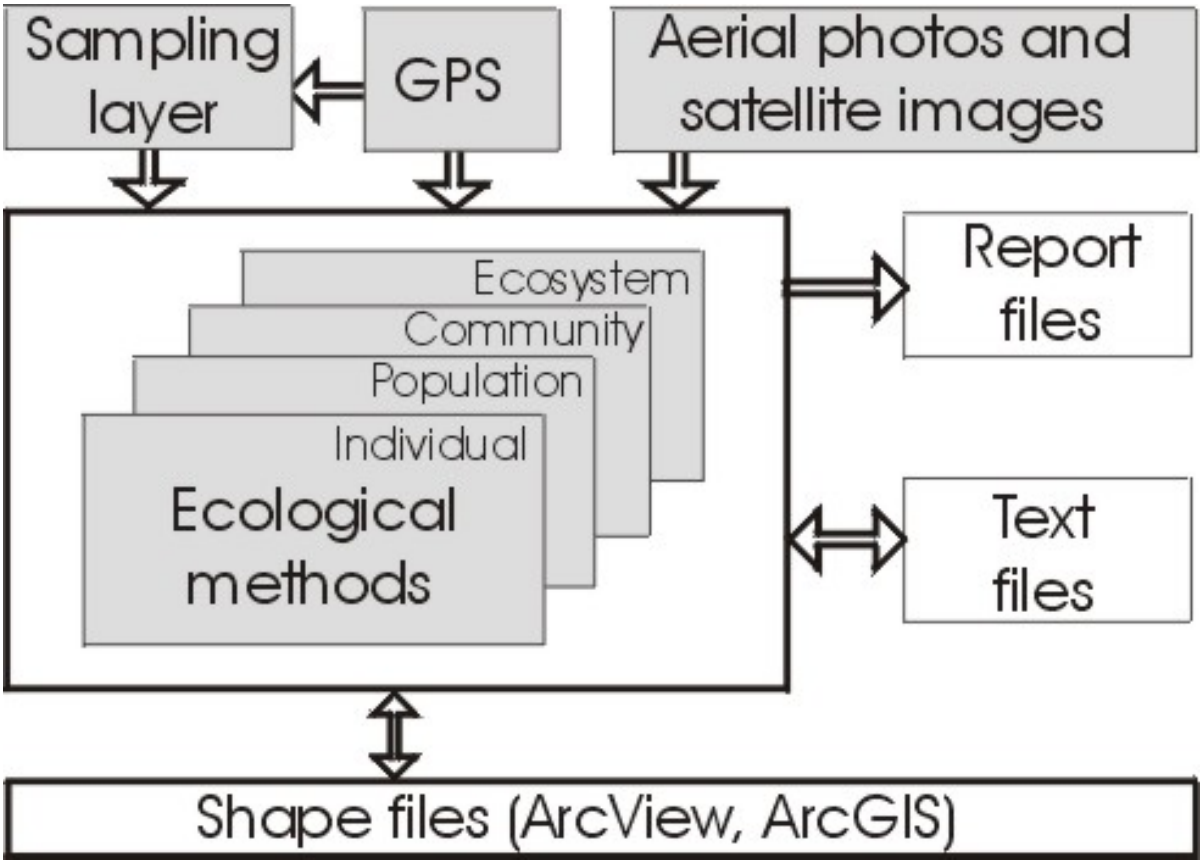

Figure 1. EcoPad: The general schema.

The sampling layer is used for collection of individual samples. Data are collected into the three forms, figure 2. The first form contains identification codes, date and time of an observation or capture, a name of the specie and its habitat, a name of the observer and the location (X ,Y, Z). The second form is used for the input of the values, which can represent the variables in nominal, ordinal, interval or ratio scales. Their meaning depends on a practical case. The third form is completed by other information (comments and images). The location (X, Y, Z) can be captured via the GPS connected to the Pocket PC, by placing the pen on the map layer or by typing appropriate coordinates into the form. All the data are stored as attributes of the sampling theme.

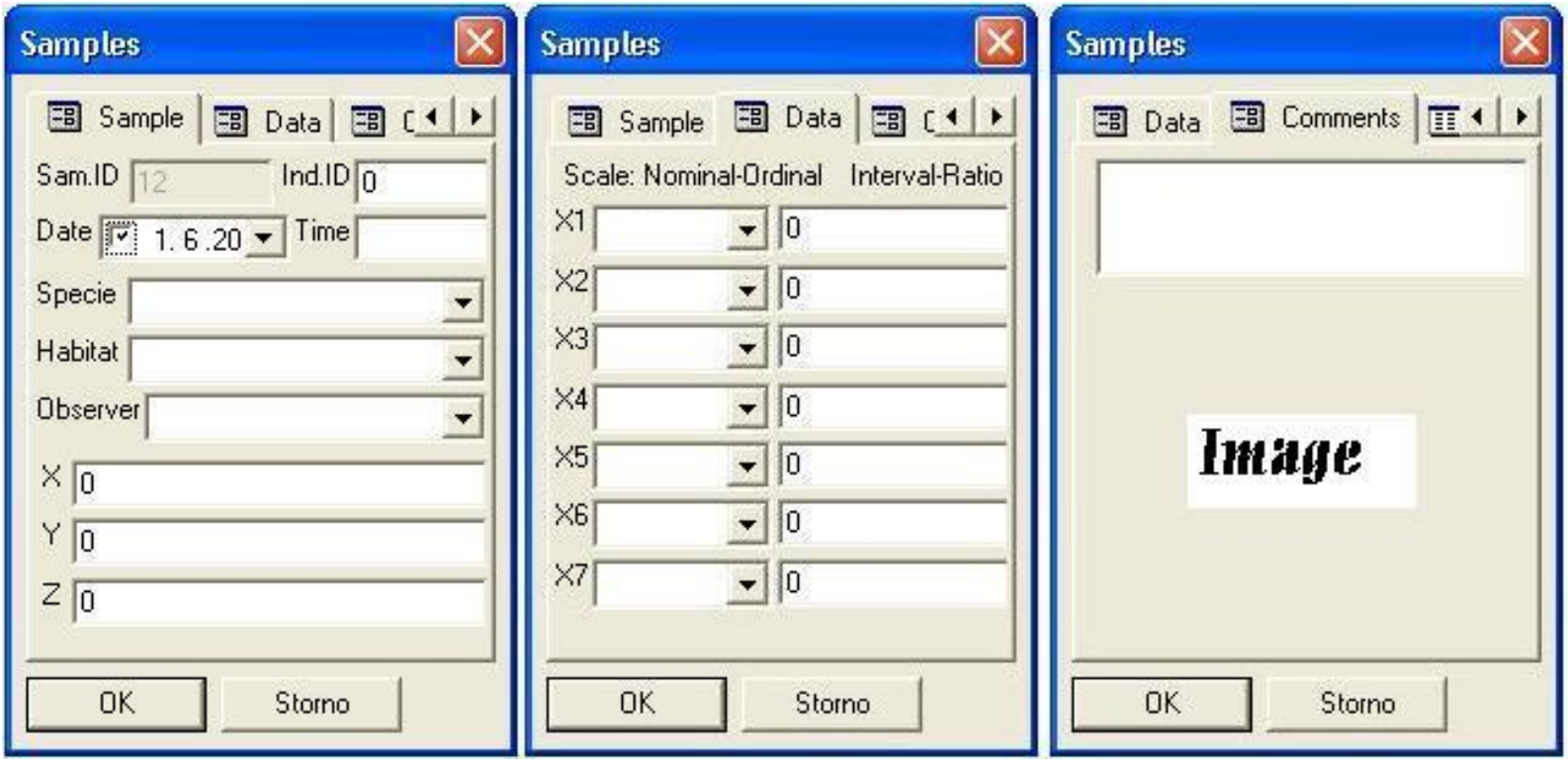

Figure 2. User forms of the sampling layer.

In addition to the sampling layer, other vector and raster layers (thematic maps, aerial and satellite images) are usually used to complete the project and help the observer to identify the position and the right place for observations. Consequently, data from the sampling layer are transferred into the modules, which carry out calculations and estimates based on ecological methods for individuals, populations, communities and ecosystems. Each method has its own form and data format for saving of input and output data. The input data can be also placed directly into the form and stored in the file without the transfer from the sampling layer. Even though the data are saved in the text format, the report files can be built to show the input data and the results in a more user friendly way. The data management of each implemented ecological method is illustrated in figure 3.

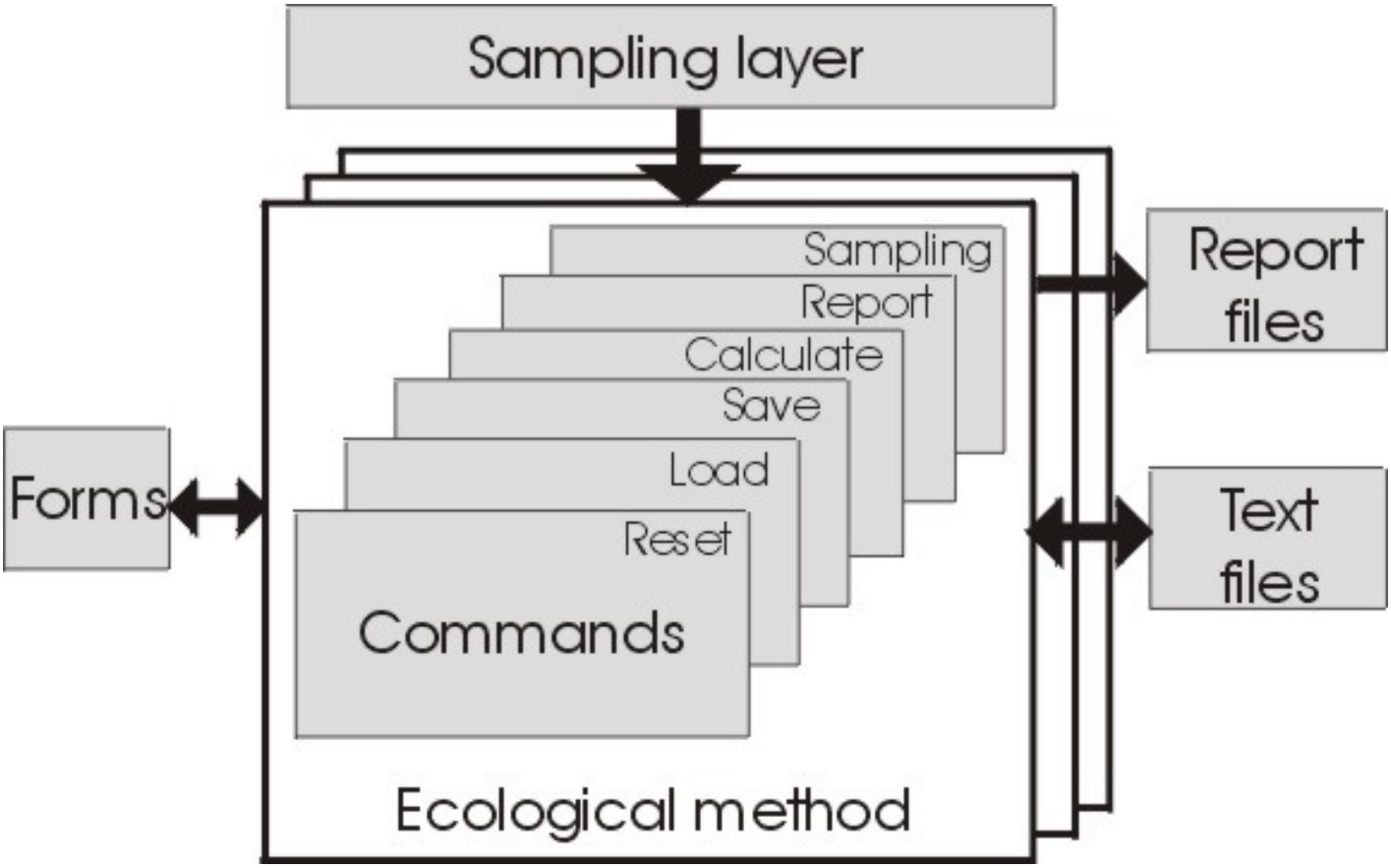

Figure 3. Data management in the frame of each ecological method.

The Reset command sets the initial values of the inputs. The Load and Save commands are used for data store in the frame of each ecological method. The Calculation command starts the computing. Input and output data can be stored by the Report command into the text file. The Sampling command transfers data from the sampling layer into the form of a selected ecological method. As an example, the forms of three ecological methods (estimates of the population abundance with the mark-recapture techniques) are presented in figure 4. The button SAMPLE is used to transfer data from the selected theme (Sampling or another theme). The next part of the form contains the input menu. The bottom part of the form is reserved for the data outputs, which are set after the calculation. The buttons of individual commands are placed in the middle part of the forms.

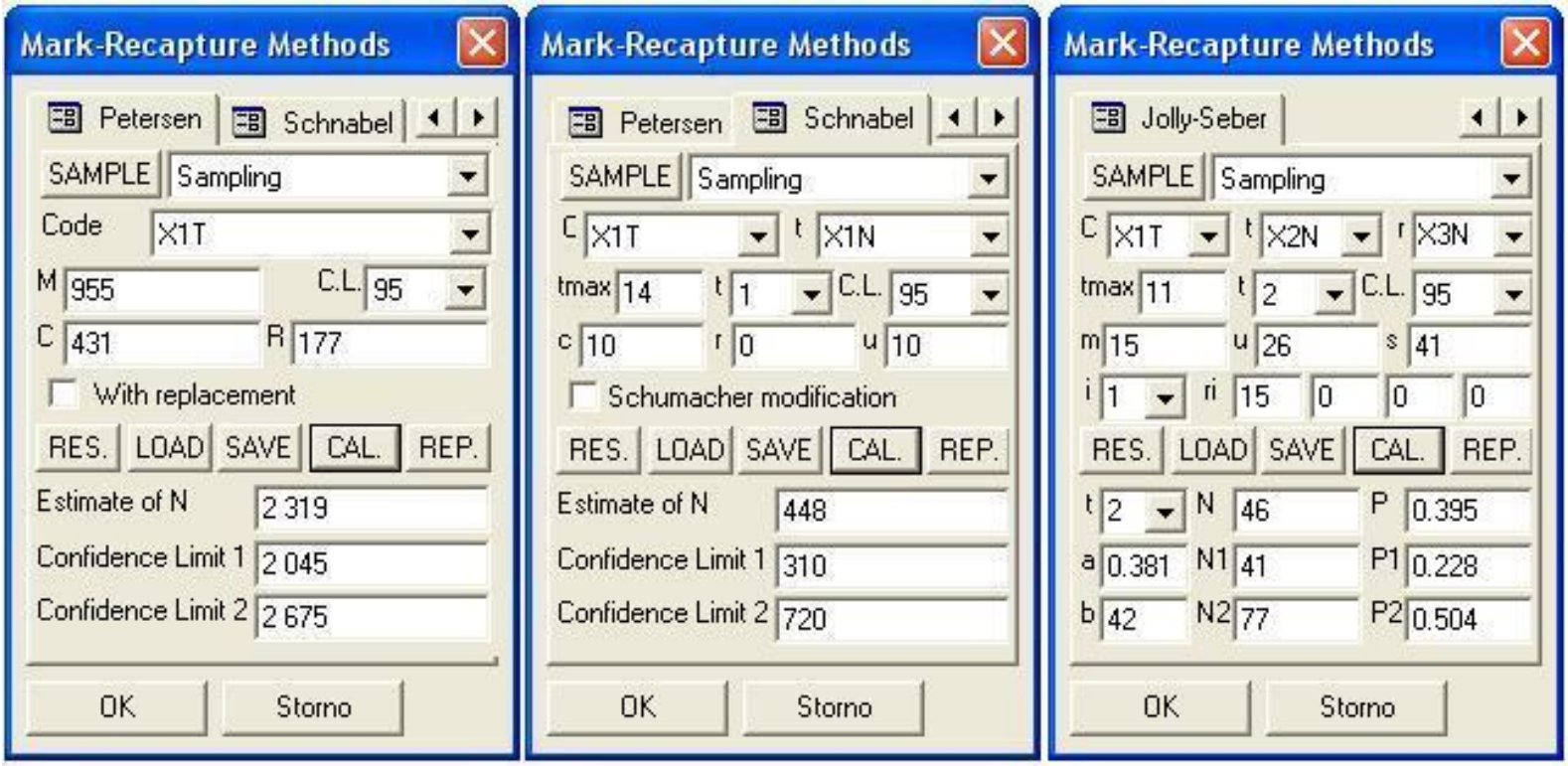

Figure 4. Examples of the forms of the mark-recapture techniques: estimates described by Petersen, Schnabel-Schumacher and Jolly-Seber (tested and processed by data from Krebs, 1999).

Each ecological method is grouped into the appropriate toolbar. In addition to the standard toolbars, the toolbars for capturing of data by the sampling layer, from the GPS and via the network from the GIS servers are added. Their individual buttons start the functions built in the ArcPad. All the ecological methods are placed into two toolbars, which are mostly concerned on methods focused on ecological populations and communities, figure 5. In reality, each button by itself represents a class of ecological methods.

Figure 5. User toolbars in the ArcPad environment.

Ecological Analysis and Modeling

There is a wide range of methods in ecology. They are used for estimates of ecological parameters, calculation of empirical rules and simulation of ecological models. The set of ecological methods implemented in the ArcPad environment together with the above described data management is called EcoPad. Considering to that other

methods focused on individuals are directly included in the scripts of the sampling layer, the main attention is dedicated to the description of the methods in ecology of populations and communities. The studies focused on ecosystem modeling are partly included in the methods dedicated to the communities. Due to the speed and the capacity of the present Pocket PCs, the lager dynamic models tend to be simulated in the standard PC's environment.

Population: data analysis and modeling

The population as one of the basic ecological units is described by a number of ecological parameters. Mostly, the abundance or density represents the main focus of ecologists. The methods in the frame of the EcoPad include mark-recapture techniques, removal methods, quadrat counts, line transects and distance methods. Production parameters together with the mortality, migration and growth are calculated in the next modules. Other part is represented by parameter estimates of growth models (exponential and logistic growth). More complex solutions are included in the simulation of dynamic models extended by spatial interactions.

The mark-recapture techniques include the methods described by Petersen, Schnabel and Schumacher for closed populations and the method made out by Jolly and Seber for opened populations. In addition to the estimates of the abundance and related parameters, the confidence limits enable to assess the variability of the results. The examples of the calculation forms are illustrated in figure 4.

Other alternatives to the mark-recapture techniques represent removal and resight methods, quadrat counts, line transect and distance methods. The methods implemented in the EcoPad include the catch-effort method (Leslie-Ricker), the line intercept technique and the Hayne estimator. The distance methods are represented by the Byth and Ripley procedure, the T-square sampling procedure and the ordered distance method. Mostly, the variability of the estimates is carried out by confidence limits on a few levels of critical values (Krebs, 1999).

Other class of methods is concerned on modeling and optimization. Among a wide range of various methods and techniques, the compensation of the biomass created by Pivnicka (1988) and the growth models (the exponential and logistic growth) together with estimates of their parameters are implemented in the EcoPad, figure 6. The dynamic models formed by the logistic growth are extended by the spatial interactions (diffusion, convection). The datasets are in the form of regular grids of cells. Each cell contains as properties a number or biomass together with other attributes (initial value, growth ratio, capacity, diffusion constant, convection, constant). The calculation parameters represent the spatial and time step and the time period of simulation. A part of the values in the grid can be displayed in the bottom part of the form. All the data can be shared and visualized in the theme by the ArcPad or transferred into to the ArcGIS as the shapefiles.

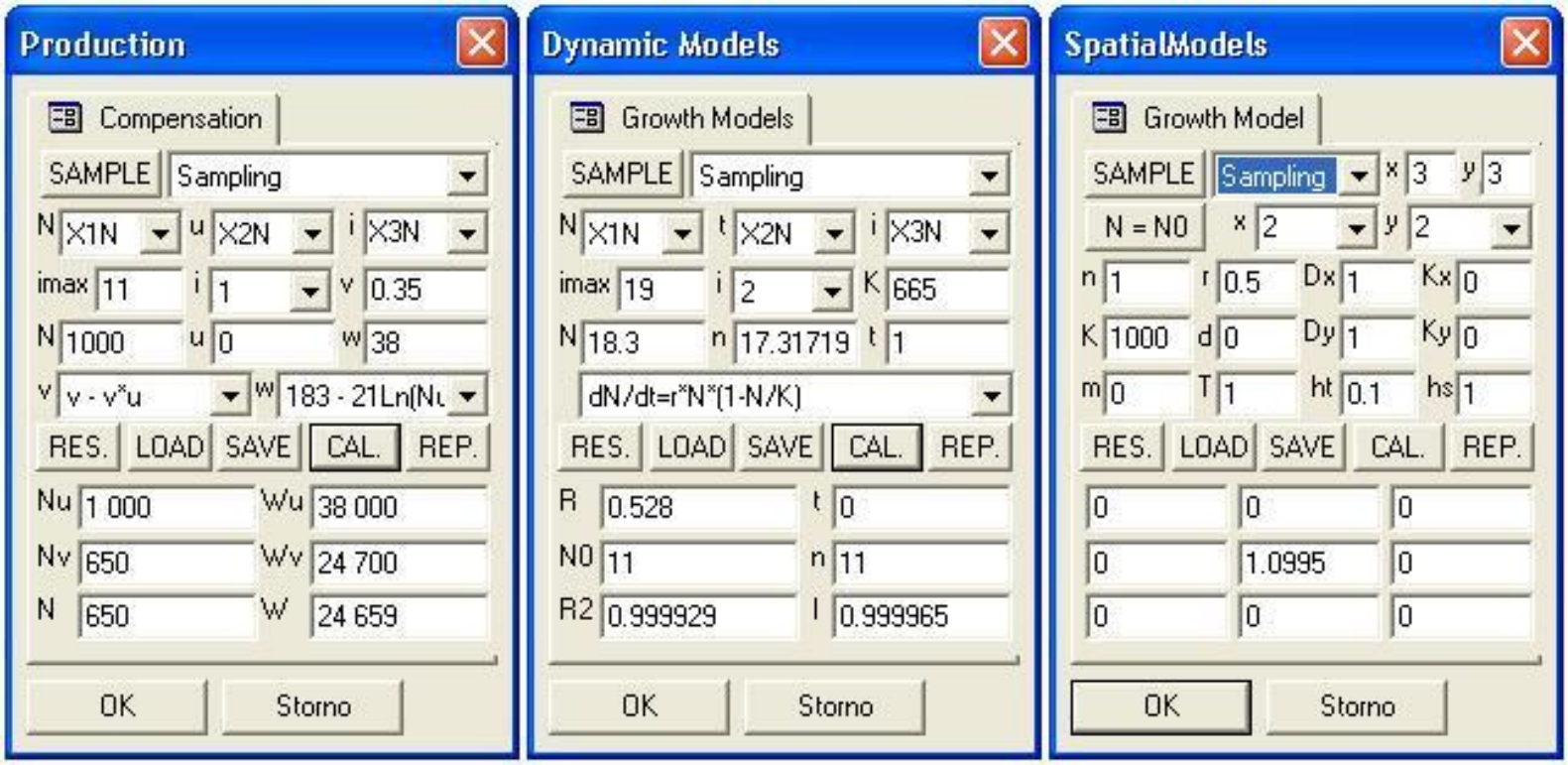

Figure 6. Compensation of the biomass, growth models and spatial growth models (tested and processed by data from Pivnicka and data reported in the frame of the research projects).

Communities: data analysis and modeling

Attempting to characterize and measure the properties of ecological communities faces a special set of statistical and modeling problems. Some of these methods and techniques are also implemented in the EcoPad. They are mostly focused on the evaluation of similarity parameters, cluster analyses and multivariate statistical techniques, diversity measures, niche characteristics and dynamic models extended by spatial interactions.

The classification of communities or their samples needs measures of similarity. The various evaluation methods of similarity have been carried out to compare communities in the best way. Also, much confusion exists about which measure to use. There are two classes of similarity measures in the EcoPad. The binary measures contain the Jaccard coefficient, the simple matching coefficient, the Sorensen coefficient and the Baroni-Urbani and Buser coefficient. The distance measures represent the euclidean distance, the Bray-Curtis measure, the Canberra metric, the percentage similarity, the Morisita's index and the Horn's index of similarity. In cases of many samples of communities to analyze, the various techniques for grouping of similar samples are used. The EcoPad contains a few clustering methods for the classification of the samples. The single, complete and average linkage clustering is implemented to classify ecological communities. The user forms of the mentioned methods are illustrated in figure 7.

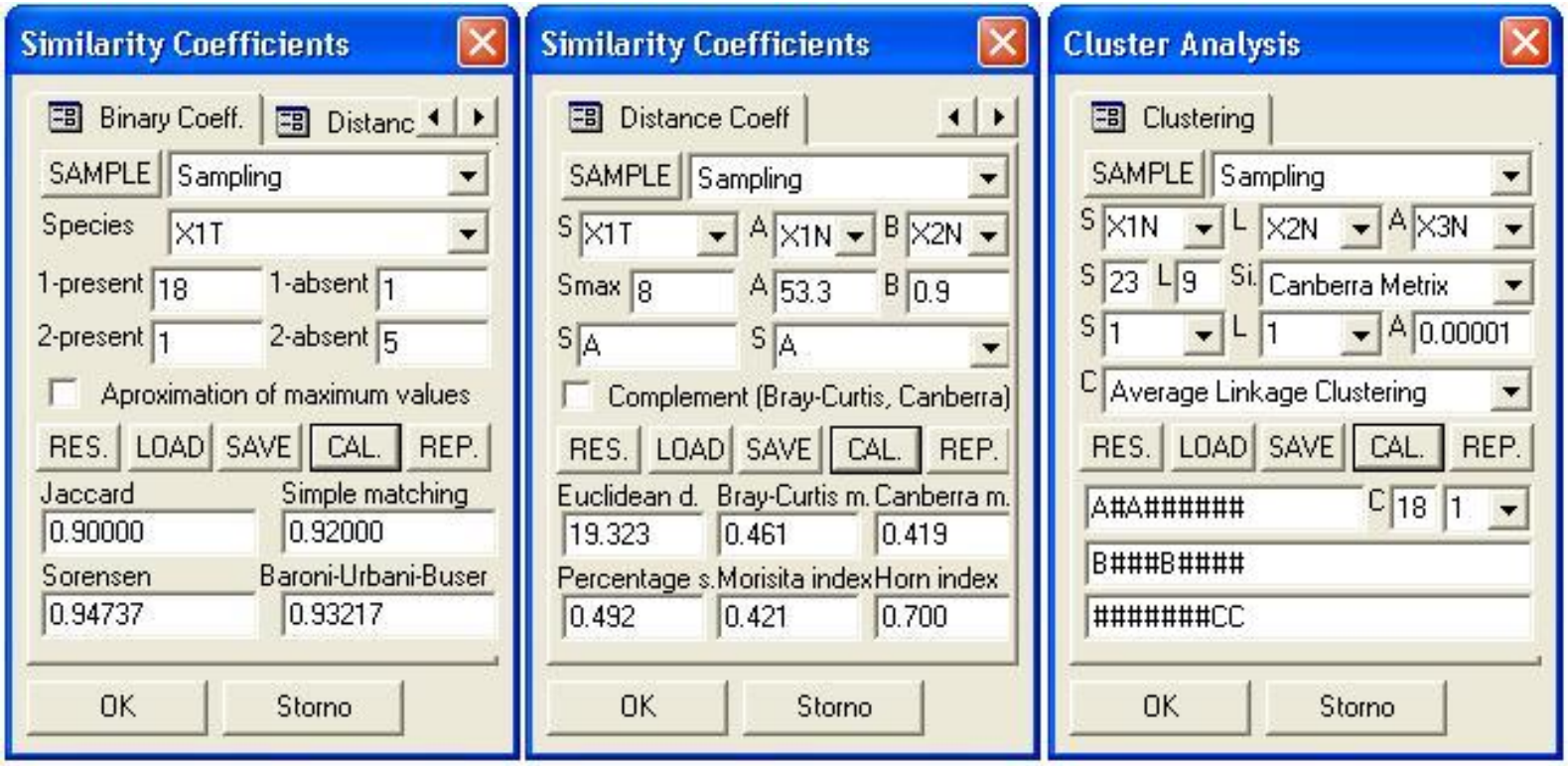

Figure 7. Similarity coefficients and clustering: the binary and distance coefficients, clustering (tested and processed by data from Krebs, 1999).

Species diversity is evaluated by many different ways. Among the various concepts of diversity, the species richness and heterogeneity is widely used. The EcoPad contains the calculation of the species richness, which is based on the rarefaction method. Evaluation of the species heterogeneity is carried out with the Simpson's index, the Shannon-Wiener function and the Brillouin's index. Measurements of niche parameters of a population to analyze community dynamics can be provided by a number of techniques. The Ecopad's methods are divided into the evaluation of the niche breadth and the niche overlap. The niche breadth, which is also called the niche width or the niche size, is calculated by the Levins's measure, the Shannon-Wiener measure and the Smith's measure. The calculation of the niche overlap is represented by the MacArthur and Levins's measure, the percentage overlap, the Morisita's measure, the simplified Morisita index, the Horn's index and the Hurlbert's index.

The group of multidimensional data analyses contains a number of more general methods, which are represented by the principal component analysis (PCA), the multiple correlation and the multiple regression, figure 8. In spite of that the methods are mostly implemented in statistical software, the EcoPad enables their integration in the frame of to the basic field calculations, ecological estimates and dynamic modeling. Another class of statistical methods forms hypothesis testing. In the EcoPad, a number of extensions can be estabilished to extend standard ecological methods.

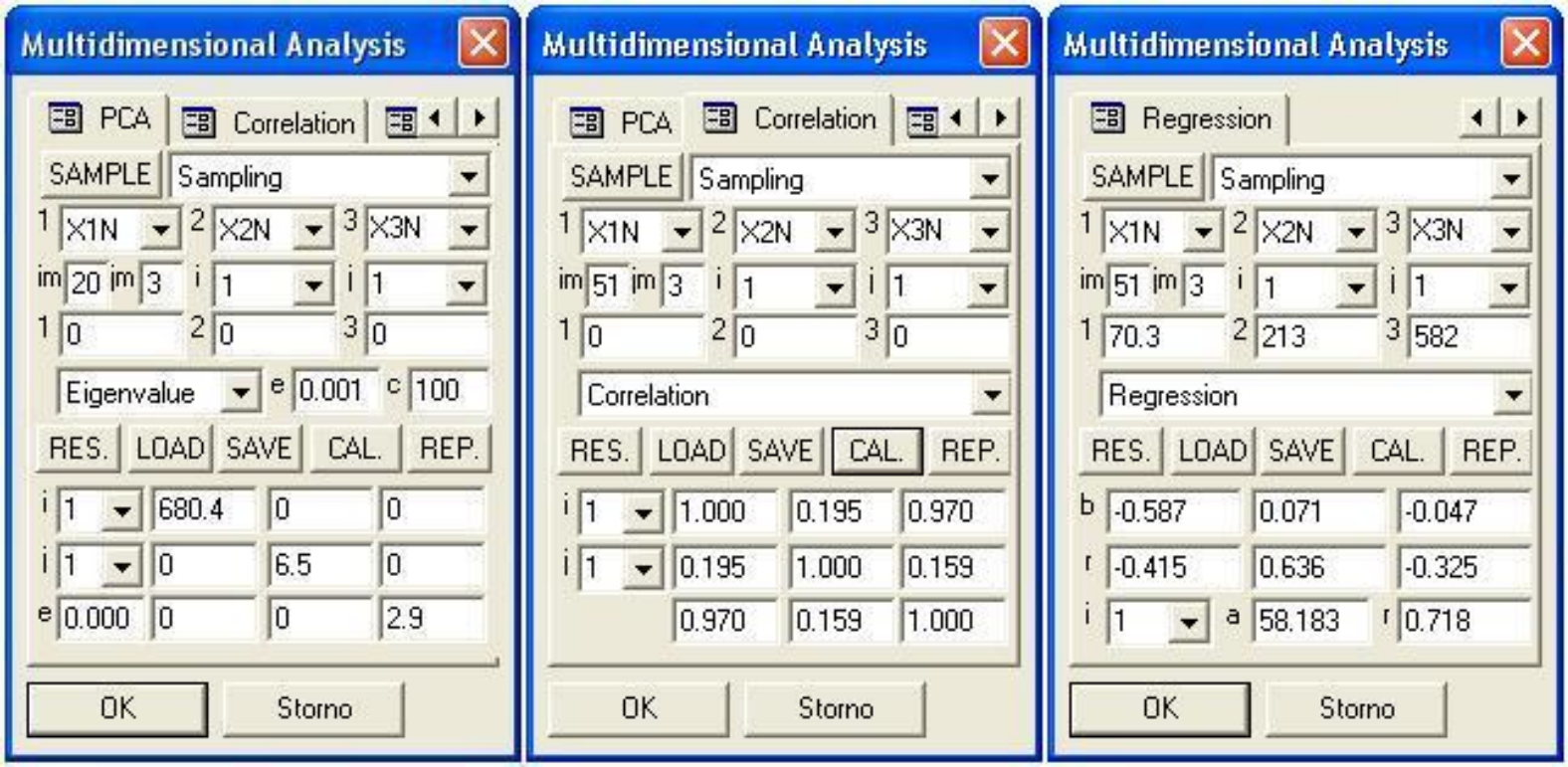

Figure 8. Multidimensional data analyses: the regression and correlation, the ordination techniques (tested and processed by data reported in the frame of the research projects).

Dynamic modeling formed by standard models, which describe community dynamics and basic types of interactions (competition, predation), is extended by spatial interactions. In addition to dynamics and spatial interactions of each population, the interactions on the level of communities or ecosystems can be subjoined. The implemented dynamic models are mostly based on competition models. In case of spatial modeling, the diffusion and convection phenomena are implemented to simulate the migration of individuals, biomass or energy. Due to the amount of data calculated in each time step, only the last data grid is stored into the output. Likewise in the spatial modeling of populations, the larger models are planned to be computed and visualized with the more powerful computer equipment (the ArcGIS extended by modules developed with ArcObjects).

A case study

As an example, mapping of the trees attacked by the insects in a basin is carried out. Considering to the climate conditions and the monocultural type of forest, the spreading of insects causes large damage. If the attacked trees are regularly observed, the migrations of the insects can be predicted by spatial dynamic models and consequently compared together with the observations. The inputs into the dynamic model is represented by aerial images and terrain sampling carried out with the EcoPad. In case of larger data grids, finally, the models are used to be solved in a more powerful computer environment. The EcoPad contains basic tools in the field of spatial dynamic modeling, which enable efficient calculations of smaller data grids. The outputs of the models are formed by individual simulation studies, which contain time series of insect densities in the data grids. The results can be displayed with symbology imported from the ArcView or the ArcGIS. The example shows integration of the mobile GIS (the EcoPad’s scripts in the ArcPad environment) together with the data from remote sensing and the GPS, figure 9. The interconnection of all these systems increases accuracy and efficiency of the estimates and predictions.

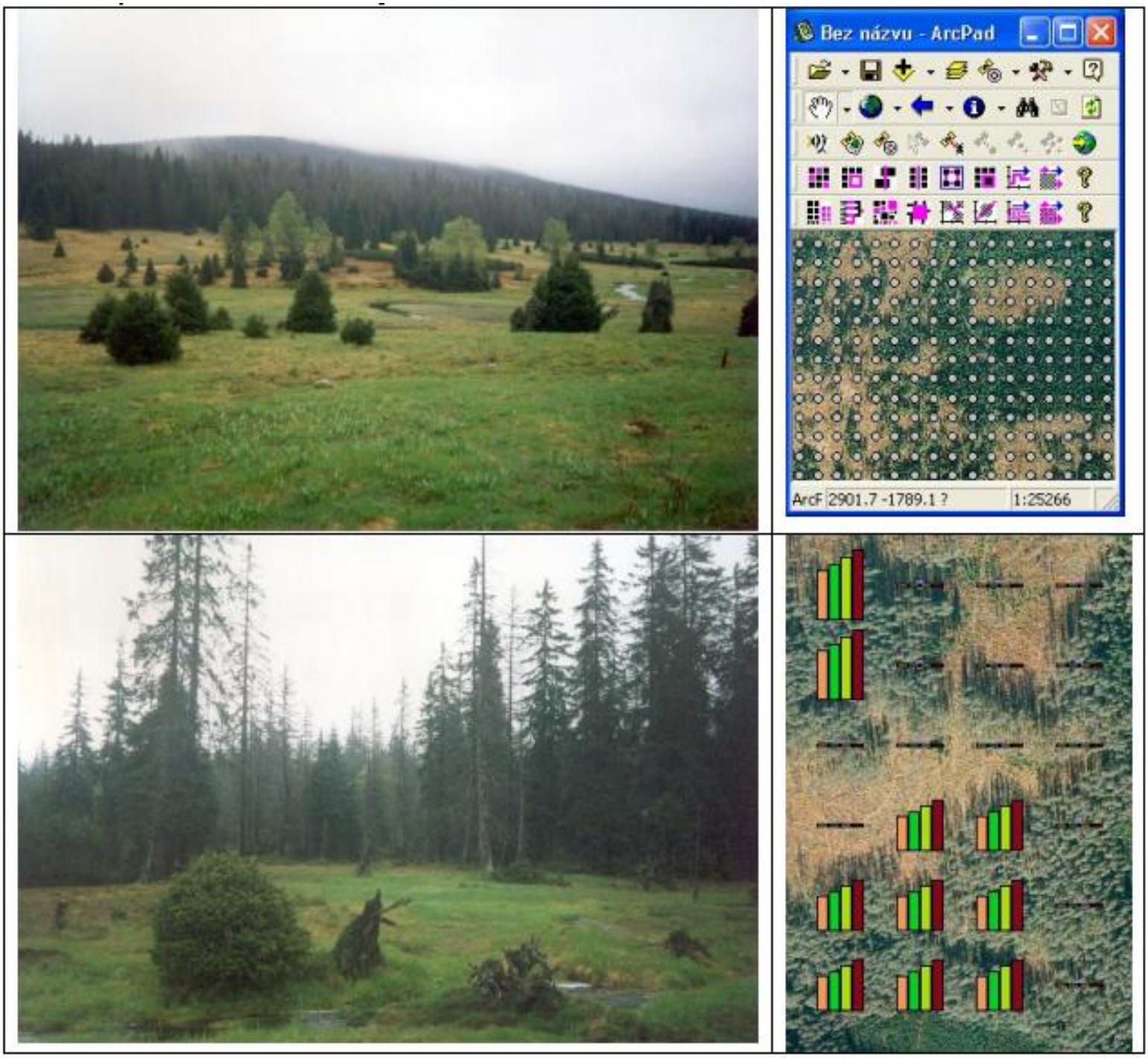

Figure 14. Spreading of the insects in forest: the simulation study in a basin carried out by the EcoPad.

Conclusion

The progress in the GISs, the mobile computing systems, the satellite and aerial images, the network interconnection represents just a part of a wide range of new technology. The tools ArcPad, which are developed in the ArcPad environment, try to implement all these new phenomena in the frame of standard ecological methodology. In addition to the basic estimates, multidimensional data analyses and spatial dynamic models are built in the EcoPad. All these tools provide a better understanding of how ecological systems are managed on a various levels of the ecological research.

References

Flowers, B.H., 1995. An introduction to numerical methods in C++. Clarendon Press, Oxford, pp. 200-272.

Heath, D., 1995. An introduction to experimental design and statistics for biology. UCL Press, London, pp. 63-303.

Johnston, C.A., 1998. Geographic information systems. Blackwell Science, Oxford, 185-201.

Kanji, G.K., 1993. 100 Statistical tests. SAGE, London, pp. 21-157.

Krebs, C.,J., 1999. Ecological methodology. Addison Wesley Longman, New York, pp. 1-496.

Krebs, C.,J., 1994. Ecology: The experimental analysis of distribution and abundance. Harper Collins, New York, pp. 198-261.

Legendre, P., Legendre, L., 1998. Numerical ecology. Elsevier, Amsterdam, pp. 387-424.

Okubo, A., 1980. Diffusion and ecological problems: Mathematical models. Springer-Verlag, Berlin, pp.1-50.

Pielou, E.C., 1977. Mathematical ecology. Wiley-Interscience, New York, pp. 7-110.

Pivnièka, K., 1988. Changes in ecological parameters and dynamics of ichtyocenoses in man-made lakes. DSc. Thesis, Charles University in Prague.

Shigesada, N., Kawasaki, K., 1997. Biological invasions: Theory and practice. Oxford University Press, Oxford, pp. 1-132.

Sokal, R.R., Rohlf, F.J., 1995. Biometry: The principles and practice of statistics in biological research. W.H. Freeman, New York, pp. 451-608.

Acknowledgements

The EcoPad (version 1.0) has been developed within the scope of the research project MSMT VS 97100 in the frame of the Campus Pack license from ESRI, Inc. The presentation of the EcoPad software at the 23rd Annual ESRI International User Conference in San Diego is partly supported by the research projects VZ MSM 113100007 and GACR 42-201827.

Lubos Matejicek
GIS Laboratory, Institute for Environmental Studies
Faculty of Natural Science, Charles University in Prague
Benatska 2,
Prague 2
128 01
Czech Republic

lmatejic@mbox.cesnet.cz