Biodiversity analysis supporting species-subspecies uncertainty in findings data

Boris Milašinović, Toni Nikolić, and Krešimir Fertalj

Abstract—The number of geocoded localities rises significantly and considerable research has been devoted to addressing spatial distribution of various types of species (endemic, threatened, invasive, etc.) and determining ecological niche of the species. An object model presented in the paper enables pairing data from GIS layers with floristic findings bound to any taxonomic rank and thus easing complex analysis regarding ecological niche and biodiversity. Biodiversity as the variety within the living species is commonly defined in relation to specific spatial unit. Due to uncertainty or impossibility to determine finding at the exact subspecies level, special care has been devoted to expressing biodiversity at the species level with data uncertainty regarding species-subspecies relation. The model structure, input and output interface and types of results and analysis have been described. An algorithm for calculation of minimal and maximum possible species count in an area has been described.

Keywords— Biodiversity, Spatial analysis, Alpha diversity, Ecological niche, Species-subspecies relation.

I. INTRODUCTION

Biological diversity or biodiversity is variety within the living world, genes, individuals, species, ecosystems [1]. This variety is commonly defined in relation to specific spatial unit, from very small area up to the whole World. One of the basic information on biodiversity and probably the most commonly used refers to the number of species that occur in the defined area [2].

The biodiversity is essential for ecological stability on planet Earth, maintenance of the biosphere in a state supportive of human life, and between others, as necessary source of material basis for humans (e.g. food, shelter, medicine, chemicals, pharmaceuticals, etc.). Information on species richness is often used for defining the so-called hot-spot areas [3][4][5], the areas particularly important as a background towards the implementation of conservation programs [6][7] and other biogeographic analysis (e.g. [8]). The same information, however, has a great significance in the scientific interpretation and understanding of natural laws that affect the distribution of biodiversity (e.g. [9][10][11]).

At the same time it is believed that each year 40,000 species disappear, among which there are mainly those that have not even been described yet. Together with them the possibilities of their application in medicine, agriculture or forestry, so far unexplored, disappear too. The historical United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in 1992 was undoubtedly the turning-point in the approach to the protection of nature and environment. On that occasion 157 countries signed the Convention on Biological Diversity - the most recent global step towards a comprehensive protection and sustainable use of natural resources. Its basic objectives are preservation and improvement of the existing biological diversity, as well as the economical use of natural resources on the principles of sustainability [12]. Owing to its unprecedented response in the world the Conference gave a powerful encouragement to nature protection and contributed to a proper appreciation of this problem area. A vast number of legislative acts around the world are adopted at different levels, and an incalculable number of practical activities around the world have been launched (see World Conservation Monitoring Centre or IUCN web sites, www.iucn.org) with the aim of preserving biodiversity, sustainable use and equitable and uniform distribution of benefits arising from the use of biodiversity as sources.

Use of GIS layers to address spatial distribution can help in assessing changes in biodiversity [13] and the object model described in this paper addresses spatial distribution of biodiversity at any taxonomic rank (e.g. species, genus, family, etc.). The model uses geocoded data on species findings, prepares input data, analyzes species distribution, calculates diversity, helps to determine species ecological profile and ecological niche and enables spatial analysis using GIS tools. As an extension to initial model presented in [14], the model is extended to take in account data uncertainty of impossibility to determine findings at the lowest taxonomic rank and to enable addition of custom functions used in producing results.

As a study area the Croatian territory has been chosen which is further elaborated in the second section. The model has defined input and output types and interfaces and can be part of a web service or a layer of an application. Model structure is described in the third section followed by description of types of analysis and results in the fourth and the fifth section. Due to uncertainty or impossibility to determine finding at the

Boris Milašinović is with the University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia (+385 1 6129864; fax: +385 1 6129915; e-mail: boris.milasinovic@fer.hr).

Toni Nikolić is with the University of Zagreb, Faculty of Science, Department of Botany, Marulićev trg 9a, 10000 Zagreb, Croatia (e-mail: toni.nikolic@botanic.hr).

Krešimir Fertalj is with the University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia (e-mail: kresimir.fertalj@fer.hr).

exact subspecies level there exists an uncertainty when expressing exact numbers of species in an area. The sixth section introduce an approach to handle this problem by defining an algorithm to find lower and upper bound of the number of species. The paper concludes summarizing benefits and drawbacks of the model with the guidelines for future development and research in this area.

II. STUDY AREA

As a study area the Croatian territory was selected. Located within 4 of the 11 recognized European bio-geographical regions: Alpine, Mediterranean, Continental and Pannonian [15] it is characterized by outstanding biodiversity [12][16][17][18]. As a test data, data on distribution of vascular plants is used from Flora Croatica Database (FCD) [19][20]. Currently more than a half million of localities data is geocoded although geocode precision varies within eleven categories ranging from low accuracy (e.g. county level) to GPS precision. It is expected that this number will rise significantly as only minor parts of herbarium collections in Croatian's museums have been digitalized due to lack of appropriate hardware and lack of funding. Also field experts would be equipped with newly developed Android application (in stage of testing) for support of field observations that will allow them to enter data from field observations during observations and directly copy data to FCD. Gradually increasing the amount of accumulated geocoded findings has provided the basis for the spatial evaluation of biodiversity [10][16][18] but until the model had been developed the largest part of analysis had to been done manually.

Species ecological indices (anatomy, life form, pollination type, humidity dynamics, etc.) have been used for analysis. Several climate layers were used providing information about temperature, precipitation, etc. also as several layers with regional data like county borders, geographical regions, etc. producing new knowledge about species leading to a few research papers in review process or already published (e.g. [21]).

III. MODEL TYPES AND INTERFACES

A. Main model classes

Fig. 1 shows the main class diagram of the model. It contains both classes used for data manipulation and input and output. An analysis can be performed on one or more thematic GIS layers either already stored on a server, or those uploaded by users, where currently only ESRI shape format [22] with polygons is supported. Each record from an ESRI shape file is read and stored in a *Polygon* class that has rectangular bounds that enables faster detection that a point does not belong to that polygon. Each shape file record is paired with corresponding record in a database file having several columns (in further text layer attributes) that provide additional information about the polygon (e.g. name of the polygon, average temperature in the polygon, etc.) These data are stored in class Polygon inside

array of values named *Attributes*.

As a prerequisite for an analysis, localities of taxa findings have to be joined with polygons from the layer(s) thus forming *JoinedShapeRecords*. Each *JoinedShapeRecord* contains a reference to a shape record and extends it with set of taxa that are contained inside the polygon and collection of points contained in the polygon. This extension is implemented using composition instead of inheritance to avoid data copying as object of type *ShapeRecord* already exists when object of type *JoinedShapeRecord* has to be constructed. This task of spatial join is equivalent to determining which point belongs to which polygon. As summarized in [23][24] two basic concepts for solving this problem are known in literature: the even–odd rule (ray-crossing method) and the winding number (angle summation algorithms). For its simplicity an algorithm using angle summation algorithm from [24] have been chosen during model implementation. As significant number of findings could have same coordinates (especially those points of lower precision) it would be inefficient to determine multiple times if the same point belong to a polygon. Due to this all finding points are merged in a list of *Localities*. A *Locality* inherits *Point* and contains a reference to all finding points at the same spot. A finding point contains coordinates (inherited from *Point* class), taxon identifier, geocode precision, finding source, finding identifier and a year of the finding. Term taxon is used instead of species allowing model to work not only with species but with any taxonomy type.

A. Types of analysis

After spatial join had been done analysis is performed on a chosen attributes from the layer(s). Only one attribute simultaneously can be chosen for an analysis and if two or more attributes are chosen from the same GIS layer, service is optimized not to do spatial join multiple times. Analysis is performed using several functions divided in two main categories based whether they calculate values for one polygon (single record functions) or set of polygons (grouped record functions) that have same attribute value. E.g., a habitat type can be formed of multiple polygons in the shape file and multiple records in the corresponding database file with the same value in column that defines habitat type. If a user is interested to find how many findings were noted in each habitat type he/she wants summarized data from all polygons belonging to the habitat.

Typical examples of functions are those that return number of findings, number of different species, number of terrain expeditions in a polygon or in a set of polygons belonging to the same attribute (e.g. to the same county) etc. Another example are characteristic/categories data functions whose number varies depending on all possible values like ecological indices functions which values is equal to the number of possible indices. Additional categories functions and other functions can be defined using interfaces described in the next section.

Results are returned as one or more data matrices where rows usually contain all possible attribute values and columns

are species and/or function names. Both single record functions and grouped record functions are used to populate matrices. The structure of the matrices varies on chosen options, number of possible values and types of analysis. A user can make two different types of analyses: ecological niche analysis and biodiversity analysis explained in detail in sections IV and V.

In addition to matrices results can be further enhanced by extending input database file from the input parameters. The database file is extended with new columns containing results of single record functions thus (with unmodified shape file and spatial index file) making a new layer that can be used in GIS tools. E.g. new columns in database file can contain information for each polygon how many field observations have been performed in the polygon, how many different species exists in the polygon, etc.
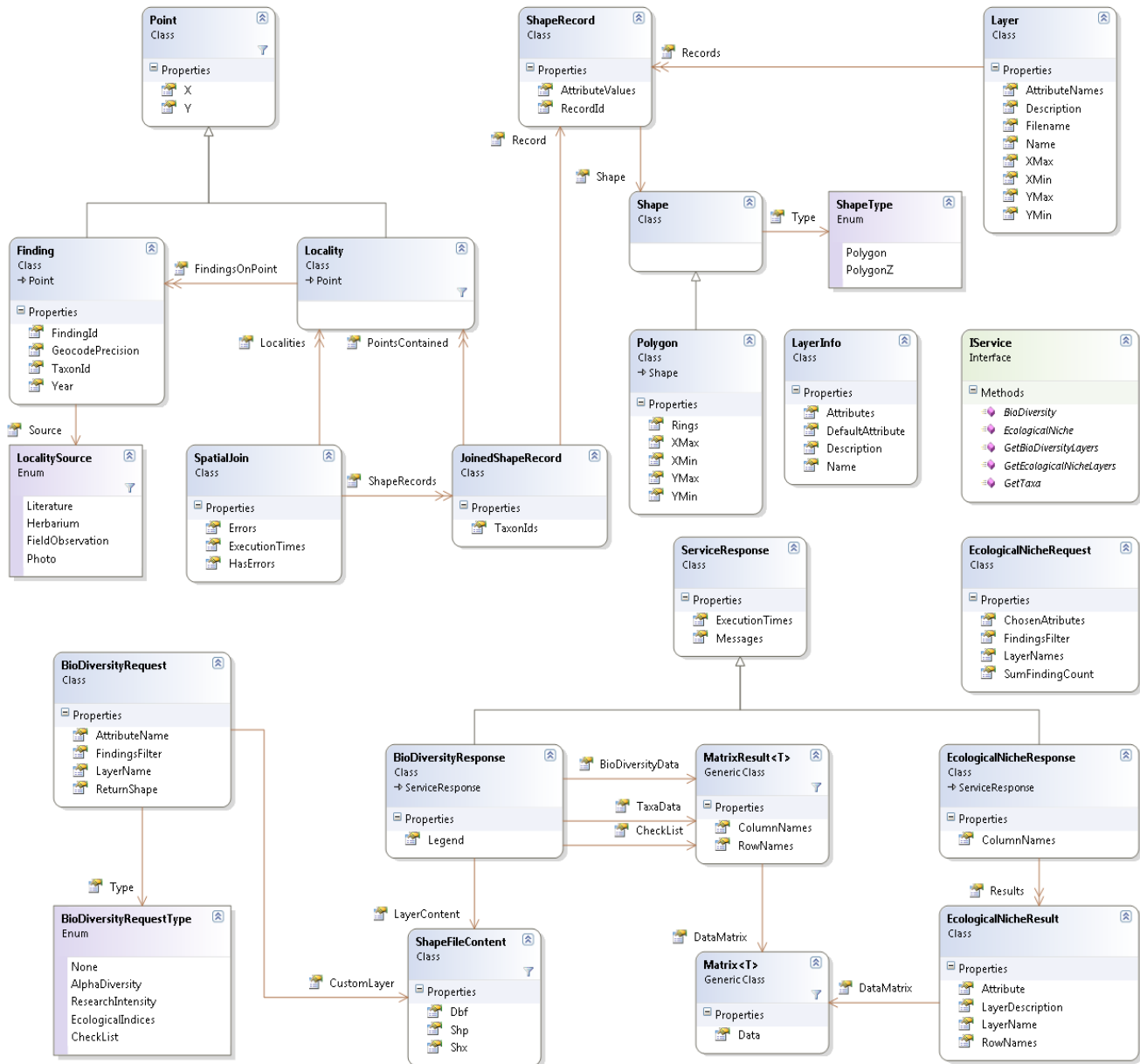

Fig. 1 Main classes of the object model

B. Data retrieval and manipulation

As a processor intensive task and prerequisite for further analysis, spatial join is an ideal candidate for being parallelized in order to obtain the benefits of the parallel computing [25]. During implementation parallelization has been done using .NET Task Parallel Library (TPL) letting TPL do divide data in appropriate thread numbers according to the number of processors and cores. As database file of a GIS layer can be extended with one or more attributes, implementation method for database file extension receives list of functions that are run in parallel. After all function values have been calculated, extension of the database file is performed. Additional performance gains are obtained by simultaneously loading layer data and findings data and by

running in parallel other minor mutually independent tasks.

Input and output are modeled using interfaces thus making model available in various usage scenarios as a web service or a layer in an application. As it only defines structure of the input data, the model is independent of concrete data loaders and service is implemented in such way that it should be independent from data retrieval as long as the data follows some biological patterns. Data retrieval is done by implementing proposed interfaces and merging them with core service implementation service using one of dependency injection techniques (e.g. [26]) when service is exposed as a web service.
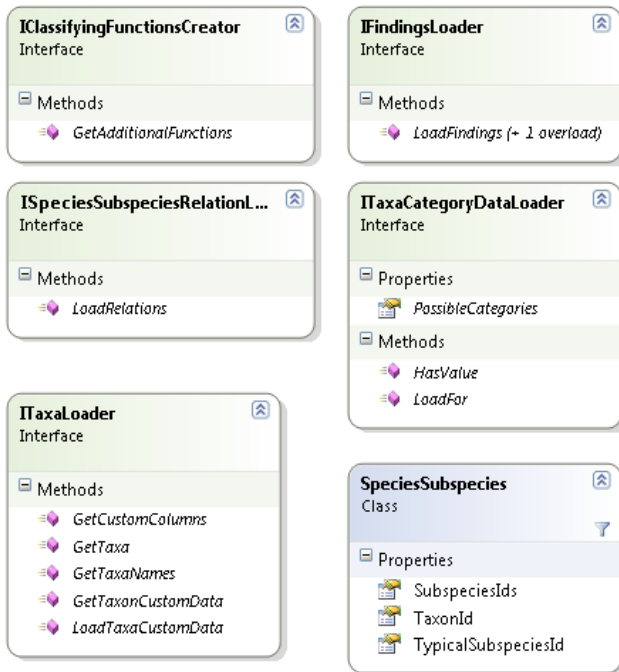

Fig. 2 Data loading interfaces

Fig. 2 shows data loading interfaces that concrete data loading implementation have to implement. Analysis is performed on findings returned from the implementation of *IFindingsLoader* interface. These can be species findings or findings of any other taxonomy type in taxonomy hierarchy. In order to calculate number of different species and to handle situations that some findings could not be determined at the lowest level in taxonomy hierarchy relation between species and subspecies have to be established and it is done using *ISpeciesSubspeciesRelationDataLoader* interface.

An implementation of the interface *ITaxaCategoryDataLoader* loads list of possible categories and values for a collection of taxa and implements characteristic function determining whether a taxon has a particular value in a category. When creating final output, not only taxa identifiers are returned, but also taxa names with additional custom data that is done by implementing interface *ITaxaLoader*. Additional classifying functions can be added by implementing *IClassyfingFunctionCreator* interface.

IV. ECOLOGICAL NICHE ANALYSIS

Purpose of ecological niche analysis is to distribute chosen species per attributes from one or more GIS layers and to find in how many polygons belonging to the same attribute a species is present without regarding how many times it has been noted.

For the simplicity of the first version of model only layers stored on server can be used. For each layer the chosen attribute would be used for grouping data and producing summary values. Besides names of chosen GIS layers and attribute names, input parameters for the calculation are findings filters consisting of species names or part of the names, endemic and invasive status, year of finding, locality source and geocode precision. For each input layer localities for chosen species are joined with polygons from the layer. Afterwards model implementation produces summary data per attribute value, e.g. for set of polygons having same attribute value. Result of this type of analysis is collection of matrices where all matrices have same number of columns and column names represents species names. Number of matrices in collection is equal to the number of input layers. Row names in a matrix represent possible attribute values of a chosen attribute in layer. A value in the matrix produced for layer l in row a and column s counts how may polygons from layer l having attribute value a contain species s.

As values for chosen attribute can be continuous this type of analysis is usually paired with further data processing where attribute values are grouped in user defined ranges and then shown on a graph like in Fig. 3 where average temperature is chosen as an attribute for analysis of three species. The graph from the figure shows in how many polygons (y-axis) having average temperature in some range (x-axis) a species has been noted. Rather than observing absolute numbers, distribution pattern have to be examined, because not all species have same outspread or not all of them are noted so frequently.
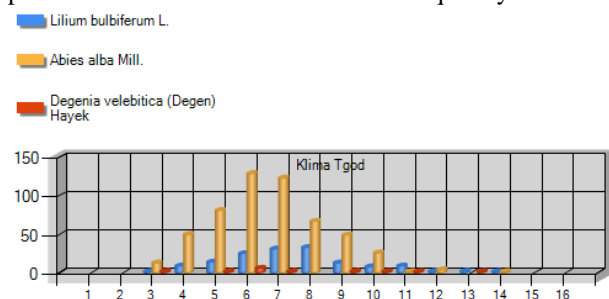

Fig. 3 Example of data from ecological niche analysis grouped in user-defined range

In addition to that, it has to be noted that interpretation of these results depends on type of layer used for analysis. E.g. if each attribute is consisted of only one polygon then described function is characteristic function and thus can have only values 0 or 1 on y axes showing whether a species is present in a polygon with a particular attribute value. Another useful result can be produced if a layer is consisted of environmental data divided in grid cells of equal size. In this case matrix

value represents how many cells with environmental value a contain species s. If there are several weather and land type layers this provides foundation for forming ecological niche of a species.

Misinterpretation occurs if areas with different attributes are formed of polygons of different size. E.g., an attribute X can be consisted of one large polygon and an attribute Y can be consisted of many small polygons with a total area even smaller than those covered by X. The maximum result for species s and the attribute X can only be 1 and for species s and the attribute Y can be equal to number of polygons of Y and such values cannot be compared and used for any analysis.

V. BIODIVERSITY ANALYSIS

Biodiversity analysis works with one GIS layer. Input parameters for biodiversity analysis are similar to those in ecological niche analysis with remark that, beside publicly available layers, a user can upload own layer. Data returned to a user contains extended ESRI database file (and original shape file and spatial index file) and three matrices: *BioDiversityData*, *CheckList* and *TaxaData* matrix. Values for the extension of database file and for *BioDiversityData* matrix are produced by functions belonging to the following types: alpha diversity, research intensity and categories data. All three types of functions are both single record functions and grouped record functions. Single record versions are used to extend database file and grouped record versions are used to produce values for *BioDiversityData* matrix.

Row names of a *BioDiversityData* matrix are formed of all possible values of a chosen attribute in the chosen layer. Column names are formed of function names and, in case of categories data, all possible values in a category. Matrix values are values of a particular function for a particular attribute. Meaning of the functions and methods of calculation are described in the following sections.

CheckList matrix represents a species checklist where row names are formed of possible attribute values and column names are formed of species names. Each value in the matrix represents number of findings of a species in an area belonging to an attribute thus giving information in what extent a species is present in an attribute.

The *TaxaData* matrix gives additional data about species and it is not related to a particular polygon but rather to a whole layer and species that are noted in the layer. It is returned as a result in order to enable further reporting on the client side.

A. α-diversity

α-diversity measures the number of different species in an area. Biologists have used several slightly different definitions for α-diversity [27]. In this paper, the term α-diversity means species diversity in a single spatial unit [28], i.e., number of different species occurring within an area of a given size [29] or the species richness of a single sampling unit [30]. This means that α-diversity has to be calculated both for each polygon from a shape file (and it would be calculated as a

number of different species in that polygon) and for each attribute (number of different species noted in all polygons of same attribute). For the latter case, calculation of a distinct union of all findings has to be done as a species can be present in more than one polygon of the same attribute but have to be counted only once per attribute value. Due to data uncertainty described in section VI two values have been calculated: lower and upper bound.

B. Research Intensity

Research intensity represents number of findings per spatial unit and is calculated separately for each locality source: field observations, literature references, herbarium collections and users' photos. It is calculated both for a single polygon and as a summary for all polygons having same attribute value. Research intensity for a source in a particular area is formed of number of findings and count of unique finding identifiers from the source in that area. E.g. in an area it could be m findings from n field observations and research intensity for observations in that area is pair (m, n). Thus research intensity for a locality source ls and a polygon p is pair (x, y) where x is number of findings from source ls inside polygon p and y is number of unique finding identifiers among those findings.

Using research intensity relation between number of terrain expeditions and number of findings can be established. Such information could help field experts to determine whether is worth going in some areas where someone has already been and to determine how many time field experts have to go to terrain to be sure (within a statistical error) that already all species in some area have been noted.

To calculate value x for research intensity of a source ls and an attribute a (contained of many polygons) sum of finding counts of all polygons having attribute value a is taken. However, the number y cannot be calculated by summing y values of each polygon belonging to the attribute as one field observation could be done in more than one polygon. Therefore, distinct union of all finding identifiers has to be done, and y is cardinality of the union set.

C. Categories Data

Functions related to categories data provides analysis similar to analyzing ecological niche of a species, but this time the accent is not on species. Purpose of this analysis is to join each polygon with each possible value from a set of categories (e.g. categories can be formed of possible ecological indices values). Assigning a number N for a polygon p and a value V where V is one of the possible distinct values in a category C means that in the polygon p there are N species that have value V for category C.

Similar calculation can be done for a set of polygons having same attribute value with notable difference that there are two valid approaches. The first one is to count different species and the second is to sum values already assigned to a particular polygon from a set. Both approaches have biological explanation. Data about number of different species that are located in some area maybe will not give enough information

as information gained by summing individual values of each polygon from the group. E.g. suppose that a layer with climate data is formed of 1km^2 cells and that area where temperature is 5 is present in m cells and m species is distributed in those m cells in such way that neither two species are in the same cell. Then, the number of species in area with temperature 5 is m and total sum is also m. But, if every species is present in every of m cells then number of species is m but the total sum is m^2 and the second approach brings additional information to interpretation of results, while the first approach cannot distinguish these two situations.

As the different species number has to be found and data uncertainty occurs, same as for α-diversity, possible lower and upper bound have to be calculated. As there are two valid approaches for calculation categories data for an area, thus four values have been produced for each possible category value.

VI. DATA UNCERTAINTY

A. Species and Subspecies relation

Some species may have subspecies and due to nature of biological work there are situations that some findings could not be determined at the lowest level in taxonomy hierarchy. It can be due to inexperience of the field observer, lack of details in the literature or in herbaria, etc. Thus some findings are noted at the species level and some on the subspecies level and taking cardinality of the union of all observed taxa cannot be applied because it will produce larger α-diversity then it really is. E.g. suppose that in a polygon the following species and subspecies are noted: *Thalictrum minus* L., *Thalictrum minus* L. ssp. *minus* (typical) and *Thalictrum minus L.* ssp. *olympicum* (Boiss et Heldr.) Strid. α-diversity of that polygon is not 3, but 2 as *Thalictrum minus L.* in the study area appears either as *Thalictrum minus* L. ssp. *minus* or as *Thalictrum minus* L. ssp. *olympicum* (Boiss et Heldr.) Strid.

Due to this, exact number cannot be always calculated so only lower and upper bound on different species number can be established. Due to this, in any situation that number of different species has to be calculated (e.g. calculation of α-diversity) two values (minimal and maximum possible) are produced. In certain situations these two values could be the same.

B. Decision tree for α-diversity increment

Relation between a species and its subspecies can be such that

- A species does not have subspecies
- A species has typical subspecies and might have other subspecies
- A species has subspecies but does not have a typical one

During calculation process α-diversity is incremented for one for every subspecies in the list of taxa belonging to an observed polygon (or set of polygons) and for every species that does not have subspecies as these two situations surely increase α-diversity. Uncertainty occurs for findings of species that has subspecies.

Decision tree from Fig. 4 decide whether α-diversity will be incremented when a species with subspecies is observed. It can be assumed (although there is a risk that assumption may not be correct) that if species has typical subspecies (left branch of the decision tree) that the finding of the species is probably finding of the typical subspecies. Therefore if a polygon contains the typical subspecies then α-diversity is not increased for the observed species. Contrary, if typical subspecies exists but it is not noted in the polygon α-diversity is increased for one.

E.g. using previous example that would mean that every finding of *Thalictrum minus* L. is a finding of *Thalictrum minus* L. ssp. *minus.* If both *Thalictrum minus* L. ssp. *minus* and *Thalictrum minus* L. are present in the polygon then α-diversity is incremented for *Thalictrum minus* L. ssp. *minus* (as it is not uncertain data) but not for *Thalictrum minus* L. In case that *Thalictrum minus* L. ssp. *minus* is not present in the polygon then presence of *Thalictrum minus* L. increases α-diversity.
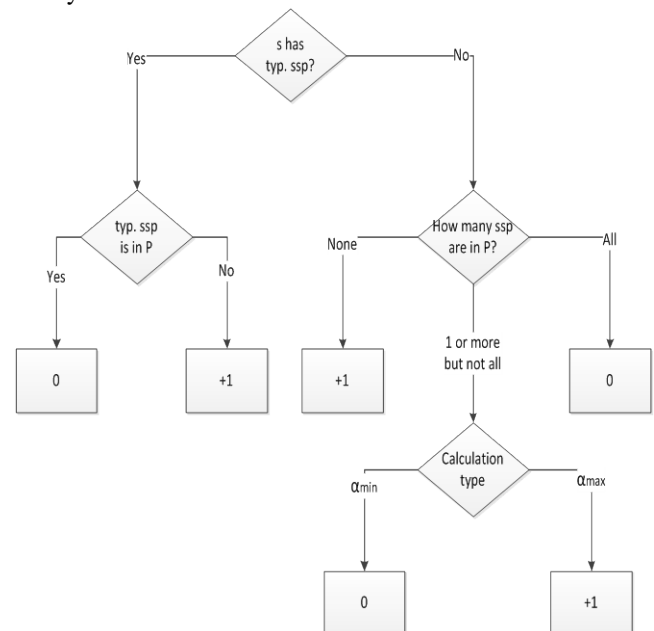

Fig. 4 Decision tree for α-diversity increment for a species with subspecies

If a species does not have the typical subspecies, but have other subspecies then there are two indisputable cases: when the polygon does not contain neither one species' subspecies or if it contains all known subspecies of the species. In the former case α-diversity must be incremented as this species is really unique species in the polygon. In the latter case α-diversity must not be incremented because findings of the species are just finding of one of its subspecies and all subspecies are already noted and α-diversity is incremented for each of them.

Uncertainty occurs if the polygon contains only some of the subspecies and we cannot for sure determine if the species occurrence in that polygon should be occurrence of one of

already present subspecies or it is an another one. As there is no typical subspecies any assumption easily could be incorrect.

Due to aforementioned reasons exact value of α-diversity cannot be calculated for some polygons and attributes. Therefore, for α-diversity and any other functions that rely on different species number two values are calculated, lower and upper possible values. α_{min} is the worst case scenario where it is assumed that record bound to species that has subspecies is one of already recorded subspecies in that polygon and α-diversity is not incremented. The opposite situation is α_{max}, the optimistic, best case scenario where α-diversity is incremented assuming that the finding of the species is finding of one of not already noted subspecies.

VII. CONCLUSION

In recent years, considerable research has been devoted to mapping the flora distribution, spatial analysis and biodiversity calculation. Complex analyses were performed manually and thus were work intensive. Developed model enables pairing data from species localities and data from GIS layers, makes some analysis automated and eases complex analysis by providing many new aggregated data. Thus it enables new knowledge about species in Croatia. There are currently a few new research papers in revision process addressing spatial distribution of endemic, threatened and invasive species in Croatia and its relationship to conservation efforts, determining ecological niche of the species, locating hot spots and under investigated parts of country and finding patterns for species distribution. These parts of botanic research in Croatia have been insufficiently examined and this model enabled further research in those fields.

It is important to note that interpretation of results depends on type of a layer used for the analysis and species outspread. Having disproportionally formed polygons for attributes could lead to incomparable results. Also, if findings are not precisely bound to the lowest taxonomy level (subspecies) then in analysis that rely on the number of different species in an area it is not always possible to produce an exact number. However, by establishing relations between species and their subspecies lower and upper bound can be established that depending of the data sample could be the same in many areas.

Improvements of the proposed model could be achieved by developing various input methods using web interfaces and exposing model as a service to wider specter of users. At the present moment analysis are rather infrequent but the calculation is very processor intensive. As most of the algorithm process is already parallelized that means that simultaneously requested analysis must be done sequential. If the number of users would rise, a cloud solution would be appropriate in order to increase scalability. The proposed model in that case could be split in two worker roles (for data retrieval and calculation) and one web interface role. As the data retrieval and calculation is done without tight coupling with concrete loaders and with as minimum knowledge of semantic meaning of findings and taxa as possible the

proposed model should not undergo major changes.

REFERENCES

[1] B. Groombridge, Ed. *Global Biodiversity. Status of the Earth's Living Resources.* A Report Compiled by the World Conservation Monitoring Centre, London: Chapman et Hall, 1992.

[2] D. L. Hawkshworth, and M. T. Kalin-Arroyo, "Magnitude and distribution of biodiversity," in *Global biodiveristy assessment*, V. H. Heywood, Ed. Cambridge: Cambridge University Press, 1995, pp. 107-191.

[3] F. Médail, and P. Quézel, "Hot-spots analysis for conservation of plant biodiversity in the Mediterranean basin,", *Annals of the Missouri Botanical Garden*, vol. 84, pp. 112–127, 1997.

[4] F. Médail, and P. Quézel, "Biodiversity hotspots in the Mediterranean Basin: Setting global conservation priorities," *Conservation Biology*, vol. 13, pp. 1510-1513, 1999.

[5] F. E. Zachos, and S. J. Habel, ed. *Biodiversity hotspots. Distribution and protection of conservation priority areas*. Berlin: Springer-Verlag, 2011.

[6] D.J. Coates, and A. Kenneth, "Priority setting and the conservation of Western Australia's diverse and highly endemic flora," *Biological Conservation*, vol. 97, pp. 251-263, 2001.

[7] M. E. Soulé, "Conservation: tactics for a constant crisis." *Science*, vol. 253, pp. 744-750, 1991.

[8] M. V. Lomolino, B. R. Riddle, and J. H. Brown, *Biogeography*. 3th ed. Sunderland: Sinauer Ass., Inc, 2006.

[9] L. Bragazza, "Conservation priority of Italian Alpine habitats: a floristic approach based on potential distribution of vascular plant species," *Biodiversity and Conservation*, vol. 18, no. 4, pp 2823-2835, 2009.

[10] I. Dobrović, T. Nikolić, S. D. Jelaska, M. Plazibat, V. Hršak, and R. Šoštarić, "The evaluation of floristic diversity of Medvednica Nature Park (Northwest Croatia)," *Plant Biosystems,* vol. 140, no. 3, pp. 234-244, 2006.

[11] F. Essl, M. Staudinger, O. Stohr, L. Schratt-Ehrendorfer, W. Rabitsch, and H. Niklfeld, "Distribution patterns, range size and niche breadth of Austrian endemic plants," *Biological Conservation,* vol. 142, no. 11, pp. 2547-2558, 2009.

[12] I. Martinić, "An overview of the state of biological and landscape diversity in Croatia with the strategy and action plans," Ministry of Environment Protection and Physical Planning, Zagreb, 2000.

[13] R. Grabaum et al. "Use of GIS and Field Site Network for Assessing Changes in Biodiversity," in *Proc. 5th WSEAS International Conference on Environment, Ecosystems and Development*, Venice, 2006, pp. 89-93.

[14] B. Milašinović, T. Nikolić, and K. Fertalj, "An Object Model for Biodiversity Analysis," in *Latest Trends in Information Technology,* A. Anderson, Y. Hung-Yen, P. Varacha (Ed.), WSEAS Press, 2012, pp. 273-278.

[15] U. Pinborg, Ed. *Europe's biodiversity – biogeographical regions and sea*. European Environment Agency, Sweden, 2002.

[16] S. D. Jelaska, et al., "Terrestrial Biodiversity Analyses in Dalmatia (Croatia): A Complementary Approach Using Diversity and Rarity," *Environmental Management*, vol. 45, no. 3, pp. 616-625, 2010.

[17] T. Nikolić, "The diversity of Croatian vascular flora based on the Checklist and CROFlora database," *Acta Botanica Croatica* vol. 60, no. 1, pp. 49-67, 2001.

[18] T. Nikolić et al., "Plant species diversity of Adriatic islands: An introductory survey," *Plant Biosystems,* vol. 142, no. 3, pp. 435-445, 2008.

[19] T. Nikolić, Ed, Flora Croatica Database, Available: http://hirc.botanic.hr/fcd

[20] K. Fertalj, T. Nikolić, T. Helman, V. Mornar, and D. Kalpić, "Flora Croatica Database Application," in *Mathematics and Computers in Modern Science*, N. Mastorakis, Ed. New York: WSEAS Press, 2000, pp. 175-182.

[21] T. Nikolić, and B. Milašinović, "Distribution pattern, range, ecology, hotspots and conservation of Croatian endemic plants," in *Book of Abstracts Evolution of Balkan Biodiversity*, Zagreb, 2012. pp. 35-36.

[22] (1998) ESRI Shapefile Technical Description, Available: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

[23] E. Haines, "Point in polygon strategies," in *Graphic Gems IV*, P. Heckbert, Ed. Boston: Academic Press, 1994, pp. 24–46.

[24] K. Hormann, and A. Agathos, "The point in polygon problem for arbitrary polygons," *Computational Geometry*, vol. 20, no. 3, pp. 131-144, 2001.

[25] G. Popov, N. Mastorakis, and V. Mladenov, "Calculation of the Acceleration of Parallel Programs as a Function of the Number of Threads," in *Latest Trends in Computer*, vol. 2, N. Mastorakis, V. Mladenov, Z. Bojkovic (Ed.), WSEAS Press, pp. 411-414.

[26] Ninject – Open source dependency injector for .NET, Available: http://www.ninject.org/

[27] J. Navar, "Computer Programs for Estimating the Alpha Diversity: The Basis of Sustainable Management Plans to Conserve Diversity," in *Proc. 5th WSEAS International Conference on Environment, Ecosystems and Development*, Venice, 2006, pp. 301-305.

[28] A. E. Magurran, *Measuring biological diversity*, Oxford: Blackwell Pub., 2004.

[29] R. Hengeveld, P.J. Edwards, and S.J. Duffield, *Characterization of Biodiversity: Biodiversity from an ecological perspective, Global Biodiversity Assessment.* Cambridge: Cambridge University Press, 1997, pp. 88–106.

[30] J. S. Gray, "The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf," *Journal of Experimental Marine Biology and Ecology*, vol. 250 no. 1-2, pp. 23-49, 2000.

Boris Milašinović is a research associate at the Department of Applied Computing at the Faculty of Electrical Engineering and Computing, University of Zagreb.

He graduated in 2001 at the Department of Mathematics at Faculty of Science, University of Zagreb. He received MSc degree in 2006 and Ph. D. degree in 2010 in Computing at the Faculty of Electrical Engineering and Computing. His main research interests include software development and workflow management. Currently he lectures a couple of undergraduate courses in Computing and he is teaching assistant to one graduate course in Computing.

Toni Nikolić is an associated professor at the Department of Botany at the Faculty of Science (Division of Biology), University of Zagreb.

He graduated in 1982, received MSc degree in 1986 and PhD in 1989, all at the Faculty of Science, University of Zagreb. Postdoctoral fellowships he was done in Botanische Institute, Karl-Franzens Universitat, Graz, Austria and Royal Botanical Garden, Kew, Great Britain. Main research interest are biogeography, taxonomy and phylogeny of the vascular flora. Currently he lectures a couple of undergraduate courses, e.g. General Botany, Systematic Botany, Flora Croatica, etc.

Krešimir Fertalj is a full professor at the Department of Applied Computing at the Faculty of Electrical Engineering and Computing, University of Zagreb. Currently he lectures a couple of computing courses on undergraduate, graduate and doctoral studies.

His professional and scientific interest is in computer-aided software engineering, complex information systems and in project management. He participated in a number of information system designs, implementations and evaluations.

Prof. Fertalj is member of ACM, IEEE, PMI, and Croatian Academy of Engineering.