Data collection and advanced statistical analysis in phytotoxic activity of aerial parts exudates of *Salvia* spp.

M. Giacomini,*,1 A. Bisio,2 E. Giacomelli,2 S. Pivetti,1 S. Bertolini,1 D. Fraternale,3 D. Ricci,3 G. Romussi,2 N. De Tommasi4

1Dipartimento di Informatica, Sistemistica e Telematica, Università di Genova, Italia,
2Dipartimento di Chimica e Tecnologie Farmaceutiche e Alimentari, Università di Genova, Italia,
3Dipartimento di Scienze dell'Uomo, dell'Ambiente e della Natura, Università di Urbino, Italia,
4Dipartimento di Scienze Farmaceutiche,Università di Salerno, Italia.

Article

Abstract: In order to define the phytotoxic potential of *Salvia* species a database was developed for fast and efficient data collection in screening studies of the inhibitory activity of Salvia exudates on the germination of *Papaver rhoeas* L. and *Avena sativa* L.. The structure of the database is associated with the use of algorithms for calculating the usual germination indices reported in the literature, plus the newly defined indices (Weighted Average Damage, Differential Weighted Average Damage, Germination Weighted Average Velocity) and other variables usually recorded in experiments of phytotoxicity (LC50, LC90). Furthermore, other algorithms were designed to calculate the one-way ANOVA followed by Duncan's multiple range test to highlight automatically significant differences between the species. The database model was designed in order to be suitable also for the development of further analysis based on the artificial neural network approach, using Self-Organising Maps (SOM).

Introduction

The use of crop residues, cover crops and plant residues with a high content of allelochemicals in the soil, as well as extracts of allelopathic plants, can provide new tools for the selective weed control through the release of allelochemicals (Bhowmik et al., 2003; Dayan et al., 2000; Xuan et al., 2005) thus contributing to the maintenance and the development of sustainable agro-ecosystems. As Salvia exudates have been shown to possess inhibitory activity on the germination of *Papaver rhoeas* L. and *Avena sativa* L. in a screening work (Bisio et al., 2010a), further studies were planned for the isolation and identification of the active metabolites, as well as their mode of action and their interactions with soil components and organisms.

Some tools are now available to support web based collaboration and most of them are in the field of common report production or in general multiple author writing (Bramley et al., 2006; Brazma et al., 2003; Camon et al., 2004; Gogoulou et al., 2005; Halsall, 2005; Thorisson et al., 2005). These web-based models can act as a versatile and low-cost way to permit distributed collaborative work (Giacomini et al., 2009;

Giacomini & Cafferata, 2009). Then for the definition of the phytotoxic potential of *Salvia* species a database for fast and efficient data collection has been developed among the various collaborating groups to be used in the following research steps.

In relational database design, the process of organising data to minimise redundancy is called normalisation. The goal of database normalisation is to decompose relations with anomalies in order to produce smaller, well-structured relations. So structured data can be exchanged in a secure and safe way through web-based tools in order to quicken the process. The normalised structure of the data collected into the database simplifies their use for the automatic calculation of several inferential statistical parameters. It has been shown (Bisio et al., 2010b) that these parameters cannot completely describe the differences in the activity of the several Salvia species involved in this experiment. For this reason, a set of more advanced analytical approaches can be applied. Also in this case, the structured features of the collected data will enable a quick execution of the advanced analysis algorithms. Similar cases have been considered in many fields correlated to life sciences and in these

Data collection and advanced statistical analysis in phytotoxic activity of aerial
parts exudates of *Salvia* spp.
M. Giacomini et al.

cases a combination of connectionist and symbolic artificial intelligent approach have been proven to be successful (Basheer & Hajmeer, 2000; Bertone et al., 1996; Giacomini & Ruggiero, 2001; Giacomini et al., 2000; 1997; 1996). The connectionist methods aim to mimic the example based method that is typical of the human brain. There are two main possible connectionist approaches, one based on the supervised and the other based on the unsupervised artificial neural networks. The supervised nets are very efficient, but a complete and fully described set of examples has to be available at the beginning of the analysis (Rumelhart et al., 1986).When this is not possible and when fully unknown situations are feasible, the unsupervised Self Organising Maps (Kohonen, 2001) can be successfully adopted. These maps place the input patterns on a two dimensional finite plane divided into a finite number of areas (shown as hexagonal spaces, as in Figure 4). Similar input examples are placed in the same output element, or in the very near elements. In this way, after the learning process, areas allocated to specific classes can be singled out. A pattern coming from an unclassified example is classified according to the zone in which this pattern is placed by the network (Giacomini et al., 2000).

To highlight the importance of neural networks for analysing chemical data, it may suffice to mention that each year more than 1000 publications appear on the use of neural networks in chemistry (Fink & Reymond, 2007; Gasteiger et al., 2003; Krasznai et al., 2010; Selzer & Ertl, 2006).

In comparison to other neural network methods the Self-Organising Map (SOM) are transparent and the results are easily interpretable (Vracko, 2005). In fact the SOM can give a transparent design space to engineer and also help designers to analyse and reduce design space. It can save much product design time, and make optimisation easier (Chu et al., 2010). This reduction allows the SOM results to be represented by a two dimension segmented map on which clusters can be easily identified just looking at this map.

Materials and Methods

Plant material

Fresh aerial parts of *Salvia namaensis* Schinz (1), *S. fallax* Fernald (2), *S. disermas* L. (3), *S. chamaedryoides* Cav. (4), *S. confertiflora* Pohl. (5), *S. x jamensis* J.Compton (6), *S. buchananii* Hedge (7), *S. wagneriana* Polak (8), *S. scabra* Linn.fil. (9), *S. miniata* Fernald (10), *S. cacaliaefolia* Benth.(11), *S. adenophora* Fernald (12), *S. rutilans* Carrière (13) were obtained from The Regional Center For Agricultural Experimentation and Assistance (Albenga, Italy) and

Research Unit for Floriculture and Ornamental Species (Sanremo, Italy). The species were identified by Dr. Gemma Bramley and voucher specimens are deposited in Kew Herbarium (K) (Kew, Surrey, UK). Commercial seeds, of *Papaver rhoeas* L.(La Semeria-www. Lasemeria.it-Italy) and *Avena sativa* L.(Il Monastero-Sementi-Italy) were used for both Petri dish and pot assays.

Assays

Extraction of exudates, germination and growth tests

The exudates of the fresh aerial parts of the *Salvia* spp. were extracted and used for the germination and growth tests as previously reported (Bisio et al., 2010b) for both the Petri dish and the pot experiments. Four concentrations of testing solutions of Salvia exudates were used. The assay was performed in triplicate. Emergence of the radicle (≥1 mm) was used as an index of germination and it was observed daily for a period of 10 days in each replicate. Seedling growth, *i.e.* seedling height, cotyledon length and root length, were recorded with callipers. Fresh and dry weights of the seedlings were also measured.

Germination and phytotoxic activity indices

The phytotoxic effects of the tested exudates on germination were described by the use of germination indices (Chiapusio et al., 1997; Ranal & Garcia De Santana, 2006): Total Germination (GT) (Jefferson & Pennacchio, 2003; Ranal & Garcia De Santana, 2006), Speed of Germination (S), Speed of Accumulated Germination (AS) and Coefficient of the Rate of Germination (CRG) (Chiapusio et al., 1997; Ranal & Garcia De Santana, 2006). In addition, new formulas describing the phytotoxic damage were used: Weighted Average Damage (WAD), Differential Weighted Average Damage (DWAD), Germination Weighted Average Velocity (GWAV) (Bisio et al., 2010b). In order to provide a clearer description of these new indices, their mathematical formulations are listed below:

$$WAD_{s\_TARGET} = \frac{\sum_{R=1}^{3}\left(\sum_{t=1}^{N_t}\left.\frac{G_{Tcontrol}-G_{Tt}}{c_t}\right|_{TARGET}\right)}{3}$$

WAD is an index of phytotoxic activity against the target *Papaver, Avena* and *Papaver* and *Avena* together. WADs is the Weighted Average Damage caused by the considered species of *Salvia* against a

model species taken separately or the model species taken all together.

$$DWAD_s = \frac{\sum_{R=1}^{3}\left(\sum_{t=1}^{N_t}\left.\frac{|G_{Tcontrol} - G_{Tt}|}{c_t}\right|_p - \sum_{t=1}^{N_t}\left.\frac{|G_{Tcontrol} - G_{Tt}|}{c_t}\right|_a\right)}{3}$$

Where p = *Papaver* and a = *Avena*. DWAD is an index of phytotoxic activity against *Papaver* vs. *Avena*. This is the Differential Weighted Average Damage (DWAD), which has been used to evaluate which species is most active against *Papaver* and less active against *Avena*.

$$GWAV_{s\_TARGET} = \frac{\sum_{R=1}^{3}\left(\sum_{t=1}^{N_t}\left(\left.\frac{\sum_{d=1}^{N_d}\frac{|G_{Tcontrol_d} - G_{Ttd}|}{d}}{c_t}\right|_{TARGET}\right)\right)}{3}$$

Where GWAV is the Germination Weighted Average Velocity, $d = 1, 2, 3 \ldots Nd$ days of experiment ($N_d = 10$). The position of d in the formula above indicates that the germination time is a linear penalizing factor, like c. GWAV is an index against the target Papaver, Avena and Papaver and Avena together.

The lethal concentrations needed to reduce germination by 50% (LC50) and by 90% (LC90) were also considered (Bisio et al., 2010b).

Database design methodologies

The entity-relation (E-R) approach was adopted (Batini et al., 1992) for the conceptual representation of information, whereas the relational database approach was used to describe the logic scheme of this system. Another possibility to describe the concepts at the basis of these tools is the ontological modelling (Kalinichenko et al., 2003), but in this project the contributions come from a quite broad spectrum of knowledge make the formation of a solid and shared ontology extremely difficult (Webster, 2002). Moreover, the E-R model is closer than other models to an almost automatic translation into logical schema within a real Data Base Management System (DBMS) (Giacomini et al., 2009).

Statistical analysis

Three replicates were maintained for each treatment in a completely randomised manner. Untransformed data was used to calculate the average and the standard deviation. For the analysis of the variance (one-way ANOVA), the percentage germination data was transformed into arcsine-square root to meet the requirements of the test (Laterra & Bazzalo, 1999). The statistical significance was set at $p<0.05$ for all the tests, in order to determine significant differences among the averages.

Information engineering methodologies

Experimental data organisation was designed through the use of the Microsoft® SQL Server 2008, a relational model database server produced by Microsoft®. In addition, the Microsoft® Visual Studio .NET 2010 was used to produce the web application. A program written using MATLAB® environment, in particular the SOM Toolbox, was developed for the artificial neural networks analysis. This function package implements the Kohonen's Self-Organising Map (SOM) algorithm, which was created by the Laboratory of Computer and Information Science (CIS) of the Department of Computer Science and Engineering at the Helsinki University of Technology. The SOM Toolbox functions were used to initialise and train the network and then to display the SOM results. Another essential part of this package is based on the k-means function:

$$IDX = kmeans(X, k)$$

The application of k-means clustering provide a partition of the points in the n-by-p data matrix X into k clusters. The rows of X correspond to the points, while the columns correspond to variables. The kmeans returns a n-by-1 vector IDX containing the cluster indices of each point. By default, the kmeans uses squared Euclidean distances. When X is a vector, the kmeans treats it as an n-by-1 data matrix, regardless of its orientation (The MathWorks, 1994-2011).

Results and Discussion

Figures 1 and 2 show respectively the database Entity-Relation model and the related database diagram. The entity-relation (E-R) model is closer than other models to an almost automatic translation into logical schema within a real Data Base Management System (DBMS). Therefore the E-R approach (Batini et al., 1992) was adopted for the conceptual representation of information, whereas the relational database approach was used to describe its scheme.

The GENUS and SPECIES ranks were considered in relation to the taxonomic data. The main entity is the EXPERIMENT, that collects all the experimental parameters, *e.g.* the pair of species (a *Salvia* and *Papaver rhoeas* L. species or a *Salvia* and

Data collection and advanced statistical analysis in phytotoxic activity of aerial
parts exudates of *Salvia* spp.
M. Giacomini et al.

Avena sativa L. species) involved in the test and the type of culture used (Petri dish or pot). The VALUE entity was created to collect the germination and growth experiment results.

The database diagram (Figure 2) shows the relationships and the transposition of model entities into tables. The redundancy of data was minimised due to this organisation. The results of the experimental germination are automatically placed in the database structure, together with the morphological data collected on the subsequent growth of seedlings (seedling height, cotyledon length, root length, seedling weight). In addition, the database was associated to a specific website through which the user can view the results of the germination indices usually reported in the literature (Total Germination - GT, Speed of Germination - S, Speed of Accumulated Germination - AS, Coefficient of the Rate of Germination - CRG),and the newly defined indices (Weighted Average Damage - WAD, Differential Weighted Average Damage - DWAD, Germination Weighted Average Velocity - GWAV) (Bisio et al., 2010b) and other variables usually recorded in experiments of phytotoxicity (LC50, LC90). This specific web-based tool was designed to highlight automatically significant differences between the species, it implements the algorithms to calculate the one-way ANOVA followed by Duncan's multiple range test (Duncan, 1955).

The analysis of the variance (ANOVA) is used to express the observed variance as a sum of several values, each one referable to different causes. Thus providing point estimates of the unknown population variance. The k populations considered were independent, normal and homoscedastic. The results are useful in order to perform a test of statistical validity for the difference of the k averages. The one-way analysis of variance involves a single independent variable (or treatment) that is presumed to have some effect on the dependent measured variable. In applying the one-way ANOVA k populations (or groups) of equal size n were considered; on each element of a group the same treatment was performed and then the associated dependent variable was measured. A test of statistical validity can be performed on the ANOVA results, in which the null hypothesis is based on the assumption that the means of the k populations are equal.

Duncan's multiple-range test is performed if the ANOVA has produced a rejection of H_0. It has the same assumptions as the ANOVA, and requires equal sample sizes. The test compares the range of any subset p of the sample means with the calculated statistic called the *least significant range*, denoted by R_p:

$$R_p = r_p \sqrt{\frac{MSW}{n}}$$

where r_p is called the *least significant Studentised range*, MSW is the mean square within the ANOVA, and n is the sample size.

If the range of the subset is less than R_p, then all the sample means in the subset are taken to be equal. On the contrary, if the range of the subset is greater than R_p, then the sample means at the extremes of the range are considered to be significantly different, and thus their associated population means are regarded to be different.

The tests should be done in a standard order. First, the range from the largest to smallest mean is compared with the appropriate R_p value. Then the range from the largest to the second smallest; next from largest to the third smallest; and so on, up to the comparison for the range from largest to second largest. The next comparison should be the range from the second largest
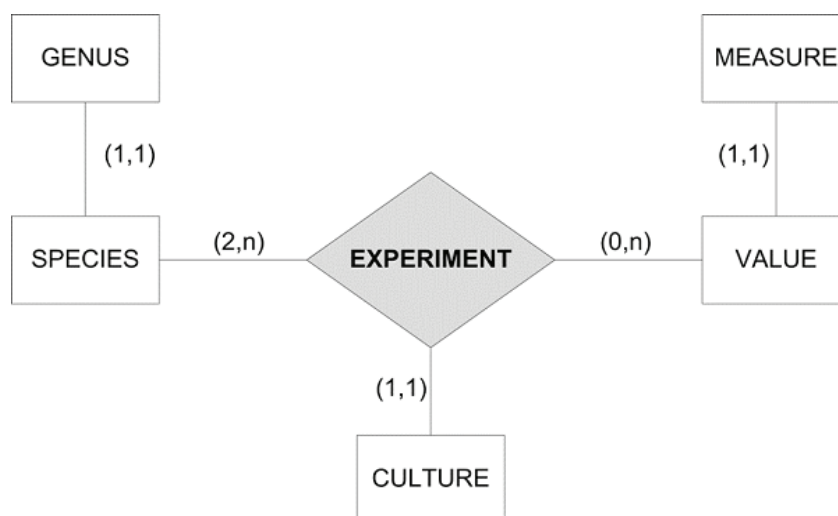

Figure 1. Database Entity-Relation model. It shows the main entities and relations of the designed database.

to the smallest and so on. These range comparisons with R_p are continued until all the relationships between the means are determined (Bernstein & Bernstein, 1999).

In Figure 3 an example of a test result is shown: a Petri dish germination experiment is reported.

As the results of this set of analysis did not show significant differences among the active species, a new approach not based on inferential statistics was taken into consideration. Further analysis was based on the artificial neural network methodology. In addition previous experience of some authors (Bertone et al., 1996; Giacomini et al., 2000; 1997) evidenced that this approach can be very valuable in scenarios where a great amount of quantitative data is used to yield evidence based conclusions.

Due to the nature of the relational database it is possible to apply native methods of data mining to identify the hidden relationships and connections within the full set of data; providing that these algorithms are fed by highly accurate data, from the information content point of view. For this reason an algorithm associated with the database has been created to filter the incoming data according to their significance information, both within and among the attributes. The filtering criteria are based on the awareness that the first days after sowing germinal values are considered null and that the variables fresh weight and dry weight are highly correlated. As a result, some rows and columns were excluded because of a variation coefficient smaller than

10% (within one element), and of a correlation greater than 90% with another element (only one element over two are maintained in the set).

Amongst all possible data mining algorithms, that are present in literature, neural networks have been applied in the form of Self-Organising Maps, also called Kohonen maps. The SOMs are single layer feed-forward neural networks where the output neurons are organised in grids and represent a class of neural-network algorithms in the unsupervised-learning category (Kohonen, 2001). The SOM has spread into numerous fields of science and technology as an analysis method, for example, visualisation of statistical data and document collections, process analysis, diagnostics, monitoring and biomedical applications, including diagnostic methods and data analysis in bioinformatics (Kohonen, 2001).

After the application of SOMs, k-means algorithm was performed to define neuron clusters. It is well established, from literature, that the result of applying *k*-means algorithm depends greatly on the centroid initial choice (Bradley & Fayyad, 1998). For this reason, five neurons were randomly selected as centroids (*i.e.* maximum five clusters), to avoid any bias. This clustering procedure was repeated 100 times using the skimmed files, setting the epochs number to 200 for the first phase (rough training phase) and to 1000 for the second phase (fine tuning phase). The number of neurons was set at 90 and the maximum
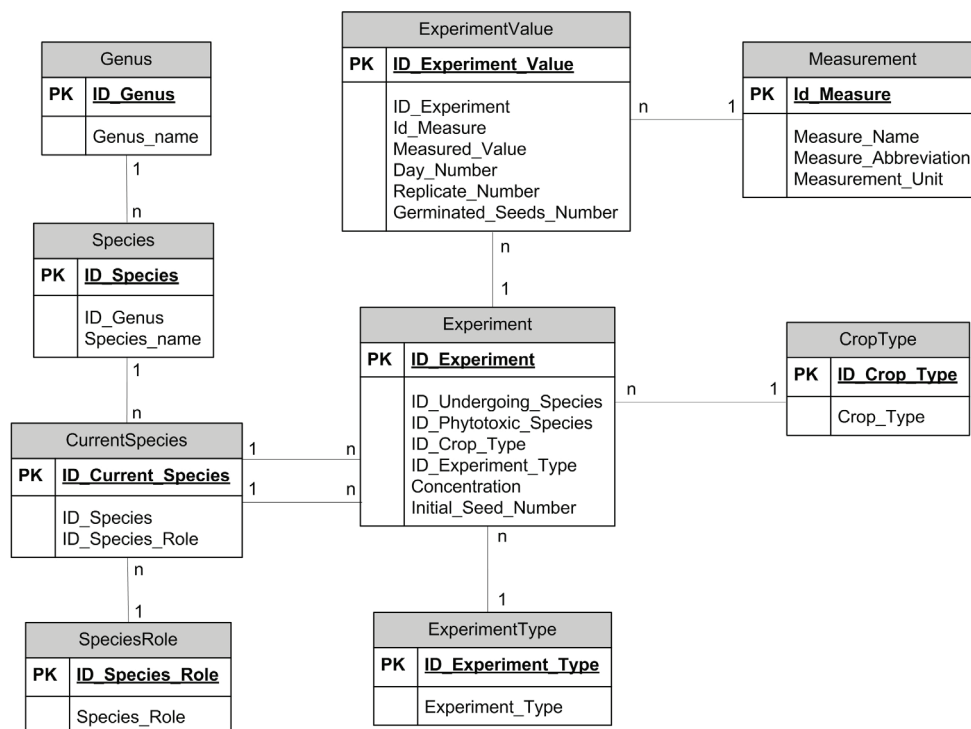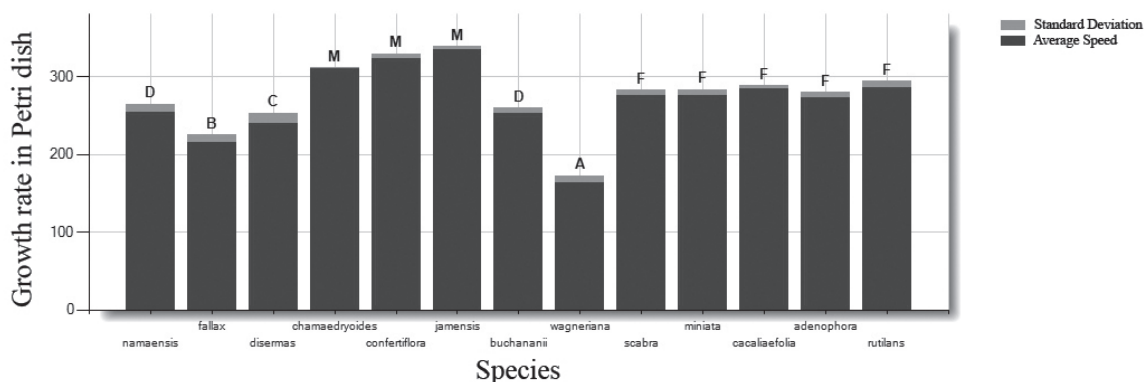

Figure 2. Database design diagram.

Figure 3. The growth rate ($\frac{mL}{day \cdot \mu g}$) ANOVA analysis of each *Salvia* species in Petri dish experiments. The evaluation of germination speed of *Papaver* and *Avena* is considered in response to the various *Salvia* exudates (*Salvia namaensis*, 1; *S. fallax*, 2; *S. disermas*, 3; *S. chamaedryoides*, 4; *S. confertiflora*, 5; *S. x jamensis*, 6; *S. buchananii*, 7; *S. wagneriana*, 8; *S. scabra*, 9; *S. miniata*, 10; *S. cacaliaefolia*, 11; *S. adenophora*, 12; *S. rutilans*, 13). The mean GWAV plus standard deviation is reported. The letters A, B, C, D, F and M indicates the groups of species defined by the ANOVA application followed by the Duncan multiple range test.
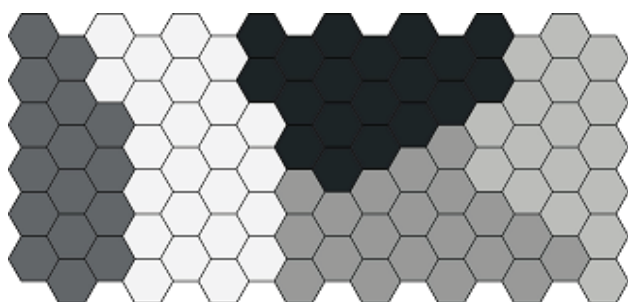
Figure 4. A Kohonen map: an example of a clustering result. This map shows a range of grey-scale colours, but normally each cluster has a random colour. The SOM algorithm randomly assigns the cluster, so the specific colour is not relevant, but the boundaries of each area are important because they represent the clustering.

number of clusters to search at five. An example of the clustering test results is shown in Figure 4.

Since five was the maximum number of clusters that was searched for, the algorithm in some cases found three or four clusters, when convergence was reached in advance. Table 1 provides a summary of clustering tests results for the types of culture and species used.

The clustering was found to be fairly homogeneous although the tests are independent of each other. As referred to previously, the order of neurons is randomised for each test to ensure that the tests have the same input data but in random order, so that the operation of clustering does not depend on the order of the data input.

As shown in Table 1 most of the tests related to *Avena* pot and *Papaver* Petri dish experiments gave only two clusters. This limited statistical evidence of the real presence of different clusters, has led to the development of an algorithm based on the Kohonen map topology

Table 1. Percentage of cluster distribution among the 100 performed tests.

No. of clusters	Experiment type			
	Avena (Petri)	*Avena* (pot)	*Papaver* (Petri)	*Papaver* (pot)
2	-	96%	80%	-
3	26%	-	14%	-
4	17%	-	1%	17%
5	57%	4%	5%	83%

and inspired by the Indicator Value Analysis method of Dufrène and Legendre (Dufrène & Legendre, 1997) in order to obtain a ranking of the activity of the various *Salvia* species. The Indicator Value Analysis is a statistical analysis widely applied in biogeography, that looks for the species typical of areas of endemism. The Indicator Value Analysis introduces an index that combines the specificity and fidelity method to quantify the Indicator Species significance: indicator species are the ones mainly found in a single area of endemism and present in the major part of the subzones constituting the area (Dufrène & Legendre, 1997).

$$IndVal_i = \max[\ IndVal_{ij}\ ] = \max[\ A_{ij} \cdot B_{ij} \cdot 100] = \max\left[\frac{N_{sites_{ij}}}{N_{sites_i}} \cdot \frac{N_{sites_{ij}}}{N_{sites_j}} \cdot 100 \right]$$

where *IndVal_i* is the Indicator Value for the species *i; j* is the Operational Geographic Unit (OGU); N_{sites} is the number of sites.

Previous experience of some authors, shows how Indicator Species Analysis can be applied to obtain a quantitative estimate for the endemic relevance of a species in a certain area of endemism. Due to the analogy between the two fields, this methodology has been applied also to

the phytotoxicity study explained in this work (Pivetti et al., 2009). The obtained results are reported in Chart 1.

Also this new set of calculation rely on the relational structure of the DB. The present tool can be considered as a prototype for a fast and efficient management of such type of data. This further application of the tool will be able to also prove its versatility.

Conclusions

The system presented is based on a database structure that has been applied to this particular study. Therefore it represents the basis for a valid interaction between the users from different backgrounds. A clear structure and organisation of the information are highly important features in order to avoid redundancy, assure data consistency and encourage data communication. The authors sustain that this system is suitable also for other similar cases in the context of botany. In fact, for several years, these tools have been used in this field, for example in synthesis of nanostructures (Giacomini et al., 2009).

The first approaches on this study were published in the Crop Protection journal (Bisio et al., 2010b). Consequently the system proposed allows the user to appreciate directly obtained results, through the use of a very complicated statistical tool, as well as the ANOVA or the SOM, without the need for a thorough knowledge of these tools. In fact this tool provides an equally high level of security as MATLAB®, for example, without interacting with highly sophisticated functions. This interface is used in order to explain the results obtained by these tools.

Acknowledgment

References

Basheer IA, Hajmeer M 2000. Artificial neural networks: fundamentals, computing, design, and application. *J Microbiol Meth 43*: 3-31.

Batini C, Ceri S, Navathe SB 1992. *Conceptual database design. An entity-relationship approach* Redwood City, California, USA: The Benjamin/Cummings Publishing Company.

Bernstein S, Bernstein R 1999. *Elements of statistics II: Inferential Statistics* Boulder, Colorado, USA: McGraw-Hill Professional.

Bertone S, Giacomini M, Ruggiero C, Piccarolo C, Calegari L 1996. Automated systems for identification of heterotrophic marine bacteria on the basis of their fatty acid composition. *Appl Environ Microbiol 62*: 2122-2132.

Bhowmik PC, Inderjit 2003. Challenges and opportunities in implementing allelopathy for natural weed management. *Crop Prot 22*: 661-671.

Bisio A, Fraternale D, Damonte G, Giacomini M, Giacomelli E, Salis A, Russo E, Cafaggi S, Ricci D, Romussi G, De Tommasi N 2010a. Attività fitotossica di specie di Salvia. *FITOMED 2010 - VI Congresso Intersocietà sulle Piante Medicinali Cagliari.*

Bisio A, Fraternale D, Giacomini M, Giacomelli E, Pivetti S, Russo E, Caviglioli G, Romussi G, Ricci D, De Tommasi N 2010b. Phytotoxicity of *Salvia* spp. exudates. *Crop Prot 29*: 1434-1446.

Bradley P, Fayyad U 1998. Refining initial points for K-means clustering. *The Fifteenth International Conference on Machine Learning,* p. 91-99.

Chart 1. Ranking of *Salvia* species according to the method of Dufrène & Legendre (1997).

Avena		*Papaver*	
Petri dish	Pot	Petri dish	Pot
S. jamensis	S. namaensis	S. chamaedryoides	S. confertiflora
S. chamaedryoides	S. jamensis	S. cacaliaefolia	S. buchananii
S. confertiflora	S. fallax	S. namaensis	S. rutilans
S. namaensis	S. chamaedryoides	S. disermas	S. jamensis
S. cacaliaefolia	S. confertiflora	S. wagneriana	S. cacaliaefolia
S. miniata	S. scabra	S. miniata	S. wagneriana
S. rutilans	S. miniata	S. adenophora	S. adenophora
S. buchananii	S. rutilans	S. fallax	S. disermas
S. scabra	S. adenophora	S. jamensis	S. scabra
S. adenophora	S. wagneriana	S. scabra	S. namaensis
S. disermas	S. disermas	S. confertiflora	S. miniata
S. wagneriana	S. buchananii	S. buchananii	S. chamaedryoides
S. fallax	S. cacaliaefolia	S. rutilans	S. fallax
Control	Control	Control	Control

Bramley R, Chiu K, Devadithya T, Gupta N, Hart C 2006. Instrument monitoring, data sharing, and archiving using common instrument middleware architecture (CIMA). *J Chem Inf Model 46*: 1017-1025.

Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA 2003. ArrayExpress a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res 31:* 68-71.

Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R 2004. The Gene Ontology Annotation (GOA) Database: sarin knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res 32*: 262-266)

Chiapusio G, Sanchez AM, Reigosa MJ, Gonzalez L, Pellissier F 1997. Do germination indices adequately reflect allelochemical effects on the germination process? *J Chem Ecol 23*: 2445-2453.

Chu X-Z, Gao L, Qiu H-B, Li W-D, Shao X-Y 2010. An expert system using rough sets theory and self-organizing maps to design space exploration of complex products. *Expert Syst Appl 37*: 7364-7372.

Dayan FE, Romagni JG, Duke SO 2000. Investigating the mode of action of natural phytotoxins. *J Chem Ecol 26*: 2079-2094.

Dufrène M, Legendre P 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol Monograph 67*: 345-363.

Duncan DB 1955. Multiple range and multiple F tests. *Biometrics 11*: 1-42.

Fink T, Reymond JL 2007. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: Assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *J Chem Inf Model 47*: 342-353.

Gasteiger J, Teckentrup A, Terfloth L, Spycher S 2003. Neural networks as data mining tools in drug design. *J Phys Org Chem 16*: 232-245.

Giacomini M, Pastorino L, Caneva Soumetz F, Mielczarski JA, Mielczarski E, Rangelow I, Gotszalk T, Glezos N, Huq E, Ruggiero C 2009. Data modeling for tools and technologies for the analysis and synthesis of NANOstructures (TASNANO) project. *JITR 2*: 49-70.

Giacomini M, Cafferata M 2009. WEB Based Tool for the Organization of Knowledge in Neurology. In: Schlegel ODaWC (Ed.). *World Congress 2009 IFMBE Proceedings* Munich: Springer & Verlag, p. 285-288.

Giacomini M, Ruggiero C 2001. Application of artificial neural networks in medicine and biology. In: Mohan RM (Ed.). *Res. Adv. in Biomedical* Eng. Trivandrum, Kerala, India: Global Reasearch Net Work, p. 137-147.

Giacomini M, Ruggiero C, Calegari L, Bertone S 2000. Artificial neural network based identification of environmental bacteria by gas-chromatographic and electrophoretic data. *J Microbiol Meth 43*: 45-54.

Giacomini M, Ruggiero C, Bertone S, Calegari L 1997. Artificial Neural Network Identification of Heterotrophic Marine Bacteria Based on Their Fatty-Acid Composition. *IEEE Trans Biomed Eng 44*: 1185-1191.

Giacomini M, Ruggiero C, Maillard M, Lillo FB, OE V 1996.

Objective evaluation of two markers of HIV-1 infection (p24 antigen concentration and CD41 cell counts) by a self organising neural network. *Med Inform 21*: 215-228.

Gogoulou A, Gouli E, Grigoriadou M, Samarakou M 2005. ACT: A Web-Based Adaptive Communication Tool. *Computer Supported Collaborative Learning 2005: The Next 10 Years!* Taipei, p. 180-189.

Halsall F 2005. The World Wide Web. *Computer Networking and the Internet*. 5th ed. Harlow, Essex, England: Pearson Education U.K.

Jefferson LV, Pennacchio M 2003. Allelopathic effects of foliage extracts from four Chenopodiaceae species on seed germination. *J Arid Environ 55*: 275-285.

Kalinichenko L, Missikoff M, Schiappelli F, Skvortsov N 2003. Ontological Modelling. *Proceedings of the 5th Russian Conference on Digital Libraries* RCDL2003, St.-Petersburg, Russia.

Kohonen T 2001. *Self-Organizing Maps*. 3rd ed. New York, New York, USA: Springer.

Krasznai E, Borics G, Várbíró G, Abonyi A, Padisák J, Deák C, Tóthmérész B 2010. Characteristics of the pelagic phytoplankton in shallow oxbows. *Hydrobiologia 639*: 173-184.

Laterra P, Bazzalo ME 1999. Seed-to-seed allelopathic effects between two invaders of burned Pampa grasslands. *Weed Res 39*: 297-308.

Pivetti S, Giacomini M, Casazza G, Minuto L 2009. Data mining for floral distribution pattern detection. In: Schlegel ODaWC (Ed.). *World Congress 2009 IFMBE Proceedings* Munich: Springer & Verlag, p. 230-233.

Ranal MA, Garcia De Santana D 2006. How and why to measure the germination indices? *Rev Bras Bot 29*: 1-11.

Rumelhart DE, Hinton GE, Williams RJ 1986. Learning internal representations by error propagation. In: Press M (Ed.). *Parallel Distributed Processing* Cambridge, Cambridgeshire, England.

Selzer P, Ertl P 2006. Applications of self-organizing neural networks in virtual screening and diversity selection. *J Chem Inf Model 46*: 2319-2323.

Thorisson GA, Smith AV, Krishnan L, Stein LD 2005. The International HapMap Project Web site. *Genome Res 15*: 1592-1593.

Vracko M 2005. Kohonen artificial neural network and counter propagation neural network in molecular structure-toxicity studies. *Curr Comput-Aided Drug Des 1*: 73-78.

Webster C 2002. A methodology for incorporating web technologies into a computer-based patient record, with contributions from cognitive science. *Int J Med Inform 68*: 39-47.

Xuan TD, Tawata S, Khanh TD, Chung IM 2005. Decomposition of allelopathic plants in soil. *J Agron Crop Sci 191*: 162-171.

*Correspondence

M. Giacomini
Dipartimento di Informatica, Sistemistica e Telematica, Università di Genova
Via all'Opera Pia 16145 Genova, Italia
Mauro.Giacomini@dist.unige.it
Tel. +39 010 353 6546
Fax: +39 010 353 2154