A MATHEMATICAL METHOD TO ESTIMATE THE UNDERGROUND DEVELOPMENT OF THE CLONAL SPECIES

Anna Jakubska[1], Małgorzata Dudkiewicz[2], Paweł Jankowski[3], Radosław Sikora[4]

[1]Department of Biodiversity and Plant Cover Protection
Institute of Plant Biology, University of Wrocław
Kanonia 6/8, 50-328 Wrocław, Poland
e-mail: Ajak@biol.uni.wroc.pl
[2]Department of Experimental Design and Bioinformatics
Faculty of Agriculture and Biology, Warsaw University of Life Sciences
Nowoursynowska 159, 02-776 Warszawa, Poland
e-mail: malgorzata_dudkiewicz@sggw.pl
[3]Department of Biometry
Faculty of Applied Informatics and Mathematics, Warsaw University
of Life Sciences
Nowoursynowska 159, 02-776 Warszawa, Poland
e-mail: pawel_jankowski@sggw.pl
[4]Television Division, Institute of Radioelectronics
Faculty of Electronics and Information Technology, Warsaw University
of Technology Nowowiejska 15/19, 00-665 Warszawa, Poland
e-mail: R.Sikora@ire.pw.edu.pl

Summary

One of the most serious problems for field researchers investigating clonal species is to distinguish separate plants (genets) of which the population is made up. In the case of many endangered species the difficulty is connected with the restrictions imposed by their protected status. We present an application of the graph theory Minimal Spanning Tree model combined with the variance analysis formulae for estimation of the underground development of the rhizome which as we believe, can be used for many clonal species. We analyse a practical application of the

method for an exemplary *Epipactis helleborine* (L.) Crantz population. Hypothetical structure of its underground rhizome connections is pictured showing the growth scheme of the population.

1. Introduction

The Minimal Spanning Tree (MST) model has been used in numerous fields of modern biology, for instance to perform the multiple sequence threading, in protein fold recognition and in homology modelling (Taylor, 1997; Aszodi *et a*l. 1997), to identify the history of transmission of a virus in epidemiological investigations (Spada *et al.*, 2004; Chiou *et al,.* 2006). The MST calibration algorithm is also well suited to calibrate MS spectra of protein samples resulting from two-dimensional gel electrophoretic separation (Wolski *et al.*, 2005). In this article we propose its application in botanical and ecological studies.

One of the most serious problems for field researchers investigating clonal species is to distinguish separate plants (genets) of which the population is made up. Obviously a direct observation of the underground connections among the visible aboveground stems is not always possible as is the case for the protected species. One is forced then to apply indirect approaches such as genes or external features comparison. These methods can be very expansive, difficult and time consuming. We present a mathematical method which in the case of some clonal species may be used instead. As the method was developed for the orchid species, our approach takes into account the so called "occurrence fluctuation" or dormancy phenomenon reported for many orchids. The model is based purely on the observation of the stems spatial distribution, distances among them and the sequence of their appearances and disappearances.

The method will be described in details in the following section of the article. Next, we describe an application of the model for the protected *Epipactis helleborine* (L.) Crantz population observed in the Eastern Sudety Mountains (SW Poland).

2. The model

In our model we apply notions introduced by the graph theory, see for instance (Diestel, 2005). The basic objects in graph theory are graphs which are sets of points called vertices connected by lines called edges. Graphs may connect all or a set of vertices and may include loops defined as edges which start and end on the same vertex. In many applications the so called weighted graphs are used. In a weighted graph each edge has a number (weight) assigned representing for instance cost, time or length.

This work is based on the Euclidean graphs being weighted graphs with the Euclidean lengths defining the edge weights. The spanning tree is a graph connecting all vertices with no loops allowed. There may be many possible spanning trees for the given set of vertices. The one with the smallest total weight (defined as the total length of the graph edges in the case of a Euclidean graph) is called the Minimal Spanning Tree (MST). In the case of some clonal species, the specimen is build of stems connected by rhizomes. For each separate plant the stems can be identified as vertices and rhizome segments (internodes) connecting two stems as edges of a graph (Fig.1).


Fig.1. Inflorescense of *Epipactis helleborine* L. (Crantz) and the general assumptions of the model: the aboveground stems as graph vertices and the rhizomes creating connecting edges

Our model is designed for long-term observations and takes into account the so called "occurrence fluctuation" which signifies the possibility of individual stems to disappear one year and reappear another one, with the possible period of dormancy of a few years. The approach is based on the assumption that each rhizome "chooses the shortest way" among its stems. It results from the observation that the most economical search for nutrients by any plant is in its proximity and that a plant does not return to a place already utilized. The shortest way among the stems means that from all possible combinations of stem connections we choose one with the shortest sum of all the rhizome segments. Hence, each plant creates its own MST based on the Euclidean rhizome lengths. The other model assumptions are the following:

1. Rhizomes may branch.
2. The stem which does not reappear in any of the observed years is treated as dead and is removed from the model.
3. The stem which disappears given year but reappears in any of the future years is treated as dormant and remains in the model.
4. The part of a rhizome connecting two stems is removed from the model if it is a final part of a rhizome branch with a dying stem at its end (for graphical explanation see Fig. 2a) unless it is a part of a branch developing during the next year (see Fig. 2b) meaning that though the stem has died the rhizome connection has not been broken. The part of the rhizome connecting a stem which dies but lies in the middle of a branch is not removed from the model because one can not judge whether this connection has been broken or not (see Fig. 2c).
5. We do not assume the maximal length of the rhizome segment (an internode between two connected stems).
6. The model is two dimensional which means that it does not take into account any irregularities of the ground level.

The analysis is divided into two phases. First, one should consider all segments visible during the time of observation longer then the period affected by the "occurrence fluctuation" phenomenon to obtain rough approximation of the rhizomes course. Next, in the second phase, each year one adds new stems and checks whether any of the old stems and rhizome segments died and should be removed from the model.

In the first phase we divide stems visible during first years of observations into clusters forming separate rhizomes. We apply the Prim algorithm (Prim, 1957) of the graph theory and iteratively create the MST composed of the stems. We begin the iteration allowing for connections between the stems of any length and producing a MST made of all stems. Next, in each step of iteration we forbid all connections longer or equal to the maximal connection in the MST ob-
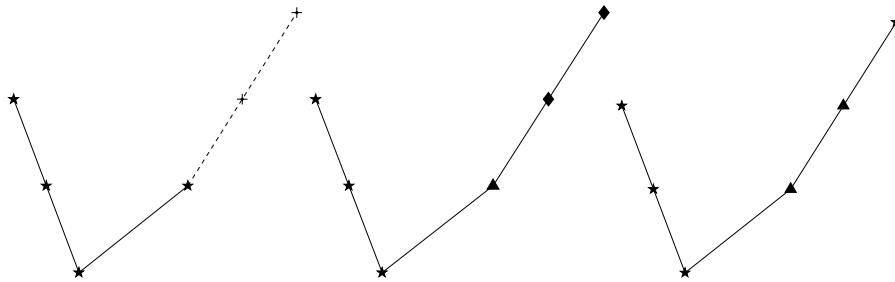
Fig. 2. The plots demonstrate when a rhizome segment connecting two stems is to be removed according to the model described within the text of the article. The stars and diamonds denote old and new stems observed given year, respectively. The plus signs denote the stems which disappeared given year and were not observed in future. Finally, the upper triangles signify stems which disappeared this or one of former years but we can not judge whether their connection with other stems *via* rhizome was broken

tained in the previous step. That way in each step we obtain a different number of separate MSTs which we identify with separate rhizomes. In order to choose the final step of iteration we use the formulae of the classical one-way ANOVA analysis in the same way as it is applied in the cluster segmentation procedures. In the Agglomerative Hierarchical Clustering or k-means methods the appropriate number of clusters is usually chosen basing on the comparison of the variability of the clustered objects within and between the clusters (given by the *F* ratio). In presented approach clustering is performed through MSTs and the variability of the objects is given by their Euclidean distances. Hence, in each step of iteration we calculate the *F* ratio of the sum of squared distances within rhizomes to the sum of squared distances between the rhizomes:
within:

$$q_r = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (p_{ij} - \overline{p}_i)^2 , \qquad (2.1)$$

between:

$$q_g = \sum_{i=1}^{k} \left(p_i - \overline{p} \right)^2 n_i , \qquad (2.2)$$

$$F = \frac{q_r (k-1)}{q_g (N-k)} , \qquad (2.3)$$

where: N - number of all stems (equivalent of number of all observations in classical one-way ANOVA), k - number of rhizomes (equivalent of number of independent groups used in comparison), n_i - number of stems in the i-th rhizome, p_{ij} – location of the j-th stem in the i-th rhizome defined as coordinates of the given stem in Cartesian coordinate system of the experimental field, $\overline{p}_i$ – average location of the stems in the i-th rhizome –, $\overline{p}$ – average location of all stems. The final set of MSTs is chosen from the results of the iteration for which the F value is minimal. Hence MSB and MSW values calculated for every possible configuration of stems clusters were used to obtain F statistics which allowed us to evaluate the probability of every given configuration.

In the second phase of the analysis each year we add new stems appearing for the first time to the rhizome structure already developed during the former years. We choose a stem for which we find an old stem such that the distance between them is smaller than for any other pair of an old and new stem. That way we also find a rhizome to which the chosen stem should be connected. We repeat this procedure until all new stems have been chosen.

3. Example application

The method described in this article has been applied to the protected orchid, *Epipactis helleborine* (L.) Crantz, population observed in the Eastern Sudety Mountains (SW Poland). The study site has been chosen in the Krowiarki Range of Eastern Sudety Mountains in the south of Poland. An area 12x18 meters containing a population of *E. helleborine* was selected for a study in July 1997. The positions of individuals were mapped between 1997 and 2006. We present in this article results of first four years of the experiment. The observations were carried out under field conditions during the peak of flowering period. Position of each stem was precisely measured and marked on a grid.

As the main tool for computer data analysis we applied Matlab 6.5 (The Math Works, Inc, Natick, Massachusetts).

The species belonging to the *Epipactis* family fulfil the assumptions of our model. Fig. 3 shows the typical rhizome, Fig. 4 a rhizome connecting two stems, which begins to fragmentize and Fig. 5 a new branching of the rhizome of an *Epipactis helleborine* (L.) Crantz population. In Fig. 6 we present a photograph of an *Epipactis palustris* (L.) Crantz stem with its system of branching rhizomes.

The *Epipactis helleborine* stems may disappear and reappear even after 18 years of dormancy but such long periods of absence are very rare (Light and

MacConaill, 2006). We take 3 years as the length of the first phase of the model in which all stems observed during this time are analysed together. Hence, we combine years 1997-1999 of the experiment. Next we add stems which appeared for the first time in year 2000 and check whether any of the old stems and rhizome segments died and should be removed from the model. We obtain a hypothetical map of the underground development of the examined population presented in Fig. 7. As can be seen new stems extend the lines of the old rhizome showing a reasonable scheme of the clonal growth. We distinguish 13 individual plants forming the population. Such a situation corresponds with the growth pattern of one of the populations genetically examined in NE Poland, on the slope of the forested hill at the Białe Lake (Brzosko *et. al,* 2004). In both cases the observed low level of genetic diversity can be explained by vegetative spread, chosen cause of favourable habitat conditions (low cover of other species).
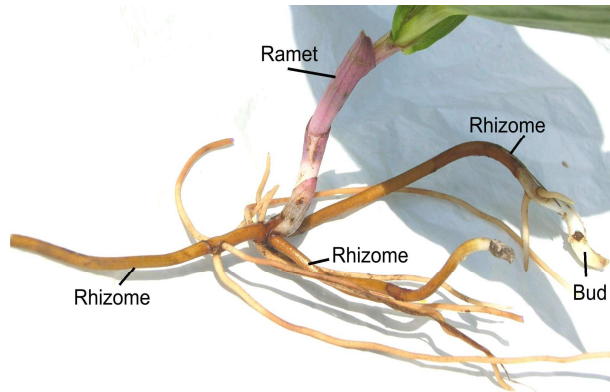

Fig. 3. The initiation of the rhizome fragmentation (*Epipactis helleborine* (L.) Crantz)

Fig. 4. A rhizome connecting two stems of *Epipactis helleborine* (L.) Crantz

Fig. 5. Branching of the rhizome *of Epipactis helleborine* (L.) Crantz


Fig. 6. Branched rhizome, bud and ramet of *Epipactis palustris* (L.) Crantz, other species of *Epipactis* genus

4. Conclusions

This article presents a new approach to prediction of the underground growth of the rhizomial species on the example of protected orchid, *Epipactis helleborine* (L.) Crantz. We used a simple mathematical model based on the graph theory which allowed us creating hypothetical maps of the rhizome distribution. We believe that this method if proved valid for a given species can be complementary to the expensive and time consuming genetic methods.
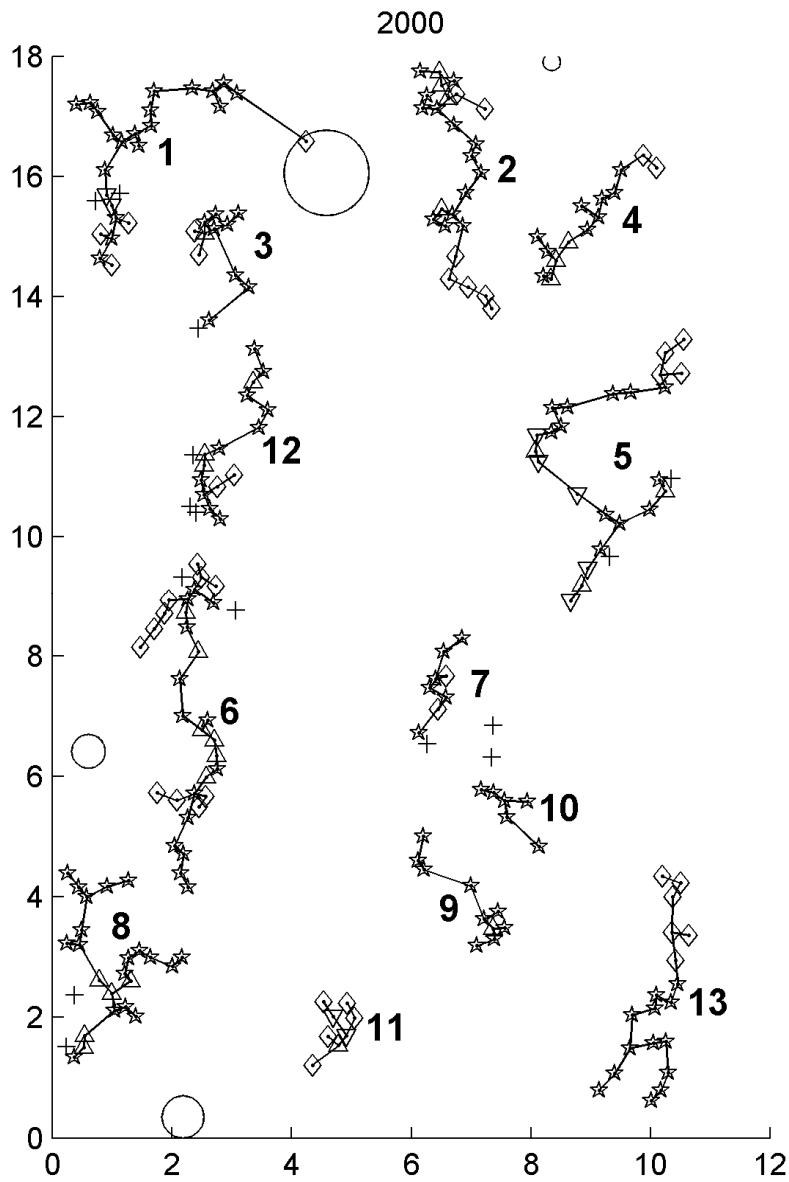
Fig. 7. Map of the rhizomes for the year 2000. The stars, diamonds and lower triangles denote old and new stems observed given year and old stems absent given year, respectively. The upper triangles signify stems which disappeared this or one of former years but we can not judge whether their connection with other stems via rhizome was broken. Finally, the plus signs denote the stems which do not reappear in any future year of observations and are treated as dead. The big circles denote trees.

References

Aszódi A., Munro R.E., Taylor W.R. (1997). Protein modeling by multiple sequence threading and distance geometry. *Proteins, Suppl* 1, 38-42.

Brzosko E., Wróblewska A., Talalaj I. (2004). Genetic variation and genotypic diversity in *Epipactis helleborine* populations from NE Poland. *Plant Systematics and Evolution* 248, 57-69.

Chiou C.S., Liao J.C., Liao T.L., Li C.C., Chou C.Y., Chang H.L., Yao S.M., Lee T.S. (2006). Molecular epidemiology and emergence of worldwide epidemic clones of *Neisseria meningitidis* in Taiwan. *BMC Infectious Diseases* 6, 25.

Diestel R. (2005). *Graph Theory*, 3 edn. Springer.

Light M.H.S. , MacConaill M. (2006). Appearance and disappearance of a weedy orchid, *Epipactis helleborine. Folia Geobotanica* 41, 77–93.

Prim R.C. (1957). Shortest connection networks and some generalisations. *Bell System Technical Journal* 36, 1389-1401.

Spada E., Sagliocca L., Sourdis J., Garbuglia A.R., Poggi V., De Fusco C., Mele A. (2004). Use of the Minimum Spanning Tree model for molecular epidemiological investigation of a nosocomial outbreak of Hepatitis C Virus infection. *J. Clin. Microbiol.* 42, 4230–4236.

Taylor W.R. (1997). Multiple sequence threading: an analysis of alignment quality and stability. *J. Mol. Biol.* 269, 902-43.

Wolski W.E, Lalowski M., Jungblut P., Reinert K. (2005). Calibration of mass spectrometric peptide mass fingerprint data without specific external or internal calibrants. *BMC Bioinformatics* 6, 203.

Zozio T., Allix C., Gunal S., Saribas Z., Alp A., Durmaz R., Fauville-Dufaux M., Rastogi N., Sola C. (2005). Genotyping of *Mycobacterium tuberculosis* clinical isolates in two cities of Turkey: description of a new family of genotypes that is phylogeographically specific for Asia Minor. *BMC Microbiol* 26, 5-44.

MATEMATYCZNA METODA WYZNACZANIA PRZEBIEGU KŁĄCZA ROŚLIN KLONALNYCH

Streszczenie

Jeden z najpoważniejszych problemów, na jakie napotykają badacze gatunków roślin klonalnych w trakcie obserwacji terenowych stanowi problem wyróżnienia indywidualnych roślin (indywidualnych klonów), z jakich składa się dana populacja. W przypadku wielu zagrożonych gatunków trudność związana jest z tym, iż posiadają one statut roślin chronionych niepozwalający

na bezpośrednią obserwację kłączy. W pracy prezentujemy zastosowanie modelu Minimalnego Drzewa Rozpinającego znanego z teorii grafów w powiązaniu z wzorami zaczerpniętymi z analizy wariancji do opisu rozwoju podziemnego kłączy. Jak wierzymy nasze podejście może być zastosowane dla wielu gatunków klonalnych. W celu ilustracji przedstawianej metody prezentujemy jej praktyczne zastosowanie dla przykładowej populacji storczyka odmiany *Epipactis helleborine* (L.) Crantz. Wynikiem analizy jest hipotetyczna struktura podziemnych połączeń kłączy obrazująca rozwój populacji.

Słowa kluczowe: Minimalne Drzewo Rozpinające, *Epipactis helleborine* (L.) Crantz, rozwój populacji

Klasyfikacja AMS 2000: 92D25, 05C90, 05C05