Plant Disease Outbreak – Prediction by Advanced Data Analysis

Outi Mäyrä[1*], Mika Ruusunen[1], Marja Jalli[2], Lauri Jauhiainen[2], Kauko Leiviskä[1]

[1]Control Engineering, University of Oulu, P.O.Box 8000, FI-90014, Finland; *outi.mayra@oulu.fi
[2]Natural Resources Institute Finland, Humppilantie 14, 31600 Jokioinen, Finland

Abstract. In crop farming, pests and plant diseases cause losses for both grain yield and quality. The fungicides are used for yield protection. With the accurate plant disease prediction system, the farmers could optimize the chemical spraying and save money and environment. The plant disease modelling is challenging and widely discussed topic in modern agriculture. This short paper introduces the novel point of view for the net blotch prediction. In this approach, the weather measurements are combined with the historical net blotch observations and the data based prediction system is introduced. The estimation scheme is demonstrated with the example of the selected observation field data.

Introduction

The net blotch is a common barley disease caused by fungus *Pyrenophora teres*. It exists all over the world and affects both grain yield and malt quality [Bogacki et al. 2010]. For example, in Finland the net blotch was present in 86% of the investigated barley fields in 2009 (Jalli et al. 2011).

The weather has an influence on the occurrence of plant diseases and the different pathogens need different conditions leading into the infection. The prediction of the plant diseases can be a valuable tool for optimizing the use of the fungicides, but the reliable prediction is complicated. [Hardwick 2002].

The minimized usage of chemicals save money and environment and with the accurate prediction, the chemicals could be sprayed only when needed. The data analysis and modelling, but also knowledge about plant diseases, are the components for reliable disease forecasting [Hardwick 2006].

Cunniffe et al. (2015) discussed thirteen challenges in plant disease modelling. The authors focused particularly on disease prediction and control with epidemiological models. In that article, the challenges are partitioned into three groups: Modelling the plant host, modelling the pathogen, and modelling for control.

In this short paper, the data based modelling with feature generation is applied to the estimation of net blotch occurrence. The main principles of the prediction system are shortly described and the results are discussed. Generally, the aim of this research is the adequate prediction accuracy and the simple model structure for prediction the selected plant disease as in (Mäyrä et al. 2018).

1 Materials and Methods

Two different datasets – the weather data from the open database of Finnish Meteorological Institute (FMI) and the data of the net blotch observations are combined and utilized for net blotch prediction. The observations of the net blotch are collected and pre-processed by The Natural Resources Institute Finland (Luke) during the years 1991 – 2015. The data included the information about the net blotch observations in twenty different localities with the varying time scale. The fields, which are used as an example here, locate in the Southern part of Finland. The weather conditions and the beginning date of the growing season alternate between years during the observation period. To make different years comparable, the data was normalized and the time step 0 is the beginning of the growing season in every data set. The whole analysis and evaluation of the results are performed in the Matlab® software environment. The principle of this study – the utilization of data fusion and advanced data analysis is presented in Figure 1.
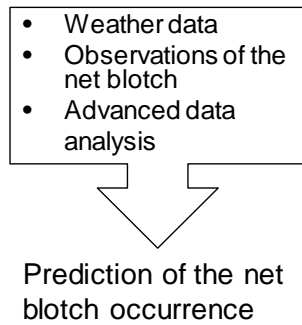
- Weather data
- Observations of the net blotch
- Advanced data analysis

Prediction of the net blotch occurrence

Figure 1: The principle of the net blotch prediction; the data fusion and the advanced data analysis.

The yearly weather measurements were divided into separate groups according to the occurrence of the net blotch. In addition, one group included the weather data of those years, when the net blotch was not observed. Later, the years with no net blotch is called as 'normal situation'. The other dataset included weather measurements of those years, when the net blotch appeared. The variables included in the data analysis were

- the place of observation,
- the date of observation,
- the rainfall per day [mm],
- the average temperature per day, (T,avg) [°C],
- the daily minimum temperature, (T,min) [°C] and
- the daily maximum temperature, (T,max) [°C].

The datasets, which were used in this research, included the weather observations from the latest years, specifically four years with, and four years without the existence of the net blotch.

The statistical characteristics of the weather data were first computed to find out any differences between the data sets with or without the net blotch. The statistical characteristics are presented in Table 1.

The feature generation was performed to improve the information content of the data. More about the features and their use are presented, for example, by Dash and Liu (1997) Garcia-Torres et al. (2016), and Pérez-Rodríguez et al. (2015). In this study, feature generation technique presented by Ruusunen (2013, p. 50) was utilized. Briefly, in this method the features are generated by fusing the above listed variables with different mathematical operations. The feature values in the beginning of the growing season are then summed cumulatively. The aim is to separate the cumulative summed feature values according to the net blotch appearance. Weather conditions favouring the net blotch occurrence can be predicted with weather measurements

and suitable features. For this case, the most suitable features were found by exhaustive search and visual inspection. Also T-test was tested in the feature selection step. Further research comprising several observation fields and the feature evaluation has to be automated for example by using T-test. The modelling procedure is demonstrated step by step in Figure 2.
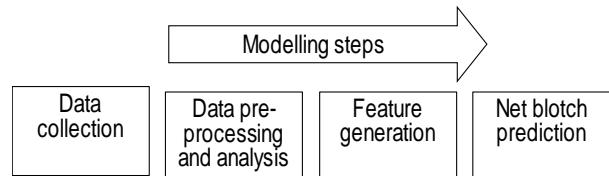
Modelling steps

Data collection → Data pre-processing and analysis → Feature generation → Net blotch prediction

Figure 2: Modelling steps.

2 Results and Discussion

As presented in the Table 1., the datasets seem statistically similar. To find out those years that differs from the normal situation, the feature generation step was performed.

Obs. field	Mean	Std	Median
	No / Yes	No / Yes	No / Yes
Rain fall	1,2 / 1,6	4,2 / 5,5	0,0 / 0,0
T, avg	11,8 / 11,8	5,7 / 5,6	12,5 / 13,0
T, min	6,6, / 6,7	5,6 / 5,9	7,3 / 7,5
T, max	17,3 / 17,1	6,4 / 6,1	17,8 / 17,9

Table 1: The statistical characteristics of the observation field data. The characteristics are presented as per variables (the left column) according to the occurrence of the net blotch (No / Yes).

The grouping of the weather data sets (with or without the net blotch) was next studied with the combination of the feature generation and cumulative sum. Among the total of 110 generated alternatives, the suitable feature for this case was chosen as

$$(x+y)/y, \qquad (1)$$

where x is the rainfall per day and y is the average value of daily temperature.

The years without net blotch observations are presented with the solid lines and the years when the net blotch appeared are plotted with the dash lines in Figure 3. The time scale (x-axis) is 10 days and the time step 0 is the beginning of growing season.
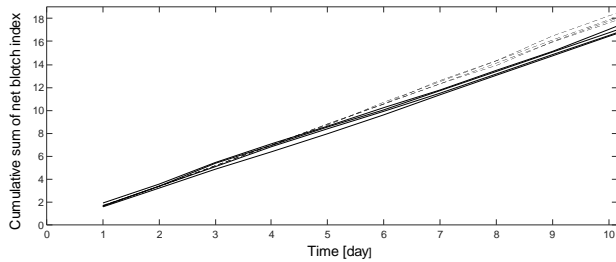
Figure 3: The datasets are grouped by using feature generation. The years with no net blotch (solid lines), the years with the net blotch (dashed lines).

In the presented example, grouping and separation of the data sets according to the appearance of the net blotch was possible by using the feature generation technique. As plotted in Figure 3, the charts forms own "pipes" and the years with the net blotch separate this way from the normal situation.

The results strongly indicate that the presented method is a simple but applicable tool for the prediction of the net blotch occurrence. The simple structure is easy to adapt to another observation fields thanks to non-parametric features. On the other hand, the evaluation of the presented method still needs more example cases. The scopes of the further research are the generalization and validation of the presented prediction method with data sets from different localities. Also the usability in the prediction of the different plant diseases has to be studied.

Acknowledgements

References

[1] Bogacki P, Keiper FJ, Oldach KH. 2010. Genetic structure of South Australian Pyrenophora teres populations as revealed by microsatellite analyses. Fungal Biology 114, 834–841.

[2] Jalli M, Laitinen P, Latvala S. 2011. The emergence of cereal fungal diseases and the incidence of leaf spot diseases in Finland. Agricultural and Food Science 20 1: 62-73.

[3] Hardwick NV. (2002) Weather and plant diseases. Weather, Vol 57, Issue 5, p. 184-190.

[4] Hardwick NV. (2006) Disease forecasting. The Epidemiology of Plant Diseases, Second edition, published by Springer in The Netherlands, p. 239-267.

[5] Cunniffe NJ, Koskella B, Metcalf CJE, Parnell S, Gottwald TR, Gilligan CA. (2015) Thirteen challenges in modelling plant diseases. Epidemics, Challenges in Modelling Infectious DIsease Dynamics 10, 6–10. https://doi.org/10.1016/j.epidem.2014.06.002

[6] Mäyrä O, Ruusunen M, Leiviskä K. (2018) Data based modelling and feature generation as a tool for plant disease forecasting. Accepted for the CAFEi2018 e-Proceeding (2018) The 4th International Conference on Agricultural and Food Engineering on 6th -8th November, 2018 at Kuala Lumpur, Malaysia.

[7] Dash M, Liu H. (1997) Feature selection for classification. Intelligent Data Analysis, Vol 1, p. 131-156.

[8] García-Torres M, Gómez-Vela F, Melián-Batista B, Moreno-Vega JM. (2016) High-dimensional feature selection via feature grouping: A Variable Neighborhood Search approach. Information Sciences, Vol. 326, p. 102–118.

[9] Pérez-Rodríguez J, Arroyo-Peña AG, García-Pedrajas N. (2015) Simultaneous instance and feature selection and weighting using evolutionary computation: Proposal and study. Applied Soft Computing, Vol. 37 p. 416–443.

[10] Ruusunen M. (2013) Signal Correlations in Biomass Combustion - an Information Theoretic Analysis. Acta Univ Oulu C 459, 2013. PhD Thesis.