

# Plant disease outbreak – prediction by advanced data analysis

Outi Mäyrä, University of Oulu, Control Engineering, outi.mayra@oulu.fi  
Mika Ruusunen and Kauko Leiviskä, University of Oulu, Control Engineering  
Marja Jalli and Lauri Jauhiainen, Natural Resources Institute Finland

## Abstract

In crop farming, the plant diseases cause losses for both grain yield and quality. The fungicides are one tool for yield protection. With the accurate plant disease prediction system the farmers could optimize the chemical spraying in crop farming and save both money and environment. In this work, the existing information about weather observations and the data of selected plant disease is combined in the novel way to predict the plant disease outbreak to help farmers spray the fields at the right time.

## Weather observations, the FMI open database

- The weather observations were collected from FMI open database
- The data included the location of information, date, rainfall per day and temperature; daily minimum, maximum and average
- The correlation between a single weather observation and the plant disease outbreak can not be found.

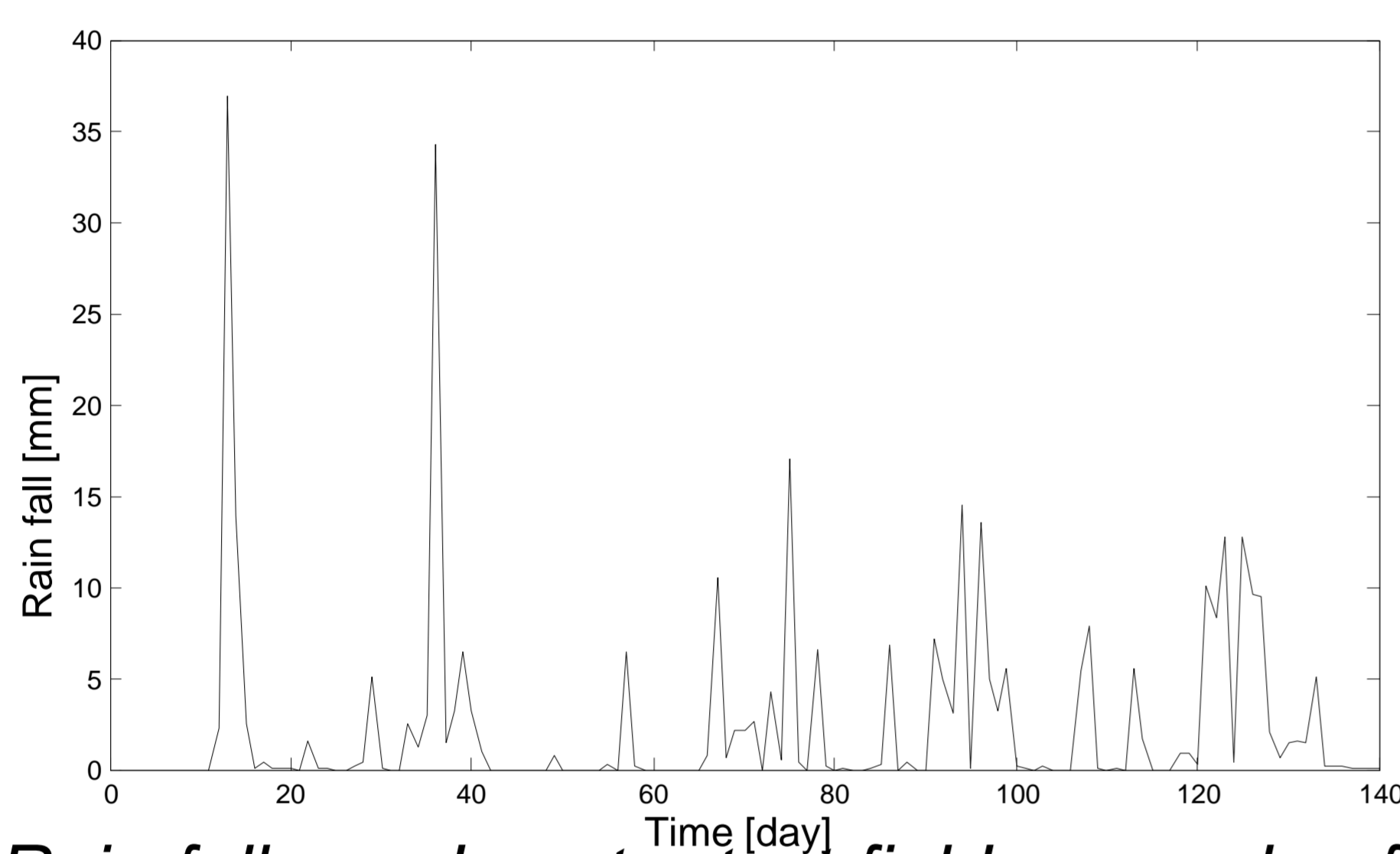


Figure 1. Rain fall per day at a test field, example of data collected from FMI open database.

## Certain plant disease data from Natural Resources Institute Finland

- The information of the selected plant disease is collected by Natural Resources Institute Finland
- The data included the location of information, date, information about the plant disease occurrence and the index describing the amount of plant disease (0-4)
- The data is collected during 1990–2014 from several fields in Finland

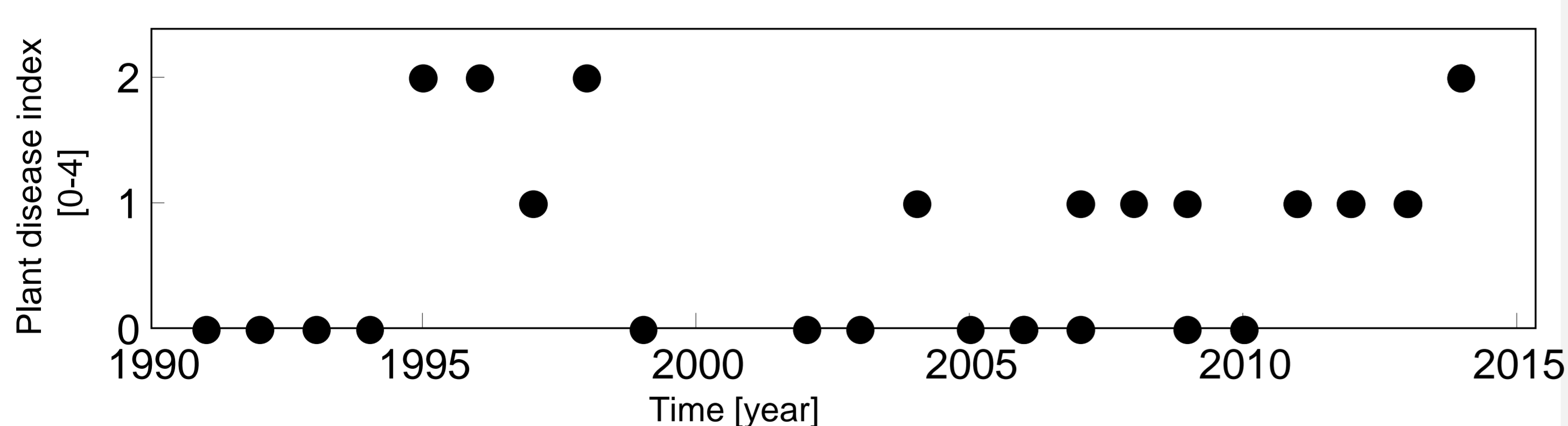


Figure 2. The occurrence of selected plant disease at a test field in 1990–2014. About the plant disease indexes, 0 = not at all and 4 = a lot.

## Advanced data analysis, feature generation and classification

- Feature generation is performed by combining variables with different mathematical operations [1]
- The suitable feature is selected by ranking the generated features
- The weather data timing is performed and the time frame starts from the beginning of the growing season

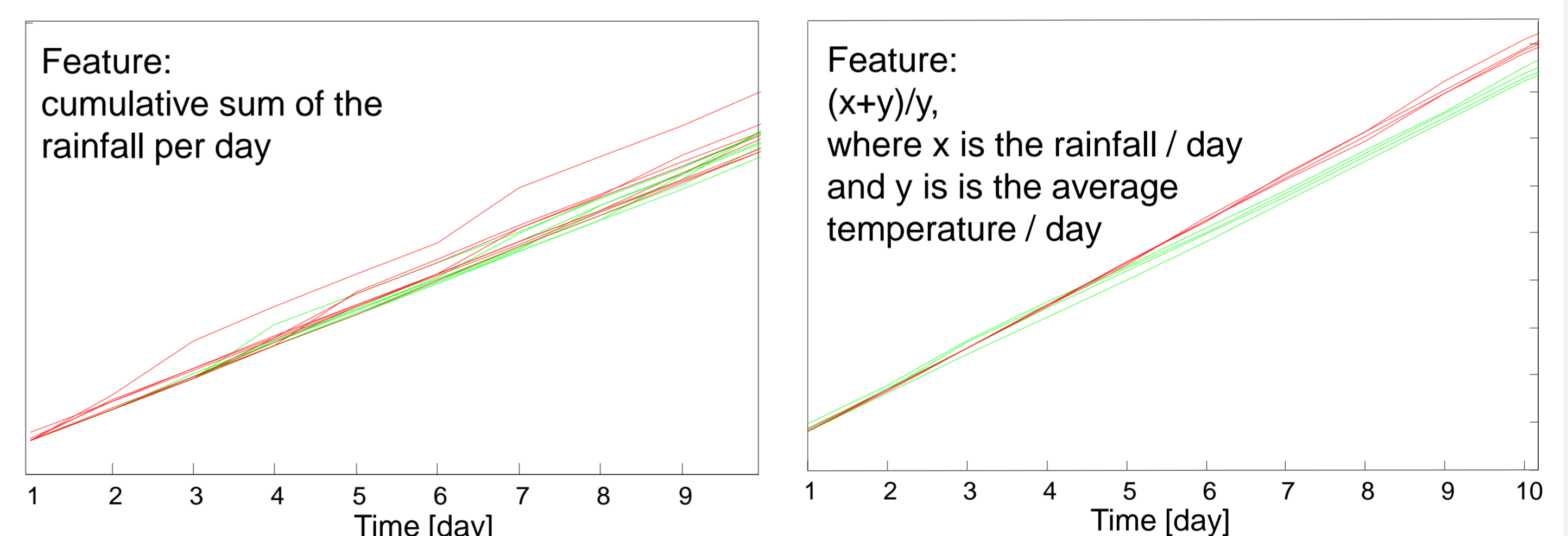


Figure 3. Feature selection. The unfeasible feature on the left and the suitable one on the right.

## The prediction of the selected plant disease, validation of the prediction system

- The prediction system is validated with the year 2015 data
- The grey area describes the weather conditions unfavourable to selected plant disease according to the feature
- The black line predicts the year 2015 situation correctly

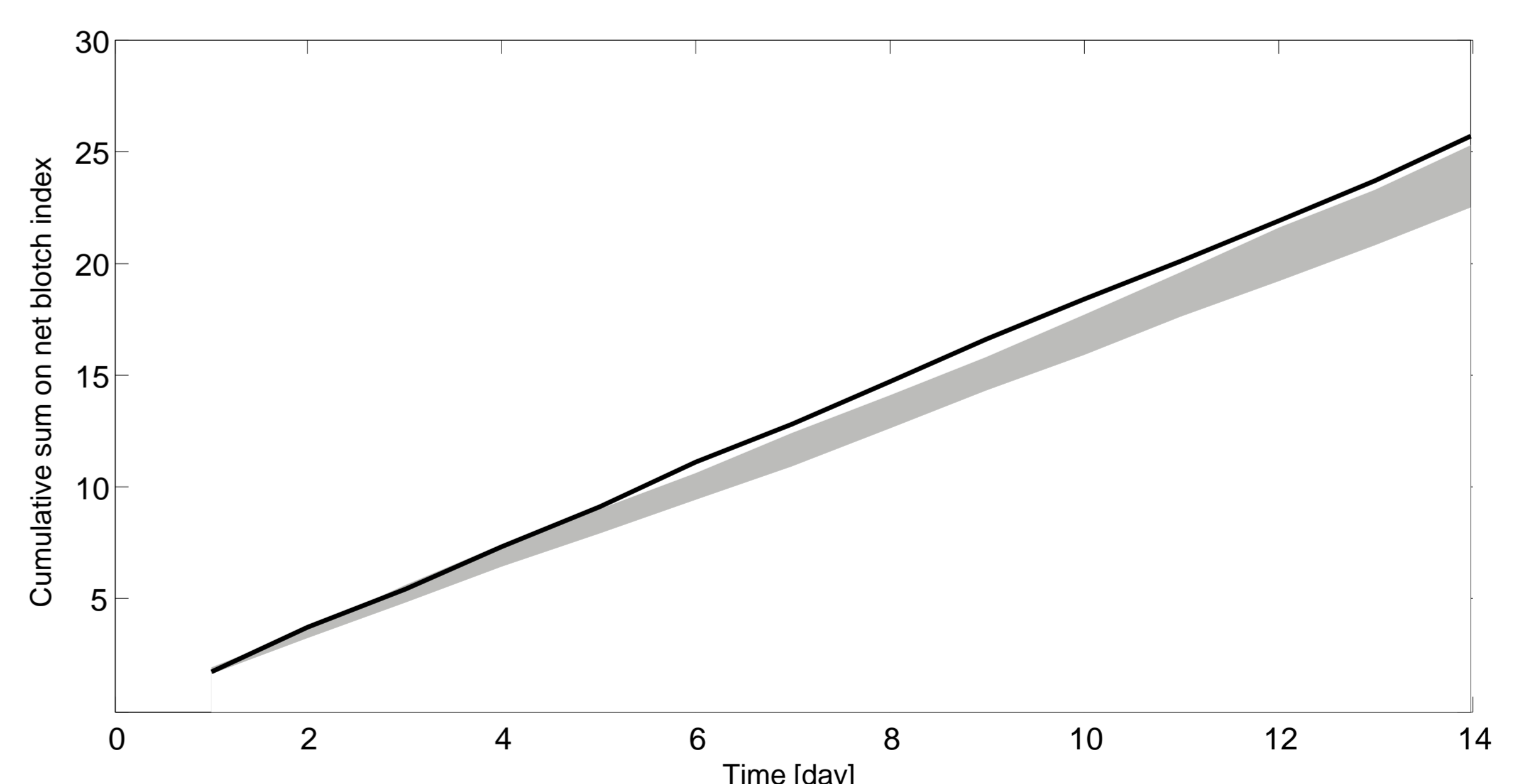


Figure 4. The outbreak of selected plant disease is predicted with generated feature. Validation with the weather data from 2015.

References:

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