Master Thesis

submitted within the UNIGIS MSc programme
Interfaculty Department of Geoinformatics - Z_GIS
University of Salzburg

Meteorological data modelling for the early prediction of Alternaria solani in potato crops to support an operational decision support system

by

Carina Jank, BSc
104890

A thesis submitted in partial fulfilment of the requirements of
the degree of
Master of Science (Geographical Information Science & Systems) – MSc (GIS)

Advisor:

Dr. Michael Riffler

Innsbruck, 25/06/2020

Science pledge

By my signature below, I certify that my thesis is entirely the result of my own work. I have cited all sources I have used in my thesis and I have always indicated their origin.

Innsbruck, 25/06/2020

Place, Date

Signature

Acknowledgment

I would like to thank the management team of GeoVille GmbH who gave me the opportunity to write my master thesis in the frame of the EO PLUG-IN project. This not only enabled me to work on an extremely interesting topic, I was also able to work technically almost without any restrictions thanks to the company's resources.

Specially I want to thank Samuel Barrett who always supported me throughout the long-lasting process of this thesis and not only was an extremely helpful technical supervisor but also gave me meaningful thought impulses for my entire scientific approach. Furthermore, I want to thank Michael Riffler who helped me with the challenges on the meteorological data handling and calculations and stated an excellent scientific supervisor who always encouraged me to not get lost on the never-ending researches of my thesis.

Finally, I wish to show my gratitude to my family, friends and all my colleagues at GeoVille GmbH, but specially to Christina Hirzinger and Eva Haas, who were my mental support throughout the whole process. Thank you for always convince me to not lose my motivation especially throughout the last year and always taking the load of me as good as you could to make it possible for me to focus on my thesis.

Preface

This thesis is the final work of my Master study in Geoinformatics at the UNIGIS Salzburg. It is a documentation of my research work, which was done between April 2018 and March 2020. It presents the results of a study for the early detection of Alternaria solani in potato crops to potentially support an operational decision support system. This thesis is written as a manuscript-based thesis as it addresses a gap in the current state-of-the-art, and it is hence, of high interest to publish the results of this scientific work. The manuscript is written for a possible publication in the ISPRS International Journal of Geo-Information. This open-access journal has a strong focus on applied geoinformatics and numerous publications with an agricultural background and, therefore, states a suitable framework for this research. This thesis contains two main sections. The manuscript (section 1) considers all specifications of the ISPRS journal and is ready for the submission to be published. The report in section 2 outlines all further technical and scientific developments and analyses done within this research work, that could not be included in more detail in the manuscript.

Table of contents

List of Figures

List of Tables

List of acronyms

AOI	Area of interest
API	Application programming interface
CAP	Common Agricultural Policy
DSS	Decision support system
EEA	European Environmental Agency
EO	Earth observation
ESA	European Space Agency
GIS	Geo Information System
HLB	Hilbrands Laboratorium B.V.
LIAB	Lab in a box system
LPIS	Land parcel identification systems
NWP	Numerical weather prediction
P-Days	Physiological days
RF	Random forest classifier
RH	Relative humidity
SDI	Spectral disease index

1. Manuscript

Research Article

Meteorological data modelling for the early prediction of Alternaria solani in potato crops to support an operational decision support system

Carina JANK[1], Samuel BARRETT[2] and Michael RIFFLER[3] *

[1] GeoVille GmbH; jank@geoville.com

[2] GeoVille GmbH; barrett@geoville.com

[3] GeoVille GmbH; riffler@geoville.com

* Correspondence: GeoVille GmbH, Sparkassenplatz 2, 6020 Innsbruck; jank@geoville.com; Tel.: +43-(0)512-46202114 (C.J.)

Abstract: This Article shows the research done by the authors on the predictive factor of meteorological parameters to early detect and predict early blight caused by Alternaria solani in potato crops. These analyses should on one hand support an existing handheld disease diagnoses system released by HLB and on the other hand determine the feasibility to develop an operational decision support system for the early prediction of Alternaria solani in potato crops, as currently no such operational system exist. Therefore, a regional numerical weather prediction model for the Netherlands was queried to gather meteorological parameters. By using daily temperature values it was possible to implement a specific crop growth model, called P-Days model, as the occurrence of Alternaria solani in potato crops is highly linked to the growth status of the plant. As the occurrence of Alternaria solani in potato crops is dependent on the development stage of the crop, the temperature based P-Days model was used to determine disease susceptibility. Additionally, the relative humidity was calculated on daily basis for the growing season 2018 in the aera of interest (AOI). These calculations were done not only to obtain thermal based analyses as Alternaria risk factor but also the disease pathogen is highly related to the relative humidity for a certain period before the disease occurs. It was possible to adjust the 300 P-Days model using the regional meteorological parameters and to successfully validate these results with the provided laboratory approved samples of Alternaria occurrence in potato plants without any controlled conditions. This led to the conclusion that no in-field measurements are necessary to apply the P-Days model, which is an important finding for the application in an operational disease risk model.

Keywords: Alternaria solani, potato, decision support system, physiological day, forecasting

---

1.1. Introduction

Worldwide the potato (*Solanum tuberosum*) is the third most consumed main nutritious crop after wheat and rice. Northwestern and Central European countries like Germany, France, the United Kingdom, Belgium and the Netherlands are especially efficient potato producers with yields higher than 40 tons per ha [1]. Consequently, in the Netherlands, crop production and especially the whole sector of potato production is of high economic value. Yield losses of potato crops in this region can add up to 30-40% and cause high economic impact not only to single farmers but as well to the whole potato value chain [2]. About 85% of plant disease damages are caused by fungal diseases worldwide [3]. Therefore, it is of common interest throughout the potato industry to lower the occurrence of crop diseases. Several studies [4 – 8] have already proven the possibility of the early prediction of late blight (*Phytophthora infestans*). These studies show the possibility of early prediction of late blight by calculating occurrence risk values in potato crops based on several meteorological parameters like temperature, relative humidity and precipitation. Late blight risk and susceptibility maps can be provided as decision support system (DSS) to improve spraying plans and to lower the occurrence of late blight significantly [9]. Beside these current developments and successes, the early blight disease caused by Alternaria solani states another reason for reduced potato yield, as it is widespread in nearly all potato production regions.

Early blight is characteristically visible through brown spots on the foliar of the potato crop [10]. The susceptibility of a potato crop getting infected by Alternaria solani spores increases remarkably with the physiological aging of the plant. Consequently, the occurrence of Alternaria solani is age-conditioned [10, 11]. Different models were analyzed and compared in this study by their predictive factors and accuracies to forecast Alternaria solani. Escuredo, O. & Seijo, M. C. [13] as well as Gent, D. H. & Schwarz, H. F. [14] confirmed that the physiological days model (P-days) is a good indicator for Alternaria infections of potato crops in specific circumstances. P-days model was initially developed by Sands, P.J. et al. [15] to model the development of potato tubers and foliar. Furthermore, Pscheidt, J.W. & Stevenson, W. R. [16] suggested to use this model to predict the susceptibility of potato crops for infection by Alternaria solani.

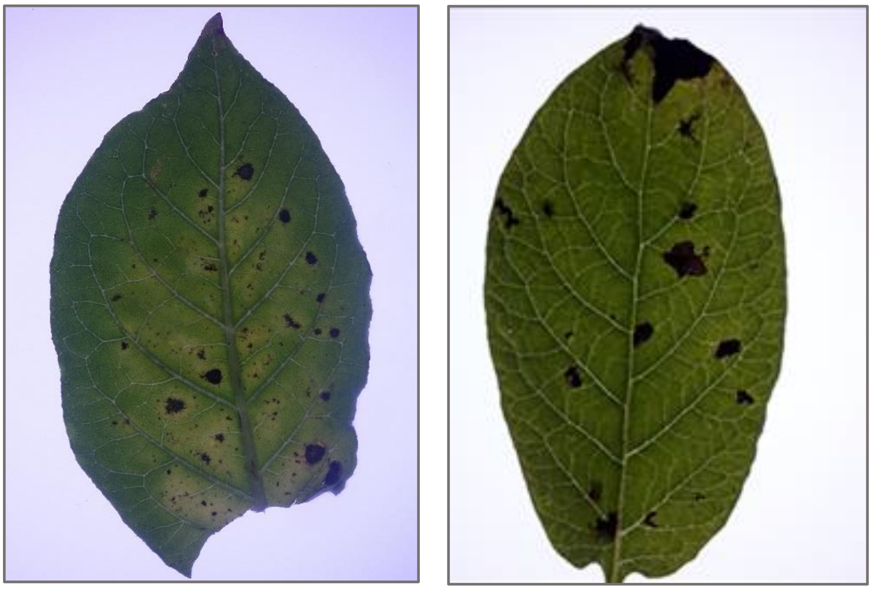
Other authors showed in their studies the possibility to predict the occurrence of the disease with models based on meteorological parameters like relative humidity, temperature and precipitation [11, 16]. Abuley, I.K. & Nielsen, B.J. [18] evaluated different disease prediction models based on meteorological data and assumed, as in several other studies [18 – 20], that it is theoretically possible in a scientific framework to predict early blight occurrence in potato crops, but none of these models is yet operational.

To provide a regional operational system, it is necessary to accurately predict the risk of early blight occurrence without in-situ field measurements or observations. Therefore, meteorological data from regional weather stations should gain an adequate accuracy to obtain the susceptibility of a crop for getting infected by Alternaria solani. In this study, the main scope is (I) to determine the predictive power of meteorological data for the occurrence of early blight caused by Alternaria solani as well as (II) the feasibility to develop an operational susceptibility and risk model for early blight prediction using a regional weather model.

The present study focuses on the calculation of the P-Days model and it's valuable predictive factor for the occurrence of Alternaria solani (section 1.2.1). Additionally, further developments using a regional weather model like the analyses of the calculated relative humidity (section 1.2.2) and the development of a random forest classifier (section 1.2.3) were done to more precisely evaluate their predictive potential. The promising results of these methods show the feasibility to establish an operational decision support system for the early prediction of Alternaria solani in potato crops as demonstrated in section 1.3.

1.2. Materials and Methods

The study was done in cooperation with a Dutch partner organization, Hilbrands Laboratorium B.V. (HLB), for the growing seasons 2018 and 2019 and limited to the area of the Netherlands due to the availability of data. For this study, over 200 laboratory-approved samples of Alternaria and non-Alternaria diagnosed potato crop foliar was made available as input data. The samples were taken by local farmers in the Netherlands at their potato fields using the HLB LeafSpot diagnosis system. This system was initially developed as a diagnostic tool for disease monitoring in potato crops. It is a computer vision model to provide a farmer with a specific diagnosis for the affected crop. Therefore, the user takes a photo of the foliar and uploads the image by using the LeafSpot smartphone app. A computer vision algorithm should automatically analyze if the crop is affected by Alternaria solani, ozone damage or nutrient deficiency for instance. This system is currently challenging this differentiation of the several possible causes for the characteristic brown spots on the potato foliar. Figure 1 demonstrates that the visual characteristics of a potato foliar affected by ozone damage or Alternaria solani are nearly identical. Hence, it is of high interest to include a risk and susceptibility model for Alternaria disease to improve the differentiation of these causes.



(a) (b)

Figure 1. *LeafSpot images of (a) an ozone-damaged foliar and (b) Alternaria solani. Source: LeafSpot database (HLB)*

The samples, including GPS coordinates, date and the laboratory-diagnosed analyses whether the crop was infected by Alternaria or not, are a unique dataset for investigating Alternaria prediction developments. As currently no operational Alternaria disease risk model for potato crops exists, the availability of laboratory-approved diagnosis of Alternaria occurrence in natural habits under natural (i.e. non-controlled) conditions is of enormous value for this study. These samples are used to validate the P-Days model and were used as test and training data for the additionally developed random forest classifier.

The meteorological information is requested from a numerical weather prediction (NWP) model provided by a Dutch national weather data provider. The model comes with a spatial resolution of 3 km. In this study, historical weather data of 2018 were gathered while the NWP allows to predict the required parameters up to 10 days into the future (depending on the required accuracy of the model for the usage in an operational DSS). The minimum, maximum and mean daily temperature, vapor pressure as well as the daily amount of precipitation were taken into consideration for all following calculations.

1.2.1. P-Days

The P-Days for the growing season 2018 were calculated using Equation 1 regarding the defined temperature threshold (Equations 2 to 4) as developed by Sands, P.J. et al. [15]:

$$P - Days = \frac{1}{24} * \left[(5 * PTmin) + 8\left(\frac{2*PTmin}{3} + \frac{PTmax}{3}\right) + 8\left(\frac{PTmin}{3} + \frac{2*PTmax}{3}\right) + (3 * PTmin)\right] \quad (1)$$

Where:

$$P(T) = 0 \qquad if\ T < 7°C\ or\ T \geq 30°C \quad (2)$$

$$P(T) = 10 * \left[1 - \frac{(T-21)^2}{(21-7)^2}\right] \qquad if\ 7°C \leq T < 21°C \quad (3)$$

$$P(T) = 10 * \left[1 - \frac{(T-21)^2}{(30-21)^2}\right] \qquad if\ 21°C \leq T < 30°C \quad (4)$$

The calculated physiological days (P-Days) are thermal-based calculations of the growing stadium of the plant at a specific field parcel and, therefore, are taken as promising indicators for the susceptibility and risk to get infected by Alternaria solani. The model considers daily minimum temperatures (PTmin) of 7° Celsius and 30° Celsius as daily maximum temperatures (PTmax), as outside these temperature thresholds, Alternaria solani is not spreading amongst these crops [14]. Initially, the P-days model was developed for the usage of in-field temperature measurements, but Gent, D. H. et al. [14] compared in their study the results of the P-Days model by using in-field temperature measurements and regional weather data. The study suggested that the P-Days model could potentially also provide accurate results in using regional weather data. Hence, in-field measurements should theoretically not be necessary for the prediction of risk and susceptibility of Alternaria infection. Several studies [12, 15, 20] took 300 P-Days as the threshold for the occurrence of Alternaria disease. This threshold needs to be defined for each region where the P-days model should be applied. In this study, meteorological data from the regional weather model were used for the calculation of the P-Days units for the growing season 2018. The calculations of the P-Days were done for every sample starting at the date of emergence. The date of emergence was defined manually for these calculations due to the low amount of samples. For an operational decision support system the definition of the date of emergence could be done automatically using optical satellite data but further details on this approach are outside the scope of this study.

The laboratory-approved samples of the diagnosed Alternaria diseases were used to evaluate whether the 300 P-Days is accurate as a threshold for the Netherlands or not.

1.2.2. *Relative humidity*

Different studies have already shown a positive correlation between the occurrence of Alternaria and relative humidity [21, 22]. Early blight pathogen is highly favored by relative humidity over 90% [23, 24]. Hence, the relative humidity for the duration of the growing season for each sample provided was calculated. Therefore, the saturated vapor pressure (e_s) was calculated using Teten's formula (Equation 5) [26].

$$e_s = e_0 * exp\left[\frac{b*(T-T_1)}{T-T_2}\right] \tag{5}$$

Where:

$e_0 = 0.611$ kPa

$b = 17.2694$

$T_1 = 273.16\ °K$

$T_2 = 35.86\ °K$

The mean daily temperature (T) needed for this calculation was requested from the regional weather provider. The relative humidity (RH) of each day was calculated with the usage of the mean daily vapor pressure (e), provided by the weather API, divided by the saturated vapor pressure (e_s), see Equation 6 below [22].

$$RH\ [\%] = \frac{e}{e_s} \tag{6}$$

1.2.3. Machine Learning - Random Forest Classifier

In addition to the P-Days and relative humidity calculations, a random forest classifier was developed. Besides relative humidity, several studies stated the possibility to early predict Alternaria solani using temperature and wind speed as predictive parameters. Therefore, especially the meteorological conditions within the last 30 days before the occurrence of Alternaria solani are of high relevance [11, 21]. A random forest classifier was applied to analyze the correlation of temperature as well as the wind speed and the occurrence of early blight. The random forest classifier, initially developed by Breiman, L. [27], is a machine learning algorithm based on decision trees. The structure of these decision trees is based on nodes and edges, which represent the features and decisions. Each decision is a chance to reduce entropy and allows the clustering of the data [28]. By splitting the samples into training and test data gives the classifier the possibility to learn from its decisions and getting more accurate the more training and test data are made available. This is the general concept of each machine learning algorithm [29]. The laboratory-approved samples were taken as training and test data for the random forest classifier. The provided samples for the growing season 2018 and 2019 not only contained information if the crop foliar was diagnosed by Alternaria solani disease in the laboratory, but also provided coordinates and dates, when the samples were gathered. Therefore, they represent the occurrence of Alternaria disease without any controlled conditions. This means that the disease appears in the plant without controlled external influences.

1.2.4. Earth observation data

Despite the previously described methods about early predicting Alternaria solani in potato crops using meteorological parameters and models, crop diseases can also be early detected by using remotely sensed optical data from satellite sensors [8]. The occurrence of a specific disease usually causes stress in the crop and is visible throughout a lowered chlorophyll production of the plant. This lower chlorotic lesions consequently causes a reduction of the Normalized Difference Vegetation Index (NDVI) [30]. Various recent studies [30 – 36] have already focused on the detection of crop diseases using optical satellite data. All current studies have stated that optical satellite data has a high potential for the early detection of crop diseases. However, the development of different techniques and methodologies would be required for the detection of a particular disease in a specific crop. For instance, Zhang, M. et al. [38] have already demonstrated the ability to diagnose late blight on tomato plants using specific spectral signatures. This has also been stated by several other studies on diverse crops [4 – 8]. The approach of using spectral signatures and band calculations of optical satellite data, like the NDVI, may be reliable for detecting late blight in potato crops, but compared to late blight for the detection of early blight the NDVI does not seem to be a suitable indicator. For instance, Golhani, K. et al. [3] stated that the NDVI is not a reliable indicator for the detection of crop diseases, as no specific wavebands represent the chlorophyll reduction specifically caused by disease-induced crop stress. Hence, it is suggested by several other studies [7, 38 – 42] to develop a spectral disease index (SDI), as specific indicator for a defined disease in a particular crop. The approach of developing a specific disease index, using hyperspectral imaging, is described by Mahlein, A.K. et al. [39]. In this work, the SDI for several sugar beet diseases under controlled conditions was identified by using a handheld hyperspectral scanner. Mirik, M. et al. [7] as well have used a hyperspectral field spectrometer and a digital camera to detect stress in several plants just to outline two of the above mentioned successful works on spectral-based disease detection. Despite all these studies using the SDI for detecting several diseases in various crops, no specific SDI for the detection of Alternaria solani in potato crops has been developed yet. Though, the development of an early blight specific SDI for potato plants would state the fundamental basis for further analyses of the early detection of early blight caused by Alternaria solani using optical satellite data [37, 43].

1.3. Results

1.3.1. P-Days and relative humidity

The P-Days units were calculated for every sample for each day, starting from the date of emergence until the end of the growing season. The results of these daily-based calculations show that Alternaria solani only occurred after 300 P-Days in the area of interest. For this analysis, the provided samples of the laboratory-approved Alternaria occurrence were taken into consideration. Figure 2 visualizes these results and shows the cumulative number of Alternaria diagnoses by the P-Days value on the date of diagnosis. It is clearly visible that Alternaria corresponds to the crop growth status.

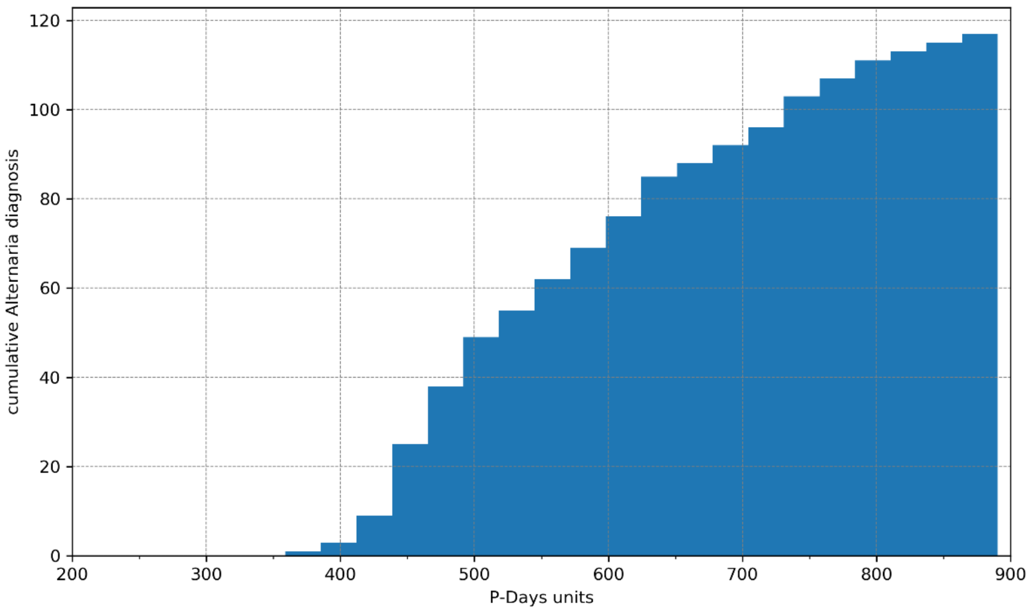


***Figure 2**. Cumulative Alternaria solani diagnosis in potato crops compared to the P-Days units*

These results show that the P-Days model, with a threshold of 300 P-Days, can also be applied for the Netherlands. This observation confirms the results of the previously mentioned studies, which already suggested a threshold of 300 P-Days [12, 15, 20]. However, therein it is suggested to reanalyze and redefine the threshold, depending on the area of interest. Another relevant result from this calculation proves that regional weather models also seem to be suitable for the P-Days calculations and no in-field weather stations and measurements are required.

The calculations of the mean daily relative humidity values were compared to the occurrence of the Alternaria. As a result, this analysis showed that the relative humidity was quite high before the first occurrence of Alternaria-diagnosed samples. Therefore, this parameter appears to be significant for the definition of the risk factor of the occurrence of early blight in potato crops. Due to the small amount of available laboratory-approved samples, it was not possible to accurately define a suitable threshold for the relative humidity as an additional decisive factor for the susceptibility of potato crops, getting affected by Alternaria solani. Nevertheless, these analyses show the possibility to define a practicable Alternaria risk model using regional weather data and that no in-field observations and measurements are needed. This is a crucial finding to develop an operational decision support system.

1.3.2. Random Forest Classifier

The random forest classifier based on the laboratory approved samples shows a predictive power, although it only relies on the input values of about 200 samples. The amount of the mean daily precipitation, as well as the mean daily wind speed and mean daily temperature, were used as training and test data for the random forest classifier. All these analyzed parameters show a prediction accuracy of about 60%. This shows that more respective samples are needed to further test and train the random forest classifier. However, this supporting approach shows promising results for the risk and susceptibility modeling.

The figure below (Figure 3) shows the correlation of the mean daily temperature and mean daily wind speed for the Alternaria diagnosed samples. The plot shows that in our area of interest, the disease only occurred between about 17 and 21 °C. Furthermore, a correlation to the windspeed seems to exist;

i.e., an increase of windspeed lowers the temperatures at which Alternaria solai occurs. In addition, it can be observed that slightly more Alternaria-diagnosed samples occur with increasing wind speed. This analysis corresponds to several previous studies on predicting Alternaria disease based on meteorological parameters, which came amongst others to the conclusion that wind speed is as well a significant predictive factor [11, 21].

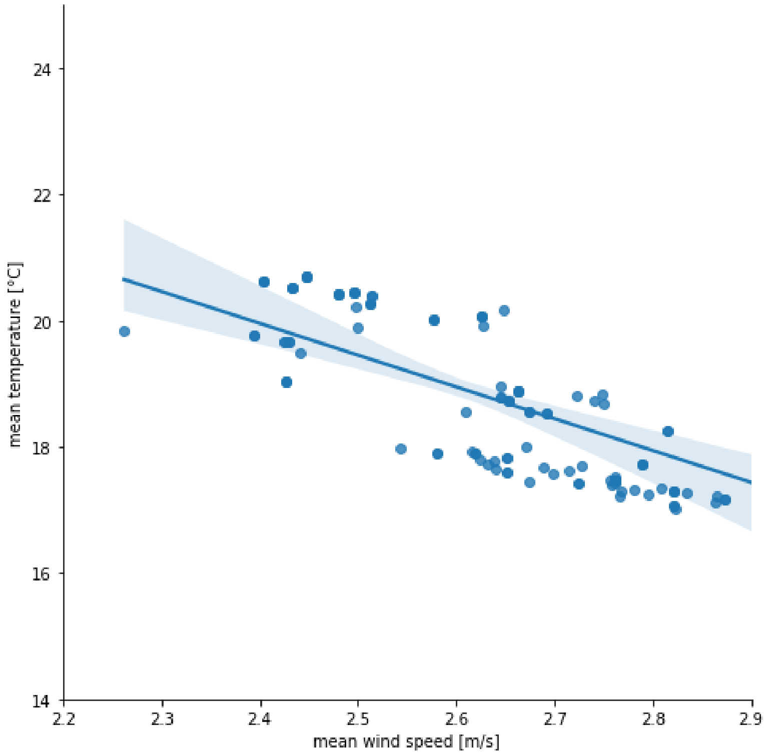


Figure 3. *Correlation of mean daily temperatures and mean wind speed*

Figure 3 also shows that the temperature thresholds, as applied to the P-Days modeling, correspond to these analyses. Consequently, a machine learning algorithm like the random forest classifier seems promising as a predictive modeling approach for this application.

To more precisely demonstrate the meteorological thresholds, as obtained by the available samples diagnosed with Alternaria solani, the table below (Table 1) shows the main parameters of interest and their analyzed values for the time range of 30 days before the observation of the Alternaria disease.

Table 1. Analyzes of meteorological parameters 30-days before Alternaria solani occurrence

Meteorological parameters	Min. value	Max. value	Mean value
Air temperature [°C]	4.23	37.43	18.53
Relative humidity [%]	52.62	92.45	72.11
Precipitation [mm]	1.49	11.35	5.63
Wind velocity [m/s]	2.26	3.49	2.70

1.4. Discussion

Several studies in the past have already established that the 300 P-Days model is valuable to early predict the occurrence of Alternaria disease and, furthermore, can be used to improve spraying plans and lower yield losses on potato crops [12, 15, 20]. Although the P-Days model developed by Sands, P.J. [15] was initially used to predict potato yield. The calculations start at the date of emergence and use daily minimum and maximum temperatures to simulate the growth status of the crop. Gent, D. H. et al. [46] and Pscheidt, J.W. et al. [16] reported the possibility of using the P-Days model for the prediction of the susceptibility of the occurrence of Alternaria solani on potato crops. The susceptibility of potatoes is directly related to the crop maturity, in general the disease does not occur until the crop is nearly fully developed [10]. Therefore, using P-Days as an indicator for the early prediction of the disease and to support a decision support system is a promising approach. Even though it is possible to accurately predict the susceptibility of Alternaria disease on potato crops, the presence of the spores and other climatic favorable conditions are not considered by the P-Days model. Therefore, to run an operational system to more accurately predict not only the theoretical susceptibility but also to include other favorable meteorological factors like precipitation or relative humidity, the implementation of a machine learning algorithm is suggested. The development of Alternaria solani is favored by wet conditions and directly related to the duration of leaf wetness within a certain period [10]. These additional factors could be involved into a prediction model by querying regional weather data and including them to a random forest classifier, as done in this study to show the possible positive impact of this approach for a decision support system. The machine learning algorithm based on the provided samples for Alternaria diagnosis in the season 2018 and 2019 showed the predictive factor of this model using regional weather data. For this study, only a small number of laboratory-approved samples was made available, therefore, the random forest classifier could not be developed to obtain precise results. However, even with the small number of samples, the results demonstrate the predictive value of weather data for predicting Alternaria disease occurrence.

Currently, no operational decision support system for predicting the susceptibility of Alternaria disease exists. At the moment, all relevant studies [13 - 14, 16, 18 – 20, 45], rely on controlled conditions or in-field measurements. Therefore, the existing statements on implementing the P-Days model for this usage are scientifically based and are not ready for an operational implementation. All of the previous studies included in-field temperature measurements to calculate the P-Days units for the designated crop field. Furthermore, the date of emergence which states the starting date of the calculations for the P-Days developments were defined manually by in-field observations [16, 45]. In this study weather data were gathered by a regional weather data provider and therefore, showed the possibility of developing an operational system including these data in the prediction model. The daily relative humidity seems to be an additional predictive parameter to be included to the development of an operational system. Due to the small amount of laboratory approved samples, it was not possible to define a suitable threshold or model for the risk assessment of Alternaria in potato crops.

1.5. Conclusions

This study showed the possibility to develop a risk and susceptibility model using meteorological and optical satellite data. Previous studies have already successfully developed risk assessments for the occurrence of Alternaria disease on potato crops, but all of them were based on in-field measurements. Hence, these models were not suitable for an operational system. Other studies rely on the P-Days calculations in addition to the needed in-field observations. The P-Days calculation is a suitable indicator for the susceptibility of a crop getting infected by Alternaria solani but does not take other relevant conditions as relative humidity or wind speed into consideration. The results of this study show that it is possible to accurately calculate the P-Days units using a regional weather model. It is also possible to define the date of emergence using optical satellite data as starting date for the P-Days calculations. To include information about the relative humidity, which is of high relevance for the occurrence of Alternaria, the regional weather model is also suitable and provides useful additional information. Even though a higher amount of laboratory approved Alternaria diagnosed samples including positioning and dates would be needed to accurately define the parameters and thresholds for such a risk and susceptibility model for the occurrence of early blight caused by Alternaria solani on potato crops, it is possible to develop an operational decision support system using the mentioned methods.

Additionally, an operational decision support system for the early detection of Alternaria solani could also be enhanced using optical satellite data. To include these earth observation data to an operational system, the development of a specific SDI needs to be done as described in section 1.2.4. These developments are outside the scope of this study but represent an interesting basis on which future studies could be dedicated to.

Author Contributions: Conceptualization, Carina Jank and Michael Riffler; Methodology, Carina Jank and Samuel Barrett; Validation, Carina Jank and Samuel Barrett; Writing-Original Draft Preparation, Carina Jank; Writing-Review & Editing, Carina Jank; Visualization, Carina Jank and Samuel Barrett; Supervision, Michael Riffler and Samuel Barrett. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Space Agency's (ESA) InCubed program, grant number 4000116994_16_I_NB.

1.6. References

[1] H. Campos and O. Ortiz, Eds., *The Potato Crop: Its Agricultural, Nutritional and Social Contribution to Humankind*. Cham: Springer International Publishing, 2020.

[2] T. Rakotonindraina *et al.*, 'Modeling of Yield Losses Caused by Potato Late Blight on Eight Cultivars with Different Levels of Resistance to *Phytophthora infestans*', *Plant Dis.*, vol. 96, no. 7, pp. 935–942, Jul. 2012, doi: 10.1094/PDIS-09-11-0752.

[3] K. Golhani, S. K. Balasundram, G. Vadamalai, and B. Pradhan, 'A review of neural networks in plant disease detection using hyperspectral data', *Inf. Process. Agric.*, vol. 5, no. 3, pp. 354–371, Sep. 2018, doi: 10.1016/j.inpa.2018.05.002.

[4] L. J. Dowley and European Association for Potato Research, Eds., *Phytophthora infestans 150*. Dublin: Published for the European Association for Potato Research by Boole in conjunction with Teagasc, 1995.

[5] W. E. Fry, 'Evaluation of Potato Late Blight Forecasts Modified to Incorporate Host Resistance and Fungicide Weathering', *Phytopathology*, vol. 73, no. 7, p. 1054, 1983, doi: 10.1094/Phyto-73-1054.

[6] Vaibhav K. Singh, Shailbala, and V.S. Pundhir, 'Forecasting Models: An Effective Tool for Potato Late Blight Management', 2012, doi: 10.13140/RG.2.1.1856.9369.

[7] M. Mirik *et al.*, 'Using digital image analysis and spectral reflectance data to quantify damage by greenbug (Hemitera: Aphididae) in winter wheat', *Comput. Electron. Agric.*, vol. 51, no. 1–2, pp. 86–98, Apr. 2006, doi: 10.1016/j.compag.2005.11.004.

[8] I. H. Newton, A. F. M. Tariqul Islam, A. K. M. Saiful Islam, G. M. Tarekul Islam, A. Tahsin, and S. Razzaque, 'Yield Prediction Model for Potato Using Landsat Time Series Images Driven Vegetation Indices', *Remote Sens. Earth Syst. Sci.*, vol. 1, no. 1–2, pp. 29–38, Jun. 2018, doi: 10.1007/s41976-018-0006-0.

[9] D. Shtienberg, 'Modelling: the basis for rational disease management', *Crop Prot.*, vol. 19, no. 8–10, pp. 747–752, Sep. 2000, doi: 10.1016/S0261-2194(00)00099-5.

[10] J. E. Waals, L. Korsten, and T. Aveling, 'A review of early blight of potato', *Afr Plant Prot*, vol. 7, pp. 91–102, Jan. 2001.

[11] W. E. Fry and D. Shtienberg, 'Integration of host resistance and fungicide to manage potato diseases', *Can. J. Plant Pathol.*, vol. 12, no. 1, pp. 111–116, Mar. 1990, doi: 10.1080/07060669009501049.

[12] M. Burch and E. Levetin, 'Effects of meteorological conditions on spore plumes', *Int. J. Biometeorol.*, vol. 46, no. 3, pp. 107–117, Aug. 2002, doi: 10.1007/s00484-002-0127-1.

[13] O. Escuredo, M. C. Seijo, M. Fernández-González, and I. Iglesias, 'Effects of meteorological factors on the levels of Alternaria spores on a potato crop', *Int. J. Biometeorol.*, vol. 55, no. 2, pp. 243–252, Mar. 2011, doi: 10.1007/s00484-010-0330-4.

[14] D. H. Gent and H. F. Schwartz, 'Validation of Potato Early Blight Disease Forecast Models for Colorado Using Various Sources of Meteorological Data', *Plant Dis.*, vol. 87, no. 1, pp. 78–84, Jan. 2003, doi: 10.1094/PDIS.2003.87.1.78.

[15] P. J. Sands, C. Hackett, and H. A. Nix, 'A model of the development and bulking of potatoes (Solanum Tuberosum L.) I. Derivation from well-managed field crops', *Field Crops Res.*, vol. 2, pp. 309–331, Jan. 1979, doi: 10.1016/0378-4290(79)90031-5.

[16] J. W. Pscheidt, 'Comparison of Forecasting Methods for Control of Potato Early Blight in Wisconsin', *Plant Dis.*, vol. 70, no. 10, p. 915, 1986, doi: 10.1094/PD-70-915.

[17] W. E. Fry and D. Shtienberg, 'Integration of host resistance and fungicide to manage potato diseases', *Can. J. Plant Pathol.*, vol. 12, no. 1, pp. 111–116, Mar. 1990, doi: 10.1080/07060669009501049.

[18] I. K. Abuley and B. J. Nielsen, ‘Evaluation of models to control potato early blight (Alternaria solani) in Denmark’, *Crop Prot.*, vol. 102, pp. 118–128, Dec. 2017, doi: 10.1016/j.cropro.2017.08.012.

[19] A. Angelosante Bruno *et al.*, ‘Estimation of fungal spore concentrations associated to meteorological variables’, *Aerobiologia*, vol. 23, no. 3, pp. 221–228, Aug. 2007, doi: 10.1007/s10453-007-9066-y.

[20] R. D. Davidson, A. J. Houser, and R. Haslar, ‘Control of Early Blight in the San Luis Valley, Colorado’, *Am. J. Potato Res.*, vol. 93, no. 1, pp. 43–49, Feb. 2016, doi: 10.1007/s12230-015-9482-4.

[21] R. O. C. Arana, L. Zambolim, and L. C. Costa, ‘Potato early blight epidemics and comparison of methods to determine its initial symptoms in a potato field.’, *Rev. Fac. Nac. Agron. Medellin Bogota*, vol. 60, no. 2, pp. 3877–3890, 2007.

[22] B. Tomassetti *et al.*, ‘Mapping of Alternaria and Pleospora concentrations in Central Italy using meteorological forecast and neural network estimator’, *Aerobiologia*, vol. 29, no. 1, pp. 55–70, Mar. 2013, doi: 10.1007/s10453-012-9262-2.

[23] A. Grinn-Gofroń and A. Mika, ‘Selected airborne allergenic fungal spores and meteorological factors in Szczecin, Poland, 2004–2006’, *Aerobiologia*, vol. 24, no. 2, pp. 89–97, Jun. 2008, doi: 10.1007/s10453-008-9088-0.

[24] J. R. Pelletier, ‘Characterization of Resistance to Early Blight in Three Potato Cultivars: Incubation Period, Lesion Expansion Rate, and Spore Production’, *Phytopathology*, vol. 79, no. 5, p. 511, 1989, doi: 10.1094/Phyto-79-511.

[25] D. (aro Shtienberg, D. Blachinsky, G. Ben-Hador, and A. Dinoor, ‘Effects of growing season and fungicide type on the development of Alternaria solani and on potato yield’, *Plant Dis. USA*, 1996, Accessed: Feb. 15, 2020. [Online]. Available: http://agris.fao.org/agris-search/search.do?recordID=US9632724.

[26] R. B. Stull and C. D. Ahrens, *Meteorology for scientists and engineers*, 2nd ed. Pacific Grove, CA: Brooks/Cole, 2000.

[27] L. Breiman, ‘Random forests’, *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[28] T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, and L. Plümer, 'Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance', *Comput. Electron. Agric.*, vol. 74, no. 1, pp. 91–99, Oct. 2010, doi: 10.1016/j.compag.2010.06.009.

[29] D. L.Gupta, A. K. Malviya, and S. Singh, 'Performance Analysis of Classification Tree Learning Algorithms', *Int. J. Comput. Appl.*, vol. 55, no. 6, pp. 39–44, Oct. 2012, doi: 10.5120/8762-2680.

[30] F. Martinelli *et al.*, 'Advanced methods of plant disease detection. A review', *Agron. Sustain. Dev.*, vol. 35, no. 1, pp. 1–25, Jan. 2015, doi: 10.1007/s13593-014-0246-1.

[31] C. H. Bock, P. E. Parker, A. Z. Cook, and T. R. Gottwald, 'Visual Rating and the Use of Image Analysis for Assessing Different Symptoms of Citrus Canker on Grapefruit Leaves', *Plant Dis.*, vol. 92, no. 4, pp. 530–541, Apr. 2008, doi: 10.1094/PDIS-92-4-0530.

[32] S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, 'A review of advanced techniques for detecting plant diseases', *Comput. Electron. Agric.*, vol. 72, no. 1, pp. 1–13, Jun. 2010, doi: 10.1016/j.compag.2010.02.007.

[33] A.-K. Mahlein, U. Steiner, C. Hillnhütter, H.-W. Dehne, and E.-C. Oerke, 'Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases', *Plant Methods*, vol. 8, no. 1, p. 3, 2012, doi: 10.1186/1746-4811-8-3.

[34] M. Prabhakar, Y. G. Prasad, and M. N. Rao, 'Remote Sensing of Biotic Stress in Crop Plants and Its Applications for Pest Management', in *Crop Stress and its Management: Perspectives and Strategies*, B. Venkateswarlu, A. K. Shanker, C. Shanker, and M. Maheswari, Eds. Dordrecht: Springer Netherlands, 2012, pp. 517–545.

[35] J. Clevers, L. Kooistra, and M. van den Brande, 'Using Sentinel-2 Data for Retrieving LAI and Leaf and Canopy Chlorophyll Content of a Potato Crop', *Remote Sens.*, vol. 9, no. 5, p. 405, Apr. 2017, doi: 10.3390/rs9050405.

[36] F. Baret, V. Houles, and M. Guerif, 'Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management', *J. Exp. Bot.*, vol. 58, no. 4, pp. 869–880, Nov. 2006, doi: 10.1093/jxb/erl231.

[37] S. N. Dlamini, A. Beloconi, S. Mabaso, P. Vounatsou, B. Impouma, and I. S. Fall, ‘Review of remotely sensed data products for disease mapping and epidemiology’, *Remote Sens. Appl. Soc. Environ.*, vol. 14, pp. 108–118, Apr. 2019, doi: 10.1016/j.rsase.2019.02.005.

[38] M. Zhang, Z. Qin, X. Liu, and S. L. Ustin, ‘Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing’, *Int. J. Appl. Earth Obs. Geoinformation*, vol. 4, no. 4, pp. 295–310, Nov. 2003, doi: 10.1016/S0303-2434(03)00008-4.

[39] A.-K. Mahlein *et al.*, ‘Development of spectral indices for detecting and identifying plant diseases’, *Remote Sens. Environ.*, vol. 128, pp. 21–30, Jan. 2013, doi: 10.1016/j.rse.2012.09.019.

[40] R. H. J. Heim, I. J. Wright, A. P. Allen, I. Geedicke, and J. Oldeland, ‘Developing a spectral disease index for myrtle rust (*Austropuccinia psidii*)’, *Plant Pathol.*, vol. 68, no. 4, pp. 738–745, May 2019, doi: 10.1111/ppa.12996.

[41] M. A. Izzuddin, A. Seman Idris, M. N. Nisfariza, A. A. Nordiana, H. Z. M. Shafri, and B. Ezzati, ‘The development of spectral indices for early detection of *Ganoderma* disease in oil palm seedlings’, *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 6505–6527, Dec. 2017, doi: 10.1080/01431161.2017.1335908.

[42] D. Ashourloo, A. A. Matkan, A. Huete, H. Aghighi, and M. R. Mobasheri, ‘Developing an Index for Detection and Identification of Disease Stages’, *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 851–855, Jun. 2016, doi: 10.1109/LGRS.2016.2550529.

[43] M. T. Kuska and A.-K. Mahlein, ‘Aiming at decision making in plant disease protection and phenotyping by the use of optical sensors’, *Eur. J. Plant Pathol.*, vol. 152, no. 4, pp. 987–992, Dec. 2018, doi: 10.1007/s10658-018-1464-1.

[44] Z. Qin and M. Zhang, ‘Detection of rice sheath blight for in-season disease management using multispectral remote sensing’, *Int. J. Appl. Earth Obs. Geoinformation*, vol. 7, no. 2, pp. 115–128, Aug. 2005, doi: 10.1016/j.jag.2005.03.004.

[45] B. Tomassetti *et al.*, ‘Mapping of Alternaria and Pleospora concentrations in Central Italy using meteorological forecast and neural network estimator’, *Aerobiologia*, vol. 29, no. 1, pp. 55–70, Mar. 2013, doi: 10.1007/s10453-012-9262-2.

[46] D. H. Gent and H. F. Schwartz, 'Validation of Potato Early Blight Disease Forecast Models for Colorado Using Various Sources of Meteorological Data', *Plant Dis.*, vol. 87, no. 1, pp. 78–84, Jan. 2003, doi: 10.1094/PDIS.2003.87.1.78.

[47] G. Počivavšek and M. Ljuša, 'CHARACTERISTICS OF THE LAND PARCEL IDENTIFICATION SYSTEM (LPIS) AS THE MAIN SUBCOMPONENT OF THE AGRICULTURE INFORMATION SYSTEM', Sep. 2013.

[48] D. F. Wanjura, D. R. Upchurch, S. J. Maas, and J. C. Winslow, 'Spectral Detection of Emergence in Corn and Cotton', p. 15.

[49] A. R. Huete, D. F. Post, and R. D. Jackson, 'Soil spectral effects on 4-space vegetation discrimination', *Remote Sens. Environ.*, vol. 15, no. 2, pp. 155–165, Mar. 1984, doi: 10.1016/0034-4257(84)90043-9.

[50] T. Zhang, J. Su, C. Liu, W.-H. Chen, H. Liu, and G. Liu, 'Band selection in sentinel-2 satellite for agriculture applications', in *2017 23rd International Conference on Automation and Computing (ICAC)*, Huddersfield, United Kingdom, Sep. 2017, pp. 1–6, doi: 10.23919/IConAC.2017.8081990.

[51] Ying Gao, J. P. Walker, M. Allahmoradi, A. Monerris, Dongryeol Ryu, and T. J. Jackson, 'Optical Sensing of Vegetation Water Content: A Synthesis Study', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 4, pp. 1456–1464, Apr. 2015, doi: 10.1109/JSTARS.2015.2398034.

[52] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, 'Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content', *Sensors*, vol. 11, no. 7, pp. 7063–7081, Jul. 2011, doi: 10.3390/s110707063.

2. Report

2.1. Analysis of Laboratory approved Samples

The provided laboratory-approved samples are the output of the LeafSpot App, released by HLB. Farmers, using this app, are able to upload an image of the crop foliar. Depending on the characteristic spots on the crop foliar, the farmer gets a diagnosis if the crops are affected by a certain disease, i.e. Alternaria solani or if the spot is caused by nutrient deficiencies. To validate these image-based identifications, some samples were additionally laboratory diagnosed. For the calculations in this study, only the laboratory approved samples were taken into consideration as they state a reliable in situ dataset. These samples were initially analyzed by their locations and diagnoses. The overall 217 provided samples are separated to 125 Alternaria diagnosed samples and 92 non-Alternaria diagnosed ones. Furthermore, the samples were analyzed regarding the Land Parcel Identification System (LPIS) dataset for the Netherlands. The LPIS dataset provides the exact geographical position and extent of a field parcel including the agricultural use of each unique parcel. LPIS is one of the core Geo Information Systems (GIS) used within the EU Common Agricultural Policy (CAP) [1]. The LPIS field parcels are provided by several agricultural categories amongst others also potato field parcels are provided by this dataset. The analysis showed that only 66 out of the total 125 Alternaria diagnosed samples are inside one of these LPIS fields regardless the type of usage and only 35 samples are within potato LPIS fields. In total 151 samples are outside this LPIS field parcels. All these analyses are demonstrated in Figure 4 below.

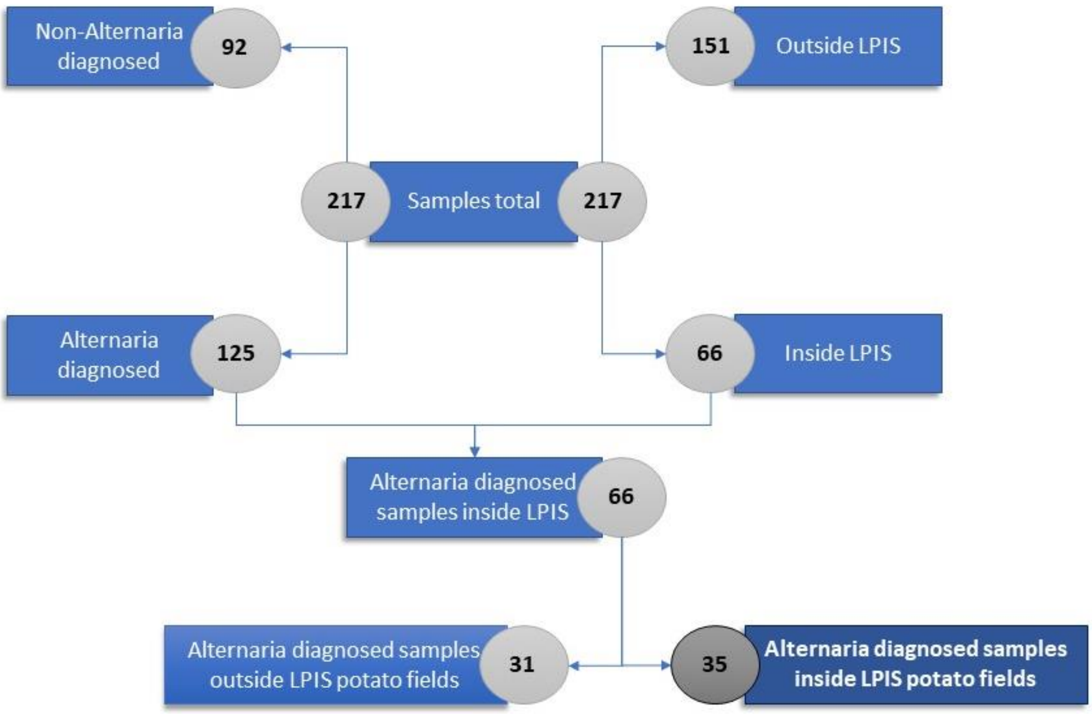


Figure 4. *Visualization of the sample analysis*

This indicates that several issues regarding the location of the provided samples exist. Detailed analyses show that many samples are located within urban areas. Consequently, they cannot be linked to the suitable field parcel. One possible reason for this issue could be that farmers take their samples within their potato field but upload the data and image at another location, for example, when they are back home, as the sample location of Figure 5 visualizes. Hence, no detailed analysis on the meteorological conditions for the area where the samples were taken can be done. This led to discussions whether in future the uploaded samples have to be linked to a specific field parcel of each farmer or not.

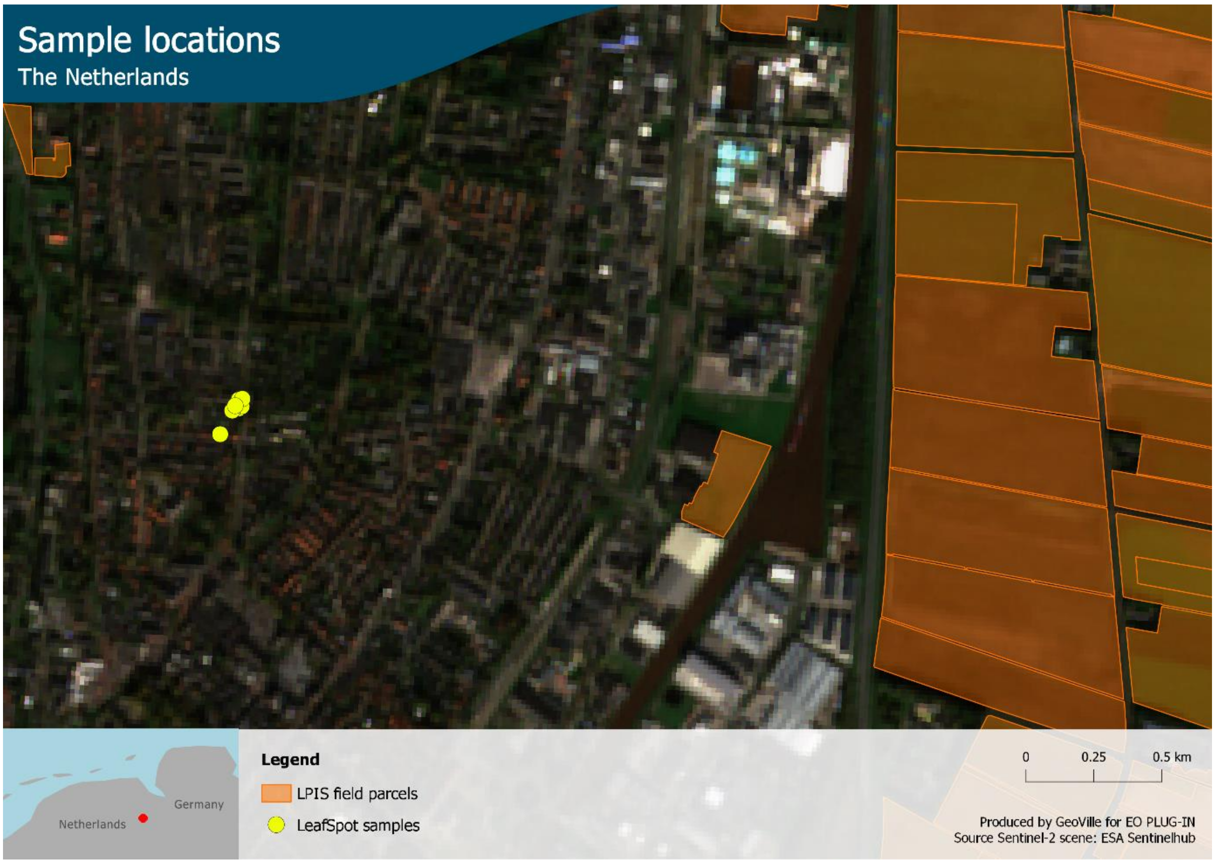


***Figure* 5**. *Subset of samples located in urban area*

2.2. Weather API

The numerical weather prediction (NWP) model queried for the model calculation was provided by HERMESS BV. The model data were gathered by an application programming interface (API) using the geographical positions of each sample. The meteorological weather model provides different parameters on a daily basis, for instance, minimum and maximum temperature, mean temperature, mean wind speed, sum of precipitation and mean vapor. The model is based on 3 km spatial resolution and provides forecast for up to 10 days into the future. The HERMESS weather data provider is calculated for the area of the Netherlands. For all regions outside the Netherlands, the NASA weather data provider is also implemented in this infrastructure to guarantee a wider spread scalability of the model. The API for querying the NWP was developed in another project and requested for this study by using the GPS coordinates of each sample and a defined start and end date. The query of the meteorological API for each sample was done as described in the code subset 1 below. The regarded geodata frame (GDF) includes all provided samples containing their coordinates and dates of observations as well as the diagnoses, as described in section 2.1. The function `point_request` was developed in the scope of another project and made available for these developments. The detailed technical reporting of this function is outside the frame of this technical report.

```
> for j, i in tqdm(GDF.iterrows(), total = len(GDF)):
>     point = point_request(i["geometry"], start_date, end_date)
>     tmp = [{ky: vl if type(vl) is not str else datetime.datetime.strptime
            (vl,"%Y%m%d").date() for ky,vl in j.items()} for j in point]
>     meteo_data = [k for k in tmp if k["DAY"] <= end_date and k["DAY"] >=
             start_date]

>     TMIN = np.array([j["TMIN"] for j in meteo_data])
>     TMAX = np.array([j["TMAX"] for j in meteo_data])
>     TEMP = np.array([j["TEMP"] for j in meteo_data])
>     VAP = np.array([j["VAP"] for j in meteo_data])
```

(Code subset 1)

As start date of the calculations, the date of emergence was used. The definition of the date of emergence is explained in the next chapter (section 2.3). For sample points, where the date of emergence was not possible to define, the 23rd of May was assumed as start date, as this date indicates the date of emergence in the majority of the tested fields and is, therefore, a reliable starting date. The end date was defined by the gathering date of each sample, as the particular conditions before a disease occurs or not is relevant and not the day afterwards. For the following, calculations of P-Days units and relative humidity, the meteorological parameters of minimum, maximum and mean temperature as well as vapor pressure were queried from this API.

2.3. Date of emergence

The date of emergence defines, in the agricultural context, the date a plant gets visible on bare soil due to the emergence of the crop foliar. Using remote sensing and RGB images this means a field turns from its brown color to a more greenish spectral range. The normalized difference vegetation index (NDVI) relies on the basis that a healthy crop reflects more intensively in the the near infrared (NIR) (~700 – 1300 nm) and green (~520 - 560 nm) wavelength of the visible spectrum and absorbs more red (VIR) (~630 nm – 690 nm) and blue (~450 – 500 nm) light, which causes that healthy crops appear green in the humans eyes. If the NIR band of an optical sensor is taken as red channel, the NIR spectrum gets visible als red color [2, 3]. The significant increase of the calculated NDVI using Sentinel-2 data is a reliable indicator for the crop growth and, therefore, is also for the date of emergence. To calculate the NDVI, the VIR and NIR bands of the Sentinel-2 optical satellite is used. A valid band composition to calculate the NDVI is using Band 8a for NIR and Band 4 as VIR [4].

To define the date of emergence, not only the markable increasing of NDVI is a useful parameter, also the leaf area index (LAI) is a reliable indicator. The LAI indicates the leaf coverage and therefore, strongly correlates with the NDVI values [5]. The LAI is also calculated using the VIR and NIR wavelength of Sentinel-2 [6]. Further details of using NDVI and LAI for detecting crop stress, especially in potato crops for detecting early blight caused by Alternaria solani, are discussed in section 1.2.4 of the manuscript.

The figure below (Figure 6. NDVI and LAI calculation for a LPIS field parcel) shows the calculation of the NDVI and LAI timeseries for a specific LPIS field parcel. For this timeseries, several cloud-free

Sentinel-2 observations were available. The LAI development shows a significant increase beginning at 26th of May 2018. Additionally, the NDVI values are also heightened at this time period. The true color Sentinel-2 images and the invers red color bands show the foliar growing beginning at 26th of May. All these indicators lead to the conclusion that for this specific LPIS field (ID 222542) parcel the date of emergence can be defined for the 26th of May 2018.

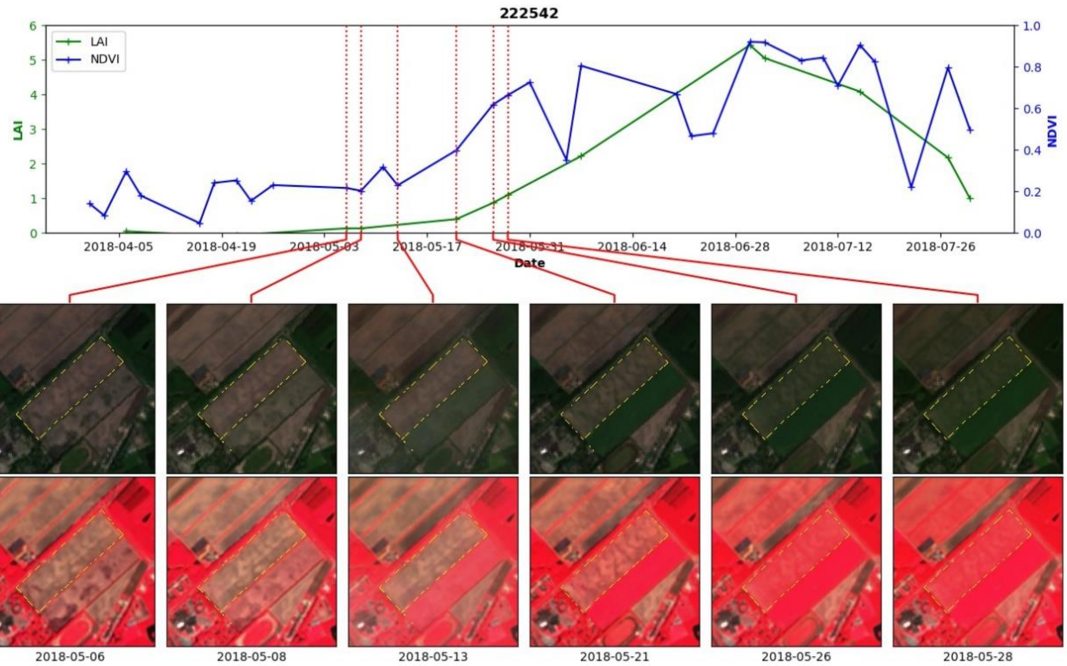


***Figure 6**. NDVI and LAI calculation for a LPIS field parcel ID 222542, Source: GeoVille GmbH*

Another example of a LPIS field parcel (ID 689729) shows similar characteristics for the date of emergence as demonstrated in Figure 7. At the 21st of May 2018, the LAI significantly increases as well as the NDVI. Consequently, the invers red band combination of the Sentinel-2 scene on 21st of May 2018 shows an increase of red intensity in this field parcel, which is caused by a more intense chlorophyll production in this field and, therefore, indicates the date of emergence.

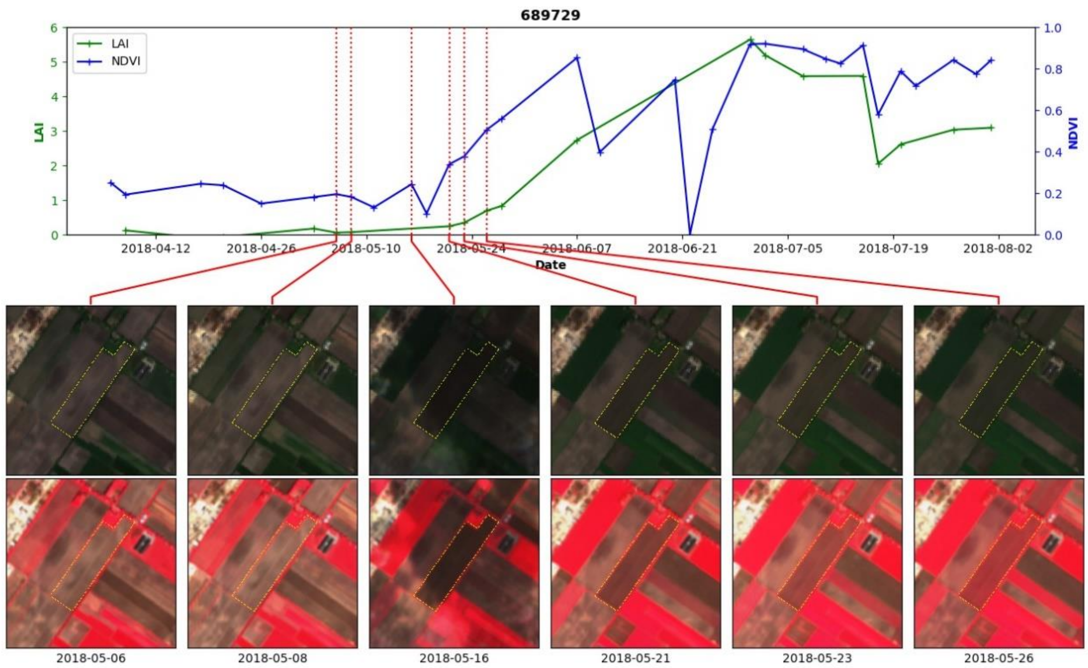


***Figure 7**. NDVI and LAI calculation for a LPIS field parcel ID 68729, Source: GeoVille GmbH*

2.4. Calculations of P-Days units

The P-Days calculations corresponding to the mentioned formula (1) in chapter 1.2.1 of the manuscript and the temperature conditions regarding formulas (2 - 4) are done using the programming language Python 3.6. For this approach, a function for temperature values was developed. This funtion includes the temperature thresholds and the specific temperature calculations for each threshold range as code subset 2 of the risk modelling code shows:

```
> import numpy as np

> def P(T):
>         if T <= 7:
>             return 0
>         elif T > 30:
>             return 0
>         elif T <= 21:
>             return 10*(1-((T - 21)**2)/((21 - 7)**2))
>         else:
>             return 10*(1-((T - 21)**2)/((30 - 21)**2))
> def p(t):
>         return np.array(list(map(P,t)))

> PDays = (1/24)*((5*p(TMIN)) + 8*p(((2*TMIN)/3) + (TMAX/3)) + 8*p(((2*TMAX)/3)
           + (TMIN/3)) + 3*p(TMAX))
```

(Code subset 2)

The daily minimum and maximum temperature values as needed for these calculations are queried using the meteorological API, as already described in section 2.2. This P-Days calculations are done for each Alternaria diagnosed sample starting at date of emergence of the specific field parcel until the observation date of the disease. Using this approach, it is possible to analyze the limit of P-Days units where Alternaria solani was first observed in the Netherlands and to define the 300 P-Days threshold for this region.

2.5. Calculations of relative humidity

The relative humidity (RH) is calculated based on the formula described in chapter 1.2.2 of the manuscript. Code subset 3 below shows the calculations as implemented in Python. The variables of the mean daily temperature (TEMP) and mean daily vapor pressure (VAP) are queried via the meteorological API as technically explained in chapter 2.2. The temperature values have to be converted to the temperature unit Kelvin as they are given in degree Celsius by the meteorological API. Furthermore, the vapor pressure is recalculated from Hectopascal (hPa) to Kilopascal (kPa) to obtain the daily relative humidity in percent using this approach.

```
> import numpy as np

> es = (0.611 * (np.exp((17.2694*((TEMP + 273.16) - 273.16))/((TEMP + 273.16) -
       35.86))))
> e = (VAP*0.1)
> RH = ((e/es)*100)
```

(Code subset 3)

2.6. Random Forest Classifier

The Random Forest classifier as delineated in chapter 1.2.3 is also done with Python using the Scikit-Learn library. Code subset 4 demonstrates an example of one of the various random forest testings that has been done. In this example, the RF testing is done using relative humidity (RH) as feature and the diagnosis as decision parameter. The test size in this example for the machine learning algorithm is estimated at 0.30. Furthermore, the number of trees (`n_estimators`) in this RF test is set to 20.

```
> from sklearn.model_selection import train_test_split
> from sklearn.ensemble import RandomForestClassifier

> x = gdf["RH"].values
> y = gdf["diagnosis"].values

> for j in range(10):
>     x_train, x_test, y_train, y_test = train_test_split(x, y, random_state = j,
      test_size = 0.30)
>     model = RandomForestClassifier(n_estimators=20,
>                                    criterion = "gini",
>                                    max_depth = None,
>                                    max_features = "auto",
>                                    random_state = 0)
>     model.fit(x_train, y_train)
```

(Code subset 4)

A subset of the tests done with the RF classifier are documented in Table 2. Overview of the random forest classifier tests. This documentation shows the various tested parameters and the obtained accuracies for each test. The best accuracies are achieved using a test size of 55%. To conclude the various random forest testing, all described tests show an accuracy of 60 to 65%. There is a slight but not a significant variation in these obtained accuracies of the different testings. A possible reason could be the low amount of the input data. All the used data cover the same growing period of 2018 in a similar area. If samples from other years in with different weather conditions would made available and included to this random forest classifier the accuracies on the different test parameters could potentially differentiate more significant and more precise information on the predictive power of each parameter could be evaluated.

Table 2. Overview of the random forest classifier tests

Test ID	Test parameters	Timerange [Days]	Splitting size	n_estimators	Accuracy [%]
1	sumRain, meanTMAX	30	0,5	20	60
2	sumRain, meanTMAX	30	0,3	20	60
3	sumRain, meanTMAX	30	0,7	20	60
4	sumRain, meanTMAX	30	0,55	20	60
5	sumRain, meanTMAX	30	0,55	25	65
6	sumRain, meanTMAX	30	0,55	15	65
7	sumRain, meanTMAX	30	0,55	10	60
8	meanRain, meanTMAX	30	0,55	15	65
9	meanRain, meanTMAX, meanTMIN	30	0,55	15	65
10	sumRain, meanTMAX, meanTMIN	30	0,55	15	65
11	meanRain, mean_Temp	30	0,55	15	62
12	meanRain, mean_Temp	30	0,55	20	65
13	meanRain, mean_Temp	30	0,55	10	65
14	meanRain, mean_Temp, mean_VAP	30	0,55	20	62
15	meanRain, mean_Temp, mean_VAP	60	0,55	20	60

2.7. References

[1] G. Počivavšek and M. Ljuša, 'Characteristics of the Land Parcel Identification System (LPIS) as main subcomponent of the agriculture Information system', Sep. 2013.

[2] D. F. Wanjura, D. R. Upchurch, S. J. Maas, and J. C. Winslow, 'Spectral Detection of Emergence in Corn and Cotton', *Precicion Agriculture*, vol 4, pp. 385-399, Dec. 2003, doi: 10.1023/A:1026323408710.

[3] A. R. Huete, D. F. Post, und R. D. Jackson, „Soil spectral effects on 4-space vegetation discrimination", Remote Sens. Environ., Bd. 15, Nr. 2, S. 155–165, März 1984, doi: 10.1016/0034-4257(84)90043-9.

[4] T. Zhang, J. Su, C. Liu, W.-H. Chen, H. Liu, und G. Liu, „Band selection in sentinel-2 satellite for agriculture applications", in 2017 23rd International Conference on Automation and Computing (ICAC), Huddersfield, United Kingdom, Sep. 2017, S. 1–6, doi: 10.23919/IConAC.2017.8081990.

[5] Ying Gao, J. P. Walker, M. Allahmoradi, A. Monerris, Dongryeol Ryu, und T. J. Jackson, „Optical Sensing of Vegetation Water Content: A Synthesis Study", IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., Bd. 8, Nr. 4, S. 1456–1464, Apr. 2015, doi: 10.1109/JSTARS.2015.2398034.

[6] J. Delegido, J. Verrelst, L. Alonso, und J. Moreno, „Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content", Sensors, Bd. 11, Nr. 7, S. 7063–7081, Juli 2011, doi: 10.3390/s110707063.