Multifactorial Evolutionary Prediction of Phenology and Pests: Can Machine Learning Help?

Francisco José Lacueva-Pérez[1][a], Sergio Ilarri[2,3][b], Juan José Barriuso Vargas[4,5][c], Gorka Labata Lezaun[1] and Rafael Del Hoyo Alonso[1][d]

[1]*ITAINNOVA - Instituto Tecnológico de Aragón, PT. Walqa, Huesca, Spain*

[2]*Department of Computer Science and Systems Engineering, University of Zaragoza, Zaragoza, Spain*

[3]*I3A, University of Zaragoza, Zaragoza, Spain*

[4]*CITA - Universidad de Zaragoza, Zaragoza, Spain*

[5]*AgriFood Institute of Aragon (IA2), Zaragoza, Spain*

Keywords: Smart Farming, Phenology Forecast, Machine Learning, Big Data, Remote Sensing.

Abstract: Agriculture is a key primary sector of economy. Developing and applying techniques that support a sustainable development of the fields and maximize their productivity, while guaranteeing the maximum levels of health and quality of the crops, is necessary. Precision agriculture refers to the use of technology to help in the decision-making process and can lead to the achievement of these goals. In this position paper, we argue that machine learning (ML) techniques can provide significant benefits to precision agriculture, but that there exist obstacles that are preventing their widespread adoption and effective application. Particularly, we focus on the prediction of phenology changes and pests, due to their important to ensure the quality of the crops. We analyze the state of the art, present the existing challenges, and outline our specific research goals.

1 INTRODUCTION

Advancing towards the sustainable development goals requires a paradigm shift in the management of farming (Kamilaris et al., 2017). Agriculture is traditionally managed on the base of the farmer's experience (Ip et al., 2018). The application of Internet of Things (IoT), Big Data and Artificial Intelligence (AI) technologies to convert real-time field data into information and knowledge is transforming the practices of farmers by supporting data-driven decision making (Lokers et al., 2016). Smart-Farming (SF) and Precision Agriculture (PA) are terms coined to refer to the application of the IoT, Big Data and AI to farm management (Zhang et al., 2002). However, SF considers larger geographical areas (province, region, country, etc.) (Bacco et al., 2019) while PA considers smaller ones (even at specific plant level) (Pham and Stack, 2018).

In this position paper, we argue that Machine Learning (ML) plays a key role in PA, although there are challenges and obstacles that are preventing its widespread adoption from a practical point of view. We focus on the phenology and pest incidence aspects of PA.

The structure of the rest of this paper is as follows. Section 2 presents the context of this research. Section 3 presents a study of the state of the art. In Section 4, we present an analysis of the existing challenges and expected trends. Then, in Section 5, and based on our previous analysis, we explain our current research goals related to the topic of this paper. Finally, in Section 6, we indicate our conclusions and prospective lines for future work.

2 BACKGROUND

We present here the background of our work. In Section 2.1, we explain the growing importance of phenology for PA, which is the focus of our research. In Section 2.2, we indicate the role of technologies related to IoT, Big Data, and AI for modelling the phenology of plants and modelling pests.

[a] https://orcid.org/0000-0003-0998-2939
[b] https://orcid.org/0000-0002-7073-219X
[c] https://orcid.org/0000-0003-2980-5454
[d] https://orcid.org/0000-0003-2755-5500

Lacueva-PÃl'rez, F., Ilarri, S., Vargas, J., Lezaun, G. and Alonso, R.
Multifactorial Evolutionary Prediction of Phenology and Pests: Can Machine Learning Help?.
DOI: 10.5220/0010132900750082
In *Proceedings of the 16th International Conference on Web Information Systems and Technologies (WEBIST 2020)*, pages 75-82
ISBN: 978-989-758-478-7

2.1 The Increasing Interest in Phenology Understanding

Phenology studies the relations between the life cycles of plants and environmental changes (Tang et al., 2016). Reaumur (Reaumur, 1735) was the first who indicated a relationship between the temperature and phenology. While the phenology is affected by multiple factors like the climatic conditions, soil features, and water stress (Tang et al., 2016), the temperature and the solar radiation are the most influential ones. Thus, most phenology models only consider them. However, the complete impact of other factors on the evolution of phenology and pests still remains unknown (Wang et al., 2019).

Phenology is an ubiquitous phenomenon. It is a highly-sensitive indicator of the impact of Climate Change (CC) on plants (Tang et al., 2016). Factors that affect the phenology have also an impact on the incidence of pests (fungi, insects, etc.) (Zhao et al., 2013). Thus, the interest in studying, monitoring and creating accurate prediction models for phenology is increasing, because they are used for analyzing and foreseeing the impact of CC (Schwartz et al., 2003), but also because they contribute to the implementation of SF and PA solutions (Kamilaris et al., 2017).

2.2 Phenology Modelling Technologies

Phenology monitoring and forecasting projects deal with multi-sourced field observations, having different formats (e.g., structured data or hiperspectral images), produced at different speed and frequency (e.g., climatic observations every 30 minutes or satellite images every 5 days), and creating high volumes of data (e.g., in the case of hiperspectral images). According to De Mauro's definition of Big Data (De Mauro et al., 2016), these projects are Big Data projects.

Phenology-related time-series contain measures about seasonally-variable conditions or field observations (e.g., the temperature, the phenology, etc.) as well as seasonally-stable variables (e.g., field limits, kind of soil, etc.) (Barnard, 2018). The relevant data sources are classified in four main categories (Zeng et al., 2020): *Human observations (HO)*; *Near-surface observations (NSO)*; *Remote sensing observations (RSO)*; and *Gridded observations/forecast model data (GOF)* (Taylor and White, 2020).

Considering the area covered, HO provide manually-recorded data about the phenology, pests and some determinant factors in a specific area of study. This data supports creating local models (Rötzer and Chmielewski, 2001). The generality of the models increases using NSO, networks of static climatic sensors and red / green / blue (RGB) and hyperspectral cameras (Zeng et al., 2020). Whenever NSO measurements are provided with the required frequency, in particular images, they can be used for creating regional models (Zeng et al., 2020). RSO are RGB or hyperspectral images taken by drone/airborne/satellite-mounted hyperspectral cameras (Mohanty et al., 2016). Both RSO and GOF provide high spatial resolution data of wide areas at a regular frequency, thus enabling the development of regional and global models (Zeng et al., 2020).

Big Data technologies allow the automation of data capture from these sources and its transformation for creating phenology and pests models (L'heureux et al., 2017). The combination of data sources can increase the knowledge of the biological processes on which phenology depends (Tang et al., 2016). This can help to expand the number of variables that can be used by the models, both "classical" and AI-based, improving their quality (Kamilaris et al., 2017). Moreover, thanks to their precision (Rötzer and Chmielewski, 2001), HO are used for the calibration of NSO and RSO (Zhao et al., 2013).

ML techniques, in particular Deep Learning (DL) techniques, are used to extract patterns from multi-sourced big data sets (time series, images, etc.). These detected patterns are used to classify information, improve the quality of data sets, or predict future values (Najafabadi et al., 2015). Big Data and ML are applied to determine the phenology evolution of plants (Wolfert et al., 2017) as well as the life cycles of the diseases that can affect them (Mahlein, 2016).

Web technologies allow linking all these data and the needed services (Lokers et al., 2016). They support the access and exchange of data (data concerning climatic stations, satellite images, etc., are provided through REST services) (L'heureux et al., 2017), they provide computing power and storage on the cloud (Kamilaris et al., 2017), and they can make AI models and AI results available through open REST services (Wolfert et al., 2017).

3 STATE OF THE ART

This section provides an overview of the state of the art of Big Data and AI technologies used for monitoring and predicting phenology and pests. The first three subsections describe the use of the data provided by each of the types of data sources identified. In Section 3.4, ML algorithms used for monitoring and prediction are reviewed. Finally, Section 3.5 presents existing data-driven projects development models.

3.1 Works Exploiting Field Observation Data (HO)

Plant phenology is determined by 3 main types of factors (Zhao et al., 2013): individual characteristics (e.g., age); environmental factors (e.g., location); and management practices (e.g., pruning) (Reynolds et al., 2018).

Classical approaches statistically analyse the phenology and the influencing environmental factors, looking for correlations between them (Tang et al., 2016). However, due to the difficulties to integrate a high number of variables in human-driven calculations (Taylor and White, 2020), the models usually rely on (linear) functions of only one or two climatic variables (temperature and precipitation) (Rodríguez Galiano et al., 2016). Thus, the obtained models simplify real processes and they do not reflect the effect of various variables or past events (e.g., winter chilling) (Wang et al., 2019). In addition, some models set the start date of phenology development to a fixed calendar day. However, making this date relative to a phenology-related event (e.g., the start of dormancy) will improve the accuracy of the models (Rodríguez Galiano et al., 2016).

3.2 Works Exploiting Climatic Data (NSO and GOF)

Phenology models use NSO and GOF to try to determine how their measurements influence the phenology's observed state either using HO or RSO. In consequence, data of the predicted temperatures, among other climatic variables (e.g., precipitation data), are required for prediction (Taylor and White, 2020).

Observed data is obtained from climatic station networks in near real-time. Its measurements are accurate at their locations. However, their precision decreases due to several factors such as the distance or geographic features (like valley bottoms). This makes their use inappropriate even for modelling the phenology of fields not far away from the stations. Fortunately, current climatic models provide gridded data sets that can be used for more accurate observations (Pytharoulis et al., 2016). In order to be aligned with climatic NSO and HO, GOF data (concerning their locations, their timestamps, and units of measurement) usually needs to be transformed (Taylor and White, 2020). These transformations are applied to both observed and predicted measurements.

3.3 Works Exploiting Hyperspectral Images (RSO)

Hyperspectral images are used, at local and global scale, for monitoring the phenology (Tang et al., 2016), canopy anomalies, and the incidence of external factors (such as the impact of pests) (Barnard, 2018), as well as for determining the factors causing anomalies (Cârlan et al., 2020).

Hyperspectral cameras provide images using several spectral bands, including RGB and near-infrared bands (Barnard, 2018). Vegetation Indexes (VI) combine subsets of these bands. They use simple mathematical formulations to measure the influence of different factors over the plant's phenology behavior and the vulnerability to pests (Tang et al., 2016). While there are several VI, the most widely used are the *Normalised Difference VI (NDVI)* and *Enhanced VI (EVI)* (Tang et al., 2016).

Hyperspectral images can be provided by commercial cameras (Wang, 2017) or obtained from publicly-available repositories (Cârlan et al., 2020), such as Copernicus Scihub (ESA-European Space Agency, 2020). Nevertheless, their quality must be evaluated in order to determine their possible use. Wang provides a review of the aspects to be considered for assessing the quality of the images (Wang, 2017). For the purpose of our work, the most relevant factors are the following:

1. The spatial resolution assesses the relation between a pixel in the image and the area of the Earth's surface it captures. Fixed and drone cameras provide very high resolution (<1 meters per pixel) (Wang, 2017). Satellite images provide wider-areas pictures but having poorer resolution (>75 meters per pixel) (ESA-European Space Agency, 2020). This poses challenges for the accurate monitoring of the phenology of fruit trees, like vineyards, based on canopy observations. Due to canopy discontinuity (Barnard, 2018), VI can be contaminated by the vegetation surrounding the trees (e.g., weeds) and / or the soil composition (Cârlan et al., 2020).

2. The time resolution determines the frequency at which a picture can be made available. Whereas fixed cameras can capture images several times a day, drones and airborne cameras do not fly regularly; on the other hand, satellites can provide images at a regular frequency lower than a week (usually between 1 and 15 days) (Reed et al., 2009).

3. Satellite images are affected by fog and cloud occlusions as well as by the position of the sun

when captured (projected shadows) (Stöckli et al., 2008). Low-quality images can be detected and corrected using methods like the ones suggested by (Duarte et al., 2018) and (Borgogno-Mondino et al., 2018), respectively. Moreover, the quality of the images can be improved by combining observations from different constellations of satellites (Nguyen and Henebry, 2019).

3.4 AI-based Modelling Approaches

ML and DL algorithms are used for different purposes when applied during the creation of monitoring or forecasting phenology models. We provide here a short review of representative approaches.

(Sirsat et al., 2019) used random forests to select climatic variables to be used in yield prediction (Längkvist et al., 2014). DATimeS (Belda et al., 2020) provided a set of of 30 different algorithms for filling gaps in satellite image time series. This set includes, among other algorithms, bagging trees, adaptive regression splines, boosting trees and k-nearest neighbors regression, among others. Multiple linear regressions, artificial neural networks (ANN) and Random Forest (RF) methods have been compared for the prediction of the incidence of rice pests using phenology RSO (Skawsang et al., 2019).

A comprehensive review of existing literature on the application of DL to solve agriculture problems (including image processing) was presented in (Kamilaris and Prenafeta-Boldú, 2018). According to this work, Convolutional Neural Networks (CNN) improve the performance of other ML classification methods such as Support Vector Machines (SVM), ANN, and RF. While CNNs can be used for time series classification problems, Recurrent Neural Networks (RNN) perform better. RNN's architectures (Long Short Time Memory –LSTM–, SVMs, three-unit LSTMs, and CNNs+LSTMs) tend to exhibit a dynamic temporal behavior.

Cheng obtained similar results (Cheng, 2018). He compared the performance of two DL models (a 3-D fully CNN and a Siamese Network) and an ML model (a sigmodial regression network) to determine the transition dates of vegetable images stored in Phenocam (a database of static RGB cameras). The DL algorithms led to more accurate results. Besides, their accuracy can be further improved by combining CNN with RNN.

3.5 Data-driven Project Development Methodologies

CRISP-DM (Cross Industry Standard Process for Data Mining) (Wirth and Hipp, 2000) has been the *de facto* standard process for developing data mining and knowledge discovery projects for more than 20 years. However, it has to be redefined to be able to fit the exploratory nature of data-science projects. (Martínez-Plumed et al., 2019) proposed a new process model, the *Data Science Trajectories model (DST)*. DST is based on CRISP-DM; it considers its process components (business understanding, data understanding, data preparation, data modelling, evaluation, and deployment) but modifies the data management aspects. DST defines four different data management processes (Acquisition, Simulation, Architecting, and Release) and it also adds exploratory processes (Data Source, Goal, Product, Data Value, Result, and Narrative). Nevertheless, the main contribution of DST is the concept of *trajectories*. Instead of imposing an iterative sequence of all the processes, trajectories allow each project to define its own path on processes considering the restriction of being an acyclic directed graph over activities.

4 CHALLENGES

In the following, we discuss some perspectives of interest that should attract further research in the near future.

Multidisciplinary Approaches to Phenology Modelling.

Existing phenology models have proved their validity for specific locations and species / varieties, although they have to be readjusted to work well in different scenarios (Zhang et al., 2002). Besides, the complexity and lack of full awareness of the phenology processes, and the fact that current models just consider a subset of the available variables, complicate the improvement of the models' accuracy and generality (Wang et al., 2019).

Better and more data can lead to improved models. However, the complexity of handling data also increases (e.g., due to the dimensionality increase) (Holzinger, 2018). This increment makes the use of human-driven calculations difficult. Consequently, computational models have become the preferred modeling techniques for plant biology (Prusinkiewicz and Runions, 2012), making multidisciplinarity a requirement for creating multifactorial phenology models (Tang et al., 2016).

Multidisciplinary approaches leverage different points of view to solve a problem. However, this increases the complexity. There are gaps within the agronomic field (e.g., the different phenology scales complicate the comparison of observations and results) (Tang et al., 2016). However, good-quality information eases solving these gaps (Holzinger, 2018) (e.g., by using an ontology to map the existing scales) (Costa et al., 2019).

Agronomic knowledge gaps increase the difficulty of explaining non-fully-known phenology processes to technological experts. Moreover, the potential understanding of these problems based on available data is further complicated by data uncertainty, including its unpredictability, non-linearity and non-homogeneity in time (Holzinger, 2018). These issues define Big Data problems (De Mauro et al., 2016).

Finally, on the computing side, AI results have to be valuable and useful for users. They have to be presented in such a way that they are understandable (Taylor and White, 2020). Data presentation has to allow perceiving the presented facts in the context they happen, easing taking decisions (Holzinger, 2018). This also applies to AI models. Explainable AI (XAI) is a very state-of-the-art trend aiming to explain models in an understandable way, but also trustworthy, to the users. XAI aims to answer questions about how and why certain results are obtained. This will allow users to justify and explain their data-driven decisions (Hoffman et al., 2018) and, in some scenarios, it will allow them to demonstrate that they comply with applicable legislation (Holzinger, 2018).

Methodologies for Creating SF and PA Systems.
A common methodology for creating phenology and pest AI models would ease the execution of both research and development projects. However, to the best of our knowledge, a specific methodology for the agronomic sector still does not exist.

Therefore, it is relevant to work on the development of a methodology for filling the gaps between the agronomic and data science worlds. This methodology could be developed based on the DST process model (Martínez-Plumed et al., 2019), in cooperation with the agronomists. The goal is to build appropriate physical and AI models of phenology and pests. Both types of models will share data. Therefore, it is important to clearly and easily describe how the data is obtained, transformed, presented as well as how the results of the models are validated.

Regarding the quality of the models, a set of measurements, which can be obtained from models created by agronomists and data scientists, must be agreed upon, in order to make their results comparable. It is also desirable to set common criteria for differentiating training and validation data sets. This will allow to validate the models, in particular their generality (Kamilaris and Prenafeta-Boldú, 2018).

By making the whole processes visible, DST trajectories (Martínez-Plumed et al., 2019) can ease the automation of the data logistics and pipeline activities. This automation will also make it possible to efficiently create (or retrain) models which will fit better to CC (Taylor and White, 2020).

Development and Exploitation of Phenology and Pest Forecasting Models.
While statistical and ML approaches are used for developing these models (Tonnang et al., 2018), the real-time monitoring and short-term prediction of phenology, in particular using satellite images, remains challenging (Zeng et al., 2020). The challenges previously described obviously have influence on this one. Nevertheless, low-level issues related with Big Data and AI can be bigger influencers.

Multi-factorial phenology and pest forecasting models integrate incomplete long-term time series data sets related to phenology, pests and relevant external factors (Tang et al., 2016). Mixing these data sets is necessary for a better understanding of the mechanisms behind the specific phenology or pest as well as for creating general CC reactive models of phenology and pests (Tang et al., 2016). Besides, the quality and size of the data sets are going to determine the accuracy of the models created (Ball et al., 2017); therefore, capturing and obtaining good data becomes crucial. However, this is not easy when high-dimensional and multi-temporal and spatial-scaled data is used (Zeng et al., 2020).

Our study of the state of the art shows different ways to deal with problems concerning the data sets. First, some works show techniques to deal with discontinuous series of field observation data and satellite images for phenology states (Wang, 2017). Second, other studies try to provide solutions to the potential lack of some phenology states' observations, due to their short duration in comparison with the phenology observation frequency (Kamilaris and Prenafeta-Boldú, 2018). Third, others authors propose methods for improving the poor resolution of satellite images (Zeng et al., 2020). However, we were not able to find any study offering a complete solution to these problems.

5 RESEARCH GOALS

Our work is being developed in the context of a European research project GRAPEVINE (Munné and Del Hoyo, 2019). It aims to use high-performance computing to go further in the prevention of grapevine pests, providing useful information to fight pests at the right moment.

Our use case study will be developed in the context of vineyards, considering several wine grape varieties. Models will be created and validated in two pilots placed in distant Mediterranean regions. The first pilot will consider garnache, mazuela, tempranillo, and chardonnay varieties, whereas the second pilot will focus on Greek Xinomavro, Negoska, Roditis and Assyrtiko, Sauvignon blanc, Merlot, Chardonnay, and Cabernet sauvignon varieties. Due to their incidence in Europe, pests to be considered are three pathogenic fungi (*Plasmopara viticola*, *Botrytis cinerea* and *Uncinula necator*) and two butterflies (*Lobesia botrana* and *Sparganothis pilleriana*).

Our work is being developed in cooperation with agronomic collaborators. Whereas they are going to create what we call the *physical models* (i.e., the creation of classical phenology and pest models applying statistical analysis on HO and NSO), we are going to use this data and also RSO and GOF for creating ML models, that we call the *logical models*.

The physical and logical models will be created following a two-step process: *creating the phenology* models and *using them to create pest models*. This reduces the complexity of predicting the incidence of pests (Chuine et al., 2013), taking advantage of the direct relation between the pest incidence risk level and the phenology state when it is produced (Ribeiro et al., 2020). Moreover, some treatments are only effective if they are applied in advance to a given phenology stage within a given time window (i.e., 2 or 3 weeks). Finally, we can provide more value to farmers by allowing them to schedule tasks of brief duration that demand a high volume of (human) resources (e.g., for cropping) (Taylor and White, 2020).

From the scientific and technical perspectives, our goal is twofold: to generate a methodology for developing PA systems using ML; and to create ML methods for predicting the phenology and pest behavior in a horizon of 2 or 3 weeks using that methodology.

Our proposed methodology is going to be based on the DST process model (Martínez-Plumed et al., 2019). We will define DST trajectories to document the process that the data follows when captured, transformed, used for creating the models, used for validating the models, and used in real conditions.

For creating the models, we are going to use data from different sources. The first considered source is HO recorded by agronomists working in pilot areas. They provide field seasonally-stable variables such as the field's location and its boundaries. However, the data they weekly capture recording the phenology state and pest incidence is more important. Besides, we will try to use Copernicus hyperspectral images (their derived VI) for monitoring the phenology and pests that appear on leaves. Environmental variables will be captured from the climatic station networks deployed in the regions of interest. They will provide the temperature, precipitation, solar radiation and wind measurements at a frequency of up to 30 minutes. The station's data could be complemented with GOF data, which is going to be provided by another member of the consortium (Pytharoulis et al., 2016). Its models will also provide climatic prediction data.

Big Data technologies will be used to clean, combine and transform raw data in a data set that can be used for creating multi-factorial phenology and pest prediction models. This might require using ML algorithms to improve the quality of the different data sets, in particular in the case of a hyperspectral image data set (see Section 4). Our idea is to create a comprehensive data set containing all the available raw data as well as derived data like VI. This data set will be a continuous time series in order to be able to feed the models with previous season's data, and so allowing our models to learn about the influence of past events.

ML algorithms will be applied with different objectives. First, they could be used for determining the impact of the considered variables on the phenology and pests. Second, they will be used to create the models for the phenology and pests. The processes that we will follow, the problems and challenges that we will overcome, and the solutions that we will obtain, will be included in our methodology. This will help us to communicate during project development but also will ease the diffusion and reuse of our work.

Finally, although it is currently in an early state of development, we will follow the evolution of XAI to complete our methodology with techniques that allow to explain our models.

6 CONCLUSIONS AND FUTURE WORK

Providing multi-factorial intelligence phenology models is challenging, both for humans and for machines. Machine learning techniques, such as deep learning, offer many benefits over more traditional modelling approaches; data may be highly dimensional, heterogeneous, and contains spatio-temporal structure, but can still be able to generate useful location-specific models with minimal human input (Lee et al., 2020).

In this position paper, we have analyzed the challenges for the widespread adoption of machine learning techniques for precision farming, studied the state of the art, and presented our research goals in relation to this topic. Our future steps involve the collection and integration of heterogeneous data related to our case study, the analysis and evaluation of machine learning models using those data, and the proposal of a methodology that helps advance the application of machine learning for precision agriculture.

ACKNOWLEDGEMENTS

This research is part of the work in progress of *GRAPEVINE –hiGh peRformAnce comPuting sErvices for preVentIon and coNtrol of pEsts in fruit crops*. GRAPEVINE is co-financed by the European Union's Connecting Europe Facility (CEF) – Telecommunications Sector Agreement under Grant Agreement No. INEA/CEF/ICT/A2018/1837816. We also thank the support of the project TIN2016-78011-C4-3-R (AEI/FEDER, UE) and the Government of Aragon (Group Reference T64_20R, COSMOS research group).

REFERENCES

Bacco, M., Barsocchi, P., Ferro, E., Gotta, A., and Ruggeri, M. (2019). The digitisation of agriculture: a survey of research activities on smart farming. *Array*, 3:100009.

Ball, J. E., Anderson, D. T., and Chan, C. S. (2017). Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing*, 11(4):042609.

Barnard, Y. (2018). *The use of technology to improve current precision viticulture practices: predicting vineyard performance*. PhD thesis, Stellenbosch University.

Belda, S., Pipia, L., Morcillo-Pallarés, P., Rivera-Caicedo, J. P., Amin, E., De Grave, C., and Verrelst, J. (2020). DATimeS: A machine learning time series GUI toolbox for gap-filling and vegetation phenology trends detection. *Environmental Modelling & Software*, page 104666.

Borgogno-Mondino, E., Novello, V., Lessio, A., Tarricone, L., and de Palma, L. (2018). Intra-vineyard variability description through satellite-derived spectral indices as related to soil and vine water status. *Acta Horticulturae*, 1197:59–68.

Cârlan, I., Mihai, B.-A., Nistor, C., and Große-Stoltenberg, A. (2020). Identifying urban vegetation stress factors based on open access remote sensing imagery and field observations. *Ecological Informatics*, 55:101032.

Cheng, Z. (2018). Detecting phenological transition dates of vegetation based on multiple deep learning models. Master's thesis, TU Delft.

Chuine, I., de Cortazar-Atauri, I. G., Kramer, K., and Hänninen, H. (2013). Plant development models. In *Phenology: an Integrative Environmental Science*, pages 275–293. Springer.

Costa, J. M., Marques da Silva, J., Pinheiro, C., Barón, M., Mylona, P., Centritto, M., Haworth, M., Loreto, F., Uzilday, B., Turkan, I., et al. (2019). Opportunities and limitations of crop phenotyping in southern European countries. *Frontiers in Plant Science*, 10:1125.

De Mauro, A., Greco, M., and Grimaldi, M. (2016). A formal definition of big data based on its essential features. *Library Review*, 65(3):122–135.

Duarte, L., Teodoro, A. C., Monteiro, A. T., Cunha, M., and Gonçalves, H. (2018). QPhenoMetrics: An open source software application to assess vegetation phenology metrics. *Computers and Electronics in Agriculture*, 148:82–94.

ESA-European Space Agency (2020). Copernicus open access hub. https://scihub.copernicus.eu/. Accessed: 2020-07-23.

Hoffman, R. R., Mueller, S. T., Klein, G., and Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608*.

Holzinger, A. (2018). From machine learning to explainable AI. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, pages 55–66. IEEE.

Ip, R. H., Ang, L.-M., Seng, K. P., Broster, J., and Pratley, J. (2018). Big data and machine learning for crop protection. *Computers and Electronics in Agriculture*, 151:376–383.

Kamilaris, A., Kartakoullis, A., and Prenafeta-Bold'u, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143:23–37.

Kamilaris, A. and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90.

Längkvist, M., Karlsson, L., and Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11–24.

Lee, M. A., Monteiro, A., Barclay, A., Marcar, J., Miteva-Neagu, M., and Parker, J. (2020). A framework for predicting soft-fruit yields and phenology using embedded, networked microsensors, coupled weather models and machine-learning techniques. *Computers and Electronics in Agriculture*, 168:105103.

L'heureux, A., Grolinger, K., Elyamany, H. F., and Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access*, 5:7776–7797.

Lokers, R., Knapen, R., Janssen, S., van Randen, Y., and Jansen, J. (2016). Analysis of big data technologies for use in agro-environmental science. *Environmental Modelling & Software*, 84:494–504.

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, 100(2):241–251.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Orallo, J. H., Kull, M., Lachiche, N., Quintana, M. J. R., and Flach, P. A. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*.

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7:1419.

Munné, R. and Del Hoyo, R. (2019). Grapevine-hiGh performance comPuting sErvices for preVention and https://grapevine-project.eu/. Accessed: 2020-07-15.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., and Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1):1.

Nguyen, L. H. and Henebry, G. M. (2019). Characterizing land use/land cover using multi-sensor time series from the perspective of land surface phenology. *Remote Sensing*, 11(14):1677.

Pham, X. and Stack, M. (2018). How data analytics is transforming agriculture. *Business Horizons*, 61(1):125–133.

Prusinkiewicz, P. and Runions, A. (2012). Computational models of plant development and form. *New Phytologist*, 193(3):549–569.

Pytharoulis, I., Kotsopoulos, S., Tegoulias, I., Kartsios, S., Bampzelis, D., and Karacostas, T. (2016). Numerical modeling of an intense precipitation event and its associated lightning activity over northern greece. *Atmospheric Research*, 169:523–538.

Reaumur, R. A. F. (1735). Thermometric observations made at Paris during the year 1735, compared to those made below the equator on the isle of Mauritius, at Algiers and on a few of our American islands. *Mémoires de l'Académie royale des sciences (Paris)*.

Reed, B. C., Schwartz, M. D., and Xiao, X. (2009). Remote sensing phenology. In *Phenology of Ecosystem Processes*, pages 231–246. Springer.

Reynolds, M., Kropff, M., Crossa, J., Koo, J., Kruseman, G., Molero Milan, A., Rutkoski, J., Schulthess, U., Sonder, K., Tonnang, H., et al. (2018). Role of modelling in international crop research: overview and some case studies. *Agronomy*, 8(12):291.

Ribeiro, H., Piña-Rey, A., Abreu, I., Rodríguez-Rajo, F. J., et al. (2020). Integrating phenological, aerobiological and weather data to study the local and regional flowering dynamics of four grapevine cultivars. *Agronomy*, 10(2):185.

Rodríguez Galiano, V. F., Sánchez Castillo, M., Dash, J., Atkinson, P., and Ojeda Zújar, J. (2016). Modelling interannual variation in the spring and autumn land surface phenology of the European forest. *Biogeosciences*, 13:3305–3317.

Rötzer, T. and Chmielewski, F.-M. (2001). Phenological maps of Europe. *Climate Research*, 18(3):249–257.

Schwartz, M. D. et al. (2003). *Phenology: an integrative environmental science*. Springer.

Sirsat, M. S., Mendes-Moreira, J., Ferreira, C., and Cunha, M. (2019). Machine learning predictive model of grapevine yield based on agroclimatic patterns. *Engineering in Agriculture, Environment and Food*, 12(4):443–450.

Skawsang, S., Nagai, M., K Tripathi, N., and Soni, P. (2019). Predicting rice pest population occurrence with satellite-derived crop phenology, ground meteorological observation, and machine learning: A case study for the central plain of Thailand. *Applied Sciences*, 9(22):4846.

Stöckli, R., Rutishauser, T., Dragoni, D., O'keefe, J., Thornton, P., Jolly, M., Lu, L., and Denning, A. (2008). Remote sensing data assimilation for a prognostic phenology model. *Journal of Geophysical Research: Biogeosciences*, 113(G4).

Tang, J., Körner, C., Muraoka, H., Piao, S., Shen, M., Thackeray, S. J., and Yang, X. (2016). Emerging opportunities and challenges in phenology: a review. *Ecosphere*, 7(8).

Taylor, S. D. and White, E. P. (2020). Automated data-intensive forecasting of plant phenology throughout the United States. *Ecological Applications*, 30(1):e02025.

Tonnang, H. E., Makumbi, D., and Craufurd, P. (2018). Methodological approach for predicting and mapping the phenological adaptation of tropical maize (Zea mays L.) using multi-environment trials. *Plant Methods*, 14(1):108.

Wang, H. (2017). *Crop assessment and monitoring using optical sensors*. PhD thesis, Kansas State University.

Wang, Y., Case, B., Rossi, S., Dawadi, B., Liang, E., and Ellison, A. M. (2019). Frost controls spring phenology of juvenile smith fir along elevational gradients on the southeastern tibetan plateau. *International Journal of Biometeorology*, 63(7):963–972.

Wirth, R. and Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Fourth International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, pages 29–39. Springer.

Wolfert, S., Ge, L., Verdouw, C., and Bogaardt, M.-J. (2017). Big data in smart farming–a review. *Agricultural Systems*, 153:69–80.

Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., and Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, 237:111511.

Zhang, N., Wang, M., and Wang, N. (2002). Precision agriculture—a worldwide overview. *Computers and Electronics in Agriculture*, 36(2-3):113–132.

Zhao, M., Peng, C., Xiang, W., Deng, X., Tian, D., Zhou, X., Yu, G., He, H., and Zhao, Z. (2013). Plant phenological modeling and its application in global climate change research: overview and future challenges. *Environmental Reviews*, 21(1):1–14.