Weather-data-based model: an approach for forecasting leaf and stripe rust on winter wheat

Victor M. Rodríguez-Moreno1 | Alejandro Jiménez-Lagunes2 | Juan Estrada-Avalos3 | Jorge E. Mauricio-Ruvalcaba4 | José S. Padilla-Ramírez4

1National Institute of Research on Forestry, Agriculture, and Livestock (INIFAP). Experimental Field Station Pabellón, National Laboratory of Modeling and Remote Sensing, Pabellón de Arteaga, Mexico
2State Committee of Plant Health at Sonora. Pedro Villegas, 51, Col Del Razo Hermosillo, Mexico
3National Institute of Research on Forestry, Agriculture, and Livestock (INIFAP). National Center of Disciplinary Research. Km 6.5 right margin Sacramento channel, Gómez Palacio, Mexico
4National Institute of Research on Forestry, Agriculture and Livestock (INIFAP). Experimental Field Station Pabellón, National Laboratory of Modeling and Remote Sensing, km 32.5 highway Ags-Zac, Pabellón de Arteaga, Mexico

Correspondence
Rodríguez-Moreno, Victor M., National Institute of Research on Forestry, Agriculture, and Livestock (INIFAP). Experimental Field Station Pabellón, National Laboratory of Modeling and Remote Sensing, Pabellón de Arteaga, Mexico.
Email: rodriguez.victor@inifap.gob.mx

Abstract
Classification and regression trees (CARTs) for data analysis, an hourly weather dataset, and a 3 year field incidence and severity dataset of winter wheat rust were integrated to forecast pathogens’ presence/absence. The field dataset of incidence and severity was collected for three production cycles. Measured records of 88 Automatic Meteorological Stations and the indirect weather dataset generated in the Weather Research and Forecasting environment interpolated to each Automatic Meteorological Station location were analysed in the Python ecosystem. The focal point of the analysis was the severity of the disease. The analysis of direct weather data revealed the association of leaf rust severity with a night temperature of <14.25°C and global radiation of <521.67 W·m⁻², while the estimated dataset showed that its severity is better explained by the dew point temperature of <13.7°C and a mean temperature of <19.06°C. The direct dataset also indicated that stripe rust severity was associated with relative humidity of <88.73%, global radiation of <597.39 W·m⁻² and dew point temperature of <16.09°C, whereas the estimated data revealed that pathogen severity is better explained by a model composed of a dew point temperature of <14.6°C, night temperature of <20.4°C and a maximum temperature of <27.9°C. The severity and intensity analysis indicated the pathogen's preference for non-dry ambient conditions and the preference of stripe rust pathogen for humid and warmer temperatures than leaf rust. The weather thresholds of both pathogens, and CART analysis, unveiled that winter wheat rust can be forecasted. This constitutes the foundation of a more efficient extension programme based on the internet of things.

KEYWORDS
CART, forecasting, leaf rust, stripe rust, winter wheat rust, WRF

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. Meteorological Applications published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.
1 | INTRODUCTION

Rusts are among the most economically significant fungal diseases in cereal crops worldwide. The fungus can mutate to new strains that potentially can attack even previously resistant varieties (Bergmark, 2018). Rust diseases can rapidly develop under optimal weather conditions. The disease severity and yield losses are affected by environmental factors such as temperature and moisture (Chen, 2005). The basis for effective control of any disease requires an understanding of the disease cycle, and the climatic and other environmental factors influencing it, as well as the particular requirements of the host plant (Maloy, 2005). Lifecycles of the pathogen vary considerably between production regions when it is associated with temperature and moisture conditions. These two factors, however, remain as the major elements limiting regional production and the yield of wheat species. The characterization of the pathogen lifecycle through climate indices indicates how these factors affect its survival rate, time of infection, latency period and sporulation rate. These bio indices are fundamental datasets that feed forecasting models (Coakley and Line, 1982).

Three rust diseases occur on wheat: stem rust (StR), leaf rust (LR) and stripe rust (SR). A particular species of *Puccinia* causes each one. Their names originated from their appearance on the plant. LR is the most common; it occurs every year and is generally found on leaves but may also infect glumes and awns. SR is not as prevalent as other rusts because many wheat varieties are resistant to this disease; it occurs primarily on stems but can also be found on leaves, sheaths, glumes, awns and even seed. StR is distinguished by the presence of light yellow and straight-sided pustules that occur in stripes on leaves and heads (Marsalis and Goldberg, 2017).

Each rust fungi species has particular environmental requirements that include a film of water on the leaf surface from intermittent rains or heavy dews and temperatures conducive for the germination and growth of the pathogen (Marsalis and Goldberg, 2017). Relative humidity, air temperature and precipitation are critical conducing factors to the infection and progress of winter wheat LR (Junk et al., 2016). Hot summers and dry weather conditions are least conducive for *Puccinia striiformis* f. sp. *tritici*; above 25°C the fungus does not produce spores and above 29°C the pathogen dies (Evans et al., 2008). Minimum, optimum and maximum temperatures for causing infection are 0, 11 and 23°C, respectively (Hogg et al., 1969). LR inoculation is optimal when the overnight temperature and humidity predict a dew period or if rain is imminent (Kolmer, 2013); spore germination occurs after 4–8 hr at 20°C under 100% atmospheric humidity (Hu and Rijkenberg, 1998; Zhang et al., 2003).

StR infection requires high humidity during 4–6 hr at 10–15°C, with increasing time required at lower and higher temperatures (Murray et al., 2005); minimum, optimum and maximum temperatures for spore germination are 2, 15–24 and 30°C, respectively (Hogg et al., 1969); StR favours hot days (25–30°C), mild nights (15–20°C) and wet leaves from rain or dew. Severe epidemics and losses can occur when the flag leaf is infected before anthesis (Chester, 1946).

In Mexico, the forecasting models for predicting plant diseases based on meteorological data have been considered as an information and communication technology (ICT) tool resembling modern extension programmes. Predictive modelling techniques involve using deep data mining and probability tools in climate and weather forecasting as early warning systems for predicting outcomes. In plant pathology, a statistical approach named Window Pane was proposed in 1982 (Coakley and Line, 1982) for identifying and quantifying disease–environment relationships. Gouache et al. (2015) report that Window Pane analysis systematically calculates synthetic climate variables over overlapping time frames (ranging from a few weeks to a few months in length) for every growing season. After that, correlation coefficients between each environmental variable and the observed “target variable” (such as the disease level at the end of a season) are computed (Gouache et al., 2015). This approach is used to identify critical periods in which the variations in specific climatological variables lead to variations in disease expression (Gouache et al., 2015). These results serve to improve the understanding of the studied pathosystem and provide useful knowledge for managing the disease (Gouache et al., 2015). According to Campbell and Neher (1994), incidence assessment is one of the most difficult challenges because it is based on recognizing symptoms or signs when they become visible. The proportion of healthy to unhealthy plants depends greatly on technical interpretation and the correct association between expressed symptoms, pathogen identification and specificity of the plant host. Severity data are the maximum reached from observations of well-trained field technicians. A naked eye observation method is generally used to decide a disease’s severity during the production cycle, but these results are subjective and it is not possible to measure disease extent precisely.

The lifecycle of an organism is ruled by specific environmental conditions that regulate exchange processes. To explore the numerical relationship between data of meteorological variables and disease severity, most disease–weather models use regression analysis; however, the use of data mining techniques is also widespread. Data mining is the collection of exponentially growing techniques, which are used to find some useful
information, patterns and knowledge from already given data (Aggarwal, 2015). The applications of data mining are uncountable; it is used in almost every aspect of life to provide valuable information to decision-makers (Chattamvelli, 2009). Data mining techniques can be broadly distinguished into two types: predictive and descriptive (Patel and Patel, 2014). According to Ayub and Moqurrab (2018), classification is one of the renowned data mining subfields that allocate data in a group for targeting predefined taxonomies or classes. Some major applications of data mining are healthcare, market analysis, finance, education, manufacture engineering, corporation surveillance and agriculture (Chattamvelli, 2009). Among the data mining classifiers, decision trees are one of the most intuitive algorithms as a result of their strength to forecast and assign probabilities, e.g. winter wheat rust presence/absence. Decision trees classify the dataset by using trees (Hill et al., 2014), and the final result is a tree with decision nodes (attributes) and leaf nodes (class label) (Amir, 2018). The whole idea is to approximate the target class (presence/absence) by analysing the dataset (weather dataset).

The objective of this study was to explore the use of the classification and regression tree (CART), a particular type of nonlinear predictive model, to identify key weather–disease links. The affinities, the near real-time weather dataset and web-tool computing analysis techniques will provide invaluable information to forecast the presence/absence of winter wheat rust in the frame of 5 days. This study strives to model how winter wheat rust epidemics are correlated with weather variables using the analysis of a field serial dataset and two meteorological datasets: the direct dataset registered by an Automatic Weather Stations (AWS) network and the estimated weather data obtained by the Weather Research and Forecasting (WRF) model.

2 | MATERIALS AND METHODS

2.1 | Dataset

2.1.1 | Field data

Wheat rust field data was obtained from farm growers during a 3 year whole field campaign of winter wheat (production cycles 2013–2014, 2014–2015 and 2015–2016). This campaign started in autumn–winter of each production cycle. Registered variables were coordinate data, sampling date, observed incidence and severity. The recorded data and the materials used in preparation of the field operation until the field database was obtained are shown in Figure 1. Campaigns were undertaken by expert technicians using standardized identification tools.

The winter wheat rust severity was estimated as the proportion of total green leaf area affected. The field campaign was under the responsibility of the State Committee of Plant Health at Sonora (CESAVE) and included the detection of LR, SR and foliage aphids. The field dataset was derived from 500 wheat production farms in the Yaqui-Mayo-Fuerte Valleys of Sonora, Mexico (Figure 2). Some of the agronomics of four of the five most recommended varieties of wheat for the area of study are presented in Table 1. Field data involved 100 plants in 10 randomly selected sites inside a farm where 10 individuals showed disease symptoms. These farms were randomly selected. Disease severity was assigned a value of 10% if at least one of 10 plants inspected had disease symptoms.

Sample sites were not the same from one cycle to other, because field technicians were specifically looking for the presence of disease symptoms. Two types of rust were identified, LR and SR. As observed in Figure 1, LR was detected in the southern region of the valley, closer to the border with the state of Sinaloa, another important Mexican agricultural region; however, SR was located throughout the Yaqui-Mayo Valley, a cooler region.
Differences in the incidence and severity between epidemics of LR and SR on winter wheat rust during the production cycle were noticed, as seen in Figure 3 and Table 2. The dominant field disease was SR and this may be due to the less warm prevailing temperatures. Comparing both wheat rust diseases, SR was noted as having a light inter-year effect with its maximum severity in March 2016 (70%), February 2015 (20%) and March 2014 (50%). The maxima for LR were more consistent throughout production cycles, in April 2014 (10%), February 2015 (25%) and less than 2% for 2016 (Figure 3).

In Figure 3 it is consistently noted that during the February and March winter period wheat rust registered a maximum in incidence and severity. This may foster support for issuing mitigation programme recommendations by the month and an integrated pest management regimen promoting programmes that maximize the efficiency of cultural practices including the time, frequency and dosage of pesticide applications.

Noting the difference in optimal environmental temperatures, these results suggest that the incidence and severity of winter wheat rust in this region must have their own optimal parameters, which this study attempts to establish using the proposed methodology and data analysis.

**TABLE 1** Agronomic characteristics of wheat genotypes

| Wheat variety     | Gleaning (days) | Physiological maturity (days) | Plant height (cm) | Winter wheat rust       | Yield (t ha⁻¹) |
|-------------------|-----------------|-------------------------------|-------------------|-------------------------|----------------|
| Cirno C2800       | 72              | 108                           | 65                | LR resistant            | 5.6–6.3        |
|                   |                 |                               |                   | SR resistant            |                |
| Borlaug 100       | 81              | 120                           | 97                | LR resistant            | 7.5            |
|                   |                 |                               |                   | SR resistant            |                |
| Villa Juarez F2009| 76              | 119                           | 96                | LR resistant            | 7.4            |
|                   |                 |                               |                   | SR resistant            |                |
| Tacupeto F2001    | 81              | 121                           | 103               | LR susceptible          | 7.4            |
|                   |                 |                               |                   | SR susceptible          |                |

Data provided by PIEAES A.C., Board of Trustees for Agricultural Research and Experimentation of the State of Sonora, A.C.
2.1.2 Climate data

The serial dataset was obtained from two sources: (a) a “measured” dataset origin in the REMAS network, a group of 88 AWSs with state spatial distribution in Sonora, Mexico, and (b) an “indirect” dataset of WRF origin, with dx = 13 km extrapolated data corresponding to each AWS location. Both datasets were collected hourly and included nine variables: rain (mm), minimum temperature \( T_{\text{min}} \) (°C), maximum temperature \( T_{\text{max}} \) (°C), mean temperature \( T_{\text{mean}} \) (°C), maximum wind speed \( W_{\text{max}} \) (km·h\(^{-1}\)), wind direction \( W_{\text{dir}} \) (° azimuth), dewpoint temperature \( T_{\text{DWPt}} \) (°C), relative humidity (RH) (%), night temperature \( T_{\text{night}} \) (°C) and global radiation \( \text{Rad}_{\text{G}} \) (W·m\(^{-2}\)).

A descriptive flowchart of the weather dataset (direct and indirect) is shown in Figure 4.

| Year | Disease | Month | Incidence \((n)\) | Severity (%) |
|------|---------|-------|-------------------|--------------|
| 2014 | LR      | January | 1 | 0.02 |
|      |         | February | 20 | 0.04 |
|      |         | March | 24 | 0.08 |
|      |         | April | 13 | 0.10 |
|      | SR      | January | – | – |
|      |         | February | 110 | 0.28 |
|      |         | March | 189 | 0.50 |
|      |         | April | 15 | 0.13 |
| 2015 | LR      | January | – | – |
|      |         | February | 17 | 0.25 |
|      |         | March | 38 | 0.11 |
|      |         | April | – | – |
|      | SR      | January | 5 | 0.10 |
|      |         | February | 23 | 0.20 |
|      |         | March | 21 | 0.08 |
|      |         | April | – | – |
| 2016 | LR      | January | 1 | 0.01 |
|      |         | February | 7 | 0.02 |
|      |         | March | 7 | 0.02 |
|      |         | April | – | – |
|      | SR      | January | 4 | 0.01 |
|      |         | February | 160 | 0.40 |
|      |         | March | 84 | 0.70 |
|      |         | April | 1 | 0.10 |

\( n \) stands for disease incidence (count); incidence numbers correspond to unhealthy plants visibly affected with wheat rust symptoms; severity numbers indicate the amount of tissue affected expressed as a percentage.
As expected for any dataset recorded in field campaigns, some missed and aberrant data were observed; these items were excluded from the analysis. The indirect dataset (WRF origin) had not lost data. The indirect rain data were estimated based on RAINC and RAINNC parameterizations. RAINC is produced by the cumulus (convective) parameterization; the multi-scale Kain–Fritsch was implemented following the recommendations of Kain (2004). RAINNC is produced by the cloud microphysics scheme as delineated by Lin et al. (1983) where they separate a total cloud water variable into five predicted hydrometeors (cloud water, cloud ice, rain, snow and graupel). On a scale >10 km, RAINNC is mostly from stratiform clouds. At a high resolution <5 km, the grid scale may resolve some cumulus clouds, and their precipitation will be included in RAINNC. For domains with resolution <2–5 km, no cumulus scheme is used because most or all convection will be resolved at the grid scale. In this case, RAINC = 0.

In this context, since there are nine variables, all 512 different combinations were explored; it was found that there were several nearly identical results.

### 2.2 Dataset analysis and exploration

A setup was established on Python v. 3.6 handling three datasets. The first contains meteorological records (AWS network), the second estimated meteorological records (of WRF origin) and the last embodies the field data campaign (LR and SR incidence and severity). The very first step of the analysis correlated, estimated and registered meteorological records with field dates where winter wheat rust was observed. The procedure was to match both serial datasets and then locate the entire match with a rule of thumb n minus fourth day for weather variables, which in theory precedes the pathogen's presence. Then, this rule of thumb was used to forecast pathogen presence on the next 5 days. This procedure is consistent with the conceptual clustering proposed by Michalski and Stepp (1983) along with Fisher (1987) where the learner tries to discover concepts by grouping the observations in a “meaningful” way. This dataset was then extracted from its original file for future analysis.

Then, the relation between derived weather variables and disease observations was calculated. Due to their simplicity and efficiency, tree-based regression models were chosen to analyse the serial dataset. Regression trees are a technique for dealing with subjects having a large number of variables and cases by using a fast divide and conquer greedy algorithm that recursively partitions the given training data into smaller subsets (Rainho-Alves-Torgo, 1999).

Here, measured data are referred to as AWS origin, estimated data as WRF origin and a weather factor as a weather variable transformed by a weather function. The weather functions here were (a) the number of...
consecutive days where weather data conform to a declared threshold, the range of the weather factor, and (b) the accumulation rate above or below a specific threshold.

2.3 Statistical analysis

CART analysis, also known as decision trees, is a non-parametric simple yet powerful analytical tool, which helps determine the most important variables in a particular dataset for obtaining a robust explanatory model. It does not require a functional form, and the number of classes is known a priori. Variables are not selected in advance; it can handle datasets with a complex structure including either categorical or continuous variables. A simple split of data into a presence (winter wheat rust identified in the field) or absence (no disease) classification was used to analyse the weather data and the field data together. In the CART procedure, the field data of severity were indicated as a response variable and all meteorological data as explanation variables. The classified outputs were useful to discern whether the meteorological variables analysed can predict an epidemic (1) or the absence (0) of the pathogen.

A three-component model was proposed as a starting point for forecasting the pathogen’s presence. All components are associated with temperature (Figure 5).

From Figure 5, $T_{\text{mean}}$ was calculated as the average of the maximum and minimum temperatures; $T_{\text{night}}$ stands for the temperature during the night, calculated from 8:00 pm to 7:00 am of the next day; and $T_{\text{DWpnt}}$ is indicated as a minimum of 5°C. For forecasting the presence of winter wheat rust a rule of thumb was declared: five consecutive days with prevailing meteorological conditions, as declared in the model, will trigger an alert warning for a priori pathogen presence. These guidelines resemble a logit regression where only two results are possible: presence (1) or absence (0). The scheme (presence/absence) is robust to predict the occurrence (1) and the non-occurrence (0) of the disease based on compliance with the meteorological conditions that favour its presence (Figure 6 and Table 3).

In Table 3, the most robust outputs of the model were the extreme diagnosed conditions represented here with five successive outputs of the same characters (1s or 0s). Intermediate categories should be interpreted by users in accordance with their experience in the management of the disease and the wheat production system.

3 RESULTS AND DISCUSSION

3.1 Winter wheat rust models

The proposed models to associate disease incidence and severity of winter wheat rust were highly correlated with weather data ($r = 0.89$ and $r = 0.87$, respectively). Models that best described the relationship of weather variables with winter wheat rust epidemics also enhanced some differences in weather variables, the dataset origin and the pathogens (Figures 7 and 8).

As shown in Figure 7, LR severity was hierarchically related to $T_{\text{DWpnt}}$, $T_{\text{mean}}$ and $T_{\text{min}}$, agreeing with the measured dataset. The indirect dataset unveiled that LR severity was explained by the variables $T_{\text{night}}$, $T_{\text{mean}}$ and $\text{RadG}$. The variable link between the two databases was $T_{\text{mean}}$, which was located at the same second graded level of both the trees. These results provide evidence of the disease severity of winter wheat rust based on weather variables. One model to forecast the occurrence of LR based on air temperature, RH and rain has been proposed (Rader et al., 2007). Contrasting with this model, as shown in the two decision trees (Figures 7 and 8), LR severity is conditioned to a range of values of variables associated with air temperature, but not with those related to humidity such as rain or RH.
In contrast to LR prediction models, no common weather variables between databases were observed that might explain SR severity (Figure 8). The measured database yielded a hierarchical variables distribution of RH, RadG and $T_{DWpnt}$, while the indirect database unveiled $T_{DWpnt}$, $T_{night}$ and $T_{max}$. Temperature and RH are the weather variables necessary to explain the development of SR (Sandhu et al., 2017) but these results show that more weather variables are necessary.

### TABLE 3  Model outputs associated with the risk of the presence of winter wheat rust

| Model output | Diagnosis          |
|--------------|--------------------|
| 1 1 1 1 1    | High presence risk |
| 1 1 1 1 0    | Mid-presence risk  |
| 1 1 1 0 0    | Presence risk      |
| 1 1 0 0 0    | Low presence risk  |
| 1 0 0 0 0    | Very low presence risk |
| 0 0 0 0 0    | No risk of presence |

3.2 | Weather variables’ association with disease severity

Connection with non-dry environments served to show that weather variables are associated with disclosing a pathogen’s presence. One difference between the two causal organisms is precisely their ability to adapt to different temperature thresholds. There were some differences in optimal environmental conditions for each pathogen. In a technical report of *Crop Science* (2018, April 2), SR appears earlier in the season because development is enhanced by the cool, moist weather early in the growing season; LR, however, is more prevalent in later spring when temperatures are warmer; and SR develops optimally between 7.2 and 12.2°C. The Cereal Disease Laboratory of the Agricultural Research Service (2018) mentioned that the optimal environmental range for SR is 0–25°C while for LR the optimal range is between 20 and 25°C, with nights between 15 and 20°C. Based on the number of nodes in the decision tree's structure, the indirect dataset had fewer nodes than in the measured dataset. This fact *per se* is enough to suggest that the indirect model would be more parsimonious than the measured dataset. When

![FIGURE 7](image-url)  
Leaf rust severity as explained by weather variables. Left, direct dataset; right, indirect dataset

![FIGURE 8](image-url)  
Stripe rust severity as explained by weather variables. Left, direct dataset; right, indirect dataset
the severity of the rust caused by both pathogens is explained, LR resulted in temperature preferences lower than SR; the variables $T_{DWPn}$ and $T_{night}$, as well as RH and $T_{DWPnt}$ were the most significant for LR and SR, respectively. These results partially support the assertion of Hau and de Vallavieille-Pope (2006) that SR fungus favours cold temperatures; the difference is that RH is a function of both how much moisture the air contains and the temperature. However, Temizyurek and Dadaser-Celik (2018) documented that the relationships between water temperature and RH and wind speed were comparatively weak.

Along with good and stable weather conditions for promoting the outbreak and the biological cycle of winter wheat rust pathogens, forecasting their presence has experienced a paradigm, which takes the form of probability distributions (Fraley et al., 2011).

3.3 Early warning systems for crop protection

As a mechanism to influence yield losses, epidemic patterns are very important (Savary et al., 2006). In plant disease epidemiology wheat is a crop suitable for robust estimation of epidemic patterns because its production cycle is expressed in heat units and chill units. Proper dissemination of the outputs of predictive systems to farmers is also important for effective disease management. The specific array in which the dataset was integrated, the selection of weather variables, the starting initial conditions of the model and the analysis of the results as seen in regression trees might constitute the foundation of an early warning system to improve preparedness of farmers to handle the emerging winter wheat rust disease in this large agricultural region.

Plant disease forecasting systems often provide information to farmers for management decisions to avoid initial inoculum or to slow the rate of an epidemic (Esker et al., 2008). Interactive retrieval of disease incidence and severity and weather data in near real time provides beyond state-of-the-art research employing new diagnostic tools to enable rapid and precise identification of rust pathogens, along with planning activities for production systems. This research has the potential for a breakthrough in the area of integrated pest management by going beyond field sampling campaigns and determining the risk of the presence of winter wheat rust based on the internet of things.

3.4 ICT framework

In two decades, massive technological development has transformed people’s lives and provided them with endless opportunities to get access to critical information to support decision-making processes. ICT tools have been transforming traditional agricultural extensionism by promoting the flow of information based on digital agricultural programmes. The value of the framework’s information depends on many factors, particularly farmers’ perceived risk at the beginning of the growing season and the accuracy of the system’s forecast (Roberts et al., 2006). In developing countries, due to lack of early warning systems, farmers do not apply preventive measures to reduce the risk of yield loss of wheat. Warning systems consist of concatenated modules of datasets, data mining techniques for analysis, web services and mobile apps developments, and an operative computer environment running under the internet of things frame. In the very short term, ICT might constitute the new paradigm of agricultural extension programmes, sharing datasets and related information from research institutions through extension agents to farmers and policymakers. For developing countries, radio messages, pamphlets, brochures and face meetings will soon be replaced by short message service (SMS) alert messages, as a more successful approach for agricultural extension programmes.

3.5 Dissemination of information

This paper has unveiled how digital agriculture is gaining space in traditional agriculture. Modelling datasets of weather data, field data and machine-learning techniques to forecast diseases constitutes the foundation of an integrated pest management. The key element to reach success in preventing losses due to winter wheat rust is to secure users’ continuous access to the forecast through web services and developments for mobile devices so that they can be informed on the current and future risks. A couple of models to forecast winter wheat rust (LR and SR) for regional and local dispersion and environmental suitability for infection have been developed based on weather variables and high-performance computing resources. There are many applications for big data analytics in agriculture that correlate weather patterns, actual or past disease dynamics, and the recommendation of optimal use of pest management solutions; however, early warning systems are not in common use in developing countries where traditional agricultural practices prevail (FAO, 2018). In developing countries this will raise a “natural” rejection barrier from farmers towards digital agriculture. Traditional farmers need to be convinced that the forecast model predicts the winter rust disease with accuracy. In this adoption process researchers and promoters of its use have to develop an effective promotion programme.
4 | CONCLUSIONS

The presence of leaf rust (LR) and stripe rust (SR) was successfully predicted through two weather-based models. Models unveiled that temperature-associated variables are the most important to explain and forecast the presence of LR, while SR is better explained with a humid and warmer environment. Thresholds of temperature and relative humidity of both pathogens, the specific design and structure of the weather-based datasets, the incidence and severity field campaign, data mining techniques for analysis, and farmers’ involvement to feedback with field condition reports of presence/absence of the disease improve the efficiency of the web and apps platforms to predict winter wheat rust disease. These components are ideal to start information and communication technology support developments to provide users, farmers, researchers and policymakers with useful information through web services and mobile applications. These elements are the foundation of an early warning system.

ACKNOWLEDGEMENTS

The authors greatly acknowledge the technical support of INIFAP. They also wish to thank Wally Weerts for his objective review and suggestions on English grammar, and Rufina Hernández Martinez for her contribution on restructuring the paper. They also wish to thank two anonymous reviewers whose comments were very valuable for improving the paper.

COMPLIANCE WITH ETHICAL STANDARDS

This study was funded by INIFAP and the State Committee of Plant Health at Sonora. Rodriguez-Moreno, V.M., Estrada-Avalos, J., Padilla-Ramirez, J.S., Mauricio-Ruvalcaba, J.E. and Jimenez-Lagunes A. declare that they have no conflict of interest.

This paper does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

Aggarwal, C.C. (2015) Data Mining: The Textbook. New York, NY: Springer International Publishing. https://doi.org/10.1007/978-3-319-14142-8.

Amir, A. (2018) Decision tree (CART) algorithm in machine learning. Chapter 4: Decision tree (CART). Available at: https://medium.com/machine-learning-researcher/decision-tree-algorithm-in-machine-learning-248fb7de819e [accessed 24 October 2019]

ARS. Cereal Disease Lab. (2018) Wheat stripe rust. Available at: https://www.ars.usda.gov/midwest-area/stpaul/cereal-disease-lab/docs/cereal-rusts/wheat-stripe-rust [accessed 27 April 2018]

Ayub, U. and Moqrrab, S. (2018) Predicting crop diseases using data mining approaches: Classification. 1-6. https://doi.org/10.1109/ICPESG.2018.8384523

Bergmark, L. (2018) Understanding diseases. Mid-West fertilizer. Available at: https://www.midwestfertilizerinc.com/News/Understanding-Wheat-Diseases [accessed 10 August 2017]

Campbell, C.L. and Neher, D.A. (1994) Estimating Disease Severity and Incidence. In: Epidemiology and Management of Root Diseases. Berlin, Heidelberg: Springer, pp. 117–147. https://doi.org/10.1007/978-3-642-85063-9_5.

Chattamvelli, R. (2009) Data Mining Methods. Oxford, England: Alpha Science International.

Chen, X.M. (2005) Epidemiology and control of stripe rust [Puccinia striiformis f. sp. tritici ] on wheat. Canadian Journal of Plant Pathology, 27, 314–337.

Chester, K.S. (1946) The Nature and Prevention of the Cereal Rists as Exemplified in the Leaf Rust of Wheat. Waltham, MA: Chronica Botanica, p. 269.

Coakley, S.M. and Line, R.F. (1982) Prediction of stripe rust epidemics on winter wheat using statistical models. Phytopathology, 72, 1006.

Esker, P.D., Sparks, A.H., Campbell, L., Guo, Z., Rouse, M., Silwal, S.D., Tolos, S., Van Allen, B. and Garrett, K.A. (2008) Ecology and Epidemiology in R: Disease Forecasting. The Plant Health Instructor. https://doi.org/10.1094/PHI-A-2008-0129-01.

Evans, K., Israelasc, C., Barnhill, J. and Pace, M. (2008) Wheat Stripe Rust (Puccinia striiformis f. Sp. tritici). Fact sheet. Utah: Utah State University. Available at: http://utahpests.usu.edu/IPM/htm/field-crops/insects-and-diseases/wheat-stripe-rust [Accessed: 26th November 2018].

FAO. (2018) ICT’s for sustainable agriculture. Technologies for agricultural information sharing. Part II. Available at: https://www.fao.org/docrep/019/i3557e/i3557e01.pdf [accessed 27 April 2018]

Fisher, D.H. (1987) Knowledge acquisition via incremental conceptual clustering. Machine Learning, 2(1), 139–172. https://doi.org/10.1007/BF00114265.

Fraley, C., Raftery, A., Gneiting, T., Slaughter, M. and Berrocal, V. (2011) Probabilistic weather forecasting in R. Contributed Research Articles, 3(1), 55–63.

Gouache, D., Léon, M.S., Duyme, F. and Braun, P. (2015) A novel solution to the variable selection problem in window pane approaches of plant pathogen climate models: development, evaluation and application of a climatological model for brown rust of wheat. Agricultural and Forest Meteorology, 205, 51–59.

Hau, B. and de Vallavieille-Pope, C. (2006) Wind-dispersed diseases. In: Cooke, B., Jones, D. and Kaye, B. (Eds.) The Epidemiology of Plant Diseases. Dordrecht: Springer.

Hill, M.G., Connolly, P.G., Reutemann, P. and Fletcher, D. (2014) The use of data mining to assist crop protection decisions on kiwifruit in New Zealand. Computers and Electronics in Agriculture, 108, 250–257.

Hogg, W.H., Hounam, C.E., Mallik, A.K. and Zadoks, J.C. (1969) Chronica Botanica, p. 269. Waltham, MA: Chronica Botanica, p. 269.

Hu, G. and Rijkenberg, F.H. (1998) Subcellular localization of beta-1,3-glucanase in Puccinia recondita fsp tritici-infected wheat leaves. Planta, 204, 324–334.

Junk, J., Kouadio, L., Delfosse, P. and El Jarroudi, M. (2016) Effects of regional climate change on brown rust disease in winter wheat. Climatic Change, 135(3–4), 439–451. https://doi.org/10.1007/s10584-015-1587-8.

Kuo, C.C. and Hsu, C. (2016) Machine learning-248fb7de819e [accessed 24 October 2019]
Kain, J.S. (2004) The Kain–Fritsch convective parameterization: an update. Journal of Applied Meteorology, 43, 170–181.

Kolmer, J. (2013) Leaf rust of wheat: pathogen biology, variation and host resistance. Forests, 4, 70–84. https://doi.org/10.3390/F4010070.

Lin, Y.L., Farley, R.D. and Orville, H.D. (1983) Bulk parameterization of the snow field in a cloud model. Journal of Climate and Applied Meteorology, 22, 1065–1092.

Maloy, O.C. 2005 Plant disease management. The plant health instructor. Available at: https://doi.org/ 10.1094/PHI-I-2005-0202-01 http://aulavirtual.agro.unlp.edu.ar/pluginfile.php/23847/mod_resource/content/1/Manejo%20de%20Enfermedades%20en%20Plantas.pdf [accessed 29 November 2019]

Marsalis, M.A., and Goldberg, N.P. (2017) Leaf, stem, and stripe rust disease of wheat. NMS. Available at: https://aces.nmsu.edu/pubs/_a/A415/welcome.html [accessed 29 November 2019]

Michalski, R.S. and Stepp, R.E. (1983) Learning from observation: conceptual clustering. In: Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. (Eds.) Machine Learning: An Artificial Intelligence Approach. Tioga: Palo Alto, CA, pp. 331–363.

Murray, G, Wellings, C., Simpfendorfer, S., and Cole C. 2005. Stripe rust: understanding the disease in wheat. Available at: https://www.dpi.nsw.gov.au/_data/assets/pdf_file/0006/158964/stripe-rust-in-wheat.pdf [accessed 26 November 2018]

Patel, H. and Patel, D. (2014) A brief survey of data mining techniques applied to agricultural data. International Journal of Computer Applications, 9, 6–8.

Rader, T., Racca, P., Jorg, E. and Hau, B. (2007) PUCREC/PUCTRI a decision support system for the control of leaf rust of winter wheat and winter rye. EPPO Bulletin, 37(2), 378–382.

Raínho-Alves-Torgo, L.F. (1999) Inductive learning of tree-based regression models. Thesis, Faculdade de Ciências da Universidade do Porto, Setembro

Roberts, M.J., Schimmelpfennig, D., Ashley, E. and Livingston, M. (2006) The Value of Plant Disease Early-Warning Systems. A Case Study of USDA’s Soybean Rust Coordinated Framework. Economic Research Report No: (ERR, 18), Washington, DC: US Department of Agriculture, Economic Research Service.

Sandhu, S.K., Dhaliwal, L.K. and Pannu, P.P.S. (2017) Effect of meteorological and micrometeorological parameters on stripe rust (Puccinia striformis f. Sp. tritici) severity of wheat. International Journal of Current Microbiology and Applied Sciences, 6(7), 4512–4520. https://doi.org/10.20546/ijcmas.2017.607.470.

Savary, S., Teng, P.S., Willocquet, L. and Nutter, F.W. (2006) Quantification and modelling crop losses: a review of purposes. Annual Review of Phytopathology, 44, 89–112.

Crop Science. (2018) Wheat rust diseases. Available at: https://www.cropsscience.bayer.us/learning-center/articles/wheat-rust-diseases#phcontent_5_divAccordion [accessed 27 Apr 2018]

Temizyurek, M. and Dadaser-Celik, F. (2018) Modelling the effects of meteorological parameters on water temperature using artificial neural networks. Water Science and Technology, 77(6), 1724–1733. https://doi.org/10.2166/wst.2018.058.

Zhang, L., Meakin, H. and Dickinson, M. (2003) Isolation of genes expressed during compatible interactions between leaf rust (Puccinia triticina) and wheat using cDNA-AFLP. Molecular Plant Pathology, 4, 469–477.

How to cite this article: Rodríguez-Moreno VM, Jiménez-Lagunes A, Estrada-Avalos J, Mauricio-Ruvalcaba JE, Padilla-Ramírez JS. Weather-data-based model: an approach for forecasting leaf and stripe rust on winter wheat. Meteorol Appl. 2020; 27:e1896. https://doi.org/10.1002/met.e1896