Check for updates

Received: 12 October 2023 | Accepted: 23 February 2024

DOI: 10.1111/ppa.13902

Plant Pathology WILEY

ORIGINAL ARTICLE

Predicting airborne ascospores of *Sclerotinia sclerotiorum* through machine learning and statistical methods

Jonathan Reich[1,2] | Debra McLaren[3] | Yong Min Kim[3] | Owen Wally[4] | Dmytro Yevtushenko[5] | Richard Hamelin[2] | Syama Chatterton[1]

[1]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, Alberta, Canada

[2]Faculty of Forestry, University of British Columbia, Vancouver, British Columbia, Canada

[3]Brandon Research and Development Centre, Agriculture and Agri-Food Canada, Brandon, Manitoba, Canada

[4]Harrow Research and Development Centre, Agriculture and Agri-Food Canada, Harrow, Ontario, Canada

[5]University of Lethbridge, Lethbridge, Alberta, Canada

Correspondence
Syama Chatterton, Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, 5403-1 Ave. South, P.O. Box 3000, Lethbridge, Alberta T1J 4B1, Canada.
Email: syama.chatterton@canada.ca

Abstract

A main biological constraint of dry bean (*Phaseolus vulgaris*) production in Canada is white mould, caused by the fungal pathogen *Sclerotinia sclerotiorum*. The primary infectious propagules of *S. sclerotiorum* are airborne ascospores and monitoring the air for inoculum levels could help predict the severity of white mould in bean fields. Daily air samples were collected in commercial dry bean fields in Alberta, Manitoba and Ontario and ascospores were quantified using quantitative PCR. Daily weather data was obtained from in-field weather stations. The number of ascospores on a given day was modelled using 63 different environmental variables and several modelling methods, both regression and classification approaches, were implemented with machine learning (ML) (random forests, logistic regression and support vector machines) and statistical (generalized linear models) approaches. Across all years and provinces, ascospores were most highly correlated with ascospore release from the previous day (r ranged from 0.15 to 0.6). This variable was also the only variable included in all models and had the greatest weight in all models. Models without this variable had much poorer performance than those with it. Correlations of ascospores with other environmental variables varied by province and sometimes by year. A comparison of ML and statistical models revealed that they both performed similarly, but that the statistical models were easier to interpret. However, the precise relationship between airborne ascospore levels and in-field disease severity remains unclear, and spore sampling methods will require further development before they can be deployed as a disease management tool.

KEYWORDS
aerobiology, disease forecasting, dry bean, machine learning, modelling, white mould

1 | INTRODUCTION

White mould, caused by the fungal pathogen *Sclerotinia sclerotiorum*, poses a significant biological constraint to dry bean (*Phaseolus vulgaris*) production across Canada (Boland & Hall, 1987; Harding et al., 2017; Pesticide Risk Reduction Program, 2017). In spring, prolonged cool temperatures (15–25°C) (Clarkson et al., 2007; Fall et al., 2018; Hao et al., 2003; Wu & Subbarao, 2008) and moist soil conditions (−100 to 0 kPa) (Clarkson et al., 2007; Hao et al., 2003; Sun & Yang, 2000) promote the germination of soilborne sclerotia

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2024 His Majesty the King in Right of Canada and The Authors. *Plant Pathology* published by John Wiley & Sons Ltd on behalf of British Society for Plant Pathology. Reproduced with the permission of the Minister of Agriculture and Agri-food Canada.

and the formation of apothecia. Each apothecium can release hundreds of thousands of ascospores, the primary infectious propagule of *S. sclerotiorum*, per day over the course of several days (Bourdôt et al., 2001; Clarkson et al., 2003; Schwartz & Steadman, 1978). These airborne ascospores land on senescing plant tissue, germinate and invade the plant cell walls, leading to diseases in hundreds of plant species (Bolton et al., 2006; Willbur et al., 2019).

Dry bean production in Canada is predominantly concentrated in southwestern Ontario (40% of production in 2021), southern Manitoba (29%) and southern Alberta (26%) (Statistics Canada, 2023). In southern Alberta, where dry bean fields are irrigated, the environmental conditions for apothecia formation and white mould development are consistently favourable. Consequently, disease incidence levels can reach 100% in some fields (Reich et al., 2019). In contrast, dry bean production in Manitoba and Ontario occurs only on dryland. Although white mould remains the primary disease of concern in these provinces, disease levels tend to be lower than in Alberta, with prevalence reaching up to 40% in recent years (Kim et al., 2017, 2019).

Predicting epidemics caused by *S. sclerotiorum* has remained notoriously difficult across a range of host crops (Reich & Chatterton, 2023). Consequently, dry bean growers in southern Alberta often resort to applying two or three fungicides every season as a preventative measure, regardless of perceived or actual disease risk. In Ontario and Manitoba, fungicide application is less frequent and relies partially on weather forecasts that indicate conducive conditions for white mould development. Previous approaches to predicting diseases caused by *S. sclerotiorum* have often focused on predicting apothecia presence as an indication of pathogen intensity (Willbur et al., 2018) or have emphasized weather and agronomic factors associated with final disease outcomes (Mila et al., 2004; Twengström et al., 1998). However, a potential drawback of these approaches is their limited ability to provide real-time evaluation of the risk of disease, as observations or predictions of apothecia or disease may come too late to make timely management decisions (i.e., fungicide application), or predictions based solely on environmental factors do not account for actual pathogen levels in the field. In this context, adopting preventative sprays may be a logical risk reduction strategy.

In certain pathosystems, real-time monitoring of airborne inoculum has proven useful in informing effective disease management strategies. For instance, in rust (*Puccinia* spp.) diseases of wheat and late blight of potato caused by *Phytophthora infestans*, early detection of airborne spores or sporangia can alert growers to the presence of the pathogen, enabling timely prevention measures and helping to avoid unnecessary early season fungicide applications (Aboukhaddour et al., 2020; Araujo et al., 2021; Rosete et al., 2021). Similarly, a regional warning system based on airborne spore levels has been developed for blackleg disease of canola (*Leptosphaeria maculans*), which alerts growers when the risk of disease is high (Jędryczka et al., 2008, 2013). A similar system has been developed to notify growers of several diseases of brassicas in the UK (https://www.syngenta.co.uk/brassica-alert). At finer scales, regional networks of spore collectors have played a crucial role in managing powdery mildew (*Podosphaera aphanis*) of strawberry (Van der Heyden et al., 2014) and Botrytis leaf blight (*Botrytis squamosa*) of onion (Carisse et al., 2009). These systems rely on frequent collection and analysis of samples, providing real-time information on the most judicious and effective use of fungicides based on airborne inoculum levels.

Although ascospores clearly play an integral role in diseases caused by *S. sclerotiorum*, ascospore liberation, escape and transport are complex phenomena and remain some of the most poorly understood parts of the life cycle of many airborne fungi (Aylor, 1990). The factors influencing these processes comprise many complex interactions, and Mahaffee et al. (2023) offer a general framework for understanding these processes. Removal of ascospores from the apothecium may be a result of passive (wind), active (possibly through puffing or 'spore plumes'; Clarkson et al., 2003) or mechanical (rain, physical disturbances like leaves shaking) dispersion. Escape from the crop canopy is influenced by the strength of turbulence, ambient winds and the porosity of the canopy (Aylor, 1999; de Jong et al., 2002); and transport above the crop canopy is influenced by factors such as wind, topography and physical barriers (Mahaffee et al., 2023). Although the development of apothecia typically takes place over several weeks (Clarkson et al., 2007; Hao et al., 2003; Sun & Yang, 2000), and each apothecium can release ascospores for a period of several days (Clarkson et al., 2003; Schwartz & Steadman, 1978), it is likely that ascospore release itself occurs at a scale of hours or days rather than weeks, and is more acutely related to diurnal patterns, disturbance events or sudden changes in environment than long-term environmental conditions (Archer & Yuan, 2004; Bourdôt et al., 2001; Clarkson et al., 2003; de Jong et al., 2002; Savage et al., 2013). This framework for spore transportation can help guide sampling efforts aimed at informing growers on the disease risk posed by *S. sclerotiorum* in a growing season.

A complimentary or alternative approach to frequent sample collection and analysis is to develop models that can predict airborne pathogen levels based on environmental characteristics. This approach offers advantages such as reduced labour and cost, as well as the ability to forecast several days into the future based on weather forecasts (Smith et al., 2018). However, the development of such models still requires a large and robust dataset, along with statistical methods capable of uncovering potentially complex or nonlinear relationships between environmental variables and the predicted outcome. In recent years, machine learning (ML) algorithms have gained attention as useful tools for prediction in addition to more traditional linear and nonlinear regression methods often employed in plant pathology (de Oliveira Aparecido et al., 2020; Hamer et al., 2020; Shahoveisi et al., 2022).

The choice of the best statistical or ML model often depends on the characteristics of the dataset and the underlying assumptions regarding data distributions. Consequently, there may be no a priori best approach to model fitting. In the case of airborne ascospores of *S. sclerotiorum*, numerous complex biological and environmental factors influence ascospore liberation and dispersal,

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

making this situation an excellent candidate for comparing several modelling techniques. Therefore, the objectives of this research were to (a) assess the utility of spore samplers for collecting ascospores of *S. sclerotiorum* across the three dry bean-growing regions in Canada, (b) identify the environmental drivers of ascospore release in *S. sclerotiorum* and (c) compare statistical and ML methods for predicting daily ascospore levels in commercial dry bean fields.

2 | MATERIALS AND METHODS

2.1 | Field selection

For each province, the fields used for ascospore surveillance were selected with the help of industry agronomists, dry bean growers and researchers who identified fields that, historically, have had either low or high incidences of white mould (Reich et al., 2023). After identifying growers who were willing to have spore samplers deployed in their fields throughout the growing season, the fields to survey were selected to represent a range of disease potential (i.e., low incidence to high incidence expected).

2.2 | Spore sampling in Alberta

Four growing seasons (2018–2021) were dedicated to monitoring commercial dry bean fields in southern Alberta. In each monitored field, we deployed the Burkard 7-day volumetric cyclone sampler (Burkard Manufacturing) to collect daily samples for 24 h each day (00:01–23:59). The Burkard samplers intake air at a rate of 16.5 L min^{-1} (about 1 m^3 h^{-1}) and deposit aerosols in 1.5 mL microcentrifuge tubes. The samplers were programmed to change vials every day for the period of a week, and vials were collected and replaced weekly. In 2018, to assess the within-field distribution of ascospore dispersal, we sampled three commercial bean fields from June to August. These fields were located near Cranford, Vauxhall and Bow Island, which are 40–60 km away from each other. In each field, we positioned three samplers as follows: one sampler with the intake orifice 1 m above the ground and above the canopy, another sampler adjacent to the first one with the intake orifice 0.15 m above the ground and below the canopy and a third sampler 40 m from the first one, perpendicular to the predominant direction of wind and with the intake orifice also 1 m above ground. All samplers in commercial fields were placed inside the field at least 40 m from the edge and on the east side of the field (the lee side of prevailing winds). Samples were collected weekly and stored at −20°C until DNA extraction, as described below.

From 2019 to 2021, spore samplers were deployed from July to August in five to seven commercial bean fields near Vauxhall and Rolling Hills. Each field had at least one Burkard sampler, and in each year, one field had two samplers set up 40 m apart. Sampler setup followed a similar protocol for 2018 except no samplers were placed below the canopy: the intake orifice on each sampler was 1 m above the ground, and samplers were at least 40 m from the edge of the field and located on the leeside of predominant winds. Additionally, during each of these years (2019–2021), three Burkard samplers were placed within 150 m of each other at the Lethbridge Research and Development Centre (LeRDC) in Lethbridge, Alberta. These samplers were placed in experimental research plots of dry bean, canola and wheat and used to compare ascospore loads from host versus nonhost crops.

To assess the utility of existing spore trapping networks in Alberta to monitor for *S. sclerotiorum* spores, we analysed air samples collected by the Potato Growers Association (PGA) of southern Alberta in 2019. This group currently maintains a spore monitoring network for *P. infestans* under the direction of Dr Yevtushenko at the University of Lethbridge. The network consists of Burkard samplers deployed in or near commercial potato fields that are active for 3 h every afternoon. We selected a subset of samples to analyse from this network by identifying two samplers that were closest to a dry bean field containing a spore sampler from this study, as described above. DNA extraction and spore quantification were performed as described below for samples from bean fields.

2.3 | Spore sampling in Manitoba and Ontario

In 2019, 2020 and 2021, spore sampling was conducted in Manitoba and Ontario. In Manitoba, in 2019, one site was located at each disease nursery at the Carman Research Farm and the Morden Research and Development Centre; in 2020, only the Morden site was monitored; and, in 2021, the Carman and Morden sites, as well as an additional commercial field, were monitored. In Ontario, one sampler was placed in a disease nursery at the Harrow Research and Development Centre and two samplers were in nearby commercial bean fields each year. While the disease nurseries were irrigated, the commercial fields were grown without supplemental irrigation. For each field site, one spore sampler was deployed following the same configuration as described above for the commercial sites in Alberta.

2.4 | Comparison of spore samplers and sampler types for the collection of *S. sclerotiorum* ascospores

In 2019, we conducted a comparison to assess the relative collection abilities of different sampling methods for *S. sclerotiorum* ascospores. A GRIPST-2009 rotation impaction sampler (Aerobiology Research Laboratories) was installed directly beside a Burkard cyclone sampler, with the rotating arm at the same height as the intake of the Burkard sampler. The samplers were placed in a Sclerotinia stem rot of canola research plot at the LeRDC. The GRIPST-2009 sampler operated at 10% sampling capacity (sampled for 1 min every 10 min) and sampled from about 08:00 to16:00 Monday to Friday and samples were collected daily. The Burkard sampler collected for 24 h every day as described previously. To facilitate comparison of spore counts between samplers, data were transformed to ascospores per cubic metre per day, as described in the statistical analysis section below.

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.5 | Weather data

At each spore sampling site, a WatchDog Mini (model 2450) weather station (Spectrum Technologies Inc.) was mounted on PVC pipe and installed directly beside the Burkard spore sampler above the canopy. Weather stations recorded hourly air temperature (T, °C), relative humidity (RH, %), precipitation (Precip, mm), soil moisture (SoilWC, % vol/vol), soil temperature (SoilT, °C) and dew point (DP, °C). Vapour pressure deficit (VPD, mmHg) was calculated from the raw data according to Xu et al. (2000).

2.6 | Disease surveys

The dataset described in this research article was generated as part of a larger project involving white mould disease surveys. Efforts to relate white mould levels in commercial fields to airborne inoculum are presented by Reich et al. (2023). Because the main goal of this article is to determine what factors influence daily ascospore counts, disease survey data are not addressed in detail here but are referenced in the Discussion.

2.7 | DNA extraction

DNA was extracted from all aerosol samples using the protocol described by Ferencova et al. (2017). DNA was extracted directly from the 1.5 mL microcentrifuge tubes once they were retrieved from the samplers as follows: 50 mg of acid-washed 425–600 μm glass beads (Sigma-Aldrich), 100 μL sterile distilled water and 10 μL of 0.1 ng mL^{-1} synthetic tuna DNA (Li et al., 2015) were added to each vial. The tuna DNA was included as an exogenous internal control and amplified in a multiplex quantitative PCR (qPCR) assay with *S. sclerotiorum* DNA. Therefore, it acted as a control for false negatives; if the tuna DNA did not amplify, the sample would be removed from analysis. Samples were then placed in a TissueLyser II (Qiagen) for 30 s to dislodge particles from the sides of the vials. Following bead beating, 100 μL of a 10% (wt/vol) Chelex:TE mixture (Chelex 100, 50–100 mesh, Sigma-Aldrich; 10 mM Tris, pH 8.0, 0.1 mM EDTA, IDT) was added to each sample. Samples were then vortexed for 10 s and placed in a block heater at 95°C for 20 min. At 5, 10, 15 and 20 min, samples were vortexed for 3 s. Following the final vortex, samples were centrifuged at 16,000*g* for 1 min. Supernatant was collected and stored in new 1.5 mL tubes at −20°C.

2.8 | Generation of standard curve and qPCR analysis

Ascospores of *S. sclerotiorum* were obtained from InnoTech (Edmonton, Canada). The standard curve was generated by extracting DNA from five-fold serial dilutions of ascospores as described by Reich et al. (2017). The primers and probes for the tuna DNA were described by Li et al. (2015), while the primers and probes for *S. sclerotiorum* were developed by Reich et al. (2016). The *S. sclerotiorum* assay was specifically designed for use in environmental samples and reliably identifies *S. sclerotiorum* from closely related fungi such as *Botrytis cinerea* and *Sclerotinia trifoliorum*. The reaction conditions for the multiplex qPCR assay followed those described by Reich et al. (2017), and the specific primers, probes and reaction conditions given in Table S1. A standard curve was included for every batch of samples analysed, and all samples were run in technical triplicates.

2.9 | Data quality control

Preliminary data cleaning was performed for data involving ascospore counts from qPCR. Histograms of C_t values for the exogenous internal control were examined and generally followed a normal distribution. Samples for which the internal control did not amplify, or for which the C_t value was outside the normal distribution, were removed from subsequent analysis. For modelling purposes, for fields that had more than one Burkard sampler, ascospore counts were averaged to get a single value per day per field.

2.10 | Comparison of ascospore levels among provinces and months

To determine whether airborne ascospore levels varied by province, year within province or month within province-year, a series of one-way analyses of variance (ANOVAs) was performed. The logarithm of the daily number of ascospores plus 1 (log[no. ascospores + 1]) was used as the response variable. The ANOVAs were performed using the aov and post hoc means separation was performed using Tukey's HSD with TukeyHSD, both implemented in the 'stats' package (R Core Team, 2020).

2.11 | Comparison of samplers and sampler types

To evaluate the representativeness of each spore sampler in relation to the sampled field, Pearson's correlation coefficients were calculated on the logarithm of the daily number of ascospores plus one (log[no. ascospores + 1]) using the rcorr from the 'Hmisc' package (Harrell Jr, 2020) for fields where multiple samplers were present in the same field. The log transformation was applied to normalize the ascospore data and minimize the weight of the few large outliers (i.e., the days with huge amounts of ascospores) in analysis. When analysing means of ascospores, the means were calculated based on the log-transformed values, representing geometric means instead of arithmetic means. To compare the GRIPST-2009 rotation impaction sampler ascospore counts with the Burkard ascospore counts, all values were transformed to ascospores per cubic metre. The formula for calculating the volume of air sampled for the

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

GRIPST-2009 samplers was determined by multiplying the number of minutes each sampler was active by the sampling volume per minute (0.047 $m^3 min^{-1}$, according to the GRIPST-2009 user manual). Similarly, for samples collected from Burkard samplers (including those from the PGA network), sampling volumes were calculated as 0.99 $m^3 h^{-1}$ (or 16.5 L min^{-1} as per manufacturer specifications) of sampling. Total number of ascospores per day were then divided by the volume of air sampled per day to result in the number of ascospores per cubic metre per day.

2.12 | Predictor variables

Weather data was summarized at daily levels (midnight to midnight) to match the frequency and duration of ascospore collection, and because previous work suggests that spore release may be dependent on minor fluctuations throughout the day (Archer & Yuan, 2004; Reich et al., 2017). The theoretical framework of spore liberation and escape described above implies that, while spore production may be a result of days- or weeks-long environmental conditions (e.g., apothecia production), spore transportation is more likely dictated by short-term environmental conditions and changes. As a result, we focused our analysis on environmental conditions that were near the day of spore capture rather than long-term conditions, which would be more indicative of apothecia and ascospore production rather than ascospore release.

As there is little understanding or consensus on the biological and environmental drivers underlying ascospore release under field conditions (Reich & Chatterton, 2023), the environmental variables were summarized in several ways. First, one group of predictor variables comprised daily mean, minimum and maximum values of T, RH, VPD, Precip, DP, SoilWC and SoilT. To account for the potential influence of the previous day on spore release, another group of predictor variables comprised all variables mentioned previously at a time lag of 1 day. Also included in this group of predictor variables was the number of ascospores released the previous day (logSs_t1), as some airborne fungal pathogens (e.g., *B. cinerea*) have been shown to have high temporal correlations in spore release (Xu et al., 2000). Finally, to account for the possibility that spore release is caused by sudden or dramatic changes in environmental conditions, a final group of predictor variables was calculated as the differenced values between one day and the previous day for each variable (e.g., the difference of the maximum temperature on day t and the maximum temperature on day $t - 1$).

Several other predictor variables were derived, including a binary variable indicating whether it rained on the day of ascospore collection, and two variables that summarized the maximum change in RH over 2- or 3-h periods on the day of ascospore collection (Reich et al., 2017). Variables that exhibited no variation in the values were excluded from the analysis (e.g., the minimum precipitation on any given day was always 0 because these values were recorded every hour and no day had continuous precipitation). In total, 63 predictor variables were considered for model development.

2.13 | Data preparation for modelling

To examine linear associations between the number of ascospores and environmental conditions, Pearson correlations between log(no. ascospores + 1) and each environmental predictor were performed. The correlations revealed highly heterogeneous relationships between ascospores and environmental variables among provinces, discussed in more detail in the Results section (Table S2). This finding indicated that different provinces may require different models for ascospore prediction based on their unique field and climatic conditions. Consequently, the modelling was focused on the Alberta dataset due to the relatively small size and high number of missing values from Manitoba and Ontario.

2.14 | Variable (feature) selection for modelling

Variable selection methods can enhance model accuracy by addressing issues associated with multicollinearity and overfitting (Cai et al., 2018). Eight variable selection approaches were evaluated in the ML models to determine the most effective methods. The tested variable sets included all variables, variables selected from the LASSO algorithm, variables with a low (<10) variance inflation factor (VIF; algorithm provided by Beck, 2013), variables with low correlations ($r < |0.7|$) with all other variables, variables from same day, variables from previous day and variables that represented differenced values.

2.15 | Modelling approaches

Both ML and classical statistical methods were assessed in each of the regression and classification approaches (Hastie et al., 2008). In the regression approaches, the daily log(number of ascospores + 1) was used as the response variable. For the classification approaches, a range of arbitrary daily ascospore threshold cut-offs were modelled as the response, because there is currently little empirical evidence relating *S. sclerotiorum* airborne inoculum levels to disease expression (Reich & Chatterton, 2023). For the classification thresholds, a positive day was defined as a daily ascospore level greater than each threshold of 100, 200, 500 and 1000 ascospores per day and coded as a 1 and 0 otherwise. For model validation, a test dataset of one field per year ($n = 4$ field-years) was arbitrarily chosen to be removed from model development, because the purpose of these predictive models is to predict ascospore loads in new locations and years. Model training was therefore performed on the remaining field-years ($n = 17$). In total, the training dataset had 855 data points and the test dataset had 191 data points. Because there were many data points with missing

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

soil temperature and moisture readings, they were not included in the model development but were included for correlation analyses.

2.15.1 | ML models

ML models were implemented using Python v. 3.8.5 (Python Software Foundation) and the Scikit-learn package (Pedregosa et al., 2011). Random forests (RF_reg) were utilized for regression approaches, while random forests (RF_class), logistic regression (LR) and support vector machines (SVM) were employed for classification approaches. A pipeline was constructed for model development, incorporating scaling (StandardScaler) and imputing (IterativeImputer) for numeric variables and onehot encoding (OneHotEncoder) and imputing (SimpleImputer) for categorical variables. This pipeline was executed for each iteration of the 5-fold cross-validation. Hyperparameter tuning for the ML algorithms was performed via 5-fold cross-validation for each model within the RandomizedSearchCV function for RF models and with the GridSearchCV function for LR and SVM models. For RF models, the hyperparameters considered were the number of trees per forest, the depth of the forest and (for classification only) whether the classes were balanced. LR hyperparameters included a regularization parameter (C) and whether the classes were balanced. SVM hyperparameters consisted of C (as for LR), the number of features used (gamma) and whether the classes were balanced. Each model was optimized to maximize the area under the receiver operator characteristic curve (AUC), which takes into account the proportion of true positive and false negative predictions.

2.15.2 | Statistical models

Statistical models were developed in R v. 4.0.3 (R Core Team, 2020). The set of variables selected by the VIF algorithm consistently performed best in ML models; therefore, this set of variables was employed for the statistical models. Generalized linear models (GLM) were used for both regression and classification approaches. In regression, the log(number of ascospores + 1) was directly modelled with an identity link function. For classification, binary LR models with a logit link function were developed for each of the thresholds described above. Variable selection in the statistical methods was conducted using forward, backward and stepwise methods. Unlike the ML workflow, imputation was not included in the statistical model workflow, resulting in the removal of incomplete cases from the analysis, which accounted for approximately 10% of the data points.

2.16 | Evaluation of models

All models were developed on the training dataset and evaluated on the test dataset. Therefore, the diagnostics focused on evaluating the model performance only on the test set. For regression approaches, the models were evaluated with r^2 (calculated as the square of the Pearson correlation between actual and predicted values) and mean-squared error (MSE). For classification approaches, models were evaluated with accuracy, sensitivity (recall), specificity, precision, F1 score and AUC (DeVries et al., 2021). The classification metrics were calculated from a confusion matrix as follows:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

$$precision = \frac{TP}{TP + FP}$$

$$F_1\ score = 2 \times \frac{precision \times sensitivity}{precision + sensitivity}$$

where TP, number of true positives, TN, number of true negatives, FP, number of false positives and FN, number of false negatives.

3 | RESULTS

3.1 | Summary of ascospore levels

Ascospores were observed throughout the growing season in all provinces. The levels of ascospores varied among fields, provinces, months and years (Figure 1). In all field-years, ascospores were continuously released over the course of the season and levels of ascospores showed variation among fields (Figures S1, S2, S3). The highest recorded number of daily ascospores was 38,199 in Alberta on 18 August 2020. Mean daily ascospore levels differed significantly by province, with Manitoba having the greatest numbers (396), followed by Alberta (262) and Ontario (25) (values back-transformed because statistical analysis was performed on log-transformed variables; all pairwise comparisons significant at $p < 0.01$). Within each province, differences in average ascospore levels were also observed among most years sampled, except 2018 and 2021 in Alberta and 2019 and 2020 in Manitoba (all other pairwise comparisons significant at $p < 0.01$; Figure 1). In Alberta, mean ascospore levels varied by month in all years except 2019, with August consistently exhibiting the highest average ascospore levels. In Manitoba and Ontario, the results were more varied, with fewer significant differences among months and mean maximum ascospore levels occurring across months (Figure 1).

3.2 | Correlations of ascospores among samplers and sampler types

In 2018, each of the three commercial fields in Alberta was equipped with three Burkard samplers, and ascospore counts were assessed

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

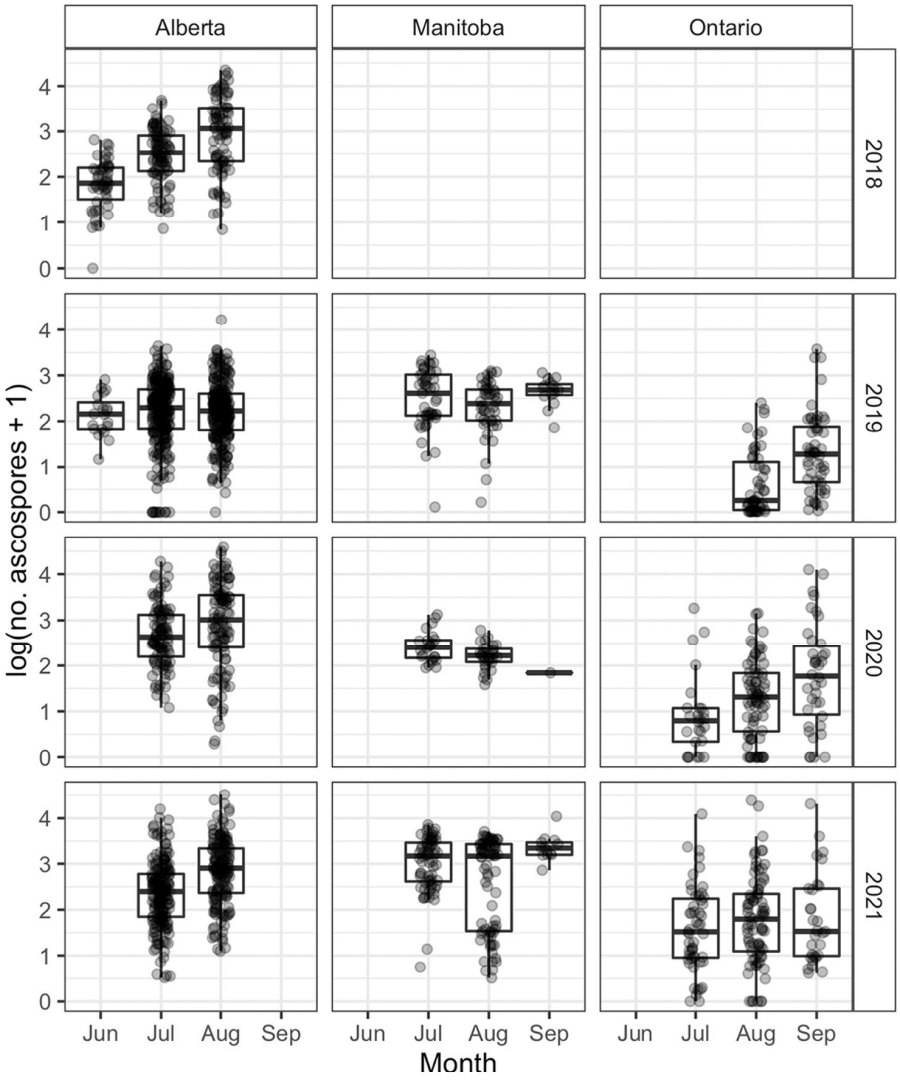

FIGURE 1 Boxplots of the log(no. ascospores + 1) by month for each province and year in the present study. The boxes represent the first and third quartile, and the horizontal line is the median. Vertical lines do not exceed 1.5×interquartile range; dots outside of these lines represent outliers.

TABLE 1 Summary of spore samplers placed side-by-side, one above the canopy and one below, in three commercial dry bean fields in Alberta in 2018.

	Above		Below	
Location	Mean	*SD*	Mean	*SD*
Cranford	39.7	4.5	369.5	9.7
Taber	39.9	2.7	22.1	3.8
Vauxhall	85.6	5.6	157.6	5.4

Note: Mean and *SD* calculations were performed on the log(number of ascospores per day + 1) (i.e., the geometric mean was calculated), and the data presented here have been back-transformed to represent number of ascospores per day.

for correlation among these samplers. Pearson correlations of log(ascospores + 1) among samplers within fields were all less than 0.50 except for one comparison with $r=0.81$ ($p<0.0001$). For other comparisons, the r values ranged from 0.15 to 0.46, although not all comparisons were statistically significant. Samplers placed beside each other, with one sampler above the canopy and another sampler below the canopy, did not have stronger correlations with each other compared to correlations with samplers 40 m away (data not shown). In 2018, in two of the fields surveyed, the samplers below the canopy collected, on geometric average, 2–10 times the number of ascospores as those above the canopy, but in the third field, the sampler below collected on geometric average, half of the ascospores as the sampler above the canopy (Table 1).

In 2019, no significant correlations were found between two Burkard samplers in the same commercial field approximately 40 m apart ($r=0.22$, $p=0.40$). Among the three Burkard samplers at LeRDC, all positioned within approximately 150 m of each other, two comparisons were statistically significant ($r=0.51$, $p<0.001$; $r=0.28$, $p=0.049$), while the other comparison was not ($r=-0.04$, $p=0.78$). In 2020 and 2021, correlations among samplers were similarly low or inconsistent for both commercial fields and LeRDC sampler datasets (data not shown).

In the research plot where a Rotorod and Burkard sampler were positioned beside each other, no correlation was found between ascospores per cubic metre per day ($r=0.02$, $p=0.9$). The correlations between ascospore counts from samples collected by the PGA network and any of the nearest dry bean fields 2–30 km away were similarly low ($r=-0.10$ to 0.07). On average, samples from the PGA

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

network contained 1/50th the amount of ascospores compared to samplers from nearby bean fields.

3.3 | Correlations of ascospores with environmental variables

A detailed analysis of correlations at a fine scale, by year and province, is presented in Table S2. Across all years and provinces, the log(number of ascospores + 1) was positively correlated with the number of ascospores from the previous day (range of Pearson's *r*: 0.15–0.6, depending on the year and provinces; $p < 0.001$ for most provinces and years). However, the direction and significance of correlations differed for other variables among provinces. Within each province, the direction of correlation (positive or negative) was often consistent, but the magnitude and sometimes statistical significance varied by year. For example, in Alberta, daily mean, maximum and minimum RH were negatively associated with ascospore numbers in 2018, 2019 and 2021, but in 2020 these relationships were non-significant. In Manitoba, there was no association between measures of RH and daily ascospore levels, and in Ontario the direction was positive and significant in only 2020. Similar variability was observed with all environmental correlations.

When data were pooled across years and analysed by province, certain trends emerged. In Alberta (Table 2, Table S2), ascospores were positively associated with ascospores from the previous day; max and mean temperature on the same day and previous day; VPD variables on the same day and previous day; and the difference in RH from the previous day. Ascospores were negatively associated with RH variables on the same day and mean RH from the previous day; Precip on the same day and previous day; DP on the same day; and SoilWC variables on the same day and previous day. Daily ascospores were not correlated with differenced variables, which represent changes in the environment.

In Manitoba (Table 3, Table S2), SoilT data was unavailable for the duration of this study. Daily ascospore levels were positively correlated with ascospores from the previous day; SoilWC variables on the same day and previous day; Precip from the same day; and minVPD on the same day and previous day. Ascospores were negatively correlated with differences in RH on the same day and from the previous day. Ascospores were not correlated with T, RH or differenced variables.

In Ontario (Table 4, Table S2), daily ascospore counts were positively correlated with ascospores from the previous day and minVPD, although the statistical significance of the latter was marginal ($p = 0.047$, unadjusted *p*-value). Ascospores were negatively correlated with DP variables on the same day and previous day; Precip on the same day; SoilT variables on the same day and previous day; and maxRH and maxVPD on the same day and previous day. Ascospores were not correlated with T, RH, or differenced variables.

TABLE 2 Top 10 Pearson correlations of daily log(number of ascospores + 1) with daily environmental variables collected by in-field weather stations for combined data for 2018–2021 in Alberta.

Variable	*r*	*p*
logSs_1d	0.361	<0.001
MeanVPD	0.120	<0.001
MeanRH	−0.118	<0.001
MaxVPD_1d	0.118	<0.001
MaxT_1d	0.116	<0.001
MeanVPD_1d	0.115	<0.001
MaxVPD	0.101	<0.001
MinRH	−0.100	<0.001
MinDP	−0.098	<0.001
MinRH_1d	−0.098	<0.001

Note: logSs_t1, the log(number of ascospores + 1) from the previous day; VPD, vapour pressure deficit; RH, relative humidity; T, air temperature. Variables with a '_1d' represent the value from the previous day. For the complete list of correlations, refer to Table S2.

3.4 | ML models

When comparing variable selection methods, it was found that selecting variables based on the VIF algorithm outperformed other methods in seven of the 12 scenarios (3 ML methods × 4 ascospore thresholds; Table S3). Therefore, only the results obtained using the VIF method are presented and discussed. However, it is worth noting that variable selection methods using LASSO or variables with low correlations yielded very similar outcomes to the VIF method. These three methods vastly outperformed the other variable datasets, including the dataset with all predictor variables (Table S3).

For RF_Reg model, the R^2 value of the predicted values on the test set was 0.29 (Table 5). The most important variable in this algorithm was logSs_t1, which was over 25 times more important than the next most important variable in this model (Figure 2). All other variables included in the RF_Reg model comprised comparatively little (<15%) of importance.

For classification approaches, RF_Class, LR and SVM models had similar AUC scores across thresholds (Table 6). However, they varied more in their F1 scores, which is a combination of sensitivity and precision (Table 6). At thresholds of 100, 200 and 500 ascospores per day, SVM had higher sensitivity and lower precision than RF_Class and LR, but these trends were reversed at a threshold of 1000 ascospores per day. For specificity, all models differed greatly at thresholds of 100 or 200 ascospores per day but were very similar for thresholds of 500 and 1000 ascospores per day. In all models, the most important variable was logSs_t1, the log-number of ascospores from the day before, and, as with regression approaches, this variable was dramatically more important (6–8 times more important) than the next most important feature (Figure 2).

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Top 10 Pearson correlations of daily log(number of ascospores + 1) with daily environmental variables collected by in-field weather stations for combined data in 2019–2021 in Manitoba.

Variable	r	p
logSs_1d	0.632	<0.001
MinWC_1d	0.224	0.001
MeanWC_1d	0.204	0.004
MaxWC_1d	0.175	0.013
MinWC	0.171	0.014
DiffRH_0d	−0.155	0.004
MeanWC	0.153	0.027
DiffRH_1d	−0.142	0.008
MinVPD_1d	0.130	0.014
MaxWC	0.127	0.067

Note: Variables are as for Table 1, with the addition of WC, soil moisture (water content). Variables that have a 'Diff' preceding them represent differenced values (e.g., 'DiffRH_0d' represents the difference between min and max RH of the same day whereas 'DiffRH_1d' represents the difference between the min and max RH of the previous day). For the complete list of correlations, refer to Table S2.

TABLE 4 Top 10 Pearson correlations of daily log(number of ascospores + 1) with daily environmental variables collected by in-field weather stations for combined data in 2019–2021 in Ontario.

Variable	r	p
logSs_1d	0.304	<0.001
MinDP_1d	−0.231	<0.001
MeanDP_1d	−0.194	0.001
MinSoilT	−0.181	0.001
MinDP	−0.161	0.007
MeanSoilT	−0.158	0.003
MeanDP	−0.147	0.013
MaxSoilT	−0.131	0.013
MaxDP_1d	−0.130	0.030
MinSoilT_1d	−0.130	0.014

Note: Variables are as for Tables 1 and 2, with the addition of SoilT, soil temperature; DP, dew point. For the complete list of correlations, refer to Table S2.

3.5 | Statistical models

For statistical model development, the variables selected based on the low VIF criterion were included as predictors. As a result, these variables had low collinearity. In all cases, the forward and stepwise model selection procedures converged on the same model. In general, these models tended to be more parsimonious than, and had similar deviance and AIC as, those determined by backward model selection procedures (Tables 7, 8 and 9). For regression approaches, models were highly similar for all diagnostics, with R^2 values of 0.29 and MSE of 0.5 (Table 5). The variable with the most influence in the model was the logSs_t1, with approximately 20 times more weight in the model than the next important variable (Table 7).

For classification approaches, as with the linear regression, the forward and stepwise models tended to be more parsimonious than the backward selection models (Table 8). Within each spore threshold, all models had similar deviance and AIC values (Table 9). The variables included in each model varied by threshold and selection procedure, but logSs_t1 was the only variable included in all models (Table 8). The relative weight of this variable ranged from approximately 10–40 times that of the next important variables in the model.

The AUC values for all thresholds and models showed minimal variation (range from 0.74 to 0.78), but the F1 scores did vary substantially. At thresholds of 100 or 200 ascospores per day, the F1 values ranged from 0.69 to 0.74. However, at thresholds of 500 or 1000 ascospores per day, the F1 value ranged from 0.41 to 0.60, depending on the model and threshold (Table 9). These differences in F1 scores were primarily the result of greater sensitivities at lower thresholds, because precision remained relatively constant (about 0.62) across all models and datasets. However, specificity (the proportion of true predicted negatives) declined at these higher thresholds, which accounts for the relatively stable AIC value across all threshold values.

3.6 | Comparison of ML and statistical approaches

In terms of regression models, both ML and statistical models had similar R^2 (about 0.29) and reductions in MSE (reduced by 0.23). It is noteworthy that all models identified logSs_t1 as the most important predictor. However, the GLM determined by forward and stepwise methods was the most parsimonious of all models, with TotalPrecip, MaxVPD_1d and DiffMaxVPD as the three other significant variables.

For classification models, both ML and statistical approaches demonstrated comparable performance in diagnostics like AUC and F1 scores. However, statistical approaches exhibited greater sensitivity (a measure of true positives) and lower specificity (a measure of true negatives) than ML approaches, particularly at the lower thresholds of 100 and 200 ascospores per day (Tables 6 and 9). Consequently, a trade-off may exist between these approaches depending on the prioritization of accurately predicting true positives or true negatives.

4 | DISCUSSION

This research was designed to assess the utility of spore samplers for collecting ascospores of *S. sclerotiorum* and to predict the occurrence of ascospores using environmental data and a range of statistical and ML approaches. Regarding the first objective, our findings revealed inconsistent and relatively low correlations of ascospores among samplers. Similarly, the correlation between ascospores collected from the Burkard sampler and the Rotorod sampler was also very low. However, it is important to note that

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Evaluation metrics for the generalized linear model (GLM) and machine learning (ML, random forest) regression models.

	Deviance	AIC	R^2	MSE
ML (random forest)				
Null	–	–	−0.063	0.692
VIF-variable set	–	–	0.285	0.466
GLM				
Null	465.44	1784.1	0	0.742
Backward	321.23	1519.4	0.294	0.504
Forward/stepwise	325.24	1520.8	0.294	0.503

Note: Deviance and Akaike information criterion (AIC) are those for the training dataset in the GLM development and R^2 and mean-squared error (MSE) are diagnostics on the test dataset.

even Burkard samplers placed right beside each other had low correlations, which could be due to the patchy distribution of *S. sclerotiorum* in fields (Jones et al., 2011; Wutzki et al., 2019) or an inherent limitation of spore sampling technology (Jackson & Bayliss, 2011). A persistent challenge in spore sampling projects is determining how representative a single air sample is, because spore transport and collection depend on a complex interaction of factors such as fungal biology, plant architecture and spacing, wind direction and speed, and sampler location and height (Jackson & Bayliss, 2011; Mahaffee et al., 2023; Van der Heyden et al., 2021). While we are not aware of many studies that compare nearby samplers of the same type (as in our subset of Burkard samplers here), several studies have compared the collection efficiencies of different Rotorod and Burkard samplers for pollen and fungal spores. In studies comparing pollen concentrations Rotorod and Burkard samplers were highly correlated with each other ($\rho = 0.92$) (Crisp et al., 2013; Peel et al., 2014), although the Burkard samplers collected many more pollen grains than the Rotorod. In a study investigating relative collection efficiency for *Venturia inaequalis* spores, the Rotorod collected 37% of the number of spores that the Burkard sampler collected, but no correlations were computed (Aylor, 1993).

Regarding the modelling objectives of this study, when ascospores were modelled using statistical and ML algorithms, the most influential factor associated with daily ascospore counts across all provinces, years and models was consistently the number of ascospores released during the previous day. The inclusion of environmental variables in the models contributed comparatively little to the model predictions (Tables 7 and 8, Figure 2) and models excluding ascospores from the previous day had much lower predictive ability. These findings suggest that for an ascospore prediction network to be established, continual quantification of airborne ascospores would be necessary to enable reliable predictions into the near future. As ascospore release patterns exhibited multiple peaks and valleys throughout the growing season, continual monitoring of ascospores would also be crucial for detecting sudden, high ascospore days that might arise from random processes. While this continual monitoring may seem like an insurmountable challenge, significant progress has been made in the development of automated spore samplers that can quantify airborne fungal spore species (de la Pasture, 2018), and so the implementation of this type of monitoring—if useful—may be quite feasible. Interestingly, the strong temporal autocorrelation of ascospore quantities was observed in all years and provinces, indicating that while microclimatic factors influencing ascospore liberation may vary by location (discussed below), the robust autocorrelation of daily ascospore quantities could be inherent to the biology of this fungus.

Several ML and statistical modelling approaches were employed to predict ascospore levels in southern Alberta, and their performances were all similar. Their similarities could be attributed to the dominance of a single non-environmental variable (logSs_t1) in all models, which may indicate that ascospore release and transportation in *S. sclerotiorum* is less driven by conditions like mean, minimum or maximum T and RH than previously thought. In light of the theoretical framework of spore liberation and dispersal described in the introduction, these findings may not be that surprising, because that framework focuses primarily on air turbulence and canopy characteristics and not other environmental factors directly. Empirical evidence on the relationships between ascospore levels and environment is scant and often contradictory. In witloof chicory, ascospores were not correlated with precipitation and most T variables from the day of collection to 7 days before, but were positively correlated with RH variables for all those periods and minimum air T on the same day and day before sampling (Leyronas et al., 2019). In a study of irrigated soybean fields, daily ascospores were not significantly correlated with any measured daily environmental variables, including air T, RH and soil T (Fall et al., 2018). In carrot, ascospore counts were uncorrelated with air T, RH or soil T values from 1 to 3 weeks before sampling but were positively correlated with SoilWC values over all these periods (Foster et al., 2011). In pastures, ascospore counts were uncorrelated with air T, solar radiation and windspeed (Bourdôt et al., 2001). Similarly, studies with a more descriptive approach have concluded that "ascospore concentrations in the air within the crop appeared to be related more to the numbers of apothecia present than weather variables" (Hartill, 1980; McCartney & Lacey, 1991). Comparatively few studies have developed models aimed at predicting ascospore levels, and none, to our knowledge, have examined ascospore temporal autocorrelation as a possibility. Perhaps, then, while apothecia and ascospore production tends to be highly driven by moisture and temperature variables, ascospore liberation and transportation is less driven by ambient

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

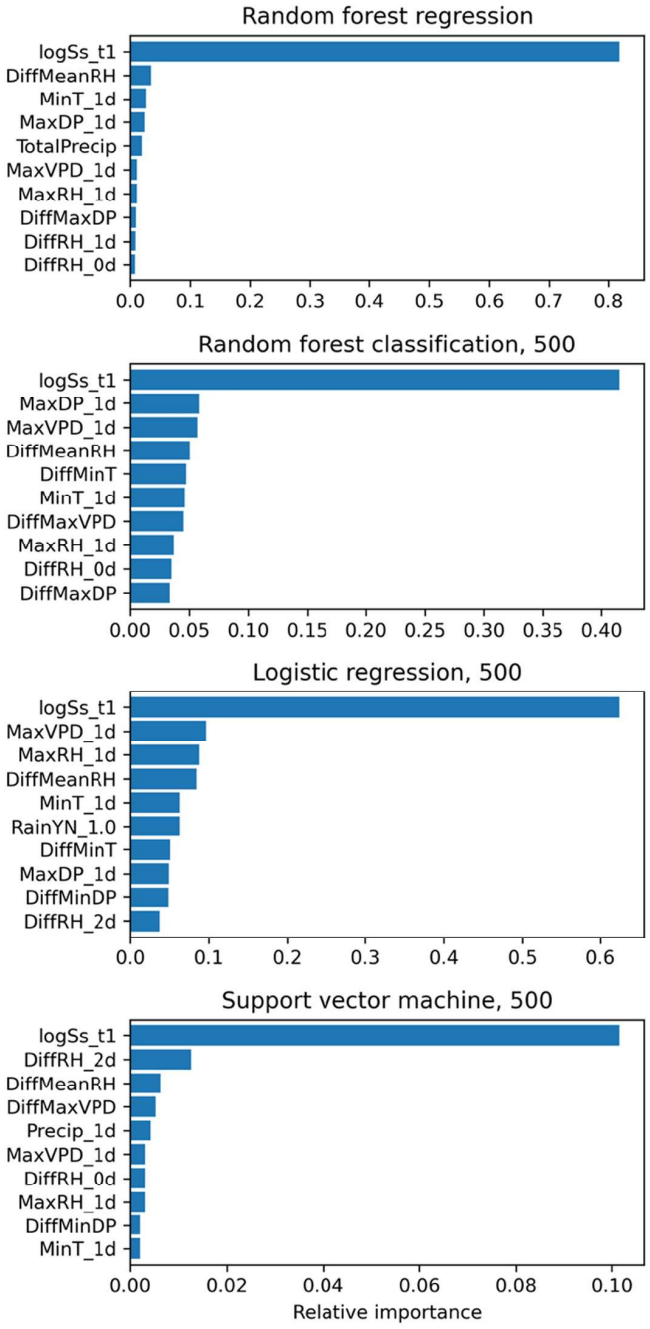

FIGURE 2 Importance of the top 10 variables included in the machine learning models implemented in this study. The top panel represents variable importances in the RF_Ref model, and the bottom three panels represent importances in the classification approaches (RF_Class, LR and SVM) for only the threshold of 500 ascospores per day. Results for all thresholds were qualitatively similar. [Colour figure can be viewed at wileyonlinelibrary.com]

environmental conditions and instead driven by biological mechanisms and complex air turbulence patterns.

During the modelling process, variable selection emerged as a crucial step; models incorporating all possible variables consistently performed poorly on the test dataset, probably due to overfitting, compared to models using only a subset of variables. From a practical standpoint, the statistical models are easier to interpret and implement compared to the ML models. The coefficients from the statistical models have a straightforward interpretation, and their presence (or absence) from the model has a clear statistical basis. Furthermore, the equations presented here can be easily implemented by anyone. In contrast, the ranking of the importance of variables in ML models, especially the RF and SVM approaches, have a more opaque interpretation (e.g., decrease in impurity). Additionally, for these models to be adopted by others, they need to be saved and made accessible to download, which may reduce accessibility.

The variability in the direction, magnitude and significance of correlations of ascospore levels with environmental conditions observed across different years and provinces suggests that *S. sclerotiorum* encounters distinct limitations in various regions. In Alberta, ascospores exhibited a positive correlation with factors related to dryness (e.g., VPD) and negatively correlated with factors related to moisture (e.g., RH and Precip). In contrast, in Manitoba, ascospores were positively associated with moisture (e.g., SoilWC and constant RH) and in Ontario, ascospores were negatively associated with higher temperatures (e.g., SoilT and DP). These contrasting patterns, especially between Alberta versus Manitoba and Ontario, may be caused by the presence of irrigation in all commercial dry bean fields in southern Alberta compared to dryland fields in Manitoba and Ontario. To corroborate this hypothesis, previous work has found that, for apothecia formation, irrigation modifies the microclimate in such a way that associations with other variables (e.g., air temperature) can be reversed between irrigated and non-irrigated plots (Willbur et al., 2018).

Notably, ascospores showed little correlation with differenced variables, which aimed to assess the relationship between ascospore counts and environmental changes. However, differenced variables often emerged as preferred predictors in variable selection methods like VIF, as they exhibited low correlation with other variables and frequently appeared as significant predictors in the final models. This trend corroborates previous research of ascospore release in seed alfalfa fields in the same region (Reich et al., 2017). For the same threshold of 100 ascospores per day, the models developed in this study demonstrated greater AUC (about 0.70) than that reported for ascospore release in seed alfalfa (AUC about 0.60). However, the earlier study did not include logSs_t1 as a potential predictor variable, highlighting the improvement achieved by incorporating this variable in the current models.

In a previous study, we found that including the number of ascospores collected in a dry bean field did not improve predictions of white mould levels in that field (Reich et al., 2023). Combined with results from this study, these findings indicate that sampling for and/or predicting airborne ascospores are unlikely to provide practical insights into the management of white mould in dry bean fields, at least in irrigated conditions. Even if airborne ascospore concentration could be measured or predicted perfectly, there is scant empirical evidence describing how ascospore levels relate to disease intensity. In one study in canola that described this relationship, the same level of airborne ascospores resulted in dramatically different levels of disease incidence depending on location and year (presumably, both indirect measures of differences in environment), and so the inoculum–disease relationship is not a straightforward one (Qandah & Del Rio Mendoza, 2012).

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Model diagnostics for 3 machine learning classification algorithms (RF = random forest, LR = logistic regression, SVM = support vector machines) and four daily threshold values (100, 200, 500 and 1000) of ascospores.

	Accuracy	Sensitivity	Specificity	Precision	F1	AUC
y100						
Null	0.613	1.000	0.000	0.613	0.760	0.500
RF	0.649	0.607	0.716	0.772	0.679	0.703
LR	0.665	0.607	0.757	0.798	0.689	0.729
SVM	0.607	0.940	0.081	0.618	0.746	0.725
y200						
Null	0.455	1.000	0.000	0.455	0.626	0.500
RF	0.675	0.632	0.712	0.647	0.640	0.761
LR	0.723	0.644	0.788	0.718	0.679	0.765
SVM	0.644	0.793	0.519	0.580	0.670	0.769
y500						
Null	0.712	0.000	1.000	NaN[a]	NaN	0.500
RF	0.754	0.436	0.882	0.600	0.505	0.772
LR	0.749	0.382	0.897	0.600	0.467	0.766
SVM	0.754	0.473	0.868	0.591	0.525	0.759
y1000						
Null	0.775	0.000	1.000	NaN	NaN	0.500
RF	0.796	0.535	0.872	0.548	0.541	0.740
LR	0.796	0.093	1.000	1.000	0.170	0.755
SVM	0.801	0.233	0.966	0.667	0.345	0.750

Note: These diagnostics are for models developed with the VIF-selected variables.

[a]NaN, not possible to calculate due to division by zero.

Many of the sampling issues encountered in this study are not inherent to this pathosystem but are, rather, characteristic of aerobiological studies, and some of these issues are described by Jackson and Bayliss (2011). In the context of the present study, further investigation is needed to determine the spatial scales at which a single sampler can accurately reflect the field environment. For instance, in the case of strawberry, one sampler per field was deemed enough to collect a representative number of *Podosphaera aphanis* spores (Van der Heyden et al., 2014). However, for a pathogen like *S. sclerotiorum*, which typically has a patchy distribution (Boland & Hall, 1988a, 1988b) and whose spores often have a sharp dispersal gradient (Bourdôt et al., 2001), collection of ascospores may rely on chance—placing the sampler in the right area of the field where apothecia emerge. In this study, correlations among samplers placed within the same field were consistently low each year. Furthermore, the appropriate height above ground for a sampler remains an unknown. In wheat, *Blumeria graminis* spores were found at much greater levels within the canopy (0.6 m above ground) than above the canopy (1.2 m) (Cao et al., 2012). In the present study, more ascospores were collected below the canopy in two of three fields where samplers were placed both above and below the canopy, but in the third field the effect was reversed (Table 1). Canopy density could be a contributing factor, although it was not measured in this study. Finally, the Burkard samplers collect about 1 m^3 of air per hour, which may not be a large enough sample to reliably infer anything about real dynamics of spore transport.

While the models presented here performed better than previously published models in similar agricultural contexts (Reich et al., 2017), future attempts to predict ascospore release may be improved by several additional avenues of research. A finer temporal scale analysis, considering hourly ascospore levels and hourly environmental variables, as shown in a non-peer-reviewed article (Archer & Yuan, 2004), may reveal important associations. Measurement of additional environmental variables such as wind could also be explored, although, as observed with other variables, wind has shown relatively poor associations with *S. sclerotiorum* life stages (Reich & Chatterton, 2023). However, investigating gust events, canopy porosity and ambient wind speeds could help elucidate airborne ascospore dynamics. Finally, a more focused attempt to link apothecia emergence with ascospore levels could be pursued. Given the challenges of scouting for apothecia in field crops, this approach could involve monitoring sclerotia deposits under various environmental conditions while simultaneously collecting airborne ascospores nearby.

Numerous attempts have been made to predict diseases caused by *S. sclerotiorum* in field crops, but the literature on predicting ascospore release remains scant. This study contributes to knowledge about airborne ascospore levels by employing numerous modelling approaches to assess the relative importance of environmental variables on airborne spore concentrations. Across years and provinces,

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 7 Coefficients for variables included the best regression models as determined by backward, forward and stepwise procedures using generalized linear models in R.

	Intercept	logSs_t1	DiffMinT	DiffMeanRH	MaxRH_1d	DiffRH_1d	TotalPrecip	DiffMaxDP	MaxDP_1d	MaxVPD_1d	DiffMaxVPD
Backward	1.386	0.527	−0.018	−0.008	−0.007	0.005	−0.013	0.025	0.016		
Forward/step	1.083	0.523					−0.013			0.010	0.008

Note: The predicted variable was the daily log(no. ascospores + 1). Model diagnostics are given in Table 5.

TABLE 8 Coefficients for variables included the best logistic regression models as determined by backward, forward and stepwise procedures in R.

Threshold	Intercept	logSs_t1	DiffMinT	DiffMaxVPD	MaxVPD_1d	TotalPrecip	Precip_1d	DiffMaxDP	MinT_1d	DiffMeanRH	MaxRH_1d	MaxDP_1d	DiffRH_1d	RainYN1
y100														
Backward	−2.616	1.205	−0.059	0.045	0.067	−0.037	0.019	0.047						
Forward/step	−2.357	1.165		0.044	0.060	−0.030								
y200														
Backward	−1.105	1.148	−0.095		0.041	−0.023		0.080	−0.083	−0.021	−0.019	0.048		
Forward/step	−2.998	1.166	−0.061		0.047	−0.023		0.070		−0.017				
y500														
Backward	−3.035	1.536								−0.021	−0.019		0.011	−0.342
Forward/step	−4.584	1.536			0.024					−0.018				−0.335
y1000														
Backward	−3.175	1.594							−0.056	−0.018	−0.017			−0.401
Forward/step	−5.212	1.603								−0.012				−0.450

Note: Successes were defined as 1 = the number of ascospores per day were greater than the shown threshold (100, 200, 500, or 1000 ascospores per day) and 0 otherwise. Model diagnostics are given in Table 9.

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 9 Model diagnostics for three model selection methods (backward, forward and stepwise) and four daily threshold values of ascospores.

	Deviance	AIC	Accuracy	Sensitivity	Specificity	Precision	F1	AUC
y100								
Null	816.2	818.2	0.602	1.000	0.000	0.602	0.752	0.500
Backward	690.6	706.6	0.626	0.903	0.206	0.633	0.744	0.736
Forward/step	696.2	706.2	0.602	0.903	0.147	0.616	0.732	0.743
y200								
Null	997.6	999.6	0.480	1.000	0.000	0.480	0.648	0.500
Backward	857.5	877.5	0.673	0.768	0.584	0.630	0.692	0.785
Forward/step	861.8	875.8	0.667	0.780	0.562	0.621	0.692	0.782
y500								
Null	1027.3	1029.3	0.690	0.000	1.000	NA[a]	NA	0.500
Backward	838.3	850.3	0.754	0.547	0.847	0.617	0.580	0.777
Forward/step	839.2	849.2	0.766	0.566	0.856	0.638	0.600	0.782
y1000								
Null	863.8	865.8	0.754	0.000	1.000	NA	NA	0.500
Backward	702.3	714.3	0.778	0.310	0.930	0.591	0.406	0.770
Forward/step	705.9	713.9	0.795	0.333	0.946	0.667	0.444	0.776

Note: Models were developed using generalized linear models and a logit-link, where successes were defined as 1 = the number of ascospores per day were greater than the shown threshold (100, 200, 500, or 1000 ascospores per day) and 0 otherwise. Deviance and Akaike information criterion (AIC) are those for the training set and other columns represent diagnostics on the test set. Refer to Table 8 for model specifications.

[a]NA, not possible to calculate due to division by zero.

and in all models developed, it was clear that ascospore release was strongly temporally autocorrelated and the environmental variables we considered contributed relatively little to ascospore levels. However, further refinement is necessary before these models can be implemented as a real-time disease risk assessment tool for growers.

ACKNOWLEDGEMENTS

We thank students and technicians Natasha Tetzlaff, Uriel Karerwa, Jason Patey, Brooke Groenenboom, Anthony Erickson, Eric Amundsen, Christine Vucurevich and Hesther Yueh for their enormous help in data collection, laboratory work and machine maintenance. We also acknowledge the help of collaborating technicians Tom Henderson, Waldo Penner and Ashley Wragg. Maintenance of spore samplers and collection of samples were performed in part by employees at ProMax Agronomy and Viterra. We also are grateful for the guidance of Dr Michael Gelbart in the development of the ML algorithms. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada and funding provided by the Canadian Agriculture Partnership Pulse Cluster program, through Agriculture and Agri-food Canada; the Alberta Pulse Growers; Manitoba Pulse and Soybean Growers; and the Ontario Bean Growers. Open Access funding provided by the Gouvernement du Canada Agriculture et Agroalimentaire Canada library.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Jonathan Reich https://orcid.org/0000-0001-8469-9733

REFERENCES

Aboukhaddour, R., Fetch, T., McCallum, B.D., Harding, M.W., Beres, B.L. & Graf, R.J. (2020) Wheat diseases on the prairies: a Canadian story. *Plant Pathology*, 69, 418–432.

Araujo, G.T., Amundsen, E., Frick, M., Gaudet, D.A., Aboukhaddour, R., Selinger, B. et al. (2021) Detection and quantification of airborne spores from six important wheat fungal pathogens in southern Alberta. *Canadian Journal of Plant Pathology*, 43, 439–454.

Archer, C. & Yuan, Z.Q. (2004) *Modelling of spore release and aternative methods of control for stem rot (Sclerotinia sclerotiorum) in beans*. Sydney, N.S.W: Horticulture Australia.

Aylor, D.E. (1990) The role of intermittent wind in the dispersal of fungal pathogens. *Annual Review of Phytopathology*, 28, 73–92.

Aylor, D.E. (1993) Relative collection efficiency of Rotorod and Burkard spore samplers for airborne *Venturia inaequalis* ascospores. *Phytopathology*, 83, 1116–1119.

Aylor, D.E. (1999) Biophysical scaling and the passive dispersal of fungus spores: relationship to integrated pest management strategies. *Agricultural and Forest Meteorology*, 97, 275–292.

Beck, M.W. (2013) *Collinearity and stepwise VIF selection*. https://beckmw.wordpress.com/2013/02/05/collinearity-and-stepwise-vif-selection/ [Accessed 28th February 2024]

Boland, G.J. & Hall, R. (1987) Epidemiology of white mold of white bean in Ontario. *Canadian Journal of Plant Pathology*, 9, 218–224.

Boland, G.J. & Hall, R. (1988a) Numbers and distribution of apothecia of *Sclerotinia sclerotiorum* in relation to white mold of white bean (*Phaseolus vulgaris*). *Canadian Journal of Botany*, 66, 247–252.

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Boland, G.J. & Hall, R. (1988b) Relationships between the spatial pattern and number of apothecia of *Sclerotinia sclerotiorum* and stem rot of soybean. *Plant Pathology*, 37, 329–336.

Bolton, M.D., Thomma, B.P.H.J. & Nelson, B.D. (2006) *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. *Molecular Plant Pathology*, 7, 1–16.

Bourdôt, G.W., Hurrell, G.A., Saville, D.J. & DeJong, D.M.D. (2001) Risk analysis of *Sclerotinia sclerotiorum* for biological control of *Cirsium arvense* in pasture: ascospore dispersal. *Biocontrol Science and Technology*, 11, 119–139.

Cai, J., Luo, J., Wang, S. & Yang, S. (2018) Feature selection in machine learning: a new perspective. *Neurocomputing*, 300, 70–79.

Cao, X., Duan, X., Zhou, Y. & Luo, Y. (2012) Dynamics in concentrations of *Blumeria graminis* f. sp *tritici* conidia and its relationship to local weather conditions and disease index in wheat. *European Journal of Plant Pathology*, 132, 525–535.

Carisse, O., Tremblay, D.M., Lévesque, C.A., Gindro, K., Ward, P. & Houde, A. (2009) Development of a TaqMan real-time PCR assay for quantification of airborne conidia of *Botrytis squamosa* and management of Botrytis leaf blight of onion. *Phytopathology*, 99, 1273–1280.

Clarkson, J.P., Phelps, K., Whipps, J.M., Young, C.S., Smith, J.A. & Watling, M. (2007) Forecasting Sclerotinia disease on lettuce: a predictive model for carpogenic germination of *Sclerotinia sclerotiorum* sclerotia. *Phytopathology*, 97, 621–631.

Clarkson, J.P., Staveley, J., Phelps, K., Young, C.S. & Whipps, J.M. (2003) Ascospore release and survival in *Sclerotinia sclerotiorum*. *Mycological Research*, 107, 213–222.

Crisp, H.C., Gomez, R.A., White, K.M. & Quinn, J.M. (2013) A side-by-side comparison of Rotorod and Burkard pollen and spore collections. *Annals of Allergy, Asthma & Immunology*, 111, 118–125.

de Jong, M.D., Bourdôt, G.W., Powell, J. & Goudriaan, J. (2002) A model of the escape of *Sclerotinia sclerotiorum* ascospores from pasture. *Ecological Modelling*, 150, 83–105.

de la Pasture, L. (2018) Theory to field – sentries set to watch over crops. *Crop Production Magazine*, 20, 42–45.

de Oliveira Aparecido, L.E., de Souza, R.G., da Silva Cabral De Moraes, J.R., Cts, C. & de Souza, P.S. (2020) Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *International Journal of Biometeorology*, 64, 671–688.

DeVries, Z., Locke, E., Hoda, M., Moravek, D., Phan, K., Stratton, A. et al. (2021) Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability. *The Spine Journal*, 21, 1135–1142.

Fall, M.L., Willbur, J.F., Smith, D.L., Byrne, A.M. & Chilvers, M.I. (2018) Spatiotemporal distribution pattern of *Sclerotinia sclerotiorum* apothecia is modulated by canopy closure and soil temperature in an irrigated soybean field. *Plant Disease*, 102, 1794–1802.

Ferencova, Z., Rico, V.J. & Hawksworth, D.L. (2017) Extraction of DNA from lichen-forming and lichenicolous fungi: a low-cost fast protocol using Chelex. *The Lichenologist*, 49, 521–525.

Foster, A.J., Kora, C., McDonald, M.R. & Boland, G.J. (2011) Development and validation of a disease forecast model for Sclerotinia rot of carrot. *Canadian Journal of Plant Pathology*, 33, 187–201.

Hamer, W.B., Birr, T., Verreet, J.A., Duttmann, R. & Klink, H. (2020) Spatio-temporal prediction of the epidemic spread of dangerous pathogens using machine learning methods. *ISPRS International Journal of Geo-Information*, 9, 44.

Hao, J.J., Subbarao, K.V. & Duniway, J.M. (2003) Germination of *Sclerotinia minor* and *S. sclerotiorum* sclerotia under various soil moisture and temperature combinations. *Phytopathology*, 93, 443–450.

Harding, M.W., Daniels, G.C., Burke, D.A., Pugh, C.A. & Nielson, J.M. (2017) White mould on dry bean in Alberta in 2016. *Canadian Plant Disease Survey*, 97, 185–187.

Harrell Jr FE, 2020. *Hmisc: Harrell Miscellaneous*. Available from: https://cran.r-project.org/web/packages/Hmisc/index.html [Accessed 28th February 2024]

Hartill, W.F.T. (1980) Aerobiology of *Sclerotinia sclerotiorum* and *Botrytis cinerea* spores in New Zealand tobacco crops. *New Zealand Journal of Agricultural Research*, 23, 259–262.

Hastie, T., Tibshirani, R. & Friedman, J. (2008) *The elements of statistical learning*. New York: Springer.

Jackson, S.L. & Bayliss, K.L. (2011) Spore traps need improvement to fulfil plant biosecurity requirements: spore traps for biosecurity. *Plant Pathology*, 60, 801–810.

Jędryczka, M., Brachaczek, A., Kaczmarek, J., Dawidziuk, A., Kasprzyk, I., Maczyńska, A. et al. (2013) System for forecasting disease epidemics (SPEC)–decision support system in polish agriculture, based on aerobiology. *Alergologia Immunologia*, 9, 89–91.

Jędryczka, M., Kaczmarek, J., Dawidziuk, A. & Brachaczek, A. (2008) System for forecasting disease epidemics—aerobiological methods in polish agriculture. *Aspects of Applied Biology*, 89, 1–6.

Jones, S.J., Gent, D.H., Pethybridge, S.J. & Hay, F.S. (2011) Spatial characteristics of white mould epidemics and the development of sequential sampling plans in Australian bean fields: spatial characteristics of white mould epidemics. *Plant Pathology*, 60, 1169–1182.

Kim, Y.M., McClaren, D.L., Conner, R.L., Penner, W.C. & Kerley, T.J. (2019) Diseases of field bean in Manitoba in 2017. *Canadian Plant Disease Survey*, 99, 168–170.

Kim, Y.M., McLaren, D.L., Conner, R.L., Penner, W.C. & Kerley, T.J. (2017) Disease of field bean in Manitoba in 2016. *Canadian Plant Disease Survey*, 97, 188–190.

Leyronas, C., Benigni, M., Leignez, S., Duffaud, M., Villeneuve, F. & Nicot, P.C. (2019) Characterization of *Sclerotinia sclerotiorum* airborne inoculum, the widespread agent of white mould disease. *Aerobiologia*, 35, 243–252.

Li, M., Ishiguro, Y., Kageyama, K. & Zhu, Z. (2015) A simple method for normalization of DNA extraction to improve the quantitative detection of soil-borne plant pathogenic oomycetes by real-time PCR. *Letters in Applied Microbiology*, 61, 179–185.

Mahaffee, W.F., Margairaz, F., Ulmer, L.D., Bailey, B.N. & Stoll, R. (2023) Catching spores: linking epidemiology, pathogen biology, and physics to ground-based airborne inoculum monitoring. *Plant Disease*, 107, 13–33.

McCartney, H.A. & Lacey, M.E. (1991) The relationship between the release of ascospores of *Sclerotinia sclerotiorum*, infection and disease in sunflower plots in the United Kingdom. *Grana*, 30, 486–492.

Mila, A.L., Carriquiry, A.L. & Yang, X.B. (2004) Logistic regression modeling of prevalence of soybean Sclerotinia stem rot in the north-central region of the United States. *Phytopathology*, 94, 102–110.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011) Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.

Peel, R.G., Kennedy, R., Smith, M. & Hertel, O. (2014) Relative efficiencies of the Burkard 7-day, Rotorod and Burkard personal samplers for *Poaceae* and *Urticaceae* pollen under field conditions. *Annals of Agricultural and Environmental Medicine*, 21, 745–752.

Pesticide Risk Reduction Program. (2017) *Crop profile for dry bean in Canada, 2015*. Ottawa, ON: Agriculture and Agri-Food Canada.

Qandah, I.S. & Del Rio Mendoza, L.E. (2012) Modelling inoculum dispersal and Sclerotinia stem rot gradients in canola fields. *Canadian Journal of Plant Pathology*, 34, 390–400.

R Core Team. (2020) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available from https://www.r-project.org/ [Accessed 28th February 2024]

Reich, J., Alexander, T.W. & Chatterton, S. (2016) A multiplex PCR assay for the detection and quantification of *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *Letters in Applied Microbiology*, 62, 379–385.

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Reich, J. & Chatterton, S. (2023) Predicting field diseases caused by *Sclerotinia sclerotiorum*: a review. *Plant Pathology*, 72, 3–18.

Reich, J., Chatterton, S. & Johnson, D. (2017) Temporal dynamics of *Botrytis cinerea* and *Sclerotinia sclerotiorum* in seed alfalfa fields of southern Alberta, Canada. *Plant Disease*, 101, 331–343.

Reich, J., McLaren, D.L., Kim, Y.M., Wally, O., Yevtushenko, D., Hamelin, R. et al. (2023) Occurrence of ascospores and white mould caused by *Sclerotinia sclerotiorum* in dry bean fields in Alberta, Canada. *Plant Disease*, 107, 3754–3762.

Reich, J., Tetzlaff, T., Harding, M.W. & Chatterton, S. (2019) White mould of dry bean in southern Alberta in 2018. *Canadian Plant Disease Survey*, 99, 132–133.

Rosete, Y.A., To H, Evans, M., White, K., Saleh, M., Trueman, C. et al. (2021) Assessing the use of DNA detection platforms combined with passive wind-powered spore traps for early surveillance of potato and tomato late blight in Canada. *Plant Disease*, 105, 3610–3622.

Savage, D., Barbetti, M.J., MacLeod, W.J., Salam, M.U. & Renton, M. (2013) Temporal patterns of ascospore release in *Leptosphaeria maculans* vary depending on geographic region and time of observation. *Microbial Ecology*, 65, 584–592.

Schwartz, H.F. & Steadman, J.R. (1978) Factors affecting sclerotium production of, and apothecium production by, *Sclerotinia sclerotiorum*. *Phytopathology*, 68, 383–388.

Shahoveisi, F., Riahi Manesh, M. & del Río Mendoza, L.E. (2022) Modeling risk of *Sclerotinia sclerotiorum*-induced disease development on canola and dry bean using machine learning algorithms. *Scientific Reports*, 12, 864.

Smith, D.L., Kerns, J.P., Walker, N.R., Payne, A.F., Horvath, B., Inguagiato, J.C. et al. (2018) Development and validation of a weather-based warning system to advise fungicide applications to control dollar spot on turfgrass. *PLoS One*, 13, e0194216.

Statistics Canada. (2023) Table 32-10-0359-01: Estimated areas, yield, production, average farm price and total farm value of principal field crops, in metric and imperial units. Available from: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3210035901 [Accessed 28th February 2024]

Sun, P. & Yang, X.B. (2000) Light, temperature, and moisture effects on apothecium production of *Sclerotinia sclerotiorum*. *Plant Disease*, 84, 1287–1293.

Twengström, E., Sigvald, R., Svensson, C. & Yuen, J. (1998) Forecasting Sclerotinia stem rot in spring sown oilseed rape. *Crop Protection*, 17, 405–411.

Van der Heyden, H., Dutilleul, P., Charron, J.B., Bilodeau, G.J. & Carisse, O. (2021) Monitoring airborne inoculum for improved plant disease management. A review. *Agronomy for Sustainable Development*, 41, 40.

Van der Heyden, H., Lefebvre, M., Roberge, L., Brodeur, L. & Carisse, O. (2014) Spatial pattern of strawberry powdery mildew (*Podosphaera aphanis*) and airborne inoculum. *Plant Disease*, 98, 43–54.

Willbur, J.F., Fall, M.L., Bloomingdale, C., Byrne, A.M., Chapman, S.A., Isard, S.A. et al. (2018) Weather-based models for assessing the risk of *Sclerotinia sclerotiorum* apothecial presence in soybean (*Glycine max*) fields. *Plant Disease*, 102, 73–84.

Willbur, J.F., McCaghey, M., Kabbage, M. & Smith, D.L. (2019) An overview of the *Sclerotinia sclerotiorum* pathosystem in soybean: impact, fungal biology, and current management strategies. *Tropical Plant Pathology*, 44, 3–11.

Wu, B.M. & Subbarao, K.V. (2008) Effects of soil temperature, moisture, and burial depths on carpogenic germination of *Sclerotinia sclerotiorum* and *S. minor*. *Phytopathology*, 98, 1144–1152.

Wutzki, C.R., Berger Neto, A., Grabicoski, E.M., Henneberg, L., Sartori, F.F. & Jaccoud Filho, D.S. (2019) Within-field variability and spatial analysis of white mould and soybean crop attributes in southern Brazil. *Tropical Plant Pathology*, 44, 104–111.

Xu, X., Harris, D.C. & Berrie, A.M. (2000) Modeling infection of strawberry flowers by *Botrytis cinerea* using field data. *Phytopathology*, 90, 1367–1374.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Reich, J., McLaren, D., Kim, Y.M., Wally, O., Yevtushenko, D., Hamelin, R. et al. (2024) Predicting airborne ascospores of *Sclerotinia sclerotiorum* through machine learning and statistical methods. *Plant Pathology*, 73, 1586–1601. Available from: https://doi.org/10.1111/ppa.13902

13653059, 2024, 6, Downloaded from https://bsppjournals.onlinelibrary.wiley.com/doi/10.1111/ppa.13902 by University Of Lethbridge, Wiley Online Library on [19/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License