Stem Characteristics and Local Environmental Variables for Assessment of Alfalfa Winter Survival

Md Saifuzzaman [a,*], Viacheslav Adamchuk [a], Maxime Leduc [b]

[a]Department of Bioresource Engineering, McGill University, Montreal, Quebec H9X 3V9, Canada; md.saifuzzaman@mail.mcgill.ca (M.S.); viacheslav.adamchuk@mcgill.ca (V.A.)

[b]Fondateur et Gestionnaire de projet, Jasons Systèmes Fourragers, Canada; mleduc@jasonssystemesfourragers.com

A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

Abstract.

Alfalfa (Medicago sativa L.) is considered the queen of forage due to its high yield, nutritional qualities, and capacity to sequester carbon. However, there are issues with its relatively low persistency and winter survival as compared to grass. Winter survival in alfalfa is affected by diverse factors, including the environment (e.g., snow cover, hardiness period, etc.) and management (e.g., cutting timing, manure application, etc.). Alfalfa's poor winter survival reduces the number of living plants, delays plant development, and diminishes field productivity. To better understand poor winter survival and persistency in alfalfa and assess winter damage, this research aimed to develop an assessment tool for Canadian growers. In addition, a prediction model was designed to consider and account for the variability and potential risks. Both field measurements and remote sensing approaches were incorporated into the assessment tool.

Soil samples, stem counts, and height were collected from 192 farms in four provinces – Nova Scotia, Quebec, Ontario, and Manitoba. The field sampling design used time-series vegetation indices in the k-means clustering procedure. A randomized design was implemented in each cluster. The stem count samples were measured from each site in the Spring and Fall of 2021. The soil texture was mainly loam, which varies across the provinces. A total of 1612 targeted soil samples were collected from the four regions. The sampling points were then positioned using the iPad GPS. Lab-measured soil micro-and macro-nutrients were pH, soil organic matter (SOM), phosphorus (P), potassium (K), cation exchange capacity (CEC), Magnesium (Mg), Manganese (Mn), Zinc (Zn), and Calcium (Ca). Many regions also used soil and stem characteristics for winter risk assessment grids. The initial field assessment scores were evaluated based on suitable parameters (i.e., stand age, soil pH and potassium levels, harvest frequency, and cultivar type) of all agro-ecological zones for the status of potential risks. Both historical field measurements and topographic datasets were used for the assessment model. Descriptive statistical analysis and correlation between stem characteristics and topographic variables, together with the

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 15th International Conference on Precision Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2018). Title of paper. In Proceedings of the 15th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

knowledge of soil nutrients, enhanced our understanding of the spatial heterogeneity of alfalfa production areas. A random forest regression model was applied. Model parameters were developed to determine the number of essential variables and regression trees to be used in the training phase; this resulted in optimal model performance and scenario maps.

The prediction model and data-driven decisions pose challenges only with soil chemical analysis in assessing winter mortality, identifying potential agronomic and environmental factors and their potential for improvement. The emerging risk assessment tools and the application of generalized models considering all potential factors described in regional guidelines will assist Canadian forage growers in improving their productivity by using alternative management practices, including species selection and soil recommendations, using information on survival rates and persistency, to increase financial returns.

Keywords.

Alfalfa, Winter Survival, Precision Agriculture, Models, Risk assessment, Prediction

Introduction

Alfalfa production systems are well adapted in diverse agro-climatic regions due to their economic importance of high yield, nutritional qualities, and capacity to sequester carbon. Profitable forage production depends on favorable conditions along with a selection of well-drained growing fields. However, there are many issues with its lower persistency and winter survival as compared to grass. [1,4]. Winter survival in alfalfa is affected by diverse factors, including environmental (*e.g.*, snow cover, hardiness period, etc.) and management (*e.g.*, cutting management, manure application, etc.). Alfalfa's poor winter survival reduces the number of living plants, delays plant development, and diminishes field productivity [2], which negatively affects to the production system of many regions.

Due to the impact of diverse drivers and anomalies in the production system, alfalfa fields manage distinctively in seasons and different regions. According to regional guidelines, many management decisions affect alfalfa growth and persistency. This management includes variety selection, seeding rate, stand establishment, harvesting time, soil fertilizer and irrigation management, and managing moisture conditions [1]. Besides these field managements, diverse climatic and topographic conditions must be emphasized to manage the production fields better and enhance winter survival [4]. Also, alfalfa growing in the summer is a different structure that winter hardens alfalfa. To better understand poor winter survival and persistency in alfalfa and assess winter damage, this research aimed to develop an assessment tool for Canadian growers. In addition, a prediction model was designed to understand the variability and potential risks. Both field measurements and remote sensing approaches were incorporated into the assessment tool.

Materials and methods

2.1 Study site and sample data collection

Soil samples, stem counts, and stem heights during sampling were collected from 192 farms (478 fields) in four provinces: Nova Scotia, Quebec, Ontario, and Manitoba (Fig. 1 and Table 1). The field sampling design used time-series vegetation indices in the k-means clustering procedure. A randomized design was implemented in each cluster of the farms and fields. The stem count samples were measured from each site (composite of 3 data points, often called a landmark) in the Spring and Fall of 2021. Soil samples were also collected from the same site as the stem count locations. The stem count measurements were taken from a rectangular frame used in each landmark position (Fig. 2). The soil texture was mainly loam, maintaining moderate drainage conditions [2].

Table 1: Field statistics and data collection sites

Summary	**Statistics**
Total fields	478 (out of 1040)
Total farms	192 (out of 250)
Number of producers	192
Number of advisors	33

Fig 1. Study area and data collection sites (red points): alfalfa stem count, yield, and soil sample locations in Spring & Fall 2021.

2.2 Field measurements – Stem count

Stem count

The stem count samples were measured from each sampling site in the Spring and Fall of 2021. Sample measurement locations and placement of the quadra are shown in Fig. 2.

(a) (b) (c)

Fig 2. Landmark (no. 2 out of 3 with a distance of 1 m) placed in the sampling site (a), measurements were taken with in the red rectangle (30x30 cm) paced on the corner of each landmark (b), and each stem was counted from the individual plant (c).

Soil Sampling

A total of 1612 targeted soil samples were collected in 2021 and 2022. The sampling points were then positioned using the iPad GPS. Each soil sample was obtained from a composite of 8-10 soil cores 17 cm (7 in.) deep mixed in a sampling bucket and then processed for lab analysis. Lab-measured soil micro-and macro-nutrients were pH, soil organic matter (SOM), phosphorus (P), potassium (K), cation exchange capacity (CEC), Magnesium (Mg), Manganese (Mn), Zinc (Zn), Calcium (Ca), Aluminium (Al), Boron (B), Copper (Cu), Iron (Fe), Integral Suspension Pressure (ISP), Saturation of K/Mg/Ca, and lime index. Except pH, SOM, CEC, ISP, all properties were analyzed by Mehlich III extraction method.

2.3 Other datasets

According to regional guidelines, cultivar, seeding rate, harvesting time, soil nutrients and moisture, fertilizer application, and drainage conditions were important assessment parameters. Historical crop and field management practices, weather, and topographic datasets were used for the assessment model. Those datasets were – cropping practices, soil history, temperature and rainfall, topographic data, and the following management practices:

Crop history and stand age

Alfalfa's stand age was calculated by field cropping practices each year. Therefore, we only considered forage crops from 2018 to 2021.

Soil history and texture

Soil history and nutrient data were recorded at the field level from 2018 to 2020. In addition, the land suitability and soil texture map, published by the Research and Development Institute for the Agri-environment (IRDA), Quebec, contains valuable information at the farm level for the producer.

Management practices

We also collected field management data: manure, fertilizer (N-P-K), and lime applications at the field scale. Seeding rates and cultivar data were also collected in 2021.

Topographic data

Topographic parameters including: elevation, topographic wetness index (TWI), slope, and aspect ratio, were calculated from light detection and ranging (LiDAR) data of high-resolution digital elevation model (HRDEM) data, Canada. Drainage condition and performance were initially evaluated from the generated TWI.

Weather data

We used historical daily weather data, maximum and minimum temperature, and cumulative rainfall from the Hobolink weather station located at every farm. We also used open-source weather data from Meteostat to fill in the missing Hobolink data. Growing degree days (GDD) were calculated for alfalfa production.

2.4 Data preprocessing:

Data preprocessing:

Stem count and soil samples were collected from the exact sample location (composite of 3 data points, often called a landmark). Timestamps, seasonal variations, locations, the distance between data points, and other variable measurements were evaluated simultaneously in the preprocessing steps. By removing outliers, various data features were added to the raw data for geospatial adjustment.

Statistical analysis and outlier detection:

Potential outliers and null values of the sample measurements (latitude and longitude, and unique ID of the fields along with test data point, other comments from the producer) were identified in this step. Data filtering was applied after generating the histogram, data distribution curve, and descriptive statistics. In this study, different environmental variables were considered for building the input and training datasets used by the model. General statistical analysis and correlation matrices of the selected variables were used to determine targeted variables in the following sections.

Stem values were counted mostly between 10 and 90 (Fig. 3). The stem count for the Fall and Spring seasons had a similar distribution pattern in all of the provinces (Fig. 4). Soil chemical analysis was compared to three provinces only since analysis data were not available for Nova Scotia. In terms of pH measurements, Manitoba was more alkaline than the other three provinces (Fig. 5). In micronutrient analysis, Quebec had a more diverse soil nutrient distribution than the other provinces.

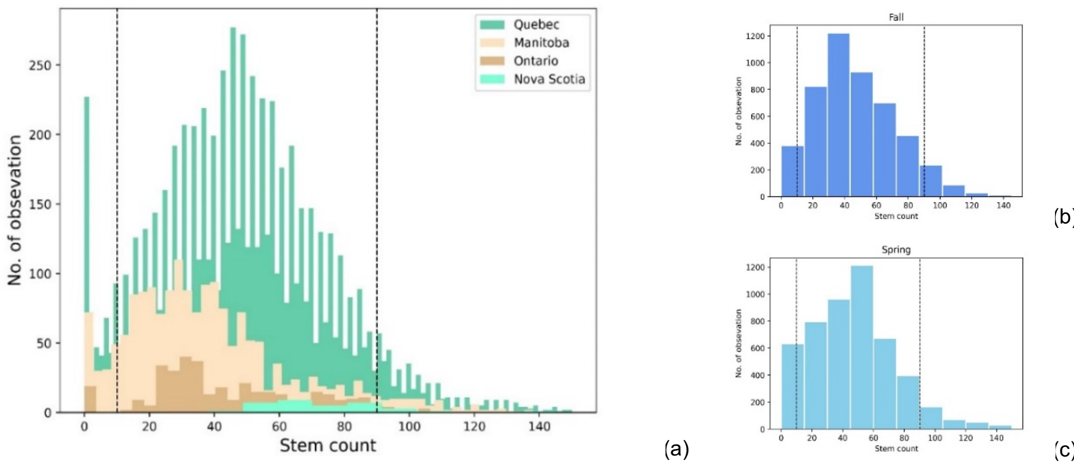

Fig 3. Stem count by province (a) and by seasonal distribution by Fall (b) and Spring (c).

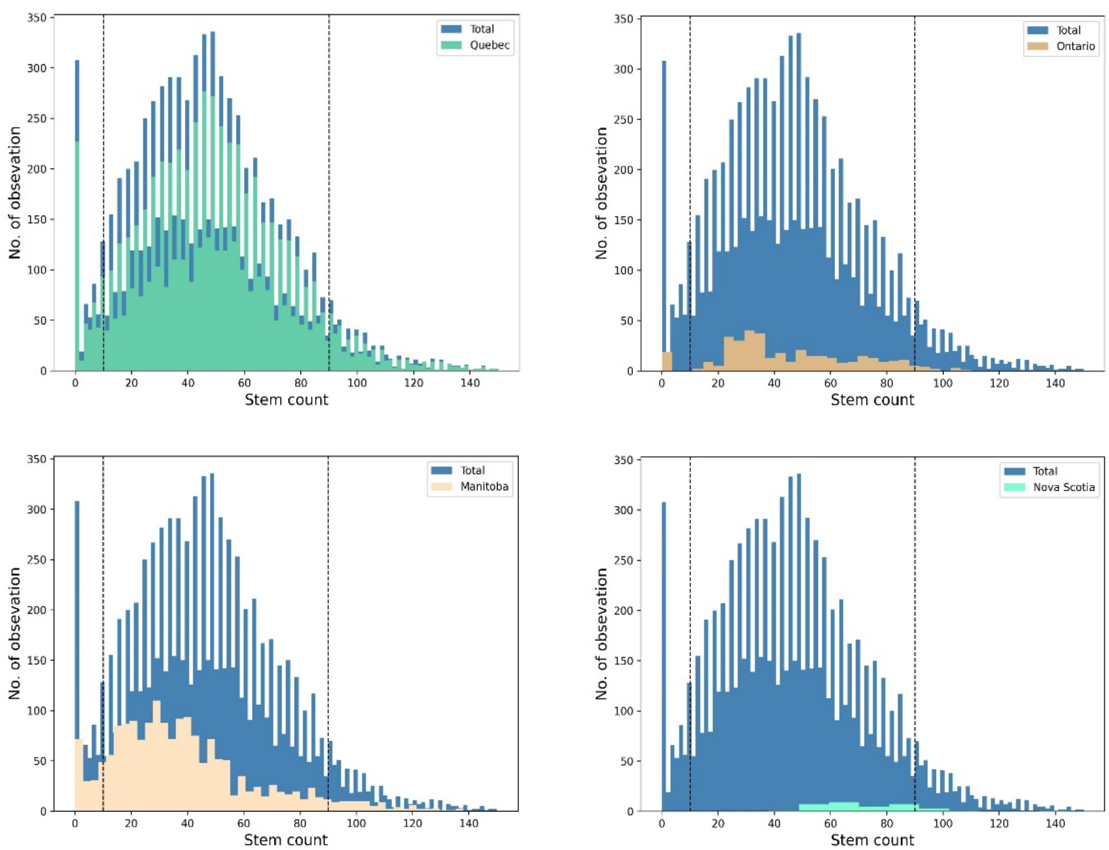

Fig 4. The total stem count of this study was compared to each Province (Clockwise: Quebec, Ontario, Nova Scotia, and Manitoba).

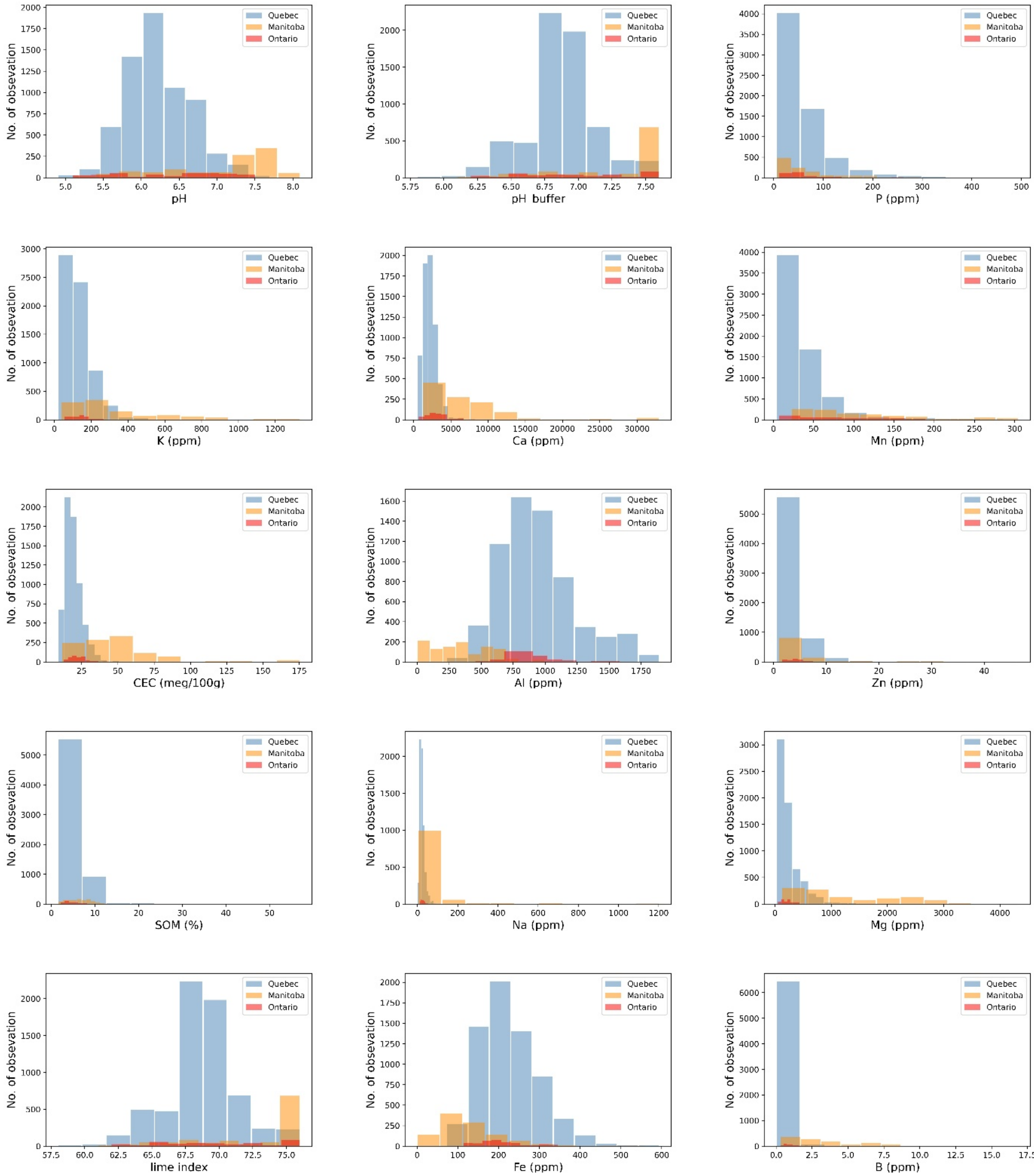

Fig 5. Soil nutrients at the field scale compared to different provinces in Canada.

2.5 Regression forest model and parameter optimization

Regression forest (RF) design

Descriptive statistical analysis and correlation between stem characteristics and soil nutrients enhanced our understanding of the spatial heterogeneity of alfalfa production areas. A random forest regression model was designed and implemented for stem count prediction (Fig. 6). Model parameters were developed to determine the number of essential variables and regression trees to be used in the training phase; this resulted in optimal model performance and scenario maps.

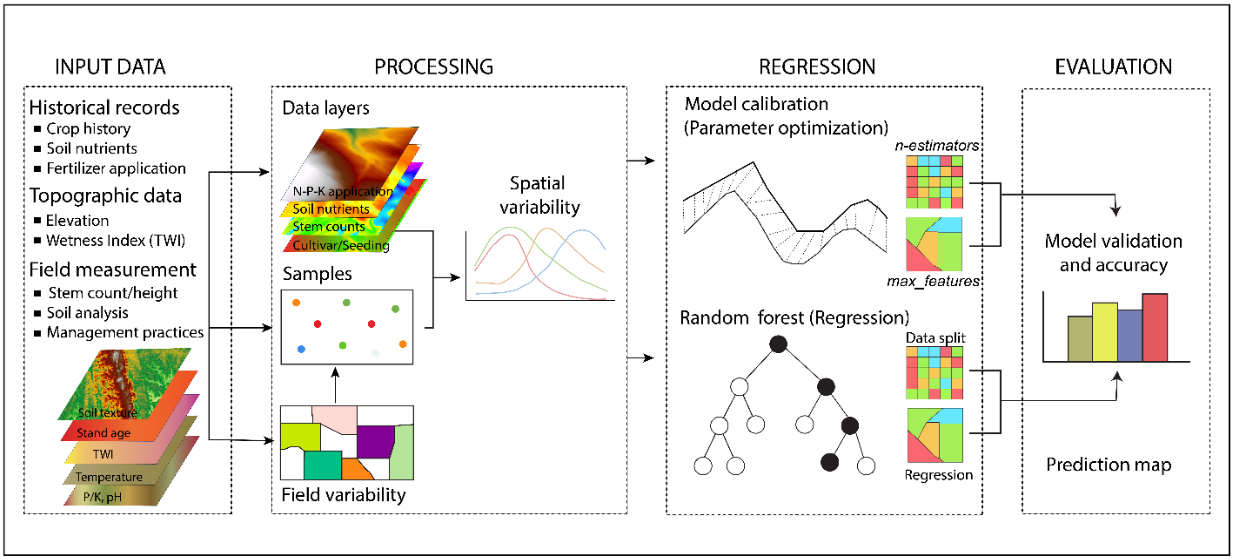

Fig 6. Model design and process diagram.

For the regression procedure, the random forest built *k* trees, where the predicted values were the average of all individual tree predictions. However, it does not handle the data beyond the training samples. Random forest regression creates a set of *K* trees [Tx_i,….., Tx_k)], where x = [x_i,….., x_β], is a β-dimension of the input vector which forms a forest. The predicted values were obtained by the aggregation of the results of all individual trees. The following equation provides the random forest regression predictor:

$$f(x) = \frac{\sum_{k=1}^{k=K} T_k(x)}{K} \tag{6}$$

Random forest builds a set of regression trees (*K*) and averages the predictions of individual trees to make a final prediction [5]. Where *k* is the individual bootstrap sample, and T_k is the individual learner or decision tree.

Parameter optimization

The training dataset checked different combinations of soil properties in the training stage to fit a regression model and determine the parameters of the random forest model. The model used all of the Fall datasets in training, and there was no test dataset. In this case, the model score is 0.79, and the validation (out of bag - OOB) score is 0.56. If we used 20% in test data, the validation mean absolute error (MAE) of the stem count was 13. After the hyper-parameter optimization (‘n estimators': 100; 'max features': 5; 'bootstrap': True; 'random state':35; ‘min samples leaf’:4), the RF training and validation score were 0.80 and 0.54, respectively. Based on the training and test data, the relative importance of all soil variables is shown in Fig. 7. For the sensitivity analysis, it is important to evaluate the model on a dataset using k-fold cross-validation.

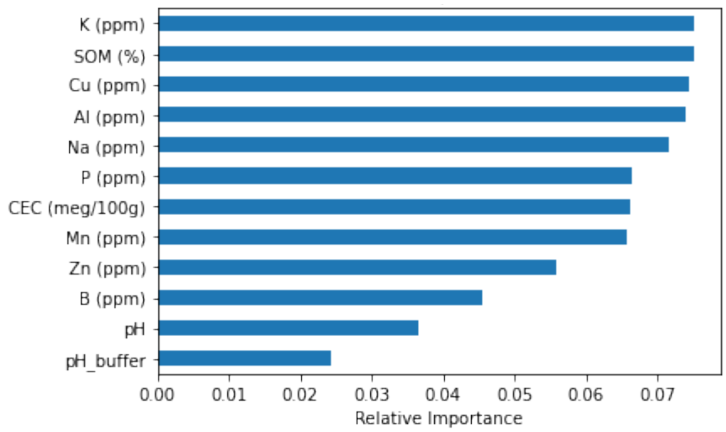

Fig 7. Feature importance of all variables for the regression model before optimization

Results and Discussion

Stem analysis

The stem count and descriptive statistics are shown in Tables 2 and 3. Higher mean values (μ = 73.91) of stem count were observed in Nova Scotia than the other three provinces, while maximum stem count was observed in Quebec. A maximum stem count of 165 was found in the Spring season, while the maximum stem count in the Fall season was 145. The average stem count was above 45 in all provinces, which was not considered in limiting yield of the fields according to the regional guideline [4].

Table 2: Descriptive statistics for stem count data of differences

Province	Min	Max	Mean	Std
Quebec	0	165	45.87	24.46
Nova Scotia	38	145	73.91	19.85
Ontario	0	110	44.98	23.56
Manitoba	0	160	38.61	25.86

Table 3: Descriptive statistics for stem count data of two seasons

Season				
Spring	0	165	43.68	25.28
Fall	0	145	45.62	24.42
Total	0	165	44.64	24.88

Soil analysis

The lab-measured soil analysis data were processed and selected for the prediction model. Ranges between the maximum (max) and minimum (min) values for the soil properties varied throughout the whole field and farm data (Table 4). The range, standard deviation - Std (σ), and mean (μ) for each soil parameter showed a large variability in the different provinces. The measured pH values varied between 4.9 and 8.1 (lower values of σ = 0.41 and μ = 6.24 in Quebec than in other provinces). For the SOM measurements, the range varied highly between 1.5% and 57.1% in the whole dataset (σ = 2.9%), which was observed in Quebec. The standard deviation value of Na measurements was larger in Manitoba (σ = 121.42 ppm) than in Ontario (σ = 9.01 ppm). Less variability was found in P measurements (μ = 57.80 ppm in Quebec; μ = 58.88 ppm in Ontario, while K showed high variability (μ = 339.24 ppm in Manitoba; μ = 145.17 ppm in Ontario.

Table 4: Descriptive statistics of soil nutrients in the three provinces of Canada.

Al (ppm)				
Province	Mean	Std	Min	Max
Manitoba	337.02	235.07	0.00	995.00
Ontario	878.98	198.28	454.00	1584.00
Quebec	941.84	290.07	228.00	1895.00

SOM (%)				
Province	Mean	Std	Min	Max
Manitoba	6.25	2.54	1.60	12.20
Ontario	4.82	1.80	2.10	12.20
Quebec	5.25	2.91	1.50	57.10

B (ppm)				
Province	Mean	Std	Min	Max
Manitoba	3.22	2.53	0.30	14.30
Ontario	1.01	0.43	0.10	2.20
Quebec	0.60	0.58	0.00	16.90

P (ppm)				
Province	Mean	Std	Min	Max
Manitoba	62.33	57.84	6.25	295.98
Ontario	58.88	46.73	11.61	262.05
Quebec	57.80	49.20	4.91	495.09

CEC (meg/100g)				
Province	Mean	Std	Min	Max
Manitoba	47.96	28.93	11.10	175.90
Ontario	22.64	5.93	12.50	42.50
Quebec	19.02	5.55	8.80	50.50

K (ppm)				
Province	Mean	Std	Min	Max
Manitoba	339.24	260.18	35.71	1337.05
Ontario	145.17	55.33	55.80	319.64
Quebec	128.07	75.76	19.64	840.63

Ca (ppm)				
Province	Mean	Std	Min	Max
Manitoba	6868.12	5478.43	1211.16	33037.50
Ontario	3131.23	1415.20	691.07	8201.34
Quebec	2167.28	875.13	516.07	7389.73

Ca_sat (%)				
Province	Mean	Std	Min	Max
Manitoba	68.45	14.89	37.00	94.00
Ontario	67.10	17.70	22.00	96.00
Quebec	56.19	13.13	18.00	97.00

Cu (ppm)				
Province	Mean	Std	Min	Max
Manitoba	3.04	2.22	0.00	11.79
Ontario	2.58	2.03	0.63	21.96
Quebec	2.34	1.63	0.26	15.50

K_sat (%)				
Province	Mean	Std	Min	Max
Manitoba	2.23	1.61	0.00	7.80
Ontario	1.74	0.75	0.60	4.00
Quebec	1.73	0.88	0.20	9.50

Fe (ppm)				
Province	Mean	Std	Min	Max
Manitoba	130.02	80.97	0.00	558.16
Ontario	206.80	56.58	113.99	345.46
Quebec	227.21	71.67	72.78	594.82

KMgCa_sat (%)				
Province	Mean	Std	Min	Max
Manitoba	92.31	12.38	46.10	100.00
Ontario	77.66	17.70	28.00	100.00
Quebec	67.56	14.06	20.40	100.00

ISP (%)				
Province	Mean	Std	Min	Max
Manitoba	15.50	12.19	1.40	49.30
Ontario	6.73	5.40	1.60	40.80
Quebec	6.56	6.05	0.50	47.50

Mg_sat (%)				
Province	Mean	Std	Min	Max
Manitoba	21.63	11.44	5.21	51.51
Ontario	8.82	4.62	2.61	17.81
Quebec	9.64	5.24	1.67	30.34

Lime index				
Province	Mean	Std	Min	Max
Manitoba	73.09	4.26	61.00	76.00
Ontario	69.93	4.22	62.00	76.00
Quebec	68.57	2.67	58.00	76.00

Na (ppm)				
Province	Mean	Std	Min	Max
Manitoba	61.48	121.42	0.00	1204.91
Ontario	25.67	9.01	10.71	50.89
Quebec	22.82	12.30	0.00	89.29

Mg (ppm)				
Province	Mean	Std	Min	Max
Manitoba	1254.23	894.94	117.86	4337.05
Ontario	223.39	103.83	53.57	591.96
Quebec	238.33	193.33	29.02	1437.50

Zn (ppm)				
Province	Mean	Std	Min	Max
Manitoba	5.77	7.71	0.90	45.80
Ontario	4.24	1.82	1.60	11.50
Quebec	3.44	2.47	0.50	46.60

Mn (ppm)				
Province	Mean	Std	Min	Max

pH				
Province	Mean	Std	Min	Max

Manitoba	107.57	71.34	22.00	305.20
Ontario	82.43	61.97	7.40	266.90
Quebec	35.42	33.60	3.50	286.00

Manitoba	7.02	0.72	5.10	8.10
Ontario	6.40	0.69	5.10	7.50
Quebec	6.24	0.41	4.90	7.70

Buffer pH				
Province	Mean	Std	Min	Max
Manitoba	7.31	0.43	6.10	7.60
Ontario	6.99	0.42	6.20	7.60
Quebec	6.86	0.27	5.80	7.60

Alfalfa is growing in an optimum condition of soil pH between 6.5 and 8.0 [2]. Lower pH level (<6.0) was mostly found in Quebec. The soil sample analysis observed lower K values (<80 ppm), mainly in Manitoba and Quebec. According to the regional guideline, lower levels (<80 ppm) of soil exchangeable K caused more winter damage.

Correlation analysis

According to the relationship between the predictor variables (soil analysis) found in Fig. 8, most variables were minimal colinear and considered for the prediction model. Pairwise relationships between the stem count, soil properties, and their strengths are shown in a correlogram (Fig. 9). SOM correlated positively (r = 0.16) with Spring Stem count measurements. However, no systematic correlations of the micronutrients were found with Fall stem measurements.

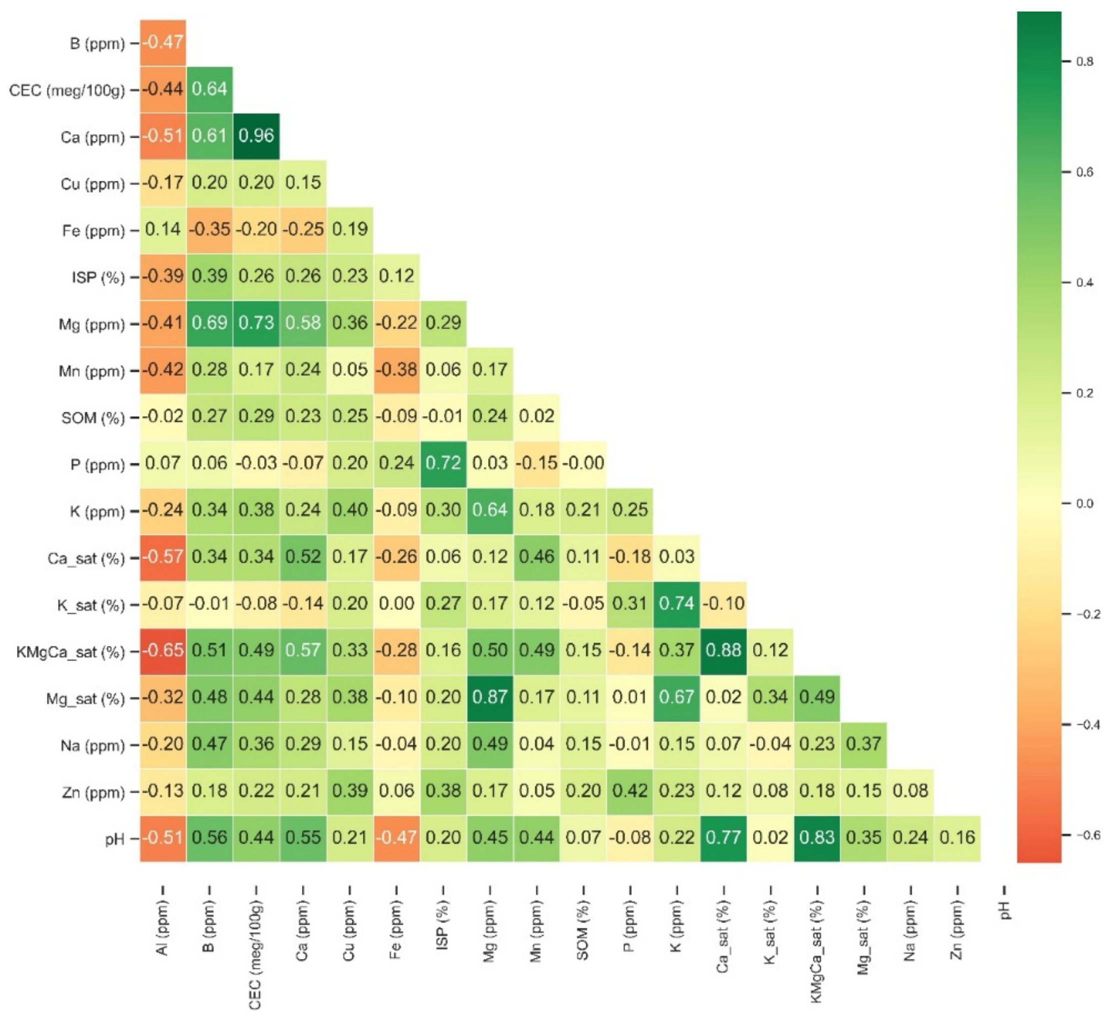

Fig 8. Correlogram between soil nutrients of four provinces.

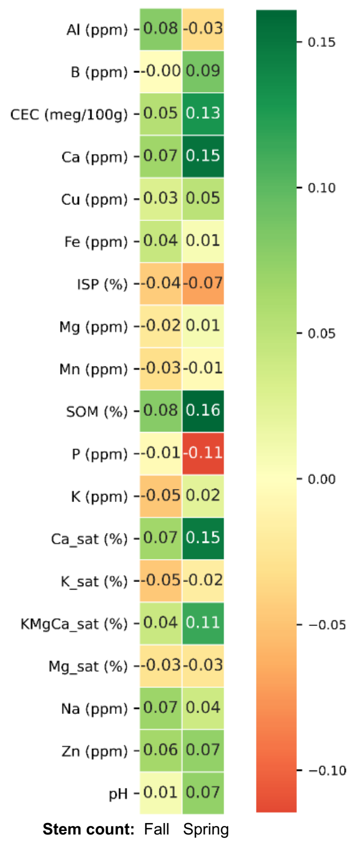

Fig 9. Correlation between soil nutrients and stem count (Fall and Spring 2021 measurements)

Development of the random forest regression

Training datasets were prepared for the model prediction after a rigorous statistical assessment of the field measured data. In the RF regression model, from the original data (8242 data points) of soil properties, 80% of the dataset was randomly selected for the training data, and 20% served as a test set and a final validation set. About 60% of the training data were randomly selected for developing forest model estimators and evaluating the parameters in the trained model. About 20% of the data were selected for cross-validation and performance evaluation of the regression model estimator at this initial stage. In this study, approximately one-sixth of the data points (1374 out of 8242 sample datasets) served for the regression models' final validation and accuracy assessment. In the regression model, the predictor variables (number of sample points, n= 1374) that have a different effect in stem count assessment.

The sensitivity analysis of individual variables was evaluated by the degree of contribution when the RF model split a node for decision. This study tested a single approach variable reduction (default settings). The RF model evaluated the relative importance of 21 variables. Less influential variables (micro-nutrients) were removed for testing the model performance. Fig. 10 showed the relative variable importance when selected nutrients were considered for predicting the stem count for winter assessment. The selected variables were pH, soil organic matter (SOM), phosphorus (P), potassium (K), cation exchange capacity (CEC), Magnesium (Mg), Calcium (Ca), and Aluminum (Al). The n-estimators values were selected from a range in the trained model. The optimum value for n-estimators with a ten-fold (k = 10) CV procedure was within a range from 120 to 230 for the different stem count values. Based on the initial results, the optimum value of n-estimators was 100, where R^2 improved for Spring and Fall stem prediction. After several runs, R^2 reached the maximum level (R^2 = 0.47) in the independent cross-validation phase when the number of dominant variables among the soil chemical analysis was selected.

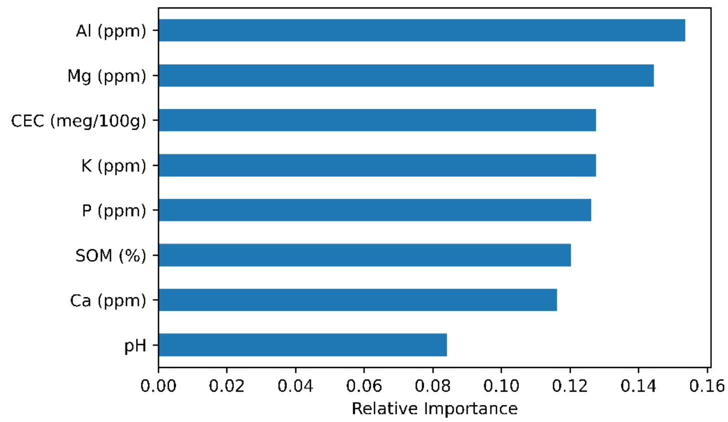

Fig 10. Feature importance of the selected variables for the regression model

Conclusion

The prediction model and data-driven decisions pose challenges in assessing winter mortality, identifying influential agronomic and environmental factors and their potential for improvement. The RF regression (training and testing) analysis indicates that soil variability determined using field measurements and methods improved the construction of precise prediction models for winter assessments. The emerging risk assessment tools and the application of generalized models will assist Canadian forage growers in improving their productivity by using alternative management practices, including species selection and soil recommendations, by using the information on survival rates and persistency to increase financial returns [6].

Besides many challenges in the data collection system, only reliable field measurements deploy to build a numeric simulation for improving alfalfa's winter assessment. Along with the RF model, internal model validation and independent cross-validation would increase accuracy and efficiency for the alfalfa stand assessment. Immediate research work will be emphasized on other variables – weather variables, drainage characteristics, seeding rate, and cultivar type – along with feature scaling and error optimization of the numeric simulation, which will be better off the decision support tool. The developed algorithm and model will improve the prediction methods and provide tools for a decision support system in any dynamic production system across the provinces. Further research will validate and implement results through a set of case studies, after which the findings will be disseminated among the agricultural farming communities. Thus, erroneous data removal techniques and supervised machine learning prediction frameworks could be implemented as web applications to facilitate appropriate site-specific agronomic and environmental decisions.

Acknowledgments

The authors give special thanks to the CFGA and Agriculture and Agri-Food Canada and IRDA Quebec for data support and cooperation. We are grateful to the Dr. Gilles Bélanger and Dr. Dan Undersander for their valuable suggestions and contribution to the project. In addition, we would like to thank graduate students and postdocs for data processing at various levels.

References

[1] J. Lauzon, G. F. Tremblay, G. Bélanger *et al.*, "Alfalfa and timothy nutritive value in contrasting agroclimatic regions," *Agronomy Journal,* vol. 111, no. 3, 2019.

[2] A. L. Laycook, and D. A. Price, "Factor influencing forage quality: Environmental influences on nutrional value of forage plants," in Range and Wildlife Habitat Evaluation: A Research Symposium, 1970.

[3] N. Tantalaki, S. Souravlas, and M. Roumeliotis, "Data-Driven Decision Making in Precision Agriculture: The Rise of Big Data in Agricultural Systems," *Journal of Agricultural & Food Information,* vol. 20, no. 4, pp. 344-380, 2019/10/02, 2019.

[4] D. Undersander, N. Martin, D. Cosgrove *et al.*, *Alfalfa management guide*, 2011.

[5] M. Saifuzzaman, "Optimization of geospatial data modelling for crop production by integrating proximal soil sensing and remote sensing data," McGill University Libraries, [Montreal], 2020.

[6] S. Orloff, *Choosing appropriate sites for alfalfa production*, Chapter 2, University of California, Oakland, California, 2007.

Nomenclature

Aluminum (Al)

Bore (B)

Calcium (Ca)

cation exchange capacity (CEC)

Copper (Cu)

High-resolution digital elevation model (HRDEM)

Integral suspension pressure (ISP)

Iron (Fe)

Light detection and ranging (LiDAR)

Magnesium (Mg)

Manganese (Mn)

Maximum value (Max)

Minimum value (Min)

Out of bag (OOB)

Phosphorous (P)

Potassium (K)

Random Forest (RF)

Saturation of Ca, K, Mg, K+Mg+Ca)

Sodium (Na)

Soil organic matter (SOM)

Standard deviation (Std)

Topographic wetness index (TWI)

Zinc (Zn)