scientific reports

OPEN

A probabilistic approach to predicting alfalfa's winter survival using local conditions, weather and management factors

Md Saifuzzaman[1]✉, Viacheslav I. Adamchuk[1], Maxime Leduc[2], Karem Chokmani[3], Hamed Etezadi[1] & Philippe Seguin[4]

Alfalfa (*Medicago sativa L.*) produce relatively high yields of high-quality forage and contributes to carbon sequestration. Still, its winter survival is often lower than that of grasses, reducing plant density and field productivity. Winterkill is influenced by environmental factors (e.g., snow cover, temperature fluctuations, hardening period) and management practices (e.g., cutting time, fertilization, drainage). To assess these factors and improve persistence, this study developed a quantitative assessment tool for Canadian forage producers. A numerical simulation framework integrated soil, weather, and field management variables to evaluate yield variability and winterkill risk. Data were collected from 225 farms across Nova Scotia, Quebec, Ontario, and Manitoba using a randomized hierarchical sampling design. Soil samples were analyzed in commercial laboratories, and stem density was measured each spring and fall over three years. Descriptive statistics linking stem characteristics with soil and topographic features, weather conditions, and management practices, including soil nutrient levels, revealed a decline in mean stem counts from 49 in spring 2021 to 37 in spring 2023. To illustrate performance, weighted scores and persistence analyses were used to define model parameters across three distinct scenarios: optimal, average, and worst-case. The risk-assessment tool offers decision support to Canadian forage growers, enhancing productivity through informed management, species selection, and soil recommendations.

Keywords Alfalfa, Environmental factors, Field management, Persistency, Probability model, Risk assessment

Abbreviations

Al	Aluminum
BS	Base saturation
B	Bore
BD	Bulk density
Ca	Calcium
CEC	Cation exchange capacity
CRAAQ	Centre de référence en agriculture et agroalimentaire du Québec
Cu	Copper
DSS	Decision support system
GDD	Growing degree days
HRDEM	High-resolution digital elevation model
ISP	Integral suspension pressure
Fe	Iron
LiDAR	Light detection and ranging
Mg	Magnesium

[1]Department of Bioresource Engineering, McGill University, Montreal, QC, Canada. [2]Agronomy, Jasons Forage Systems, Montreal, Canada. [3]Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Quebec, Canada. [4]Department of Plant Science, McGill University, Montreal, QC, Canada. ✉email: md.saifuzzaman@mail.mcgill.ca

Mn	Manganese
Max	Maximum value
Min	Minimum value
NDVI	Normalized difference vegetation index
OMAFRA	Ontario Ministry of Agriculture, Food and Rural Affairs
P	Phosphorous
K	Potassium
IRDA	Research and Development Institute for the Agri-environment
K + Mg+Ca	Saturation of Ca, K, Mg
Na	Sodium
SOM	Soil organic matter
Std.	Standard deviation
TWI	Topographic wetness index
Zn	Zinc

Alfalfa (*Medicago sativa* L.) is vital for livestock industries across diverse agro-climatic regions, providing essential nutrients for animals while enhancing soil health through carbon sequestration and improved structure[1,2]. However, alfalfa's vulnerability in colder regions, particularly its reduced winter survival and stand persistence compared to grasses, remains a significant constraint. Yield losses are commonly linked to harsh environmental conditions such as prolonged snow cover, ice sheeting, and freeze-thaw cycles, which elevate the risk of frost damage and impair root viability by affecting crown temperature, a critical factor in crown survival[3–6]. These limitations present both agronomic and economic challenges, particularly in northern regions where consistent forage availability is critical.

The economic impact of alfalfa winterkill includes reduced yields and increased reseeding resulting in higher management costs that could potentially discourage some from producing alfalfa in affected regions[7–9]. In addition to the key environmental stressors, genetic factors such as dormancy levels and metabolite accumulation play a vital role in winter hardiness, with specific cultivars exhibiting greater resilience[10]. Moreover, agronomic practices, including the timing of harvest, fertilization management, and adequate field drainage, significantly influence plant survival[11]. Maintaining soil fertility, optimal pH levels, and water availability also supports plant resilience against winter stress[12,13]. To enhance alfalfa's persistence in cold climates, an integrated strategy that combines adaptive management, environmental monitoring, and genetic selection is essential.

Advances in precision agriculture, utilizing remote and proximal sensing, have encouraged the development of innovative assessment tools for evaluating alfalfa winter persistence[14]. Remote sensing techniques, complemented by ground validation methods, offer a holistic approach to monitoring plant health and predicting winter survival rates[15]. Moreover, recent advances in remote sensing and predictive modelling present novel avenues for monitoring alfalfa fields and forecasting winter survival rates[16,17]. The integration of soil and plant data via geographic information systems (GIS) can help establish correlations between soil texture, moisture levels, and topography and alfalfa survival rates, providing a basis for precision management practices[18]. Machine learning models, such as random forest and neural networks, have demonstrated efficacy in integrating environmental, management, and genetic data to forecast winter survival outcomes with notable accuracy[7]. Despite these technological advancements, existing studies often focus on limited spatial scales and lack the capacity to integrate a wider array of environmental parameters and regional dynamics effectively[19]. This gap underscores the need for a sophisticated decision support system (DSS) that can encompass broad geographic scales and diverse data sets. Such a system would not only improve predictive accuracy but also enhance the strategic planning and management of alfalfa crops across various climatic zones, thereby making a significant contribution to precision farming.

The integration of precision agriculture technologies further strengthens field management by enabling real-time monitoring and data-driven decisions that help reduce the effects of adverse winter conditions on plants[20–22]. However, current predictive methodologies often fall short in accounting for the diversity of field management practices required to align with regional agricultural guidelines[23]. Traditional methods, including soil chemical analysis and growing degree days (GDD)-based plant growth indicators, while providing essential information, often fail to capture the complex interactions that influence alfalfa's winter resilience and productivity. This limitation highlights the urgent need to develop a user-friendly decision support system (DSS) that incorporates expert assessment criteria to address limiting factors in the production environment and enables user-defined field scenarios. Given the seasonal and regional variability in alfalfa production, a more holistic, comprehensive, and adaptive approach is required[24,25]. This highlights the value of numerical simulation in facilitating more informed and adequate decision-making in alfalfa cultivation. This research proposes a DSS as a predictive, comprehensive tool to support agronomic decision-making and improve alfalfa management across diverse geographic regions.

A DSS was developed using a likelihood-based probabilistic framework, Bayesian in nature, to support non-parametric data exploration and address the complex challenges of managing alfalfa winter survival under diverse agronomic and environmental conditions. The model estimates the likelihoods of outcomes (e.g., stem count probabilities) conditioned on observed soil and management variables, using field-derived likelihood functions. This approach boasts several advantages: it handles missing parameters, seamlessly integrates numerical and categorical variables, and effectively eliminates irrelevant covariates through the use of weighting factors. Additionally, the analytical framework is well-suited for analyzing non-parametric data, as it accommodates variables without predefined distribution patterns. It also establishes hierarchical relationships within the response variables and among the predictors.

Addressing the spatial and seasonal variability of alfalfa alongside the specific conditions controlled by regional dynamics, field management, and established regional guidelines poses a multifaceted challenge for producers and agricultural advisors. The diversity in climatic and topographic conditions across production regions requires region-specific strategies to optimize growth and enhance winter survival capabilities[26]. This complexity is further compounded by alfalfa's phenological changes during the summer months, a crucial phase for developing plant growth hardiness and supporting the physiological processes necessary for winter hardening[27].

This research aims to develop an advanced field assessment and visualization tool designed for Canadian alfalfa growers. By integrating field measurements, soil and topographic data, and dynamic management practices, this newly developed tool will address regional agroclimatic challenges. It seeks to enhance winter survival and plant persistence by identifying key risk factors and optimizing field conditions. By combining predictive modelling with practical management insights, the tool will improve winter damage assessment and mitigation strategies. Ultimately, this system will enhance alfalfa resilience and productivity, providing growers with data-driven support for informed decision-making and promoting more sustainable, region-specific agricultural practices.

Materials and methods

Probability emerges as a crucial tool for evaluating and responding to winter survival challenges on Canadian farms. This method involves continually updating beliefs about an event as new field data becomes available, facilitating the adjustment of field practices and management decisions based on both fresh information and existing knowledge[12,28]. To implement this, a similarity matrix was employed to construct these distributions across distinct field scenarios, following Wisconsin recommendations[1], and leveraging dynamic field records and user input. The similarity matrix, a nonparametric weighted data visualization tool, integrates user-defined parameters derived from expert systems, such as the Wisconsin scoring system.

The Wisconsin scoring system, developed by the agronomy department at the University of Wisconsin, USA[3], serves as a robust method for assessing winter survival. This scoring method serves as a foundational element for evaluating alfalfa fields and selecting model input parameters, categorizing them into three major groups: soil characteristics, weather conditions, and field management practices. The assessment model integrates field data with ancillary information, such as soil texture, topography, historical crop records, nutrient profiles, and weather data (temperature and rainfall). The Wisconsin scores, combined with data analytics for numeric simulation, were elucidated through a flow diagram (Fig. 1). Ultimately, this approach yields multiple scenarios to inform management decisions aimed at enhancing the winter survival of alfalfa crops.

Model predictions were dynamically adjusted using a data-driven probability framework, in which outcome likelihoods (e.g., stem count categories) were estimated from the frequency of similar cases in the historical dataset. This non-parametric, similarity-weighted approach enabled flexible updates to predictions as new information on soil nutrients, weather, or management practices became available. Seasonal stem count probabilities were visualized using stacked bar graphs, while category-wise distributions across scenarios were shown using standard bar charts. Together, these tools quantified the effects of interacting factors on alfalfa winter survival, providing reliable decision support for site-specific forage management.

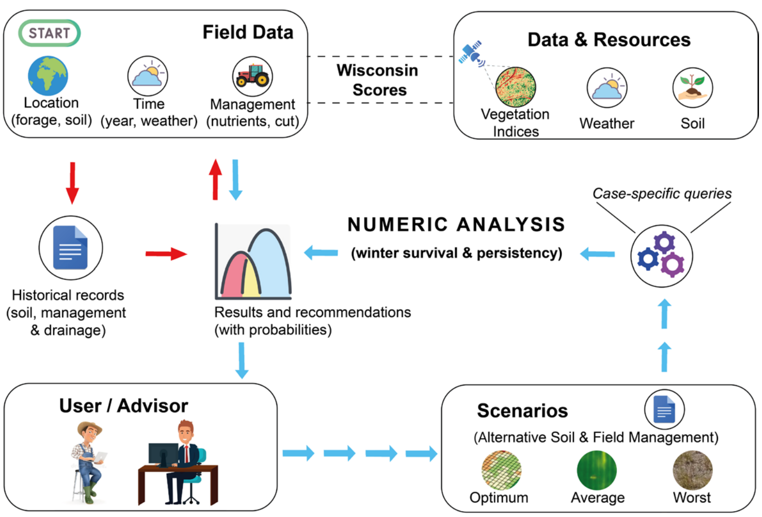

Fig. 1. Flow diagram illustrating the steps of the decision support tool. The tool seamlessly integrates field data with online resources, merging historical records and management parameters evaluated through Wisconsin scores for numerical analysis. The blue arrows represent outputs, queries, and feedback loops and the red show how raw and historical data flow toward the analysis engine. The resultant scenarios generated helped refine management decisions to enhance alfalfa's winter survival.

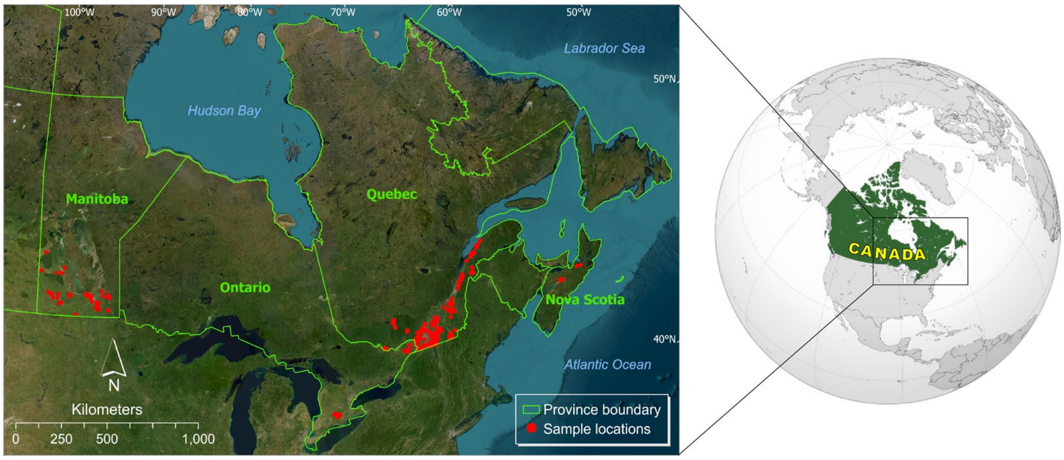

Fig. 2. The study area and data collection sites, marked by red dots, from four Canadian provinces. These represent the locations for alfalfa stem counts and soil sample collections conducted in both spring and fall from 2021 to 2023.

Provinces	Sample locations	Total fields	Total farms	Number of producers	Number of advisors
Quebec	1498	501	174	118	38
Ontario	125	15	12	9	4
Nova Scotia	34	4	4	4	3
Manitoba	448	46	35	35	11
Total	2105	566	225	166	56

Table 1. Information on data collection (i.e., alfalfa stem counts and soil sampling) conducted in four provinces in Canada.

Study sites
Data and samples were collected from 566 fields located on 225 farms located in four provinces (Nova Scotia, Quebec, Ontario, and Manitoba; Fig. 2; Table 1). Most fields were well-drained, having both surface and subsurface drainage systems, and contained various mixtures of alfalfa and perennial grasses.

Field measurements
Alfalfa stem analysis
A hierarchical sampling framework was implemented across four Canadian provinces to assess spatial and temporal variability in alfalfa productivity[29]. Fifty-six farm advisors, selected for their experience with fields of varying productivity, identified 566 fields across 225 farms, ensuring accessibility for repeated field surveys and long-term collaboration. Vegetation indices, primarily the Normalized Difference Vegetation Index (NDVI), derived from Sentinel-2 multispectral imagery, were used to assess within-field variability and guide the placement of sampling locations. A total of 2,105 GPS-referenced sampling locations were established. Each sampling location consisted of three closely spaced replicates, referred to as landmarks. The number of sample locations varied based on field size and productivity gradients observed in previous growing seasons. Alfalfa stem counts were recorded using permanent 30 × 30 cm rectangular quadrats at each landmark during spring (May to July) and fall (August to November) of 2021, 2022, and 2023. The sampling layout and quadrat placement are illustrated in Fig. 3. These seasons are hereafter referred to as spring 2021 (S0), fall 2021 (F0), spring 2022 (S1), fall 2022 (F1), and spring 2023 (S2).

Alfalfa stem counts were assessed biannually from spring 2021 to spring 2023 across four provinces: Quebec, Manitoba, Ontario, and Nova Scotia. The results (Table 2; Fig. 4) show distinct temporal and spatial variations. Stem counts demonstrated similar distributions between spring and fall (Fig. 4b and c, between the vertical dotted lines), with the majority ranging from 10 to 90, and the most frequent values between 20 and 50. Ontario recorded the highest mean stem count ($\mu = 57$), while Nova Scotia had the lowest ($\mu = 19$). The overall mean stem count declined from 49 in spring 2021 to 37 in spring 2023. Quebec exhibited the maximum observed count (234) in spring 2021 and 158 in fall 2021. Standard deviations for spring counts ranged from 25.9 to 29.9. In contrast, fall counts showed lower variability (Std. ≈ 25). Except for Nova Scotia ($\mu = 19$), all provinces maintained mean stem counts above 40, consistent with agricultural guidelines for productive forage stands. These findings reveal spatial variability and temporal decline in stem density, supporting winter survival modelling and region-specific management.

Fig. 3. Landmark configuration for alfalfa stem count sampling: **(a)** Example of a sampling location, which consists of three points (landmarks) spaced 1 m apart at a given field site; **(b)** measurement area delineated by a red rectangular quadrat placed at the corner of each location; **(c)** enumeration of individual plant stems conducted within each delineated quadrat area.

Landmark	Mean	Std.	Min	Max	Landmark	Mean	Std.	Min	Max
Spring 2021					Fall 2021				
S21_1	49.6	27.8	0.0	188	F21_1	47.4	24.9	0.0	152
S21_2	49.9	28.0	0.0	234	F21_2	47.2	24.9	0.0	158
S21_3	48.7	27.1	0.0	178	F21_3	46.0	24.6	0.0	145
Spring 2022					Fall 2022				
S22_1	44.8	29.9	0.0	210	F22_1	37.3	24.8	0.0	100
S22_2	43.5	29.0	0.0	175	F22_2	36.9	24.7	0.0	100
S22_3	43.4	29.2	0.0	151	F22_3	35.5	24.5	0.0	100
Spring 2023									
S23_1	36.9	25.9	0.0	100					
S23_2	37.5	26.8	0.0	100					
S23_3	36.4	25.9	0.0	100					

Table 2. Summary statistics of alfalfa stem count (stems per m^2) performed during three consecutive years in 566 fields across 225 farms in four Canadian provinces (i.e., Nova Scotia, Quebec, Ontario, and Manitoba), including mean, standard deviation (Std.), minimum (Min), and maximum (Max), were calculated. Data were collected from spring 2021 to spring 2023 across three landmarks, with each measurement taken at a 1-meter interval.

This variability assessment enabled the creation of persistence classes using the Wisconsin scoring system, grouping alfalfa fields into four categories based on stem count dynamics: strong crop (≥ 40 stems), poor crop (< 50 across seasons), uncertain crop (high temporal fluctuation), and lost crop (< 10 in any season). The assessment also supports the creation of various winter assessment scenarios. Temporal stem count patterns revealed anomalies with agronomic stress, providing a quantitative basis for winter risk evaluation and informed management.

Soil analysis

Equally, a total of 2,105 soil samples were collected in 2021 and 2022 from the exact location where plant counts were conducted to support parallel agronomic assessments. 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 laboratory analysis. Lab-measured soil properties 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 for pH, SOM, CEC, and ISP, all properties were analyzed by the Mehlich III extraction method[30]. Due to missing or invalid values at specific locations, only 1,715 out of 2,105 laboratory results were retained for analysis. Missing values at 390 sample points were imputed using the field average, assuming a normal distribution for each property (pH, SOM, P, and K).

Soil properties across various Canadian provinces (Quebec, Nova Scotia, Ontario, and Manitoba) exhibit distinct variations that could significantly influence agricultural practices and alfalfa growth in these regions. According to the various alfalfa assessment criteria used by the USA Extension Program[1], four land-measured soil properties, pH, SOM, P, and K, were used for the model. Soil chemical analysis was compared across four provinces, despite more soil samples from Quebec being analyzed than from Nova Scotia.

The laboratory results from soil samples collected as part of this study were directly incorporated into the development of a predictive model to assess alfalfa winter survival. This analysis revealed substantial variability

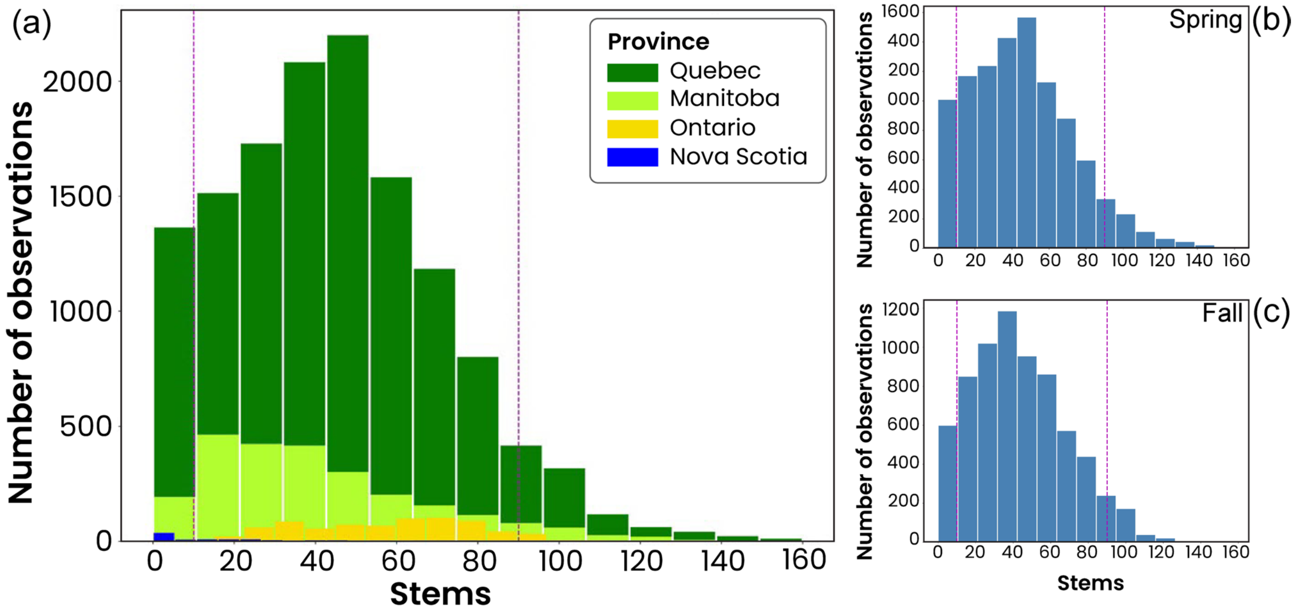

Fig. 4. Alfalfa stem counts recorded at multiple Canadian locations from 2021 to 2023, with generalized average values ranging from 10 to 90 stems per m², indicated by the vertical dotted line. The data are presented by province **(a)**, and by seasonal distribution: spring **(b)** and fall **(c)**.

	SOM				pH				*P* (kg/ha)				K(kg/ha)			
Provinces	Mean	Std.	Min	Max	Mean	Std.	Min	Max	Mean	Std.	Min	Max	Mean	Std.	Min	Max
Quebec	5.15	2.97	1.5	57.1	6.25	0.42	4.9	7.7	130.76	105.14	11.02	1110.98	284.72	174.21	44.08	1886.36
Nova Scotia	4.30	1.94	1.7	9.3	6.44	0.42	5.5	7.1	154.18	87.69	35.06	381.68	272.45	87.67	144.26	555.99
Ontario	4.72	1.71	2.1	12.2	6.53	0.65	5.1	7.5	122.52	103.51	26.05	588.05	319.70	122.42	107.19	717.28
Manitoba	5.77	2.57	1.6	14.6	6.96	0.75	4.6	8.1	130.09	111.57	14.03	664.18	754.69	537.59	80.14	3000.35

Table 3. Four key soil chemical properties—pH, soil organic matter (SOM), phosphorus (P), and potassium (K)—were measured for alfalfa, with their variation analyzed across four Canadian provinces: Quebec, Nova Scotia, Ontario, and Manitoba.

in soil parameters across different provinces, as evidenced by the range, standard deviation (σ), and mean (μ) of each parameter (Table 3).

Across regions, pH ranged from 4.6 to 8.1 (Table 3), with Manitoba showing the widest variation and more alkaline soils, while Quebec remained more acidic ($\mu = 6.25$; min = 4.9). Soil pH values showed moderate variability (μ 6.25–6.96; Std. 0.42–0.75) and remained within agronomically relevant ranges, with no evidence of extreme skewness. These characteristics support the use of field-level mean imputation for missing values without substantially affecting model outcomes. In worst-case conditions, pH dropped to 5.0, below the optimal range of 6.5–8.0 (Bélanger et al., 2006)[2]. Potassium (K) levels varied from 272 kg ha^{-1} in Nova Scotia to 755 kg ha^{-1} in Manitoba, with deficiencies (< 180 kg ha^{-1}) increasing winterkill risk. Soil organic matter (SOM) ranged from 1.5 to 57.1%, with Manitoba showing the highest mean ($\mu = 5.77\%$). Phosphorus (P) was relatively stable, ranging from 123 to 154 kg ha^{-1}, supporting the modelled optimal scenario of $P = 400$ kg ha^{-1}.

Auxiliary data

According to regional guidelines, soil nutrients and moisture, fertilizer application, cultivar, seeding rate, harvesting time, and drainage conditions are all factors that play a role on alfalfa winter survival[31], thus information on soil texture, topographic datasets, historical crop records, available nutrients, weather data (including temperature and rainfall), and field management practices were also recorded for each field. Field management practices recorded included cropping practices (crop history and age), N-P-K application, and drainage quality. These data served as preparatory inputs within the modelling framework, and their processing, including visualizations and statistical analyses, enabled effective variable screening and model tuning to identify the most influential predictors.

Topographic data

Topographic parameters, including the topographic wetness index (TWI), slope, and aspect ratio (Fig. 5), were calculated from a high-resolution digital elevation model (HRDEM) derived from light detection and ranging

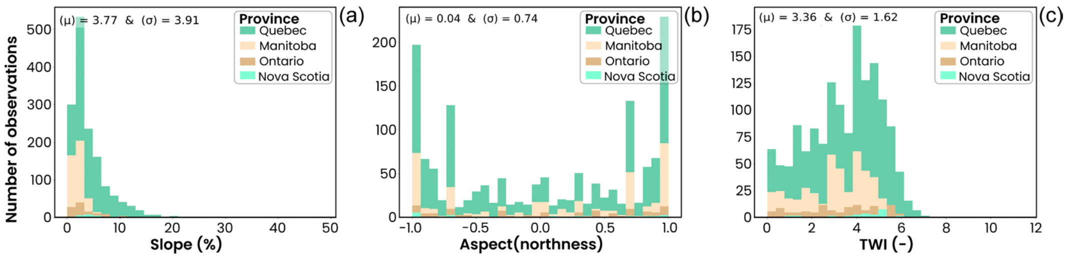

Fig. 5. Topographic parameters derived from a digital elevation model illustrating variability across four Canadian provinces. The parameters shown are **(a)** slope (percentage), **(b)** aspect (northness) (+/-), and **(c)** topographic wetness index (TWI) (-).

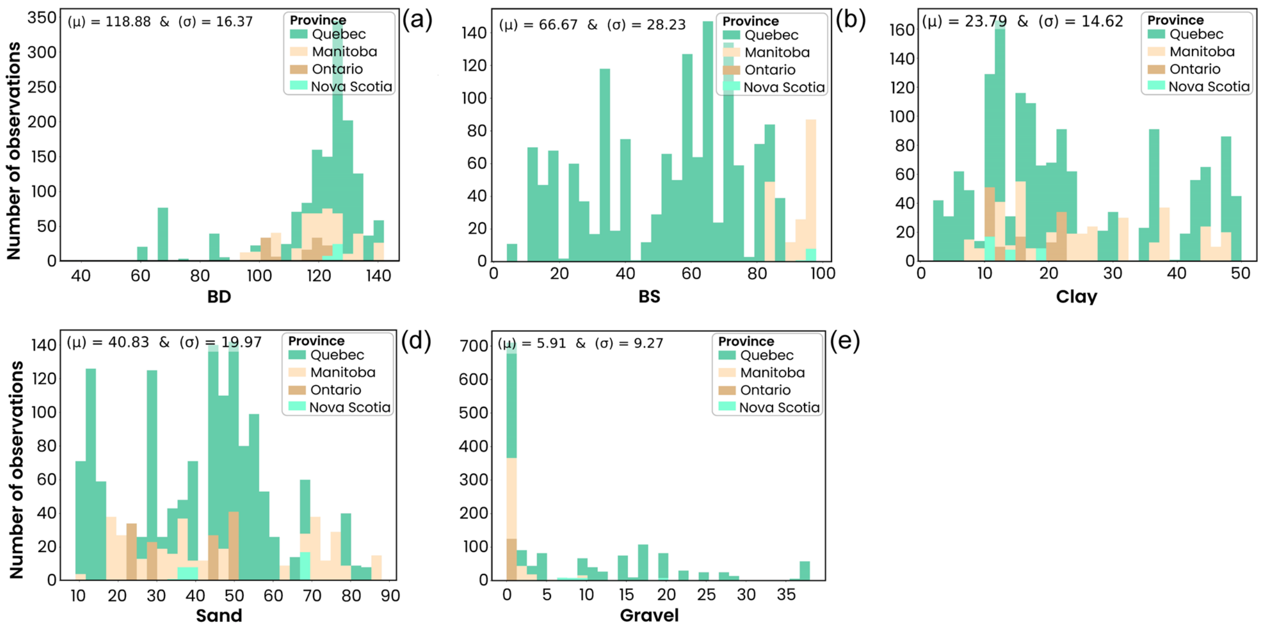

Fig. 6. Soil texture parameters for fields across four Canadian provinces—Quebec, Manitoba, Ontario, and Nova Scotia—illustrating **(a)** soil bulk density (BD): 0–120 cm (g/cm^3), **(b)** base saturation (BS): 0–120 cm (%), **(c)** clay content: 0–120 cm (%), **(d)** sand content: 0–120 cm (%), and **(e)** gravel content: 0–120 cm (%). Histograms illustrate regional variability, with Quebec appearing dominant due to a higher number of samples, while province-specific parameter ranges are preserved through scaling and weighting.

(LiDAR) data. Slope was derived from the altitude ranging from 6 to 560 m, and the digital elevation model (DEM), and subsequently classified into five categories (1 to 5 ranging from 0.57 to 48.56% of slope) based on the standard deviation of the observed data (Fig. 5a). Aspect refers to the compass direction of a slope. It can be derived from trigonometric functions applied to the slope angle. Aspect was calculated as a continuous variable by computing the cosine, referred to as "northness" of the farm landscape, which ranges from 1 (northward) to -1 (southward) (Fig. 5b). TWI was derived by utilizing the flow direction of the slope (Fig. 5c), as previously described[32].

$$\mathrm{Aspect}_{\cos} = cos\left(\frac{aspect \times \pi}{180}\right) \tag{1}$$

Where, Aspect$_{\cos}$ is the cosine value of the aspect angle in degrees from 0 to 360 (North) in the clockwise direction.

Soil texture
Land suitability and soil texture maps provided by Quebec's Research and Development Institute for the Agri-environment (IRDA) offer valuable information crucial for farm-level decision-making. According to regional guidelines, our approach incorporates base saturation (BS), bulk density (BD), and percentages of sand, gravel, and clay (Fig. 6). Gravel content (% by volume of particles ≥ 2 mm) was included as a key soil physical property due to its influence on water-holding capacity, and alfalfa stand longevity across variable field conditions. Soil

sampling was conducted at various depths, reaching a maximum of 120 cm. This study employed a reference point-based extraction method to obtain these parameters for each field.

Soil physical properties varied substantially across the four provinces. Bulk density averaged 118.9 g cm^{-3}, typically ranging from 100 to 140 g cm^{-3} (Fig. 6a). Base saturation (BS) showed a mean of 66.7% with high variability (Std. = 28.2%) (Fig. 6b). Clay content averaged 24% (Std. = 14.6%), sand 41% (Std. = 19.9%), and gravel 5.9% (Std. = 9.3%), with Quebec exhibiting the widest variation (Fig. 6c, 6d, and 6e). Texture values range from 2% to 57% (optimal ≈ 50%) for clay, up to 80% (optimal ≈ 35%) for sand, and 20% to 80% (optimal ≈ 80%) for BS, highlighting the firm regional heterogeneity that influences alfalfa growth potential.

Weather data

Historical daily weather data, including maximum and minimum temperatures and cumulative rainfall, were retrieved from the high-resolution 1-km daily meteorological dataset (Met1km), which is downscaled from the WorldClim2 dataset in Canada[33]. Growing Degree Days (GDD), an indicator of heat accumulation, are utilized in computing weather indices to determine optimal harvesting dates of each field and track the various growth stages of alfalfa[34]. This approach incorporates several weather indices as input parameters, including GDD at the first cut, total precipitation, GDD before fall/winter kill, GDD at the last fall cut, daily average fall rainfall, and GDD-based plant hardiness (Fig. 7).

$$\text{GDD5} = (\text{T}_{max} + \text{T}_{min}) \ / \ 2 \ -- \ \text{T}_{base} \tag{2}$$

Where, GDD5 is the growing degree days with base 5 °C. T_{max} is the daily maximum temperature (°C). T_{min} is the minimum temperature (°C). T_{base} is base temperature (°C), usually 5 °C for cool-season crops like alfalfa.

Climatic variables showed distinct regional patterns across the provinces. The GDD uncovered for the first cut was highest in Quebec, with a mean around 380 °C days, indicating early season growth conditions. GDD accumulation before the last fall cut showed a broader distribution (mean ~ 657 °C days), with Manitoba's data clustered more tightly around higher values (~ 800 °C days), suggesting extended growing seasons. GDD before fall/winter kill had a lower overall mean (~ 141 °C days), with Quebec contributing most data points under 200 °C days. In terms of GDD hardiness, values were centered around 125 °C days, with dense clustering between 100 and 150 °C days across Quebec and Manitoba, reflecting the thermal accumulation for cold acclimation. Winter rainfall totals averaged 23 mm, mostly from Quebec, and showed a skewed pattern with a long tail, while average daily fall rainfall peaked around 0.28 mm, again led by Quebec observations. Nova Scotia presented fewer data points but showed higher daily fall rainfall.

Management practices

In this study, field management data collected included cropping practices (such as cropping history and age), fertilizer application (including nitrogen (N), phosphorus (P), and potassium (K)), and assessments of drainage

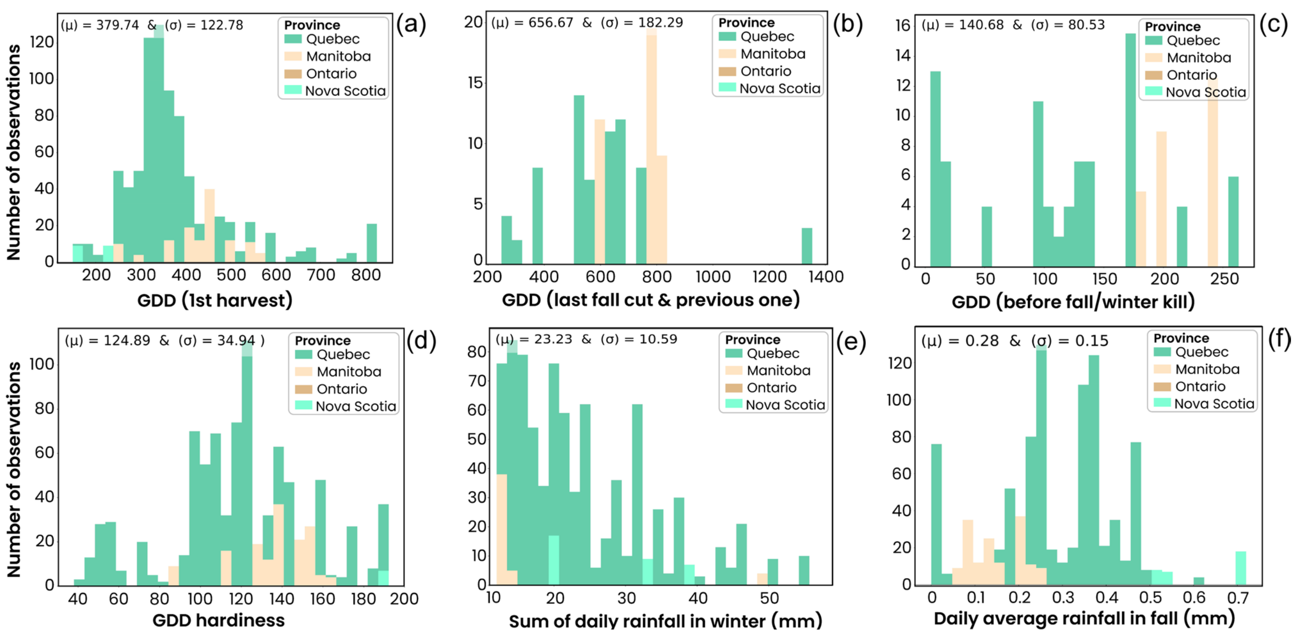

Fig. 7. Weather index calculated using Growing Degree Days (GDD), following a previously published formula[34]. This index assesses key agricultural conditions, including **(a)** the timing of the first harvest or cut, **(b)** the timing of the last fall cut and the previous one, and **(c)** pre-fall/winter hardening conditions, as well as **(d)** GDD-based plant hardiness factors. Additional rainfall parameters are included in the index, such as **(e)** total winter rainfall in mm and **(f)** daily average fall rainfall in mm.

quality. Those data were recorded after consulting with participating agricultural producers and agricultural advisors. Soil history and nutrient profiles, including manure, fertilizer (N-P-K), and lime applications at the field scale, were comprehensively documented from 2018 to 2021 and served as supplementary data for total N-P-K assessment. Additionally, seed rates and cultivar data were collected from the study farms and used for a critical analysis of winter persistence.

Total N-P-K

A detailed calculation of the nutrient budget, specifically nitrogen, phosphorus, and potassium (NPK), was conducted by incorporating data on existing soil nutrients along with manure and chemical fertilizer applications for each production year (Fig. 8). The analysis results aim to enrich the analytical depth of the model. The calculated positive nutrient budget was used as an input feature within the model. This involved multiple rounds of detailed NPK calculations and a comparative analysis of the model's outputs. Ultimately, a positive NPK strategy was adopted. This approach considers only the aggregate NPK levels ascertained through soil tests, as well as the NPK quantities derived from accumulated applications of both chemical and manure fertilizers across each production cycle and annually. It is essential to note that nutrient depletion factors were excluded from the calculation of the positive soil nutrient budget. While such depletion factors are conceptually quantifiable, they vary widely by region, soil type, crop growth stage, and environmental conditions, and require detailed site-specific calibration or modelling (e.g., process-based nutrient balance models). This comprehensive nutrient budget analysis was performed for each year of the experimental period, spanning the first two production cycles from 2021 to 2022.

Drainage quality

Practitioners evaluated the drainage quality for each farm and subsequently categorized it for inclusion in the model. This assessment was based on a combination of historical records of surface drainage conditions and GPS-verified positioning for each field, complemented by the producer's responses regarding the type of drainage present. This qualitative evaluation was then converted into numeric values for model input parameters, with "Well drained" assigned a value of 1, "Excessively drained" a value of 2, and "Poorly drained" a value of 3.

Seeding rates and cultivar

Although seeding dates and rates for each season were documented, a consistent seeding rate was applied across all fields, with rates ranging from 5 to 50 kg/ha to meet the seasonal productivity needs of the alfalfa. This data was used solely to support discussions regarding the model's application and was not directly integrated into the model itself. For cultivating alfalfa across all four provinces, cultivars selected were both winter-hardy and highly resistant to disease.

Data preprocessing and outlier detection

Timestamps, seasonal variations, geographic locations, and spatial distances were evaluated during preprocessing. Outliers were identified through latitude and longitude, null values, inconsistencies in field IDs, and comments from producers/advisors. Data filtering involved descriptive statistics, histograms, and distribution curves.

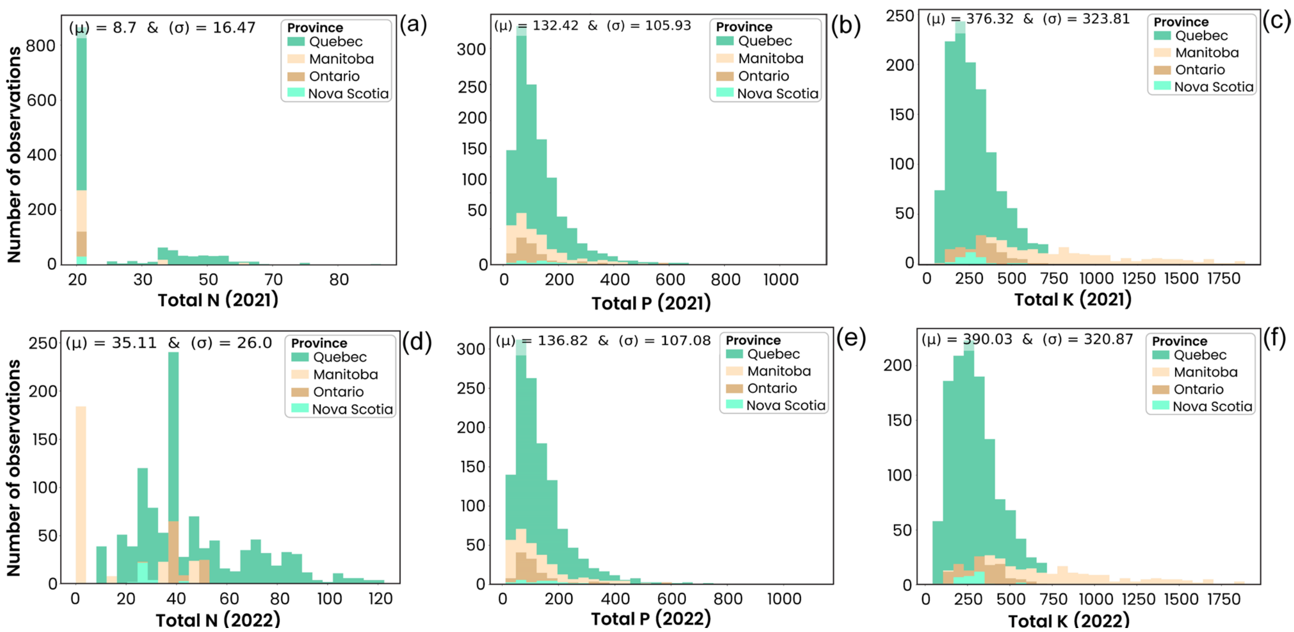

Fig. 8. Nutrient budget for nitrogen, phosphorus, and potassium (N-P-K), integrating data on existing soil nutrients and manure and chemical fertilizer applications for the 2021 production season (**a**, **b**, **c**) and the 2022 production season (**d**, **e**, **f**).

Following outlier removal, additional geospatial features were integrated. Tabular data without geographic references was georeferenced using ArcGIS Pro/Online by assigning coordinates to each sampling location. These spatially referenced data were then organized by region. Stem count statistics were calculated at the landmark level, allowing for detailed spatial and temporal analysis across seasons and years, which is essential for detecting localized patterns and informing region-specific alfalfa management strategies.

Descriptive statistics and variable selection

To assess regional patterns in alfalfa stem counts, summary statistics (mean, standard deviation, minimum, and maximum) were calculated by province. A stacked histogram illustrated the distribution of observations across Quebec, Ontario, Manitoba, and Nova Scotia, highlighting spatial variability. Environmental variables were incorporated into the model's input and training datasets. A custom Python module was developed for preprocessing, outlier detection, and statistical analysis, utilizing open-source libraries such as Pandas, GeoPandas, NumPy, and Matplotlib[35]. Model parameters were derived from both field and online sources. Table 4 summarizes the dataset types, variable descriptions, and final statistical outputs.

Parameters	Variable type & description	Mean	Std.	Min	Max
Soil texture					
BD	Continuous: Soil Bulk density: 0–120 cm (g/cm3)	118.88	16.37	38.00	142.00
BS	Continuous: Base saturation of soil: 0–120 cm (%)	66.67	28.24	4.00	100.00
Clay	Continuous: Clay percentage in soil: 0–120 cm (%)	23.79	14.63	2.00	57.00
Gravel	Continuous: Gravel percentage in soil: 0–120 cm (%)	5.91	9.27	0.00	38.00
Sand	Continuous: Sand percentage in soil: 0–120 cm (%)	40.83	19.98	5.00	88.00
Soil properties					
SOM	Continuous: Soil organic matter (%)	5.21	2.83	1.50	57.10
pH	Continuous: Soil pH	6.39	0.58	4.60	8.10
P	Continuous: Soil phosphorus (kg/ha)	130.48	105.93	11.02	1110.98
K	Continuous : Soil potassium (kg/ha)	369.42	325.45	44.08	3000.35
Nutrients available					
Applied_K_21	Continuous: Applied potassium (kg/ha) of 2021	6.90	19.20	0.00	110.00
Applied_N_21	Continuous: Applied nitrogen (kg/ha) of 2021	8.70	16.48	0.00	93.14
Applied_P_21	Continuous: Applied phosphorus (kg/ha) of 2021	1.94	7.83	0.00	82.00
Total_K_21	Continuous: Total potassium of 2021	376.32	323.90	44.08	3000.35
Total_N_21	Continuous: Total nitrogen of 2021	8.70	16.48	0.00	93.14
Total_P_21	Continuous: Total phosphorus of 2021	132.42	105.96	11.02	1110.98
Applied_K_22	Continuous: Applied potassium (kg/ha) of 2022	13.71	21.24	0.00	90.00
Applied_N_22	Continuous: Applied nitrogen (kg/ha) of 2022	26.40	18.40	0.00	82.80
Applied_P_22	Continuous: Applied phosphorus (kg/ha) of 2022	4.40	9.54	0.00	56.25
Total_K_22	Continuous: Total potassium of 2022	390.03	320.96	44.08	3000.35
Total_N_22	Continuous: Total nitrogen of 2022	35.11	26.01	0.00	122.54
Total_P_22	Continuous: Total phosphorus of 2022	136.82	107.11	11.51	1120.48
Topography					
Slope_class	Categorical: Slope categories 1–5 from slope (%)			1	5
Aspect(northness)	Continuous: Aspect (cosine): northness	0.04	0.74	-1.00	1.00
TWI (-)	Continuous: Topographic wetness index	2.89	1.62	0.00	11.00
Weather					
GDD at first Cut	Continuous: Degree days cumulated at the first cut	379.74	122.85	122.31	826.12
Seasonal precipitation	Continuous: Sum of daily rainfall in winter (mm)	23.23	10.60	4.62	56.80
GDD before fall/winter kill	Continuous: Degree days before fall/winter kill	140.68	80.88	4.31	263.17
GDD last fall Cut	Continuous: Degree days of the last fall cut	656.67	183.01	247.48	1353.18
Daily Av. Fall rain	Continuous: Daily average rainfall in fall (mm)	0.28	0.15	0.00	0.72
GDD plant hardiness	Continuous: Degree days of hardiness	124.89	34.96	38.01	203.00
Drainage quality					
Drainage category	Categorical: Categories of the drainage quality 1–3			1	3

Table 4. Soil chemical properties, texture, soil nutrients and their availability, topography, agroclimatic weather indices, and drainage quality were evaluated for model development. The table presents the description and types (continuous and categorical) of the input variables, along with their summary statistics, including mean, standard deviation (Std.), minimum (Min), and maximum (Max).

Data assessment by Wisconsin scoring

Wisconsin scoring and stem categories

Wisconsin scoring system, along with the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) winter assessment grid and the Centre de référence en agriculture et agroalimentaire du Québec (CRAAQ), was integrated to evaluate different field parameters and classify alfalfa stem counts (Fig. 9). This assessment grid named 'Expert System' and their scoring scales diverge across the three grid systems. The stem counts were categorized according to the parameters and scores advised by each framework. Many regions also used soil and stem characteristics to assess winter risk in their grids.

In this study, the initial field assessment scores were evaluated against suitable parameters across all agro-ecological zones to assess potential risk status. The assessment considers various controllable factors, including alfalfa stand age, soil pH and potassium levels, number and frequency of harvests, soil moisture levels, drainage quality, cultivar selection, winter hardiness, and disease resistance scores. The missing parameters of each field were assessed based on all available factors (Fig. 9a). This system categorizes persistence into four categories, each with a weighted score ranging from 3 to 33. Using these weighted scores, the Expert System assigns corresponding risk levels of very high (17–33), high (13–17), moderate (8–12), and low (3–7) to support targeted management decisions. In this study, four persistence classes were defined as follows: strong (stem count > 40), poor (stem count < 50), uncertain (variable values across seasons), and lost (stem count < 10) (Fig. 9b).

Wisconsin scoring and persistency classes

According to the University of Wisconsin Extension[26], the expert system provides two key guidelines for evaluating alfalfa's persistency, potential yield, and risk of winter injury. The first guideline is based on stem density, recommending that producers maintain crops with a stem density above 55 (as this density does not limit yield) and consider reseeding when density falls below 40 (as this severely limits yield). The second guideline utilizes a scoring system ranging from 3 to 33, as discussed in the methodology section. This scoring system categorizes risk into four levels: (i) scores of 3 to 7 indicate low to below-average risk, corresponding to our classification of a strong crop; (ii) Scores of 8 to 16 denote average risk, aligning with our classification of uncertain stem count; (iii) scores of 17 to 27 suggest a high average risk, applicable to our classification of poor crop; and scores of 28 or higher represent a very high risk, indicating a significant likelihood of crop loss. While this scoring system serves as an early warning tool, our model is further enhanced by incorporating more detailed seasonal or temporal variations. This enhancement would improve the model's utility in predicting alfalfa's persistence and managing risks associated with winter survival.

Probability distribution model and winter assessment

Similarity index

The similarity approach is optimal for extensive datasets. Similarity indices, as demonstrated by their widespread application in machine learning, data mining, and recommendation systems across various domains, including the Google search engine[36], play a crucial role. One strategy involves calculating a similarity index for all records in the dataset and incorporating it into the modelling process. The value in each cell of the matrix quantifies the degree of similarity between the corresponding pair of objects. This method prioritizes records with higher similarity during model evaluation, enabling the computation of distances. This distance measures how closely a database record, a trial, aligns with the conditions provided by the user, which are also referred to as features, inputs, or parameters[21]. For the analysis, the model incorporated all user input values, which were derived from databases that record field variability in soil, weather, and field management, as detailed in Table 5. Additionally,

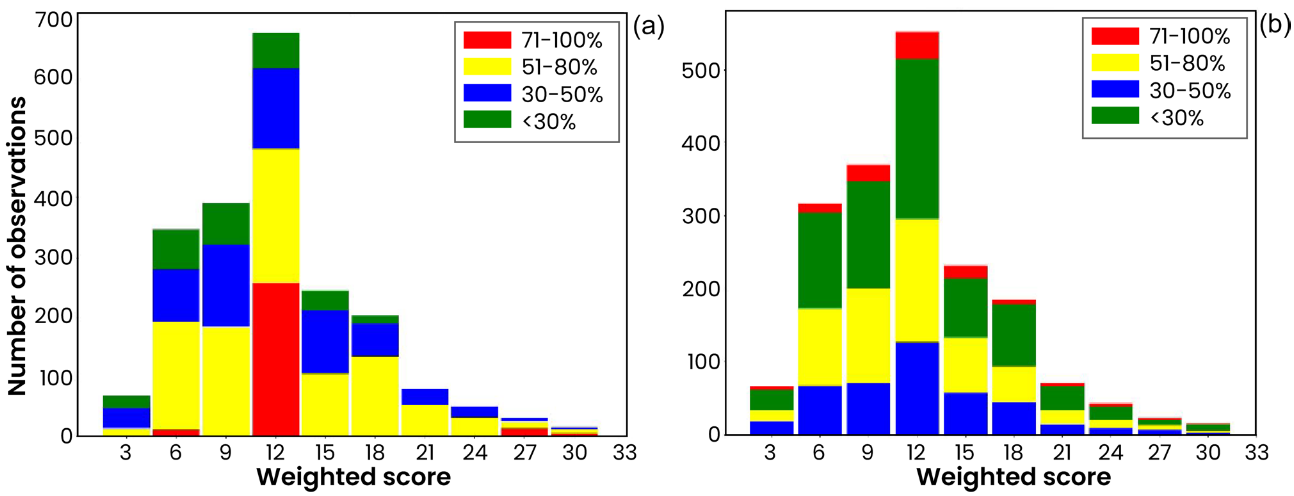

Fig. 9. Field management data assessment and persistence classifications using Wisconsin scores: **(a)** analysis of the percentage of missing data for management parameters required to compute Wisconsin weighted scores across multiple fields, and **(b)** categorization into four persistence classes, with each class assigned its corresponding Wisconsin score.

Parameters	Optimum case (I)	Average case (II)	Worst case (III)
Soil texture			
BD (g/cm3)	100	119	142
BS (g/cm3)	80	60	20
Clay (g/cm3)	20	10	57
Gravel (g/cm3)	5	15	35
Sand (g/cm3)	35	50	80
Soil properties (lab analysis)			
SOM (%)	5	5	2
pH	7	6	5
P (kg/ha)	400	200	30
K (kg/ha)	1500	300	150
Nutrients available			
Applied_K_21 (kg/ha)	100	40	20
Applied_N_21 (kg/ha)	60	40	10
Applied_P_21 (kg/ha)	50	10	30
Total_K_21	2000	800	100
Total_N_21	60	40	10
Total_P_21	600	200	100
Applied_K_22 (kg/ha)	80	40	10
Applied_N_22 (kg/ha)	60	40	10
Applied_P_22 (kg/ha)	30	25	10
Total_K_22	2000	800	100
Total_N_22	100	60	20
Total_P_22	600	300	100
Topography			
Slope_class (%)	1	2	5
Aspect(northness)	-1	0	1
TWI (-)	2	3	4
Weather			
GDD at 1 Cut	500	20	300
Precipitation in winter (mm)	40	25	10
GDD before fall/winter kill	250	150	50
GDD last fall Cut	800	600	200
Daily av. Fall rain (mm)	0.10	0.30	0.60
GDD plant hardiness	200	100	50
Drainage quality			
Drainage category	1	2	3

Table 5. User-defined input values for model parameters, derived from field data and expert knowledge, for three arbitrary scenarios—optimum, average, and worst. These scenarios were used to estimate alfalfa winter survival probability and to assess the model's performance in each case.

the Wisconsin Scoring Guide and regional guidelines from Ontario and Quebec, Canada, were utilized to validate the user input values for the test scenarios.

When selected features of database records contain missing values, one method for handling these gaps is to randomly insert new values derived from the population distribution specific to each feature. Although numerous methods exist to impute missing values with synthetic data, the integrity of the results may be compromised if a substantial number of values are missing for a given feature. In this study, records were excluded if they lacked most of the data points for the similarity assessment feature in any given field. Furthermore, additional parameters were not considered because field-specific data for pH and Soil Organic Matter (SOM) were largely unavailable, as they were absent from the soil lab analysis results during the similarity assessments. Aside from these exclusions, most parameters were incorporated, even when some data points were missing. This inclusion was achieved through a straightforward weighted method, which calculated the ratio of available parameters to the total parameters to adjust the similarity index. For the subsequent numerical modelling analysis, a Python module integrating several libraries (openpyxl, matplotlib, numpy, and scipy) was developed.

To simplify the presentation, let's assume that the record representing the user context record is u. The similarity λj, u between any j_{th} record (field measured stem count) and the user-specified record u (user context parameters/scenarios) for all k features can be calculated as:

$$\lambda_{j,u} = \prod_{k=1}^{K}\left(1 - \delta\lambda_k\left|\frac{x_{k,j} - x_{k,u}}{x_{k,\ max} - x_{k,min}}\right|\right)^q \cdot \frac{Np}{Nt} \tag{3}$$

Where,

$\delta\lambda k$ is the percentage of 0 to 1 range (weight) affiliated with the *kth* feature, indicating its importance for similarity. *Xk*, *j* is the value of the *kth* feature of record *j* (field estimation in the database), *xk*, *u* is the value of the *kth* feature of record *u* (representing the user context) and *q* is the power of similarity (high value reduces the influence of records that do not match user inputs identically). *Np* is count of available parameters where values are present. *Nt* is count of total parameters used in the model.

Alfalfa stem count class and error distribution

According to the calculated statistics, human and recording errors in the stem count values were minimized. To determine landmark-specific variability and classify measurement errors, the actual stem count values were group into 10-unit increments. All errors were computed for each reference point across the five seasons, and the mean and standard error were subsequently calculated. Finally, the following formulas were applied for probability calculations.

Errors to probability approach

$$\varepsilon_{i,j} = (Y_i - Y_j)\,.\lambda_j \tag{4}$$

Where,

$\varepsilon_{I,\,j}$ is the weighted stem estimation error for a combination of model *ith* predictive model and *jth* observed values, *Yi* is the predicted stem count (by category) estimate for *ith* model (derived from Eq. 3), *Yj* is the actual stem count (average over field replicates) for *jth* record in the database, *λj* indicates similarity coefficient between the user-defined scenarios and the *jth* record in the observed value. This scalar (ranging from 0 to 1) reflects how closely the *jth* database entry matches the current user-defined scenario in terms of key environmental and management factors.

Further, the distribution of estimation errors was used to calculate the probability of zero deviation between the model-predicted and observed stem counts. This was done by applying the normal probability density function as described previously[37], for each modelled scenario, treating the estimated stem count as the output of the '*NumericAg*' model:

$$f(Y)_i = \frac{e^{\frac{(\varepsilon - avg\ (\varepsilon_i))^2}{2*std\ (\varepsilon_i)^2}}}{\sqrt{2*\pi*std\ (\varepsilon_i)^2}} \tag{5}$$

Where, f(Y)i is the probability density of 0 from the frequency of errors at the *ith* model calculated by providing the average *avg (εi)* and standard deviation of errors *std (εi)*.

Additionally, it is needed to standardize the raw probability values so that the sum of *p(Y)I* = 1:

$$p(Y)_i = \frac{f(Y)_i}{\sum_{i=1}^{N} f(Y)_i} \tag{6}$$

Where, *p*(*Y*)*i* is the normalized probability (i.e., the probability assigned to the model *i*) after standardization, ensuring that the total probability sums to 1 across all *N* models, and $f(Y)_i$ is the raw likelihood or score assigned to the predicted stem count Y_i for the *i*th model. This normalization step facilitates direct comparison among models and aligns with standard probabilistic modelling practices.

Results
Model illustration outcomes

Table 5 summarizes three probability distribution scenarios for alfalfa stem counts under varying soil, weather, and management conditions. The optimum scenario (I) assumes favourable conditions, and shows a 60–80% probability of high stem counts (80–100). The average scenario (II) reflects moderate fertility, mixed weather conditions, and average management, with a 40–60% probability of stem counts ranging from 40 to 60. The worst scenario (III), representing poor soil, adverse weather, and suboptimal management, shows a 60–80% probability of low stem counts (0–20). These user-defined scenarios provide a statistical basis for assessing growth variability across environmental conditions. Stacked probability graphs (Figs. 10, 11 and 12, left panels) illustrate seasonal trends across five periods—spring 2021 (S0), fall 2021 (F0), spring 2022 (S1), fall 2022 (F1), and spring 2023 (S2)—using stem count categories from red (0 stems) to gray (100 stems), with probabilities scaled from 0 to 100%. Corresponding bar charts (right panels) depict the modelled probability distributions under each scenario, integrating topographic, climatic, and management influences. Together, these results demonstrate the model's capacity to simulate seasonal variability in winter survival and growth outcomes under real-world field conditions.

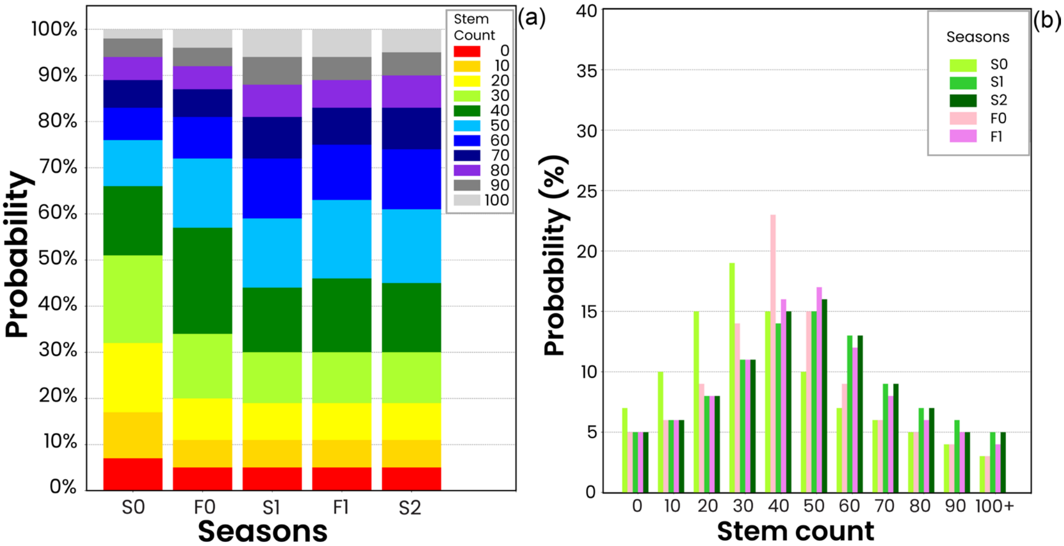

Fig. 10. **(a)** Stacked diagram showing the probability distribution of alfalfa stem counts under the optimum scenario across five seasons: spring 2021 (S0), fall 2021 (F0), spring 2022 (S1), fall 2022 (F1), and spring 2023 (S2), allowing seasonal comparisons. **(b)** Corresponding bar chart illustrating the probability distribution of stem counts (categorized from 0 to 100+) under varying field conditions across the same seasons.

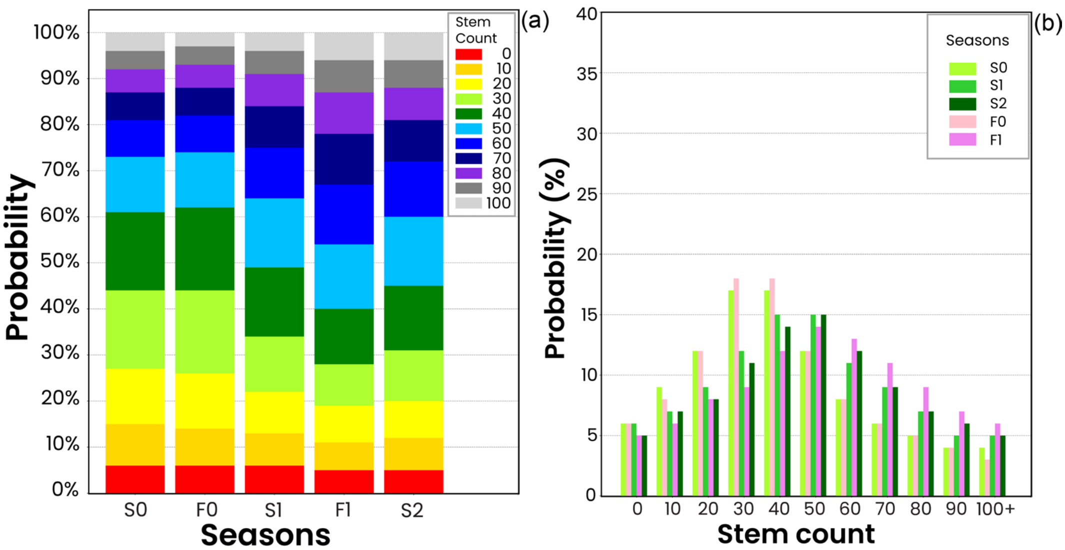

Fig. 11. **(a)** Stacked diagram showing the probability distribution of alfalfa stem counts under the average scenario across five seasons: spring 2021 (S0), fall 2021 (F0), spring 2022 (S1), fall 2022 (F1), and spring 2023 (S2), enabling seasonal comparison. **(b)** Corresponding bar chart illustrating the probability distribution of stem counts (categorized from 0 to 100+) under average growing conditions across different seasons.

Scenario I: optimum condition

Figure 10a shows the probability distribution of alfalfa stem counts across five seasons (spring 2021 to 2023) optimal growth conditions. High stem counts (80–100), represented by violet to grey segments, denote winter survival and appear most prominently in S1, where they comprise ~ 20% of the total. Mid-range stem counts (30–80, green to blue) dominate across all seasons, representing typical stands under moderate winter conditions. To identify high-performance periods, the upper stem counts (70–100) were compared across seasons. An increase in high stem-count segments in fall 2022 suggests improved survival, likely due to better field practices, milder winters, or the use of more resilient varieties. In contrast, smaller segments may indicate reduced plant survival. Overall, the distribution suggests that management practices implemented before spring and fall 2022 were the most effective in promoting alfalfa health post-winter.

Figure 10b shows the probability distribution of alfalfa stem counts across five seasons under optimal growth conditions. The analysis emphasizes stem count ranges of 30–60, associated with strong post-winter recovery.

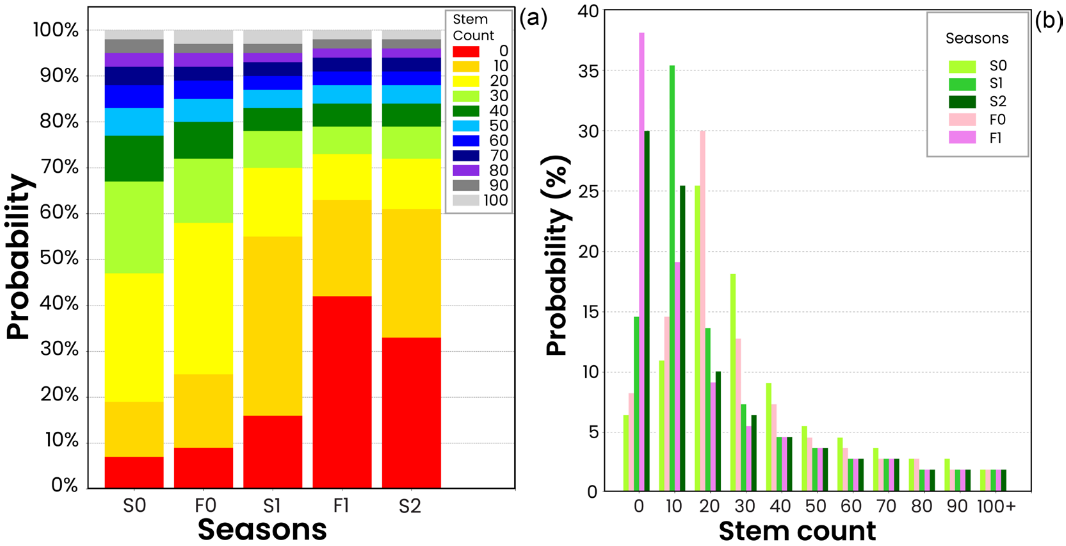

Fig. 12. **(a)** Stacked diagram showing the probability distribution of alfalfa stem counts under the worst-case scenario across five seasons: spring 2021 (S0), fall 2021 (F0), spring 2022 (S1), fall 2022 (F1), and spring 2023 (S2), allowing seasonal comparison. **(b)** Corresponding bar chart illustrating the probability distribution of stem counts (categorized from 0 to 100+) under poor growing conditions across different seasons.

Fall 2021 (F0) and fall 2022 (F1) show elevated probabilities in this range, with F1 reaching 12–17% for the 40–60 stem categories. The 40-stem category in F0 records the highest probability at 23%, identifying it as the most favourable condition. Mid-range stem counts (40–50) remain consistent across seasons, making them as optimal for winter resilience. Multiple peaks in F0 and F1 highlight favourable climatic, soil, and management conditions during fall periods, while lower and higher stem categories exhibit reduced probabilities. Overall, these results demonstrate that mid-range stem counts are the most probable outcomes under optimal growing conditions.

Scenario II: average condition

Figure 11a presents the probability distribution of alfalfa stem counts across five seasons under average field management conditions. Mid-range categories (40–60 stems; green to blue) indicate moderate plant health and winter resilience. Spring 2022 (S1) exhibits the highest probabilities within this range, followed by Fall 2021 (F0) and Fall 2022 (F1). These mid-range categories account for 30–60% of total probabilities, reflecting consistent but moderate field performance. When expanded to include 30–70 stems, probabilities range from 25% to 85% across seasons, confirming the dominance of mid-range outcomes. Overall, these results indicate stable alfalfa performance under average management and environmental conditions, with minimal extremes in growth or winter survival responses.

Figure 11b presents the probability distribution of alfalfa stem counts across five seasons (S0, F0, S1, F1, S2) under average conditions. Mid-range stem counts (40–60) dominate, representing typical crop health and winter survival. Spring and fall 2021 (S0, F0) show the highest probabilities within this range, while spring and fall 2022 (S1, F1) display distinct peaks with probabilities up to ~15%. Slightly higher values beyond 60 stems in fall 2022 indicate improved field outcomes. Overall, probabilities outside the 40–60 range remain low, confirming stable yet moderate alfalfa performance across seasons.

Scenario III: worst condition

Figure 12a shows the stacked probability distribution of alfalfa stem counts across five seasons (S0, F0, S1, F1, S2) under worst-case conditions. Low stem categories, 0 (red) and 10 (orange), indicate poor plant health and low winter survival. These segments dominate fall 2022 (F1) and spring 2023 (S2), reflecting heightened risk. Spring 2023 further confirms these concerns, with a cumulative probability of 72% for the 0 and 10 stem count categories, reflecting a high likelihood of winter damage. The 10-stem category alone peaks at 50% in spring 2022 (S1). These results identify fall 2022 and spring 2023 as critical periods associated with weak stands and elevated vulnerability to winter damage.

Figure 12b presents the probability distribution of alfalfa stem counts across five seasons (S0, F0, S1, F1, S2) under worst-case conditions, emphasizing low stem counts (0–20) associated with poor winter survival. Fall 2022 (F1) exhibits the highest probability for zero stems (42%), indicating severe post-winter damage. Spring 2022 (S1) peaks at the 10-stem category (38%), reflecting limited regrowth and reduced resilience. Spring 2023 (S2) also shows high probabilities for 0–10 stems, coupled with declining frequencies in the 30–80 stem range. In this scenario, stem counts above 40 are rare (< 5%), confirming generally weak stands. These distributions identify fall 2022 and spring 2023 as the most critical periods for crop failure, with low survival rates linked to adverse conditions. The data highlight key seasonal vulnerabilities in alfalfa persistence under suboptimal management and environmental stress.

Discussion

Across the four modelled locations, several key features demonstrated strong potential for predicting alfalfa survival. In Quebec, where most samples were collected, higher organic matter (SOM > 5%), moderate base saturation (BS ~ 70%), and well-drained clay-loam soils with subsurface drainage infrastructure were associated with higher stem counts. In contrast, Ontario fields—which were under-sampled—showed greater pH variability, with several sites below pH 5.5, which correlated with reduced survival. Nova Scotia exhibited low K and high soil compaction, and acidic soils (pH < 5.2) were frequent, which aligned with poor stem recovery in spring 2023. Among soil parameters, pH, soil organic matter (SOM), and potassium (K) emerged as the most influential. Fields with pH > 6.6 and SOM > 5% consistently showed higher stem count probabilities, reflecting reduced winter injury and enhanced resilience—particularly in Quebec and Manitoba. Potassium levels above 1500 kg ha^{-1} significantly improved freezing tolerance and root vigour[38,39], particularly important in fall-to-spring transitions (e.g., S1 to S2). In contrast, severely low pH (< 5.0) and K concentrations (< 200 kg ha^{-1}) at certain Ontario and Nova Scotia sites aligned with higher winterkill risk.

Topographic and management-related features also contributed to predictive strength. For example, north-facing slopes, elevated bulk density, and the absence of subsurface drainage were associated with poor survival in steeper or heavier soil regions[40,41]. The Topographic Wetness Index (TWI ≈ 2) and moderate clay content (~ 24%) emerged as key indicators of sites with balanced water retention and drainage[42]. Tillage and seeding practices further influenced outcomes: regions with consistent field management (e.g., drainage infrastructure in Quebec, consistent fall fertilization in Manitoba) showed notably better survival, reinforcing the predictive utility of these management variables[43].

Seasonal weather variables significantly influenced alfalfa winter survival across provinces. In spring, growing degree days (GDD) at first harvest emerged as a key predictor of early biomass accumulation, particularly elevated in Quebec and Manitoba. In the fall, GDD before the last cut and daily rainfall were important drivers of regrowth and moisture stress[44]. Variables such as GDD before winter kill, accumulated hardiness units, and total fall/winter precipitation were critical for overwinter survival. Under the optimum scenario, favorable thermal accumulation, balanced moisture levels, and timely harvests led to a high predicted probability of winter survival across both fall and spring seasons, reflecting strong stand resilience under well-managed and climatically suitable conditions. Collectively, these weather variables captured both the growing season dynamics and overwintering responses essential for accurately modeling alfalfa productivity under Canada's variable climatic conditions[45].

The *NumericAg* model integrated soil, weather, and management factors to assess their combined influence on alfalfa performance. It accounted for provincial patterns by adjusting parameter weights and applying within-field variability controls, as evidenced by consistent mid-range predictions (30–70 stem count class) under average case scenarios (Fig. 11). Higher stem counts (60–100) observed in season S2 are indicative of improved management or favourable environmental conditions. The model scenarios (optimum, average, and worst) were constructed with known expectations, with average conditions anticipated to fall between the extremes. While Quebec's over-representation helped establish a robust similarity-weighted baseline, regional differences in topography, aspect, and management practices remain critical for model adaptation. These results offer valuable preliminary insights, comprehensive model validation and broader applicability require extensive cross-validation using independent datasets.

While this version of the model employs probabilistic reasoning using similarity-weighted likelihoods (e.g., stem counts conditioned on soil and management features), future iterations will incorporate complete Bayesian updating, allowing model parameters to evolve as new seasonal data accumulate, thereby improving the precision of site-specific survival predictions. While this study focused on field and environmental variables, integrating historical and expert data could further reduce uncertainty. Future calibration against the USDA winter survival rating scale (1–6) may strengthen links between stem count probabilities and varietal resilience. The reliability of such extensions will depend on incorporating region-specific parameters, including agroecological zones, topography, soil type, microclimate, and management intensity, to ensure robust, adaptive modelling of alfalfa persistence across Canadian agroecological zones.

Estimating nutrient depletion factors, such as leaching, volatilization, and crop uptake, remains a critical next step for improving nutrient budget accuracy in future model iterations. In parallel, model development should prioritize enhancing spatial representation, especially in under-sampled provinces such as Ontario and Nova Scotia. Incorporating additional predictors such as root traits, seeding rates, crop age, and stress-based weather indices could enhance predictive accuracy. Further, model performance can be improved through formal sensitivity analysis, cross-validation with new datasets, and refinement of error estimation algorithms. Finally, deploying the model as a web-based decision-support tool with real-time data input and remote sensing integration would enable scalable, site-specific recommendations and broader adoption across Canada's forage production systems.

Conclusion

The predictive model and data-driven assessment tools developed in this study provide a comprehensive framework for evaluating alfalfa winter survival, highlighting key agronomic and environmental variables that influence plant survival. Soil variability, topographic features, and adaptive management strategies informed by real-world field measurements are key parameters for building accurate and region-specific prediction models. Despite regional sampling imbalances, normalization techniques and within-field variability controls helped ensure robust model performance. Tools such as the Wisconsin scoring system and provincial guidelines enhance the integration of covariate data into practical field assessments. Seasonal probability distributions of alfalfa stem counts provide straightforward visual insights for growers and advisors, supporting more informed decisions on planting schedules, fall cutting, and input management. These visual tools also support comparisons across years, which can refine local management practices and guide long-term agronomic planning. For Canadian forage producers, adopting generalized decision-support models and scenario-based tools presents new opportunities to optimize productivity through evidence-based interventions, including variety selection and site-specific soil amendments under variable agro-climatic conditions. By understanding survival and persistence trends, producers can better align agronomic practices with economic and sustainability goals. These findings can support government agencies and policymakers in designing region-specific forage management strategies, optimizing resource allocation, and improving agri-environmental programs for sustainable agricultural planning.

Data availability

The data supporting the findings of this study are available from Mon Système Fourrager; however, restrictions apply to their availability, as they were used under license for the current research and are not publicly accessible. Data are, however, available from the authors upon reasonable request and with permission of Mon Système Fourrager.

Received: 20 August 2025; Accepted: 22 January 2026

References

1. Undersander, D., Grau, C., Cosgrove, D. & Neal Martin, J. D. Alfalfa stand assessment: is this stand good enough to keep? (2011). http://forage.msu.edu/wp-content/uploads/2014/07/WI-A3620-IsThisAlfalfaStandGoodEnoughToKeep-Undersander-etal_2011.pdf.
2. Bélanger, G. et al. Winter damage to perennial forage crops in Eastern canada: causes, mitigation, and prediction. *Can. J. Plant. Sci.* **86**, 33–47 (2006).
3. Sheaffer, C. C. et al. Leaf and stem properties of alfalfa entries. *Agron. J.* **92**, 733–739 (2000).
4. Castonguay, Y., Laberge, S., Brummer, E. C. & Volenec, J. J. Alfalfa winter hardiness: a research retrospective and integrated perspective. *Adv. Agron.* **90**, 203–265 (2006).
5. Wells, M. S., Martinson, K. L., Undersander, D. J. & Sheaffer, C. C. A survey investigating alfalfa winter injury in Minnesota and Wisconsin from the winter of 2012–2013. *Forage Grazinglands* **12**, 963 (2014).
6. Poudel, K. et al. Quantifying winter survival of alfalfa (*Medicago sativa* L). *Agron. J.* **116**, 170–179 (2024).
7. Brock Porth, C. et al. Remote sensing applications for insurance: a predictive model for pasture yield in the presence of systemic weather. *North. Am. Actuar. J.* **24**, 333–354 (2020).
8. Kinoshita, R., Rossiter, D. & van Es, H. Spatio-temporal analysis of yield and weather data for defining site-specific crop management zones. *Precis Agric.* **22**, 1952–1971 (2021).
9. Ma, H. & Jiang, P. Effects of nitrogen fertilization combined with subsurface irrigation on alfalfa yield, water and nitrogen use efficiency, quality, and economic benefits. *Front Plant. Sci* **15**, 748 (2024).
10. Brummer, E. C., Shah, M. M. & Luth, D. Reexamining the relationship between fall dormancy and winter hardiness in alfalfa. *Crop Sci.* **40**, 971–977 (2000).
11. Cunningham, S. M., Volenec, J. J. & Teuber, L. R. Plant survival and root and bud composition of alfalfa populations selected for contrasting fall dormancy. *Crop Sci.* **38**, 962–969 (1998).
12. de Nijs, P. J., Berry, N. J., Wells, G. J. & Reay, D. S. Quantification of biophysical adaptation benefits from climate-smart agriculture using a bayesian belief network. *Sci. Rep.* **4**, 6682 (2014).
13. Eltarabily, M. G. et al. Simulated soil water distribution patterns and water use of alfalfa under different subsurface drip irrigation depths. *Agric Water Manag.* **293**, 4152 (2024).
14. Kim, J. Y. Software design for image mapping and analytics for high throughput phenotyping. *Comput. Electron. Agric.* **191**, 106550 (2021).
15. Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A. & Schwartz, M. D. Shifting plant phenology in response to global change. *Trends Ecol. Evol.* **22**, 357–365 (2007).
16. Chen, Y. et al. Application of image-based phenotyping tools to identify QTL for in-field winter survival of winter wheat (*Triticum aestivum* L). *Theor. Appl. Genet.* **132**, 2591–2604 (2019).
17. Strullu, L. et al. Simulation using the STICS model of C&N dynamics in alfalfa from sowing to crop destruction. *Eur. J. Agron.* **112**, 125948 (2020).
18. Wood, S. et al. Validation of predictive equations of pre-harvest forage nutritive value for alfalfa–grass mixtures. *Agron. J.* **110**, 950–960 (2018).
19. Volenec, J. J. & Nelson, C. J. Environmental aspects of forage management. In *Forages: An Introduction to Grassland Agriculture* (2017).
20. Cortés, U., Sànchez-Marrè, M., Ceccaroni, L., R-Roda, I. & Poch, M. Artificial intelligence and environmental decision support systems. *Appl. Intell.* **13**, 77–91 (2000).
21. Shinde, S. Development of a numeric on-line decision support system for crop fertilizer optimization. *Dissertations & Theses @ McGill University (Canada)* (2017).
22. Tantalaki, N., Souravlas, S. & Roumeliotis, M. Data-driven decision making in precision agriculture: the rise of big data in agricultural systems. *J. Agric. Food Inf.* **20**, 344–380 (2019).
23. McCauley, D. M., Nackley, L. L. & Kelley, J. Demonstration of a low-cost and open-source platform for on-farm monitoring and decision support. *Comput. Electron. Agric.* **187**, 106284 (2021).
24. Buxton, D. R. & Fales, S. L. ASA, CSSA, SSSA,. Plant environment and quality. In *Forage Quality, Evaluation, and Utilization* (ed. Fahey, G. C.) (1994).
25. Hall, M. H., Nelson, C. J., Coutts, J. H. & Stout, R. C. Effect of seeding rate on alfalfa stand longevity. *Agron. J.* **96**, 717–722 (2004).
26. Undersander, D. et al. *Alfalfa Management Guide* (Wiley, 2021).
27. Thompson, N. M., DeLay, N. D. & Mintert, J. R. Understanding the farm data lifecycle: collection, use, and impact of farm data on U.S. Commercial corn and soybean farms. *Precis Agric.* **22**, 1685–1710 (2021).
28. Thorsen, S. M. & Höglind, M. Modelling cold hardening and dehardening in timothy: sensitivity analysis and bayesian model comparison. *Agric. Meteorol.* **150**, 1529–1542 (2010).
29. Saifuzzaman, M. et al. Clustering tools for integration of satellite remote sensing imagery and proximal soil sensing data. *Remote Sens.* **11**, 1036 (2019).
30. Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA). Soil Fertility Handbook. Queen's Printer for Ontario, Ch. 5 (2022).
31. Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA). Risk of alfalfa winterkill (1991). https://www.ontario.ca/page/risk-alfalfa-winterkill.

32. Saifuzzaman, M., Adamchuk, V., Biswas, A. & Rabe, N. High-density proximal soil sensing data and topographic derivatives to characterise field variability. *Biosyst Eng.* **211**, 19–34 (2021).
33. Zhang, Y., Qian, B. & Hong, G. A long-term, 1-km resolution daily meteorological dataset for modeling and mapping permafrost in Canada. *Atmosphere* **11**, 526 (2020).
34. Bélanger, G. et al. Climate change and winter survival of perennial forage crops in Eastern Canada. *Agron. J.* **94**, 1120–1130 (2002).
35. Python Software Foundation. Python (Version 3.x) [Computer software] (2024). https://www.python.org.
36. Page, L. & Brin, S. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **30**, 107–117 (1998).
37. Evans, M., Hastings, N. & Peacock, B. Probability density function and probability function. *Stat. Distrib.* **2000**, 9–11 (2000).
38. Darapuneni, M. K. et al. Potassium and sulfur fertilizer sources influence alfalfa yield and nutritive value and residual soil characteristics in an arid, moderately low-potassium soil. *Agronomy* **14**, 526 (2024).
39. Dhakal, C. & Lange, K. Crop yield response functions in nutrient application: a review. *Agron. J.* **113**, 5222–5234 (2021).
40. Hakl, J., Pisarčik, M., Fuksa, P. & Šantrůček, J. Potential of Lucerne sowing rate to influence root development and its implications for field stand productivity. *Grass Forage Sci.* **76**, 378–389 (2021).
41. Orloff, S. Choosing appropriate sites for alfalfa production. In *Alfalfa Management Guide, Ch. 2* (Univ. of California, 2007).
42. Elsamet, R. M., Serag, A. & Abdel-Razek, M. The influence of tile drainage spacing on water-table fluctuation and alfalfa crop production in the Northeast nile Delta, Egypt. *Menoufia J. Soil. Sci.* **8**, 185–197 (2024).
43. Fick, G. W. & Mueller, S. C. Alfalfa: quality, maturity, and mean stage of development. *Cornell Univ. Info Bull.* **217**, 1–16 (1989).
44. Gao, L., Su, J., Chen, C., Tian, Q. & Shen, Y. Increases in forage legume biomass as a response to nitrogen input depend on temperature, soil characters and planting system: a meta-analysis. *Grass Forage Sci.* **76**, 309–319 (2021).
45. Lauzon, J. et al. Alfalfa and timothy nutritive value in contrasting agroclimatic regions. *Agron. J.* **111**, 1371–1380 (2019).

Acknowledgements
The authors give special thanks to the Canadian Forage and Grassland Association (CFGA) and Agriculture and Agri-Food Canada and Institut de recherche et de développement en agroenvironnement (IRDA) Quebec for data support and cooperation. We are grateful to Dr. Gilles Bélanger and Dr. Dan Undersander for their valuable suggestions and contributions to the project. Additionally, we would like to express our appreciation to Felipe Hoffmann Silva Karp and other graduate students for their assistance with data processing at various stages.

Author contributions
MS and VIA conceived and designed the study. MS, VIA, and ML developed an analytical framework and methodology. MS, VIA, ML, and HE performed data analysis and investigations. MS, VIA, ML, and KC created visualizations and contributed to the interpretation of the results. MS drafted the initial manuscript. MS, VIA, PS, ML, and KC reviewed, revised, and approved the final version of the manuscript for submission.

Funding
This research was supported in part through the Mitacs Accelerate program with support from the Canadian Forage Grassland Association (Fredericton, NB, Canada), Agriculture and Agri-Food Canada (AAFC), and Ministère de l'Agriculture, des Pêcheries et de l'Alimentation (MAPAQ).

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.