American Journal of Plant Sciences, 2022, 13, 1247-1260
https://www.scirp.org/journal/ajps
ISSN Online: 2158-2750
ISSN Print: 2158-2742

Early Nutrient Diagnosis of Kentucky Bluegrass Combining Machine Learning and Compositional Methods

Abdo Badra, Léon Etienne Parent

Department of Soils and Agri-Food Engineering, Université Laval, Québec, Canada
Email: leonet965@gmail.com

How to cite this paper: Badra, A. and Parent, L.E. (2022) Early Nutrient Diagnosis of Kentucky Bluegrass Combining Machine Learning and Compositional Methods. *American Journal of Plant Sciences*, **13**, 1247-1260.
https://doi.org/10.4236/ajps.2022.139085

Received: June 11, 2022
Accepted: September 20, 2022
Published: September 23, 2022

Abstract

Kentucky bluegrass (*Poa pratensis* L.) is the most common perennial turfgrass species grown on playgrounds, municipal and residential lawn areas, and golf tees, fairways and roughs. Fertilization is the most efficient way to improve and maintain turfgrass aesthetic quality. Tissue diagnosis can guide fertilization, but tissue concentration ranges are biased by not taking into consideration nutrient inter-relationships, carryover effects and other key features. The centered log-ratio transformation reflects nutrient interactions in plants and avoids statistical biases. Machine learning (ML) models relate the target variable to the key features *ex ante*, and can predict future events from prior knowledge. The objective of his study was to predict turfgrass quality from key features and rank nutrients in the order of their limitations. The experimental setup comprised four N, three P, and four K rates applied on permanent plots during three consecutive years. Soils were a loam and an USGA sand. Eleven elements (N, S, P, K, Ca, Mg, B, Cu, Zn, Mn, Fe) were quantified in clippings collected during spring, summer and autumn every year. Turfgrass quality was categorized as target variable by color rating. Concentrations were centered log-ratioed (*clr*) partitioned into four quadrants in the confusion matrix generated by the xgboost ML model. The area under curve (AUC) and model accuracy were high to predict turfgrass color from the nutrient analyses of clippings collected in the preceding season, facilitating the seasonal adjustment of the fertilization regime to sustain high turfgrass quality. We provide a computational example to run the ML model and rank nutrients in the order of their limitations.

Keywords

Centered Log Ratio, Data Set, Machine Learning, Turfgrass Foliage Color, Turfgrass Shoot Density, Xgboost

1. Introduction

Kentucky bluegrass (*Poa pratensis* L.) is the most common perennial turfgrass species grown on playgrounds, golf tees, fairways and roughs, and municipal and residential lawns [1]. Soils are generally of sandy or loamy texture. Fertilization is the most efficient way to improve and maintain turfgrass aesthetic quality in terms of shoot density and foliage color [1] [2].

The demand for high-quality turfgrass often prompts golf managers to over-fertilize. Nitrogen increases turfgrass foliage color and shoot density non-linearly, but decreases turfgrass belowground biomass linearly [1]. The excessive nitrogen (N) fertilization may inhibit the growth of rhizomes, stolons, and roots [3] [4] [5] [6] and increase the risk of pest damage [7]. Because turfgrass color is easier to monitor than the labour-intensive measure of belowground biomass or shoot density, Badra *et al.* [8] elaborated an easy-to-use rating for foliage color or greenness in the narrow range of 7.0 to 8.9 based on the color chart of the Royal Horticultural Society (London).

The nutrient status of perennial turfgrass stands is assessed by nutrient tests. By diagnosing each nutrient independently of others, it is assumed that all nutrients but the one being addressed are equal or at their optimum levels (*ceteris paribus* assumption), which is a nonsense by just looking at the unequal results of soil or tissue tests [9]. Moreover, the interpretation of tissue tests does not take into consideration the seasonal variations in nutrient contents, carryover effects of stored carbohydrates and nutrients, and nutrient interactions. The results of laboratory analyses to assess nutritional deficiencies, sufficiency or excesses are not predictive and often arrive too late in the season for early adjustment of fertilization. The impacts of seasonal variations and carryover effects on turfgrass quality could be addressed *ex ante* using predictive machine learning (ML) models [10] [11].

Nutrient data are compositional in nature, *i.e.*, they are intrinsically inter-related and multivariate, and "resonate" on each other due to closure of the bounded compositional space expressed as measurement unit or scale [12]. In contrast, statistical analysis assumes that the data are distributed in the real space (±∞) [13]. If computed from data restricted to the compositional space (e.g., 0% - 100%), the standard deviation (or variance) has no statistical relevance and may even lead to absurd results such as confidence intervals exceeding the lower or upper limits of the predefined compositional space [14] [15]. The results of laboratory analyses expressed as raw concentrations could be log-ratio transformed to allow scanning the real space (±∞) and run statistical analysis unbiasedly [12] [13] [16] [17].

The centered log ratio (*clr*) is the natural logarithm of the ratio between any component and the geometric mean of others. The sources of "resonance" among nutrients quantified in analytical report are dual interactions [18], partial replacement (e.g., K-Na) [19], dilution [20] and cross-talks [21]. The *clr* integrates dual nutrient ratios into a single multi-ratio expression and is a convenient means to diagnose nutrients unbiasedly [21].

The ML methodology can be combined with log-ratio transformations to model compositional data [22]. Several combinations of key features have been used to model the mineral nutrition of perennials [10] [11] and to derive *clr* nutrient standards for diagnostic purposes [23] [24] [25]. As perennial, a turfgrass stand can be monitored every season across several years to allow diagnosing the plant nutrient status from laboratory analyses made available one season in advance, and to predict the quality of the stand during the following season for early adjustment of the fertilization program.

We hypothesized that the aesthetic quality of turfgrass as foliage color can be predicted accurately from features collected at least one season in advance. The objective of this paper was to develop ML models and *clr* nutrient standards to predict turfgrass aesthetic quality for an early adjustment of the fertilization program. Such prediction requires an experimental design maintained during several consecutive seasons and years.

2. Material and Methods

2.1. Experimental Design

The data were collected on an experimental site established in L'Acadie, Quebec, Canada [1] [8]. The turfgrass stands were composed of equal proportions of Kentucky bluegrass cultivars Baron, Argyle, Gnome, and Regent. Soils were a St-Blaise-Macdonald loamy soil (Haplaquent) and a normalized USGA sandy soil. Soil physical and chemical analyses, sampling procedure, and climatic conditions were reported in [1]. The soils were analyzed at the onset of the experiment for grain-size distribution by sedimentation, pH in water (1:1), Mehlich3-extractible nutrients [26] and Walkley-Black carbon content [1]. Soil pH was 6.17 ± 0.27 in the loam and 7.10 ± 0.16 in the sand. Mehlich-III extracts averaged 68 mg P kg^{-1}, 63 mg K kg^{-1}, 165 mg Mg kg^{-1}, 2587 mg Ca kg^{-1} and 1131 mg Al kg^{-1} in the loam, and 58 mg P kg^{-1}, 39 mg K kg^{-1}, 143 mg Mg kg^{-1}, 1925 mg Ca kg^{-1} and 333 mg Al kg^{-1} in the sand. Carbon content was 17.4 g C kg^{-1} in the loam and 10.4 g C kg^{-1} in the sand.

The stand was deemed uniform during the summer following the establishment one year earlier. The trial was a factorial combination of four rates of N (0 or 50, 100, 200, and 300 kg N ha^{-1}), three rates of P (0 or 21.8, 43.7 and 87.3 kg P ha^{-1}), and four rates of K (0 or 41.7, 83.3, 166.7, and 250 kg K ha^{-1}) applied at eight occasions during three consecutive years. The stands were irrigated to avoid drought stress and facilitate the incorporation of fertilizer granules into the soil. Foliage color was rated during eight consecutive periods, *i.e.*, in summer and autumn of year 1, and in the spring, summer and autumn of year 2, and in the spring, summer and autumn of year 3. Color ratings were recorded weekly and summarized as the median value across weekly evaluations preceding and including each clipping harvest.

2.2. Plant Analyses

Clippings were collected at the same time as foliage color was rated, two days following fertilization to avoid collecting fertilizer granules in mower's basket. Mowing height was 38 mm for clipping harvest on areas of 2.38 by 0.56 m, and 50 mm otherwise (2 - 3 times per week as a regular maintenance). Harvests started in April or May and ended in September or October. Clippings were weighed then dried at 70˚C. Total carbon was not quantified in the clipping biomass but may range between 360 and 395 g C kg^{-1} [27] or 355 to 418 g C kg^{-1} [28], averaging 380 g C kg^{-1}. Total N was determined by micro-Kjeldahl. Tissues were acid-digested [29]. The dissolved nutrients were quantified by plasma emission spectroscopy.

2.3. Log Ratio Transformation

The foliage compositional simplex comprised eleven elements and a filling value as follows:

$$\begin{aligned} S^D = \{ & [\mathrm{N},\mathrm{S},\mathrm{P},\mathrm{K},\mathrm{Mg},\mathrm{Ca},\mathrm{B},\mathrm{Fe},\mathrm{Mn},\mathrm{Zn},\mathrm{Cu},\mathrm{F_v}];[\mathrm{N}>0,\mathrm{S}>0,\mathrm{P}>0,\\ & \mathrm{K}>0,\mathrm{Mg}>0,\mathrm{Ca}>0,\mathrm{B}>0,\mathrm{Fe}>0,\mathrm{Mn}>0,\mathrm{Zn}>0,\mathrm{Cu}>0,\mathrm{F_v}>0];\\ & \mathrm{N}+\mathrm{S}+\mathrm{P}+\mathrm{K}+\mathrm{Mg}+\mathrm{Ca}+\mathrm{B}+\mathrm{Fe}+\mathrm{Mn}+\mathrm{Zn}+\mathrm{Cu}+\mathrm{F_v}=1000\ \mathrm{g}\cdot\mathrm{kg}^{-1}\} \end{aligned} \tag{1}$$

where F_v is the filling value between measurement unit (here, 1000 g·kg^{-1}) and the sum of the individual nutrient concentrations expressed in g·kg^{-1}. The centered log ratios were computed as follows Equation (2) [16]:

$$clr_i = \ln\left(\frac{c_i}{G}\right) \tag{2}$$

where ln is the natural logarithm, c_i is i^{th} concentration, and G is the geometric mean across components of the whole including the filling value, computed as follows (3):

$$G = \sqrt[12]{\mathrm{N}\times\mathrm{S}\times\mathrm{P}\times\mathrm{K}\times\mathrm{Mg}\times\mathrm{Ca}\times\mathrm{B}\times\mathrm{Fe}\times\mathrm{Mn}\times\mathrm{Zn}\times\mathrm{Cu}\times\mathrm{F_v}} \tag{3}$$

The geometric mean across components avoids over-optimistic assumptions on equal (Law of minimum) or optimum (Law of optimum) levels of other nutrients [30]. Equation (2) is a combination of pairwise ratios reflecting dual interactions, as follows for N Equation (4):

$$\begin{aligned} clr_i &= \ln\frac{\mathrm{N}}{G} = \ln\left(\frac{\mathrm{N}^{12}}{\mathrm{N}\times\mathrm{S}\times\mathrm{P}\times\mathrm{K}\times\mathrm{Mg}\times\mathrm{Ca}\times\mathrm{B}\times\mathrm{Fe}\times\mathrm{Mn}\times\mathrm{Zn}\times\mathrm{Cu}\times\mathrm{F_v}}\right)^{\frac{1}{12}}\\ &= \ln\left(\frac{\mathrm{N}}{\mathrm{N}}\times\frac{\mathrm{N}}{\mathrm{S}}\times\frac{\mathrm{N}}{\mathrm{P}}\times\frac{\mathrm{N}}{\mathrm{K}}\times\frac{\mathrm{N}}{\mathrm{Mg}}\times\frac{\mathrm{N}}{\mathrm{Ca}}\times\frac{\mathrm{N}}{\mathrm{B}}\times\frac{\mathrm{N}}{\mathrm{Fe}}\times\frac{\mathrm{N}}{\mathrm{Mn}}\times\frac{\mathrm{N}}{\mathrm{Zn}}\times\frac{\mathrm{N}}{\mathrm{Cu}}\times\frac{\mathrm{N}}{\mathrm{F_v}}\right)^{\frac{1}{12}}\\ &= \frac{1}{12}\left(\ln\frac{\mathrm{N}}{\mathrm{S}}+\ln\frac{\mathrm{N}}{\mathrm{P}}+\ln\frac{\mathrm{N}}{\mathrm{K}}+\ln\frac{\mathrm{N}}{\mathrm{Mg}}+\ln\frac{\mathrm{N}}{\mathrm{Ca}}+\ln\frac{\mathrm{N}}{\mathrm{B}}+\ln\frac{\mathrm{N}}{\mathrm{Fe}}+\ln\frac{\mathrm{N}}{\mathrm{Mn}}+\ln\frac{\mathrm{N}}{\mathrm{Zn}}+\ln\frac{\mathrm{N}}{\mathrm{Cu}}+\ln\frac{\mathrm{N}}{\mathrm{F_v}}\right) \end{aligned} \tag{4}$$

The N/P ratio that relates protein production to internal energy [31] and the N/K ratio that highlights the role of K in regulating the N transfer from roots to shoot [19] should be properly balanced in turfgrass [32]. Other nutrient interactions and cross-talks are documented in [18] and [21].

2.4. Statistical Analysis

Three relational models were tested as follows:

$$Y_t = f\left(X_t\right), \tag{5}$$

$$Y_t = f\left(X_{t-1}\right), \tag{6}$$

$$Y_t = f\left(X_{t-2}\right), \tag{7}$$

where Y_t is foliage color rating at time *t* (the season), whereas X_t, X_{t-1} and X_{t-2} are features collected at times *t*, *t* − 1 and *t* − 2, respectively. Equation (5) represents the current diagnosis where laboratory analyses are calibrated against plant performance at time *t* [8]. Equation (6) predicts future plant performance at time *t* from the results of laboratory analyses of tissues sampled at time *t* − 1. Equation (7) predicts future plant performance at time *t* from laboratory analyses of tissues sampled at time *t* − 2.

Features were periods, soil texture, NPK fertilization, tissue nutrients and clipping biomass. Soils were a loam and an USGA sand. Periods were spring, summer and autumn. Because there are *D* − 1 degrees of freedom in the *D clr* variables adding up exactly to zero [33], one redundant *clr* variable (here, clr_{F_v}) was removed from the ML model. The target variable was the category of turfgrass color (within or outside the range between 7.0 and 8.9) for high quality turfgrass. We used the Orange Data Mining software 3.32 to process the data (https://orangedatamining.com/download/#windows). The most accurate ML model was extreme gradient boosting (xgboost) using 40 trees, learning rate of 0.100 and a limiting depth to three individual trees. Data were partitioned in the confusion matrix as follows:

1) True negative (TN) specimens showing balanced mineral nutrition and high turfgrass quality.

2) False negative (FN) specimens showing balanced mineral nutrition but low turfgrass quality (Type II error).

3) True positive (TP) specimens showing imbalanced mineral nutrition but high turfgrass quality.

4) False positive (FP) specimens showing imbalanced mineral nutrition and low turfgrass quality (Type I error).

Model accuracy was measured as follows:

$$\text{Accuracy}\left(\%\right) = 100 \times \frac{\text{VN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \tag{8}$$

The area under curve (AUC) is another criterion for model performance. It has been suggested that AUC of 0.5 has no diagnostic interest [34]. The model would be little informative in the AUC range of 0.5 to 0.7, moderately informative if 0.7 ≤ AUC < 0.9 and very informative if AUC ≥ 0.9.

Nutrient standards were computed as mean and standard deviation of the *clr* values of true negative (TN) specimens. The *clr* indices were computed to rank nutrients, as follows [16]:

$$clr_{x_i}\ index = \frac{clr_{x_i} - clr_{x_i}^*}{SD_{x_i}^*}$$

where clr_{x_i} is the *clr* value for component x_i, and $clr^*_{x_i}$ and $SD^*_{x_i}$ are the corresponding seasonal *clr* nutrient standards.

3. Results and Discussion

3.1. Model Accuracy

There were 3072 observations to run model 1. Model 2 processed 2688 observations because one of the eight periods must be sacrificed at both the loamy and the sandy sites to account for carryover effects. Model 3 processed 2304 observations. The most successful combination of features comprised soil texture, season, clipping biomass, nutrients and NPK fertilization (**Table 1**). The model was little affected after excluding clipping biomass. Soil texture was an important feature, particularly in models 2 and 3 (scenario no. 3 vs. scenario no. 2). The minimum dataset to predict color rating comprised soil texture, season and NPK fertilization provided by the field manager, as well as nutrient analyses from the preceding season supplied by the laboratory.

Table 1 shows that collecting tissue samples one or two periods in advance to predict color rating provided higher area under curve (AUC) compared to nutrient analysis and concomitant color rating at the time of tissue sampling. Models were moderately to very informative. The AUC and accuracy were comparable to other perennials [10] [11] [23] [24] [25].

3.2. Nutrient Standards

Model 2 allows laboratory analyses to be available in time for the early correction of the fertilizer program. Nutrient standards for turfgrass, computed for

Table 1. Effect of selected datasets on the area under curve (AUC) and accuracy of xgboost to predict turfgrass color from features.

Features included in models	Model 1		Model 2		Model 3	
	AUC[§]	Accuracy[§]	AUC	Accuracy	AUC	Accuracy
1) Soil, season, clippings, nutrients, NPK fertilization	0.963	0.908	0.980	0.917	0.983	0.924
2) Soil, season, nutrients, NPK fertilization	0.962	0.904	0.979	0.918	0.982	0.924
3) Season, nutrients, NPK fertilization	0.956	0.893	0.965	0.881	0.970	0.885
4) Soil, season, clippings, nutrients	0.889	0.809	0.932	0.817	0.925	0.801
5) Soil, season, nutrients	0.842	0.762	0.902	0.767	0.903	0.765
6) Season, clippings, nutrients	0.882	0.794	0.915	0.796	0.908	0.782
7) Season, nutrients	0.834	0.754	0.880	0.748	0.877	0.736
8) Nutrients	0.833	0.750	0.866	0.727	0.863	0.722

[§]The highest AUC or accuracy value is 1.

every season using the feature combination no. 7 in model 2, were not stationary across seasons (**Table 1**). The means and standard deviations of the *clr* values for TN specimens are presented in **Table 2**. Between spring and autumn, tissue S tended to increase and tissue N and P to decrease by more than 0.2 *clr* unit. The quartile concentration ranges for tissue N, P and S did not overlap, indicating seasonal effects on plant nutrition that are attributable in part to differential nutrient mobility [35] or nutrient dilution in the growing plant [20].

Table 2. Nutrient standards for the preceding season (t − 1) in model $Y_t = f\left(X_{t-1}\right)$ to reach color ratings of 7.0 to 8.9 during the next season (t).

Nutrient	N	Mean	SD$	LQ†	HQ†	N	Mean	SD	LQ	HQ
		clr		g·kg⁻¹			*clr*		g·kg⁻¹	
		Spring					Summer			
N	151	3.661	0.224	36.3	44.5	267	3.266	0.147	26.3	32.5
S	151	0.688	0.178	1.7	2.6	267	0.978	0.152	2.6	3.3
P	151	1.623	0.222	4.9	5.7	267	1.357	0.159	4.0	4.9
K	151	3.377	0.071	28.4	31.9	267	3.298	0.223	26.3	35.6
Mg	151	0.652	0.104	1.8	2.1	267	0.639	0.232	1.6	2.8
Ca	151	1.526	0.109	4.4	5.0	267	1.582	0.184	4.6	6.5
B	151	−4.676	0.644	0.005	0.024	267	−4.075	0.249	0.016	0.022
Fe	151	−2.122	0.242	0.096	0.153	267	−2.243	0.189	0.104	0.136
Mn	151	−2.975	0.197	0.047	0.059	267	−3.056	0.223	0.044	0.063
Zn	151	−3.742	0.295	0.021	0.027	267	−3.771	0.193	0.022	0.031
Cu	151	−4.804	0.290	0.006	0.012	267	−4.689	0.269	0.009	0.013
F_v	151	6.793	0.083	-	-	267	6.715	0.088	-	-
		Autumn					Overall			
N	531	3.464	0.232	28.6	40.9	949	3.440	0.247	28.6	40.9
S	531	0.927	0.255	2.2	3.2	949	0.903	0.239	2.2	3.2
P	531	1.497	0.165	4.2	5.5	949	1.478	0.195	4.2	5.5
K	531	3.364	0.102	27.8	32.5	949	3.348	0.147	27.8	32.5
Mg	531	0.535	0.095	1.7	2.1	949	0.583	0.157	1.7	2.1
Ca	531	1.293	0.147	3.6	5.1	949	1.411	0.204	3.6	5.1
B	531	−4.258	0.231	0.012	0.020	949	−4.273	0.387	0.012	0.020
Fe	531	−2.247	0.191	0.098	0.131	949	−2.226	0.205	0.098	0.131
Mn	531	−2.996	0.208	0.045	0.060	949	−3.009	0.213	0.045	0.060
Zn	531	−3.652	0.213	0.022	0.032	949	−3.700	0.229	0.022	0.032
Cu	531	−4.706	0.289	0.008	0.012	949	−4.717	0.286	0.008	0.012
F_v	531	6.778	0.083	-	-	949	6.763	0.095	-	-

$Standard deviation; †Lower and higher quartiles, respectively.

Adams [32] reported optimum N/K ratio varying between 1.0 and 1.4 and optimum N/P ratio in the range of 5.7 to 9.0 for turfgrass. The N/K and N/P ratios of TN specimens used to run Model 2 averaged 1.06 ± 0.26 and 6.69 ± 1.34, respectively, across seasons. While a concept of optimum N:K:P:Ca:Mg proportions in plants stable during the exponential growth of woody perennials has been elaborated by [36], such diagnostic approach appeared to be illusionary for turfgrass due to seasonal variations attributable to large seasonal variation in clipping removal. The highest rate of removal occurred in summer (**Table 3**) where nutrient balances were most impacted compared to spring conditions (**Table 2**).

Nutrient standards averaged across seasons differed from those presented by [8] because the FP specimens were excluded by the confusion matrix to derive the *clr* standards. False positive specimens comprise cases of luxury consumption or suboptimal nutrient levels leading to high aesthetic turfgrass quality despite nutrient imbalance. The ML model can isolate the high-quality and nutritionally balanced TN specimens in the confusion matrix.

3.3. Steps to Conduct the Compositional Nutrient Diagnosis of Turfgrass

We propose following five steps from data collection to diagnosis:

1) Collect nutrient concentration data using the same measurement unit ($g \cdot kg^{-1}$) or scale (dry matter basis) to constitute a large and diversified dataset of turfgrass experimental and observational data as reference file.

2) In a separate file, organize data for the specimens under diagnosis using the same features and measurement units as those documented in the reference data set.

3) Run the ML model to classify the diagnosed specimens into the high- (color rating between 7.0 and 8.9) or low-quality (outside range) categories.

4) If turfgrass quality is declared low by the ML model, compare *clr* values to *clr* standards and rank the *clr* indices from the most negative (relative shortage) to the most positive (relative excess).

5) Draw a histogram of *clr* indices to illustrate the diagnosis.

Table 3. Seasonal changes in clipping biomass.

Season	Mean	SD[§]
	$g \cdot m^{-2}$	
Spring	8.8	1.4
Summer	26.3	5.8
Autumn	15.2	7.0
Overall	17.3	8.6

[§]Standard deviation.

3.4. Computational Example

Turfgrass clippings are collected on an USGA sand in the spring. The manager reports a fertilizer regime of 300 kg N $ha^{-1}\cdot yr^{-1}$, 100 kg P_2O_5 $ha^{-1}\cdot yr^{-1}$, and 300 kg K_2O $ha^{-1}\cdot yr^{-1}$. Soil properties were comparable to those of the one used in the above experiment. The manager asks whether the fertilization could be reduced to prevent grass degradation, the eutrophication of surrounding ponds, and, overall, the apparently imbalanced or excessive fertilization and nutrient waste.

The Orange 3.32 Machine Learning Software processed the data using object-oriented algorithms (**Figure 1**). The software can be easily operated by field managers and crop advisers. The reference dataset is first retrieved. Variables are selected (select columns) as features or target to run the model. Several learners are available in Orange 3.32. After sampling 100% of the data, run the ML model using stratified cross-validation, setting the number of folds at 10. Use 40 trees to run xgboost. Note that data can also be divided into calibration and validation data sets to assess model accuracy. Data are partitioned in a confusion matrix (**Figure 2**).

The mineral analysis of clippings is presented in **Table 4**. Data to be diagnosed are saved in a separate file. The diagnosed specimen is classified by the ML model as high- or low-quality specimen, and nutrients are ranked in an order of limitation. The xgboost model predicted a higher probability to reach high than low aesthetic quality at the time of sampling the grass in the spring. Using Excel, the *clr* indices in **Table 4** can be presented in a histogram to illustrate relative nutrient imbalance even if the specimens appeared to be well rated (**Figure 3**). False positive specimens could be rebalanced.

Table 4. Composition of clippings to be diagnosed by the ML model.

Component	Concentration	*clr*	*clr* mean	*clr* SD	*Clr index*
	$g\cdot kg^{-1}$	unitless			
N	37.9	3.662	3.661	0.224	0.004
S	2.7	1.011	0.688	0.178	1.815
P	5.3	1.695	1.623	0.222	0.324
K	28.7	3.384	3.377	0.071	0.099
Mg	1.7	0.557	0.652	0.104	−0.913
Ca	4.3	1.485	1.526	0.109	−0.376
B	0.008	−4.802	−4.676	0.644	−0.196
Fe	0.084	−2.450	−2.122	0.242	−1.355
Mn	0.045	−3.074	−2.975	0.197	−0.503
Zn	0.034	−3.355	−3.742	0.295	1.312
Cu	0.007	−4.964	−4.804	0.290	−0.552
F_v	919.2	6.850	6.793	0.083	0.687

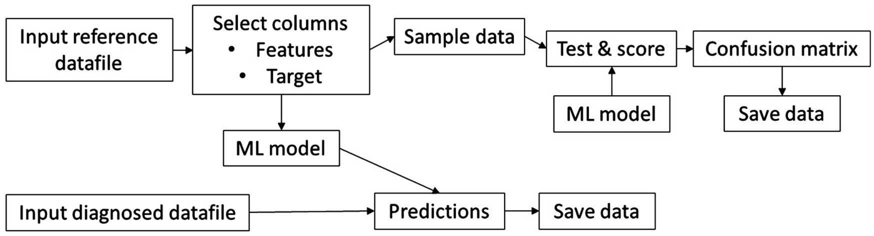

Figure 1. Caneva of the orange 3.32 data mining software (https://orangedatamining.com/download/#windows) to diagnose the aesthetic quality of turfgrass stands.

	Predicted	
Actual	TN	FP
	FN	TP

Figure 2. Confusion matrix of the ML model partitioning the predicted and the actual target variable (turfgrass color category) into True Negative (TN), False Negative (FN), False Positive (FP) and True Positive (TP) specimens.

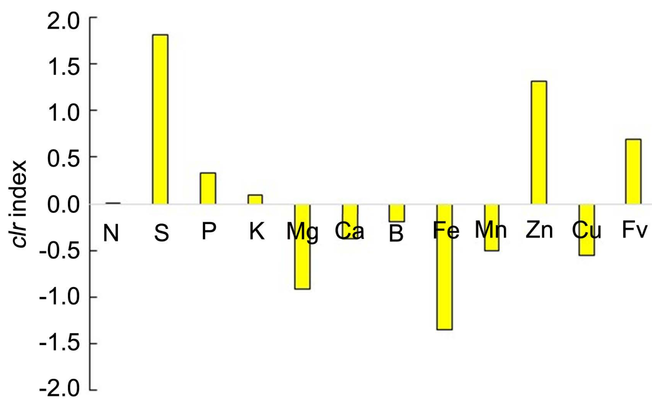

Figure 3. Graphical representation of nutrient ranking for an actual low-quality turfgrass. The *clr* indices are unitless.

The nitrogen appeared to be well balanced (*clr* value near zero), yet at relatively higher level under spring conditions compared to summer and autumn (**Table 2**). However, N fertilization should still sustain turfgrass aesthetic quality by fertilization due to high removal rate of the clippings at high N application rate. Tissue N tended to decrease in the summer (**Table 2**) due in part to higher rate of clipping removal that reduced N storage (**Table 3**). Because turfgrass response to added N is nor-linear, achieving a color rating of 7.0 in an USGA sandy soil was found to be ≥245 kg N ha^{-1} [1], a potential gain of 55 kg N ha^{-1} compared to current N management. The P and K fertilization could also be reduced substantially without quality loss because turfgrass was little or not responsive to P and K additions at those levels of soil test [1].

The apparent S excess indicated that potassium sulfate could be replaced by the less expensive potassium sources. Less potassium could also have a positive impact on Mg and Ca acquisition by the plant by alleviating antagonisms [18].

Excess sulfur may also impact the transfer of metals from the roots to the shoot through cross-talks, depending on the plant [21]. The Zn appeared to be in relative excess and could be skipped from the fertilization program.

4. Conclusions

Plant nutrient diagnosis has been conducted traditionally using sufficiency ranges of nutrient concentrations. However, nutrients that "resonate" on each other within the plant system are inherently multivariate data and should be analyzed as combinations specific to the specimen under diagnosis. Nutrients were centered log-ratio transformed to account for nutrient inter-relationships before running the xgboost ML model. The diagnostic model combined xgboost to classify the diagnosed specimen and *clr* indices to rank nutrients in the order of limitation. In contrast with the traditional diagnostic approaches, ML models can include several other features and are predictive, allowing to anticipate turfgrass quality one season earlier in order to adjust the fertilization program.

The xgboost ML model was accurate in predicting turfgrass color rate based on prior knowledge of features documented in a reference dataset (e.g., soil texture, season, irrigation, NPK fertilization, clipping biomass and tissue nutrient composition). Nutrient standards were elaborated for the preceding season to forecast turfgrass quality in the following season and readjust fertilization accordingly. Readjusting regularly the turfgrass fertilization program could be profitable both economically and, most importantly, environmentally.

The dataset used in this paper was specific to a single turfgrass species grown on two sites and cannot be generalized to a larger diversity of playgrounds, golf tees, fairways and roughs, and municipal and residential lawn areas. However, additional experimental and observational data could be acquired collaboratively to build up the large and diversified data sets on nutrient requirements of turfgrass ecosystems needed to run ML models.

Acknowledgements

This research was conducted at the L'Acadie Research station of Agriculture and Agri-Food Canada (St-Jean-sur-Richelieu). We are grateful to the Quebec Ministry of Agriculture, Fisheries and Food for chemical plant analyses, and Hydro Agri Canada laboratory for chemical and physical soil testing. We thank Hydro-Agri Canada (a division of Norsk-Hydro in Norway), Pelouses Richer Boulet, Groupe Richer, Comptoir Richelieu, Bigelow sand, l'Institut Québécois du Développement de l'Horticulture Ornementale, Technique Gazon Inc., Agriculture and AgriFood Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC-DG 2254) for providing financial support and facilities.

Author Contribution Statement

Conceived and designed the experiments: AB, LEP; performed the experiment:

AB; analyzed the data: AB, LEP; wrote the manuscript: AB, LEP.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] Badra, A., Parent, L.E., Desjardins, Y., Allard, G. and Tremblay, N. (2005) Quantitative and Qualitative Responses of an Established Kentucky Bluegrass (*Poa pratensis*, *L.*) Turf to N. P. and K Additions. *Canadian Journal of Plant Science*, **85**, 193-204. https://doi.org/10.4141/P03-125

[2] Landschoot, P. (2018) Turfgrass Fertilization. A Basic Guide for Professional Turfgrass Managers. Penn State Extension, University Park, 21 p.

[3] Watschke, T.L. and Waddington, D.V. (1974) Effect of Nitrogen Source Rate and Timing on Growth and Carbohydrates of Merion Kentucky Bluegrass. *Agronomy Journal*, **66**, 691-696. https://doi.org/10.2134/agronj1974.00021962006600050025x

[4] Watschke, T.L. and Waddington, D.V. (1975). Effect of Nitrogen Fertilization on the Recovery of Merion Kentucky Bluegrass from Scalping and Wilting. *Agronomy Journal*, **67**, 559-563. https://doi.org/10.2134/agronj1975.00021962006700040027x

[5] Christians, N.E., Martin, D.P. and Karnok, K.J. (1981). The Interactions among Nitrogen Phosphorus and Potassium on the Establishment Quality and Growth of Kentucky Bluegrass (*Poa pratensis*, *L.* "Merion"). *Proceedings of the* 4*th International Turfgrass Research Conference*, Guelph, 19-23 July 1981, 341-348.

[6] Bocker, P. and Opitz von Boberfeld, W. (1974) Influence of Various Fertilizers on Root Development in a Turfgrass Mixture. *Proceedings of the* 2*nd International Turfgrass Research Conference*, Blacksburg, 19-21 June 1973, 99-103. https://doi.org/10.2135/1974.proc2ndintlturfgrass.c15

[7] Martinez, D.A., Loening, U.E., Graham, M.C. and Gathorne-Hardy, A. (2021) When the Medicine Feeds the Problem; Do Nitrogen Fertilisers and Pesticides Enhance the Nutritional Quality of Crops for Their Pests and Pathogens? *Frontiers in Sustainable Food Systems*, **5**, Article ID: 701310. https://doi.org/10.3389/fsufs.2021.701310

[8] Badra, A., Parent, L.E., Allard, G., Tremblay, N., Desjardins, Y. and Morin, N. (2006) Effect of Leaf Nitrogen Concentration versus CND Nutritional Balance on Shoot Density and Foliage Color of an Established Kentucky Bluegrass (*Poa pratensis*, *L.*) Turf. *Canadian Journal of Plant Science*, **86**, 1107-1118. https://doi.org/10.4141/P05-242

[9] Nowaki, R.H.D., Parent, S.É., Cecílio Filho, A.B., Rozane, D.E., Meneses, N.B., Santos da Silva, J.A., Natale, W. and Parent, L.E. (2017) Phosphorus Over-Fertilization and Nutrient Misbalance of Irrigated Tomato Crops in Brazil. *Frontiers in Plant Science*, **8**, Article No. 825. https://doi.org/10.3389/fpls.2017.00825

[10] Parent, L.E., Jamaly, R., Atucha, A., Parent, J.E., Workmaster, B.A., Ziadi, N. and Parent, S.É. (2021) Current and Next-Year Cranberry Yiels Predicted from local Features and Carryover Effects. *PLOS ONE*, **16**, Article ID: e0250575. https://doi.org/10.1371/journal.pone.0250575

[11] Nava, G., Reisser Júnior, C., Parent, L.E., Brunetto, G., Moura-Bueno, J.M., Navroski, R., Benati, J.A. and Barreto, C.F. (2022) Esmeralda Peach (*Prunus persica*)

Fruit Yield and Quality Response to Nitrogen Fertilization. *Plants*, **11**, Article No. 352. https://doi.org/10.3390/plants11030352

[12] Aitchison, J. (1986) The Statistical Analysis of Compositional Data. Chapman and Hall, London.

[13] Filzmoser, P., Hron, K. and Reimann, C. (2009) Univariate Statistical Analysis of Environmental (Compositional) Data: Problems and Possibilities. *Science of Total Environment*, **407**, 6100-6108. https://doi.org/10.1016/j.scitotenv.2009.08.008

[14] Diaz-Zorita, M., Perfect, E. and Grove, J.H. (2002) Disruptive Methods for Assessing Soil Structure. *Soil & Tillage Research*, **64**, 3-22.

[15] Pawlowsky-Glahn, V., Egozcue, J.J. and Tolonasa-Delgado, R. (2011) Lecture Notes on Compositional Data Analysis. http://www.compositionaldata.com/material/others/Lecture_notes_11.pdf

[16] Parent, L.E. and Dafir, M. (1992) A Theoretical Concept of Compositional Nutrient Diagnosis. *Journal of the American Society of Horticultural Science*, **117**, 239-242. https://doi.org/10.21273/JASHS.117.2.239

[17] Greenacre, M. (2021) Compositional Data Analysis. *Annual Review of Statistics and its Application*, **8**, 271-299. https://doi.org/10.1146/annurev-statistics-042720-124436

[18] Wilkinson, S.R., Grunes, D.L. and Sumner, M.E. (2000) Nutrient Interactions in Soil and Plant Nutrition. In: Sumner, M.E, Ed., *Handbook of Soil Science*, CRC Press, Boca Raton, D89-D112.

[19] Marschner, H. (1986) Mineral Nutrition of Higher Plants. Academic Press, London, Orlando.

[20] Jarrell, W.M. and Beverly, R.B. (1981) The Dilution Effect in Plant Nutrition Studies. *Advances in Agronomy*, **34**, 197-224. https://doi.org/10.1016/S0065-2113(08)60887-1

[21] Gallardo, K., Vigani, G., Brunel-Muguet, S., Trouverie, J., Salon, C. and Ourry, A. (2019) Disentangling the Complexity and Diversity of Cross-Talk between Sulfur and Other Mineral Nutrients in Cultivated Plants. *Journal of Experimental Botany*, **70**, 4183-4196. https://doi.org/10.1093/jxb/erz214

[22] Parent, S.É. (2020) Why We Should Use Balances and Machine Learning to Diagnose Ionomes. *Authorea*, **1**, 20 p.

[23] Neto, A.J. L., Deus, J.A. L., Rodrigues Filho, V.A., Natale, W. and Parent, L.E. (2020) Nutrient Diagnosis of Fertigated "Prata" and "Cavendish" Banana (*Musa* spp.) at Plot-Scale. *Plants*, **9**, Article No. 1467. https://doi.org/10.3390/plants9111467

[24] Betemps, D.L., Paula, B.V. de, Parent, S.É., Galarça, S.P., Mayer, N.A., Marodin, G.A. B., Rozane, D.E., Natale, W., Melo, G.W.B., Parent, L.E. and Brunetto, G. (2020) Humboldtian Diagnosis of Peach Tree (*Prunus persica*) Nutrition Using Machine-Learning and Compositional Methods. *Agronomy*, **10**, Article No. 900. https://doi.org/10.3390/agronomy10060900

[25] Paula, B.V., de Arruda, W.S., Parent, L.E. and Brunetto, G. (2020) Nutrient Diagnosis of Eucalyptus at Factor-Specific Level Using Machine Learning and Compositional Methods. *Plants*, **9**, Article No. 1049. https://doi.org/10.3390/plants9081049

[26] Mehlich, A. (1984) Mehlich 3 Soil Test Extractant: A Modification of Mehlich 2 Extractant. *Communications in Soil Science and Plant Analysis*, **15**, 1409-1416. https://doi.org/10.1080/00103628409367568

[27] Garnier, E. and Vancaeyzeele, S. (1994) Carbon and Nitrogen of Congeneric Annual

and Perennial Grass Species: Relatinships with Growth. *Plant Cell & Environment*, **17**, 399-407. https://doi.org/10.1111/j.1365-3040.1994.tb00308.x

[28] Bümane, S., Poiša, L., Čubars, E. and Platače, R. (2015) The Analysis of Carbon Content in Different Eneergy Crops. 25*th NJF Congress Nordic View to Sustainable Rural Development*, 16-18 June 2015, Riga, 156-160.

[29] Isaac, R. and Johnson, W. (1976) Determination of Total Nitrogen in Plant Tissue Using a Block Digestor. *Journal of the Association of Official Analytical Chemists*, **59**, 98-100. https://doi.org/10.1093/jaoac/59.1.98

[30] De Wit, C.T. (1992) Resource Use Efficiency in Agriculture. *Agricultural Systems*, **40**, 125-151. https://doi.org/10.1016/0308-521X(92)90018-J

[31] Güsewell, S. (2004) N : P Ratios in Terrestrial Plants: Variation and Functional Significance. New Phytologist, **164**, 243-266. https://doi.org/10.1111/j.1469-8137.2004.01192.x

[32] Adams, W.A. (1978) Effects of Nitrogen Fertilization and Cutting Height on the Shoot Growth. Nutrient Removal and Turf Grass Composition of an Initially Perennial Ryegrass Courbdominant Sports Turf. *Proceedings of the* 3*rd International Turfgrass Research Conference*, Munich, 11-13 July 1977, 343-350. https://doi.org/10.2135/1974.proc3rdintlturfgrass.c40

[33] Aitchison, J. (1994) Principles of Compositional Data Analysis. Multivariate Analysis and Its Applications IMS Lecture Notes Monograph Series Vol. 24, Institute of Mathematical Statistics, Ohio, 73-81. https://doi.org/10.1214/lnms/1215463786

[34] Delacour, H., Servonnet, A., Perrot, A., Vigezzi, J.F. and Ramirez, J.M. (2005) La courbe ROC (receiver operating characteristic): Principes et principales applications en biologie clinique. *Annales de Biologie Clinique*, **63**, 145-154. [In French].

[35] Hill, J. (1980) The Remobilization of Nutrients from Leaves. *Journal of Plant Nutrition*, **2**, 407-444. https://doi.org/10.1080/01904168009362788

[36] Ingestad, T. (1987) New Concepts on Soil Fertility and Plant Nutrition as Illustrated by Research on Forest Trees and Stands. *Geoderma*, **40**, 237-252. https://doi.org/10.1016/0016-7061(87)90035-8