Open Access

SCHOLARENA

Journal of Horticulture Science and Forestry

Review Article

Micronutrient Soil and Tissue Test Interpretation for Vegetable Crops in Eastern Canada

Léon Etienne Parent[1,2*] and Melissa Quinche[3]

[1]*Department of Soils, Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brazil*
[2]*Department of Soils and Agrifood Engineering, University Laval, Québec (Qc), Canada*
[3]*SOLINOV inc., Brossard (Qc), Canada*

[*]**Corresponding author:** Léon Etienne Parent, Visiting Professor, Department of Soils, Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brazil and Emeritus Professor, Department of Soils and Agrifood Engineering, Université Laval, Québec (Qc), Canada, Tel: 4188365872, E-mail: leon-etienne.parent@fsaa.ulaval.ca

Citation: Léon Etienne Parent, Melissa Quinche (2021) Micronutrient Soil and Tissue Test Interpretation for Vegetable Crops in Eastern Canada. J Hortic Sci For 3: 101

Abstract

Micronutrient soil and tissue test interpretation is challenging in vegetable production due to limited field calibration, large number of factors impacting micronutrient availability, and analytical methods more or less correlated between them. Besides, several NPK fertilizer trials have been conducted to calibrate soil test. Because yield potential varied widely and soil and tissue test have been conducted, micronutrients concentration intervals could be computed at high-yield level. On the other hand, tissue nutrient reference values can also be extracted at high-yield level from large observational datasets. Our objective was to delineate micronutrient reference values for high-yielding vegetable crops from experimental and observational datasets. There were 290 observations from fertilizer trials for muck vegetables, 650 for crucifers and 7968 for potatoes, to derive soil test reference values for micronutrients. There were 1005-3372 observations to derive tissue test reference values for vegetables and potato. The datasets were processed using Random Forest in classification mode where yield cutoffs were set at regional averages. Classification accuracies were 0.665-0.881 for soil tests and 0.652-0.851 for tissue tests, allowing to set apart nutritionally balanced, high-yielding, true negative specimens. There were wide concentration compatibility intervals for high-yield crops at regional scale. High concentration values were attributable to contamination by fungicides and to variation in soil properties. True negative neighbors allowed conducting local diagnosis of defective specimens using tools of compositional data analysis. Perturbation vectors ranked nutrient status from shortage to excess to guide recovery from nutrient imbalance.

Keywords: Broccoli; Cauliflower; Cabbage; Potato; Carrot; Celery; Onion; Machine Learning; Nutrient Balance

Introduction

There are several micronutrients involved in plant nutrition [1,2] and disease control. Micronutrient shortage impacts the productivity of many sensitive vegetable crops [4]. While vegetable crops are grown in a large variety of mineral and organic soils, micronutrient soil testing have been conducted in limited field trials [5]. It is not easy to find responsive sites to conduct fertilization trials [6] and to interpret soil test results considering the large number of soil features impacting micronutrient availability such as organic matter content, pH, texture, and soil test P [5,7]. Moreover, micronutrient extraction methods may not correlate well [6,8,9] On the other hand, the plant integrates site-specific genetic, managerial, and environmental factors [10].

To derive soil and tissue test sufficiency ranges, simple response models such as Liebig's law of the minimum, Liebscher's law of the optimum, Mitscherlich's law of diminishing returns or statistical and boundary lines have been fitted to experimental data under the *ceteris paribus* assumption [11-13]. The *ceteris paribus* assumption may not hold at the step of assembling the results of field experiments where factors other than the ones being varied may have limited yield [14]. Wallace and Wallace (1993) suggested a law of the maximum where all factors, not only nutrients, are optimized. However, high crop yield depends on high soil quality [15] and other factors [16]. Low or excessive micronutrient soil test levels may limit the attainment of yield potential across seasons [5]. Micronutrient reference values can be retrieved from experimental and observational datasets to allow diagnosing the results of soil and tissue tests [6,17] Machine learning (ML) methods can address the complexity of soil-plant relationships by accounting for the growth-impacting features [18-21].

Plants grow normally across a large number of nutrient combinations [22]. There has been much controversy on how to diagnose tissue nutrients. Paula *et al.* (2020) distinguished nutrient-pasting diagnosis against "critical" concentration values, where each nutrient is addressed separately [11,23,24] from "optimum" nutrient combinations [12, 25,26]. Sufficiency ranges of micronutrients can be delineated using boundary lines [11]. Beaufils (1973) analyzed yield-nutrient relationships to compute statistics on the nutrient status (concentrations and dual ratios) of high-yielding crops. The latter two approaches did not discard false positive specimens showing suboptimal concentration levels, contamination by dust or fungicides, or luxury consumption. The confusion matrix of machine learning methods can set apart specimens as true negative (nutritionally balanced, high-yielding), from false negative (nutritionally balanced, low-yielding), false positive (nutritionally imbalanced, high-yielding), and true positive (nutritionally imbalanced, low-yielding) [27].

Nutrient diagnosis using regionally derived sufficiency ranges may fail for not accounting for nutrient interactions [27]. The numerous nutrient interactions in the plant system [28] support the nutrient balance concept. Nutrient interactions can be viewed as physiologically meaningful nutrient combinations where nutrients resonate on each other within tissue compositional entities [26]. A sound numerical solution to closed interactive systems was provided by "Compositional Data Analysis" (CoDa) methods [29,30]. Because the geometry of closed compositions is Euclidean, a Euclidean distance can be computed between two compositions, one abnormal and one normal, allowing to compare two nutrient compositional entities rather than examining nutrients separately.

Machine learning and compositional methods provide a comprehensive approach to analyze the relationships between crop yield and micronutrient soil and tissue tests across numerous combinations of growth-impacting features. We hypothesized that (1) regional references for micronutrient soil and tissue tests can be computed accurately from experimental and observational datasets using machine learning methods, and (2) local micronutrient diagnosis accounting for numerous combinations of crop-impacting features may differ from regional diagnosis across factors. Our objective was to facilitate interpreting, by combining machine learning and compositional methods, micronutrient soil and tissue tests at regional and local scales to reach high yields of vegetables grown in organic and mineral soils.

Material and Methods

Datasets

Three soil datasets of experimental data collected in Quebec, Canada (Figure 1) related crop yield to soil test: (1) 290 observations from NP fertilizer trials on vegetable crops grown in organic soils (muck vegetables), (2) 650 observations from NPK fertilizer trials on crucifers grown in gleysols and brunisols, and (3) 7968 observations from NPK fertilizer trials on potato (*Solanum tuberosum L.*) crops grown in podzols and brunisols. Muck vegetable were carrot (*Daucus carota L*. subsp. sativus), celery (*Apium graveolens*), Chinese cabbage (*Brassica rapa* subsp. chinensis), lettuce (*Lactuca sativa* L.), onion (*Allium cepa* L.), and potato (*Solanum tuberosum L.*). Crucifers were broccoli (*Brassica oleracea* var. italica), cauliflower (*Brassica oleracea* var. botrytis) and cabbage (*Brassica oleracea* var. capitata).

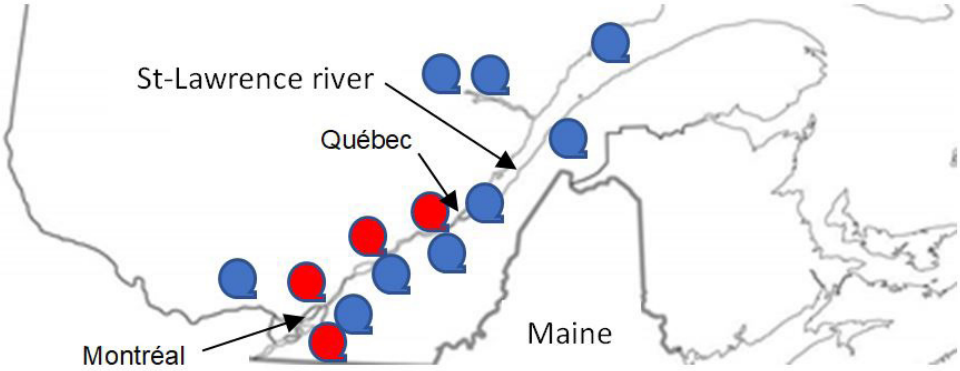

Figure 1: Sites where field fertilizer trials were conducted during the 1993-2017 period (blue and red symbols) and tissue samples were collected during the 1985-1991 period (red symbol) in vegetable cropping systems in Quebec, Canada.

There were four tissue datasets of observational data collected in southwestern Quebec (Figure 1) and relating yield to tissue test at four growth stages (Table 1). For carrot and onion, the green aboveground part was collected at stages A to C, followed by the uppermost mature leaf at stage D. For potato, the first mature leaf from top was sampled across the season. Yields were measured in three 2-m long beds per plot. For celery, the petiole of the highest leaf was selected [31]. Celery was harvested after 745-955 cumulated degree-days (≥ 7 °C), foliage excess chopped, and yield adjusted to 1000 degree-days, assuming proportional yield gain [32]. Celery yield was reported as kg per plant due to the large variation in plantation density (32000-74000 plants ha^{-1}).

Crop	#data	Growth stage			
		A	B	C	D
Carrot	1221	4-5 leaf	6-7 leaf	8-10 leaf	Root enlargement
Celery	1196	30 cm high	50 cm high	-	-
Dry onion	1025	2-3 leaf	4-5 leaf (leek stage)	6-8 leaf	Bulb enlargement
Potato	3192	20 cm high	Floral bud stage	10% blooming	Before vine killing

Table 1: Growth stages and number of data in the Quebec tissue dataset

Methods of analysis

The soil datasets reported soil test results from composite samples collected in the 0-20 cm layer at each site. Samples were dried at 45 °C. Soil pH was measured in water. Soil tests were conducted using the Mehlich-3 method (Mehlich 1984). Nutrients were quantified by ICP-OES (Inductively coupled plasma-optical emission spectrometer, Perkin Elmer, Waltham, Massachusetts). Total carbon and nitrogen were quantified by combustion (CNS-Leco 2000, St-Louis, MO) after drying at 105 °C.

The tissue datasets reported tissue test results from 10-20 composite leaf samples collected on 30 m by 30 m plots delineated in commercial fields. Dust and other sources of contaminants cannot be removed completely from plant tissue, causing difficulties in interpreting analytical results [33]. Washing to decontaminate is not recommended unless absolutely necessary due to potential loss of soluble elements and further contamination [34,35] and to avoid damaging the fragile fresh tissues. Samples were dried at 70 °C. Nitrogen was determined by microKjeldahl or combustion (CNS-Leco 2000, St-Louis, MO). Other nutrients were quantified by ICP-EOS after dry-ashing at 550 °C followed by dissolution in dilute HCl.

Machine learning

In the soil datasets, only experimental sites reporting crop yield, soil test for micronutrients and related features (texture, C, P, pH, Al and Fe) were retained to run the machine learning model. To calibrate soil test from experimental datasets, the highest marketable yield at each site was selected as site-specific yield potential. In the tissue datasets, only observational data that reported all macro- and micronutrients and yield were retained.

Marketable yields were related to features using the Orange 3.23 data mining freeware. To run machine learning model in classification mode, yield cutoffs between low- and high-yielding crops were set at Quebec 2015 average for onion (46.6 Mg ha^{-1}), carrot (37.7 Mg ha^{-1}), celery (44.7 Mg ha^{-1}, averaging 0.86 kg $plant^{-1}$), lettuce (24.7 Mg ha^{-1}), broccoli (13.6 Mg ha^{-1}), cauliflower (20.6 Mg ha^{-1}), cabbage (38.8 Mg ha^{-1}), and potato (33.8 Mg ha^{-1}) [36,37]. Yield cutoff of Chinese cabbage was assumed to be similar to that of celery. For the celery tissue dataset, yield cutoff was set at 0.86 kg per plant for plant density of 52 000 plants ha^{-1}.

The Random Forest model was calibrated using stratified cross-validation (k = 10). Classification accuracy was computed as the sum of true negative and true positive (TP) (nutritionally imbalanced, low yield) specimens divided by total number of observations. True negative (TN) specimens (nutritionally balanced, high yield) were set apart using the confusion matrix. False negative (FN) specimens were those showing low yield despite nutrient balance, indicating that factors other than nutrients reduced yield. False positive specimens (FP) showed nutrient imbalance at high yield level. Nutrient intervals [38] were selected as minimum and maximum soil and tissue test values among TN specimens that represented successful conditions. Such intervals called "compatibility intervals" differed from "critical" ranges because they were derived from observations on successful specimens rather than from calibration experiments where nutrient dosage had been varied under the *ceteris paribus* assumption. The results were compared to literature values.

Compositional data analysis

To run local diagnosis on abnormal vs. normal specimens (Munson and Nelson, 1990), tissue micronutrients were arranged into balances (Figure 2). Balances are computed as isometric log-ratios (ilr) coordinates with orthonormal basis, as follows (Egozcue *et al.* 2003):

$$ilr_k = \sqrt{\frac{r_k s_k}{r_k + s_k}}\, ln\left(\frac{\sqrt[r_k]{\prod_{i=1}^{r_k} x_i}}{\sqrt[s_k]{\prod_{j=1}^{s_k} x_j}}\right)$$

where r_k and s_k are numbers of components at numerator and denominator, respectively; i and j refer to components at numerator and denominator, respectively; and $\sqrt[r_k]{\prod_{i=1}^{r_k} x_i}$ and $\sqrt[s_k]{\prod_{j=1}^{s_k} x_j}$ are geometric means of components at numerator and denominator, respectively. The conterminous successful neighbors were detected as the ones showing the closest distances (ε) from the diagnosed specimen in the Euclidean space (Parent, 2020), as follows:

$$\varepsilon = \sqrt{\sum_{k=1}^{D-1}(ilr_k - ilr_k^*)^2}$$

Where * refers to composition of the reference successful specimens. Eucidean distances could be computed separately across the micronutrient and macronutrient subsystems as shown in Figure 2 to avoid affecting the diagnosis by contaminants.

Micronutrient and macronutrient components of the diagnosed specimen were ranked against the composition of the closest successful specimens using the perturbation vector [39] to guide corrective measures enabling to reach realistic yields at local scale under nutritionally balanced conditions. The perturbation vector computes sequentially the ratio between target and diagnosed concentrations [39]. The perturbation (p) between two D-part compositions X (defective) and x (successful) is a scaling operation computed as follows [40]:

$$X = p \oplus x = \frac{[p_1 x_1 \ldots p_D x_D]}{(p_1 x_1 + \cdots + p_D x_D)} = [p_1 x_1 \ldots p_D x_D]$$

Where

$$p = X \ominus x = \left[{}^{X_1}/_{x_1}, \ldots, {}^{X_D}/_{x_D}\right]$$

There is relative nutrient excess in the defective composition where $X_i/x_i > 1$ or $X_i/x_i - 1 > 0$, and relative shortage where $X_i/x_i < 1$ or $X_i/x_i - 1 < 0$.

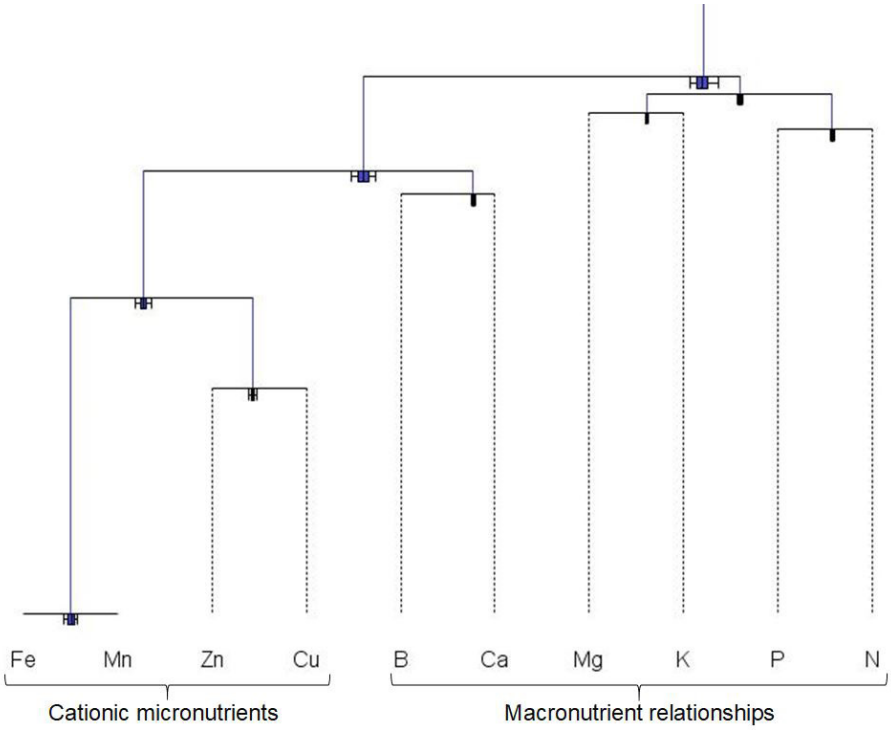

Figure 2: Nutrient balance design to project nutrient relationships in the Euclidean space as coordinates

Nutrient diagnosis

Risk analysis is the probability for a given composition to exceed yield cutoff as predicted by the machine learning model. If the machine learning model indicates that the specimen is a low-yielder, nutrients are ranked from relative shortage to relative excess using the composition of the closest successful neighbors as reference concentrations to compute the perturbation vector [38]. While the successful neighbors have soil or tissue test compositions close to those of the diagnosed specimen, its composition is adequate to reach yield levels above yield cutoff. The smallest Euclidean distance between the successful and defective compositions and the perturbation vector could guide nutrient management under comparable site conditions. Nutrient diagnosis was conducted as follows:

1. Make a prediction for the diagnosed specimen using the ML model to determine whether cutoff yield can be exceeded.
2. In the case of nutrient imbalance, compare compositions of defective to those of the closest successful specimens.
3. Select realistic yield using yields of successful specimens as benchmark.
4. Rank nutrients using the perturbation vector.
5. Assist interpretation using nutrient intervals at high yield level (optional).
6. Suggest corrective measures.

Results

Site characteristics

Site characteristics are presented in Tables 2 and 3 for the mineral and organic soil datasets, respectively. There were wide ranges of soil quality features that may impact yield potential of vegetable crops.

Feature	Crucifers	Potato
#Observations	650	7968
Textural group[1]	G1, G2, G3	G1, G2, G3
pH_{water}	5.6-7.9	4.6-7.2
	g kg^{-1}	
Carbon	8-43	1-7
	mg kg^{-1}	
P	7-394	5-592
K	45-320	10-478
Ca	526-5778	92-3750
Mg	21-758	6-487
B	-	0.2-20
Cu	1.0-6.1	0-17
Zn	1.0-6.5	0-39
Mn	7-206	1-92
Fe	119-460	49-761
Al	523-1389	343-3344

[1]G1: fine-textured soils (texture other than medium or coarse); G2: medium-textured soils (loam, silty loam, silt); G3: coarse-textured soils sand, sandy loam, loamy sands)

Table 2: Minimum and maximum values of mineral soil properties used as features

Feature	Carrot	Celery	Chinese cabbage	Lettuce	Onion	Potato
#Observations	32	42	32	90	46	48
pH_{water}	4.4-5.9	4.8-6.0	4.9-5.7	4.8-6.2	4.6-6.0	3.9-6.6
	g kg^{-1}					
Carbon	318-537	393-492	343-469	322-662	339-507	106-513
	mg kg^{-1}					
P	29-491	45-512	47-854	48-704	49-610	111-614
K	397-722	135-1431	132-1292	349-1317	145-823	240-1524
Ca	6931-12628	8754-14933	8781-15728	7756-17138	5265-13170	4809-16438
Mg	806-2090	1511-2779	1092-2562	826-2708	965-3276	612-2293
Cu	7.6-51.4	9.7-49.8	4.8-29.4	6.9-52.2	5.7-49.3	1.4-26.1
Zn	8.7-34.6	14.6-41.6	18.1-39.7	12.5-44.1	13.8-39.6	6.2-35.3
Mn	19-91	20-56	19-46	19-100	29-81	13-116
Fe	475-1216	432-1052	540-1188	396-1386	278-1627	440-1384
Al	6-2314	1-63	8-743	1-740	1-768	13-819

Table 3: Minimum and maximum values of organic soil properties used as features

Cationic micronutrients represented the major source of variation in foliar compositions as shown by potatoes grown in mineral soils and onions grown in organic soils (Figure 3). Biplot analysis supported diagnosing cationic micro- and macronutrient+B subsets separately (Figure 2). Large variations indicated numerous potential combinations of micronutrients to reach high yield at regional scale. In contrast, local diagnosis could avoid comparisons with excessively high or low levels of micronutrients. At local scale, there are successful combinations of nutrient levels at high yield level under conditions more similar to those of defective specimens.

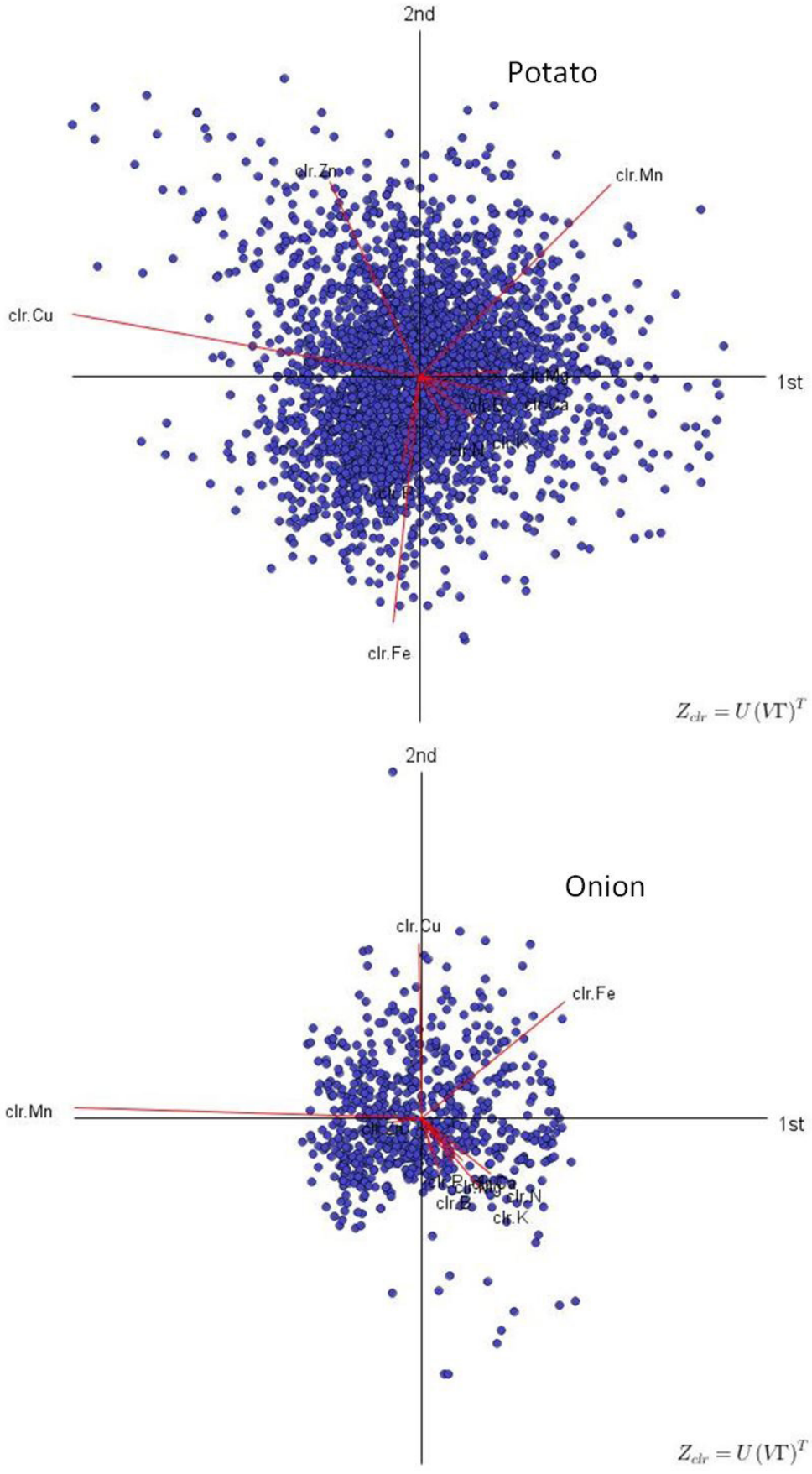

Figure 3: Sources of variation in foliar nutrient composition of potato and onion showing the predominance of cationic micronutrients Cu, Mn and Fe in biplots

Machine learning model for soil tests at regional scale

The Random Forest model related marketable yields to soil features (texture, pH, total C and analysis of soil P, Cu, Zn, Mn, Fe and Al) with classification accuracy between 0.7 and 0.8, allowing to set apart a subpopulation of TN specimens (Table 4). Model accuracy was higher for vegetables grown in organic soils compared to those grown in mineral soils. Nutrient intervals are presented in Table 5 as minimum and maximum soil test values for micronutrients and associated features for crops yielding more than their respective Quebec average. Minimum micronutrient levels were generally higher in organic than in mineral soils. Micronutrient intervals provided milestones to interpret micronutrient tests specific to crops and soils.

Crop category	**Classification** accuracy		Counts
		# Sites	# True negative specimens
Potato in mineral soils	0.665	200	52
Crucifers in mineral soils	0.726	148	39
Vegetables in organic soils	0.881	108	49

Table 4: Accuracy of the Random Forest model for vegetable crops for soil test results (texture, pH, total C, and soil test P, Cu, Zn, Mn, Fe, and Al)

#observations	Cu	Zn	Mn	Fe	Al	P	C	pH
	mg kg^{-1}						%	
	Potato in mineral soils							
52	0.3-5.9	0.6-11.3	3-58	116-673	392-2511	15-580	1.1-5.7	4.9-6.7
	Crucifers in mineral soils							
39	1.3-7.9	1.3-16.5	27-159	163-447	570-1287	12-349	1.1-4.2	6.1-7.4
	Vegetables in organic soils							
49	4.8-52.2	8.7-44.1	19-100	278-1628	1-2314	29-669	31.8-53.7	4.7-6.3

Table 5: Nutrient intervals of soil features for vegetable crops at high-yield level

Machine learning model for tissue tests at regional scale

Model accuracy and number of TN specimens were highest for carrot and lowest for onion (Table 6). Lower accuracy is due to higher numbers of FN and FP specimens in the onion and potato datasets. The celery dataset showed intermediate results.

Crop	Classification acuracy	Balanced specimens		Imbalanced specimens	
		TN	FN	FP	TP
Carrot in organic soils	0.853	649	102	30	119
Celery in organic soils	0.791	383	86	39	90
Onion in organic soils	0.652	206	122	119	245
Potato in mineral soils	0.700	379	342	885	2485

Table 6: Data partitioning and accuracy of the Random Forest model for tissue composition (N, P, K, Mg, Ca, B, Cu, Zn, Mn, Fe)

Micronutrient intervals at high-yield level are presented per growth stage in Table 7. They differed from published concentration ranges, indicating the need to develop region-specific reference values. Upper bounds of intervals were often extremely high compared to literature values, indicating contamination of the foliage by dust or fungicide applications that may vary among growth stages. Lower bounds than published values indicated differences in environment, management or genetics.

Micronutrient	Growth stage (see Table 1)				
	A	B	C	D	Mid/half-growth[1]
	Range (mg kg^{-1})				
	Carrot in organic soils				
n	111	121	118	299	
B	22-54	26-55	27-78	24-66	29-60
Cu	6-31	2-43	4-22	3-72	4.5-7.0
Zn	33-164	23-209	29-179	21-233	20-50
Mn	24-588	29-872	11-481	22-759	190-325
Fe	92-3162	71-633	65-284	53-1862	120-335
	Celery in organic soils				
n	193	190			
B	37-148	42-220			30-50
Cu	2-593	4-67			5-8
Zn	9-603	27-216			20-50
Mn	13-721	6-106			200-300
Fe	23-275	21-179			20-40
	Onion in organic soils				
n	51	61	30	64	
B	14-37	18-51	23-42	20-43	30-45
Cu	8-36	4-22	4-36	4-38	
Zn	40-147	33-538	32-144	21-360	10-15
Mn	27-808	25-681	43-774	49-903	
Fe	85-649	59-510	58-444	47-606	20-40
	Potato in mineral soils				
n	108	125	92	17	
B	15-47	15-45	17-175	16-139	30-40
Cu	6-30	3-53	4-25	4-20	-
Zn	23-116	19-102	17-480	19-165	20-40
Mn	37-541	38-558	53-1510	94-541	30-50
Fe	136-3263	108-2690	110-1139	156-437	70-150
	Deficient[2]	Sufficient or normal[2]	Excessive or toxic[2]		
B	<15	20-100	200+		
Cu	<4	5-20	20+		
Zn	<20	25-150	400+[3]		
Mn	<20	20-50	500+		
Fe	<50	50-250	unknown		

[1]Geraldson and Tyler (1990)

[2]Ranges for recently mature leaves estimated across plants species (Stevenson 1986)

[3]144-222 mg Zn kg-1 for Chinese cabbage and celery (Long et al. 2003)

Table 6: Micronutrient intervals per growth stage of four vegetables at high yield level

Local diagnosis using tissue tests

Compositions were analyzed by the machine learning model as combinations of features before focusing on individual nutrients using the perturbation vector and concentration intervals in Table 7. Local diagnosis was conducted at the 6-8 leaf stage on two defective onion specimens containing, respectively:

1. Specimen #1: 40 g N kg^{-1}, 6.8 g P kg^{-1}, 76.4 g K kg^{-1}, 5.9 g Mg kg^{-1}, 32.6 g Ca kg^{-1}, 0.032 g B kg^{-1}, 0.012 g Cu kg^{-1}, 0.252 g Zn kg^{-1}, 0.216 g Mn kg^{-1}, and 0.105 g Fe kg^{-1};

2. Specimen #2: 36 g N kg^{-1}, 3.0 g P kg^{-1}, 44.5 g K kg^{-1}, 2.6 g Mg kg^{-1}, 19.3 g Ca kg^{-1}, 0.023 g B kg^{-1}, 0.005 g Cu kg^{-1}, 0.030 g Zn kg^{-1}, 0.315 g Mn kg^{-1}, and 0.195 g Fe kg^{-1}.

The Random Forest prediction model returned risk analysis with 15% probability for specimen #1 (20.9 Mg ha^{-1}) and 10% probability for specimen #2 (36.9 Mg ha^{-1}) to reach yield higher than Quebec average of 46.6 Mg ha^{-1}. Hence, nutrient imbalance apparently depressed yield. Successful neighbors showing the closest Euclidean distances showed yield potentials of 76 Mg ha^{-1} for specimen #1 and 51-53 Mg ha^{-1} for specimen #2. The two balance subsets in Figure 2 were diagnosed separately to set apart trace metals. By addressing subsets separately, the large influence of highly variable levels of cationic micronutrients on Euclidean distance can be avoided. Nutrients were ranked as perturbation vectors for each balance subset.

For specimen #1, leaf Zn appeared to be at excessive level while macronutrients were adequately balanced (Figure 4). Because yield of specimen #1 was so low, Zn likely reached toxic level, well outside the Zn concentration interval in Table 7 at the 6-8 leaf stage (stage C). Hence, there was a need to abate Zn levels or modify the cropping system. There was shortage of most nutrients in the leaf of specimen #2 compared to its closest successful neighbor (Figure 5). Nutrient imbalance can be re-established by changing the fertilization regime. Paradoxically, nutrient concentrations were all within the large nutrient intervals at high yield level (Table 7).

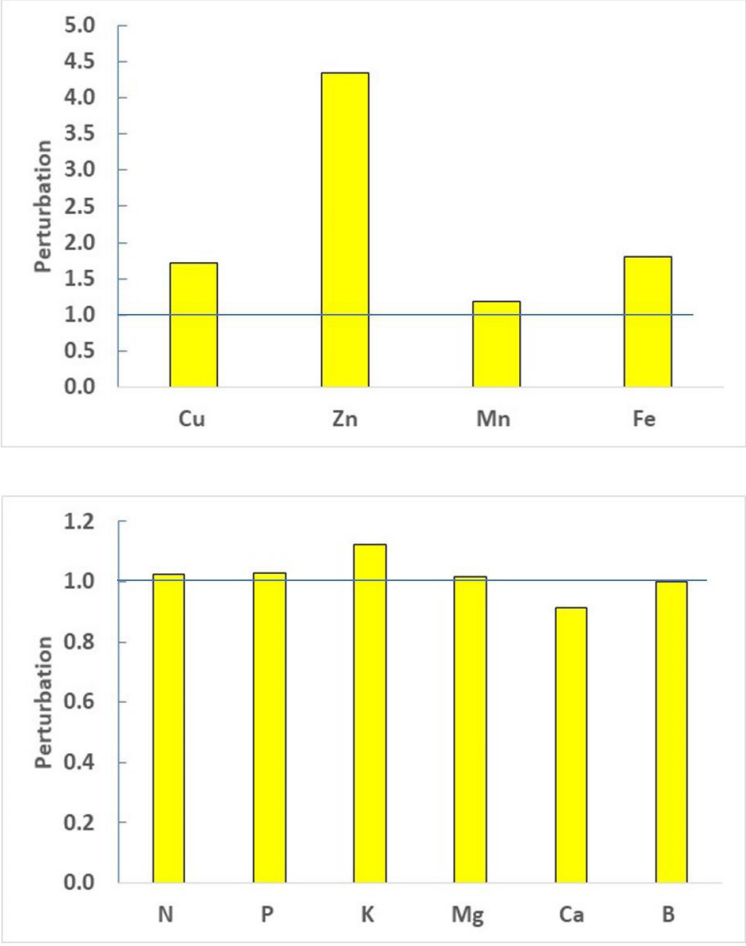

Figure 4: Local nutrient diagnosis of onion specimen #1

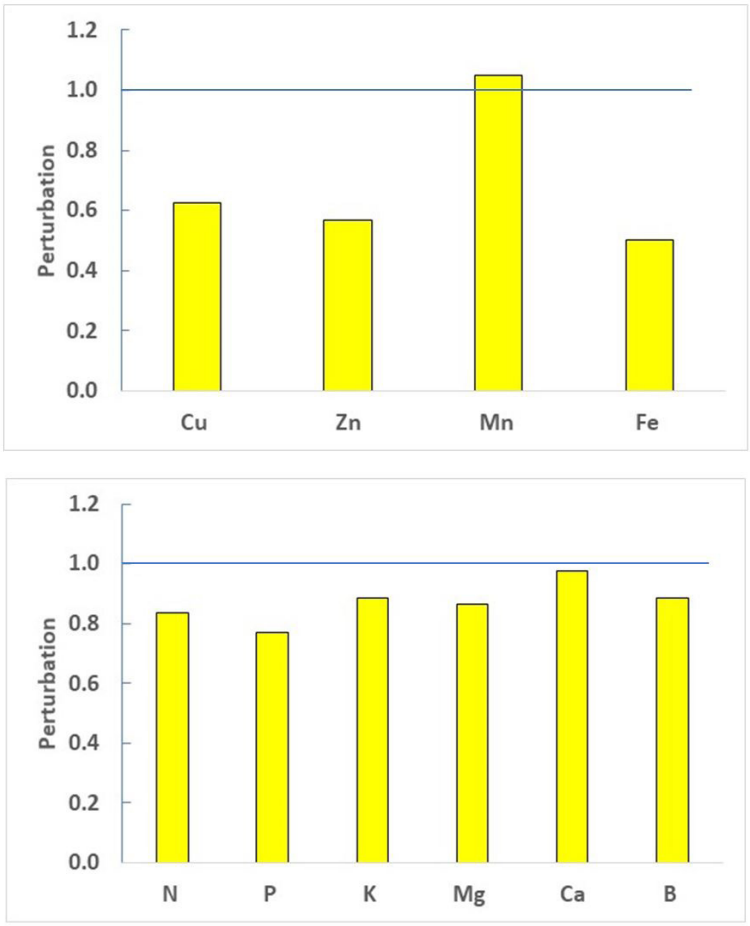

Figure 5: Local nutrient diagnosis of onion specimen #2

Discussion

Ceteris paribus assumption

The comparison between abnormal and normal specimens growing in the same or similar fields has a long tradition in agronomy [10,14]. Indeed, it is logical to think that for crop specimens growing under similar conditions (*ceteris paribus*), one crop being defective and others successful, successful nutrient compositions showing close Euclidean distances from that of the defective one can inform on site-specific nutrient requirements to reach high yield. Diagnosing compositions rather than concentrations differs from the current one-by-one diagnosis using nutrient standards computed separately across regional factors. Local diagnosis of unique nutrient combinations is facilitated by access to large datasets, predictive machine learning models and tools of compositional data analysis [27].

While micronutrient soil test results are often difficult to interpret because they are related to several soil properties such as pH, texture and organic matter and sesquioxide contents [5], and nutrient interactions [28]. Machine learning methods can integrate factors regulating soil's capacity to supply micronutrients and can address several combinations of features simultaneously. They are appealing modelling approaches to interpret the results of soil and tissue tests for micronutrients and predict crop yield from features. The confusion matrix sets apart nutritionally balanced high-yielding specimens to derive regional concentration intervals and local perturbation vectors.

Because micronutrient shortage is frequent in vegetable crops (Lucas 1982) and macronutrients were not limiting at experimental sites, we focused on micronutrients as probable reasons for lower crop performance, allowing to derive concentration intervals at high-yield level. On the other hand, we related tissue tests to absolute crop yields in observational datasets, like [12,41,42] and others did. We projected tissue compositions into the Euclidean space, and considered two balance subsets, one for macronutrients, generally supported by fertilizer trials, and the other for cationic micronutrients largely impacted by fungicide applications and soil properties. The differences between nutrient balance subsets of two compositions were diagnosed across their respective Euclidean dimensions.

Nutrient diagnosis using regional concentration intervals may differ from diagnosis using the local closest successful compositions and realistic attainable yields as references, supporting the need to acquire large datasets for fine-tuning nutrient diagnosis at factor-specific level. The ML models can connect myriads of factor-impacting factors that may occur at local scale. Additional local features such as cultivar, meteorological indices, soil quality indices, fertilizer dosage, pest management and soil classification, could be documented in the datasets to increase the accuracy of ML models and improve yield predictions. Where fertilizer trials are available, site-specific response curves can be generated by the machine learning model [19, 21,39] Otherwise, regional guidelines computed across factors (Martens and Westermann 1991) [43] could be adjusted upward or downward as corrective measure.

Micronutrient soil test interpretation

Concentration intervals found in this research were larger than those reported in the literature and depended on crop and soil types, indicating genetic x environment interactions. The potato cropping system in mineral soils showed the highest variation in soil test Zn and Mn values. Muck vegetables showed the highest Cu and Zn minimum soil test values, and potato grown in mineral soils, the smallest, due in part to considerable differences in soil carbon contents and other reactive surfaces involved in the micronutrient cycles [29]. Note that concentration intervals at high yield level could also vary if yield cutoffs other than regional averages had been selected.

Organic soils showed high variation in soil test Fe and Cu. Indeed, there is large spatial variability in Fe oxy-hydroxide content in organic soils as related to soil genesis and the formation of "bog ore" [44]. Naturally occuring high total Cu up to 1745 mg Cu kg^{-1} did not affect onion yield and leaf concentrations of P, K, Ca, Mg and Mn [45]. Differential Cu dosage can also result in highly variable soil test Cu. Copper may be applied in large amounts in recently reclaimed organic soils to support crop yield by tackling the "reclamation disease" and, later during soil ripening, for partial control of subsidence through mitigation of enzyme activities [46]. The contamination of vegetables through inappropriate use of fertilizers and pesticides (e.g., copper sulfate) represents a risk for human health. For example, Chinese cabbage may show high Cu content without any visible symptom of toxicity, indicating potential risk to human health. It is therefore important to monitor Cu fertilization to reduce the risk for human health from the bioaccumulation of copper in vegetable crops.

Micronutrient tissue test interpretation

Leaf Cu, Fe, Mn and Zn contents represented the largest source of variation in tissue nutrients inherited primarily from fungicide applications and, as discussed above, contrasting soil properties or management. Indeed, high micronutrient concentrations left on leaf surface through fungicide applications are not physiologically active in the plant but still contribute to crop yield by tackling plant diseases. It was thus justified although not mandatory to address tissue test diagnosis using two different nutrient balance subsets to isolate cationic micronutrient balances from others. To illustrate how to conduct local diagnoses, two defective onion specimens showing low probability to reach high yield levels were compared to their respective closest successful specimens. Yield targets documented for successful specimens indicated considerable yield gain potential after recovery from nutrient imbalance.

One defective specimen (#2) showed general nutrient shortage as may be encountered in newly reclaimed organic soils. The other defective specimen (#1) showed 252 mg Zn kg^{-1} at the 6-8 leaf stage and evidence of Zn toxicity compared to the closest successful composition. In comparison, toxic Zn levels occurred naturally in New-York organic soils, reaching 397-2053 mg Zn kg^{-1} in onion tops for unproductive organic soils compared to 44 mg Zn kg^{-1} for normally productive soils [47]. Any Zn toxicity could be partially mitigated by heavy liming that may, in turn, cause other collateral damages such as Mn shortage [4]. Although upper limits of Zn intervals often exceeded 397 mg Zn kg^{-1} across onion growth stages, such Zn excess for false positive specimens was not toxic and was likely due to contamination by fungicides.

On the other hand, Mn toxicity may occur in potato crops depending on cultivar and soil pH (Ouellette and Genereux 1965a, 1965b), especially at soil pH less than 4.8, close to the minimum pH of 4.6 found to reach above average provincial yield (Table 1). The toxicity of Mn could also be favored by acidifying fertilizers such as superphosphate (Boyer 1980.) While foliar Mn levels exceeding 400 mg Mn kg^{-1} may lead to leaf necrosis, crop yield may decrease markedly where foliar Mn levels exceed 200 mg Mn kg^{-1}, well below the uppermost value of the interval in Table 6. Where foliar Mn of high-yielding crops exceeded apparent Mn toxicity, contamination by Mn-containing fungicides likely occurred.

Conclusion

In the present study, soil and tissue tests were assembled into datasets for potato and crucifers grown in mineral soils, and for several vegetables grown in organic soils. The main sources of variation in tissue compositions were cationic micronutrients, indicating large variability in soil genesis and managerial practices likely impacting crop yield. We derived micronutrient reference values under the hypothesis that micronutrients can contribute to differential crop productivity. Due to variable numbers of false negative and false positive specimens, the Random Forest model classification accuracy varied between 0.665 and 0.881 for soil tests and between 0.652 and 0.851 for tissue tests. Reference soil and tissue test values were retrieved as concentration intervals at high yield levels from the population of true negative specimens set apart by the confusion matrix of the Random Forest model.

Concentration intervals for soil and tissue tests at high yield level were highly variables and differed from literature values, indicating that numerous nutrient combinations can return high yields. Using two nutrient balance subsets, one for micronutrients and another for the remaining nutrients, local diagnosis compared compositions of defective specimens to the closest successful compositions in the Euclidean subspaces. The resulting local diagnosis may differ from regional diagnosis by accounting for local factors. The diagnostic tools for micronutrients developed in the present study could sustain vegetable-producing systems.

Author Contributions

LEP and MQ assembled and modelled the data, and wrote the paper.

Funding

This project was funded by the Quebec Ministry of Agriculture, Fisheries and Food (MAPAQ) and the Natural Science and Engineering Research Council of Canada (NSERC-DG-2254).

Conflicts of Interest

The authors have declared that no competing interests exist.

References

1. Knezek BD, Ellis BG (1980) Essential micronutrients IV: copper, iron, manganese, and zinc. In Applied soil trace elements, Ed. Davies, B.E., John Wiley & Sons, NY: 259-86.

2. Welch RM, Shuman L (1995) Micronutrient Nutrition of Plants. Crit. Rev. Plant Sci **14**: 49-82.

3. Graham RD, Webb MJ (1991) Micronutrients and disease resistance and tolerance in plants. In Micronutrients in Agriculture, 2nd edn Eds. Mortvedt, J.J., Cox, F.R., Shuman, L.M., Welch, R.M., Soil Sci. Soc. Book Ser. Am. #4, Madison WI: 329-70.

4. Lucas RE (1982) Organic soils (Histosols). Formation, distribution, physical and chemical properties, and management for crop production. Research Report, Michigan State University Agricultural Experimental Station and Cooperative Extension Service, East Lansing MI, and Agricultural Experiment Stations, University of Florida, Gainesville FL 77.

5. Sims TJ, Johnson GV (1991) Micronutrient soil tests. In Micronutrients in Agriculture, 2nd Ed. Eds Mortvedt, J.J., Cox, F.R., Shuman, L.M., Welch, R.M., Soil Sci. Soc. Book Ser. Am. #4, Madison WI 421-76.

6. Mortvedt JJ (1984) Micronutrient soil test correlations and interpretation. In Soil testing: correlating and interpreting the analytical results; 4th ed., Ed. Stelly, M., Am. Soc. Agron. Spec. Publ. 29, American Society of Agronomy, Madison WI; 89-117.

7. Moraghan JT, Mascagni HJ (1991) Environmental and soil factors affecting micronutrient deficiences and toxicities. In Micronutrients in Agriculture, 2nd Ed., Eds Mortvedt, J.J., Cox, F.R., Shuman, L.M., Welch, R.M., Soil Sci. Soc. Book Ser. Am. #4, Madison WI: 371-425.

8. Levesque MP, Mathur SP (1988) Soil tests for copper, iron, manganese, and zinc in Histosols: 3. A comparison of eight extractants for measuring active and reserve forms of elements. Soil Sci **145**: 215-21.

9. Vocasek FF, Friedericks JB (1994) Soil micronutrient extraction by Mehlich-3 compared to $CaCl_2$-DTPA. Commun. Soil Sci. Plant Anal 25: 1583-93.

10. Munson RD, Nelson WL (1990) Principles and Practices in Plant Analysis. In Soil Testing and Plant Analysis, Ed. Westerman, R.L., Soil Science Society of America, Madison WI: 359–87.

11. Webb RA (1972) Use of the boundary line in the analysis of biological data. J Hort Sci 47: 309-19.

12. Beaufils ER (1973) Diagnosis and Recommendation Integrated System (DRIS), Soil Science Bulletin #1. Pietermaritzburg, South Africa: Dept. Soil Science and Agrometeorology, Univ. Natal.

13. Wit CT de (1992) Resource Use in Agriculture. Agric. Syst. 40: 125–51.

14. Nelson LA, Anderson RL (1984) Partitioning of soil test-crop response probability. In Soil testing: correlating and interpreting the analytical results, 4th ed., Ed. Stelly, M., Am. Soc. Agron. Spec. Publ. 29, American Society of Agronomy, Madison WI: 19–28.

15. Doran JW, Parkin TB (1994) Defining and assessing soil quality. In Defining soil quality for a sustainable environment, Eds. Doran, J.W., Coleman, D.C., Bezdicek, D.F., Stewart, B.A.; Soil Sc. Soc. Am. Spec. Publ. 15, Madison WI: 3-21.

16. Liliane TN, Charles MS (2020) Factors Affecting Yield of Crops. In Agronomy - Climate Change & Food Security, Chapter 2, Ed. D. Amanullah, IntechOpen, London UK.

17. Ulrich A, Hills FJ (1967) Principles and practices of plant analysis. In Soil Testing and Plant Analysis, Ed. Stelly, M., Hamilton, H., Part II. Soil Science Society of America, Madison, WI: 11–24.

18. Parent SÉ (2020) Why we should use balances and machine learning to diagnose ionomes. Authorea 2020.

19. Parent SÉ, Dossou-Yovo W, Ziadi N, Tremblay G, Pellerin A, et al. (2020a) Corn response to banded phosphorus fertilizers with or without manure application in Eastern Canada Agron J 112: 2176-87.

20. Coulibali Z, Cambouris AN, Parent SÉ (2020a) Cultivar-specific nutritional status of potato (*Solanum tuberosum L.*) crops. PLoS ONE 15: e0230458

21. Coulibali Z, Cambouris AN, Parent SÉ (2020b) Site-specific machine learning predictive fertilization models for potato crops in Eastern Canada. PLoS ONE 15: e0230888.

22. Bayens J (1967) Nutrition des plantes de culture ou physiologie appliquée aux plantes agricoles. Ed. E. Nauwelaerts : Louvain, Belgium.

23. Macy P (1936) The quantitative mineral nutrient requirements of plants. Plant Physiol. 11: 749–64.

24. Ulrich A (1952) Physiological basis for assessing the nutritional requirements of plants. Ann. Rev. Plant Physiol. 3: 207-28.

25. Lagatu H, Maume L (1934) Le diagnostic foliaire de la pomme de terre. Ann. L'école Natl. Agron. Montp. 22: 150–8.

26. Parent LE, Dafir MA (1992) Theoretical Concept of Compositional Nutrient Diagnosis. J Am Soc Hortic Sci 117: 239–42.

27. Betemps DL, Paula BV de, Parent SÉ, Galarça SP, Mayer NA (2020) 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: 900.

28. Wilkinson SR (2000) Nutrient Interactions in Soil and Plant Nutrition. In Handbook of Soil Science, Sumner, M.E., ed., CRC Press, Boca Raton, FL: D89–D112.

29. Stevenson FJ (1986) Cycles of soils. Carbon, nitrogen, phosphorus, sulfur, micronutrients. Wiley, NY.

30. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric Logratio Transformations for Compositional Data Analysis. Math Geol 35: 279–300.

31. Tremblay N, Auclair P, Parent LE, Gosselin A (1993) A multivariate diagnosis approach applied to celery. Plant Soil 15: 39-43.

32. Tremblay N, Parent LE, Gosselin A (1990) Élaboration de normes DRIS provisoires pour des transplants de céleri. Phytoprotection 71: 129-36.

33. Mehlich A (1984) Mehlich-3 soil test extractant: a modification of Mehlich-2 extractant. Commun. Soil Sci Plant Anal 15: 1409-16.

34. Jones JB, Case VW (1990) Sampling handling and analyzing plant tissue samples. In Soil Testing and Plant Analysis, 3rd ed., Ed. Westerman, R.L., Soil Science Society of America: Madison, WI, USA: 389–427.

35. Jones JB (1991) Plant tissue diagnosis in micronutrients. In Micronutrients in Agriculture, 2nd ed., Eds. Mortvedt, J.J., Cox, F.R., Shuman, L.M., Welch, R.M., Soil Sci. Soc. Book Ser. Am. #4, Madison WI: 329-70.

36. Déziel MH (2017) Portrait-diagnostic sectoriel des légumes frais au Québec. MAPAQ, Québec, Canada.

37. Déziel MH (2019) Portrait-diagnostic sectoriel de l'industrie de la pomme de terre au Québec. MAPAQ, Québec, Canada.

38. Paula BV de, Squizani Arruda W, Parent LE, Araujo EF, et al. (2020) Nutrient Diagnosis of Eucalyptus at the Factor-Specific Level Using Machine Learning and Compositional Methods. Plants 9: 1049.

39. Parent SÉ, Lafond J, Paré MC, Parent LE, Ziadi N (2020b) Conditioning Machine Learning Models to Adjust Lowbush Blueberry Crop Management to the Local Agroecosystem. Plants *9:*1401.

40. Egozcue JJ, Pawlowsky-Glahn V (2005) Groups of Parts and Their Balances in Compositional Data Analysis. Math. Geol. 37: 795–828.

41. Parent LE, Cambouris AN, Muhawenimana A (1994) Multivariate diagnosis of nutrient imbalance in potato crops. Soil Sci. Soc Am J 58:1432-8.

42. Khiari L, Parent LE, Tremblay N (2001) The phosphorus Compositional Nutrient Diagnosis range for potato. Agron J 93: 815-19.

43. Martens DC, Westermann DT (1991) Fertilizer applications for correcting micronutrient deficiencies. In Micronutrients in Agriculture, 2nd ed., Eds Mortvedt, J.J., Cox, F.R., Shuman, L.M., Welch, R.M., Soil Sci. Soc. Book Ser. Am. #4, Madison WI: 549-92.

44. Guérin J, Parent LE, Abdelhafid R (2007) Agri-environmental thresholds using Mehlich-III soil phosphorus saturation index for vegetables in Histosols. J Environ Qual 36: 975-82.

45. Mathur SP, Bélanger A, Valk M, Preston CM, Knibbe E, et al. (1983) A study of onions grown in microplots on three organic soils each containing four levels of copper. Can J Soil Sci 63: 221-8.

46. Karam A (2003) Retention of copper in Cu-enriched organic soils. In Organic soils and peat materials for sustainable agriculture, Eds. Parent, L.E., Ilnicki, P., CRC Press, Boca Raton FL: 137-50.

47. Staker EV (1943) Progress report on the control of zinc toxicity in peat soils. Soil Sci Soc Am Proc 7: 387-92.

48. Aitchison J (1986) The Statistical Analysis of Compositional Data, Monographs on Statistics and Applied Probability; Chapman & Hall Ltd.: London UK.

49. Agueh V, Degbey CC, Sossa-Jerome C, Adomahou D, Paraiso MN, et al. (2015) Niveau de contamination des produits maraîchers par les substances toxiques sur le site de Honéyiho au Bénin. Int J Biol Chem Sci 9: 542-51.

50. Boyer I (1980) Toxicité apparente de divers engrais phosphates et toxicité manganique induite. Cahiers ORSTOM Pédologie 18: 297-04.

51. Geraldson CM, Tyler KB (1990) Plant analysis as an aid in fertilizing vegetable crops. In: Soil Testing and Plant Analysis, Ed. Westerman, R.L., Soil Science Society of America, Madison, WI: 549–62.

52. Lima Neto AJ, Deus JAL de, Rodrigues Filho VA, Natale W, Parent LE (2020) Nutrient Diagnosis of Fertigated "Prata" and "Cavendish" Banana (Musa spp.) at Plot-Scale. Plants 9: 1467.

53. Murphy LS, Walsh LM (1972) Correction of micronutrient deficiencies with fertilizers. In Micronutrients in Agriculture, Mortvedt, J.J, (ed.), Soil Sci. Soc. Am, Madison WI: 347-87.

54. Ouellette G, Généreux H (1965a) Influence de l'intoxification manganique sur six variétés de pomme de terre. Can. J Soil Sci 45: 24-32.

55. Ouellette G, Généreux H (1965b) Influence du pH et des éléments fertilisants sur l'intoxification manganique de la pomme de terre. Can. J Soil Sci 45: 347-53.

56. Parent SÉ, Parent LE, Rozane DE, Hernandes A, Natale W (2012) Nutrient balance as paradigm of soil and plant chemometrics. In Soil Fertility, Ed. Issaka, R. N., IntechOpen Ltd., London UK: 83–114.

57. Schnug E, Heym J, Achwan F (1996) Establishing critical values for soil and plant analysis by means of the boundary line development system (bolides). Commun. Soil Sci. Plant Anal. 27: 2739-48.

58. Wallace, A. and Wallace, G.A. 1993. Limiting Factors, High Yields, and Law of the Maximum. Hortic Rev 15: 409–48.

59. Xiong ZT, Wang H (2005) Copper toxicity and bioaccumulation in Chinese cabbage (Brassica pekinensis Rupr.). Environ. Toxicol, 20: 188-94.