

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,800

Open access books available

142,000

International authors and editors

180M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Machine Learning, Compositional and Fractal Models to Diagnose Soil Quality and Plant Nutrition

Léon Etienne Parent, William Natale and Gustavo Brunetto

Abstract

Soils, nutrients and other factors support human food production. The loss of high-quality soils and readily minable nutrient sources pose a great challenge to present-day agriculture. A comprehensive scheme is required to make wise decisions on system's sustainability and minimize the risk of crop failure. Soil quality provides useful indicators of its chemical, physical and biological status. Tools of precision agriculture and high-throughput technologies allow acquiring numerous soil and plant data at affordable costs in the perspective of customizing recommendations. Large and diversified datasets must be acquired uniformly among stakeholders to diagnose soil quality and plant nutrition at local scale, compare side-by-side defective and successful cases, implement trustful practices and reach high resource-use efficiency. Machine learning methods can combine numerous edaphic, managerial and climatic yield-impacting factors to conduct nutrient diagnosis and manage nutrients at local scale where factors interact. Compositional data analysis are tools to run numerical analyses on interacting components. Fractal models can describe aggregate stability tied to soil conservation practices and return site-specific indicators for decomposition rates of organic matter in relation to soil tillage and management. This chapter reports on machine learning, compositional and fractal models to support wise decisions on crop fertilization and soil conservation practices.

Keywords: Datasets, factor-specific management, fractal analysis, machine learning, nutrient balance, soil quality

1. Introduction

With the world population expected to reach more than 9×10^9 people by 2050, the food demand must increase by 70% in a situation where yield average of several staple crops is expected to decline [1]. More than 95% of our food is produced on soil [2]. Despite the general perception that soil is an abundant resource, the reality is that the soil resource is degrading at fast rate as a result of salinization, erosion, compaction, contamination, structure collapse, acidification, loss of organic matter and biological activities, as well as land allocation to urban and industrial development. Gains in technology alone will not suffice to compensate the harmful agricultural practices thought heroically to maintain soil productivity and farm viability on the long run. Understanding comprehensively how agroecosystems build and function worries more. Two centuries ago, German scientist Alexander von Humboldt warned that

management of living systems must be based on the rigorous collection of contextual facts and local knowledge [3]. His thoughts translate today into data acquisition from diverse sources, data mining and data processing methods to assist making wise decisions on how to manage soils properly at local scale.

The land is the basic resource for food production. There is a need to develop soil quality criteria and implement them where it matters most. [4] attributed large disparities in decision-making thought naively to manage soils properly to unequal, insufficient or inadequate collection of information, widespread ignorance on how agroecosystems function, lack of understanding on how factors interact, and the wrong perception that business-oriented economic and social values outweigh environmental damages or beneficial ecosystem services. Indeed, high crop productivity relies on positive interactions between climatic, managerial and edaphic factors [5]. Data must be integrated into comprehensive decision-making models to manage complex systems sustainably. High-quality and diversified information reduces the risk of making wrong decisions based on regional averages rather than at the right interaction level at field scale [6, 7]. Judicious decisions on locally acceptable actions should rely on well-documented facts and sound knowledge of environmental conditions. Besides traditional means to diagnose soil-plant systems, progress on data acquisition tools includes proximate and remote sensing, high-throughput laboratory technologies or on-the-go data acquisition kits of precision agriculture.

Several diagnostic models support decisions on soil and nutrient management. While soil properties and plant compositions have been addressed as separate variables in reductionist models [8], empirical-mechanistic models were developed to synthesize more data, balancing untestable and testable concepts [9–11]. This required not only sufficient data input, but also calibrating empirical coefficients and validating the results in a wide variety of environments. More recently, modern tools of artificial intelligence allowed to process large and diversified datasets in relation with ecosystem performance based on Alexander von Humboldt's principles of biogeography [3].

On the other hand, soil and plant analytical data are inherently multivariate compositional data constrained to the measurement unit, posing a serious numerical problem of “resonance” within the constrained space of compositions, such as 100% or the unit of measurement [12]. Ternary diagrams were the first representations of the closed space of three interrelated variables [13]. [14] related tissue N, P and K concentrations in a ternary NPK diagram to delineate the space of successful tissue compositions. It was not until [12] that ternary diagrams formed the basis of an emerging and appealing field of mathematics called “Compositional Data Analysis” (CoDa). CoDa rely on log ratio transformations. [15] developed means to project compositions as coordinates in the Euclidean space. The CoDa concepts corrected computational errors and fallacies in earlier plant and soil diagnostic models [16, 17].

On the other hand, the fractal theory has been useful to address the geometry of soil aggregation [18] and the kinetics of carbon decomposition in soils [19]. Fractal kinetics assigned to time a coefficient between 0 and 1 to explain the reduction in decomposition rate due to reduced contact between organic matter particles and their immediate environment resulting from aggregate buildup with time [19]. Fractal coefficients also provided a description of aggregate fragmentation patterns upon mechanical stress and avoided computational errors reported in classical synthetic measures of aggregation [20].

Machine learning, compositional and fractal modeling tools can process large and diversified soil-plant datasets that allow conducting side-by-side comparisons between failure and success. We hypothesized that well-informed models can assist

making wise decisions on soil and nutrient management at local scale. In this chapter, we address carbon sequestration and factor-specific fertilization to sustain soil productivity and support resource conservation actions.

2. Datasets

2.1 Growth-limiting factors

Field trials to document practices are conducted under the assumption that all factors but the ones being varied are equal or at optimum levels. Liebscher's law of the optimum stated that "a production factor which is in minimum supply contributes more to production, the closer other production factors are to their optimum" [8]. The law of the maximum aimed to optimize controllable factors given the impossibility to modify factors that are not controllable in the present state of knowledge and technology [21]. A provisional list of growth-impacting factors is provided in **Table 1**.

Noncontrollable factors under field conditions (more than 20)	Partially controllable factors under field conditions (more than 40)
1. Day-night temperatures	1. Soil available essential and beneficial nutrients: N, P, K, Ca, Mg, S, Fe, Zn, Mn, Cu, Mo, B, Na, Ni, Se, Si ...
2. Precipitations	2. Soil salinity and sodicity (Na leaching)
3. Radiation	3. Soil pH
4. Wind	4. Soil organic matter and carbon sequestration
5. Slope of the land	5. Soil texture
6. Altitude and latitude	6. Surface crusting potential
7. Number of frost-free days	7. Soil tillage
8. Number of chilling hours	8. Plowing depth
9. Photoperiod	9. Soil aggregation
10. Light intensity	10. Fertilization
11. Percent sunshine	11. Liming
12. Radiation	12. Irrigation
13. Relative humidity	13. Gypsum amendment
14. Precipitations	14. Water table level
15. Air contamination	15. Soil moisture
16. Soil texture	16. Serpentine characteristics
17. Cation exchange capacity	17. Pest management (insects, rodents, birds, other wild animals, plant diseases, soil-borne diseases, weeds, ...)
18. Phosphorus sorption capacity	18. NH ₄ :NO ₃ ratio
19. Micronutrient sorption capacity	19. Water and wind erosion
20. Carbon dioxide level	20. Plant population
21. Soil genesis and stratification	21. Planting date
22. Soil profile thickness	22. Soil aeration
23. Soil rockiness and stoniness	23. Soil water permeability
24. Etc.	24. Cultivar
	25. Crop rotation
	26. Toxicity from trace elements
	27. Evapotranspiration
	28. Seed bed preparation
	29. Crop residues
	30. Pesticide residues
	31. Growth regulators
	32. Date of harvest
	33. Quality of irrigation water
	34. Fertilizer placement, source, rate, timing

Table 1.
 Partial list of noncontrollable and partially controllable growth-limiting factors [21, 22].

Nutrient interactions impact crop yield through synergism, antagonism, dilution, excess, toxicity or crosstalks. Nutrient interactions are addressed as pairwise ratios [23]. Nutrient crosstalks occur where change in sulfur availability alter tissue compositions of micronutrients [24]. An extreme case of nutrient excess is toxicity where vital processes are affected. In field experiments, synergism is also viewed as positive interaction occurring where plant response is greater by combining two nutrients than from individual effects [25]. A list of nutrient interactions is presented in **Table 2**.

Face to the formidable task to optimize tens of growth-limiting factors and myriads of factor interactions, most of them being unknown, each case under study could rather be viewed as unique combinations of factors. For successful cases in the neighborhood, most factors are equal except those impacting the performance of defective specimens, facilitating side-by-side comparisons.

Nutrient	Interaction
N	Positive: NH ₄ with NO ₃ , P, Fe, Mn, Zn; NO ₃ with Ca, Mg, K; P, K ↑ if (NH ₄) ₂ SO ₄ Negative: NH ₄ with Ca, Mg, K; NO ₃ with Fe, Mn, Zn; Ca, Mg, Cu, Mn, Zn ↓ if (NH ₄) ₂ HPO ₄ Concentration ↑ if N deficient: P, K, Ca, Mg, S, B, Fe, Mn
P	Concentration ↓ if P deficient: N, P, K, Ca, Mg Concentration ↑ if P in excess: N, P, Ca, Mg, B, Mo Concentration ↓ if P in excess: K, Cu, Fe, Mn, Zn, Se
K, Ca, Mg	<ul style="list-style-type: none"> • $\sum K + Ca + Mg \approx constant$, hence competition at absorption sites • Antagonisms: K ↑, Ca and Mg ↓; Mg ↑, K ↓ more than Ca ↑, K ↓ • Effect on soil aggregation: Ca > Mg where [Mg] is low • Effect on soil degradation: Na > K > Mg
K	<ul style="list-style-type: none"> • Synergism: K-NO₃ • Competition for plant absorption and in clay minerals: K-NH₄ • Antagonism: P reduces negative K effect on Mg; if K ↑, Na, B, Mn, Mo, Zn ↓
Ca	<ul style="list-style-type: none"> • Synergism: N ↑, Ca ↑, especially if NO₃-N • Antagonism: NH₄, K, Na, Mg ↑ if Ca ↓; • Ca demand ↓ if Cd, Al, Cu, Fe, Mn, Zn ↓ • Ca(CO₃)₂ ↑ pH, CaSO₄ pH →; CaSO₄ could neutralize Al³⁺ in the subsoil • Ca(OH)₂: formation of “pouzzolane” cementing clayey soils
Mg	<ul style="list-style-type: none"> • Synergism: if Mg ↑, P, B, Fe, Mn, Mo, Na, Si ↑ • Antagonism: NH₄ ↑, Mg ↓
S	<ul style="list-style-type: none"> • Crosstalks with Mo, Cu, Fe, Zn, B • N:S, P:S ratios for protein synthesis
B	<ul style="list-style-type: none"> • Dilution: if N, K ↑, B ↓ • If P deficiency, B ↑; if Ca deficiency, B ↑; if B toxicity, Ca ↑
Cu	<ul style="list-style-type: none"> • Organic matter increases Cu fixation • If N, P, K ↑, Cu ↓; if Cu ↑, Fe, Mn, Zn ↓
Fe	<ul style="list-style-type: none"> • High pH, P, Ca, Cu, Zn, Mn ↑, Fe ↓ • If NH₄ ↑, Fe ↑
Mo	<ul style="list-style-type: none"> • Mo-N: reduction of NO₃ • If Mo ↑, P, Mn ↑, K, S ↓; if Mo ↓, Fe ↓
Zn	<ul style="list-style-type: none"> • If NH₄ ↑, Zn ↑; if P ↑, Zn ↓; if Zn ↑, K, Ca, Mg, S ↓
Mn	<ul style="list-style-type: none"> • If NH₄ ↑, Mn ↑; if Mn ↑, B, Mo ↑; if Mn ↑, Ni ↓

Table 2.
Nutrient interactions in soils and plant tissues [23–29].

2.2 Soil quality indicators

In Canada and Brazil as well as in other countries, soil mismanagement led to soil degradation [30, 31]. There is a great challenge to address soil problems and optimize resource-use efficiency to sustain soil productivity [32]. Soil quality impacts nutrient supply and resistance to erosion [33, 34]. [4] provided a list of biological, chemical and physical indicators of soil quality measurable at various scales of agroecosystems (**Table 3**). Biological indicators are presently the least documented but technologies of metagenomics will fill this gap in years to come [35]. Point-scale indicators can be integrated into maps to guide precision agriculture at field or subfield level. It is still difficult to evaluate soil quality uniformly among stakeholders with respect to soil threats, soil multifunctionality and ecosystem services [36].

Biological	Chemical	Physical
Point-scale indicators		
<ul style="list-style-type: none"> • Microbial biomass • Potential N mineralization • Particulate organic matter • Respiration • Earthworm counting • Microbial communities • Biological diversity • Fatty acid profiles • Mycorrhiza populations • Potential rooting depth • Root development 	<ul style="list-style-type: none"> • pH • Organic C and N • Labile organic magttger • Soil test nutrients • Electrical conductivity • Heavy metals • CEC and base saturation • Cesium-137 distribution • Xenobiotic loadings • Soil tests • Tissue tests 	<ul style="list-style-type: none"> • Aggregate stability • Aggregate-size distribution • Soil porosity and compaction • Bulk density and porosity • Penetration resistance • Shear strength • Slaking/dispersion • Water-filled pore space • Available water • Crust formation/strength • Infiltration, surface ponding • Soil structure, consistency • Profile depth • Soil stratification • Soil color, mottles
Field-, farm-, watershed-scale indicators		
<ul style="list-style-type: none"> • Crop yield • Weed infestation • Disease and insect pressure • Wild animal pressure • Nutrient deficiencies • Growth characteristics 	<ul style="list-style-type: none"> • Soil organic matter change • Nutrient loading or mining • Heavy metal accumulation • Changes in salinity • Leaching or runoff losses • Drainage, irrigation water 	<ul style="list-style-type: none"> • Topsoil thickness and color • Compaction/ease of tillage • Ponding (infiltration) • Rill and gully erosion • Surface residue cover
Regional-, national-, international-scale indicators		
<ul style="list-style-type: none"> • Productivity, yield stability • Species richness, diversity • Keystone species • Ecosystem engineering • Biomass density, abundance 	<ul style="list-style-type: none"> • Acidification • Salinization • Water quality changes • Air quality changes (dust and chemical transport) 	<ul style="list-style-type: none"> • Desertification • Loss of vegetative cover • Wind and water erosion • Siltation of rivers and lakes

Table 3.
Indicators of soil quality [4, 35].

3. Diagnostic methods

3.1 Soil test diagnosis

The sufficiency level of available nutrients (SLAN), the basic cation saturation ratio (BCSR), and soil test buildup and maintenance (STBM) are the main soil test interpretation philosophies [34]. The SLAN and BCSR addressed the relatively immobile nutrients (P, K). The STBM was used to manage N, P, and K. Critical and maintenance soil test levels were delineated from field trials.

Bray (1963) [22] assumed that (1) for nutrients relatively immobile in soils such as P and K, soils and fertilizers have nutrient-supply coefficients specific to plant species, planting patterns and rates, provided that soil and climatic conditions are similar and (2) response patterns can be described by the Mitscherlich equation. The SLAN related soil test P and K to percentage yield using the Mitscherlich-Bray equation. Alternatively, the relationship was partitioned into soil fertility classes each given a probability of response to fertilization [34, 37]. Compared to actual yield, percentage yield showed higher correlation with soil test level. Percentage yields have been first expressed as yield at 0-level of nutrient, other factors assumed to be at adequate levels, divided by yield where all factors were assumed to be at adequate levels. Percentage yields were also expressed as response ratios, i.e., $\ln(Y_{treatment}/Y_{control})$, i.e. yield gain of treatment over that of control, to run metaanalysis at regional scale [38]. Using yield percentage and probability of response, the SLAN concept assumed random effects across factors not being varied and thus hid the effects of local factors that impact crop yield.

The BCSR postulated, without proper calibration, that “ideal” cationic ratios and saturation levels should be maintained on soil cation exchange capacity to maximize yield [28]. The application of such concept to fertilization decisions failed under field conditions, most often leading to overfertilization [39]. Nevertheless, BCSR may assist making decisions on liming and lime sources to neutralize soil acidity, provide proper cementing agents bridging soil particles and improve soil aggregation [24]. In comparison, compositional data analysis methods proved to be a more appropriate approach to run statistical analysis on results of soil tests for cations and other cementing agents [29, 40].

The STBM concept has been elaborated from nutrient budgets, nutrient-use efficiency and soil P-fixing capacity as an attempt to adjust fertilization to local conditions. Expected yield and plant- and soil-specific coefficients were assessed from field observations and pot trials [41]. Soil P fixing capacity has been assessed in priority in Brazil, but coefficients estimated from literature often proved to be unrealistic, leading to overfertilization at local scale, especially for P [42].

Transferring SLAN, BCSR and STBM regional models to the local scale cannot be a straightforward operation. Growers’ heuristics is traditionally to look for successful practices developed under comparable environmental and managerial conditions as reported in their neighborhood. Alternatively, large and diversified datasets can be documented and synthesized into a diagnostic kit of features easy-to-acquire by stakeholders at reasonable cost and effort among those presented in **Tables 1** and **3**. The minimum package of facts, factors and local knowledge supporting fertilization decisions can be handled by machine learning models to diagnose growth-limiting factors and predict crop yields after correction. Thereafter, compositional data analysis can rank diagnosed components in the order of their limitations to yield to support nutrient management [43–46]. Yield can be predicted in regression mode. Besides, the classification mode can provide a list of high-yielding and balanced specimens as benchmarks for use at local scale, as well as the probability to yield more than some yield target.

3.2 Soil quality diagnosis

The interpretation of soil quality indicators requires well defined values, otherwise, the indicators cannot be used in practice to support management decisions [35]. Benchmarks could be native soil, reference sites, or successful combinations of comparable factors for agronomically or environmentally performing soils. Scores could have thresholds for (1) more than is better, (2) optimum range, (3) less than is better, or (4) undesirable range [47]. Principal component analysis (PCA), redundancy analysis (RDA), discriminant analysis and multiple regression have been used to process data.

Soil aggregation is a key indicator of soil quality. Mean weight diameter (MWD) is a common indicator of soil aggregation computed as follows:

$$\text{MWD} = \sum_{i=1}^D \bar{x}_i w_i \quad (1)$$

Where \bar{x} is aggregate diameter and w_i is the mass of the i^{th} aggregate fraction. Mean particle diameter is assessed as average sieve size between successive sieves rather measured as average particle size. The contribution of the largest fractions is inflated artificially by multiplying the fraction by its diameter.

The MWD is numerically biased, unevenly weighted, and computed from aggregate-size fractions that vary widely among studies [40]. Alternatively, patterns of aggregate fragmentation can be synthesized into fractal dimensions. It is assumed that aggregates collapse following mechanical stress into smaller fragments of similar shape. Aggregates left on each sieve are counted after subtracting the sand fraction ($> 53 \mu\text{m}$) on each sieve [40] as follows:

$$N(d_i) = M(d_i) / (d_i^3 \rho_i c_i) \quad (2)$$

Where $N(d_i)$ is the number of particles, $M(d_i)$ is the mass of aggregates of the i^{th} aggregate-size fraction, d_i is mean diameter and ρ_i is bulk density. Note that ρ_i must differ between the stronger and denser micro- and the more friable macro-aggregates. The shape coefficient c_i refers to a cube. Particle volume can be computed as x^3 , x being the average opening between two successive sieves.

The fractal dimension D_f is estimated as follows:

$$S(d_k) = \sum_i^k N(d_i) = \alpha d_k^{-D_f} \quad (3)$$

Where $S(d_k)$ is the cumulated number of particles with diameter $\leq d_k$, $N(d_i)$ is the number of particles in the i^{th} size fraction, α is a proportionality parameter, and D_f , the fragmentation fractal dimension, is a scaling factor derived from the log-log relationship between $S(d_k)$ and d_k .

The fractal model for soil aggregation is presented in **Table 4** and **Figure 1**. The fractal was found to be 2.51 (slope), indicating well aggregated soil. Fractal dimensionality is generally between 2 and 3 for the 3-D soil aggregates, but may exceed even 3, a result difficult to interpret physically. Aggregate-size fragments have contrasting friability, often showing several fractal patterns. However, the fractal dimensions have the disadvantage of being assessed from a limited number of sieves.

Carbon sequestration plays a key role to enhance soil quality and abate greenhouse gases. Because aggregates reduce the contact between the organic substrate

Sieve Class	Diameter (x)	Mass	Bulk density	N(di)	N(dk)	log(x)	log N(dk)
mm		kg	g cm ⁻³				
2.00-1.40	1.70	0.0813	1.287	0.013	0.013	0.230	-1.891
1.40-1.00	1.20	0.0659	1.326	0.029	0.042	0.079	-1.380
1.00-0.50	0.75	0.0787	1.398	0.133	0.175	-0.125	-0.757
0.50-0.425	0.4625	0.0242	1.397	0.175	0.350	-0.335	-0.456
0.425-0.25	0.3375	0.0171	1.416	0.313	0.663	-0.472	-0.178
< 0.25-0	0.125	0.0332	1.477	11.498	12.161	-0.903	1.085

Table 4. Computation of variables $\log(x)$ and $\log N(dk)$ to derive the fractal dimension of soil aggregates.

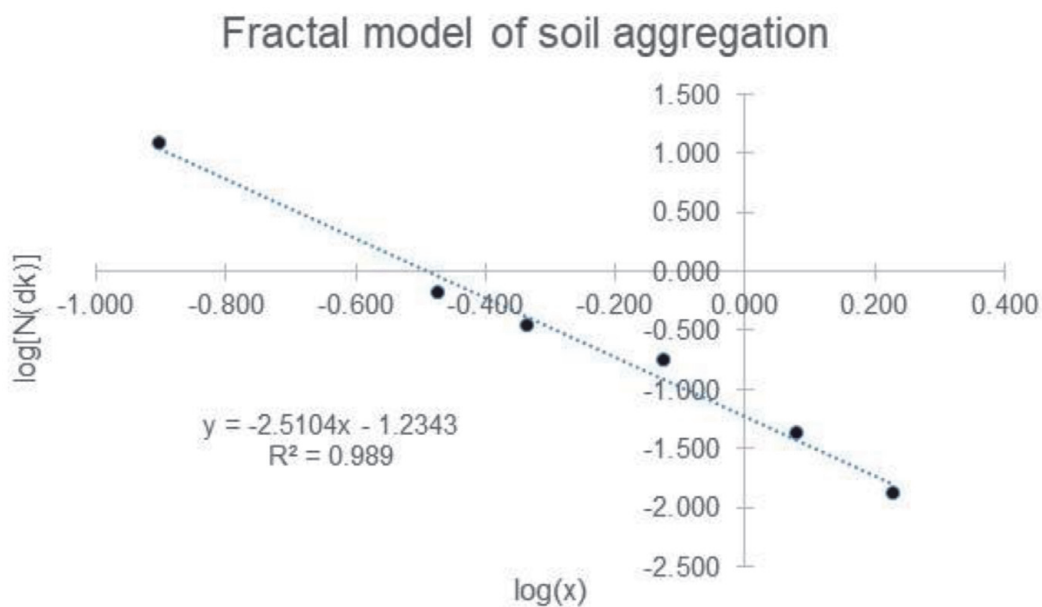


Figure 1. Fractal dimension of that soil aggregation pattern is 2.51.

and its immediate environment as they build up in soils, the decomposition rates of organic particles decrease with time, allowing organic matter to accumulate [19]. First-order kinetics of organic matter decomposition in soils $k(t)$ is controlled by fractal coefficient h as follows:

$$k(t) = k_1 t^{-h} \tag{4}$$

Where k_1 is decomposition rate at time $t = 1$ and h is fractal coefficient. If $h \rightarrow 0$, k is non-fractal and the reaction proceeds at maximum rate; if $h \rightarrow 1$, decomposition rate is fractal, indicating that protection mechanisms control reaction rate during soil agradation or degradation. [19] found fractal coefficient of 0.71 for well-aggregated soils under pasture compared to 0.45 for annual cropping and 0.25 for a degraded soil under fallow. Hence, the fractal coefficient is a measure of carbon protection mechanisms developing as soil quality increases or of loss in protection mechanisms leading to soil degradation.

The soil aggregation has also been expressed in terms of isometric log ratios (ilr) or coordinates [40]. The ilr is computed as a balance between two groups of aggregate fractions, as follows:

$$ilr = \sqrt{\frac{rs}{r+s}} \ln \left(\frac{G_1}{G_2} \right) \quad (5)$$

Where r and s are numbers of aggregate-size fractions at numerator and denominator, respectively, and G_1 and G_2 are geometric means of aggregate-size fractions at numerator and denominator, respectively. The balance dendrogram in **Figure 2** is a system of balances among five aggregate-size fractions starting with a general balance between micro- (< 0.25 mm) and macro- (> 0.25 mm) aggregates where $r = 4$ (the number of macro-aggregate fractions) and $s = 1$ (the micro-aggregate fraction). The balance between micro- and macro-aggregates in **Table 4** is computed as follows:

$$\begin{aligned} ilr_{[\text{microaggregates} \setminus \text{macroaggregates}]} &= \sqrt{\frac{5 \times 1}{5 + 1}} \ln \left(\frac{(0.0813 \times 0.0659 \times 0.787 \times 0.0242 \times 0.0171)^{1/5}}{(0.0332)^1} \right) \\ &= 0.268 \end{aligned} \quad (6)$$

Because ilr transformation allows projecting compositions into the Euclidean space, Euclidean distance ε can be computed between two soil aggregation states across ilr dimensions to indicate whether the soil is degrading or agrading, as follows [40]:

$$\varepsilon = \sqrt{\sum_{j=1}^{D-1} (ilr_j - ilr_j^*)^2} \quad (7)$$

Where j is a compositional dimension. Because computations are made on a mass basis rather than particle counts as for fractal dimensions, there is no need to make assumptions about ρ_i and c_i . The benchmark aggregation state could be defined as ultimate aggregation state where all aggregates pass through the smallest sieve size.

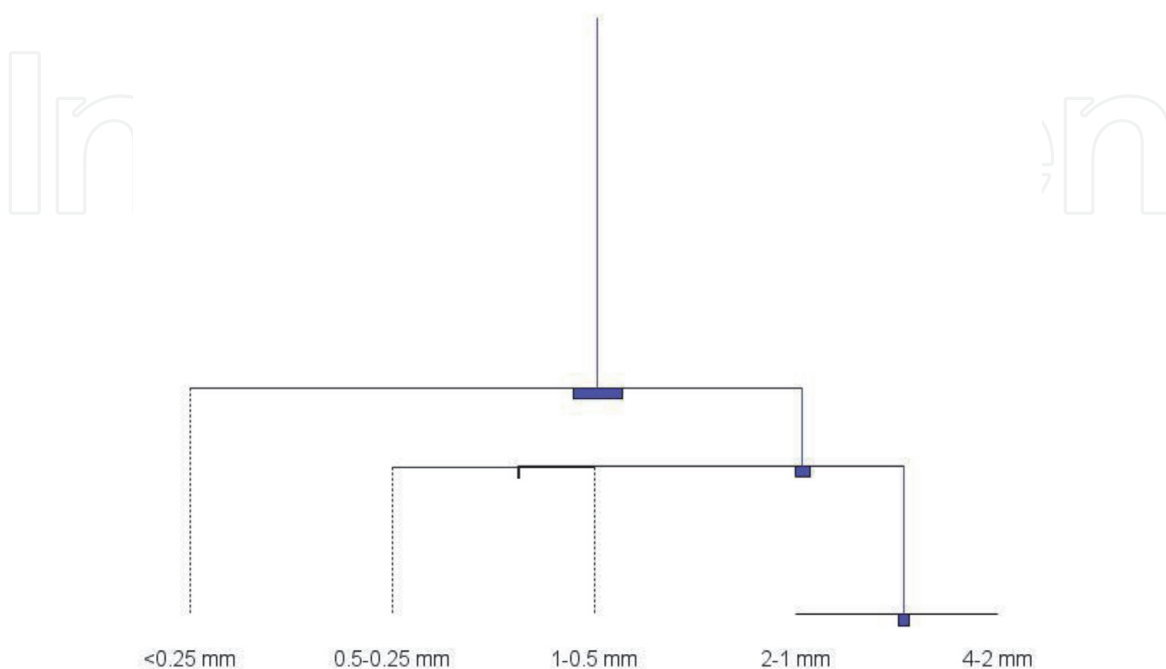


Figure 2.
 Balance dendrogram contrasting micro- and macro-aggregates and macro-aggregates.

3.3 Tissue nutrient diagnosis

Early workers proposed to classify the results of tissue tests, that are continuous variables, using concentration ranges and critical values such as poverty adjustment (deficiency), critical percentage, and nutrient sufficiency, luxury consumption or excess (including antagonism and toxicity) [48–50]. The critical percentage was the tipping point on the response curve, located at 90–95% maximum yield. Nutrients were diagnosed separately rather than as unique combinations of interactive nutrients. Although the reject/accept dichotomania led to considerable interpretation uncertainties [17], the one-nutrient-at-the-time approach is still commonly used today. [51] suggested using methods of multivariate analysis to handle tissue compositions as a whole rather than as separate components, ignoring the numerical pathologies of using inherently interrelated raw concentration values.

Dual ratios were thought to account for nutrient interactions [52]. The Diagnosis and Recommendation Integrated System (DRIS) has been elaborated to handle nutrient ratios [53, 54]. The DRIS required computing the mean and variance of dual ratios but did not fit into any method of multivariate analysis. Much earlier, [14] already developed a concept of optimum combinations of interactive nutrients within a ternary diagram (**Figure 3**). Because plants show various degrees of plasticity in response to growing conditions [55–57], they can adjust nutrient acquisition to nutrient stress [58–61]. This fits perfectly into the realm of Composition Data Analysis.

Because compositional vectors convey relative information, one should first ‘think ratios’ but, realizing that quotients are more difficult to handle than sums or differences, ‘think logratios’ [62]. Log ratios are log contrasts between components at numerator and denominator, respectively. While compositional data are constrained to the compositional space (e.g., 100%), log ratios can scan the real space, allowing to conduct statistical analyses and return confidence intervals without constraints. It was not until [12] developed the theory of Compositional data Analysis (CoDa) that ternary diagram could be expanded to more than three nutrients.

The Compositional Nutrient Diagnosis (CND) avoided several computational pathologies in DRIS such using different measurement units for macro- and micro-nutrients, pairwise rather than multivariate ratios, non-normal distribution, use of a dry matter basis as a separating component, assumed additivity of nutrient functions, non-symmetrical functions between dual ratios and their inverse, and non-symmetrical nutrient ratio and product functions. The CoDa also allowed

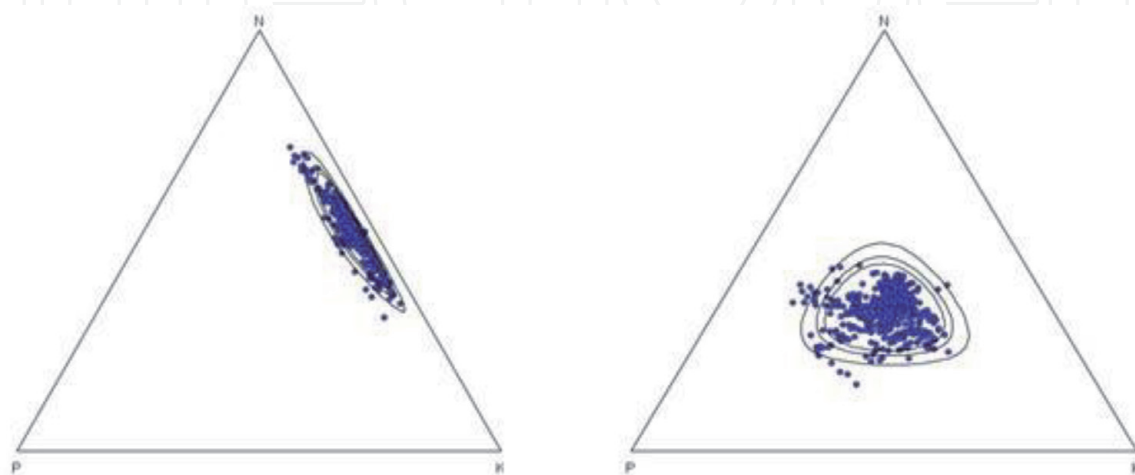


Figure 3.

Area of optimum balances between N, P and K in plant tissues uncentered (left) or centered (right) within a ternary diagram using the Codapack 2.01 freeware (ellipses with $p = 0.10, 0.05,$ and $0.01,$ respectively).

diagnosing multinutrient ratios in the Euclidean space [16] and conducting multivariate analyses in plant ionomics [58].

In CoDa, the simplex is closed to measurement unit using a filling value computed as follows:

$$F_v = 1000 - \sum_{i=1}^D c_i \quad (8)$$

Where F_v is the filling value for unit g kg^{-1} , D is the number of quantified components in the D -part composition, and c_i is concentration of each quantified part. The filling value is required to back-transform log ratio means into original concentration values. The centered log ratio [$clr = \ln(x_i/G)$] integrates all pairwise ratios into a single multinutrient expression, as follows for N:

$$clr_N = \ln\left(\frac{N}{G}\right) = \ln\left(\frac{N}{N}, \frac{N}{P}, \dots, \frac{N}{F_v}\right)^{1/D} \quad (9)$$

Where clr is centered log ratio, x_i is a component of the compositional simplex, and G is geometric mean across components including the filling value, expressed in exactly the same measurement unit. For plant tissue analysis showing 4% N, 0.325% P and 5% K, the filling value is $100\% - (4\% + 0.325\% + 5\%) = 90.75\%$. The clr value for N in that 4-part composition is computed as follows:

$$clr_N = \ln\left(\frac{4}{(4 \times 0.25 \times 5 \times 90.75)^{0.25}}\right) = -0.143 \quad (10)$$

Euclidean distance ε can be computed between two tissue states, one being diagnosed and another being used as benchmark composition, using clr or ilr as follows:

$$\varepsilon = \sqrt{\sum_{k=1}^D (clr_k - clr_k^*)^2} = \sqrt{\sum_{k=1}^{D-1} (ilr_k - ilr_k^*)^2} \quad (11)$$

The ilr has the advantage over clr that Euclidean distances can be computed across the selected Euclidean dimensions (**Figure 4**). Micronutrients can be balanced separately to avoid large variations due to tissue contamination. Moreover, macronutrients with concentrations moving in the same direction with time (N, P, K vs. Ca, Mg) [63, 64] can be set apart to address timeliness (**Figure 5**).

The CND based on clr aimed initially to replace DRIS for regional diagnosis [16, 42, 65–80]. Thereafter, a website service was made available to Brazilian growers (<https://www.registro.unesp.br/#!/sites/cnd/>). The standardized clr differences between clr values of the diagnosed (clr_j) and that of the reference subpopulation (clr_j^*) of true negative (high-yielding and nutritionally balanced) specimens weighted by the standard deviation (SD_j^*) ranked nutrients in the order of their limitation to yield, as follows [80]:

$$\text{Index}_{clr_j} = \frac{(clr_j - \overline{clr_j^*})}{SD_j^*} \quad (12)$$

At that time, the reference subpopulation was selected at regional scale using the Cate-Nelson partitioning procedure by iterating the Mahalanobis distance \mathcal{M} to maximize classification accuracy. The \mathcal{M} was computed as follows:

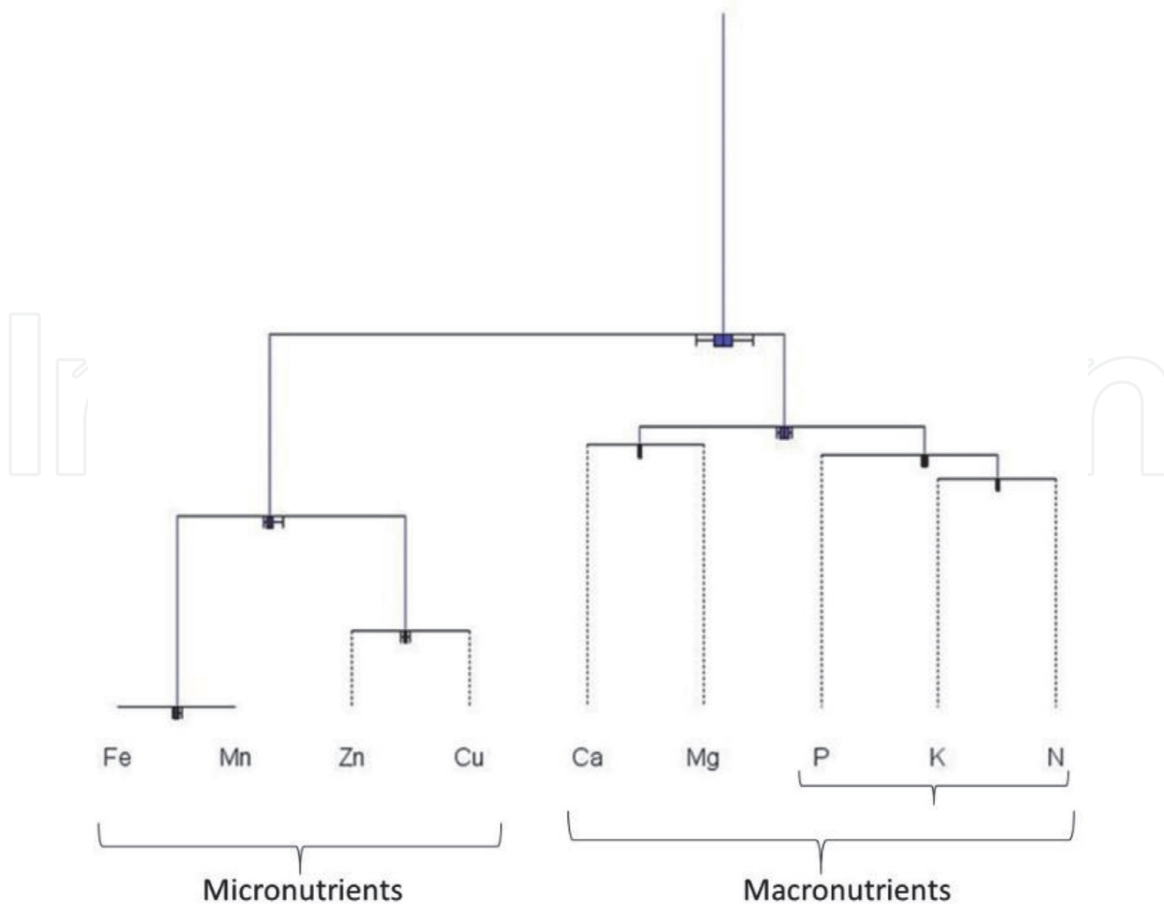


Figure 4. Balance dendrogram of tissue nutrient compositions of peach trees in southern Brazil, addressing micro- and macronutrients, then macronutrients moving in different directions with time.

$$\mathcal{M}_{ilr} = \sqrt{\sum_{j=1}^{D-1} (ilr_j - ilr_j^*) COV^{-1} (ilr_j - ilr_j^*)} \text{ or} \quad (13)$$

$$\mathcal{M}_{clr} \approx \sqrt{\sum_{j=1}^{D-1} (clr_j - clr_j^*) VAR^{-1} (clr_j - clr_j^*)} \quad (14)$$

The \mathcal{M}^2 is distributed like a χ^2 variable. The variance matrix is used where clr values are relatively independent from each other [80]. The use of D clr variables leads to singularity of the covariance matrix. This required removing one clr value, generally that of the filling value. [81] recommended using the ilr transformation rather than clr or the ordinary log transformation to conduct multivariate analysis due to the advantageous orthonormal basis of ilr variables.

The Cate-Nelson procedure returned four quadrants by point counting and thus allowed setting apart the subpopulation of true negative specimens, avoiding to include false positive specimens (high-yielding but nutritionally imbalanced) in the reference subpopulation, as was the case for DRIS and other nutrient diagnostic approaches. Quadrants are interpreted as follows:

1. True negative (TN) specimens showing high yield and nutrient balance
2. False negative (FN) specimens showing low yield despite nutrient balance (Type II error)

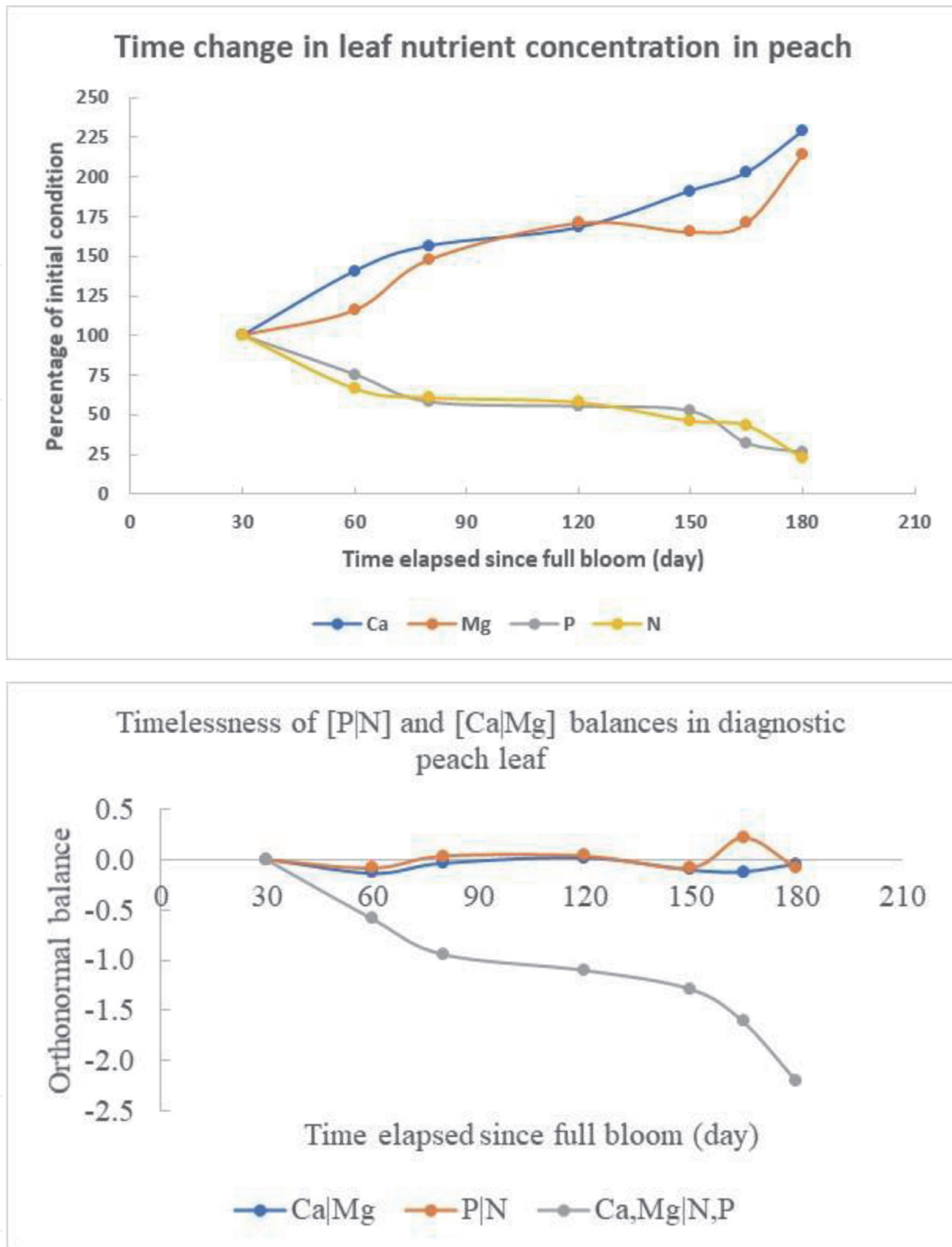


Figure 5. Time change in N, P, and K concentrations the leaf tissues of peach trees (data [64]). Balances between nutrient concentrations moving in the same direction with time are stationary (upper figure). As expected, the balance between [N, P, K] and [Ca, Mg] changes with time (lower figure).

3. True positive (TP) specimens showing low yield and nutrient imbalance
4. False positive (FP) specimens showing high yield despite nutrient imbalance (Type I error)

Model accuracy is determined as follows:

$$\text{Accuracy (\%)} = 100 \times \frac{\text{VN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \quad (15)$$

4. Machine learning methods to process large datasets

An introduction to machine learning methods is provided in [82]. “When dealing with complexity, mechanistic models become less obvious. System thinking, implying stocks and flows, becomes difficult to tune where species interact through varying functions over space and time ... most ecological patterns are nonlinear ... Another approach could rely purely on phenomenology with machine learning. Using this approach, we identify key features to predict outcomes using pattern detection”.

Machine learning is a family of methods of artificial intelligence that includes object similarity algorithms (k-nearest neighbors), decision trees (e.g., Random Forest), boosted decision trees (e.g., Gradient Boosting), multiple regression, gaussian methods, neural networks and several others, often tunable with hyperparameters. Machine learning methods can integrate numerous growth-impacting factors including soil quality indicators such as those documented by technologies of precision agriculture or supported by classical state- or industry-based agronomic models. Documenting as many growth-limiting factors as possible can decrease the number of assumptions required to diagnose nutrient problems at local scale, facilitating side-by-side comparisons. The confusion matrix generated by machine learning (ML) model in classification mode classified specimens into four quadrants by point counting, and thus allowed setting apart true negative specimens.

Compositional Data Analysis can be combined with machine learning methods to customize plant nutrient requirements for application at local scale where factor interactions shape fertilization decisions [17, 46, 83–86]. After running ML methods, it was suggested to use the *ilr* transformation to compute the Euclidean distance between the diagnosed (X) and successful (x) compositions, then compute the corresponding perturbation vector to rank nutrients in the order of their limitations to yield [44]. The perturbation vector is computed as follows [87]:

$$p = X \ominus x = \left[\frac{X_1}{x_1}, \dots, \frac{X_D}{x_D} \right], \text{ hence: } p = \left[\frac{N}{N^*}, \frac{P}{P^*}, \dots, \frac{F_v}{F_v^*} \right] \text{ or} \quad (16)$$

$$p = \left[\frac{N}{N^*} - 1, \frac{P}{P^*} - 1, \dots, \frac{F_v}{F_v^*} - 1 \right]. \quad (17)$$

The perturbation vector resembles the Deviation from Optimum Percentage [88]. Several log ratio transformation techniques other than *clr* and *ilr* are available but have not been tested yet [89].

4.1 Information flow

A flow of information from data acquisition to dataset organization and fertilizer recommendations at subfield level was described for lowbush blueberry (*Vaccinium angustifolium*) in Quebec [46], cranberry (*Vaccinium macrocarpon*) in Quebec and Wisconsin [85], and several crops in Brazil [17, 83, 84]. Nutrient diagnosis at local scale requires a well-documented dataset, an accurate machine learning model, a reliable model prediction algorithm, and a large set of ecologically diversified true negative specimens (**Figure 6**).

The bottleneck of machine learning models is knowledge gain on the learning curve. As anticipated 200 years ago by Alexander von Humboldt [3] a comprehensive understanding of living systems requires collecting facts and local knowledge

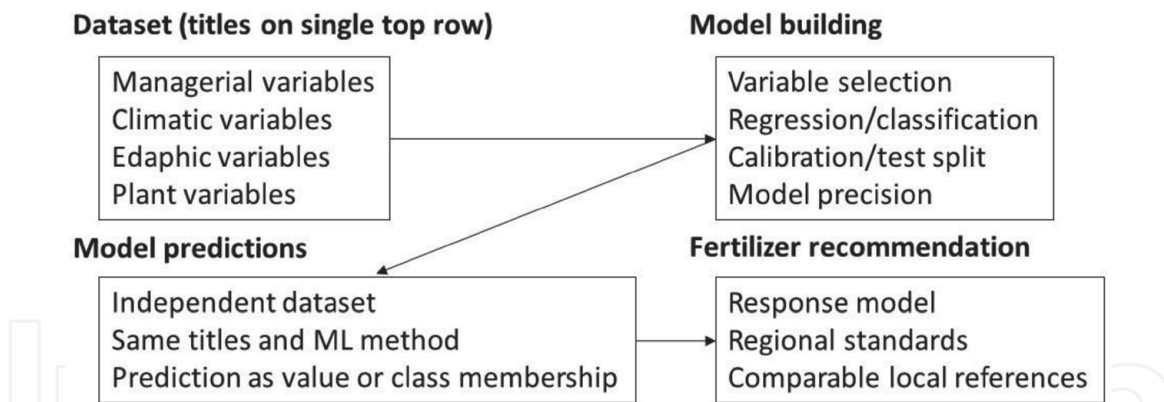


Figure 6.
 Flowchart of nutrient diagnosis in agroecosystems from data collection to fertilizer recommendations.

trustfully. Data can be observational as provided by growers, or experimental as retrieved from the published and the gray literature. Data sharing among stakeholders does not suffice to run machine learning. Data must be collected in a uniform way and cleaned from errors. Missing data could be imputed carefully or documented from other databases such as meteorological databases. Thereafter, data must be checked for their distribution to detect outliers.

A minimum dataset of meaningful features could be selected by adding or removing features (razor of Occam) without losing model accuracy during the model training process. Minimum data sets facilitate data acquisition by stakeholders at minimum cost and effort and make sense to them. The most performing machine learning model is selected. In general, the classification mode (yield class about yield cutoff) is more accurate than the regression mode. The classification mode returns the probability to exceed yield cutoff as targeted by the grower.

4.2 Local diagnosis

Features such as cultivar, rootstock, soil type or climatic conditions have been averaged to generate regional standards as “Frankenstein-built constructs” that may lead to unaccurate diagnosis at local scale where factors interact [17]. The local diagnosis often differs from regional diagnosis because the heroic assumption that “all controllable and uncontrollable factors but the ones being addressed are at equal or optimum levels” may fail at local scale. Indeed, the regional diagnosis is counter-intuitive to growers’ heuristics that compares normal to abnormal situations under similar conditions in their neighborhood [86]. Fertilizer recommendations can be customized using the fertilization regime of the closest compositional neighbors as reference, by modifying regional recommendations, from response curves, or using an optimization algorithm (Figure 7).

At local scale, the closest compositional neighbors are the true negative specimens showing similar growing conditions and the smallest compositional Euclidean distance from the diagnosed specimen. The nearest neighbors were said to be located in “Humboldtian loci or “enchanted islands”, “Ilhas Encantadas” in Portuguese, for a given set of uncontrollable factors. The grower has been pictured by [43] as a compositional parachutist manipulating nutrients as paracords to land on the closest “enchanted islands”. There, the resources to tackle controllable factors can be used parsimoniously and efficiently to reach trustful yield targets. Because the number of successful factor combinations is limited by the size and diversity of datasets, a close collaboration is required between stakeholders to collect facts and document local knowledge trustfully [6, 7, 90–94].

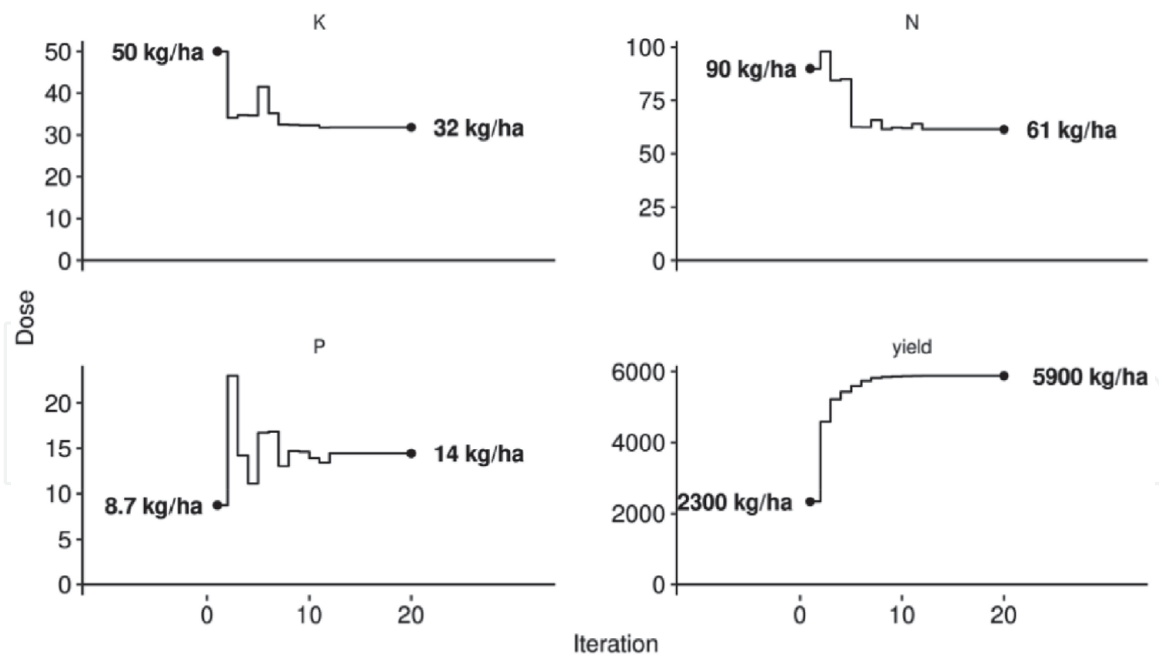


Figure 7. Fertilization recommendation using a Markov chain random walk algorithm to combine optimally N,, P and K dosage to increase yield from 2300 to 5900 kg berry ha⁻¹ for lowbush blueberry considering a set of corrected site-specific controllable factors (reproduced from [46]).

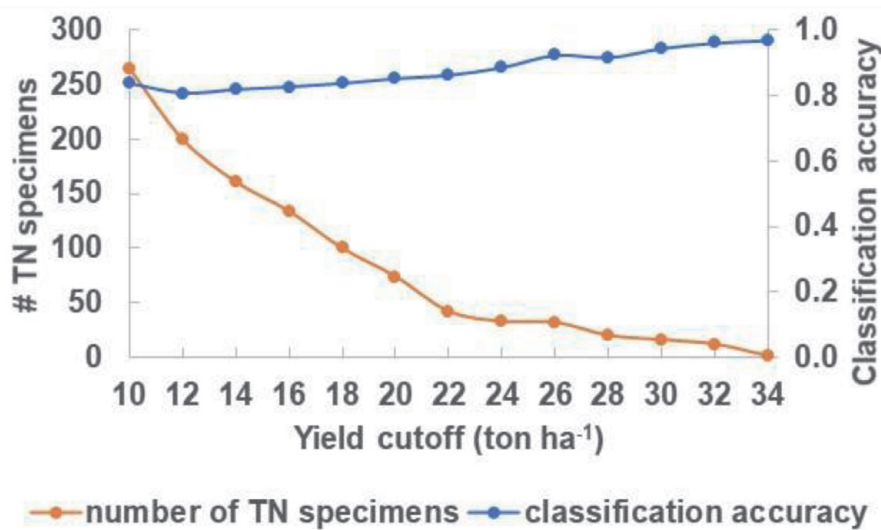


Figure 8. Dependence on yield cutoff of the number of true negative (high-yielding and nutritionmally balanced) specimens and classification accuracy.

The decision to fix a yield target in classification mode depends not only on growers' yield objective, but also on model precision and the number of true negative specimens available as close neighbors. The number of true negative specimens must be high because they provide benchmark compositions and trustfull yield targets under otherwise comparable growing conditions. As shown in **Figure 8** for the Brazilian peach tree dataset [83], classification accuracy increased slightly while the number of true negative specimens decreased exponentially as yield target increased. Smaller number of true negative specimens as benchmark compositions limits model's capacity to select local conditons close to those of the diagnosed specimen. In this case, the decision was to select 16 ton ha⁻¹ as cutoff yield, a reasonable yield objective.

5. Concluding remarks

In this chapter, we showed that fractal, compositional and machine learning models are promising alternatives to former empirical and mechanistic models to diagnose soil quality and plant nutrition at local scale and conduct side-by-side comparisons. Fractal kinetics confirmed that organic matter decomposition rates are controlled by protection mechanisms developing during organic matter transformation in soils. Site-specific coefficients can be assigned to decomposition rates under soil management practices. Compositional Data Analysis accounted for the special geometry of D -part compositions using log ratio transformations to tackle numerical bias before running numerical analyses. Machine learning methods can handle large and diversified datasets acquired through close collaboration between stakeholders. The CoDa methods can be combined with machine learning methods to diagnose nutrient imbalance and rank nutrients in the order of their limitation to yield by side-by-side comparison with successful neighbors.

This paper emphasized the need to change paradigm from the regional to the local scale to diagnose soil quality and plant nutrients and customize recommendations. Local features can be assembled in large and diversified numbers to address trustful feature combinations, then carved to a minimum data set impacting system's productivity and sustainability. Large and diversified data sets can be processed by methods of machine learning and compositional data analysis to reach the field or subfield scale. This requires collecting data uniformly and a close collaboration between stakeholders.

Author details

Léon Etienne Parent^{1,2*}, William Natale³ and Gustavo Brunetto¹


1 Departamento dos Solos, Universidade Federal de Santa Maria (UFSM), Santa Maria, Brazil

2 Department of Soils and Agrifood Engineering, Laval University, Quebec, Canada

3 Departamento de Fitotecnia, Universidade Federal do Ceará (UFC), Fortaleza, Brazil

*Address all correspondence to: leon-etienne.parent@fsaa.ulaval.ca

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Bahuguna RN, Jagadish KSV, Coast O, Wassmann R. 2014. Plant abiotic stress: temperature extremes. In: Van Alfen NK (Ed.), *Encyclopedia of Agriculture and Food Systems*, 2nd Ed, pp. 330-334, Elsevier. <https://doi.org/10.1016/B978-0-444-52512-3.00172-8>
- [2] FAO, 2015. Healthy soils are the basis for healthy food production. Available at: <http://www.fao.org/soils-2015/news/news-detail/en/c/277682/>
- [3] Keppel, G.; Kreft, H. Integration and Synthesis of Quantitative Data: Alexander von Humboldt's Renewed Relevance in Modern Biogeography and Ecology. *Front. Biogeogr.* **2019**, *11*, e43187.
- [4] Karlen, D.L., Andrews, S.S., Doran, J. W. 2001. Soil quality: current concepts and applications. *Adv, Agron.* **74**, 1-40.
- [5] Lemaire, G., Sinclair, T., Sadras, V., Belanger, G. 2019. Allometric approach to crop nutrition and implications for crop diagnosis and phenotyping. A review. *Agronomy for Sustainable Development* **39**, 2-17.
- [6] Anderson CJ, Kyveryga PM. Combining on-farm and climate data for risk management of nitrogen decisions. *Climate Risk Management* **2016**; Available from: [dx.doi.org/10.1016/j.crm.2016.03.002](https://doi.org/10.1016/j.crm.2016.03.002)
- [7] Kyveryga, P.; Caragea, P.C.; Kaiser, M.S.; Blackmer, T.M. Predicting Risk of Reducing Nitrogen Fertilization Using Hierarchical Models and On-Farm Data. *Agron. J.* **2013**, *105*, 85-94, doi:10.2134/agronj2012.0218.
- [8] De Wit, C.T. Resource Use in Agriculture. *Agric. Syst.* **1992**, *40*, 125-151.
- [9] Whisler, J.R., Acock, B., Baker, D.N., Fye, R.E., Hodges, H.F., Lambert, J.R., Lemmon, H.E., McKinion, J.M., Reddy, V.R., 1986. Crop simulation models in agronomic systems. *Adv. Agron.* **40**, 141-208.
- [10] Boote, K.J., Jones, J.W., Pickering, N.B. 1996. Potential uses and limitations of crop models. *Agron. J.* **88**, 704-716.
- [11] Brisson, N., Gary, C., Justes, E., Rocher, R., Mary, B., Ripoche, D., et al. 2003. An overview of the crop model STICS. *Europ. J. Agron.* **18**, 309-332.
- [12] Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.
- [13] Howarth, R. J. (1996). Sources for a history of the ternary diagram. *Br. J. Hist. Sci.* **29**, 337-356. doi:10.1017/S000708740003449X.
- [14] Lagatu, H., Maume, L., 1934. Le diagnostic foliaire de la pomme de terre. *Ann. Ec. Natl. Agron. Montp.* **22**, 50-158 (in French).
- [15] Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. *Math. Geol.* **35**, 279-300. <https://doi.org/10.1023/A:1023818214614>.
- [16] Parent, L.E., Dafir, M., 1992. A theoretical concept of compositional nutrient diagnosis. *J. Am. Soc. Hortic. Sci.* **117**, 239-242.
- [17] Paula, B.V. de, Arruda, W.S., Parent, L.E., Brunetto, G. Nutrient diagnosis of Eucalyptus at factor-specific level using machine learning and compositional methods. *Plants* **2020**, *9*, 1049, doi: 10.3390/plants9081049
- [18] Baveye, P., Parlane, J.Y., Stewart, B. A. 1997. *Fractals in soil science*. CRC Press, Boca Raton FL.

- [19] Parent, L.E. 2017. Fractal Kinetics Parameters Regulating Carbon Decomposition Rate under Contrasting Soil Management Systems. *Open Journal of Soil Science* 7, 111-117.
- [20] Diaz-Zorita, M., Perfect, E., Grove, J.H., 2002. Disruptive methods for assessing soil structure. *Soil and Tillage Research* 64, 3–22.
- [21] Wallace, A.; Wallace, G.A. Limiting Factors, High Yields, and Law of the Maximum. *Hortic. Rev.* **1993**, *15*, 409–448, doi:10.1002/9780470650547.
- [22] Bray, R.H. 1963. Confirmation of the nutrient mobility concept of soil-plant relationships. *Soil Sci.* 95(2), 124-130.
- [23] Wilkinson, S.R., 2000. Nutrient interactions in soil and plant nutrition. In: Sumner, M.E. (Ed.), *Handbook of Soil Science*. CRC Press, Boca Raton, FL, pp. D89–D112.
- [24] Courbet, G., Gallardo, K., Vigani, G., Brunel-Muguet, S., Trouverie, J., Salon, C., Ourry, A. 2019. Disentangling the complexity and diversity of crosstalk between sulfur and other mineral nutrients in cultivated plants. *J. Exp. Bot.* doi:10.1093/jxb/erz214.
- [25] Malavolta E. 2006. *Manual de nutrição mineral de plantas*. Ed. Agron. Ceres, São Paulo, Brazil.
- [26] Sumner, M.E. 1993. Gypsum and acid soils: the world scene. *Adv. Agron.* 51, 1-32.
- [27] Ulén, B., Etana, A. 2014. Phosphorus leaching from clay soils can be counteracted by structure liming, *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 64(5), 425-433. DOI: 10.1080/09064710.2014.920043
- [28] Chaganti, V., Culman, S.W. 2017. Historical perspective of soil balancing theory and identifying knowledge gaps: a review. *Crop Forage Turfgrass Manag.* 3, 1-7. Doi:10.2134/cftm2016.10.0072
- [29] Xu, Yan, Jimenez, M.A., Parent, S.-É., Leblanc, M., Ziadi, N., and Parent, L. E. (2017). Compaction of coarse-textured soils: balance models across mineral and organic compositions. *Frontiers in Ecology and Evolution* 28 h <https://doi.org/10.3389/fevo.2017.00083>
- [30] Sparrow HO. 1984. Soil at risk. Canada's eroding future. Standing Senate Committee on Agriculture, Fisheries and Forestry, The Senate of Canada, Ottawa ON, Canada.
- [31] Fushita AT, Camargo-Bortolin LHV, Arantes, EM, Moreira MAA, Cançado CJ, Lorandi R (2010). Fragilidade ambiental associada ao risco potencial de erosão de uma área da região geoeconômica médio Mogi Guaçu superior (SP). *Rev. Bras. Cartogr.* 63(4), 477-488.
- [32] Lal, R., Pierce, FJ. The vanishing resource. In: Lal R. and Pierce FJ, eds. *Soil management for sustainability*. Ankeny, Soil Water Conservation Society, 1991. p. 1-5.
- [33] Carter, M. R. 2002. Soil Quality for Sustainable Land Management: organic matter and aggregation interactions that maintain soil functions. *Agron. J.* 94, 38–47. doi:10.2134/agronj2002.3800.
- [34] Dahnke, W. C., and Olson, R. A. (1990). Soil test correlation, calibration, and recommendation, in *Soil testing and plant analysis, Third Edition*, ed. R. L. Westerman (Madison WI: Soil Science Society of America), 45–71.
- [35] Jeanne T, Parent S-É, Hogue R (2019) Using a soil bacterial species balance index to estimate potato crop productivity. *PLoS ONE* 14(3): e0214089.
- [36] Büsemann EK, Bongiorno G, Bai Z, Creamer RE, De Deyn G, de Goede, R,

- Fleskens L, Geissen V, Kuyper TW, Mäder P, Pulleman M, Sukkel W, van Groenigen JW, Brussaard L. 2018. Soil quality – A critical review. *Soil Biol. Biochem.* 120, 2018, 105-125.
- [37] Fitts, JW. 1955. Using soil tests to predict a probable response from fertilizer application. *Bettter Crops XXXIX*(3), 17-20.
- [38] Quinche-Gonzalez, M., Pellerin, A., Parent, L.E. 2016. Meta-analysis of lettuce (*Lactuca sativa* L.) response to added N in organic soils. *Can. J. Plant Sci.* 96(4), 670-676.
- [39] Kopitzke, P.M., Menzies, N.W. 2007. A review of the use of the basic cation saturation ratio and the “ideal” soil. *Soil Sci. Soc. Am. J.* 71(2), 259-265.
- [40] Parent, L.E., Almeida C.X. de, Parent, S.-É., Hernandez, A., Egozcue, J. J., Kätterer, T., Gülser, C., Bolinder, M. A., Andrén, O., Anctil, F., Centurion, J. F., Natale, W.. 2012. Compositional analysis for an unbiased measure of soil aggregation. *Geoderma* 179-180, 123-131.
- [41] Santos, F.C., Neves, J.C.L., Novais, R.F., Alvarez, V.V.H, Sediyaama C.S. Modeling lime and fertilizer recommendations for soybean. *R. Bras. Ci. Solo.* 2008;32: 1661-1674.
- [42] Nowaki, R.H.D., Parent, S.- É., Cecilio Filho, A.B., Rozane, D.E., Meneses, N.B., da Silva, J.A.D.S., Natale, W., Parent, L.E., 2017. Phosphorus overfertilization and nutrient misbalance of irrigated tomato crops in Brazil. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2017.00825>.
- [43] Parent, S.É. Why We Should Use Balances and Machine Learning to Diagnose Ionomes. *Authorea*, January 20, 2020, doi:10.22541/au.157954751.17355951.
- [44] Coulibali Z, Cambouris AN, Parent S-É (2020) Cultivar-specific nutritional status of potato (*Solanum tuberosum* L.) crops. *PLoS ONE* 15(3): e0230458. <https://doi.org/10.1371/journal.pone.0230458>
- [45] Parent, S.-É., Dossou-Yovo, W., Ziadi, N., Leblanc, M., Tremblay, G., Pellerin, A., Parent, L.E. 2020a. Corn response to banded P fertilizers with or without manure application in Eastern Canada. *Agronomy Journal*, DOI: 10.1002/agj2.20115
- [46] Parent, S.-É.; Lafond, J.; Paré, M.C.; Parent, L.E.; Ziadi, N. Conditioning Machine Learning Models to Adjust Lowbush Blueberry Crop Management to the Local Agroecosystem. *Plants.* 2020b;9(10): 1401. doi : 10.3390/plants9101401.
- [47] Tesfahunegn GB, 2014. Soil Quality Assessment Strategies for Evaluating Soil Degradation in Northern Ethiopia. *Applied and Environmental Soil Science* 2014, Article ID 646502, <http://dx.doi.org/10.1155/2014/646502>
- [48] Macy, P., 1936. The quantitative mineral nutrient requirements of plants. *Plant Physiol.* 11, 749–764. <https://doi.org/10.1104/pp.11.4.749>.
- [49] Ulrich, A., 1952. Physiological bases for assessing the nutritional requirements of plants. *Annu. Rev. Plant Physiol.* 3, 207–228. <https://doi.org/10.1146/annurev.pp.03.060152.001231>.
- [50] Ulrich, A., and Hills, F. J. (1967). “Principles and practices of plant analysis,” in *Soil testing and plant analysis. Part II*, eds. M. Stelly and H. Hamilton (Madison, Wisconsin: Soil Science Society of America), 11–24.
- [51] Holland, D. A. (1966). The interpretation of leaf analysis. *J. Hortic. Sci.* 41, 311–329.
- [52] Kenworthy, A.L., 1967. Plant analysis and interpretation of analysis

for horticultural crops. In: Stelly, M., Hamilton, H. (Eds.), *Soil Testing and Plant Analysis, Part II*. Soil Science Society of America, Madison, WI, pp. 59–75.

[53] Beaufilet, E. *Diagnosis and Recommendation Integrated System (DRIS)*, 1st ed.; University of Natal: Pietermaritzburg, South Africa, 1973.

[54] Walworth, J.L., Sumner, M.E., 1987. The diagnosis and recommendation integrated system (DRIS). *Adv. Soil Sci.* 6, 149–188. <https://doi.org/10.1007/978-1-4612-4682-4>.

[55] Gratani L. 2014. Plant phenotypic plasticity in response to environmental factors. *Adv. Botany* article ID 208747, <http://dx.doi.org/10.1155/2014/208747>

[56] Siebenkäs A., Schumacher J., Roscher C. 2015. Phenotypic plasticity to light and nutrient availability alters functional trait ranking across eight perennial grassland species. *AoB Plants* 7, plv029; doi:10.1093/aobpla/plv029

[57] Huang X.-Y. and Salt D.E. (2016). Plant Ionomics: From Elemental Profiling to Environmental Adaptation. *Mol. Plant.* 9, 787–797.

[58] Parent SÉ, Parent LE, Egozcue JJ, Rozane DE, Hernandez A, Lapointe L, et al. (2013b). The plant ionome revisited by the nutrient balance concept. *Front. Plant Sci.* 4, 1–10. doi: 10.3389/fpls.2013.00039.

[59] Baxter, I., 2015. Should we treat the ionome as a combination of individual elements, or should we be deriving novel combined traits? *J. Exp. Bot.* 66, 2127–2131. <https://doi.org/10.1093/jxb/erv040>.

[60] Jeyasingh PD, Goos JM, Thompson SK, Godwin CM, Cotner JB (2017) Ecological Stoichiometry beyond Redfield: An Ionomic Perspective on

Elemental Homeostasis. *Front. Microbiol.* 8:722. doi: 10.3389/fmicb.2017.00722

[61] Liu, S., Yang, X., Quan, Q., Lu, Z., Lu, J. 2020. An Ensemble Modeling Framework for Distinguishing Nitrogen, Phosphorous and Potassium Deficiencies Fertigated “Prata” and “Cavendish” Banana (*Musa* spp.) at Plot-Scale. *Plants* 2020, 9, 1467. in Winter Oilseed Rape (*Brassica napus* L.) Using Hyperspectral Data. *Remote Sens.* 2020, 12(24), 4060; <https://doi.org/10.3390/rs12244060>

[62] Aitchison J. The single principle of compositional data analysis, continuing fallacies, confusions and misunderstandings and some suggested remedies. 3rd Compositional Data Analysis Workshop, CoDawork 2008. Girona, Spain, 27-30 May 2008.

[63] Hill. J. (1980). The remobilization of nutrients from leaves. *J. Plant Nutr.* 2 (4), 407-444.

[64] Sumner, M.E. 1985. The Diagnosis and Recommendation Integrated System as a guide to orchard fertilization. Food and Fertilizer Technology Center Extension Bulletin 231, FFTC/ASPAC, Taipei, Taiwan.

[65] Parent, L.E., Natale, W., and Ziadi, N. 2009. Compositional Nutrient Diagnosis of Corn using the Mahalanobis Distance as Nutrient Imbalance Index. *Canadian Journal of Soil Science* 89:383-390.

[66] Parent, L.E., 2011. Diagnosis of the nutrient compositional space of fruit crops. *Rev. Bras. Frutic.* 33, 321–334. <https://doi.org/10.1590/S0100-29452011000100041>.

[67] Hernandez, A., Parent, S.-É., Natale, W., Parent, L.E., 2012. Balancing guava nutrition with liming and fertilization. *Rev. Bras. Frutic.* 34,

1224–1234. <https://doi.org/10.1590/S0100-29452012000400032>.

[68] Parent, L.E., Parent, S.-É., Hébert-Gentile, V., Naess, K., and Lapointe, L. 2013. Mineral balance plasticity of cloudberry (*Rubus chamaemorus*) in Quebec-Labrador. *Am. J. Plant Sci.* 4(7): 1508-1520.

[69] Parent, L.E., Parent, S.-É., Rozane, D.E., Amorim, D.A., Hernandez, A., Natale, W., 2012. Unbiased approach to diagnose the nutrient status of red guava (*Psidium guajava*). In: Santos, C.A.F. (Ed.), 3rd International Symposium on Guava and Other Myrtaceae, Petrolina, Brazil, April 23–25, 2012, pp. 145–159. <https://doi.org/10.17660/ActaHortic.2012.959.18> ISHS Acta Horticulturae, Paper #959.

[70] Parent, S.-É., Parent, L.E., Rozane, D.E., Hernandez, A., Natale, W., 2012. Nutrient balance as paradigm of soil and plant chemometrics. In: Issaka, R. N. (Ed.), Soil Fertility. IntechOpen Ltd., London, pp. 83–114. <https://doi.org/10.5772/53343>

[71] Parent, S.É., Parent, L.E., Rozane D. E. and Natale, W. 2013. Nutrient balance ionomics: case study with mango (*Mangifera indica*). *Frontiers Plant Science* 4, article 449. Parent,

[72] Parent, S.É., Barlow, P., and Parent, L.E. 2015. Nutrient balance of New Zealand kiwifruit (*Actinidia deliciosa*) at high yield level. *Communications in Soil Science and Plant Analysis* 46(1): 256-271.

[73] Deus, J. A. L., de, Neves, J. C. L., Corrêa, M. C. M., Parent, S.-É., Natale, W., Parent, L. E., 2018. Balance design for robust foliar nutrient diagnosis supervising the fertigation of banana “Prata” (*Musa* spp.). *Nature Scientific Reports* doi: 10.1038/s41598-018-32328-y

[74] Marchand, S., Parent, S.E., Deland, J.P., and Parent, L.E. 2013. Nutrient signature of Quebec (Canada) cranberry

(*Vaccinium macrocarpon* Ait.). *Rev. Bras. Frut.* 35(1):199-209.

[75] Modesto, V. C., Parent, S.-É., Natale, W., and Parent, L. E. (2014). Foliar Nutrient Balance Standards for Maize (*Zea mays* L.) at High-Yield Level. *Am. J. Plant Sci.* 5, 497–507. doi: 10.4236/ajps.2014.54064.

[76] Montes, R.M., Parent, L.E., de Amorim, D.A., Rozane, D.E., Parent, S.-E., Natale, W., Modesto, V.C., 2016. Nitrogen and potassium fertilization in a guava orchard evaluated for five cycles: effects on the plant and production. *Rev. Bras. Ci^enc. Solo.* <https://doi.org/10.1590/18069657rbcs20140532>.

[77] Souza, H.A., Parent, S.-É., Rozane, D.E., De Amorim, D.A., Modesto, V.C., Natale, W., Parent, L.E. 2016. Guava waste to sustain guava (*Psidium guajava*) agroecosystem: nutrient “balance” concepts. *Frontiers in Plant Science* 7: article 1252. DOI: 10.3389/fpls.2016.011252

[78] Rozane, D.E., Parent, L.E., Natale, W., 2015. Evolution of the predictive criteria for the tropical fruit tree nutritional status. *Cientifica* 44, 102–112. <https://doi.org/10.15361/1984-5529.2016v44n1p102-112>.

[79] Rozane, D.E., Mattos D. Jr., Parent, S. É., Natale, W., Parent, L.E. 2013. Compositional meta-analysis of Citrus varieties in the state of São Paulo, Brazil. *Scientia Agric.* 70(4):263-268.

[80] Badra, A., L.E. Parent, G. Allard, N. Tremblay, Y. Desjardins, and N. Morin. 2006. Effects of leaf nitrogen concentration versus CND nutritional balance on shoot density and foliage colour of an established Kentucky bluegrass (*Poa pratensis* L.) turf. *Canadian Journal of Plant Science* 86: 1107-1118.

[81] Filzmoser, P., Hron, K., and Reimann, C. (2009). Univariate

statistical analysis of environmental (compositional) data: problems and possibilities. *Sci. Total Environ.* 407, 6100–6108. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19740525>.

[82] Parent SÉ. 2020. Introduction to machine learning for ecological engineers. Nextjournal <https://nextjournal.com/essicolo/cc2020>

[83] 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.; Brunetto G. Humboldtian Diagnosis of Peach Tree (*Prunus persica*) Nutrition Using Machine-Learning and Compositional Methods. *Agronomy* 2020, 10, 900, doi:10.3390/agronomy10060900.

[84] Lima Neto, A.J.; Deus, J.A.L.; Rodrigues Filho, V.A.; Natale, W.; Parent, L.E. Nutrient Diagnosis of Fertigated “Prata” and “Cavendish” Banana (*Musa* spp.) at Plot-Scale. *Plants* 2020, 9, 1467.

[85] Parent LE, Jamaly R, Atucha A, Parent JE, Workmaster BA, Ziadi N, Parent SÉ. 2021. Current and next-year cranberry yields predicted from local features and carryover effects. *Plos ONE* 16(5), e0250575. <https://doi.org/10.1371/journal.pone.0250575>

[86] Munson, R.D.; Nelson, W.L. Principles and Practices in Plant Analysis. In *Soil Testing and Plant Analysis*; Westerman, R.L., Ed.; Soil Science Society of America: Madison WI, USA, 1990; pp. 359–387.

[87] Pawlowsky-Glahn, V, Egozcue, JJ. 2006. Compositional Data Analysis in the Geosciences: From Theory to Practice. Buccianti, A., Mateu-Figueras, G. and Pawlowsky-Glahn, V. (eds) Geological Society, London, Special Publications, 264, 1-10. The Geological Society of London 2006.

[88] Montañés L, Heras L, Abadía J, Sanz M. (1993) Plant analysis interpretation based on a new index: Deviation from optimum percentage (DOP), *J. Plant Nutr.*, 16:7, 1289-1308, DOI: 10.1080/01904169309364613

[89] Greenacre M. Compositional data analysis. *Ann. Rev. Stat. Appl.* 2021. 8, 271–99. <https://doi.org/10.1146/annurev-statistics-042720-124436>

[90] Parent, L. E.; Gagné, G. Guide de référence en fertilization. 2nd ed. Centre de Référence en Agriculture et Agroalimentaire du Québec (CRAAQ), Québec, Canada, 473 pp.

[91] Tremblay, N.; Bouroubi, Y.M.; Bélec, C.; Mullen, R.W.; Kitchen, N.R.; Thomason, W.E.; Ebelhar, S.; Mengel, D.B.; Raun, W.R.; Francis, D.D.; et al. Corn Response to Nitrogen Is Influenced by Soil Texture and Weather. *Agron. J.* 2012, 104, 1658–1671, doi:10.2134/agronj2012.0184.

[92] Morris, T.F.; Murrell, T.S.; Beegle, D.B.; Camberato, J.J.; Ferguson, R.B.; Grove, J.; Ketterings, Q.; Kyveryga, P. M.; Laboski, C.A.; McGrath, J.M.; et al. Strengths and Limitations of Nitrogen Rate Recommendations for Corn and Opportunities for Improvement. *Agron. J.* 2018, 110, 1–37, doi:10.2134/agronj2017.02.0112.

[93] Gibson, K.J.; Streich, M.K.; Topping, T.S.; Stunz, G.W. Utility of Citizen Science Data: A Case Study in Land-Based Shark Fishing. *PLoS ONE* 2019, 14, e0226782, doi:10.1371/journal.pone.0226782.

[94] Appenfeller, L.R.; Lloyd, S.; Szendrei, Z. Citizen Science Improves Our Understanding of the Impact of Soil Management on Wild Pollinator Abundance in Agroecosystems. *PLoS ONE* 2020, 15, e0230007, doi:10.1371/journal.pone.0230007.