

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

6,900

Open access books available

186,000

International authors and editors

200M

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



Nutrient Balance as Paradigm of Soil and Plant Chemometrics

S.É. Parent, L.E. Parent, D.E. Rozanne,
A. Hernandez and W. Natale

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/53343>

1. Introduction

Soil fertility studies aim to integrate the basic principles of biology, chemistry, and physics, but generally lead to separate interpretations of soil and plants data [1]. Paradoxically, J.B. Boussingault warned as far as in the 1830s that the balance between nutrients in soil-plant systems was more important than nutrient concentrations taken in isolation [2]. Indeed, the biogeochemical cycles of elements that regulate the dynamics of agroecosystems [3] do not operate independently [4]. However, raw concentrations of individual elements or their log transformation are commonly used to conduct statistical analyses on plant nutrients [5, 6,7], soil fertility indices [8] and C mineralization data [9, 10]. Researchers thus proposed several ratios and stoichiometric rules to relate system's components to each other when monitoring mineralization and immobilization of organic C, N, P and S in soils [4, 11], cations interacting on soil cation exchange capacity [12], nutrient interactions in plants [13, 14, 15, 16] and carbon uptake by plants [17, 18].

Different approaches have been elaborated to describe nutrient balances in soils. The nutrient intensity and balance concept (NIBC) computes ionic balances in soil water extracts [19, 20]. The basic cation saturation ratio (BCSR) concept hypothesizes that cations and acidity exchanging on soil cation exchange capacity (CEC) can be optimized for crop growth [12]. However, the BCSR has been criticized for its elusive definition of 'ideal' cationic ratios [21, 22]. In plant nutrition, [23] were the first to represent geometrically interactions between nutrients by a ternary diagram where one nutrient can be computed by difference between 100% and the sum of the other two. As a result, there are two degrees of freedom in a ternary diagram. One may also derive three dual ratios from K, Ca, and Mg but only two ratios are linearly independent because, for example, K/Mg can be computed from $K/Ca \times Ca/Mg$

and is thus redundant. Therefore, a ratio approach conveys $D-1$ degrees of freedom or linearly independent balances for a D -part composition [24].

In contrast, there are $D \times (D-1)/2$ dual ratios such as the K/Mg ratio and $D \times (D-1)^2/2$ two-component amalgamated ratios such as the K/(Ca+Mg) ratio that can be derived from a D -part composition. Most information on dual and two-component amalgamated ratios is thus redundant and the dataset is artificially inflated. In Figure 1, the number of (a) dual and (b) two-component amalgamated ratios is plotted against the number of components. With 10 components, one may compute up to 45 dual and 405 two-component amalgamated ratios, hence generating a “redundancy bubble” that inflates exponentially above D . [25] elaborated the Diagnosis and Recommendation Integrated System (DRIS) to synthesize the $D \times (D-1)/2$ dual ratios into D nutrient indices adding up to zero; therefore, there is still one redundant index closing the system to zero and computable from other indices. Applying Ockham’s razor law of parsimony to compositional data, nine degrees of freedom suffice to fully describe a 10-part composition without bias [24].

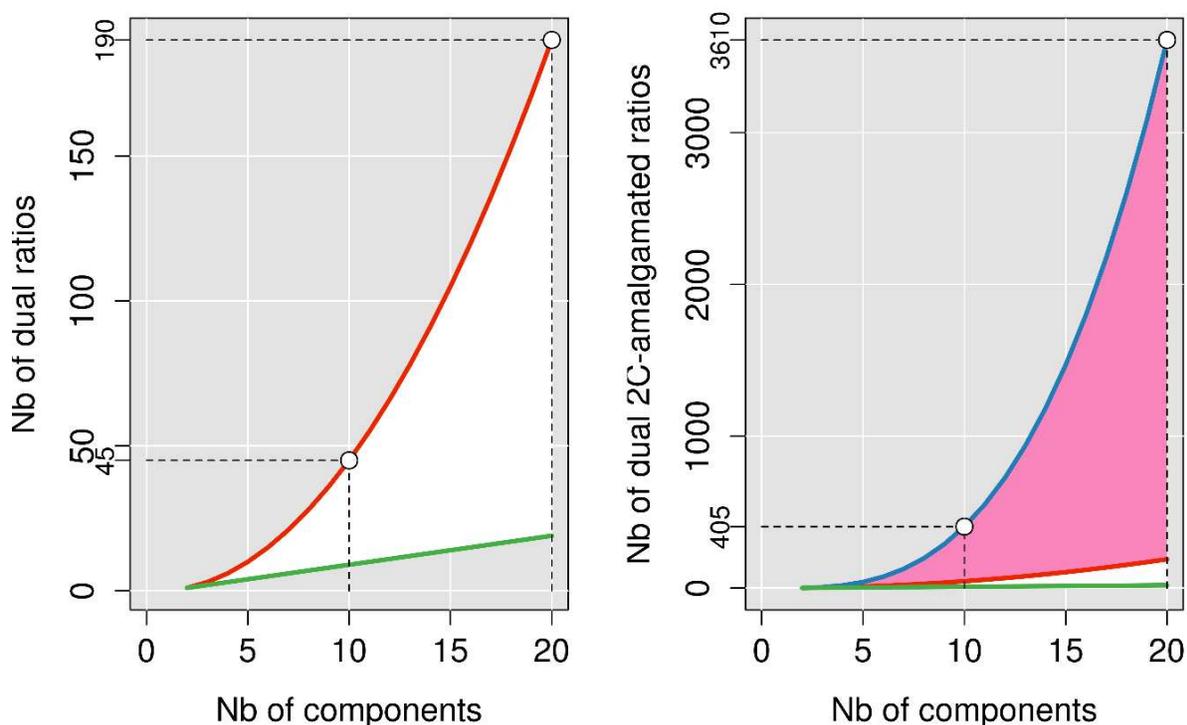


Figure 1. Number of (a) dual and (b) two-component amalgamated ratios illustrating the redundancy bubble.

To solve problems related to nutrient diagnosis in soil and plant sciences, one must first recognize that soil and plant analytical data are most often compositional, i.e. strictly positive data (concentrations, proportions) related to each other and bounded to some whole [26]. Compositional data have special numerical properties that may lead to wrong inferences if not transformed properly. Log-ratio transformations have been developed to avoid numerical biases [26, 29, 30, 31]. The balance concept presented in this chapter is based on log ratios or contrasts. Balances are computed rather simply from compositions using the isometric

log-ratio (*ilr*) transformation developed by [27]. In the literature, the nutrient balance often refers to a nutrient budget that measures the depletion or accumulation of a given nutrient in soils [28], implying exchange between compartments of some whole. In this chapter, nutrient balance is defined as dual or multiple log ratios between nutrients, implying balance between components of the same whole.

The aim of this chapter is to introduce the reader to the balance concept as applied to soil fertility studies. The first section of this chapter presents the theory common to the three subsequent subjects, which are cationic balance in soils, plant nutrient signatures and mineralization of organic residues. It is suggested that the reader gets familiar with the theory before browsing through the subject of interest.

2. Theory of CoDa

Because a change in any proportion of a whole reverberates on at least one other proportion, proportions of components of a closed sum (100%) are interdependent. Therefore, a compositional vector is intrinsically multivariate: its components cannot be analyzed and interpreted without relating them to each other [32,33]. Compositional data (CoDa) induce numerical biases, such as self-redundancy (one component is computable by difference between the constrained sum of the whole and the sum of other components), non-normal distribution (the Gaussian curve may range below 0 or beyond 100% which is conceptually meaningless) and scale dependency (correlations depend on measurement scale). Redundancy can be controlled by carefully removing the extra degree of freedom in the *D*-part composition. Scale dependency is controlled by ratioing components after setting the same scale (e.g. fresh mass, dry mass or organic mass basis) or unit of measurement (e.g. mg kg⁻¹, g dm⁻³, cmol_c kg⁻¹, etc.) across components. Compositional datasets constrained to a closed space between 0 and 100% are amenable to normality tests after projecting them into a real space using log-ratio transformations.

One of the log ratio transformations is the centered log ratio (*clr*) developed by [26]. The *clr* is a log ratio contrast between the concentration of any nutrient and the geometric mean across the compositional vector. [34] used the *clr* to convert DRIS into Compositional Nutrient Diagnosis (CND-*clr*), hence correcting inherent biases generated by DRIS. [35] and [36] modeled the time change of ion activities in soils and nutrient solutions using *clr*. However, because *clr* generates a singular matrix (the *clr* variates sum up to 0), one *clr* value should be removed (e.g. that of the filling value) in multivariate analysis. In addition, outliers may affect considerably log ratios [32]. The diagnostic power of CND-*clr* is decreased by large variations in nutrient levels (e.g. Cu, Zn, Mn contamination by fungicides) that affect the geometric means across concentrations. Nevertheless, the *clr* transformation is useful to conduct exploratory analyses on compositional data [37].

The additive log ratio or *alr* [26] computed as $\ln(x/x_D)$ is the ratio between any component *x* and a reference component *x_D*. [17] used nitrogen as reference component (N=100%) to produce a stoichiometric N:P:K:Ca:Mg rule for adjusting nutrient needs of tree seedlings. If a

tissue contains 2.50% N and 0.15% P, the Redfield N/P ratio [38] is 16.7 and the corresponding *alr* [P/N] value is $\ln(0.15/2.50) = -2.81$. Other stoichiometric rules have been proposed such the C:N:P:S rule for humus formation [4]. There are $D-1$ *alr* variables in a D -part composition because one component is sacrificed as denominator. The *alrs* are oblique to each other and are thus difficult to rectify and interpret [24]. Orthogonal balances are log ratio contrasts between geometric means of two groups of components that are multiplied by orthogonal coefficients to gain orthogonality [27]. Orthonormal balances are called ‘isometric log ratios’ coordinates or *ilr* [27] and are illustrated by a mobile and its fulcrums (CoDa dendrogram) [37]. Balances are encoded in a device called sequential binary partition that orderly allocates components to balance numerator and denominator or +/- sides of a contrast. The *ilr* of groups of components is a thus rectified ratio between their geometric means. Balances avoid matrix singularity and redundancy: there are $D-1$ independent balances in a D -part composition. The orthonormal balance concept was found to be the most appropriate technique in the multivariate [29] and multiple regression [39] analyses in geochemistry [40], plant nutrition [34, 35, 36, 41, 42], the P cycle [43], and soil quality [44, 45].

2.1. From CoDa to sound balances

The sample space of a compositional vector defined by S^D is a strictly positive vector of D nutrients adding up to some constant κ . The closure operation, \mathcal{C} , computes the constant sum assignment as follows :

$$S^D = \mathcal{C}(c_1, c_2, \dots, c_D) = \left[\frac{c_1 \kappa}{\sum_{i=1}^D c_i}, \frac{c_2 \kappa}{\sum_{i=1}^D c_i}, \dots, \frac{c_D \kappa}{\sum_{i=1}^D c_i} \right] \quad (1)$$

Where $\sum_{i=1}^D c_i$ closes the sum of components to some whole such as 1, 100%, 1000 g kg⁻¹, which allows computing a filling value to the unit of measurement. In other cases where the data do not add up to the measurement unit such as mg dm⁻³ or mg L⁻¹, the measurement unit just cancels out when components are ratioed.

In general, raw or log-transformed concentration data are analyzed statistically without any *a priori* arrangement of the data. The analyst not only processes such data through a numerically biased procedure, but also relies on a cognitively unstructured path that returns unstructured results that are barely interpretable (Figure 2).

Fortunately, recent progress in compositional data analysis provides means to elaborate structured pathways and interpret results coherently [27]. Indeed, the *ilr* technique transforms a D -part composition into $D-1$ pre-defined orthogonal balances of parts projected into a real Euclidean space [24]. Orthogonality is a special case of linear independence where vectors fall perfectly at right angle to each other [46]. The balances can thus be analyzed as additive (undistorted) variables in the Euclidean space, hence without bias. The log ratio of X/Y is also called a log contrast between X and Y because $\log(X/Y) = \log(X) - \log(Y)$. A log ratio can scan the real space ($\pm\infty$) because ratios may range from large numbers (positive log values) to small fractions (negative log values).

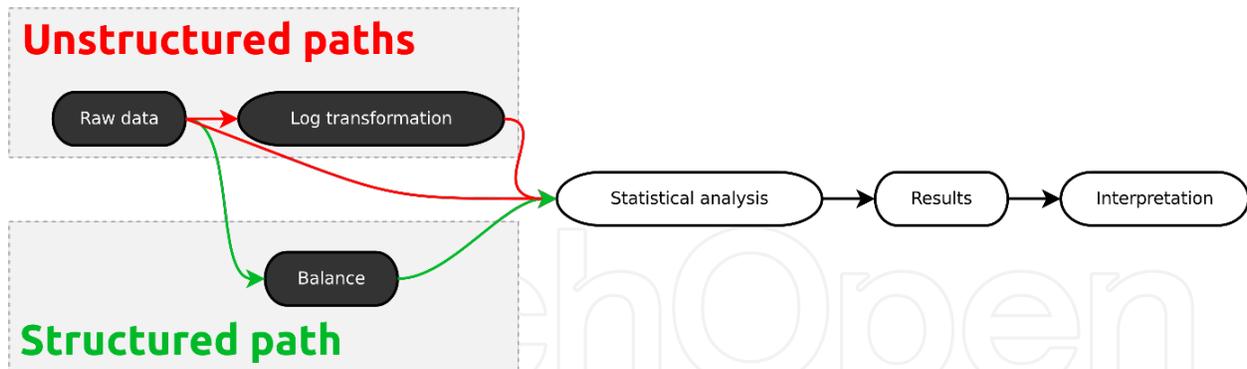


Figure 2. Unstructured or knowledge-based (structured) pathways of compositional data analysis lead to numerically biased and unbiased interpretations, respectively.

Balances can be illustrated by a CoDa dendrogram [37] where components or groups of components are balanced by analogy to a mobile and its fulcrums (Figure 3). Each part has its own weight and the balance between parts or groups of parts are the fulcrums (boxplots) equilibrating the system and computed as *ilr*. It can be shown that a relative increase in Ca concentration will change the [Ca | Mg] balance and [N,P,K | Ca, Mg] balances without affecting the ([N,P | K] and [N | P]). Transforming compositions to functional balances does not only create orthogonal real variables amenable to linear statistics; it also creates new variables whose interpretation is also of interest. Thus the interpretation of relationships between nutrients depends on how balances are conceived using the best science and management options. For example, another balance setup could be defined as [N,P | K, Ca, Mg], [N | P], [K | Ca, Mg] and [Ca | Mg].

A CoDa dendrogram (e.g. Figure 3) is interpreted as follows:

- Each fulcrum represents a balance. There are 4 balances for 5 components in Figure 3.
- If the fulcrum lies in the center of the horizontal bar, the balance is null. If it lies on the left side of the center, the mean balance is negative and left-side components occupy a larger proportion in the simplex. A fulcrum on the right side indicates a positive balance.
- Rectangles located on fulcrums are boxplots.
- The length of vertical bars represent the proportion of total variance

Nested balances are encoded in an *ad hoc* sequential binary partition (SBP) that nurtures the ties between groups of components. A SBP is a $(D-1) \times D$ matrix, where parts labelled “+1” (group numerator) are balanced with parts labelled “-1” (group denominator) in each ordered row. A part labelled “0” is excluded. The composition is partitioned sequentially at every ordered row into 2 contrasts until (+1) and (-1) subcompositions each contain a single part. The analyst can use exploratory analysis [37] or refer to current theory and expert knowledge to design the balance scheme. The CoDa dendrogram in Figure 3 is formalized by the SBP in Table 1.

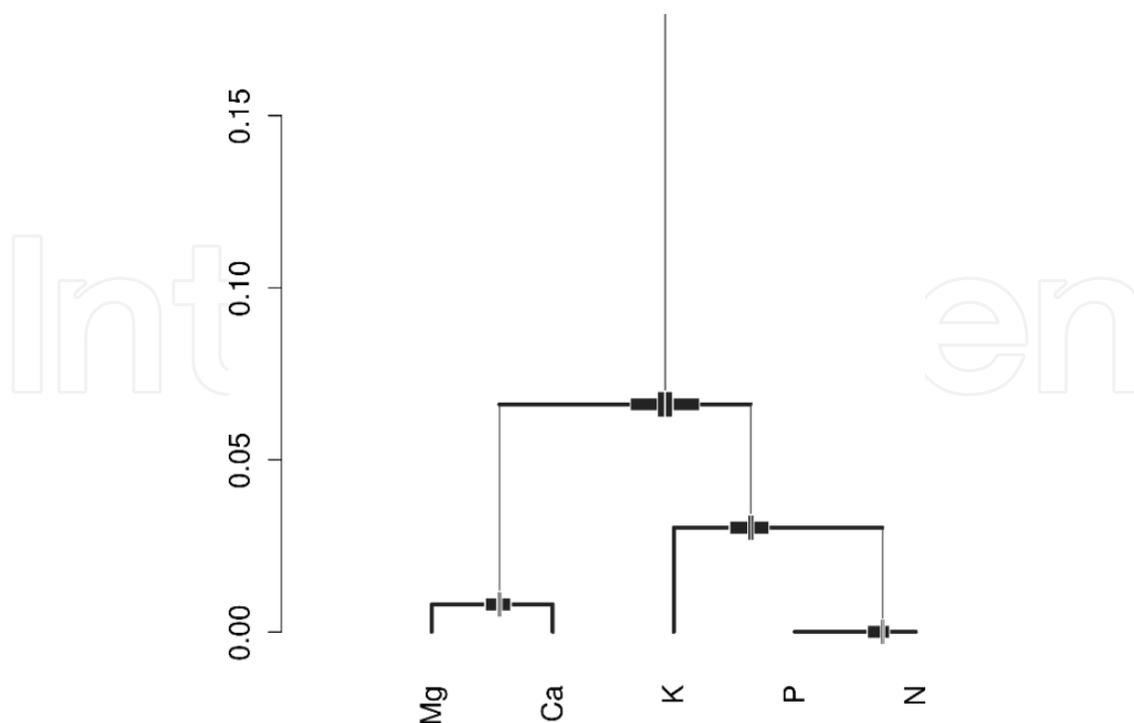


Figure 3. Balances between N, P, K, Ca, and Mg (five weight variables) are illustrated by a mobile and its fulcrums (four balance variables) where N, P, and K are contrasted with Ca and Mg, N and P with K, N with P, and Ca with Mg.

Binary partiton	Balance between groups of components					r	s	ilr computation
	N	P	K	Ca	Mg			
[N,P,K Ca,Mg]	+1	+1	+1	-1	-1	3	2	$\sqrt{\frac{3 \times 2}{3 + 2}} \ln \left(\frac{(N \times P \times K)^{1/3}}{(Ca \times Mg)^{1/2}} \right)$
[N,P K]	+1	+1	-1	0	0	2	1	$\sqrt{\frac{2 \times 1}{2 + 1}} \ln \left(\frac{(N \times P)^{1/2}}{K} \right)$
[N P]	+1	-1	0	0	0	1	1	$\sqrt{\frac{1 \times 1}{1 + 1}} \ln \left(\frac{N}{P} \right)$
[Ca Mg]	0	0	0	+1	-1	1	1	$\sqrt{\frac{1 \times 1}{1 + 1}} \ln \left(\frac{Ca}{Mg} \right)$

Table 1. Sequential binary partition defining macronutrient balances.

In Table 1, the sequential binary partition of nutrients encodes the balances between two geometric means across the + components at numerator and the – components at denominator. The orthogonal coefficient of a log contrast is computed from the number of + and – components in each binary partition. The balances between two subcompositions are orthogonal log ratio contrasts between geometric means of the “+1” and “-1” groups. The j^{th} *ilr* coordinate is computed as follows [24]:

$$ilr_j = \sqrt{\frac{rs}{r+s}} \ln \frac{g(c_+)}{g(c_-)}, \text{ with } j = [1, 2, \dots, D-1] \tag{2}$$

Where ilr_j is the j^{th} isometric log-ratio; $g(c_+)$ is geometric mean of components in group "+1", c_+ ; and $g(c_-)$ is the geometric mean of components in group "-1", c_- . Because dual ratios are nested into $g(c_+)$ and $g(c_-)$, the balances avoid generating redundant ratios. The orthogonal coefficient is computed as $\sqrt{rs/(r+s)}$ [27]. For example, a Redfield N/P ratio of 16.7 is converted to ilr as $\sqrt{1 \times 1 / (1 + 1)} \ln(16.7) = 2.02$. The ilr technique is thus not only mathematically elegant, but is also conceptually meaningful.

2.2. Dissimilarity between compositions

As a result of orthogonality, the Aitchison distance (\mathcal{A}) between any two compositions is computed as a Euclidean distance across the selected ilr coordinates as follows [47]:

$$\mathcal{A} = \sqrt{\sum_{j=1}^{D-1} (ilr_j - ilr_j^*)^2} \quad (3)$$

Where ilr_j is the j^{th} ilr of a given composition and ilr_j^* is the corresponding ilr for the reference composition. Selecting alternative SBPs to test and interpret other balances in the system under study just rotates the orthogonal axes of the ilr coordinates without affecting \mathcal{A} . The Aitchison distances computed across ilr or clr values are identical [24]. [34] rectified DRIS to fit into clr . As computed from dual ratios and nutrient indices [13] and using the same reference population as reference for computing the Aitchison distance, the DRIS nutrient imbalance index appeared to be slightly distorted and noisy (Figure 4). Tissue analyses in Figure 4 were obtained from a survey across guava (*Psidium guajava*) orchards in the state of São Paulo, Brazil. Noise and distortion between results observed in Figure 4 is attributable to numerical biases in DRIS results.

On the other hand, the Euclidean distance (\mathcal{E}) based on log transformations is biased by the difference between the geometric means times the number of parts as follows [48]:

$$\mathcal{E}^2(\ln(x), \ln(y)) = \mathcal{A}^2 + D \left(\ln \frac{g(x)}{g(y)} \right)^2 \geq \mathcal{A}^2 \quad (4)$$

In plant nutrition studies [49], the Mahalanobis distance (\mathcal{M}) may be preferred to the Euclidean distance because the former takes into account the covariance structure of the data [29] (as illustrated in Figure 5) and has χ^2 distribution [50,51]. The \mathcal{M}^2 is computed as follows:

$$\mathcal{M}^2 = (x - \bar{x})^T \times \text{COV}^{-1}(x - \bar{x}) \quad (5)$$

Where \bar{x} is the mean and COV is the covariance matrix. Both \mathcal{A} and \mathcal{M} computed across log-transformed data are higher than their counterparts computed across balances, indicating systematic upper bias using natural log compared to ilr transformations (Figure 6). Tissue analyses in Figure 6 were obtained from the same guava orchard survey as above.

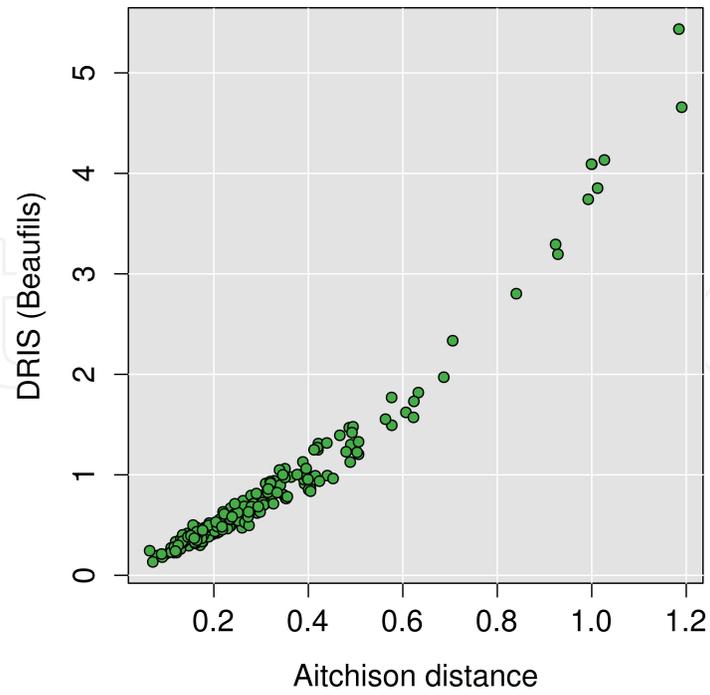


Figure 4. Distance from a reference composition computed using DRIS versus the Aitchison distance.

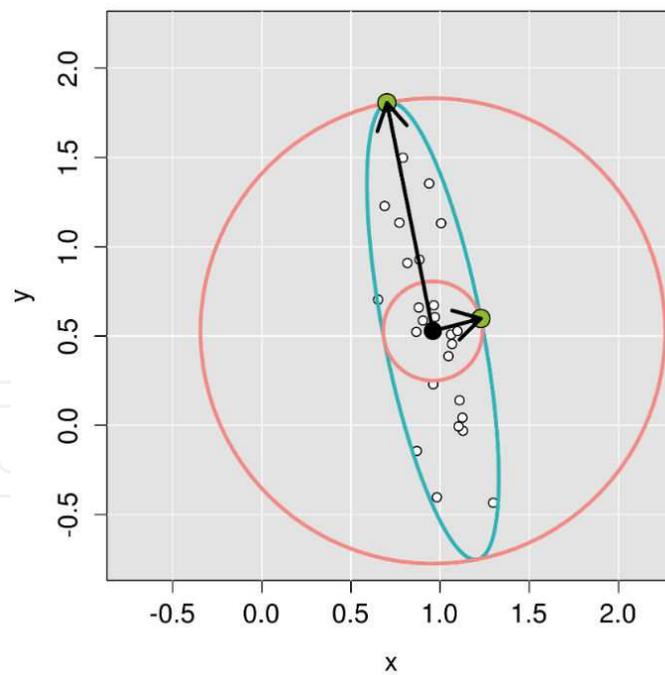


Figure 5. The Euclidean distance is circular while the Mahalanobis distance (M) is elliptical. The blue ellipse represents a line of equidistant points in terms of M that scales data to the variance in each direction. The M between green points and the center are equal. However, the Euclidean distance between each green point and the center is different, as shown by the Euclidean equidistance pink circles.

2.3. Cate-Nelson analysis

The Cate-Nelson procedure was developed as a graphical technique to partition percentage yield (yield in control divided by maximum yield with added nutrient) versus soil test [52]. The scatter diagram is subdivided into four quadrants to determine a critical test level and a critical percentage yield by maximizing the number of points in the + quadrants. This technique is analog to binary classification tests widely used in medical sciences [53] where data each quadrant are interpreted as true positive (correctly diagnosed as sick), false positive (incorrectly diagnosed as sick), true negative (correctly diagnosed as healthy) and false negative (incorrectly diagnosed as healthy). Applied to soil fertility studies, we can define four classes as follows:

- True positive (TP: nutrient imbalance): imbalanced crop (low yield) correctly diagnosed as imbalanced (above critical index).
- False positive (FP: type I error): balanced crop (high yield) incorrectly identified as imbalanced (above critical index). FP points indicate luxury consumption of nutrients.
- True negative (TN: nutrient balance): balanced crop (high yield) correctly diagnosed as balanced (below critical index).
- False negative (FN: type II error): imbalanced crop (low yield) incorrectly identified as balanced (below critical index). FN points show impacts of other limiting factors.

The performance of the test is measured by four indices:

- Sensitivity: probability that a low yield is imbalanced as $TP/(TP+FN)$
- Specificity: probability that a high yield is balanced as $TN/(TN+FP)$
- Positive predictive value (PPV): probability that an imbalance diagnosis returns low yield as $TP/(TP+FP)$
- Negative predictive value (NPV): probability that a balance diagnosis returns high yield as $TN/(TN+FN)$

The performance of the binary classification test is higher when the four indexes get closer to unity. However, the maximization of the four indexes may not be the most appropriate procedure. Indeed, agronomists are more interested in high PPV than in high specificity.

Using the Cate-Nelson graphical procedure, the TN specimens are selected as reference population after removing outliers. If the number of points is too large, yields are arranged in an ascending order and a two-group partition is computed. The sums of squares between two consecutive groups of observations are iterated as follows:

$$\text{Class sum of squares} = \frac{\left(\sum_{j=1}^{n_1} Y_{1j}\right)^2}{n_1} + \frac{\left(\sum_{j=1+n_1}^n Y_{2j}\right)^2}{n_2} - \frac{\left(\sum_{j=1}^n Y_j\right)^2}{n} \quad (6)$$

Where Y_{1j} is class 1 yields starting with the two lowest soil indices; the remaining yields are in class 2 or Y_{2j} ; and n_1 , n_2 and n are the numbers of observation in class 1, class 2 and across classes, respectively. The last member of the equation is the correction factor. The starting values for maximization of the sums of squares across \mathcal{A} or \mathcal{M} could be the *ilr* means of the upper 20 top specimens [54]. Due to yield variations between production years, the upper quartile of higher yield standardized by year of production is an additional option. Because the iterative procedure is very sensitive to extreme values, an *a posteriori* visual adjustment may be necessary to maximize the number of points in opposite quadrants.

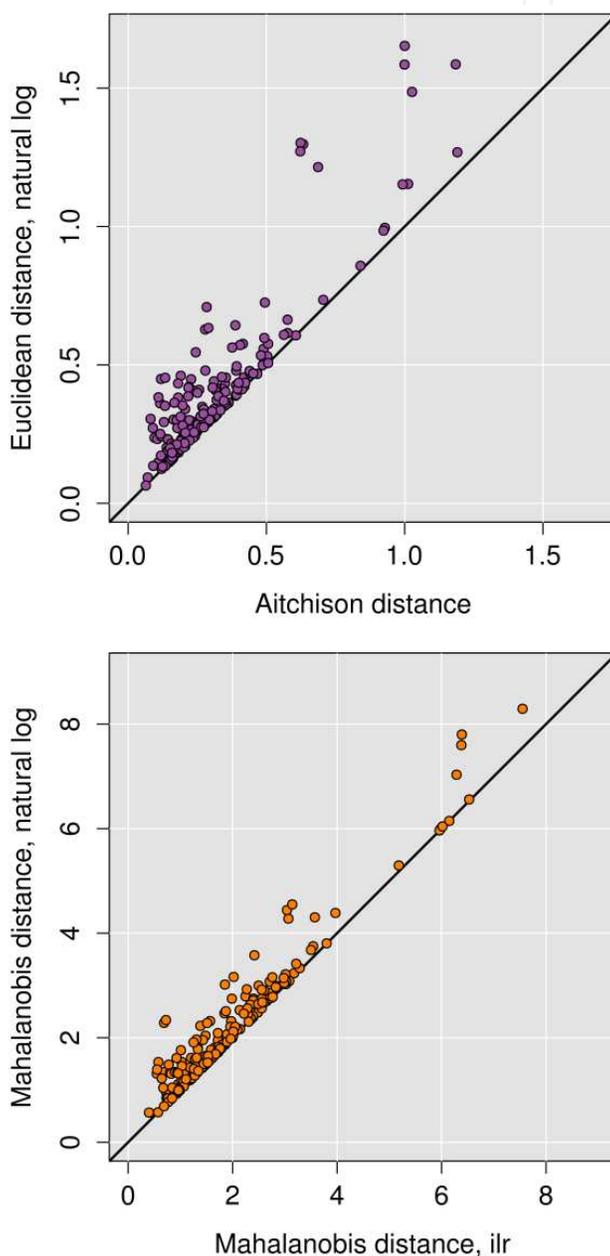


Figure 6. Relationships between the Euclidean, Aitchison and Mahalanobis distances.

2.4. Statistics

In this chapter, statistics computed across compositional data were performed in the R statistical environment [55]. Compositional data analysis was conducted using the R “compositions” package [56]. Data distribution was tested using the Anderson-Darling normality test [57] in the “nortest” package [58]. Multivariate outliers were removed using \mathcal{M} computed in the R “mvoutlier” package [59]. Linear discriminant analysis (LDA) was used as a statistical ordination technique that allows computing linear combinations of variables that best discriminate groups. Multiple regression analysis was conducted using *ilr* [39] and compared to raw data. After completing the statistical analysis, the balances could be back-transformed to the familiar concentration units using the $D-1$ *ilr* values and the sum constraint.

3. Cationic balances in tropical soils

3.1. Sequential binary partition

The percentage base saturation is the proportion of soil cation exchange capacity (CEC) occupied by a given cation. The soil compositional vector is defined as follows [12]:

$$S^4 = \mathcal{C}(K, Ca, Mg, H + Al) \quad (7)$$

As illustrated in Figure 7, the first contrast, $[K \mid Ca, Mg, H+Al]$, balances the K against divalent cations and acidity to enable adjusting the K fertilization to soil basic acid-base conditions as modified by liming.

The second contrast $[Ca, Mg \mid H+Al]$ is the acid-base contrast for determining lime requirements while the $[Ca \mid Mg]$ balance reflects the Ca:Mg ratio in soils adjustable by the liming materials. Alternative SBPs could also be elaborated such as $[K, Ca, Mg \mid (H+Al)]$, $[K \mid Ca, Mg]$ and $[Ca \mid Mg]$ balances that reflects the BCSR model of [12]. The selected sequential binary partition for cationic balances is presented in Table 2.

For example, if a soil contains $2.9 \text{ mmol}_c \text{ K dm}^{-3}$, $20 \text{ mmol}_c \text{ Ca dm}^{-3}$, $5 \text{ mmol}_c \text{ Mg dm}^{-3}$, and $23 \text{ mmol}_c \text{ H+Al dm}^{-3}$. Cationic balances are computed as follows:

$$(1) [K \mid Ca, Mg, H + Al] = \sqrt{\frac{1 \times 3}{1 + 3}} \ln\left(\frac{2.9}{\sqrt[3]{20 \times 5, \quad x23}}\right) = -1.312;$$

$$(2) [K \mid Ca, Mg] = \sqrt{\frac{1 \times 2}{1 + 2}} \ln\left(\frac{2.9}{\sqrt{20 \times 5}}\right) = -1.011; \text{ and}$$

$$(3) [Ca \mid Mg] = \sqrt{\frac{1 \times 1}{1 + 1}} \ln\left(\frac{20}{5}\right) = 0.980.$$

Note that the K fertilization would depend on soil acidity as well as levels of exchangeable Ca and Mg in the soil. We thus expect the K index and the K balance to be similarly related to fruit yield if the *ceteris paribus* assumption applies to exchangeable Ca, Mg, and acidity in this soil-plant system.

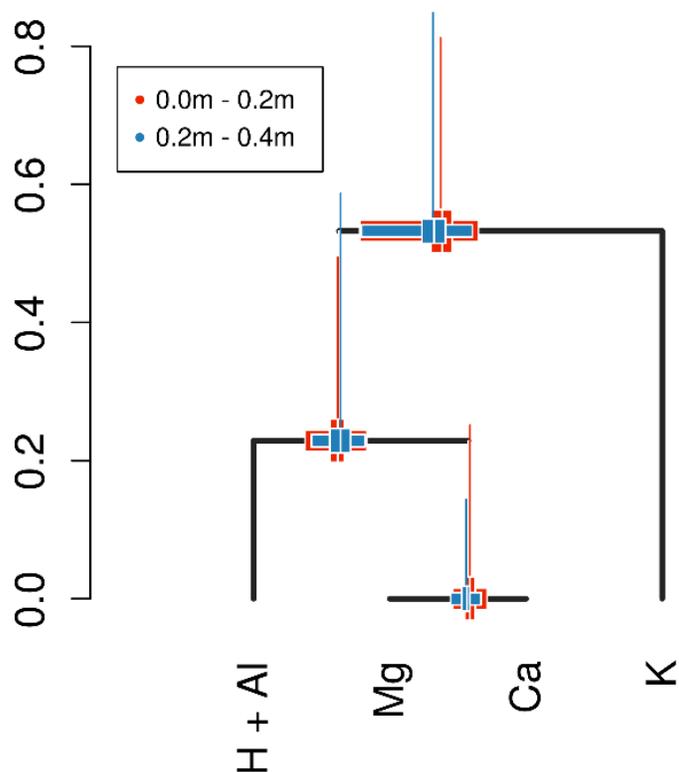


Figure 7. The cationic balances in soils can be designed to facilitate K fertilization and lime management. Each balance between two groups of ions is computed as isometric log ratio.

Partition	Cationic balances				r	s	ilr formulation
	K	Ca	Mg	H+Al			
1	1	-1	-1	-1	3	1	$\sqrt{\frac{1 \times 3}{1 + 3}} \ln\left(\frac{K}{\sqrt[3]{Ca \times Mg \times x(H + Al)}}\right)$
2	1	-1	-1	0	1	2	$\sqrt{\frac{1 \times 2}{1 + 2}} \ln\left(\frac{K}{\sqrt{Ca \times Mg}}\right)$
3	0	1	-1	0	1	1	$\sqrt{\frac{1 \times 1}{1 + 1}} \ln\left(\frac{Ca}{Mg}\right)$

Table 2. Sequential binary partition of soil cationic data

3.2. Datasets

Changes in soil cationic balances were monitored in N and K fertilizer trials established on an epieutrophic and endodystrophic soil (Red-Yellow Oxisol) [60] at São Carlos (São Paulo, Brazil). One year old plants of ‘Paluma’ guava (*Psidium guajava*) were planted. The

experiment lasted 3 yr. The N treatments in the 1st year were 0, 30, 60, 120, 180, 240 and 300 g N tree⁻¹ supplemented with 52 g P tree⁻¹ and 52 g K tree⁻¹. The initial N rates were doubled and tripled in the 2nd and 3rd years, respectively. The initial P and K doses were doubled the 2nd year. The 3rd year, rates were 240 g P₂O₅ tree⁻¹ and 360 g K₂O tree⁻¹. Fertilizers were ammonium nitrate (34% N), simple superphosphate (8.7% P) and potassium chloride (50% K). In the K trial, K was added as KCl at rates of 0, 25, 50, 100, 150 and 200 and 250 g K tree⁻¹ the 1st year and supplemented with 120 g N tree⁻¹ as ammonium sulfate (20% N) and 52 g P tree⁻¹ as triple superphosphate (19% P). The N, P, and K rates were doubled in the 2nd year. The K rates were tripled the 3rd year and supplemented with 360 g N tree⁻¹ and 105 g P tree⁻¹. The acidifying ammonium fertilizers may increase exchangeable acidity in both trials. The fertilizers were broadcast around the tree 40 cm from crown projection. Each plot comprised four trees each covering an area of 7 m × 5 m, for a total of 286 trees ha⁻¹. The experimental setup was a randomized block design with four replications. Fresh fruit yields were measured 1-3 times wk⁻¹ from January to June, starting approximately 97 d after fruit set.

Soils were sampled annually after harvest at four locations per tree in the 0-20 cm and 20-40 cm layers where most of the root system is located, then composited per plot. Soil samples were air dried and analyzed for K, Ca, Mg and (H + Al) [61]. The K, Ca and Mg were extracted using exchange resins, quantified by atomic absorption spectrophotometry and reported as mmol_c dm⁻³. Exchangeable acidity (H+Al) was quantified by the SMP pH buffer method [62] and computed using the equation of [63] to convert buffer pH into mmol_c (H+Al) dm⁻³ as follows:

$$(H + Al) = 10 \exp(7.76 + 1.053 pH_{SMP}), R^2 = 0.98 \quad (8)$$

Cation exchange capacity (CEC) was computed as the sum of cationic species. Assuming a soil bulk density of 1 kg dm⁻³, CEC averaged 5.4 cmol_c kg⁻¹.

3.3. Results

3.3.1. Influence of the K fertilization on cationic balances in soil

As shown by scatter and ternary diagrams (Figure 8), the large ellipses, that represent the distribution of cationic balances in the 0-20 and 20-40 cm layers, overlapped. However, the small ellipses (Figure 8) representing the confidence region about means differed significantly. The [K | Ca, Mg, H+Al] balance was higher in the 0-20 cm layer, indicating that more K accumulated in the surface layer as a result of surface K fertilizer applications.

Soil test K and cationic balances were averaged between the beginning and the end of the growing season to represent average soil conditions. The soil indices were related to fresh fruit yield (Figures 9a and 9b). In Figure 9, data are means of 4 replicates and bars are least significant differences.

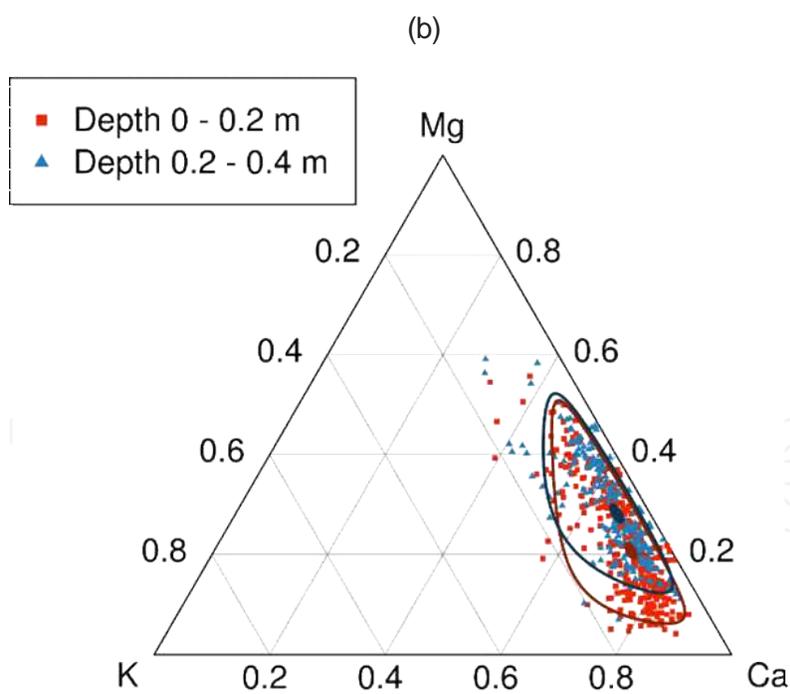
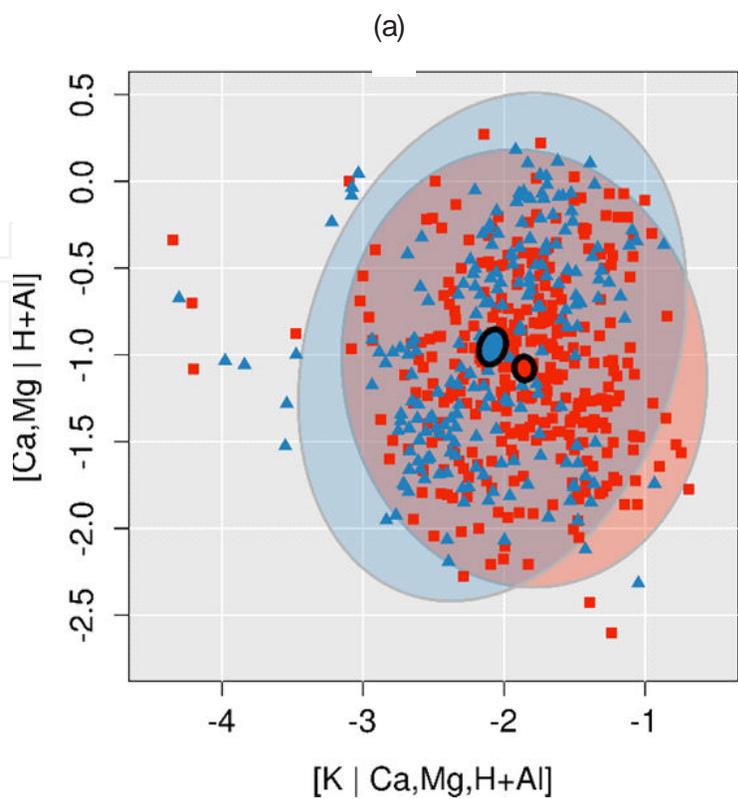


Figure 8. The K accumulated relatively more on the 0-20 cm layer following three years of K fertilization as shown by (a) two orthonormal cationic balances and (b) a K-Ca-Mg ternary diagram.

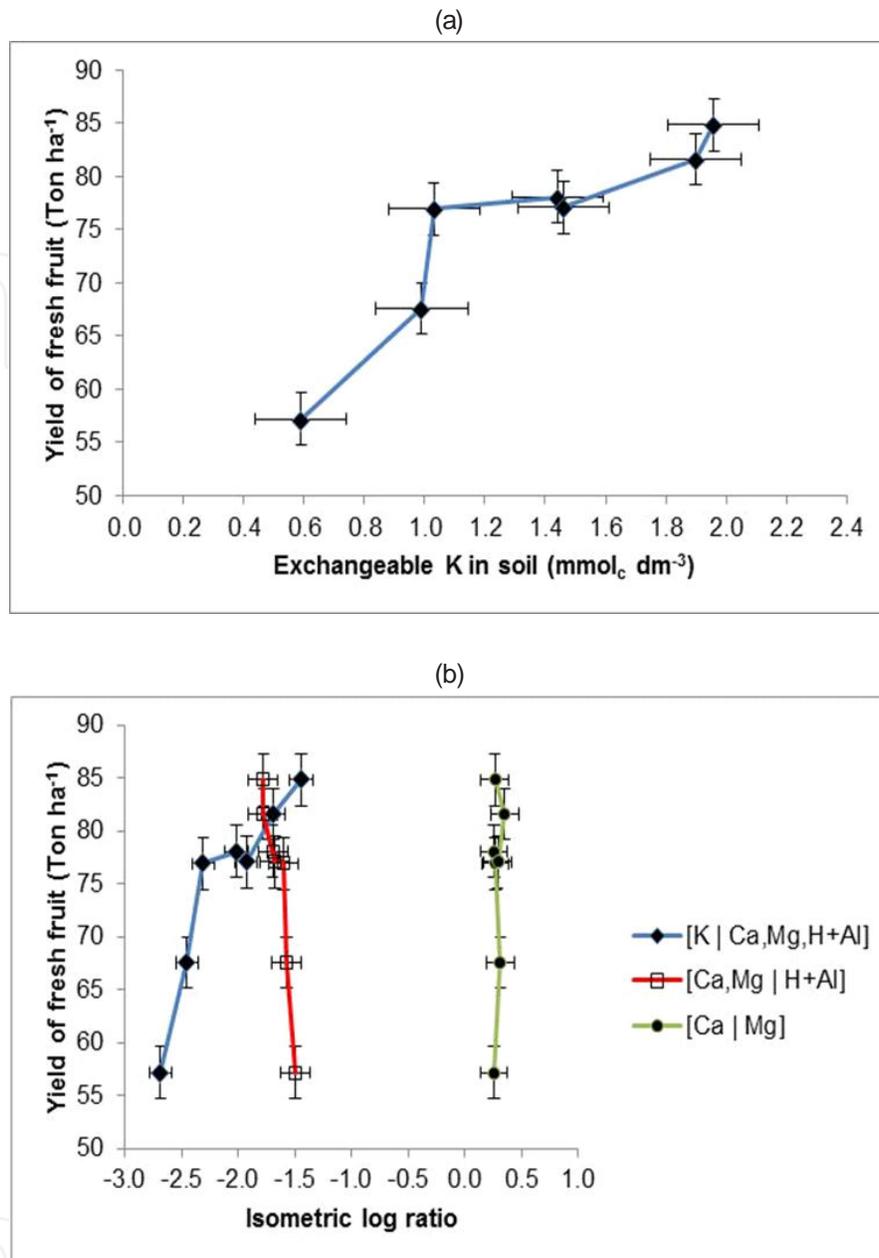


Figure 9. Exchangeable K (a), the [K | Ca, Mg, H+Al] balance (b) and fruit yield increased with added K. Other cationic balances did not change markedly.

3.3.2. Critical soil K concentration and balance in the N and K trials

The Cate-Nelson partitioning of the relationship between guava fresh fruit yield and either soil K level or the [K | Ca, Mg, (H+Al)] balance across the combined N and K fertilizer experiments indicates that the K level index classified two specimens as TN compared to four for the K balance index (Figure 10). The graphical representation of this soil-plant relationship indicates diagnostic advantage to using the K nutrient balance in rather than the K concentration.

The sensitivity, specificity, PPV and NPV criteria are presented in Table 3. We expect performance criteria to be at least 80%. Low specificity indicates that some interactions with K leading to high yield, possibly involving Ca and Mg, have been ignored. Apparently, the *ceteris paribus* assumption did not apply to this study. The fact that the balance allows to adjust the K to other cationic species may account for failure to meet the *ceteris paribus* assumption.

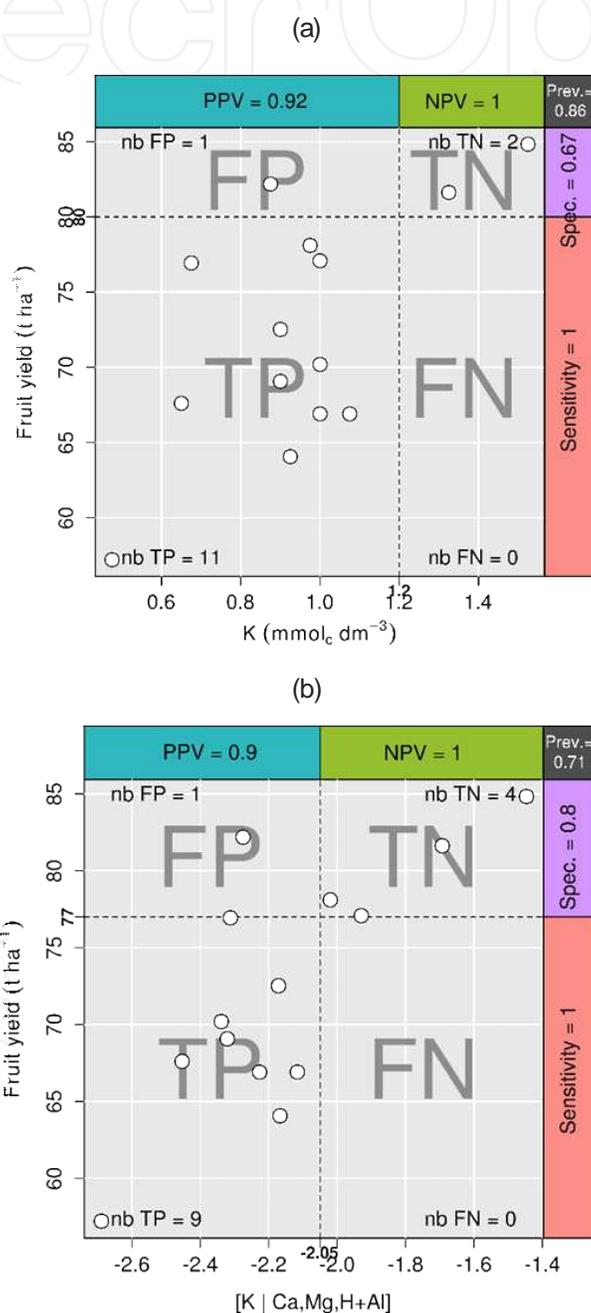


Figure 10. Cate-Nelson partitioning of the relationship between guava fresh fruit yield. critical values were (a) 1.2 $\text{mmol}_c \text{K dm}^{-3}$. TN = 2; FN = 0; TP = 11; FP = 1 and (b) -2.07. TN = 4; FN = 0; TP = 9; FP = 1.

Soil K index	Sensitivity = TP/(TP+FN)	Specificity = TN/(TN+FP)	Positive predictive value = PPV=TP/(TP+FP)	Negative predictive value = NPV=TN/(TN+FN)
	%			
K level	100.0	66.7	91.7	100.0
K balance	100.0	80.0	90.0	100.0

Table 3. Performance of K indices in terms of sensitivity, specificity, PPV and NPV

4. Multi-element Balances in plant nutrition

4.1. Sequential binary partition

Plant nutrients are classified as essential macronutrients measured in % (N, S, P, Mg, Ca, K, Cl), essential micronutrients measured in mg kg⁻¹ (Mn, Cu, Zn, Mo, B) and beneficial nutrients generally measured in mg or µg kg⁻¹ but occasionally in % (Si, Na, Co, Ni, Se, Al, I, V) [64, 65, 15]. The plant ionome is defined as elemental tissue composition as related to the genome [66]. A subcomposition of plant ionome could be defined by the following simplex for conducting statistical analysis:

$$S^D = \mathcal{C}(C, N, P, K, Ca, Mg, B, S, Cl, Cu, Zn, Mn, Fe, Mo, F_v) \quad (9)$$

Where F_v is the filling value between 1000 g kg⁻¹ and the sum of analytical data and $D = 15$, the total number of components including F_v . An SBP scheme can be elaborated based on well documented roles and stoichiometric rules provided by [17, 14, 12], who reported a large number of dual and multiple nutrient interactions in plants such as:

- Macronutrients have a stoichiometric relationship with carbon uptake;
- N with S, P, K, Ca, Mg, Fe, Mn, Zn, and Cu;
- NH₄ with K, Ca, and Mg;
- S with N, P, Fe, Mn, Mo;
- P with N, K, Ca, Mg, B, Mo, Cu, Fe, Mn, Al, and Zn;
- Cl with N and S;
- K with N, P, Ca, Mg, Na, B, Mn, Mo, and Zn;
- Ca with N, K, Mg, Na, Cu, Fe, Mn, Ni, and Zn;

- Mg with N, P, B, Fe, Mn, Mo, Na, and Si;
- B with N, P, K, and Ca;
- Cu with N, P, K, Ca, Fe, Mn, and Zn;
- Fe with N, P, Ca, Mg, Cu, Mn, Co, and Zn;
- Zn with N, P, K, Ca, Mg, S, Na, Zn, Fe, and Mn;
- Mn with N, P, K, Ca, Mg, B, Mo, Ni, and Zn;
- Mo with N, P, K, S, Fe, and Mn.

4.2. Datasets

The tissue composition can be altered by environmental and seasonal factors. A dataset of 1909 potato (*Solanum tuberosum* L. cv. 'Superior') yields and ionomes was collected at five developmental stages between 1987 and 2002 in Quebec, Canada. The first mature leaf from top was sampled at 20-cm height (n = 502), bud stage (n = 544), beginning of flowering (n = 587), full bloom (n = 213) and fast tuber growth (n = 63) and analyzed for N, P, K, Ca, and Mg. The plant nutrient signatures at each developmental stage were compared using box-plots and discriminant analysis.

A critical hyper-ellipsoid can be viewed as a particular zone of the nutrient balance space where the probability to obtain high yield is high enough to satisfy the practitioner. The points lying inside the hyper-ellipsoid would be qualified as "balanced", and those lying outside the multi-dimensional construct, as "imbalanced". The practitioner might delineate intermediate zones if needed. Fertilizer trials were conducted to monitor balance change toward optimum nutrient conditions defined by the critical ellipses. In a P trial, P treatments applied to a P deficient soil were 0, 33, 66, 98 and 131 kg P ha⁻¹. In a K trial, K treatments of 0, 50, 100 and 150 kg K ha⁻¹ were applied to a K deficient soil. The diagnostic leaf of potato was sampled at the beginning of flowering [67].

4.3. Seasonal change in nutrient compositions

The boxplots and the CoDa dendrogram illustrate the center and dispersion of nutrient balances per development stage (Figure 11). The [N, P, K | Ca, Mg] balance tended to decrease markedly during the season while the [N | P] and [Ca | Mg] balances tended to increase, and the [N,P | K] balance tended to decrease. The fast decrease in [N,P,K | Ca, Mg] balance is attributable to more N, P and K than Ca and Mg being transferred toward growing leaves during exponential growth and toward tubers during maturation. The K was more affected than N and P.

The discriminant scores (dots) and eigenvectors, as well as confidence regions at 95% level delineated the distributions of populations (large grey ellipses) and means (small color colored ellipses) across stages of plant development (Figure 12). The first axis, dominated by the Redfield [N | P] balance followed by the [N, P, K | Ca, Mg] balance, captured 92% of total inertia. It is noteworthy that the nutrient balance changed orderly from one devel-

opmental to the other. The *ilrs* can thus be described by trend equations and sample composition be detrended toward a specific developmental stage for diagnostic purposes. The seasonally increasing N/P ratio may indicate possible N or P imbalance at some point in time assuming a stationary N:P stoichiometric rule. However, the N/P ratio was found to vary widely between plant species during plant development, depending on relative growth rates [38]. The Redfield N/P ratio in eukaryotic microbes is a balance between two fundamental processes, protein and rRNA synthesis, resulting in a stable biochemical attractor toward a given protein: to RNA ratio [68]. The N/P ratio of plant biomass is used as indicator of N or P limitation but critical N/P ratios change with age and function of tissues [38]. Immature leaves of young plants assimilate and grow simultaneously and their demand for N and P follows the stoichiometric rules of basic biochemical processes such as photosynthesis, respiration, protein synthesis, DNA duplication and transcription; growth becomes restricted to active meristems such as young leaves, shoot tips and inflorescences when plants get older [38]. Mature leaves are still photosynthetically active but no longer grow, which greatly reduces the P requirements for RNA and increases the N/P ratio. Nucleic-acid P can be mobilized from older leaves and transferred to younger leaves, leading to higher N/P ratios in older leaves [69], such as the first mature leaves of potatoes used as diagnostic tissue [67].

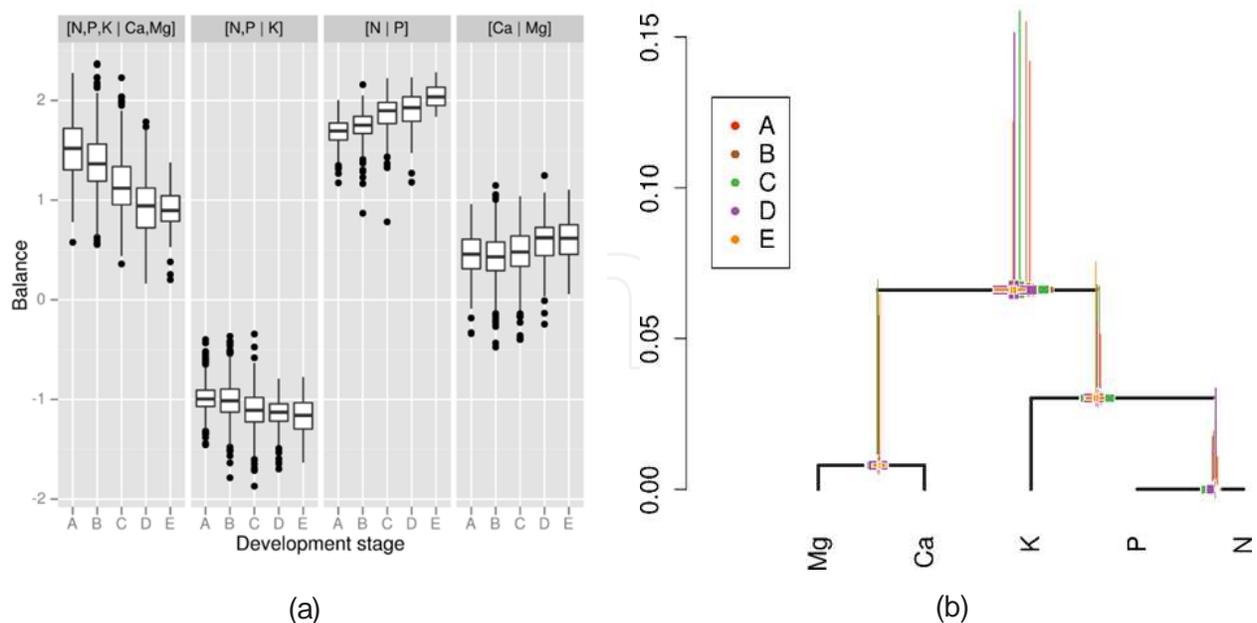


Figure 11. The (a) boxplots and (b) coda dendrogram of the four balances for five development stages.

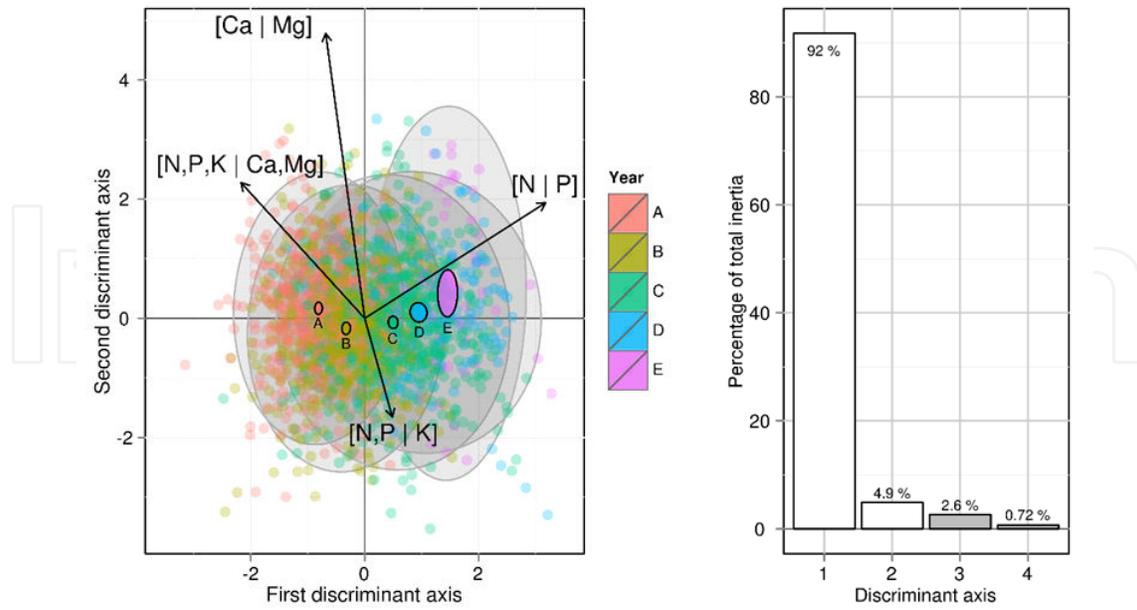


Figure 12. Discriminant analysis shows that nutrient balance changes orderly between the early and late developmental stages.

4.4. Defining reference balances for diagnostic purposes

The confidence region of optimum nutrition was defined by a 4-dimensional hyper-ellipsoid (Figure 13).

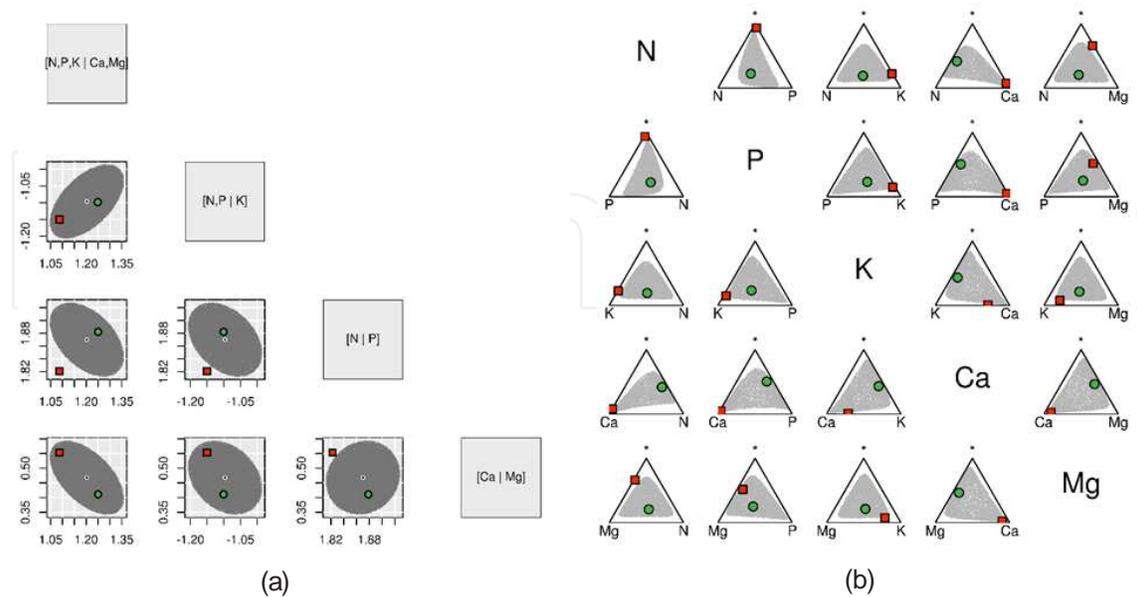


Figure 13. The ellipses define the optimum conditions of potato nutrition as found in surveyed high-yield potato stands (95% confidence regions about mean) illustrated as (a) balances and (b) scaled and centered concentrations.

The green and red points in Figure 13 represent specimens showing balanced and imbalanced nutrition, respectively. The fertilization of the potato should move nutrient signature toward the hyper-ellipsoid center. Added P perturbed the internal nutrient balance of cv. 'Superior' growing on a P deficient soil (Figure 14). The P trial showed that an addition of 98 kg P ha⁻¹ allowed the balance to penetrate into the critical ellipse.

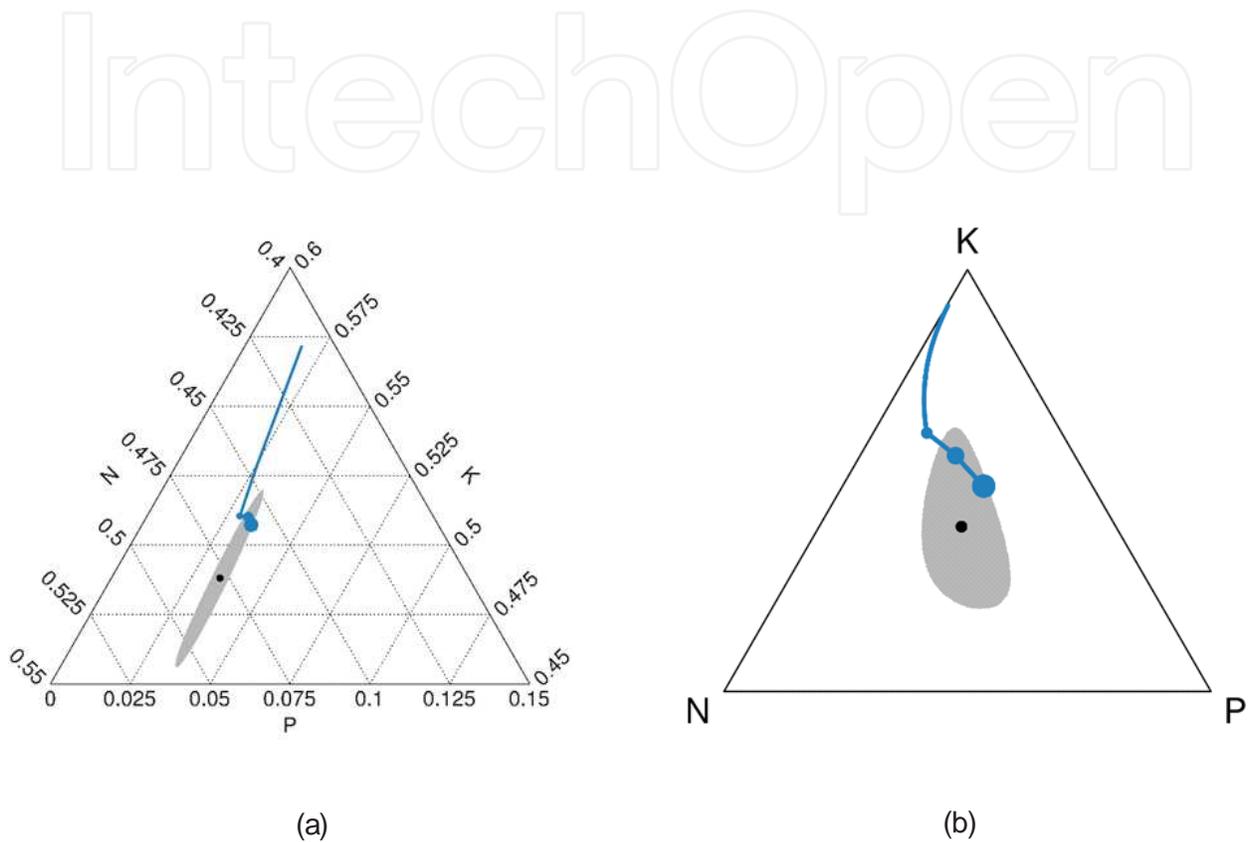


Figure 14. The P treatments applied to a P-deficient soil at rates of 0, 33, 66, 98 and 131 kg P ha⁻¹ (rates increasing as red circles enlarged) perturbed the potato ionome that moved toward the critical ellipse. The ternary diagram scaled on proportions on the left (A) was centered and rescaled on the right (B) for better appreciation of the shape of the ellipse and the trend of the fertilization effect.

In Figure 15, it can be observed that added K also perturbed the nutrient balance: the potato ionome moved toward the critical ellipse. The 2nd K rate moved the K deficient plant ionome closer to the critical ellipse, but Ca shortage maintained the crop outside the critical ellipse. From the second application rate up, the perturbation was small. In this case, the Ca was likely to be the most limiting nutrient as shown on the ternary diagram.

The perturbation on 5 nutrients can be illustrated by a matrix of ternary diagrams (Figure 16). These diagrams show 2 nutrients and an asterisk (*) representing the sum of the 3 other components. The central dot is the mean of high yielders surrounded by its 95% confidence region represented by a black line.

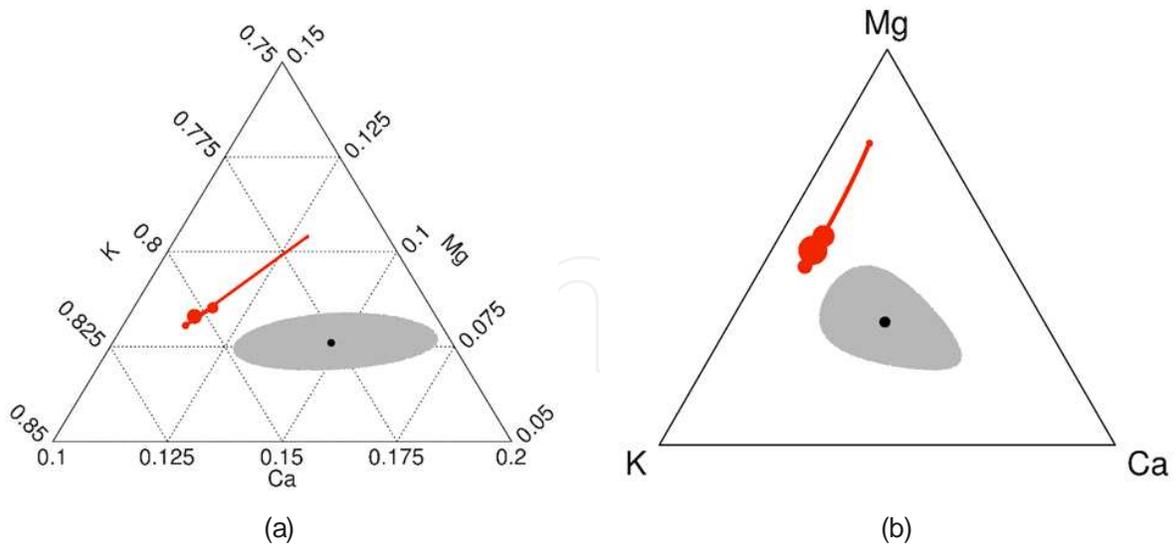


Figure 15. Ternary diagrams showing K treatments applied to a K-deficient soil at rates of 0, 50, 100 and 150 kg K ha⁻¹ (rates increasing as red circles enlarged (a) zoomed on proportions and (b) centered and scaled).

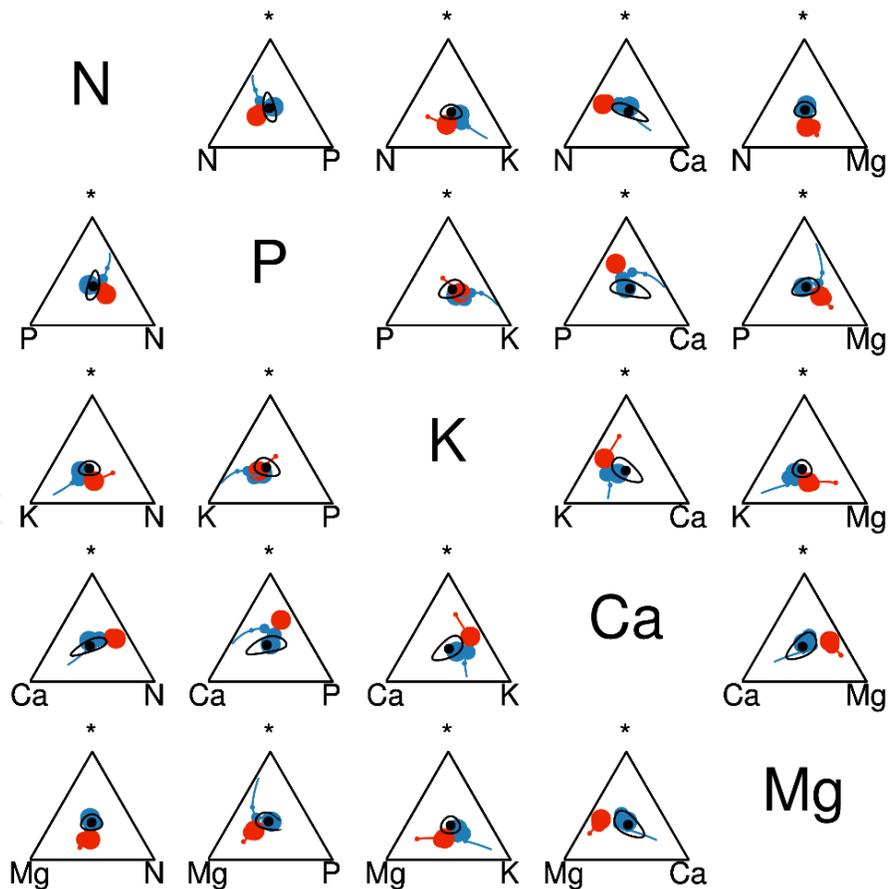


Figure 16. General view of plant ionomes in P and K deficient soils moving toward the reference ellipse for highly productive agroecosystems as P (blue) and K (red) fertilizers are added.

5. Compositional modeling of C mineralization of organic residues in soils

The carbon, nitrogen, phosphorus and sulfur cycles are interconnected in agroecosystems and often expressed using stoichiometric rules [4]. The ratio between total C and total N is the most simplified rule used in C mineralization studies but the $C_{\text{organic}}/N_{\text{organic}}$ and lignin/ N_{organic} ratios are also common. However, several biochemical components of organic matter are omitted in most studies, resulting in loss of information on the system. There are few studies on the relationship between labile or recalcitrant C and the biochemical composition of organic residues added to soil. [70] analyzed ash and N contents as well as four C fractions in organic residues representing pools of increasing resistance to decomposition. In this section, we related labile C in organic residues to this 6-part compositional vector of organic residues. The components were expressed as fractions on dry weight basis to compute a biological stability index using multiple linear regression models. The compositional vector was defined as follows:

$$S^6 = \mathcal{C}(\text{SOL}, \text{HEM}, \text{CEL}, \text{LIG}, \text{N}, \text{Ash}) \quad (10)$$

Where SOL = soluble matter, HEM = hemicellulose, CEL = cellulose, LIG = lignin, and N = total nitrogen.

Because scale dependency induces spurious correlations [71, 72, 73] and linear regression models are solved based on correlations between variables, the interpretation of regression coefficients is scale-dependent. To illustrate the problem of spurious correlations, chemical fractions were scaled on organic mass basis and analyzed using multiple linear regression.

The balance scheme reflected the C/N ratio and the order of decomposability of biochemical components (Figure 17). The SOL fraction was isolated from other biochemically labile fractions because its composition is complex, possibly including sugars, amino-sugars, amino-acids, and polypeptides as well as more recalcitrant or bacteriostatic easily solubilized polyphenols such as fulvic acids, tannic substances, resins, intermediate products, etc. The balance scheme was formalized by SBP as shown in Table 4.

<i>l/r balance</i>	SOL	HEM	CEL	LIG	Total N	Ash	r	s
[SOL,HEM,CEL,LIG,N Ash]	1	1	1	1	1	-1	5	1
[SOL,HEM,CEL,LIG N]	1	1	1	1	-1	0	4	1
[SOL,HEM,CEL LIG]	1	1	1	-1	0	0	3	1
[SOL HEM,CEL]	1	-1	-1	0	0	0	1	2
[HEM CEL]	0	1	-1	0	0	0	1	1

Table 4. Sequential binary partition of the biochemical composition of organic residues in Thuriès et al. (2002)



Figure 17. Graphical representation of the balances for the sequential binary partition defined in Table 3. SOL = soluble matter, HEM = hemicellulose, CEL = cellulose, LIG = lignin, N = total nitrogen, and ash.

The linear regression models relating labile C to bio-chemical fractions or balances showed R^2 values between 0.86 and 0.92 (Figure 18). For the 6-part (dry mass basis) and 5-part (organic matter basis) models, variation in labile C measured as evolved CO_2 was explained in part by total N and SOL as follows:

$$C_{labile} = 0.0773 + 0.5202SOL - 0.2515HEM - 0.3372CEL - 0.2882LIG + 2.3884N_{total} \quad (11)$$

$$C_{labile} = -0.1776 + 0.5585SOL + 0.1909HEM + 2.0187N_{total} \quad (12)$$

However, Equations 11 and 12 were subcompositionally incoherent. The intercept and the β coefficient for HEM showed opposite signs in equations 11 and 12 while CEL and LIG were absent in Equation 12. This incoherence is attributable to spurious correlations (Table 5). Pearson correlation coefficients among raw proportions were not consistent in terms of value, significance or sign whether the proportions were expressed on the dry mass of the organic product (including ash) or on organic matter (LOI) basis.

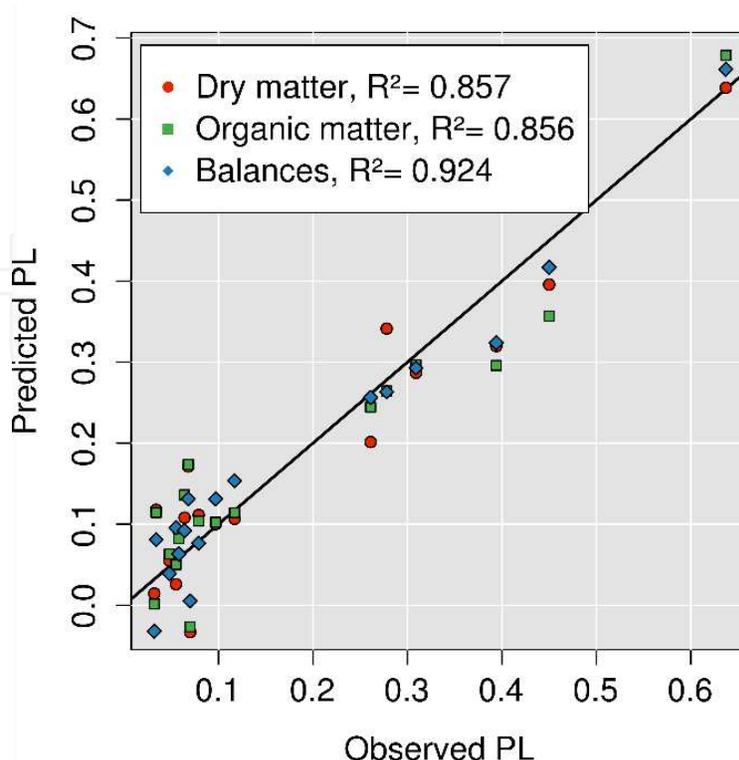


Figure 18. Regression analysis of C mineralization on dry matter (six raw components) or organic matter (five components) basis and using balances that reflects the order of decomposability of b

Component	SOL	HEM	CEL	LIG	Ash
Pearson correlation coefficient					
Dry matter basis (including ash)					
Total N	0.241	0.354	-0.462	-0.320	-0.184
SOL		-0.115	-0.232	-0.669	-0.027
HEM			-0.292	-0.293	-0.340
CEL				0.465	-0.495
Organic matter basis (loss on ignition)					
Total N	0.466	0.067	-0.637	-0.475	-
SOL		-0.194	-0.425	-0.756	-
HEM			-0.409	-0.383	-
CEL				0.376	-

SOL = soluble substances; HEM = hemicelluloses; CEL = cellulose; LIG = lignin and cutin

Table 5. Scale dependency of the correlations between biochemical components in Thuriès et al. (2002); for 17 observations, $r = 0.468$ at $P = 0.05$ and $r = 0.590$ at $P = 0.01$

On the other hand, the labile C pool was largely explained by the *ilr* balances between C sources and total N, a surrogate of the C/N ratio, the balance between labile and refractory C sources, and between two labile C pools, one being more labile (HEM) than the other (CEL). The equation was as follows:

$$C_{labile} = 0.3065 - 0.1251[SOL | HEM, CEL, LIG | N] + 0.0301[[SOL | HEM, CEL | LIG] + 0.0019[SOL | HEM, CEL] - 0.1063[HEM | CEL]] \quad (13)$$

Equation 13 shows that labile C increases with total N and higher proportions of more labile over more recalcitrant C forms. These findings indicate that the *ilr* coordinates provide a coherent interpretation of the C dynamics of organic products. The *ilrs* are not redundant, scale-invariant and free from spurious correlations.

6. Conclusions

This paper shows that the specific numerical properties of compositional data require log ratio transformations before conducting statistical analyses of soil and plant compositional data. Compared to raw concentration data, the orthonormal balances can be interpreted consistently and without numerical bias as isometric log ratio coordinates. The *ilr* approach can provide unbiased indices of nutrient balance in soils and plant tissues, biological stability of organic residues and soil quality. Well supported by techniques developed by compositional data analysts, the balance paradigm and the elaboration of its SBP schemes prompt that many concepts inherited from the past centuries be debated and revisited in soil fertility and plant nutrition.

Acknowledgements

The balance paradigm was elaborated within the plant nutrition and soil carbon modules of the research project entitled 'Implementing means to increase potato ecosystem services' (CRDPJ 385199 – 09). We acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC), the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), as well as farm partners as follows: Cultures Dolbec Inc., St-Ubalde, Québec, Canada; Groupe Gosselin FG Inc., Pont Rouge, Québec, Canada; Agriparmentier Inc. and Prochamps Inc., Notre-Dame-du-Bon-Conseil, Québec, Canada; Ferme Daniel Bolduc et Fils Inc., Péribonka, Québec, Canada.

Author details

S.É. Parent¹, L.E. Parent¹, D.E. Rozanne², A. Hernandez³ and W. Natale³

*Address all correspondence to: Leon-Etienne.Parent@fsaa.ulaval.ca

1 Department of Soils and Agrifood Engineering, Université Laval, Québec (Qc), Canada

2 Departamento de Agronomia, Unesp, Universidade Estadual Paulista, Campus de Registro, Registro, São Paulo, Brasil

3 Departamento de Solos e Adubos, Unesp, Universidade Estadual Paulista, Jaboticabal, São Paulo, Brasil

References

- [1] Sims, T. Soil fertility evaluation. In Sumner ME (ed.). Handbook of soil science. Boca Raton FL: CRC Press, 2000. p. D-113-D-153.
- [2] Epstein W., Bloom AJ. Mineral nutrition of plants: Principles and Perspectives. Sunderland MA: Sinauer Ass. Inc.; 2005.
- [3] Gliessman SR. Agroecology. Ecological processes in sustainable agriculture. Chelsea, MI: Ann Arbor Press; 1999.
- [4] Stevenson F. Cycles of soil. Carbon, nitrogen, phosphorus, sulfur, micronutrients. NY: Wiley Intersci.; 1986.
- [5] Lahner B, Gong J, Mahmoudian M, Smith EL, Abid KB, Rogers EE, Guerinot ML, Harper JF, Ward JM, McIntyre L, Schroeder, JI, Salt DE. Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. Nature Biotechnology 2003;21: 1215–1221.
- [6] Williams L, Salt DE. The plant ionome coming into focus. Current Opinion in Plant Biology 2003;12(3) 247–249.
- [7] Han WX, Fang J Y, Reich PB, Woodward FI, Wang ZH. Biogeography and variability of eleven mineral elements in plant leaves across gradients of climate, soil and plant functional type in China. Ecology Letters 2011;14: 788–796.
- [8] Schlotter D, Schack-Kirchner H., Hildebrand EE, von Wilpert K. Equivalence or complementarity of soil-solution extraction methods. Journal of Plant Nutrition and Soil Science 2012;175(2) 236-244.
- [9] Gabrielle B, Da-Silveira J, Houot S, Francou C. Simulating urban waste compost effects on carbon and nitrogen dynamics using a biochemical index. Journal of Environmental Quality 2004;33: 2333-2342.

- [10] Lashermes G, Nicolardot B, Parnadeau V, Thuriès L, Chaussod R, Guillotin ML, Li-nières M, Mary B, Metzger L, Morvan T, Tricaud A, Villette C, Houot S. Indicators of potential residual carbon in soils after exogenous organic matter application. *European Journal of Soil Science* 2009;60: 297-310.
- [11] Paul EA, Clark FE. *Soil biology and biochemistry*. NY: Academic Press; 1989.
- [12] McLean EO. Contrasting concepts in soil test interpretation: sufficiency levels of available nutrients versus basic cation saturation ratios. In M, Stelly (ed.) *Soil testing: Correlating and interpreting the analytical results*. Madison WI: American Society of Agronomy Special Publication 29; 1984. p. 39-54.
- [13] Walworth JL, Sumner ME. The Diagnosis and Recommendation Integrated System (DRIS). *Advances in Soil Science* 1987;6: 149-188.
- [14] Bergmann W. *Ernährungsstörungen bei Kulturpflanzen*. 2. Auflage. Stuttgart: Gustav Fisher Verlag; 1988.
- [15] Marshner H. *Mineral nutrition of higher plants*. NY: Academic Press; 1995.
- [16] Malavolta E. *Manual de nutrição de plantas*. Ceres, Piracicaba; 2006.
- [17] Ingestad T. New concepts on soil fertility and plant nutrition as illustrated by research on forest trees and stands. *Geoderma* 1987;40: 237-252.
- [18] Körner C. The grand challenges in functional plant ecology. *Frontiers in Plant Science* 2011;2: 1.
- [19] Geraldson CM. Intensity and balance concept as an approach to optimal vegetable production. *Communications in Soil Science and Plant Analysis* 1970;1(4) 187-196.
- [20] Geraldson CM. (1984). Nutrient intensity and balance. in Stelly, M. (ed) *Soil testing: Correlating and Interpreting Analytical Results*. Madison WI: American Society of Agronomy Special Publication 29; 1984.,p. 75-84.
- [21] Liebhardt WC. The Basic Cation Saturation Ratio Concept and Lime and Potassium Recommendations on Delaware's Coastal Plain Soils. *Soil Science Society of America Journal* 1981;52:1259-1264.
- [22] Kopittke PM, Menzies NW. A Review of the Use of the Basic Cation Saturation Ratio and the "Ideal" Soil Soil Science Society of America Journal 2007;71: 259-265.
- [23] Lagatu H, Maume L. Le diagnostic foliaire de la pomme de terre. *Annales de l'École Nationale Agronomique de Montpellier (France)* 1934;22: 50-158.
- [24] Egozcue JJ, Pawlowsky-Glahn V. Groups of parts and their balances in compositional data analysis. *Mathematical Geosciences* 2005;37: 795-828.
- [25] Beaufils ER. *Diagnosis and recommendation integrated system (DRIS)*. Soil Science Bulletin, 1, Pietermaritzburg: University of Natal; 1973.

- [26] Aitchison J. *The Statistical Analysis of Compositional Data*. London: Chapman and Hall; 1986.
- [27] Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. Isometric log-ratio transformations for compositional data analysis. *Mathematical Geology* 2003;35: 279-300.
- [28] Roy RN, Misra RV, Lesschen JP, Smaling AN.. Assessment of soil nutrient balance. Approaches and Methodologies. Rome: FAO fertilizer and plant nutrition bulletin 14; 2003. <http://www.fao.org/docrep/006/y5066e/y5066e00.htm> (accessed 20 August 2012).
- [29] Filzmoser P, Hron K. Robust statistical analysis. In Pawlowsky-Glahn V, Buccianti A (eds.) *Compositional data analysis: Theory and Applications*. NY: John Wiley and Sons; 2011. p. 57-72.
- [30]] Pawlowsky-Glahn V., Mateu-Figuerras, G., and Buccianti A.. *Compositional data analysis: Theory and Applications*. London: The Geological Society of London; 2006.
- [31] Pawlowsky-Glahn V, Buccianti A. *Compositional data analysis: Theory and Applications*. NY: John Wiley and Sons; 2011.
- [32] Filzmoser P, Hron K, Reimann C. Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Science of Total Environment* 2009;407(23) 6100-8.
- [33]] Tolosana-Delgado R, van den Boogart, KG. Linear models with compositions in R. In Pawlowsky-Glahn V, Buccianti A (eds.) *Compositional data analysis: Theory and Applications*. NY: John Wiley and Sons; 2011. p. 356-371.
- [34] Parent, LE, Dafir M. A theoretical concept of compositional nutrient diagnosis. *Journal of the American Society for Horticultural Science* 1992;117: 239-242.
- [35] Parent, LE, Cissé ES, Tremblay N, Bélair G. Row-centred logratios as nutrient indexes for saturated extracts of organic soils. *Canadian Journal of Soil Science* 1997;77: 571-578.
- [36] Lopez J, Parent LE, Tremblay N, Gosselin A. Sulfate accumulation and Ca balance in hydroponic tomato culture. *Journal of Plant Nutrition* 2002;25(7) 1585-1597.
- [37] Pawlowsky-Glahn V, Egozcue JJ. Exploring Compositional Data with the CoDa-Dendrogram. *Austrian Journal of Statistics* 2011;40(1&2) 103-113.
- [38] Güsewell S. N:P ratios in terrestrial plants: variation and functional significance. *New Phytologist* 2004;164: 243-266.
- [39] Egozcue JJ., Daunis-i Estadella J, Pawlowsky-Glahn V, Hron K, Filzmoser P. Simplicial regression. The normal model. *Journal of Applied Probability and Statistics* 2012;6(1&2) 87-108.

- [40] Buccianti A. Natural laws governing the distribution of the elements. In Pawlowsky-Glahn V, Buccianti A (eds.) *Compositional data analysis: Theory and Applications*. NY: John Wiley and Sons; 2011. p. 255-266.
- [41] Parent, LE. Diagnosis of the nutrient compositional space of fruit crops. *Revista Brasileira Fruticultura* 2011;33: 321-334.
- [42] Parent LE, Parent S.É, Rozane DE, Amorim DA, Hernandez A, Natale W. Unbiased approach to diagnose the nutrient status of guava. *Acta Horticulturae* 2012 (in press).
- [43] Abdi, D, Ziadi N, Parent, LE. Compositional analysis of phosphorus pools in Canadian Mollisols. Proceedings of the 4th International Workshop Compositional Analysis, 9-13 May 2011, San Feliu-de-Guixols, Girona, Spain. <http://www.ma3.upc.edu/users/ortego/codawork11-Proceedings/Admin/Files/FilePaper/p57.pdf> (accessed 20 August 2012).
- [44] Parent LE, Parent, S-É, Kätterer T, Egozcue JJ. Fractal and compositional analysis of soil aggregation., Girona, Spain: Proceedings of the 4th International Workshop Compositional Analysis, 9-13 May 2011, San Feliu-de-Guixols, Girona, Spain. <http://www.ma3.upc.edu/users/ortego/codawork11-Proceedings/Admin/Files/FilePaper/p5.pdf> (accessed 20 August 2012).
- [45] Parent LE, de Almeida C X, Hernandez A, Egozcue JJ, Gülser C, Bolinder MA, Kätterer T, Andrén O, Parent S.-É, Anctil F, Centurion JF, Natale W. Compositional analysis for an unbiased measure of soil aggregation. *Geoderma* 2012;179–180: 123-131.
- [46] Rodgers, J. L., Nicewander, W. A. and Toothaker, L. 1984. Linearly independent, orthogonal, and uncorrelated variables. *Amer. Statist.* 38,133-134.
- [47] Egozcue JJ, Pawlowsky-Glahn, V. Simplicial geometry for compositional data. In *Compositional Data Analysis in the Geosciences: From Theory to Practice*. Geological Society (London, England), 145-159.
- [48] Lovell D, Müller W, Tayler J, Zwart A, Helliwell C. Proportions, percentages, ppm: do the molecular biosciences treat compositional data right? In Pawlowsky-Glahn V, Buccianti A (eds.) *Compositional data analysis: Theory and Applications*. NY: John Wiley and Sons; 2011. p. 193-207.
- [49] Parent LE, Natale W, Ziadi, N. Compositional Nutrient Diagnosis of Corn using the Mahalanobis Distance as Nutrient Imbalance Index. *Canadian Journal of Soil Science* 2009;89: 383-390.
- [50] Hadi A. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society* 1992;B54: 761_771.
- [51] Hadi A. A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society* 1994;B56: 393-396.

- [52] Nelson LA, Anderson RL. Partitioning of soil test-crop response probability. in M. Stelly (ed.) *Soil testing: Correlating and interpreting the analytical results*. Madison WI; American Society of Agronomy Special Publication 29I; 1984. P. 19-38.
- [53] MedStats Club. Statistics tutorials, calculators and tools (2012). <http://www.medstats.org/sens-spec> Accessed 20 August 2012.
- [54] Lander ES, Botstein, D. Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics* 1989;121: 185-199.
- [55] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Version 2.13.1; 2011. <http://www.R-project.org> (accessed 20 August 2012).
- [56] van den Boogaart KG, Tolosana-Delgado R,, Bren, M. *compositions: Compositional Data Analysis*. R package version 1.10-2 ; 2012. <http://cran.r-project.org/web/packages/compositions/compositions.pdf> (accessed 20 August 2012).
- [57] Thode HC Jr. *Testing for Normality*. NY: Marcel Dekker; 2002.
- [58] Gross J. *nortest: Five omnibus tests for the composite hypothesis of normality*. R package version 1.0 ; 2006. <http://CRAN.R-project.org/package=nortest> (accessed 20 August 2012).
- [59] Filzmoser, P, Gschwandtner, M. *mvoutlier: Multivariate outlier detection based on robust methods*. R package version 1.9.4 ; 2011. <http://CRAN.R-project.org/package=mvoutlier> (accessed 20 August 2012).
- [60] Embrapa. *Sistema Brasileiro de classificação de solos*. 2nd Ed., Rio de Janeiro: Embrapa Solos; 2006.
- [61] Raij B van, Quaggio JA, Cantarella H, Ferreira ME, Lopes AS, Bataglia OC. *Análise química do solo para fins de fertilidade*. Campinas: Fundação Cargill; 1987.
- [62] Shoemaker HE, Mclean EO, Pratt PF. Buffer methods for determining lime requirement of soils with appreciable amounts of extractable aluminium. *Soil Science Society of America Proceedings* 1961;25: 274-277.
- [63] Quaggio JA, van Raij B., Malavolta E. Alternative use of the SMP-buffer solution to determine lime requirement of soils. *Communications in Soil Science and Plant Analysis* 1985;16: 245-260.
- [64] Callot G, Chamayou H, Maertens C, Salsac L. *Mieux comprendre les interactions sol-racine: incidence sur la nutrition minérale*. Paris : INRA; 1982.
- [65] Smith GS, Buwalda JB., Clark CJ. Nutrient dynamics in a kiwifruit ecosystem. *Scientia Horticulturae* 1988;37: 87-109.
- [66] Salt ED, Baxter I, Lahner B. Ionomics and the study of the plant ionome. *Annual Review of Plant Biology* 2008;59: 709-33.

- [67] Parent LE, Cambouris AN, Muhawenimana, A. Multivariate diagnosis of nutrient imbalance in potato crops. *Soil Science Society of America Journal* 1994;58:1432-1438.
- [68] Loladze I, Elser, JJ. The origins of the Redfield nitrogen-to-phosphorus ratio are in a homeostatic protein-to-rRNA ratio. *Ecology Letters* 2011;14: 244-250.
- [69] Usuda H. Phosphate deficiency in maize. V. Mobilization of nitrogen and phosphorus within shoots of young plants and its relationship to senescence. *Plant Cell Physiology* 1995;36: 1041–1049.
- [70] Thuriès L, Pansu M, Larré-Larrouy MC., Feller C. Biochemical composition and mineralization kinetics of organic inputs in a sandy soil. *Soil Biology and Biochemistry* 2002;34: 239-250.
- [71] Pearson K. Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society* 1897;LX: 489-502.
- [72] Tanner J. Fallacy of per-weight and per-surface area standards, and their relation to spurious correlation. *Journal of Physiology* 1949;2: 1-15.
- [73] Chayes F. On correlation between variables of constant sum. *Journal of Geophysical Research* 1960;65: 4185-4193.