NUMERICAL BIASES IN N MINERALIZATION MODELS OF ORGANIC RESIDUES

Léon Etienne Parent*[1], Serge-Étienne Parent[1], Antoine Karam[1]

[1]*Department of Soils and Agrifood Engineering, Paul-Comtois building, Laval University, Quebec (QC), Canada G1V0A6.*

Email: Leon-etienne.parent@fsaa.ulaval.ca

SUMMARY: Biochemical fractions of organic residues are compositional data that generate numerical biases and are subject to non-linear processes upon amalgamation into synthetic values. Our objective was to measure the inherent sources of distortion using percentage or fraction data in N mineralization models. Nine organic residues were incubated for 26 wk. Numerical bias was small using simple log-transformed data but amalgamation of C forms (easily extractable C, cellulose, lignin) largely distorted the C/N index. Balancing against N the sources of C according to their order of decomposability improved the N mineralization model. We suggest using the balance concept across a gradient of decomposability of biochemical fractions of organic residues in N mineralization studies.
Keywords: C/N ratio, data amalgamation, isometric log ratio, numerical bias

INTRODUCTION

Indices of N availability from organic residues such as the C/N and lignin/N ratios are widely used in C models (Parton et al., 1987). Such indices have two major limitations: (1) there are many C and N forms of contrasting decomposability in organic residues (Tremblay et al., 2010), and (2) analytical data used to described C and N forms are compositional, hence, inherently distorted (Parent et al., 2012). Organic residues contain several C and N forms associated with the soluble, hemicelluloses, cellulose, and lignin fractions of organic residues (Thuriès et al., 2002). Every organic residue can be described by a compositional vector of analytical data are intrinsically multivariate (Filzmoser et al., 2009). One component is redundant as computed by difference between 100% and others. The sources of numerical biases in compositional data (redundancy, scale dependency, and non-normal distribution) distort univariate and multivariate analysis of compositional data (Aitchison, 1986; Filzmoser et al., 2009). Another source of distortion in compositional data analysis is amalgamation whereby analytical data are added up to a synthetic value (Aitchison and Egozcue, 2005). The objective of this paper is to model without bias the N mineralization potential of organic residues.

MATERIAL AND METHODS

Briefly, eight agricultural organic residues and one urban compost were dried at 40°C, ground to < 1 mm, and analyzed by FT-NIR for biochemical fractions (soluble, hemicellulose, cellulose, lignin), by dry combustion for C and N, by ion chromatography for NH_4-N and NO_3-N after 0.01 M $CaCl_2$ extraction, and by loss on ignition (550°C) for ash content (Tremblay et al., 2010). The biosolids providing 45 mg organic N kg^{-1} soil were incubated in triplicates in two soils at room T° and under adequate moisture conditions for 26 wk. Analytical data were log or isometric log-ratio (*ilr*) transformed and analyzed statistically using R (van den Boogart et al., 2011) as well as the Excel-2007 software. The *ilr* is a log contrast balancing two subsets. The *ilr* is computed as (Egozcue

et al., 2003): $ilr = \sqrt{\frac{rs}{r+s}} \ln\left(\frac{g(c_+)}{g(c_-)}\right)$, where $\sqrt{\frac{rs}{r+s}}$ is an orthogonal coefficient, where r is the number of components for the subset at numerator and s is the number of components for the subset at denominator; $g(c_+)$ is geometric mean of components in subset at numerator; $g(c_-)$ is geometric mean of components in subset at denominator. The sequential binary partition was arranged according to decomposability of C forms (Table 1). Total C is anamalgamation of easily extractable, cellulosic and lignin sources of C. Due to orthogonality, a Euclidean distance called Aitchison distance can be computed across ilr coordinates as: $\mathcal{A} = \sqrt{\sum_{j=1}^{D-1}(ilr_j - ilr_j^*)^2}$, where ilr_j is the jth ilr of some composition and ilr_j^* is the ilr for a reference composition. The Euclidean distance (ε) between compositions x and y using simple log is biased by the difference between geometric means times number of parts as (Lovell et al., 2011): $\mathcal{E}^2(\ln(x), \ln(y)) = \mathcal{A}^2 + D\left(\ln\frac{g(x)}{g(y)}\right)^2 \geq \mathcal{A}^2$.

RESULTS AND DISCUSSION

The relationship between the Euclidean distance computed using ln-transformed fractions and the Aitchison distance across balances shows that numerical biases due to redundancy and non-normal distribution was small in this dataset (Fig. 1). Although similar performance is expected due to small numerical bias when conducting multivariate analyses such as multiple regression analysis across simple log or log-ratio transformed compositional data, incoherence between sign or magnitude of regression coefficients may occur at the step of interpreting the results (Parent et al., 2012). The C/N ratio computed by amalgamating sources of C into total C was weakly correlated to balances between N and the C pools in organic residues (Fig. 2). Amalgamation is a non-linear process in the compositional vector (Aitchison and Egozcue, 2005). Although the C/N ratio is a fair representation of N mineralization in organic residues (Fig. 3), the [EEC,CEL,LIG | N] balance provided a more accurate N mineralization model (Fig. 4). Other balances were poorly related to mineralized N (data not shown).

CONCLUSION

This work showed that recognizing that biochemical fractions of organic residues are compositional data improved the N mineralization model for nine organic residues. We thus recommend using the balance concept in N mineralization studies.

ACKNOWLEDGEMENTS

REFERENCES

Aitchison J. (1986). The Statistical Analysis of Compositional Data. Chapman & Hall, London.

Aitchison, J. and Egozcue, J.J. 2005. Compositional data analysis: where are we and where should we be heading? *Mathematical Geology* 37(7), 829-850.

Egozcue J.J., Pawlowsky-Glahn V., Mateu-Figueras G.,and Barceló-Vidal C. (2005). Isometric log-ratio transformations for compositional data analysis. *Mathematical Geology* 35, 279-300.

Filzmoser P., Hron K., and Reimann C. (2009). Univariate statistical analysis of environmental (compositional) data: Problems and possibilities. *Science of Total Environment*, 407(23), 6100-6108.

Lovell D, Müller W, Tayler J, Zwart A, Helliwell C. (2011). Proportions, percentages, ppm: do the molecular biosciences treat compositional data right? In *Compositional data analysis: Theory and Applications,V.* Pawlowsky-Glahn and A. Buccianti (eds.). John Wiley and Sons, NY, pp. 193-207.

Parent, S.-É., Parent, L.E. Rozanne, D.E. Hernandes, A., and Natale, W. (2012). Nutrient balance as paradigm of plant and soil chemometrics. InTech Publ., NY (in press).

Parton, W.J., Schimel, D.S., Cole, C.V., Ojima, D.S. (1987). Analysis of factors controlling soil organic levels of grasslands in the Great Plains. *Soil Science Society of America Journal*. 51, 1173-1179.

Thuriès L., Pansu M., Larré-Larrouy M.C.., Feller C. (2002). Biochemical composition and mineralization kinetics of organic inputs in a sandy soil. *Soil Biology and Biochemistry*, 34, 239-250.

Tremblay, M.E., Nduwamungu, C., Parent, L.E., and Bolinder, M.A. (2010). Biological Stability of C and N in Organic Products and Crop Residues using FT-NIR Reflectance Spectroscopy. *Communication in Soil Science and Plant* Analysis 41(8), 917-934.

van den Boogaart, K. G., Tolosana-Delgado, R., Bren, M. (2011) compositions: Compositional Data Analysis in R package version 1.10-2, http://CRAN.R-project.org/package=compositions (accessed 24 September 2012).

Table 1. Sequential binary C-N partition following relative decomposability of C forms; EEC = easily extractable (soluble + hemicellulose), CEL = cellulose, LIC = lignin + cutin.

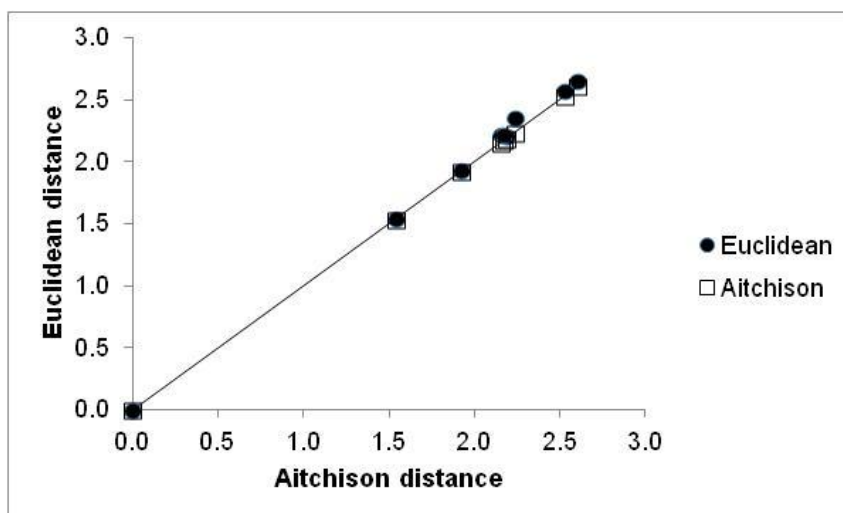| EEC	CEL	LIG	N	r	s	Balance designation	Isometric log ratio
1	1	1	-1	3	1	[EEC,CEL,LIC | N]	$\sqrt{\frac{3 \times 1}{3+1}} ln\left(\frac{\sqrt[3]{EECxCELxLIC}}{N} \right)$
1	1	-1	0	2	1	[EEC,CEL | LIC]	$\sqrt{\frac{2 \times 1}{2+1}} ln\left(\frac{\sqrt[2]{EECxCEL}}{LIC} \right)$
1	-1	0	0	2	1	[EEC | CEL]	$\sqrt{\frac{1 \times 1}{1+1}} ln\left(\frac{EEC}{CEL} \right)$

Figure 1. There is no apparent numerical bias in these ordinary log-transformed data because geometric means of compositions correspond (Eq. 3).
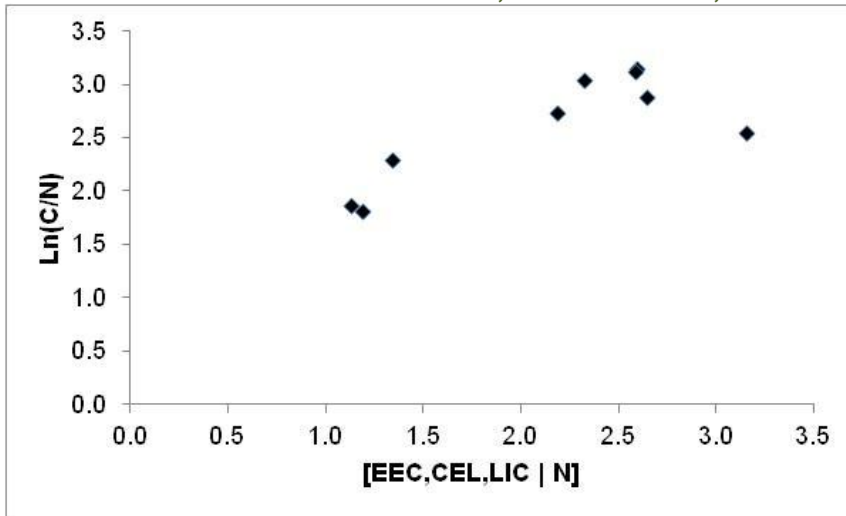
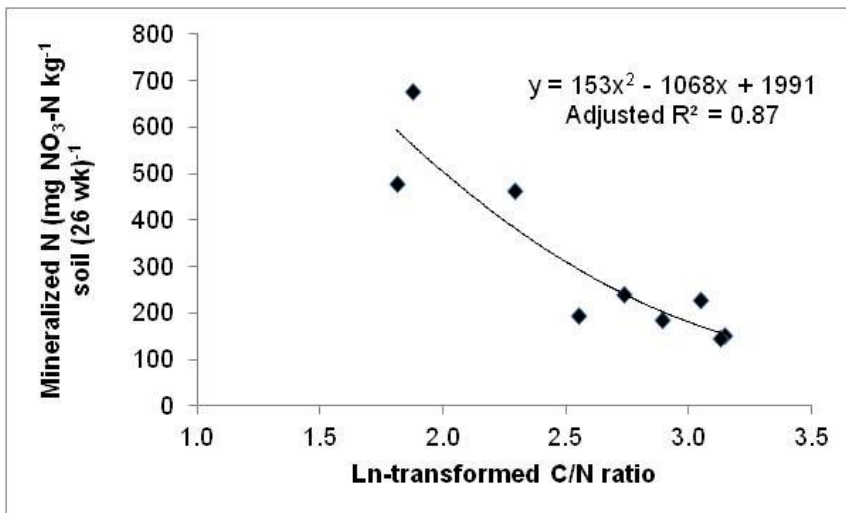Figure 2. Distortion upon amalgamating C sources into total **C.**

Figure 3. Non-linear relationship between mineralized N and the C/N ratio.
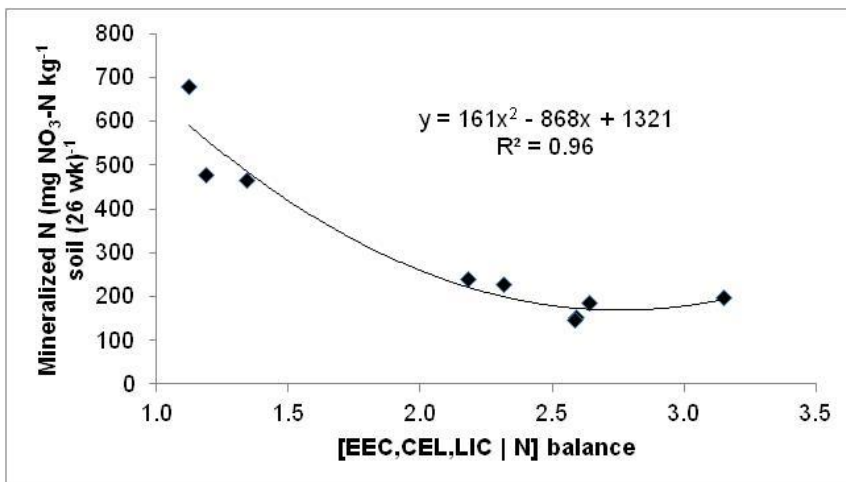

Figure 4. Non-linear relationship between mineralized N and the [EEC,CEL,LIC | N] balance. EEC = soluble + hemicellulose fractions; CEL = cellulose; LIC = lignin + cutin.