Hydrol. Earth Syst. Sci. Discuss., 9, C4913–C4926, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C4913/2012/

Hydrology and
Earth System
Sciences
Discussions

Interactive comment on "Local sensitivity analysis for compositional data with application to soil texture in hydrologic modelling" *by* L. Loosvelt et al.

J. J. Egozcue (Referee)

juan.jose.egozcue@upc.edu

Received and published: 20 October 2012

Analysis of sensitivity with respect to a compositional parameter

A comment on "Local sensitivity analysis for compositional data with application to soil texture in hydrologic modelling" by L. Loosvelt and co-authors.

by

C4913

J. J. Egozcue[1] and V. Pawlowsky-Glahn[2]

[1] Dept. Applied Mathematics III, Universitat Politècnica de Catalunya, Barcelona, Spain. e-mail: juan.jose.egozcue@upc.edu
[2] Dept. Computer Science, Universitat de Girona, Spain. e.mail: vera.pawlowsky@udg.edu

1 Introduction

To our understanding the article by L. Loosvelt and co-authors is an important contribution (Loosvelt et al., 2012). The authors identify an important problem, the sensitivity analysis to changes in compositional input parameters, and propose a way to deal with. The problem is that input parameters of hydrological models can be compositional, and variations of these parameters should be treated in an appropriate geometry. In the studied case, the input of the model TOPLAST (see references Loosvelt et al. 2012) is the *clay-sand-silt* composition characterising the soil texture. The goal was to carry out a sensitivity analysis of the output soil hydraulic parameters (SHPs) taking into account the compositional character of the input texture parameters. The presented analysis is methodologically sound and the obtained results are potentially useful for further use of TOPLAS and for improvements in the sampling techniques of soil characteristics. The merit of this contribution is daring to use appropriate compositional techniques despite of not commonly used in this context. Accordingly, the contribution states an important criticism on methods which ignore the compositional character of used data and parameters. We agree with this criticism and we would encourage the revision of methods used in geosciences, and all scientific fields, which deal with compositional data and/or compositional parameters ignoring their character and overlooking the consequences.

At this point, a definition of compositional data or parameters is worth, since the authors use a restrictive definition. However, this view has no further consequences in the paper. They use the classical idea that a composition is a vector with positive components adding to a constant. Accordingly, they are called *closed data* as it is frequently done in geosciences, e.g. Chayes (1960); Butler (1978); Chayes and Trochimczyk (1978); Buccianti and Rosso (1999). As the authors point out, an important characteristic of compositions is that the only information in a composition is contained in the ratios between the components (Aitchison, 1986). However, compositions are better thought of as equivalence classes of vectors with positive components: two of these vectors are equivalent if their components are proportional (Barceló-Vidal et al., 2001). A further step is that components of compositions do not need to add to a constant (Buccianti and Pawlowsky-Glahn, 2005; Egozcue and Pawlowsky-Glahn, 2011). Typical examples in geosciences are concentrations given in molar concentrations or in mg per liter. This kind of compositions can be changed into (closed) proportions using a *perturbation*, the addition in the simplex, without loss of information.

In Loosvelt et al. (2012), *clay-sand-silt* proportions are adequately considered as a composition –see an example in Aitchison (1986)–. However, other compositions are mentioned in the paper. For instance, the soil moisture content θ_r (m^{3}m^{-3}) can be considered as a two-part composition; also porosity can be treated as a two-part composition. It could have consequences in the sensitivity analysis proposed.

The present comment is centered in three specific points. The first one is related to the fact that the authors avoid the use of ilr-coordinates. The second one refers to some generalization of sensitivity analysis when input parameters are compositional. The third tries to show that the role of the Dirichlet distribution in the sensitivity analysis is irrelevant. These points should be considered as a positive consequence of the Loosvelt et al. (2012) contribution and they are intended to encourage further studies.

2 Use of ilr-coordinates in the simplex

The D-part simplex, $\mathcal{S}^D$, is the sample space of random compositions and the parameter space of compositional parameters. In Pawlowsky-Glahn and Egozcue (2001) and in Billheimer et al. (2001) the D-part simplex, with perturbation, powering and the Aitchison metrics (see eq. (1)–(4) in Loosvelt et al., 2012) was identified to be a $(D-1)$-dimensional Euclidean space. The corresponding geometry was called the *Aitchison geometry of the simplex*. The main consequence of this result is that any composition can be represented by its Cartesian coordinates, called *isometric log-ratio* (ilr)-coordinates (Egozcue et al., 2003). When compositions are represented using ilr-coordinates, perturbation, powering, inner-product, distance and norm in the simplex are translated into the ordinary real operations (sum, multiplication by scalars) and real metric (Euclidean inner product, distance and norm). The representation in ilr-coordinates provides a framework where the Aitchison geometry of the simplex is easily handled both for geometric and statistic computation. This practice has been named as the *the principle of working on coordinates* (Mateu-Figueras et al., 2011).

The ilr-coordinates involve log-ratios of products of components of the composition, but they can be very simple. The prototype of simple ilr-coordinates are those obtained using a *sequential binary partition* of the composition which are called *balances* (Egozcue and Pawlowsky-Glahn, 2005, 2006). In Fig. 2 of Loosvelt et al. (2012) the balances

$$x_1^* = \frac{1}{\sqrt{2}} \log \frac{x_1}{x_2} \ , \ x_2^* = \sqrt{\frac{2}{3}} \log \frac{(x_1 x_2)^{1/2}}{x_3} \ , \tag{1}$$

are used as ilr-coordinates. It shows a circle centered at the origin of these ilr-coordinates, crossed by three axes; the center corresponds to the neutral element of $\mathcal{S}^3$ (the barycenter of the ternary diagram) and the bisector lines of the angles of the ternary diagram denoted B_1, B_2, B_3. A translation produces a shifted figure (dotted lines). The shift in the ilr-coordinates is equivalent to a perturbation in the simplex. The

Fig. 1. Space of ilr-coordinates (x_1^*, x_2^*). Full lines: a circle of radius 0.5 centered at the origin and crossed by 6 axes. Dashed lines: the same figures shifted to the point $(2.2, 0.7)$. Bold lines: bisectors shown in Loosvelt et al. (2012). The axes of coordinates x_1^* and x_2^* are not shown.

Fig. 2. Ternary diagram corresponding to Fig. 1. Full lines: a circle of radius 0.5 centered at the origin and crossed by 6 axes. Dashed lines: the same lines shifted to the $\text{ilr}^{-1}(2.2, 0.7)$. Bold lines: bisectors shown in Loosvelt et al. (2012).

effect of the shift is shown in Fig. 3 (Loosvelt et al., 2012). An important feature of Fig. 3 is that the shifted bisectors B_1', B_2', B_3' appear again as straight-lines in the ternary diagram. In fact, linear segments in the ternary diagram are again transformed into linear segments after a perturbation (von Eynatten et al., 2002). However, the mentioned Figures hide some facts of the Aitchison geometry: stright-lines in the ilr-coordinate space generally correspond to curved lines in the ternary diagram. Figures 1 and 2 show a similar construction including more axes than in Loosvelt et al. (2012). In Fig. 1, the axes at angles $\pi/6$, $\pi/2$, $5\pi/6$, for both the centered circle and the shifted one, are transformed into curved axes in the ternary diagram in Fig. 2.

Our purpose with Figures 1 and 2 is to show that computing the $2M$ intersections of M axes regularly distributed on a circle is a simple task. In fact, the ilr-coordinates are

$$u_k^* = r\cos((\pi k)/M)\ ,\ v_k^* = r\sin((\pi k)/M)\ ,\ k = 0, 1, 2, \ldots, 2M\ ,$$

where r is the radius of the circle. Computing these intersections in the ternary diagram is quite demanding (see Appendix in Loosvelt et al., 2012), even for the bisector axes. If the simplicial expression of these points is needed, the ilr^{-1}-transformation can be

used. The ilr-transformation and its back transformation, in matrix notation, are

$$\mathbf{x}^* = \text{ilr}(\mathbf{x}) = V^\top \log(\mathbf{x}) \quad , \quad \mathbf{x} = \text{ilr}^{-1}(\mathbf{x}^*) = \mathcal{C}\exp[V\mathbf{x}^*]\ ,$$

where $\log$ and $\exp$ operate componentwise, $\mathcal{C}$ is the closure operator, and V is the contrast matrix associated with the specific ilr-transformation (Egozcue et al., 2011). In Loosvelt et al. (2012), the expression in the simplex of the intersection points (u_k^*, v_k^*) is

$$\text{ilr}^{-1}\begin{pmatrix} u_k^* \\ v_k^* \end{pmatrix} = \mathcal{C}\ \exp\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{6} \\ 0 & -\sqrt{2/3} \end{pmatrix} \cdot \log\begin{pmatrix} u_k^* \\ v_k^* \end{pmatrix},$$

where the matrix corresponds to V.

The algebra of ilr-transformations allows an easy generalization to more than the $M = 3$ axes (six points) taken in the perturbation circle presented in Loosvelt et al. (2012).

In a general case where the texture of the soil is described with more than 3 parts, e.g. Parent et al. (2012), the perturbation circle is not easily generalized to the sphere and hyper-spheres. A larger number of points on the hyper-sphere would be necessary and the distribution of the axes may cause difficulties.

3 Sensitivity, scale and derivatives

Local sensitivity of an output parameter θ is described by a derivative of θ with respect to input parameters, as Loosvelt et al. (2012) state. In practice, the derivative is approximated, for instance, using finite differences. A sensible question is, which is the scale of the parameter θ or, equivalently, how the difference between two values of θ is computed. For instance, the soil moisture content θ_r (m^{3}m^{-3}) could be considered as a composition of two parts: moisture, solid+gas. If this choice is taken, the difference between two values θ_{r1}, θ_{r2} is computed according perturbation-difference, i.e.

as $\mathcal{C}[\theta_{r1}/\theta_{r2}, (1-\theta_{r1})/(1-\theta_{r2})]$. The square-distance is then

$$d_A^2(\theta_{r1}, \theta_{r2}) = \left(\frac{1}{\sqrt{2}} \log \frac{\theta_{r1}}{1-\theta_{r1}} - \frac{1}{\sqrt{2}} \log \frac{\theta_{r2}}{1-\theta_{r2}} \right)^2 ,$$

which corresponds, up to a constant, to the squared difference of the *logit* transformations. The log-ratio $2^{-1/2} \log(\theta_{r1}/(1-\theta_{r1})$ is the ilr coordinate of the composition $(\theta_{r1}, 1-\theta_{r1})$. Taking the difference between two values θ_{r1}, θ_{r2} as the difference $\theta_{r1} - \theta_{r2}$, implies assuming that θ_r has the absolute scale. In this case, the difference between soil moisture content of 10^{-3} and 10^{-2} is equal to the difference between 0.300 and 0.309. In the first pair, the moisture content of the second value is 10 times the first one; for the second pair, the moisture content of the second value is only slightly greater than the first one. The option taken in Loosvelt et al. (2012) assumes that the differences for the two pairs are equal.

A way to deal with the choice of scale of an output parameter θ is to select a one-to-one function φ such that $\varphi(\theta)$ has an absolute scale. For instance, if θ_r is considered compositional, $\varphi(\theta) = 2^{-1/2} \log(\theta_r/(1-\theta_r)$; if the hydraulic conductivity K (ms^{-1}), is considered in a ratio scale, then $\varphi(K) = \log(K)$. The decision on the scale of each parameter is subjective but implications of such a decision should be carefully analysed.

Local sensitivity of an output parameter θ consists of estimating the derivative of $y = \varphi(\theta)$ with respect the soil texture $\mathbf{x}$ represented in ilr-coordinates by $\mathbf{x}^* = \text{ilr}(\mathbf{x})$. In this case, the derivative at $\mathbf{x}$ is the gradient

$$\nabla y(\mathbf{x}) = \left(\frac{\partial y}{\partial x_1^*}, \frac{\partial y}{\partial x_2^*}, \ldots, \frac{\partial y}{\partial x_d^*} \right) , \tag{2}$$

where $d = D - 1$ is the dimension of the simplex. The derivative of $y = \varphi(\theta)$ in a compositional direction $\mathbf{v}$, with ilr-coordinates $\mathbf{v}^*$ and $\|\mathbf{v}^*\| = 1$, is $D_{\mathbf{v}} y(\mathbf{x}) = \nabla y(\mathbf{x}) \cdot \mathbf{v}^*$, where the dot is the ordinary Euclidean inner product of vectors. For φ being the

identity function, the directional derivative can be approximated by the finite difference in Eq. (6) of (Loosvelt et al., 2012). It is remarkable that, once the gradient $\nabla y(\mathbf{x})$ in Eq. (2) is estimated, any directional derivative is obtained as $\nabla y(\mathbf{x}) \cdot \mathbf{v}^*$; therefore, only d derivatives are needed, one for each orthogonal axis on the ilr-coordinate space. This is not relevant for $d = D - 1 = 2$ as in Loosvelt et al. (2012). Alternatively for $D > 3$, a large number of points on the perturbation hyper-sphere can be required to compute the scalar sensitivity index. Then, using the gradient in (2) can simplify the computation of the directional derivatives dramatically.

In Loosvelt et al. (2012), the scalar sensitivity analysis continues taking absolute value of directional derivatives, computing the root mean square average of them, and finally averaging the root-mean-square values on the perturbation circle (see Eq. (8)). The appropriate order of these operations can be discussed, but it is out of the scope of this comment.

4 Use of Dirichlet distribution and measure in the simplex

In order to get a representation in the ternary diagram of sensitivity indexes, Loosvelt et al. (2012) propose to use a Dirichlet sampling providing points in which the sensitivity index is computed and afterwards interpolated. Particularly, they use the uniform distribution in the ternary diagram. For representation purposes, a deterministic uniform triangular grid is enough and would guarantee quality interpolations similar to those presented in Figure 6. We think that using the Dirichlet sampling may induce confusion to the reader. It appears as a piece of the compositional analysis, while it is simply a technique to obtain a uniform grid of points. At this point, a question arises: what is a uniform grid of points in the simplex? or, more technically, how are areas measured in the simplex?

For the three-part simplex (silt-sand-clay), the space of ilr coordinates is the two dimensional real space $\mathbb{R}^2$. The measure (area) of a rectangle is the product of the lengths of the two sides in the ordinary sense. This measure is called Lebesgue measure in $\mathbb{R}^2$. A regular uniform grid is obtained intersecting two orthogonal families of equally-spaced parallel axes. If this grid of points in the ilr-coordinate space is translated back into the simplex, the squares appear distorted in the ternary diagram (Fig. 3). The squares of the regular grid in coordinates have Lebesgue measure 1×1. In the ternary diagram they appear strongly distorted, the closer to the borders the more distortion, but they have unit Aitchison measure, i.e. they are squares of unit Aitchison side-length. The regular grids in the simplex, as that shown in Fig. 3, may be inadequate to plot surfaces similar to those shown in Fig. 6 in Loosvelt et al. (2012), but they are advisable for a plot in the ilr-coordinate space, as they do not distort measures near the borders of the ternary diagram. This comment is more than an advise on how to plot surfaces on the two-part simplex. It essentially concerns the the computation of mean values of the local sensitivity indexes. For instance, the mean value of the sensitivity index over an USDA-class of texture can be estimated as the arithmetic average of sensitivity indexes computed over all soil texture points evaluated within the USDA-class if the grid points are uniformly sampled in the ilr-coordinate space. However, this is not the case if the sampling was Dirichlet (unit parameters) or selected on a deterministic regular triangular grid on the ternary diagram.

We would suggest to perform the sensitivity analysis on a (limited) regular grid in ilr-coordinates, where both contour plots of the index and mean values are not distorted. As a traditional way of presenting results, ternary plots may be also useful, but distortion near the borders should be taken into account.

Fig. 3. Regular grid in the three-part simplex. Thick lines correspond to the axes of the ilr-coordinates in Eq. (1). Other lines are parallel axes shifted ± 1, ± 2 forming a grid of 25 points.

5 Concluding comment

The contribution by Loosvelt et al. (2012) is relevant because it fosters the problem of sensitivity analysis of a numerical model. These kind of analyses are not new, but are frequently overlooked in the standard practice. The novelty is considering the compositional character of the input parameters.

The sensitivity study of the output hydraulic parameters of the TOPLATS model with the input texture of soil leads to important conclusions. One of them is that USDA-classes seem to be too rough to characterize the soil texture. We would add that three grain classes (silt, sand, clay) are not enough for an accurate description of the soil texture, thus claiming for a revision of the code.

Our comments try to point out ways to deal with a more detailed description of the soil texture. Specifically, we remark the importance of working with ilr-coordinates to improve computation, evaluation of mean values of sensitivity indices and plotting techniques.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press). 416 p.

Barceló-Vidal, C., J. A. Martín-Fernández, and V. Pawlowsky-Glahn (2001). Mathematical foundations of compositional data analysis. In G. Ross (Ed.), *Proceedings of IAMG'01 — The sixth annual conference of the International Association for Mathematical Geology*, pp. 20 p. CD-ROM.

Billheimer, D., P. Guttorp, and W. Fagan (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association 96*(456), 1205–1214.

Buccianti, A. and V. Pawlowsky-Glahn (2005). New perspectives on water chemistry and compositional data analysis. *Mathematical Geology 37*(7), 703–727.

Buccianti, A. and F. Rosso (1999). A new approach to the statistical analysis of compositional (closed) data with observations below the "detection limit". *Geoinformatica 3*, 17–31. Litografia Editrice De Frede, Napoli (I).

Butler, J. C. (1978). Visual bias in R-mode dendrograms due to the effect of closure. *Mathematical Geology 10*(2), 243–252.

Chayes, F. (1960). On correlation between variables of constant sum. *Journal of Geophysical Research 65*(12), 4185–4193.

Chayes, F. and J. Trochimczyk (1978). An effect of closure on the structure of principal components. *Mathematical Geology 10*(4), 323–333.

Egozcue, J., C. Barceló-Vidal, J. Martín-Fernández, E. Jarauta-Bragulat, J. L. Díaz-Barrero, and G. Mateu-Figueras (2011). *Elements of simplicial linear algebra and geometry*. In: Pawlowsky-Glahn, V. and Buccianti A. (eds.), Compositional Data Analysis: Theory and Applications, Wiley, Chichester UK.

Egozcue, J. and V. Pawlowsky-Glahn (2006). Simplicial geometry for compositional data. Volume 264 of *Special Publications*. Geological Society, London.

Egozcue, J. J. and V. Pawlowsky-Glahn (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology 37*(7), 795–828.

Egozcue, J. J. and V. Pawlowsky-Glahn (2011). *Basic concepts and procedures*. In: Pawlowsky-Glahn, V. and Buccianti A. (eds.), *Compositional Data Analysis: Theory and Applications*, Wiley, Chichester UK.

Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology 35*(3), 279–300.

Loosvelt, L., H. Vernieuwe, V. R. N. Pauwels, B. De Baets, and N. E. C. Verhoest (2012). Local sensitivity analysis for compositional data with application to soil texture in hydrologic

modelling. *Hydrol. Earth Syst. Sci. Discuss. 9*, 8841–8883. doi:10.5194/hessd-9-8841-2012.

Mateu-Figueras, G., V. Pawlowsky-Glahn, and J. J. Egozcue (2011). *The principle of working on coordinates*. In: Pawlowsky-Glahn, V. and Buccianti A. (eds.), Compositional Data Analysis: Theory and Applications, Wiley, Chichester UK.

Parent, L. E., C. X. de Almeida, A. Hernandes, J. J. Egozcue, C. Gülser, M. A. Bolinder, T. Kätterer, O. Andrén, S. E. Parent, F. Anctil, J. F. Centurion, and W. Natale (2012, June). Compositional analysis for an unbiased measure of soil aggregation. *Geoderma 179–180*, 123–131.

Pawlowsky-Glahn, V. and J. J. Egozcue (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment (SERRA) 15*(5), 384–398.

von Eynatten, H., V. Pawlowsky-Glahn, and J. J. Egozcue (2002). Understanding perturbation on the simplex: a simple method to better visualise and interpret compositional data in ternary diagrams. *Mathematical Geology 34*(3), 249–257.

Suggestions

1. Throughout the paper use *baricenter* in place of *barycenter*. We would recommend this substitution.

2. Powering in the simplex is not commutative, i.e. for a real number α and a composition $\mathbf{x}$, $\alpha \odot \mathbf{x}$ is not $\mathbf{x} \odot \alpha$. The standard form is $\alpha \odot \mathbf{x}$. However, this is just a convention, and the authors can use the reverse expression, but a definition is advisable. Note that in standard real operations this is not important as both α and $\mathbf{x}$ are real and the implied operation is the commutative real multiplication.

3. We do not like to call ξ *perturbation factor*. In fact, it appears in the form $(1 \pm \xi)$, which involves standard addition. We would feel more confortable if $(1 \pm \xi)$ were called powering factor or scaling factor (it is a factor in the Aitchison geometry as it is a powering-factor). It is not easy to give an appropriate name to ξ; perhaps something like *powering rate* thus taking the name of *rate* from

economic terminology: something that has decreased 3% means multiplication-perturbation with the coefficient 0.97.

4. In Figure 2, an ilr-transformation has been used. Although different ilr-transformations consist of a rotation of the figure, we guess that the used ilr-coordinates (balances) were
$$x_1^* = \frac{1}{\sqrt{2}} \log \frac{x_1}{x_2} \ , \ x_2^* = \sqrt{\frac{2}{3}} \log \frac{(x_1 x_2)^{1/2}}{x_3} \ .$$
They should be mentioned in the caption of Fig. 2 (or in the text).

4. In Eq. (1) summation subindexes do not match.

5. Before Eq. (6), notation y_t is used. However, t is only defined after the discussion of Eqs. (6) (7) and (8). A brief description of what is t is therefore convenient before Eq. (6).

6. In page 8856, lines 4-6 (before the algorithm listing) the sentence *Note that the presented methodology does not allow to calculate the sensitivity for the baricenter as the scalar multiplication $p_0 \odot (1 \pm \xi)$ has no effect on this composition* is wrong. The operation $p_0 \odot (1 \pm \xi)$ is not needed for computation of the sensitivity index, as p_0 is a center of a circle and what is powered by $1 \pm \xi$ is a (unitary) vector placed at p_0. We would recommend to delete this confusing statement.

7. Section 3.2.2, lines 10-13. The statement is unclear, due to the concept of correlation in this context. Which correlation is referred to? A better explanation is advisable.

8. First paragraph in section 3.2.3 is difficult to read. Please, use short sentences.

9. Section 3.2.3, lines 28-31. Please, rewrite avoiding repetition of *on the other hand* and *on the contrary*.

10. Section 4, page 8866, line 16. Separate words *verythe*.

11. In Reference Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barcelo-Vidal, (2003) an accent is missing, i.e. Barceló-Vidal.