**Appendix C**

Paper: “Exploring the effects of the correlation within and between data tables on Procrustes analysis output: interpretation for ecologists”

Francy J. G. Lisboa a,b\*, Ruth J. Mitchell b, Stephen Chapman b, Jacqueline M. Pottsc, Ricardo L. L. Berbaraa

a Soil Science Department, Agronomy Institute, Federal Rural University of Rio de Janeiro, Seropédica-RJ, 23890000, Brazil

b The James Hutton Institute, Craigiebuckler, Aberdeen, AB15 8QH United Kingdom.

c Biomathematics & Statistics Scotland, Craigiebuckler, Aberdeen, AB15 8QH

\* Corresponding author: Tel/Fax 55 21 26828045; email address: agrolisboa@gmail.com



**Figure 1**. Effects of increasing number of correlated variables (6, 9, 12, and 15) across different correlation levels (0.2, 0.5, 0.7, 0.9) within the X data table on Procrustes results when use both, **X** and **Y** data tables, **as ordination axes – based matrices**. **A)** Effect on Procrustes correlation statistic *R*; **B)** Effect on significance of Procrustean relationship (*P* value); **C)** Effect on residual size ranging within the vector of relationship (Procrustean association metric: PAM); **D)** Effect on ANOVA significance by using the PAMs as response and land use type (4 levels) as categorical predictor. The **X** (soil fertility) and the **Y** (PLFA profile) data tables are derived from Lisboa et al. (2014b). The correlation within **Y** data table was held fixed (original correlation structure). Means ± 1 SE of 8 pre-Procrustes transformations related to ordination axes – based matrices simulated 100 times are shown (Fig. 1).



**Figure 2**. Effects of increasing number of correlated variables (6, 9, 12, and 15) across different correlation levels (0.2, 0.5, 0.7, 0.9) within the **X** data table on Procrustes results when use both, **X** and **Y** data tables, **as dissimilarity/distance matrices**. **a)** Effect on Procrustes correlation statistic *R*; **b)** Effect on significance of Procrustean relationship (*P* value); **c)** Effect on residual size ranging within the vector of relationship (Procrustean association metric: PAM); **d)** Effect on ANOVA significance by using the PAMs as response and land use type (4 levels) as categorical predictor. The **X** (soil fertility) and the **Y** (PLFA profile) data tables are derived from Lisboa et al. (2014b). The correlation within **Y** data table was held fixed (original correlation structure). Means ± 1 SE of 4 pre-Procrustes transformations related to dissimilarity/distance matrices simulated 100 times are shown (Fig. 1 c).



**Figure 3.** Effect of increasing the number of variables/columns correlating in the **X** data table (soil fertility) on the ordination of Procrustes residual vectors (PAMs) retained from Procrustes relationships. The **X** (soil fertility, *n*=53, *p*=15) and the **Y** (PLFA profile, *n*=53, *p*=20) data tables are derived from Lisboa et al. (2014b). Four each correlation levels (0.9, 0.7, 0.5, and 0.2) the number of correlated columns increases from 6 to 15 (total soil fertility variables). The correlation structure within the **Y** data table was held fixed (original correlation structure). In **A**, PAMs are grouped according to the number of variables correlating in X (6, 9, 12, and 15) in a given correlation level (0.9, 0.7, 0.5, and 0.2). In **B**, the PAMs are grouped based on the correlation level in a given number of correlated columns in **X**. Each symbol represents the mean PAM from 100 simulations of each pre-Procrustes transformations described in (Fig. 1 c) of the main text.

**Literature cited**

Lisboa, F. J. G; Chaer, G. M; Fernandes, M. F; Berbara, R. L. L; Madari, B. E. 2014. The match between microbial community structure and soil properties is modulated by land use types and sample origin within an integrated agroecosystem. Soil Biology and Biochemistry 78: 97-108