Visualisation and modelling of soil data using the aqp package

Pierre Roudier[1,⋆], **Dylan Beaudette**[2]

1. Landcare Research, New Zealand
2. Natural Resources Conservation Service, USDA, USA
⋆Contact author: roudierp@landcareresearch.co.nz

Keywords: Soil science, visualisation, aggregation, classification

Soils support nearly all terrestrial food webs, are composed of a dynamic mixture of organic and mineral constituents, and occur within the relatively fragile intersection between atmosphere, biosphere, and lithosphere. With the growth of the human population and a limited soil resource to feed those populations, the study of such a complex environment arises as a prime challenge (Sanchez et al., 2009). However, the importance of soils has been acknowledged for a long time. An impressive quantity of soil information has been collected to support soil survey operations, natural resource inventories, and research over at least the last 100 years. Soils are routinely sampled and studied according to characteristic layers (called *horizons*), resulting in a complex data structure, with the dimensions of location, depth, and property space.

The high dimensionality and grouped nature of large soil profile collections can complicate standard analysis, summarization, and visualization, especially within a research community that is a traditional spreadsheet user. Soil data manipulation and analysis is also complicated by difficulties associated with processing horizon data that vary widely in depth and thickness. Scalability and reproducibility are also becoming significant issues as the size of the databases grow. Finally, while the investigation of soil profile characteristics and horizon-level morphology is strongly based on visual and tactile cues, the challenge of communicating these data is traditionally addressed using written narrative or tabular form.

R is a suitable platform to address those challenges, and to develop tools that would provide soil scientists with aggregation, modelling and visualisations for soil data. An *R* package, **aqp** (Algorithms for quantitative pedology), has been developed to extend *R*'s methods to the specificities of soil data. Specialized S4 classes have recently been added to support the multivariate hierarchy of linked spatial data (e.g. coordinates), site data (e.g. landscape position), and horizon data (e.g. clay content at 10 cm). Various new aggregation and classification methods are also available, and make use of the parallelised environment provided by the **plyr** package (Wickham, 2011). Examples of the **aqp** functionalities are proposed on different soil databases, for data visualisation, analysis and classification. The future developments of the package, especially its interactions with other packages used for soil data analysis (e.g. **sp** Pebesma and Bivand, 2005), is also discussed.

References

Pebesma, E. J. and R. S. Bivand (2005, November). Classes and methods for spatial data in R. *R News 5*(2), 9–13.

Sanchez, P. A., S. Ahamed, F. Carre, A. E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, A. B. McBratney, N. J. McKenzie, M. d. L. Mendonca-Santos, B. Minasny, L. Montanarella, P. Okoth, C. A. Palm, J. D. Sachs, K. D. Shepherd, T.-G. Vagen, B. Vanlauwe, M. G. Walsh, L. A. Winowiecki, and G.-L. Zhang (2009). Digital Soil Map of the World. *Science 325*(5941), 680–681.

Wickham, H. (2011). *plyr: Tools for splitting, applying and combining data*. R package version 1.4.