Geophysical Research Abstracts
Vol. 17, EGU2015-6964-1, 2015
EGU General Assembly 2015

Soil biogeochemistry in the age of big data

Lauric Cécillon (1), Pierre Barré (2), Eric Coissac (3), Alain Plante (4), and Daniel Rasse (5)

(1) Irstea, UR EMGR Ecosystèmes Montagnards, Université de Grenoble, France (lauric.cecillon@irstea.fr), (2) Laboratoire de Géologie de l'ENS, PSL Research University, UMR8538 CNRS, Paris, France (barre@geologie.ens.fr), (3) Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université de Grenoble, France (eric.coissac@ujf-grenoble.fr), (4) Dept of Earth & Environmental Science, University of Pennsylvania, Philadelphia, USA (aplante@sas.upenn.edu), (5) Bioforsk, Norwegian Institute for Agricultural and Environmental Research, Ås, Norway (daniel.rasse@bioforsk.no)

Data is becoming one of the key resource of the XXIst century. Soil biogeochemistry is not spared by this new movement. The conservation of soils and their services recently came into the political agenda. However, clear knowledge on the links between soil characteristics and the various processes ensuring the provision of soil services is rare at the molecular or the plot scale, and does not exist at the landscape scale. This split between society's expectations on its natural capital, and scientific knowledge on the most complex material on earth has lead to an increasing number of studies on soils, using an increasing number of techniques of increasing complexity, with an increasing spatial and temporal coverage. From data scarcity with a basic data management system, soil biogeochemistry is now facing a proliferation of data, with few quality controls from data collection to publication and few skills to deal with them.

Based on this observation, here we (1) address how big data could help in making sense of all these soil biogeochemical data, (2) point out several shortcomings of big data that most biogeochemists will experience in their future career.

Massive storage of data is now common and recent opportunities for cloud storage enables data sharing among researchers all over the world. The need for integrative and collaborative computational databases in soil biogeochemistry is emerging through pioneering initiatives in this direction (molTERdb; earthcube), following soil microbiologists (GenBank). We expect that a series of data storage and management systems will rapidly revolutionize the way of accessing raw biogeochemical data, published or not. Data mining techniques combined with cluster or cloud computing hold significant promises for facilitating the use of complex analytical methods, and for revealing new insights previously hidden in complex data on soil mineralogy, organic matter and biodiversity. Indeed, important scientific advances have already been made thanks to meta-analysis, chemometrics, machine-learning systems and bioinformatics. Some techniques like structural equation modeling eventually propose to explore causalities opening a way towards the mechanistic understanding of soil big data rather than simple correlations. We claim that data science should be fully integrated into soil biogeochemists basic education schemes. We expect the blooming of a new generation of soil biogeochemists highly skilled in manipulating big data.

Will big data represent a net gain for soil biogeochemistry? Increasing the amount of data will increase associated biases that may further be exacerbated by the increasing distance between data manipulators, soil sampling and data acquisition. Integrating data science into soil biogeochemistry should thus not be done at the expenses of pedology and metrology. We further expect that the more data, the more spurious correlations will appear leading to possible misinterpretation of data. Finally, big data on soils characteristics and processes will always need to be confronted to biogeochemical theories and socio-economic knowledge to be useful.

Big data could revolutionize soil biogeochemistry, fostering new scientific and business models around the conservation of the soil natural capital, but our community should go into this new era with clear-sightedness and discernment.