Meeting: XIV MEDECOS & AEET meeting, Sevilla, January-February 2017.
Symposia: 21. Ecoinformatics: data science brings new avenues for ecology,
http://www.medecos-aeet-meeting2017.es/Symposia_396_p.htm
Requested type of contribution: oral.

Compiling a global database of sap flow measurements: the SAPFLUXNET data workflow

V. Granda, R. Poyatos, R. Molowny-Horas, M. Mencuccini, K. Steppe, J. Martínez-Vilalta

Widespread application of thermometric sap flow methods in ecological, hydrological and agronomic studies since the 1990s has generated a massive amount of quasi-continuous (i.e., from sub-daily to interannual) data of whole-plant sap flow. A global compilation of these datasets would enable unprecedented analyses of the environmental and ecological drivers of plant transpiration. However, sap flow datasets can be highly heterogeneous because of differences in methodologies and experimental designs across individual studies. Here we describe the logic behind the data ingestion process implemented in SAPFLUXNET, the first global database of sap flow measurements. This database is being compiled from the contributions by individual researchers across the globe consisting of metadata at various organisational levels (i.e. site, stand, plant...) and sap flow and environmental time series data. The actively developed and highly modular R package, '*sapfluxnetr*' performs semi-automatized data processing, standardisation and quality control (i.e., timestamps, units, geographic and species nomenclature checks). Apps created using Rmarkdown and Shiny/HTML generate interactive reports and documents to perform visual checks and to request feedback from data contributors. All the steps in our data workflow are fully documented, reproducible and version-controlled in a GitHub repository, allowing that analogous workflows ensuring quality control can be applied in similar ecological databases. Our approach will also lead to a seamless integration of the datasets in a relational database (PostgreSQL) with a user-friendly data interface (R/Shiny/HTML) to efficiently perform data queries. Overall, SAPFLUXNET constitutes a major step towards openness and reproducibility in plant physiological ecology and ecosystem ecology.