



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

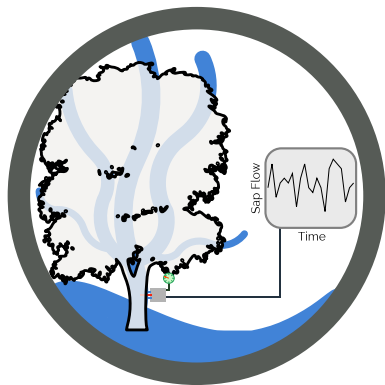
XIV MEDECOS & XIII AEET meeting

Ecoinformatics: data science brings new avenues for ecology
Symposium

Victor Granda, Rafael Poyatos, Roberto Molowny-Horas, Maurizio Mencuccini,
Kathy Steppe & Jordi Martínez-Vilalta



Centre of Ecological Research and Forestry Applications



Different **thermodynamic methods** to determine sap flow using heat as a tracer sap movement

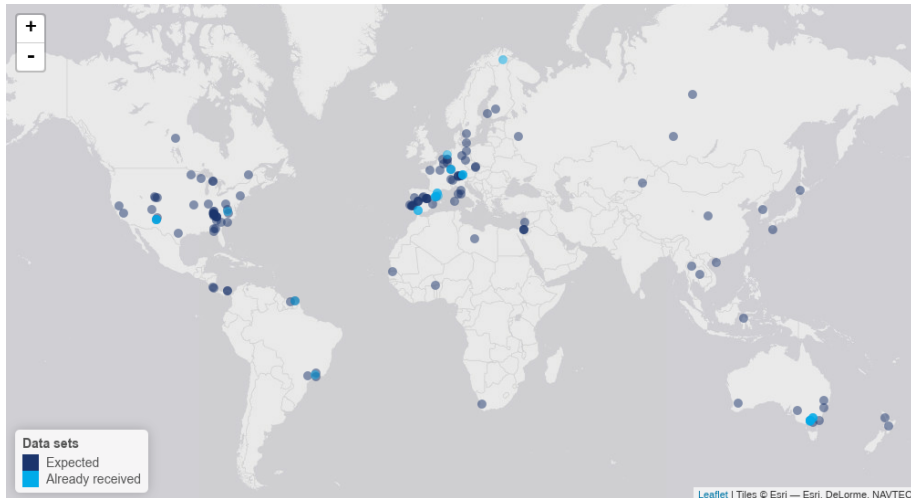
Proxy of the movement of water between the soil-plant-atmosphere continuum.

Allows **upscaling** from stem to plant and landscape level.

The time is ripe for a global database



The **SAPFLUXNET** initiative is building the first global database of plant-level sap flow measurements to analyse the environmental and physiological factors driving tree- and stand-level transpiration

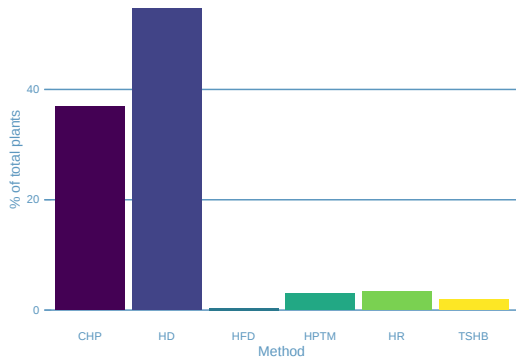




- ▶ Stem or whole-plant level
- ▶ Field conditions
- ▶ Sub-daily intervals
- ▶ Environmental data available (RH, Ta, PAR...)
- ▶ Abundant metadata (site, stand, plant, species and environmental)



High data complexity:



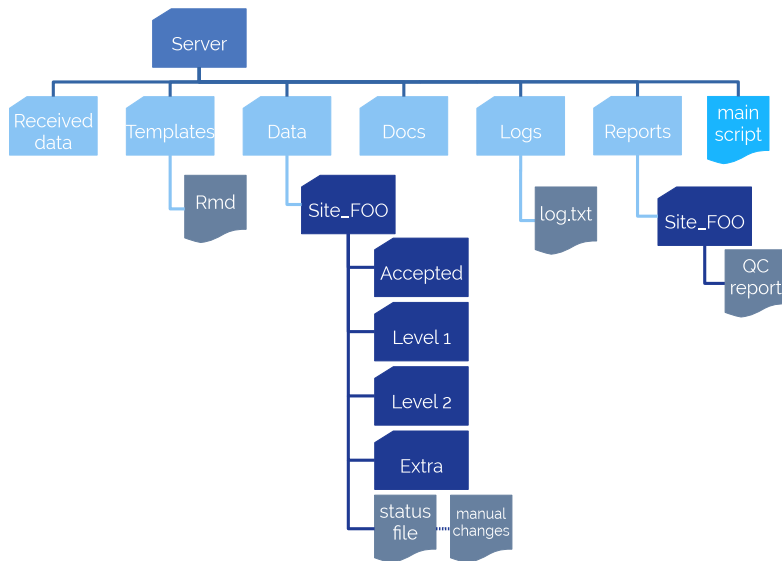
| Metadata | Items |
|---------------|-----------|
| Site | 20 |
| Stand | 16 |
| Species | 4 |
| Plant | 24 |
| Environmental | 16 |
| Total | 80 |

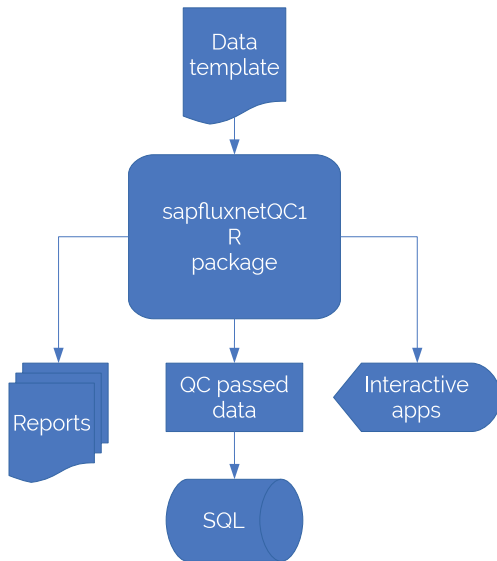


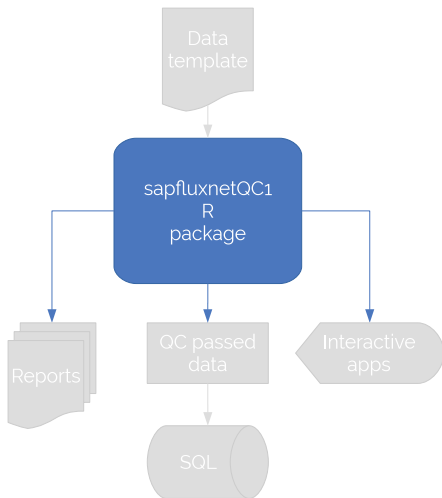
How to assay quality and store data?

We need **semi-automatic**, **reproducible** and **robust** checks to ensure the quality of the submitted datasets. Also, we need to store the data in a way that allows all essential information to be **available** in order to use the data in checks and analyses

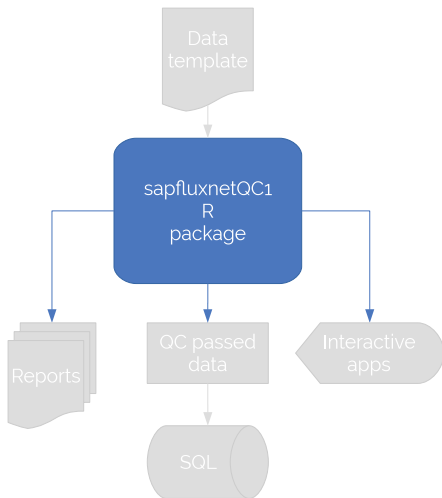
SAPFLUXNET Server Structure







- ▶ Automatic Quality Control checks (QC)
- ▶ Automatic report generation
- ▶ Storing data in special object (SfnData S4 class)
- ▶ Interactive functions allowing fine control of QC



Benefits of R as development environment

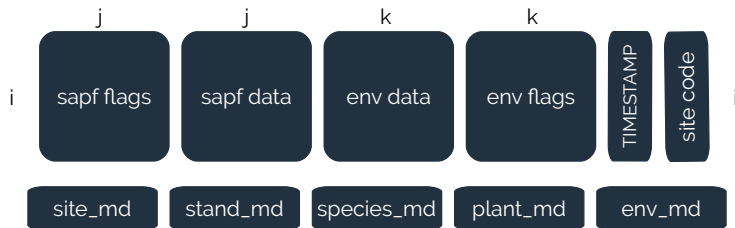
- ▶ Open
- ▶ Reproducible
- ▶ Easy maintenance and update
- ▶ Easy integration with web and SQL technologies

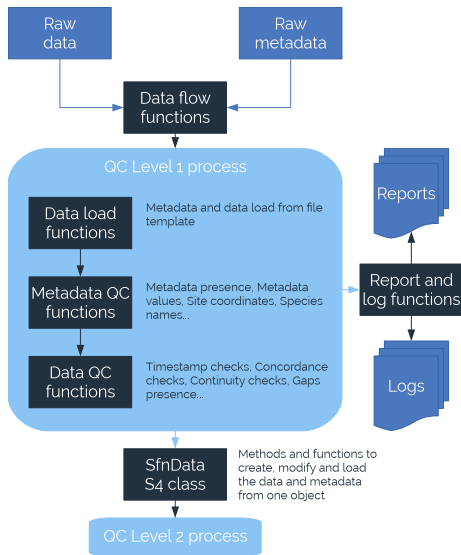


S4 classes:

- ▶ Easy way of store complete site data
- ▶ Validity checks based in fair assumptions
- ▶ Methods: `[]`, `get`, `<-`
- ▶ Scalable: allows for combining sites for more complex analyses (**whishlist**)

SfnData class:

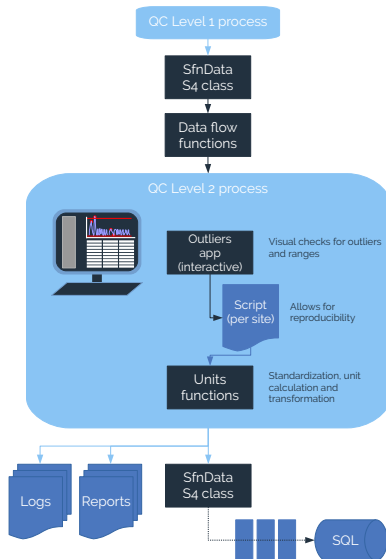




QC Level 1

General metadata and data quality checks:

- ▶ Presence/Absence of metadata and data variables
- ▶ Metadata values correctness (i.e coordinates, species names...)
- ▶ Sapflow and environmental data correctness (format, timezone, gaps, continuity...)
- ▶ Uniformization and unit transformations of data (solar time, sapflow at different levels...)



QC Level 2

Specific Data quality checks:

- ▶ Robust outlier detection
- ▶ Range checks
- ▶ Unit standardization and transformation



How to track site status?

TO DO



1. Status files
2. Apps