The hidden role of gases in trees

A Data Management Plan created using DMPTuuli

Creators: Stephen Ingram, Timo Vesala

Affiliation: University of Helsinki

Funder: Academy of Finland

Template: Academy of Finland data management plan guidelines

Project abstract:

Plants transport water continuously through vessels in their stems and branches without the aid of a mechanical pump, such as the heart. Instead, they achieve this by the transpiration of water vapour, which creates a pressure differential between the leaves and roots. Between lies the xylem tissue, which is specialized for long distance water transport under tension. However, the negative pressures induced by tension lead to embolism in the xylem, wherein whole vessels become air filled and can no longer transport water. It is known that so called ‘air seeding’ is the main cause of embolism. Additionally, it has recently been discovered that air seeding may also introduce small pockets of gas, confined to nanobubbles, into the sap, that remain stable for surprisingly long periods of time. Could it be possible that such bubbles in fact play a crucial role in the movement of gas and water within plants? In this project, we develop both a detailed tissue level network model and a more coarse whole tree level model, in which the physics of transport of all the gases are described, including advection and diffusion, and network effects. Xylem sap contains a multitude of different gases such as nitrogen, oxygen, CO2, and methane, dissolved in it. In addition, xylem sap contains volatile organic compounds (VOCs) such as monoterpenes, acetaldehyde and methanol, which have high vapor pressure and are thus present in the gas phase e.g. in bubbles. Each gas will be described uniquely, incorporating values of solubility and diffusion coefficients from the literature. Gas compounds with high solubility to water may travel a long way from their site of production axially with the xylem sap by advection. Therefore, the production and consumption of different gases will also be included in the models. X-ray tomography will be used to obtain the structure and connectivity of xylem conduits, which will be the basis of the tissue level network model. The concentrations and stem effluxes of different gases will also be measured to test and parameterize both models. Concurrently, molecular simulations will attempt to unravel the stability of nanobubbles, and their dynamic behaviour under negative pressure. These simulations will communicate with the tissue and whole tree level models. As a whole, the project will provide holistic understanding on the role of gases in tree physiological processes.

ID: 16487

Start date: 01-09-2021

End date: 31-08-2025

Last modified: 05-07-2021

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

The hidden role of gases in trees

1. General description of data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.

WP1:
Data Collected

- Data on concentration and fluxes of different gases from SMEAR II and laboratory experiments. Small amounts of data manually collected from laboratory and field measurements. Data ~10 Gb.

Data Produced

- Data on model simulations. Data ~10 Gb.

Data Reused

- Continuous data on sap flow and gas exchange at the SMEAR II station. Data ~10 Gb.

WP2:
Data Collected

- Simulation trajectories of bubble dissolution: 1-2Tb
- Simulation trajectories of ice nucleation in xylem conduits: 2-3Tb

Data Produced

- Videos of simulations: ~10Gb

Data Reused

- Simulation trajectories of lipids under negative pressure: ~2Tb
- Python scripts and output files: 100Mb

1.2 How will the consistency and quality of data be controlled?

Where files are large enough that errors in storage or download could be possible, checksums will be used to ensure integrity.

2. Ethical and legal compliance

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

GDPR will have no effect on any of our data as none of it will pertain to citizens or members of the public. Nor will we be conducting experiments on any endangered plant species.

2.2 How will you manage the rights of the data you use, produce and share?

All consortium members will sign the Research Cooperation Agreement before the project begins. Additionally, TV, SI and TH will sign the University of Helsinki "Undertaking on Transfer of Rights".

Any datasets, tables, figures, audio and video recordings produced will be available to anyone who enquires under a Creative Commons No Rights Reserved (CC0) license.

Software written for each of the Work Packages will be released under an LGPL license, or other similar copyleft/permissive license.

3. Documentation and metadata

3.1 How will you document your data in order to make the data findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?

Data produced for WP2 will be stored in the Allas object storage service provided by CSC. The directory structure and other conventions will be documented and those documents shared with the other consortium participants.

The whole tree model to be written for WP1,and the network model for WP3, will be subject to version control software (eg. git,github) such that a precise record will be kept of the changes, who made them and when.

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will the data be backed up?

As mentioned in section 1, the bulk of the molecular dynamics simulation data will be stored on the Allas object storage service provided by the CSC.

Project important documents and correspondence will be stored in Onedrive, provided by UH, or shared directly from hardware to hardware using the open source Syncthing software.

WP1 and WP3 data can be distributed amongst the hardware of the consortium members using syncthing.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

Within the University of Eastern Finland, Annamari Lauren will be responsible for the majority of data from work packages 1 and 3, as they are the only member of the consortium in that institution. Stephen Ingram and Timo Vesala will be jointly responsible for data stored in the University of Helsinki.

Sharing between the two institutions will be achieved either by password protected cloud computing services, or Syncthing as mentioned above. Syncthing is cryptographically secure as it generates a unique key to each piece of hardware data is stored on, and only shares data between authorised PCs. Therefore, it is impossible for anyone not given permission to intercept or otherwise download information.

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?

Code related to MD simulations or the the whole tree model will be available publicly on a version control system such as github. A small amount of "test" input data from our measurements will be provided to allow anyone who downloads the model to run it.

Once they have been suitably cleaned/commented, 'Medium' sized files/data (100Mb - 1Gb per file) will be stored on figshare or a similar site allowing generation of DOIs.

See the next section for longer term plans.

5.2 Where will data with long-term value be archived, and for how long?

We will apply to one of the many data repositories with persistent identifiers listed on the registry of research repositories, for example the Fairdata IDA Research Data Storage Service or the Materials Cloud Archive. We anticipate at least five years of storage.

6. Data management responsibilities and resources

6.1 Who (for example role, position, and institution) will be responsible for data management (i.e., the data steward)?

The main investigator for each work package will be responsible for the appropriate data (Annamari Lauren for WP1 and WP3, Stephen Ingram for WP2)

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

If no long term storage solution can be found (see section 5.2), several external and internal harddrives will need to be purchased for local storage of model data/backups. However, we predict that the majority will be stored in services already accessible to UH staff.

Data will need to be cleaned and the model code adequately commented etc. before upload.

No extra individuals will need to be employed for this purpose, it can be carried out by the consortium members in addition to their other duties. We predict that the majority of this work will be conducted at the beginning of the consortium, and then documentation and backup procedures will be updated throughout the next 4 years.