

EGU2020-9228

<https://doi.org/10.5194/egusphere-egu2020-9228>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



## A repeatable and reproducible modelling workflow using the Vegetation Optimality Model and RENKU

**Remko Nijzink**<sup>1</sup>, Chandrasekhar Ramakrishnan<sup>2</sup>, Rok Roskar<sup>2</sup>, and Stan Schymanski<sup>1</sup>

<sup>1</sup>Luxembourg Institute of Science and Technology, Environmental Research and Innovation, Catchment and Eco-hydrology Research Group, Belvaux, Luxembourg (remko.nijzink@list.lu)

<sup>2</sup>Swiss Data Science Center, Zurich, Switzerland

Numerical experiments become more and more complex, resulting in workflows that are hard to repeat or reproduce. Even though many journals and funding agencies now require open access to data and model code, the linkages between these elements are often still poorly documented or even completely missing. The software platform Renku (<https://renkulab.io/>), developed by the Swiss Data Science Center, aims at improving reproducibility and repeatability of the entire scientific workflow. Data, scripts and code are stored in an online repository, and Renku records explicitly all the steps from data import to the generation of final plots, in the form of a knowledge graph. In this way, all output files have a history attached, including linkages to scripts and input files used generate them. Renku can visualize the knowledge graph, to show all scientific links between inputs, outputs, scripts and models. It enables easy re-use and reproduction of the entire workflow or parts thereof.

In the test case presented here, the Vegetation Optimality Model (VOM, Schymanski et al., 2009) is applied along six study sites of the North-Australian Tropical Transect to simulate observed canopy-atmosphere exchange of water and carbon dioxide. The VOM optimizes vegetation properties, such as rooting depths and canopy properties, in order to maximize the Net Carbon Profit, i.e. the total carbon taken up by photosynthesis minus all the carbon costs of the plant organs involved. The vegetation is schematized as one big leaf for trees and one leaf for seasonal grasses, and is combined with a water balance model. Flux tower measurements of evaporation and CO<sub>2</sub>-assimilation, and remotely sensed vegetation cover are used for model evaluation, in addition to meteorological data as input for the model. A numerical optimization, the Shuffled Complex Evolution, is used to optimize the vegetation properties for each individual site by repeatedly running the model with different parametrizations and computing the net carbon profit over 20 years. The optimization was repeated several times for each site to analyze the sensitivity of the results to a range of different input parameters.

This case demonstrates an example of a complex numerical experiment with all its associated challenges concerning documenting model choices, large datasets and a variety of pre- and post-processing steps. Renku assured the repeatability and reproducibility of this experiment, by documenting this in a proper and systematic way. We demonstrate how Renku helped us to

repeat analyses and update results, and we will present the knowledge graph of this experiment.

### **References**

Schymanski, S.J., Sivapalan, M., Roderick, M.L., Hutley, L.B., Beringer, J., 2009. An optimality-based model of the dynamic feedbacks between natural vegetation and the water balance. *Water Resources Research* 45. <https://doi.org/10.1029/2008WR006841>