

HUTTON OPEN SCIENCE AWARDS

Best Hutton PhD Student Open Science Action

Using Openly Available R/RStudio Environment and Github Repository for Transparent Tracking and Recording of Bayesian Belief Networks (BBN) Modelling Decisions

Camilla Negri (EBS)

Brief description of the project (aims, objectives etc)

This project applies Bayesian Belief Networks (BBNs), which are probabilistic graphical models, to develop a decision support tool for management of phosphorus pollution in Irish agricultural river catchments. The model development follows two approaches: the use of a high-resolution water quality dataset and the use of expert elicitation where high-resolution data is lacking. Model transferability is being tested by parameterizing the model in four additional catchments with contrasting hydrology and land uses. IPCC climate change scenarios will be applied to simulate the pollution risk and inform management decisions under future change.

How does this research go beyond previous open science efforts by the participating individual (student) or team?

This research uses openly available R/RStudio environment and Github repository that allow transparent tracking and recording of modelling decisions. The model structure and specification will be made openly available to allow reproduction and model application in different contexts, with several software available (some freely) to replicate this work. Negotiations with the funder are in progress regarding data publishing. These key open science practices are not routinely part of PhD training and have required a significant learning investment by the candidate and the supervisory team. However, Camilla is now an expert and uses her knowledge to support other students and staff.

How does this research progress beyond standard expectations of open science in the respective disciplines/areas of research?

Hydrology is facing reproducibility crisis, with 1.6% of hydrology papers found to be reproducible (Stagge et al., 2019). The use of free open-source software, version control, readable coding practices (Rmarkdown) and the explicit intent to publish openly, including all code, models and data is unusual in the field and directly improves reproducibility in hydrology, challenging other researchers to adopt open-science practices. The graphical nature of BBNs is ideal for model co-development with stakeholders (farm advisers), enhancing model credibility. Combined with the open-source tools, BBNs help to bridge the gap between modelling efforts and reproducibility and effectively communicate decisions with stakeholders.

Stagge, J.H., Rosenberg, D.E., Abdallah, A.M., Akbar, H., Attallah, N.A., James, R., 2019. Assessing data availability and research reproducibility in hydrology and water resources. *Sci. Data* 6, 190030. <https://doi.org/10.1038/sdata.2019.30>

What lessons about the processes of opening up science have been learned?

1. Using an open participatory modelling framework has fostered knowledge exchange between researchers and advisers, helping to build trust and further process understanding.
2. Learning to use (and troubleshoot) open science tools has allowed Camilla to develop structured working practices, which facilitate efficient paper/report/thesis writing and publishing.
3. The acquired expertise allowed Camilla to support students and other staff to further the use of open science tools within the Institute. Some caveats with the tools have been raised regarding their longer-term maintenance, production of metadata and ancillary information; however, these are also opportunities to teamwork and facilitate handover.