Proposal for workshops – AGILE 2024

Glasgow, Tuesday 4 June 2024

https://agile-online.org/conference-2024

1. Workshop name/title (and acronym is applicable)

Spatial modelling and interpretability with Random Forest?

2. Description of the workshop by listing topic(s), objective(s) and planned outcome(s)

2.1 Topic(s)

Geospatial machine learning, Python, interpretability and explainability of machine learning models

2.2 Objective(s)

To provide an overview on how to perform spatial modelling with random forest and what interpretation methods to use to explain and improve the models.

2.3 Planned outcome(s)

The participants will learn a set of relevant methods and *Python* packages to develop, run, and interpret a decision-tree-based (RF) machine-learning model to predict spatial variables. The hands-on tutorial will be accompanied by short lecture sections to introduce and elaborate on the required background. The tutorial including scripts and Jupyter notebooks will be made accessible as reproducible package via a public GitHub repository and deposited as citable resource on Zenodo.

3. Abstract

Machine learning has been increasingly used due to its capabilities to work with large amounts of data, while having minimal assumptions on shape or distribution of variables. However, machine-learning models are often considered a black-box, and a lack of interpretability would mean that it is hard to determine if the model has found meaningful and realistic relationships between different phenomena. Thus, there is a strong need to be

able to explain and interpret machine-learning models to understand the effects and relationships of the underlying modelled processes and used covariates.

Recently, various methods to interpret the relationships between the covariates and the target variable in machine learning have been introduced. In this workshop, we will present such methods, including partial dependence plots and Shapley values, and provide tips on how to make use of these methods to build less biased, better understandable, and more robust models. The workshop will provide hands-on tasks on interpreting a machine learning model's results using the *Python* programming language.

This is how you want your workshop to be presented on the conference webpages (maximum 1000 characters)

4. Short description of the intended length (half or full day) and the format of the workshop

2h hands-on workshop in Python. The lecturers will guide the participants through ML documented and reproducible Jupyter notebooks based on the *literate programming paradigm* that will be complemented with mini lectures about the interpretability methods and things to consider when performing ML on spatial data.

5. Brief statement of the relevance of the workshop for AGILE

The workshop will give a hands-on opportunity to learn the basics of spatial machine learning using the random forest (RF) method with a focus on the interpretability of the model. The workshop is useful for all PhD students as well as researchers, who either have minimal experience with linear, statistical, or ML-based modelling or want to learn how to improve their models through interpretability methods. Understanding more deeply how to make ML models interpretable enables researchers to build better and more meaningful models.

Machine learning is increasingly used in many domains to model various spatial environmental and urban processes. While building machine learning models through widely available packages in the *Python* or *R* programming languages is nowadays easy to accomplish, making a robust and meaningful model and explaining its predictive behaviour is still challenging – it requires domain knowledge and a deeper understanding of machine learning and modeling concepts in general. AGILE is a largely academic community, and we should employ appropriate scientific rigor in ML usage as well. In this workshop, we address the interpretation of machine learning models and how to produce more st and reliable models.

6. What is the approximate number of expected participants?

25

7. Names and e-mail addresses of the organizing member(s)

Leading AGILE member (or sponsor) and contact person: Landscape Geoinformatics Lab, University of Tartu, Estonia; Contact: Holger Virro

Contributing AGILE members (including the persons involved) – at least one seconding AGILE member is needed: Marta Jemeljanova, Evelyn Uuemaa, Alexander Kmoch

Seconding AGILE member: Prof. Carsten Keßler, Bochum University of Applied Sciences

Contributing non-AGILE members (including the persons involved) – if applicable: N/A

Organizing Committee (if applicable): N/A

Programme Committee (if applicable): N/A

8. Additional information about previous workshops, if held.

Previous workshop at AGILE 2021 (virtual):
Working with Discrete Global Grid Systems (half-day), Alexander Kmoch,
(https://sisu.ut.ee/sites/default/files/landscape-geoinformatics/files/agile2021_dggs_workshop_schedule.pdf)

Previous workshops at EGU 2021:
1. Satellite image processing using Python programming
2. Spatio-temporal trend analysis of spatial climate data (temperature and rainfall) using Python
GitHub link to EGU 2021 workshops:
https://github.com/LandscapeGeoinformatics/EGU_2021_lgeo_workshops

Previous workshop at BIOGEOMON 2022:
Spatio-temporal trend analysis of spatial climate data (temperature and rainfall) using Python
GitHub link to BIOGEOMON workshop:
https://github.com/LandscapeGeoinformatics/biogeomon_2022_pangeo

9. Expected resources needed

A room with an internet connection (preferably WiFi), a large presentation screen or a projector; some accessible power sockets and distribution panels for participants and presenter; participants need to bring their own laptops; installation instructions will be sent beforehand.

10. Other information

Submission by e-mail to:

Dr Qunshan Zhao: Qunshan.Zhao@glasgow.ac.uk

Or Dr John Xiaogang Shi: John.Shi@glasgow.ac.uk