The Journal of Open Source Education

Self-Guided Decision Support Groundwater Modelling with Python

Rui T. Hugman [1], **Jeremy T. White** [1], **Michael N. Fienen** [2], **Brioch Hemmings** [3], and **Katherine H. Markovich** [1]

1 INTERA Geosciences, Perth, Western Australia, Australia **2** U.S. Geological Survey, Upper Midwest Water Science Center, Madison, WI USA **3** Wairakei Research Centre, GNS Science, Taupō, New Zealand

DOI: 10.21105/jose.00240

Software

- Review
- Repository
- Archive

Submitted: 23 October 2023
Published: 24 December 2024

Summary

The GMDSI tutorial notebooks repository provides learners with a comprehensive set of tutorials for self-guided training on decision-support groundwater modelling using Python-based tools. Although targeted at groundwater modelling, the tutorials are based around model-agnostic tools and readily transferable to other environmental modelling workflows. The tutorials are divided into three parts. The first covers fundamental theoretical concepts. These are intended as background reading for reference on an as-needed basis. Tutorials in the second part introduce learners to some of the core concepts of parameter estimation in a groundwater modelling context, as well as provide a gentle introduction to the `PEST`, `PEST++` and `pyemu` software. Lastly, the third part demonstrates how to implement highly parameterized applied decision-support modelling workflows. The tutorials aim to provide examples of both "how to use" the software as well as "how to think" about using the software. A key advantage to using notebooks in this context is that the workflows described run the same code as practitioners would run on a large-scale real-world application. Using a small synthetic model facilitates rapid progression through the workflow.

Story of the Project

The Groundwater Modelling Decision Support Initiative (GMDSI) is an industry-backed and industry-aligned initiative. Established in mid-2019, its primary goal is to enhance the role of groundwater modelling in groundwater management, regulatory processes, and decision-making. GMDSI promotes the improved use of modelling in decision support, with activities focused on industry engagement, education, practical examples, research, and software development. It also emphasizes the importance of tools for uncertainty quantification (UQ) and parameter estimation (PE) in these processes.

The roots of the materials making up the tutorial notebooks were from a traditional, week-long classroom course curriculum developed for internal training at the US Geological Survey (USGS) by a subset of the authors of this paper. After three iterations of teaching the in-person class, the authors, with support from the GMDSI, endeavored to build on the positive aspects of using Jupyter Notebooks and explore alternative teaching environments. The first major change was to add sufficient narration and explanation to the notebooks to improve possibilities for self-study. The next change was to reorganize the content from a strictly linear progression to a three-part structure. This led to a hybrid model of self-study accompanied by discussion and background lectures online.

The Journal of Open Source Education

Statement of Need

Many groundwater modelers typically rely on Graphical User Interfaces (GUIs) for their modelling needs. However, each GUI has its unique characteristics and varying degrees of compatibility with external software like PEST (Doherty, 2015) and PEST++ (Jeremy T. White et al., 2020). Creating educational materials for these GUIs would necessitate tailoring content to each GUI's specific features, obtaining cooperation from the GUI developers themselves and potentially lagging behind the latest developments. Many GUIs are commercial products as well which limits accessibility.

Decision-support modelling often demands capabilities that surpass what current GUIs can offer. Thus, the use of Python for environmental modelling has increased in recent years, due to its open-source nature, user-friendly syntax, and extensive scientific libraries. Python-based tools have been developed to facilitate UQ and PE analyses, such as pyemu (Jeremy T. White et al., 2016b; Jeremy T. White et al., 2021). pyemu is a Python package that provides a framework for implementing UQ and PE analyses with PEST and PEST++. It offers a range of capabilities, including parameter estimation, uncertainty analysis, and management optimization. Although initially designed for groundwater modelling, pyemu's methodologies are versatile and can be applied to diverse numerical environmental models, Anecdotally, we have seen that more modelers are turning to Python packages like FloPy (Bakker et al., 2016) and pyemu (Jeremy T. White et al., 2016a) for model and PEST++ setup. Unfortunately, the adoption of this approach is hindered by a steep learning curve primarily due to the scarcity of user-friendly training materials.

The GMDSI tutorial notebooks aim to address this gap by providing a comprehensive, self-guided, and open-source resource for learning decision-support modelling workflows with Python. They are designed to be accessible to a broad audience, including students, researchers, and practitioners who aim to undertake applied environmental decision-support modelling.

Contents and Instructional Design

The tutorial notebooks are structured into three main parts:

Part 0: Introductory Background

Part 0 serves as the foundation, providing essential background material. Each notebook in Part 0 is standalone and covers a unique topic. These include:

- Introduction to a synthetic model known as the "Freyberg" model (Freyberg, 1988). This model is used as a consistent example throughout the tutorial exercises, allowing learners to apply concepts in a practical context.
- An introduction to the `pyemu` Python package that is used to complement and interface with PEST/PEST++.
- Explanation of fundamental mathematical concepts that are relevant and will be encountered throughout the tutorial notebooks.

Pre-requisites for Part 0 include a basic understanding of Python, Jupyter Notebooks, and MODFLOW 6 (Langevin et al., 2022). Familiarity with git is a bonus but not fundamental.

Part 1: Introduction to `PEST` and the Gauss-Levenberg Marquardt Approach

Part 1 focuses on the Gauss-Levenberg Marquardt (GLM) approach to parameter estimation and associated uncertainty analysis in a groundwater modelling context.

Part 1 is designed to be accessible without strict sequential dependencies. Learners have the flexibility to explore its contents in any order that suits their preferences or needs. These include:

- Introduction to concepts such as non-uniqueness, identifiability, and equifinality.
- Introduction to the `PEST` control file and the PEST/PEST++ interface.
- Exploring the challenges of parameterization schemes on predictive ability, as well as how to mitigate them.
- Introducing first-order second-moment (FOSM) and prior Monte Carlo uncertainty analysis approaches.

Pre-requisites for Part 1 include a basic understanding of numerical groundwater modelling and familiarity with MODFLOW 6. Familiarity with Python and Jupyter Notebooks is assumed.

Part 2: Python-based Decision-Support Modelling Workflows

Part 2 expands on the foundational knowledge gained in Part 1 and delves into advanced topics related to ensemble-based parameter estimation, uncertainty analysis and optimization methods. These advanced topics include management optimization and sequential data assimilation and assume a highly parameterized approach, as motivated in Part 1. Topics are laid out in manner that reflects real-world workflows, with a focus on practical application of concepts and problem solving.

Learners have the option to explore various sequences, in line with real-world applied workflows, such as:

- Prior Monte Carlo analysis
- Highly parameterized Gauss-Levenberg Marquardt history matching and associated Data Worth analysis using First Order, Second Moment (FOSM) technique,
- Ensemble-based history matching and uncertainty analysis with the iterative ensemble smoother approach as implemented in `PEST++IES`,
- Sequential data assimilation with `PEST++DA`, and
- Single-objective and multi-objective optimization under uncertainty with `PEST++OPT` and `PEST++MOU`.

Each of these sequences comprises multiple notebooks to be executed in a specified order. They demonstrate how to execute the workflow, interpret results, and apply the concepts to real-world problems.

The flowchart below gives an example of a curated learning flow for a common decision support modelling application. Over time, referring back through Part 1 will provide a deeper understanding of some concepts and techniques taken for granted in the highly parameterized, largely ensemble-based approaches of Part 2.

The Journal of Open Source Education

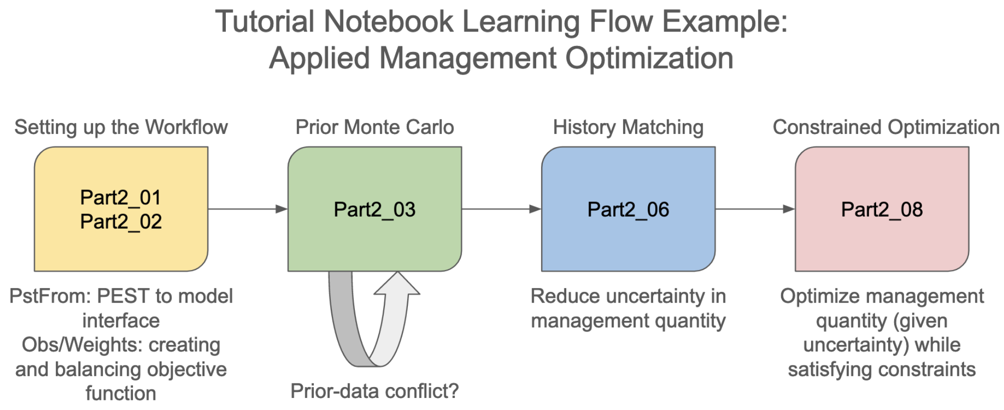

Figure 1: Example notebook learning flow demonstrating a comprehensive workflow for an applied, ensemble-based management optimization .

Pre-requisites for Part 2 include a basic understanding of PEST/PEST++ and the PEST interface, as well as familiarity with the Freyberg model. Familiarity with Python and Jupyter Notebooks is assumed.

Experience of use in teaching and learning situations

The notebooks were employed during the Applied Decision Support Groundwater modelling With Python: A Guided Self-Study Course hosted by GMDSI. This self-guided course comprised 5 online sessions, each lasting 1 to 2 hours and focused on the workflows of Part 2. During each session the instructors go through a section of the tutorials and expand on some of the concepts. Sessions were recorded and can be accessed on the GMDSI YouTube channel. Beyond the live online sessions, learners were encouraged to make use of the GitHub Discussions feature to retain a search-engine findable record of common questions that persist beyond the time frame of the course .

Feedback from the 65 students who participated in the course was anecdotal but informative. Figure 2 summarizes the responses by 34 respondents to four questions, comprising 52%. The majority of respondents indicated a preference for this hybrid self-guided/online instruction approach over an in-person week-long intensive class.

Open-ended feedback from the participants was generally positive and also included some constructive criticism. Participants appreciated the opportunity to ask questions and several reported hearing the discussion around other peoples' questions as being valuable and clarifying aspects of the material.

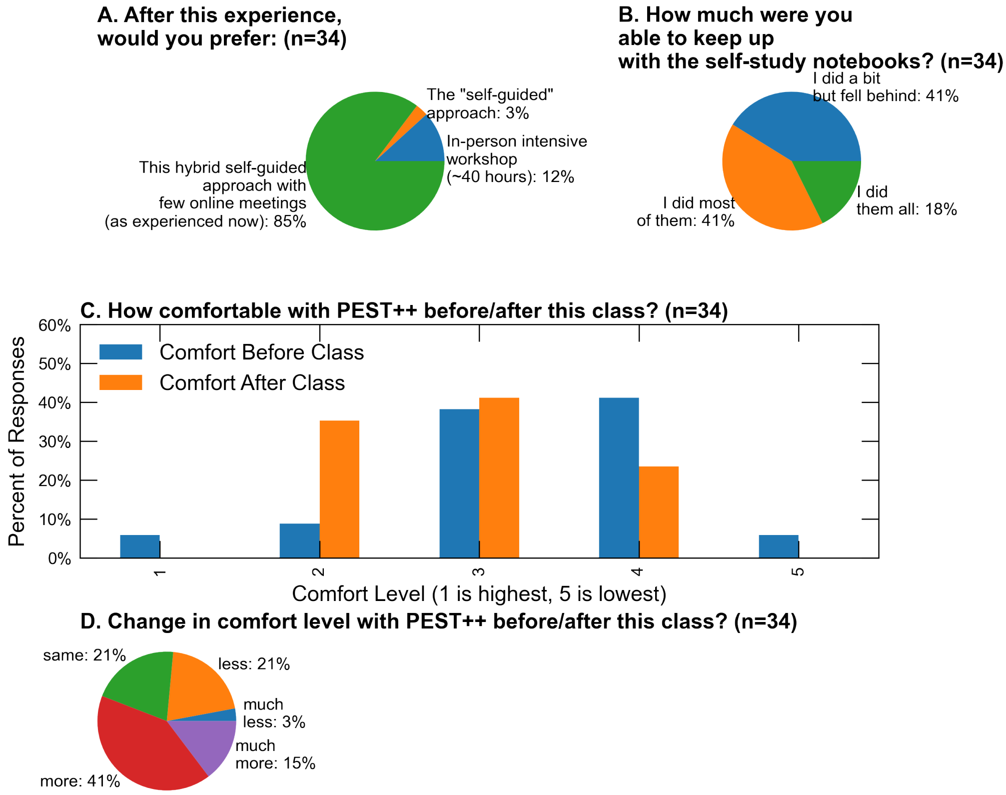

Figure 2: Summary of responses to post-course survey based on 34 responses. Panel A summarizes whether respondents would prefer an intensive in-person workshop or this hybrid option. Panel B summarizes how much of the notebooks respondents were able to complete throughout the course. Panel C summarizes respondent comfort level with PEST++ before and after the course. Panel D highlights individual changes in comfort level reported due to the course.

Acknowledgements

The tutorials were originally developed with support from the US Geological Survey (USGS) and support from USGS continues through the HyTest training project. Continued development and support is funded by the Groundwater Modelling Decision Support Initiative (GMDSI). GMDSI is jointly funded by BHP and Rio Tinto. We thank Dr. John Doherty for his tireless and pioneering efforts starting PEST and continuing to innovate and Dr. Randall Hunt for his leadership in PEST and PEST++ applications and development and contributions to the initial curriculum for this material and the early version of the notebooks. We thank Kalle Jahn (USGS), Ines Rodriguez and codyalbertross who made reviews that improved this manuscript. Lastly, we thank users and stress-testers for their valuable feedback and continued community contributions to the repository.

Disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., & Fienen, M. N. (2016). Scripting MODFLOW model development using python and FloPy. *Groundwater*, *54*(5), 733–739. https://doi.org/10.1111/gwat.12413

Doherty, J. (2015). *PEST and Its Utility Support Software*. https://pesthomepage.org/

Freyberg, D. L. (1988). AN EXERCISE IN GROUND-WATER MODEL CALIBRATION AND PREDICTION. *Groundwater*, *26*(3), 350–360. https://doi.org/10.1111/j.1745-6584.1988.tb00399.x

Langevin, C. D., Hughes, J. D., Provost, A. M., Russcher, M., Niswonger, R. G., Panday, S., Merrick, D., Morway, E. D., Reno, M. J., Bonelli, W. P., & Banta, E. R. (2022). *MODFLOW 6 Modular Hydrologic Model* (Version 6.4.1). https://doi.org/10.5066/F76Q1VQV

White, Jeremy T., Fienen, M. N., & Doherty, J. E. (2016a). A python framework for environmental model uncertainty analysis. *Environmental Modelling & Software*, *85*, 217–228. https://doi.org/10.1016/j.envsoft.2016.08.017

White, Jeremy T., Fienen, M. N., & Doherty, J. E. (2016b). A python framework for environmental model uncertainty analysis. *Environmental Modelling & Software*, *85*, 217–228. https://doi.org/10.1016/j.envsoft.2016.08.017

White, Jeremy T., Hemmings, B., Fienen, M. N., & Knowling, M. J. (2021). Towards improved environmental modeling outcomes: Enabling low-cost access to high-dimensional, geostatistical-based decision-support analyses. *Environmental Modelling & Software*, *139*, 105022. https://doi.org/10.1016/j.envsoft.2021.105022

White, Jeremy T., Hunt, R. J., Fienen, M. N., & Doherty, J. E. (2020). *Approaches to highly parameterized inversion: PEST++ version 5, a software suite for parameter estimation, uncertainty analysis, management optimization and sensitivity analysis*. US Geological Survey Techniques; Methods 7-C26. https://doi.org/10.3133/tm7C26