Model documentation in the eye of the beholder: Lessons learned from a flood risk model for a dike in the Netherlands

Brit E. van der Meijden[1] and Lieke A. Melsen[1*]
[1] Hydrology and Environmental Hydraulics, Wageningen University, Wageningen, The Netherlands

Abstract
Computer models are essential for flood risk decision-making, where documentation plays a key role in ensuring transparency and reproducibility. This study examined how different stakeholders perceive documentation practices by analysing a Dutch flood-risk model case study. We assessed perspectives from external observers (ourselves), the modellers (a consultancy company), and the model end-users (the Water Authority) through documentation analysis, model reruns, and interviews. We found a mismatch between the Water Authority's documentation goals and what was achieved with the delivered documentation. For instance, we could not reproduce the model results based on the available documentation, as crucial tacit knowledge remained implicit. Despite these shortcomings, both the consultancy company and the Water Authority were satisfied with the final products — illustrating how documentation quality is in the eye of the beholder and in this case shaped by the trust from the Water Authority in the consultancy company. However, when key goals like transparency and reproducibility are not met, accountability for model results becomes a concern, especially in decision-making contexts — highlighting the gap between perceived adequacy and broader expectations of documentation.

Keywords
model documentation; interviews; flood risk modelling

1. Introduction

Environmental computer models are of great societal importance, as they are used to support public policy decision-making (Oreskes et al., 1994; Bennett et al., 2013). In the water management domain, flood forecasting is one of the most critical model applications (Todini, 2011). Here, models have four essential functions: to predict the magnitude and probability of inundation; to provide insights in how risks may change over time, for instance as a result of climate change; to investigate the effectiveness of measures against flood risks; and to produce maps for flood risk planning and communication (Maskrey et al., 2022). Hence, computer models play a crucial role in contemporary flood risk management.

In the Netherlands, a low-lying delta-country, flood risk management is considered essential. The country is threatened by an excess of water from three sources: the sinking landscape, the increasing sea level, and water coming from two major rivers (Lintsen, 2002). As one of the most threatened countries by climate change in Europe, managing flood risks is a public interest (Wiering and Winnubst, 2017). Simulations of flood events are

Correspondence:
Contact L. Melsen at lieke.melsen@wur.nl

Cite this article as:
Van der Meijden, B.E. & Melsen, L.A.
Model documentation in the eye of the beholder: Lessons learned from a flood risk model for a dike in the Netherlands
Socio-Environmental Systems Modelling, vol. 7, 18748, 2025, doi:10.18174/sesmo.18748

Socio-Environmental Systems Modelling
An Open-Access Scholarly Journal
http://www.sesmo.org

relevant for water safety, risk management, and spatial adaptation of areas inside and outside the dikes (De Bruijn et al., 2018). As such, models to explore and mitigate flood risk are, despite their inherent uncertainty (de Moel et al., 2014), widely applied.

Since models in this context directly inform decision-making, their results have significant societal implications, necessitating a careful modeling process (Remmers et al., 2024). Several initiatives have aimed to streamline this process, including for example the Good Modelling Practice guidelines (van Waveren et al., 1999), a bottom-up effort within the Dutch hydrological modeling community, and extensive work by the Netherlands Environmental Assessment Agency on uncertainty treatment and communication (e.g. Peters et al., 2006). Also internationally, numerous initiatives propose guidelines for the modeling process, as comprehensively outlined by Jakeman et al. (2024), with model documentation playing a crucial role in all of them.

Model documentation serves various purposes, but two are most commonly highlighted: it enhances reproducibility (Hutton et al., 2016; Melsen et al., 2017; Gundersen, 2021) and it increases the transparency of the modelling process (Martinez-Moyano, 2012; Glynn et al., 2017). This is especially relevant in the context of models used to support decision-making, such as for flood risk: “Management and regulatory entities that claim to have a scientific basis for their decisions and actions ought to make accessible, documented, traceable predictions” (Glynn et al., 2017). In a decision-support setting, documentation might even have legal significance (Fisher et al., 2010; Calder et al., 2018).

The content of documentation is not the only important factor; the way it is documented also plays a crucial role. When documentation is written in highly technical language, it becomes more reliant on specialized knowledge, which can increase the costs and complexity of decision-making based on the model (Arnold et al., 2020). Bonet et al. (2014) argue that, although much of the information about environmental models is stored in individuals’ digital or biological memory, effective documentation can facilitate the generation of relevant insights and improve communication with decision-makers. However, in practice, the documentation of model workflows is often, at best, modest (Overeem et al., 2013; Jakeman et al., 2008; Janssen et al., 2020).

The aim of this study is to assess how different stakeholders involved in decision-support modelling perceive documentation practices. We achieve this by empirically examining documentation practices in a Dutch flood risk decision-support case study, considering the perspectives of modellers (from a consultancy company), end-users (at a Dutch Water Authority), and ourselves as external observers. First, we analyse the documentation of pre- and post-processing scripts and attempt to rerun the model experiment. Second, we explore the perceptions and challenges of these practices through interviews with both modellers and end-users. Finally, based on our findings, we reflect on the documentation practices, identifying key factors that shape current practices and the challenges they pose.

2. The case study: a flood risk model for a dike in the Netherlands

For water management, the Netherlands is subdivided into 21 regions, each with their own water authority. The water authorities are responsible for central water management within their region. As such, one of their legal obligations is to prepare scenarios for flood and inundation risks. If water levels are high, the water authority takes the necessary emergency response measures and provides other relevant parties, such as the Dutch Safety Regions, with information. The required information for specific response measures is obtained from flood and inundation simulation models. This study focuses on a flood risk modelling study that was conducted by a consultancy company, commissioned by a water authority (from now on referred to as ‘WA’). The study encompassed simulation of inundation resulting from a dike breach.

In a public tender in 2019, the model work for flood risk simulations was commissioned to a Dutch research and consultancy company. There were two specific requests from the WA for this tender. First of all, the WA requested the consultancy company (from now on referred to as ‘CC’) to use the D-HYDRO Suite 1D2D (Deltares, 2021). At that time (2019), this software was still under development and a beta version of the software had to be used, which was still prone to bugs and did not yet have a user-interface. Furthermore, not all functionalities for flood simulations were available. The WA wanted to explore the suitability of this software package for flood risk analysis. Secondly and following from the first requirement, the WA requested that the developed scripts and model set-up be transparent and reproducible, in order to be able to continue the development of the model at a later stage. One of the underlying reasons was so that the WA would not be dependent on the same CC for follow-up development.

The CC developed pre- and post-processing scripts to set up the model based on local data and to perform multiple model runs with the D-HYDRO Suite 1D2D. Different scenarios were explored regarding dike breaches, measures and model schematisations (such as including or excluding particular waterways). The functioning of the pre- and post-processing scripts and the results from the scenario and sensitivity study were documented in a report which was delivered to the WA. Furthermore, the software and scripts were installed on a local computer at the WA, in order to provide access to the model.

For feasibility and because the WA specifically requested documentation on these aspects, this study focuses on the documentation of the pre- and post-processing scripts that are used to set up the model and analyse its output. Consequently, we do not include the documentation of the model software D-HYDRO Suite 1D2D in this analysis.

3. Methods

The case study was examined using multiple methods. First, we analysed the scripts and the accompanying documentation. Second, we attempted to run the scripts and execute model runs. Based on these observations and attempts, interviews were conducted with both modellers and end-users.

3.1 Observations: written documentation and reproducing model runs

The first step of this study was an analysis of the documentation and the pre- and post-processing scripts. The documentation consisted of a description of the purpose of each of the scripts and a report of a sensitivity analysis that was conducted. The scripts themselves contained short comments. The documentation and the scripts were studied and mutually compared, to investigate if all choices in the scripts were substantiated in the documentation. For example, for all the numbers that were hardcoded in the scripts, we consulted the documentation to see if they were explained or justified. This was triangulated with a report on a conducted sensitivity analysis.

After gaining familiarity with the scripts and documentation, it was attempted to rerun the scripts. We aimed to rerun the scripts first with their initial values, and after that with slightly adapted values for decisions in these scripts that were identified as not well substantiated in the documentation. This served as input for the interviews. The pre-processing scripts eventually ran successfully, and several decisions in the input scripts were adapted to generate new input files. These included the decision on the width of the buffer around the dike breach (which was initially set to 60 meters, and was also run for 40 and 80 meters), the resampling method (set to 'average', and re-run with 'median' and 'nearest'), and the friction coefficient (set to 0.035, based on De Bruijn et al. (2018), re-run with 0.023).

After running the pre-processing scripts, which created the input files for the model, there was a first large test for the documentation: trying to run the model and the subsequent post-processing scripts. We attempted these model runs both on our personal computers and on a High-Performance Cluster where the same model code (albeit another version) is run for other projects.

3.2 Model set-up: interviews with modellers from the consultancy company

After studying the scripts and documentation, the modellers of the CC were interviewed about the development process of the scripts and the runs they performed with the D-HYDRO model. The interviews were semi-structured. As the name suggests, semi-structured interviews are in between a structured interview – in which all questions are pre-determined in a particular order and form – and an unstructured interview – in which both the structure and questions being asked are flexible (Kumar, 2014; Adams, 2015). The advantage of semi-structured interviews is that they efficiently collect detailed qualitative data on drivers and behaviours. At the same time, unexpected information and features of quantitative data can be acquired (O'Keeffe et al., 2016).

The interviews were conducted using a pre-determined interview guide that covered multiple topics of relevance; see Supplementary Material A. The interviews started with an introduction and questions about the experience and professional background of the interviewee. To get a sense of the perceptions and motivations of the modellers on water management models in general, the interview continued with questions about what they thought was the main purpose of modelling studies such as the one done for this case study. Subsequently, the interview questions related to the procedures regarding documentation of script development and use, previous knowledge that is assumed necessary to work with a script, and the use of scripts by other parties and persons. The questions were based on earlier (interview) studies on modelling (Melsen, 2022; Deitrick et al., 2021; Mayer et al., 2017; van Voorn et al., 2016).

All three employees from the CC who were involved in developing the preprocessing scripts for this case were interviewed. One of these three meanwhile moved to another company, but still kindly agreed to be interviewed. The interviews took around one hour, were held in Dutch, were recorded, and consisted of open questions. They were formulated with the aim to not be ambiguous, leading, double-barrelled, or based on presumptions (Kumar, 2014). Interview questions were slight reformulated after the first interview, to be more open to prompt further explanation. The first interview was, however, still included in the results. The interviews were transcribed based on the recordings, and analysed qualitatively.

3.3 Model use: interviews with persons from the Water Authority

In consultation with the WA, four employees were interviewed. One was the project leader of the modelling study, one the account manager, one a dike expert who was involved in the development process, and the last person was responsible for advice and communication in case of an emergency. In general, the same topics were discussed as with the modellers from the CC. If applicable, even the same questions were asked as in the interviews with modellers, to compare their viewpoint on specific issues, e.g., the required knowledge to use scripts like the ones in this project.

First their perceived purpose of water management models was discussed, after which the questions mainly focused on how model information is usually transferred from consultancy company to them as commissioner, their requirements regarding documentation, and possible difficulties they experience with documentation. The interview practicalities were similar to those held with the modellers from the CC: the interviews were semi-structured, recorded, transcribed, and then read and analysed. This time the interview guide was slightly adapted per interviewee since each person had a different role and different responsibilities. An overview of the interview guide can be found in Supplementary Material B, although not all questions were discussed with all interviewees.

4. Results

In this section, we first present our findings based on studying the documentation and the scripts. Secondly, we discuss the results obtained from interviewing the modellers from the Consultancy Company (CC), first on what they consider the goal of the model and second on how they perceive documentation. Finally, we present the results obtained from interviews with persons from the Water Authority (WA), what the goal of the model was for them and how they perceived the documentation of the model in this case study.

4.1 Observations: written documentation and reproducing model runs

The available documentation on the scripts and the conducted study was well-structured. The script documentation detailed each script’s name, purpose, parameters, required input, and connections to other scripts. The report on the conducted sensitivity analysis described key model design choices and showed simulation results. Although many decisions were well documented, the report did not describe the reasons for selecting the specific factors that were explored in the sensitivity analysis, nor why some other possibly relevant factors were excluded, such as the implementation of initial water height of ditches in the model, as suggested by the dike expert of the WA (later it became clear that the software did not allow this yet). Furthermore, in the comparison between the scripts and the documentation, several instances were identified where a choice in the script was not mentioned or explained in the documentation. An example was found in the file with the definitions for all cross-section profiles. The friction coefficient for all locations was set to 0.035 in a hard-coded fashion. This is different from the default value of 0.023 (De Bruijn et al., 2018), and it is not clear what the 0.035 was based on. Lastly, the link between the scripts and the model was not explained anywhere – how the scripts generated output to serve as input to the model, and how the model would be able to read these.

Furthermore, attempts to re-run the pre-processing scripts, the model, and the post-processing scripts revealed that some essential steps in the documentation were missing. For clarification of these steps, both the modellers from the CC and the model developers at Deltares needed to be consulted. Important here was that the principal author of the scripts no longer worked at the CC at the time of our study. Although colleagues of this person, who took part in the development process, knew some of the necessary information to use the scripts, no one was entirely informed. This led to a long process before a first model run could be initiated: every time a little part of the information was found, this had to be discussed with other people to get the next piece of information. An important reason for this was that, although there was a relatively extensive document available that described the functioning of the pre- and post-processing scripts, the most fundamental step of how to use the generated files to run the D-HYDRO model – was not described nor documented. Finally, it was found that this needed to be done using a researcher "interactor" version of the computational engine, since the model schematisation stemmed from a time when the model functionality was still beta. It appeared that the main model input file (the MDU file) had to be changed manually to run the model with the generated files. Also, the grid that was generated with the preprocessing scripts required a manual step that was not documented: the region of interest needed to be clipped by hand. In the end, we, the authors, did succeed in initialising a run, but not in finalising let alone reproducing model results, probably related to required software settings.

It should be emphasised that besides the observations described above, several best practices could be extracted from the script development. The available documentation was well structured, as were the scripts: all in the same way, starting with the packages to be imported, followed by the path references, the definition of parameters used, and the functions. The scripts ended by calling those functions. In addition a report explained each script, which data were needed, what was done with the data, and which output was generated. The structure for the input and output directories was made and explained in this report. Lastly, the definition of the refinement factors, which specify grid resolution, allowed for an easy adjustment of the degree of refinement and the locations to be refined. Hence, the identified best practices relate to steps that were explicit and clearly reported and documented.

In short, we observed that the documentation was incomplete concerning the justification of several modelling decisions that were made in the scripts, and incomplete to facilitate reproducibility of the model results for persons outside of the organisation (us, the authors of this study)[1]. The observed shortcomings in the documentation particularly relate to manual steps and thinking steps that were not documented, where the requirement of manual steps was partly related to the use of beta software. Because of these shortcomings, the necessary information to understand and run the scripts and model was lost when the main author of the scripts switched jobs.

4.2 Model set-up: interviews with modellers from the consultancy company

Three interviews were held with modellers that worked on the pre- and postprocessing scripts, including the main developer, who has since moved to another job. The interviews started with questions related to their background. All interviewees had similar academic and professional backgrounds, having studied 'Water engineering and management' at the same Dutch university. All of them started working for this consultancy company right after graduation. They all had experience with multiple types of hydrological and hydraulic models and focused part of their activities on operational management and advice. Their professional experience varied between four and seven years.

Concerning this specific case study, the majority of the script development was done by one person, with the help of two colleagues and in consultation with the WA. The scripts were a follow-up from an earlier study and were followed up by two similar studies. Large parts of the scripts were already developed before the WA officially commissioned the work. Following the assignment from the WA, the scripts were made more readable

[1] Confronted with these findings, the modellers and model developers estimated that, with a dedicated effort, they would probably eventually be able to reproduce the results.

and automated for the WA to be able to use them. This provides context for the results presented in the following selections.

4.2.1 Purpose of the model and scripts

This section elaborates on the purpose of the model and the scripts, according to the modellers at the CC. For all statements given, the number in brackets indicates in how many interviews (out of 3 in total) this was brought up by the interviewees. Interview quotes were selected to illustrate the main results. When asked about the goal of the hydrological model and these scripts in this project, two interviewees mentioned that the goal of hydrological modelling is to simulate an event that did not happen or was not measured yet (2) and that these scripts specifically served to quickly and automatically set up the input data for the model(2).

> "*The goal of the script is that you can set-up a model quick without many manual actions*."

The output generated by the scripts can be used to formulate mitigating measures or advice in case of a dike breach (2). Lastly, it was mentioned that the most crucial model output parameters are the maximum water depth, the maximum velocity, and the time of arrival of the water (2). The most essential objective is that the model accurately simulates reality (2). Remarks mentioned once, but seemingly relevant, are that the goal of hydrological modelling is to be able to say something about potential risks (1) and to create an operational framework (1) based on that. Moreover, script automation is meant to accommodate a fast update of the model to study different scenarios (1). The model results that follows from these scripts are compulsory for the Dutch Safety Regions to have (1) and can furthermore be used to decide which water structures need to be reinforced or adjusted (1) and to increase awareness of inhabitants (1).

Based on these results, it can be said that the modellers are aware of the potential use of flood risk models, e.g., to formulate an operating framework and inform the Dutch Safety Regions. It is also apparent that the scripts developed for this case are more seen as an automation tool rather than as component of an actual model. Ideally, such scripts would be integrated into an operational framework.

4.2.2 Documentation and information transfer

The questions related to documentation and reproducibility in the interviews with the modellers from the CC showed that these topics have low priority. The number in brackets represents how in how many interviews certain statements were made, from the three in total.

It was mentioned in two interviews that the delivery of relevant documentation is part of the completion of an assignment (2). However, there is no standard or fixed agreement on what such documentation needs to encompass or look like: the company does not have any protocols or procedures for documenting the model development process. Although all interviewees (3) were familiar with the Good Modelling Practice Handbook (van Waveren et al., 1999), the handbook itself was not actively used (2).

The reviewers all agreed (3) that substantial adaptations or choices that deviate from the standard should be documented. Confronted with the choice of setting the friction coefficient to 0.035 (rather than the standard of 0.023) resulted in the following response:

> "*It probably wasn't investigated very well, it comes from a default value and we might have discussed that we could adjust it to this value*."

The modellers estimate that the scripts are only readable for people with experience with Python and hydrological modelling (2) and are not made for a client to fully understand them (2). The client, therefore, usually also does not even receive the scripts:

> "*Usually, the only thing provided to a client is a report for which models and scripts have been used, the output, and possibly which maps were used as input*."

> "*In our experience, clients are not so interested in the scripts; they are mostly interested*

> *in the output. That is because that is more visible and tangible. We do not really discuss scripts with our clients normally*."

It is common to use a script that was developed by someone else (2), and it is perceived as important that someone with the same knowledge as the original developer of a script would get the same results when using it (2):

> "*The most important condition for documentation is that one should be able to reproduce the study*."

The response from one of the interviewees seems to suggest that reproducibility should only be achievable with a certain level of knowledge, confirmed by these quotes:

> Modeller: "*But it is not for everyone. Not everyone in the water authority could just make a flood model and run it. You need some background information*."
>
> Interviewer: "*And colleagues or partner-organisations should be able to use it in the same way as the one who develops it?*"
>
> Modeller: "*Yes, of course not every colleague because not every colleague is engaged in flood risk modelling. But the ones with the knowledge, should be able to do so, yes. They should be able to reproduce it*."

Other relevant remarks related to script development are that documentation is mainly done by versioning scripts (1). The script itself effectively serves as a log file (1). In this particular case study, the original scripts had to be adjusted because the first one included steps that only the original developer understood (1), and to reduce the number of manual steps. But still, there were manual steps involved, besides the observed steps described in Section 4.1:

> "*I think I manually adjusted the cross-sections in locations where there were bridges, or wrote a small script to do so*."

Moreover, choices are usually written down in a report, but not every parameter can be described because there are too many in a model (1). It was said to be less efficient to continue with a modelling project where someone else worked on before (1), and documenting everything is not perceived as valuable:

> "*In a small company like ours, the one continuing to improve scripts is usually also the one who worked on them before. So it is not of added value to document it every time you improve a script*."

Lastly, the level of detail of the documentation also depends on the time and budget available (1):

> "*At the start of this project, I made a consideration: with this amount of days and budget available, I can work everything out at this level of detail. But if I would have gotten twice the budget, I could have spent more time on that*."

In summary, the script developers are aware of the relevance of reproducibility and documentation of the steps they take. However, modellers develop their own modelling approaches over time, and do not feel the urge to do this in a standardised way. It is common to further develop their own scripts, not scripts initially written by others. They also assume that the people using a script or model have affinity with hydrology and programming and usually do not expect a client to look into the scripts or the model they developed. Lastly, but importantly, the level of documentation depends on the budget made available by the client.

4.3 Model use: interviews with persons from the Water Authority

The four employees of the Water Authority (WA) that were interviewed (the project leader, account manager, dike expert, and emergency response advisor) all had distinct backgrounds. They had been working at this WA for between four and ten years. Apart from the emergency response advisor, they all studied water management or fluid mechanics. The emergency response advisor had a background in safety management and education.

They all previously worked at a consultancy company and all had experience with hydrological and hydraulic models.

4.3.1 Purpose of the model and scripts

The number in brackets indicates how many of the four interviews included certain statements. Three interviewees agreed that the main purpose of the model study was to update the current, dated flood risk information for the WA (3). All interviewees mentioned that some years ago, the responsibility for providing flood risk information for an area was shifted from the provincial government to the WA (4). Other remarks made by individuals were that the scripts were made to be able to run the model (1) and that the data generated by the model is shared with relevant stakeholders, like governments and safety authorities (1). The emergency response advisor indicated that this information is used for three purposes within the WA: it is used as a basis for training, for risk analysis, and for preparation plans in case of a dike breach. The safety authority uses the information, for instance to define vital structures such as electricity stations or to safeguard evacuation routes from hospitals. The project manager indicated that the essential requirements of this study were that it was open-source, use-able for different dike cases, and, therefore, mainly automatised (1). There were two reasons underlying these requirements:

> "*On the one hand, there is the strategic argument of an equal level playing field. Meaning that [the consultancy company] would not have a competitive advantage: another party, which has not developed these scripts, should not need a lot more time to run them than [consultancy company]. And on the other hand, if something is developed using tax money, it should be open-source, in my opinion*."

4.3.2 Documentation and information transfer

The interviews with the employees of the WA produced a variety of results on the topic of documentation and reproducibility, with most statements only mentioned by one interviewee. Many steps of the model development process are based on trust (2). For example, after indicating the goal of the study, it was trusted that the developers would know how to reach that goal based on their expertise (1) and that they would know which documentation was relevant for this (1). After the scripts were developed and the model runs were performed, no formal evaluation of the delivered work was done (2). The scripts, model, and necessary dependencies were installed on a computer of the WA to demonstrate that they worked on their system (1). This was already a step ahead, since in the past, the WA would have mainly been interested in the model output, and not in the steps necessary to generate that output (1). As indicated by the project manager from the WA, a requirement of this study was that it would be in the form of a workflow. This would add to the efficiency, reproducibility, and consistency of the model study. Therefore, there were requirements concerning open-source development and documentation.

One of the requirements was 'documentation', although there were no guidelines or directives on what this documentation had to comply with (1):

> "*Because it is relatively new to us, and also because it's a pilot, we thought: yeah, well, we don't know either... So we have not set any specific preconditions or requirements for the delivery [of the documentation], except to, yes, provide documentation*."

Whether the underlying reasoning of choices should be part of such documentation was said to be dependent on the type of choice: choices concerning schematisation and thinking steps should be documented (1). The WA should be able to re-run everything and make adjustments if necessary (1). Therefore, one of the deliverables of this project was to make the tools run on a computer of the WA (1). The WA was successful in rerunning the model on the computer where the CC installed the model with the right environment, but they did not make any major adaptations to the model settings. Otherwise, the omissions in documentation that hampered re-running the model with slightly modified settings for our study, as described in Section 4.1, would have become apparent.

The interviewees from the WA not only highlighted the importance of documentation in this case study, but also in general. As discussed previously, a project financed by tax-money should, to their opinion, always be reproducible and transparent. One of the interviewees mentioned that documentation is less organised and prioritised in the Netherlands, compared to other countries where this person had worked before and where documentation-guidelines had legal significance (1). The emergency response advisor indicated that it is essential to have all relevant information available in one central place, and to not be dependent on one or two people in case of an emergency, in order to be able to run the model using the circumstances of the actual dike breach. In short, for the WA, documentation adds to a modelling study's transparency, efficiency, consistency, and reproducibility, and for this particular case study would have allowed them to run realistic scenarios in case of an emergency.

5. Discussion

In this discussion, we first examine the shortcomings of the documentation in our case study and the practical consequences. Subsequently, we place these issues in a broader context, highlighting the low priority given to documentation by both developers and commissioning parties. We explore why, despite these limitations, the modellers and end-users remained satisfied with the process and the documentation, ultimately exposing the tension between perceived sufficiency of documentation and the broader, often unmet, expectations of documentation in high-stakes decision-making.

5.1 Shortcomings of the documentation and practical consequences

Based on our own exploration of the scripts and documentation we observed that, although extensive and well-organised, the documentation was still incomplete: several motivations for modelling choices were lacking, thereby hampering transparency, and several manual steps required to re-run the model were not well-documented, thereby hampering reproducibility. In combination with the use of beta software and the main developer no longer working at the CC, meant that we, the authors, were unable to reproduce the model results.

Based on the interviews with the modellers this should, however, not come as a surprise. The modellers indicated that the model should be reproducible for someone with the same knowledge and experience as the original developers. From such a perspective, reproducibility is limited to only a small circle of people. This is in line with earlier studies on reproducibility, that emphasise the large role of tacit knowledge in reproducing an experiment (Collins, 1975; Winsberg, 2010; Ananny and Crawford, 2018; Edwards, 2010). While documentation is meant to explicate all steps for reproduction, a certain level of enculturation, where the reproducer immerses themselves in the field or lab from which they would like to reproduce their experiment, seems inevitable (Collins, 1975; Melsen, 2025). Also for this study, an internal stay of one of the authors at the consultancy company would have likely solved some of the issues encountered in reproducing the model results.

Our findings indicate that the goals to request documentation by the Water Authority were not achieved. To reiterate the goals of the WA that we identified based on our interviews: transparency, efficiency, consistency, reproducibility, and the ability to run realistic scenarios in case of an emergency. The documentation enhances transparency by justifying and explaining many model decisions. However, since not all decisions are documented, it fails to provide real accountability, as argued by Ananny and Crawford (2018). Efficiency, defined as the continuation of using the developed scripts for other case studies or project, is likely improved compared to a scenario without documentation, but given that none of the modellers at the CC had full oversight — due to the main developer switching jobs — suggests that a new modeller might prefer to start from scratch rather than first understanding and then further developing the existing scripts. The WA also desired consistency to ensure uniform methods are applied across different areas. However, the current scripts do not seem transferable, as highlighted by our own experience with adapting the settings and struggling to run the model. As for reproducibility, from the CC modellers' perspective, this is expected to be limited to a small circle given the level of expertise needed. Finally, the ability to run realistic scenarios during an emergency is hindered by the already discussed issues with consistency, transferability, and reproducibility: using these scripts for real-time emergency simulations would be difficult, as no one at the WA has the necessary expertise. As such, while the documentation has contributed to some of the WA's goals, the majority of the goals remain unachieved.

The practical consequences have remained limited so far. The WA for instance required efficiency and reproducibility in order to not depend on the same consultancy company for follow-up tenders. However, follow-up tenders were, in the end, assigned to the same consultancy company and therefore the shortcomings in the documentation did not emerge. Similarly, no dike breaches have occurred in the region, which means no real-time emergency simulations were necessary. This emphasises how the relevance of documentation remains unnoticed, only until they fail (Bowker and Star, 1999).

5.2 The low priority given to documentation

Two additional observations emerged from the interviews. First, there is limited knowledge and experience with writing documentation at the CC. For instance, there was no established procedure to consistently check whether the documentation was complete. It was noted in the interviews that the extent of documentation also largely depends on the resources allocated by the commissioner. Second, the commissioner, the WA, had limited knowledge and experience in defining specific documentation requirements and evaluating the provided documentation. While the WA was able to re-run the model results, as all necessary settings had been installed, they did not test model adaptations, nor was the documentation thoroughly reviewed. Both observations demonstrate a culture in which documentation is not prioritised — with the tender for this case-study being a first step towards giving more attention to documentation.

The low priority given to documentation is a well-documented issue in empirical studies on documentation practices. Maintenance work, including documentation, is often perceived as less valuable and rewarding than developing new, cutting-edge methods (STAR, 1999; Jackson, 2014). As a result, empirical evaluations frequently reveal shortcomings in documentation. For instance, Kajko-Mattsson (2005) defined 19 "rudimentary" documentation requirements and assessed their implementation across 18 institutes in Sweden — none met the criteria. Similarly, in their study on model use, Arnold et al. (2020) found manual steps in workflows, a lack of full transparency regarding required knowledge, and reliance on restricted-access licenses. Janssen et al. (2020) evaluated documentation and code-sharing practices in published models and found that in 2018, only 18% of models had shared code. If model documentation was available at all, it was often found to be incomplete. These findings highlight that, while documentation has the potential to enhance transparency, reproducibility, and other key criteria (Calder et al., 2018; Polhill and Edmonds, 2007; Janssen et al., 2020; Gundersen, 2021), these standards are frequently unmet.

Despite the widespread shortcomings in documentation, extensive guidelines do exist. The ODD protocol and TRACE framework, introduced by (Grimm et al., 2010; Schmolke et al., 2010; Grimm et al., 2014, 2020), are among the most comprehensive and concrete documentation recommendations to date. However, these approaches also present challenges. One major issue is the length of documentation (Müller et al., 2014; Grimm et al., 2020). For instance, the shortest TRACE notebook provided by Ayllón et al (2021) spans approximately 55 pages — raising concerns that such extensive documents could, paradoxically, hinder transparency rather than enhance it. Beyond length, other challenges persist. Polhill and Edmonds (2007) compiled a list of existing documentation standards, including those used by NASA[2], yet even well-structured guidelines struggle to capture the role of tacit knowledge. Furthermore, it is unlikely that a single documentation standard can optimally serve all models and contexts (Polhill and Edmonds, 2007). These factors highlight that, while frameworks exist, practical barriers still limit the effectiveness of documentation in real-world applications.

5.3 Model documentation in the eye of the beholder

Despite our finding that the documentation did not meet the majority of the WA's stated goals, both the WA and the CC were satisfied with their collaboration and the final product. This suggests that the perceived quality of documentation is subjective: a direct colleague may find it sufficient, while an outsider (us, researchers) may see it as lacking critical information. Meanwhile, a third party, the WA in our case, might consider it adequate and be satisfied. In this sense, the quality of model documentation is truly in the eye of the beholder.

This satisfaction, in our case study, can be attributed to several factors. First, the documentation has not yet been put to the test, as discussed in Section 5.1. Second, the WA is in the process of transitioning to a more documentation intensive approach, and as such the current documentation is seen as an improvement over previous practices. Third, the WA places trust in the expertise of the consultancy company, as expressed by the project leader from the WA:

[2] See: https://ntrs.nasa.gov/citations/19980228459

> "*Of course, you also have trust in the expertise and competence of the contractor. So, you assume that certain assumptions are stated or reported [...]*."

This highlights the knowledge asymmetry between the WA and the CC in the modelling process, making the WA particularly sensitive to unaware, tacit knowledge that the consultancy company not explicitly documents.

From these reflections, we learn that model documentation is heavily influenced by social processes such as tacit knowledge and trust. Tacit knowledge often leads to incomplete documentation, which undermines goals like reproducibility and transparency. Despite this, end-users may be satisfied with the documentation because they trust and rely on the expertise of their contractor, illustrating how the quality of model documentation is in the eye of the beholder. However, when goals like reproducibility and transparency are not achieved, it raises concerns about accountability for model results, especially in contexts where decisions are made based on the model's outputs (Funtowicz and Ravetz, 1993; Saltelli et al., 2020). Ultimately, this highlights the tension between perceived sufficiency of documentation and the broader, often unmet, expectations of documentation in high-stakes decision-making.

6. Conclusion

This study examined how different stakeholders in decision-support modeling perceive documentation practices. We analysed a flood risk model for a Dutch dike, evaluating perspectives from external observers (ourselves), model developers (a consultancy company), and model end-users (the Water Authority). Our assessment included analysing documentation of pre- and post-processing scripts, attempting to rerun the model, and conducting interviews.

We found a mismatch between the Water Authority's documentation goals — such as transparency, efficiency, consistency, and reproducibility — and what the provided documentation actually achieved. For instance, despite reproducibility being a key objective, we were unable to reproduce the model results based on the available documentation. Tacit knowledge appeared crucial for reproducibility, which remained largely implicit. Despite that the documentation did not meet the desired goals, both the consultancy company and the Water Authority were satisfied with the delivered documentation - the quality of model documentation is in the eye of the beholder. This was largely shaped by trust: the Water Authority trusted the expertise at the consultancy company, assuming that they would document whatever would be necessary. However, when key goals like transparency and reproducibility are not met, accountability for model results becomes a concern, especially in decision-making contexts. This highlights the tension between perceived documentation adequacy and the broader, often unmet, expectations of documentation in high-stakes scenarios.

Acknowledgements

The authors would like to thank the consultancy company and the water authority for their cooperation, help in trying to get the model running, time, facilitation of the process, and open attitude. Furthermore, we would like to thank the external model developers from Deltares who supported us in trying to get the model running (Arthur van Dam and Rinske Hutten). This study was performed in the context of the Dutch TKI-project 'HYDROLIB', initiated by Deltares, on creating a shared and documented software library in a community of Water Authorities, consultancy firms and research institutes. The second author thanks the Dutch Research Council for financial support through a personal Veni grant (nr. 17297, entitled "What about the modeler? The human-factor in constructing Earth and environmental predictions").

Supplementary Material

The Supplementary Material for this article can be found online at https://sesmo.org/article/view/18748/18364.

References

Adams, W. (2015). Conducting semi-structured interviews. In John Wiley & Sons, Ltd (Chap. 19, pp. 492–505). https://doi.org/10.1002/9781119171386.ch19

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20(3), 973–989. https://doi.org/10.1177/1461444816676645

Arnold, T., Guillaume, J., Lahtinen, T., et al. (2020). From ad-hoc modelling to strategic infrastructure: A manifesto for model management. Environmental Modelling & Software, 123, 104563. https://doi.org/10.1016/j.envsoft.2019.104563

Ayllón, D., Railsback, S. F., Gallagher, C., Augusiak, J., Baveco, H., Berger, U., Charles, S., Martin, R., Focks, A., Galic, N., Liu, C., van Loon, E. E., Nabe-Nielsen, J., Piou, C., Polhill, J. G., Preuss, T. G., Radchuk, V., Schmolke, A., Stadnicka-Michalak, J., Thorbek, P., & Grimm, V. (2021). Keeping modelling notebooks with TRACE: Good for you and good for environmental research and management support. Environmental Modelling & Software, 136, 104932. https://doi.org/10.1016/j.envsoft.2020.104932

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterising performance of environmental models. Environmental Modelling & Software, 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011

Bonet, F. J., Pérez-Pérez, R., Benito, B. M., Suzart de Albuquerque, F., & Zamora, R. (2014). Documenting, storing, and executing models in ecology: A conceptual framework and real implementation in a global change monitoring program. Environmental Modelling & Software, 52, 192–199. https://doi.org/10.1016/j.envsoft.2013.10.027

Bowker, G. C., & Star, S. L. (1999). Sorting things out: Classification and its consequences. The MIT Press. https://doi.org/10.7551/mitpress/6352.001.0001

Calder, M., Craig, C., Culley, D., de Cani, R., Donnelly, C. A., Douglas, R., Edmonds, B., Gascoigne, J., Gilbert, N., Hargrove, C., Hinds, D., Lane, D. C., Mitchell, D., Pavey, G., Robertson, D., Rosewell, B., Sherwin, S., Walport, M., & Wilson, A. (2018). Computational modelling for decision-making: Where, why, what, who and how. Royal Society Open Science, 5(6), 172096. https://doi.org/10.1098/rsos.172096

Collins, H. (1975). The seven sexes: A study in the sociology of a phenomenon or the replication of experiments in physics. Sociology, 9, 205–224.

De Bruijn, K., Wagenaar, D., Slager, K., de Bel, M., & Burzel, A. (2018). Leidraad voor het maken van overstromingssimulaties (Tech. Rep. No. 1200537-007-ZWS-0004). Deltares.

de Moel, H., Bouwer, L., & Aerts, J. (2014). Uncertainty and sensitivity of flood risk calculations for a dike ring in the south of the Netherlands. Science of The Total Environment, 473–474, 224–234. https://doi.org/10.1016/j.scitotenv.2013.12.015

Deitrick, A., Torhan, S., & Grady, C. (2021). Investigating the influence of ethical and epistemic values on decisions in the watershed modeling process. Water Resources Research, 57. https://doi.org/10.1029/2021WR030481

Deltares. (2021). D-HYDRO Suite 1D2D. https://www.deltares.nl/nl/software/d-hydro-suite-1d2d/

Edwards, P. N. (2010). A vast machine: Computer models, climate data, and the politics of global warming. The MIT Press.

Fisher, E., Pascual, P., & Wagner, W. (2010). Understanding environmental models in their legal and regulatory context. Journal of Environmental Law, 22(2), 251–283. https://doi.org/10.1093/jel/eqq012

Funtowicz, S., & Ravetz, J. (1993). Science for the post-normal age. Futures, September, 739–755.

Glynn, P. D., Voinov, A. A., Shapiro, C. D., & White, P. A. (2017). From data to decisions: Processing information, biases, and beliefs for improved management of natural resources and environments. Earth's Future, 5(4), 356–378. https://doi.org/10.1002/2016EF000487

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. Ecological Modelling, 221(23), 2760–2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., Liu, C., Martin, B. T., Meli, M., Radchuk, V., Thorbek, P., & Railsback, S. F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. Ecological Modelling, 280, 129–139. https://doi.org/10.1016/j.ecolmodel.2014.01.018

Grimm, V., Railsback, S. F., Vincenot, C. E., Berger, U., Gallagher, C., DeAngelis, D. L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A. S. A., Milles, A., Nabe-Nielsen, J., Polhill, J. G., Radchuk, V., Rohwäder, M. S., Stillman, R. A., Thiele, J. C., & Ayllón, D. (2020). The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism. Journal of Artificial Societies and Social Simulation, 23(2), 7. https://doi.org/10.18564/jasss.4259

Gundersen, E. (2021). The fundamental principles of reproducibility. Philosophical Transactions of the Royal Society A, 379, 20200210. https://doi.org/10.1098/rsta.2020.0210

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science? Water Resources Research, 52(10), 7548–7555. https://doi.org/10.1002/2016WR019285

Jackson, S. J. (2014). Rethinking repair. In Media technologies: Essays on communication, materiality, and society. The MIT Press. https://doi.org/10.7551/mitpress/9042.003.0015

Jakeman, A., Voinov, A., Rizzoli, A., Chen, S. H. (Eds.). (2008). Environmental modelling, software and decision support: State of the art and new perspective. Elsevier.

Jakeman, A. J., Elsawah, S., Wang, H. H., Hamilton, S. H., Iwanaga, T., Grimm, V., Koralewski, T. E., Salado, A., Jordan, R., &

Voinov, A. A. (2024). Towards normalizing good practice across the whole modeling cycle: Its instrumentation and future research topics. Socio-Environmental Systems Modelling, 6, 18755. https://doi.org/10.18174/sesmo.18755

Janssen, M. A., Pritchard, C., & Lee, A. (2020). On code sharing and model documentation of published individual and agent-based models. Environmental Modelling & Software, 134, 104873. https://doi.org/10.1016/j.envsoft.2020.104873

Kajko-Mattsson, M. (2005). A survey of documentation practice within corrective maintenance. Empirical Software Engineering, 19(1), 31–55. https://doi.org/10.1023/B:LIDA.0000048322.42751.ca

Kumar, R. (2014). Research methodology: A step-by-step guide for beginners (4th ed.). SAGE Publications Ltd.

Lintsen, H. (2002). Two centuries of central water management in the Netherlands. Technology and Culture, 43(3), 549–568.

Martinez-Moyano, I. J. (2012). Documentation for model transparency. System Dynamics Review, 28(2), 199–208. https://doi.org/10.1002/sdr.1471

Maskrey, S. A., Mount, N. J., & Thorne, C. R. (2022). Doing flood risk modelling differently: Evaluating the potential for participatory techniques to broaden flood risk management decision-making. Journal of Flood Risk Management, 15(1), e12757. https://doi.org/10.1111/jfr3.12757

Mayer, L. A., Loa, K., Cwik, B., Tuana, N., Keller, K., Gonnerman, C., Parker, A. M., & Lempert, R. J. (2017). Understanding scientists' computational modeling decisions about climate risk management strategies using values-informed mental models. Global Environmental Change, 42, 107–116. https://doi.org/10.1016/j.gloenvcha.2016.12.007

Melsen, L. A. (2022). It takes a village to run a model - the social practices of hydrological modeling. Water Resources Research, 58(2), e2021WR030600. https://doi.org/10.1029/2021WR030600

Melsen, L. A., Teuling, A. J., Torfs, P. J. J. F., & Salvadore, P. A. T. (2025). The rise of the Nash-Sutcliffe efficiency in hydrology. Hydrological Sciences Journal. https://doi.org/10.1080/02626667.2025.2475105

Melsen, L. A., Torfs, P. J. J. F., Uijlenhoet, R., & Teuling, A. J. (2017). Comment on "Most computational hydrology is not reproducible, so is it really science?" by Christopher Hutton et al. Water Resources Research, 53(3), 2568–2569. https://doi.org/10.1002/2016WR020208

Müller, B., Balbi, S., Buchmann, C. M., de Sousa, L., Dressler, G., Groeneveld, J., Klassert, C. J., Le, Q. B., Millington, J. D. A., Nolzen, H., Parker, D. C., Polhill, J. G., Schlüter, M., Schulze, J., Schwarz, N., Sun, Z., Taillandier, P., & Weise, H. (2014). Standardised and transparent model descriptions for agent-based models: Current status and prospects. Environmental Modelling & Software, 55, 156–163. https://doi.org/10.1016/j.envsoft.2014.01.029

O'Keeffe, J., Buytaert, W., Mijic, A., Brozović, N., & Sinha, R. (2016). The use of semi-structured interviews for the characterisation of farmer irrigation practices. Hydrology and Earth System Sciences, 20(5), 1911–1924. https://doi.org/10.5194/hess-20-1911-2016

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. Science, 263(5147), 641–646. https://doi.org/10.1126/science.263.5147.641

Overeem, I., Berlin, M., & Syvitski, J. P. M. (2013). Strategies for integrated modeling: The community surface dynamics modeling system example. Environmental Modelling & Software, 39, 314–321. https://doi.org/10.1016/j.envsoft.2012.01.012

Peters, E., Bier, G., van Lanen, H. A. J., & Torfs, P. J. J. F. (2006). Propagation and spatial distribution of drought in a groundwater catchment. Journal of Hydrology, 321, 257–275. https://doi.org/10.1016/j.jhydrol.2005.08.004

Polhill, J. G., & Edmonds, B. (2007). Open access for social simulation. Journal of Artificial Societies and Social Simulation, 10(3), 10. https://www.jasss.org/10/3/10.html

Remmers, J., Teuling, A. J., & Melsen, L. A. (2024). A modeller's fingerprint on hydrodynamic decision support modelling. Environmental Modelling & Software, 106167. https://doi.org/10.1016/j.envsoft.2024.106167

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Espeland, W. N., Kay, J., Lo Piano, S., Mayo, D., Pielke Jr, R., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., Stirling, A., van der Sluijs, J., & Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. Nature, 582, 482–484. https://doi.org/10.1038/d41586-020-01812-9

Schmolke, A., Thorbek, P., DeAngelis, D. L., & Grimm, V. (2010). Ecological models supporting environmental decision making: A strategy for the future. Trends in Ecology & Evolution, 25(8), 479–486. https://doi.org/10.1016/j.tree.2010.05.001

Star, S. L. (1999). The ethnography of infrastructure. American Behavioral Scientist, 43(3), 377–391. https://doi.org/10.1177/00027649921955326

Todini, E. (2011). History and perspectives of hydrological catchment modelling. Hydrology Research, 42, 73–85. https://doi.org/10.2166/nh.2011.096

van Voorn, G. A. K., Verburg, R. W., Kunseler, E. M., Vader, J., & Janssen, P. H. M. (2016). A checklist for model credibility, salience, and legitimacy to improve information transfer in environmental policy assessments. Environmental Modelling & Software, 83, 224–236. https://doi.org/10.1016/j.envsoft.2016.06.003

van Waveren, R. H., Groot, S., Scholten, H., van Geer, F., Wüsten, H., Koeze, R., & Noort, J. (1999). Vloeiend modelleren in het waterbeheer: Handboek Good Modelling Practice (Tech. Rep. No. 99-05/99.036). STOWA/RIZA. https://edepot.wur.nl/181482

Wiering, M., & Winnubst, M. (2017). The conception of public interest in Dutch flood risk management: Untouchable or transforming? Environmental Science & Policy, 73, 12–19. https://doi.org/10.1016/j.envsci.2017.03.002

Winsberg, E. (2010). Science in the age of computer simulation. University of Chicago Press.