Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR030600

Key Points:
- Fourteen hydrological modelers were interviewed about their modeling decisions
- Experience from colleagues was the main motivation to make certain decisions
- The way in which a model is selected and configured is time and place dependent

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
L. A. Melsen,
lieke.melsen@wur.nl

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

Received 15 JUN 2021
Accepted 28 JAN 2022

It Takes a Village to Run a Model—The Social Practices of Hydrological Modeling

L. A. Melsen[1]

[1]Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, The Netherlands

Abstract Computer models are frequently used tools in hydrological research. Many decisions related to the model set-up and configuration have to be made before a model can be run, influencing the model results. This study is an empirical investigation of the motivations for certain modeling decisions. Fourteen modelers from three different institutes were interviewed about their modeling decisions. In total, 83 different motivations were identified. Most motivations were related to the team of the modeler and the modelers themselves, "Experience from colleagues" was the most frequently mentioned motivation. Both institutionalization and internalization were observed: a modeler can introduce a concept that subsequently becomes the teams' standard, or a modeler can internalize the default team approach. These processes depend on the experience of the modeler. For model selection, two types of motivations were identified: experience (from colleagues or the modelers themselves), and model vision (the model has assets that align with the modeling vision). Model studies are mainly driven by context, such as time constraints, colleagues, and facilities at the institute, rather than epistemic (such as aligning with the modeling vision). The role of local context in the construction of and the value assigned to models shows that models are social constructs, making model results time, and place dependent. To account for this context in the estimation of the robustness of model results, we need a diversity of opinions, perspectives, and approaches. This requires transparent modeling procedures and an explicit modeling vision for each model study.

Plain Language Summary Even if you give different people the same recipe and ingredients, the final dish will still taste differently. The same applies to the use of computer models: different modelers will make different choices, leading to different model results. In this study, I interviewed 14 modelers to study their modeling choices. In total, 83 different reasons to make a certain decision were identified. The most frequently mentioned reason to do something in a particular way, was the experience from colleagues. Most reasons were context dependent, they were for instance related to time constraints, available resources, colleagues, and personal preferences. This makes model results time and place dependent. It is important to be aware of this when estimating how reliable model results are.

1. Introduction

Hydrological models are frequently used tools for scientific research. In 2010, about 75% of the hydrological scientific publications related to runoff were based on a model study (Burt & McDonnell, 2015). The bibliometric analysis of Addor and Melsen (2019) demonstrates a steady increase in scientific publications based on hydrological models over time. Models are thus an accepted method in scientific hydrological research.

Every hydrological model is prone to uncertainty (Oreskes et al., 1994), and in order to draw robust conclusions based on models, this uncertainty has to be made transparent (Renard et al., 2010). Transparency of the uncertainty is not only needed for scientific rigor (Gupta et al., 2012), but also to support practical applications based on the scientific insights (McMillan et al., 2017). The use of models thus comes with (societal) responsibility (Hamilton et al., 2019; Melsen, Vos, et al., 2018). Estimating and quantifying uncertainty in hydrological models is a well-established research field—although challenges remain (Liu & Gupta, 2007) and uncertainty can be framed in different ways (Guillaume et al., 2017).

Often, this field takes a technical lens: the model is taken as a starting point, and from there the uncertainty is estimated, for example, in structure, parameters, and/or data (Wagener et al., 2004), for instance, through sampling. However, this notion of uncertainty only comprehends "technical" uncertainty, while models are also prone to methodological uncertainty (Funtowicz & Ravetz, 1993). Methodological uncertainty arises from differences in approaches and methods that are evaluated as appropriate for the research question. Different modelers evaluate different models as adequate and make different decisions in the many steps that must be taken to set-up and configure a model. Some of these steps can be explored with sensitivity analysis (e.g., comparing different spatial resolutions), while other steps are taken unconcsiously, cannot be captured easily in a sensitivity analysis, or are simply taken for granted.

The first step in a model study is the development of a perceptual model. By definition, this step introduces personal preferences, because the perceptual model is "*the summary of our perceptions of how the catchment responds*" (Beven, 2012). Subsequently, the perceptual model is translated into a computational model, or an existing hydrological model is selected. Chegwidden et al. (2019) and Melsen, Addor et al. (2018) showed that hydrological projections diverge when different hydrological models are used; there can even be a change in the direction of the trend. The selection of a model is thus a crucial step for the results of the study. But also decisions beyond the model choice may influence the model results (Ceola et al., 2015; Holländer et al., 2014; Melsen et al., 2019)(Ceola et al., 2015; Holländer et al., 2014; Melsen et al., 2019), such as decisions related to the employed spatial resolution of the model or the calibration strategy.

Modeling decisions can have epistemic, ethical, and political dimensions (Beck & Krueger, 2016). Packett et al. (2020), for instance, provide several examples of how gender can play a role in hydrological model design and configuration. Babel et al. (2019) revealed, based on interviews with modelers from several fields in the Earth sciences, that habits play a large role in modeling decisions. Mayer et al. (2017) employed values-informed mental models based on interviews with climate risk scientists, to disclose how their values influence their modeling decisions. On top of all that, the technical process of modeling is prone to human errors (Menard et al., 2021). For all of these reasons, Elsawah and Jakeman (2020) describe modeling as a *socio-technical intervention*.

Given that modeling decisions impact model results, and with that the scientific conclusions based on these results, it is relevant to investigate the motivations for these decisions. This study is an exploratory empirical investigation of the motivations for several modeling decisions. It is focused around hydrological modelers: Which decisions do hydrological modelers make when selecting and configuring their model? And what is the motivation for these decisions? Based on these insights, we can further define how to estimate methodological uncertainty in hydrological models, evaluate the legitimacy of hydrological modeling as scientific method, and consider methods to enhance transparency and good modeling practice in model studies.

2. Methodology and Data

This research aimed to identify motivations for model decisions. This study was exploratory, and therefore qualitative methods were employed. Fourteen hydrological modelers were interviewed using semistructured in-depth interviews (Adams, 2015). In this format, the key questions asked to each interviewee are in principal the same, but probed toward the particular case of the interviewee, and follow-up *why* or *how* questions are asked to achieve the required depth. The interviews were held between February and June 2020, partly in real life and partly through Skype because of travel restrictions related to the global Covid-19 pandemic. All interviews were held in English.

2.1. The Interviewees

The interviewees were recruited in two ways. Two group leaders responded to a call in the AboutHydrology mailing list. They agreed to participate in this interview study, and proposed team members, who were contacted with an invitation to join the interview study. Everyone accepted this invitation. A third group leader is part of the network of the interviewer and was personally approached to participate. This group leader also provided names of team members. Four out of five team members accepted the invitation to participate. As such, the interviewees come from three different institutes, and the interviewees from the same institute are direct colleagues.

The three institutes are all located in Western Europe. One institute is a university. The department of the interviewees has nine professors, but only one of them was interviewed along with team members that are part of the subdepartment of this professor: one postdoctoral researcher and one PhD-candidate. The two other institutes are national research centers. One team has 10 permanent staff members, of which three were interviewed, along with one postdoctoral researcher and one PhD-candidate. The third team is a modeling-focused subteam of a larger department. From this team, the team leader, a senior researcher, two postdoctoral researchers, and one PhD-candidate were interviewed. All three teams have a clear modeling focus.

In total, three PhD-candidates, four postdoctoral researchers, four senior researchers, and three group leaders were interviewed, thereby capturing the spread in scientific positions. The sample contained four female modelers and 10 male modelers. Model experience varied from several months (recently started PhD-candidates) up to over 25 years (group leader). The median number of scientific publications of the interviewees up to and including the year 2020 was 19, with a minimum of zero and a maximum of 122 (and a mean of 33).

This study is exploratory. The aim is to in-depth explore certain patterns, which afterward should be tested for a larger sample size. Some saturation did occur during the coding of the interviews: very few new labels were defined for the last interviews because most motivations were already mentioned in the previous interviews. This suggests that the sample size is reasonable.

2.2. Semistructured In-Depth Interviews

The interviews were semistructured: all questions were prepared beforehand, but questions could be adapted in response to previous answers. The interview guide was not provided to the interviewees in advance. Before the interviews, the interviewees were asked to provide one of their recent scientific publications that describes a study for which they employed a hydrological model. In preparation of the interview, the interviewer read this paper in detail, and formulated questions specifically related to this study. This thus resulted in 14 interview guides, tailored to the study that each interviewee provided. For two PhD-candidates and one postdoctoral researcher, the interview was based on a draft manuscript.

The interview protocol is provided in the Supporting Information S1. The protocol was tested with two hydrological modelers before the 14 modelers selected for this study were interviewed. Based on this initial test, several questions were reformulated for clarity. The test interviews were not considered in the results.

The interviews always consisted of three parts. The first part of the interview focused on the position, background, and the experience of the modeler. Question belonging to this part were: *How would you describe your current academic position?* and *Can you describe your experience as a hydrological modeler?* This part of the interview usually took about 10 min.

The second part of the interview was specifically related to modeling decisions, and was based on the paper that the interviewee had provided. Questions belonging to this part of the interview were, for example, *What made you decide to use model [..] for this study?* and *You used the [..] downscaling technique. What made you decide to use this technique?* Some questions were asked in every interview, such as the model choice, while other questions were specific for the study at hand, such as the example question on the downscaling technique. This part of the interview usually lasted about 40 min.

In the third part of the interview, the interviewees were asked about their confidence in the model, and their perception of their influence on the model results. Furthermore, interviewees had the opportunity to add any remarks they deemed relevant. This part of the interview took about 10 min.

The shortest interview took 40 min, the longest interview 1 hr and 31 min. The total interview time for the 14 interviewees was 13.9 hr. All interviews were recorded with the consent of the interviewee, and completely transcribed after the interview. The transcriptions were sent back to the interviewees for approval. Sometimes, small additions and clarifications were added at this point. In total, 110,767 words (about 220 pages) of transcriptions were obtained.

2.3. Inductive Content Analysis

The transcribed interviews were subject to an inductive content analysis (Kyngäs, 2020; Weber, 1990). Inductive content analysis is used when limited prior knowledge is available: Concepts are defined during the analysis. This contrasts deductive content analysis, where a theory and the corresponding concepts are predefined and tested with the data. Inductive content analysis was chosen because no theories exist yet on how hydrological modelers make modeling decisions. The analysis was conducted using Atlas.ti 8 software.

Table 1
The Topics That Were Identified in the Interviews and Used to Organize the Data

Topic	Description	#
Confidence model	The confidence the modeler has in the model	14
Development of model use	How the use of models changed throughout personal career	11
Educational background	Educational background of the interviewee	14
Influence modeler	How the interviewee perceives own influence on model results	14
Model experience	Experience of the interviewee with hydrological models	14
Model use	How the interviewee perceives own model use; model user, model developer, etc.	3
Position	Professional position of the interviewee	14
Satisfied about model	Satisfaction of interviewee with the model	2
Study goal	Goal of the study	12
Decision calibration	Decisions related to calibration strategy—includes discussions on why no calibration was performed	13
Decision parameter selection	Decisions related to which parameters to optimize in calibration	3
Decision data use	Decisions related to which data to use—also includes decision on which catchment	14
Decision data use preprocessing	Decisions related to preprocessing procedure of data	10
Decision model implementation	Decisions related to how to implement model	3
Decision model selection	Decisions related to which model to use for the study	14
Decision model settings	Decisions related to specific model settings (i.e., turn on/off certain elements)	9
Decision objective function	Decisions related to objective function in calibration and evaluation	13
Decision simulation period	Decisions related to the period used for simulation	12
Decision spatial resolution	Decisions related to the spatial resolution of the model	10
Decision temporal resolution	Decisions related to the temporal resolution of the model	11
Evaluation KGE values	KGE values that interviewee considers reasonable	8
Evaluation NSE values	NSE values that interviewee considers reasonable	2
Evaluation other metrics	Other metrics for which interviewee considered reasonable values	1

Note. Topics relate to the background of the interviewee and the background of the study, to modeling decisions, and to estimates of the interviewee of what is considered a reasonable model performance. The topic labels were used to further analyze motivations for modeling decisions. Not every topic was covered in every interview (i.e., not in every study a KGE value was employed). The third column (#) indicates the number of interviews in which the topic was covered.

The first part of the analysis consisted of a topical classification of the data. Because the interviews were semi-structured, information related to specific topics could occur at multiple places throughout the interview. In the topical analysis, different parts of the interview belonging to the same topic were labeled. Topic labels were defined for topics that reoccurred frequently across the different interviews. An example of a topic label is "*model experience*"; this label was applied to all the interview parts that described the experience of the interviewee with models. In total, 22 different topic labels were defined (Table 1), but not all topics were covered in all interviews (Table 2).

The second part of the analysis consisted of the thematic classification of the data. Whereas the previous step relates to organizing the data, this step is the start of the actual content analysis. The interview transcripts were carefully read, and as soon as a motivation for a certain decision was mentioned, this text was labeled. If this motivation had been mentioned before, an already existing label was applied to the text. If a new motivation was found, a new label was defined. The labels were defined very close to the original text. An example of a response from the interviewees was: "*Classical decision, that is what is typically used to solve it, to solve this equation, so we stick to it*." The label added to this text was "*typically used*." Another example is "*We knew that this existed so that's why we used that*," which received the label "*knew existence*." After evaluating the 14 interviews, 106 different labels were identified for 727 different pieces of text. This represents an average of 52 thematic labels related to modeling decisions per interview, with a minimum of 32 and a maximum of 81 across the 14 interviews.

Table 2
The 83 Labels That Were Used to Identify Different Motivations

Classified label (total #, # of motivations)	Definition	#
External party (12, 3)		
Funding agency	Funding agency determines/allows/suggests certain decisions	9
Project end-users	End-users had certain research interests that steered direction	2
Politics	Politics played a role (for instance in which data was available, e.g., in climate projections)	1
Broader scientific community (131, 15)		
Literature	Based on the literature	33
Data availability	The data is available	29
Data quality	The data has good quality	15
Widely used	This is just the common, popular way of doing this	14
Data accessibility	The data is accessible	13
Data resolution	This was done because of the resolution of the input data	7
Discussion external developer	Discussed with the developer of the model/tool (no collaborator/coauthor)	4
Popular in another community	This is coming from another research community where it is very popular	4
Typically used	This is what is typically used	3
Best available data	This was the best available data set (at that time)	2
Reviewer asked	Something was done because the reviewer asked	2
School knowledge	This is how it is taught at school or common knowledge (e.g., specific numbers)	2
Competition	This was done to provide competition with another institute/project	1
Good documentation data	There is good documentation of these data	1
Part of community	A larger community is working with these data/tools/models	1
Scientific collaborator (38, 6)		
Discussion author team	It was discussed in the author team and this was the decision	18
Experience project partners	Decision based on experience of scientific project partners that joined the study	6
Experience external colleague	Based on experience from collaborator	5
Developed there	It was developed at the institute of coauthor/collaborator	4
Modeling protocol	It was described in a protocol	3
Research stay	Colleague or modeler did a research stay in or moved to another group	2
Institute (32, 5)		
Developed here	It was developed here in this team (also “institute” because can be more general than team)	10
Computer power	Available computer power/resources—includes memory, storage and run-time	6
Origin data	The origin of the data was relevant in relation to origin institute—for example, whether it was produced in US or in EU	6
Agreement	There is an agreement about access to data	5
Infrastructure	There is infrastructure that facilitates the use of these data/tools/models	5

Table 2
Continued

Classified label (total #, # of motivations)	Definition	#
Team (223, 19)		
Experience colleagues	Based on experience of direct colleagues	67
Previous work internal	This was shown in previous work from inside organization	23
Discussion author team	It was discussed in the author team and this came out	18
Supervisor	Supervisor decided or recommended it (in a pupil-mentor context)	17
Set-up was available already	This set-up (or the simulations) was already available from earlier work	14
Default use in the team	This is standard use in this team	13
Default model setting	Settings come along with the model set-up (also "Team" when developed there)	11
Developed here	It was developed here in this team (model and codes)	10
Default model tool	This tool is default to this model (also "Team" when developed there)	9
Not exactly known	Not exactly known how/why colleague made a certain choice	9
Experience supervisor	Based on experience from the supervisor (explicitly mentioned supervisor)	6
Available set-up was starting point	The old version was the starting point from where new improvements were made	5
Code availability	Codes for preprocessing and postprocessing of the model were available	4
Comparison previous work internal	This was done for comparison to earlier work within the group	4
Set-up will be used in follow-up study	This set-up will be used in future studies and therefore these settings were already chosen	4
Heritage	Heritage in the group from experienced people	3
Confidence in colleague	Confidence/trust in colleague	2
Script availability	A script was available to do this	2
Team leader	Team leader decided it (at a higher general level than individual supervisor)	2
Individual (246, 30)		
Individual—Personal (123, 16)		
Personal experience	Personal experience with this method/model/code/these data	34
Personal judgment	Personal expert judgment	34
Model performance was good	The model performance was good (enough)	13
Specific model aspects	Certain (conceptual) aspects of the model made that it was selected	13
Not the aim	That was not done because it was not the aim of this study	5
Personal interest	Personal interest is more in this direction (e.g., where to focus on in model development)	4
Confidence in model	The modeler has confidence in this model	3
Know the region	A region was selected because modeler knows this region	3
Like it	Personal preference, the modeler likes this method/tool/approach, without any further reason	3

10.1029/2021WR030600

Table 2
Continued

Classified label (total #, # of motivations)	Definition	#
Data requirements	Model/method was used because it had low data requirements	2
Good documentation model	This model is well documented	2
Good documentation tool	There is good documentation of this tool (e.g., calibration algorithm)	2
Method based on a lot of data	This method was selected because it is based on many data points/most data	2
Knew existence	It was used because modeler knew that it existed	1
Not well documented	Model is not well documented	1
Open source	Something was done because it could be shared/open/made available to others	1
Individual—Interaction model (51, 5)		
Run time	Run time was an argument to make a decision (does not incl. memory/storage)	21
Tested	Different options were tested and this one was chosen	16
Method lead to best model performance	This method was selected because it leads to higher model performance	6
Trial and error	Ad hoc like testing—trial and error	5
Visual inspection	Based on visual inspection	3
Individual—Pragmatic reasons (57, 7)		
Run time	Run time was an argument to make a decision	21
Lack of time	Something is not done because of time constraint	13
Pragmatism	To keep it practical	9
Time efficiency	Motivation for a decision based on efficiency in time; usually a trade-off with effort	9
Effort	It was a trade-off with effort	2
Managed to run	Model was used because modeler managed to run it	2
Not user friendly	This was not used because it was difficult to use/not user friendly	1
Individual—Nontraceable (36, 4)		
No real reason	Choices were made ad hoc, arbitrarily	17
Not exactly known	Not exactly known how/why colleague made a certain choice	9
Don't know	Not aware of what was exactly done to data/model that was used	6
Can't remember	Can't remember the reason for something anymore	4
Consequential (45, 10)		
Default model setting	Settings come along with the model set-up	11
Default model tool	This tool is default to this model (generally, calibration algorithm)	9
Consistency	To stay consistent with the model set-up	6
Available set-up was starting point	The old version was the starting point from where new improvements were made	5
Comparison previous work external	This was done for comparison to previous work in the literature	4
Comparison previous work internal	This was done for comparison to earlier work within the group	4

Table 2
Continued

Classified label (total #, # of motivations)	Definition	#
Comparison between models	This was done to fairly compare two models	2
Default tool setting	This is default setting of this tool	2
Not compatible	Other method was not compatible with model	1
Only option	It was the only option modeler had (e.g., given data gaps)	1

Note. The labels have been grouped into seven classes and four subclasses. Labels can be assigned to several classes. The frequency of occurrence of each label is indicated in the last column. The total occurrence of the different labels within the class and the number of labels is indicated as well (see also Figures 1 and 2).

The next step was to evaluate the 106 different labels and revise and group them. For example, when comparing the texts belonging to the label "*ad hoc like testing*" and the label "*trial and error*," it appeared that both labels referred to the same motivation and the labels were merged. Other labels were split. For example, the label "*for comparison to earlier work*" was split into "*for comparison to earlier work internal*," comparing it to work that was done in the same team, and "*for comparison to earlier work external*"—something was done in a certain way to be able to compare it to another method or approach used by others or described in the literature. After the evaluation, 83 different labels were left that each represent a different motivation (Table 2).

The final step was a classification of the labels. Labels that relate to each other were grouped in a class and the class was named after the overarching theme of the labels. The labels "*funding agency*," "*politics*," and "*project end-users*" all refer to external parties outside of academia and their class was therefore named "*External party*." Sometimes, labels related to multiple classes. For example, the label "*discussion author team*," that indicates that something was decided based on a discussion with the author team, belongs both to the class "*Scientific collaborator*" and the class "*Team*," because it depends on whether the authors of the study were only direct colleagues or also colleagues from other institutes.

The classification of the labels is an interpretation of the data. There are, however, also existing frameworks about values in science that might be applicable for analyzing the interviews. A relevant theory is the epistemic versus contextual values in science approach, introduced by McMullin (1982). This will be further explored in Section 3.4.

2.4. Scientific Stance of the Interviewer

The background of the interviewer can determine the direction of the interview, and therefore the results of the interview. This section shortly describes the scientific stance and the scientific background of the interviewer.

The interviewer has a scientific background in hydrological modeling, both with land-surface models and conceptual hydrological models, with a focus on uncertainty estimation. As a result of this focus, the interviewer has developed a scientific vision based on constructivism—which led to the initiation of this interview study. The technical background of the interviewer can have biased the interview questions. The interviewer, for instance, has ample experience with calibration strategies, and therefore, questions and responses to the answers related to calibration could be very detailed. On the other hand, the interviewer has little or no experience with (real-time) data assimilation. Questions and discussions related to data assimilation might therefore not have reached the same depth as questions and discussions related to calibration.

3. Results and Discussion

3.1. Framework of Motivations for Modeling Decisions

The 86 different motivations for modeling decisions across the 14 interviews were grouped into seven classes: *External party*, *Broader Scientific Community*, *Scientific Collaborator*, *Institute*, *Team*, *Individual*, and *Consequential*. Each class includes several motivations, as indicated in Figure 1 and Table 2.

The class *External party* contains motivations related to external parties outside of science. Only three motivations are associated to this class (Figure 1 and Table 2). The funding agency, as external party, can, for, instance suggest the use of certain data (I is interviewer, P3 is interviewee number 3):

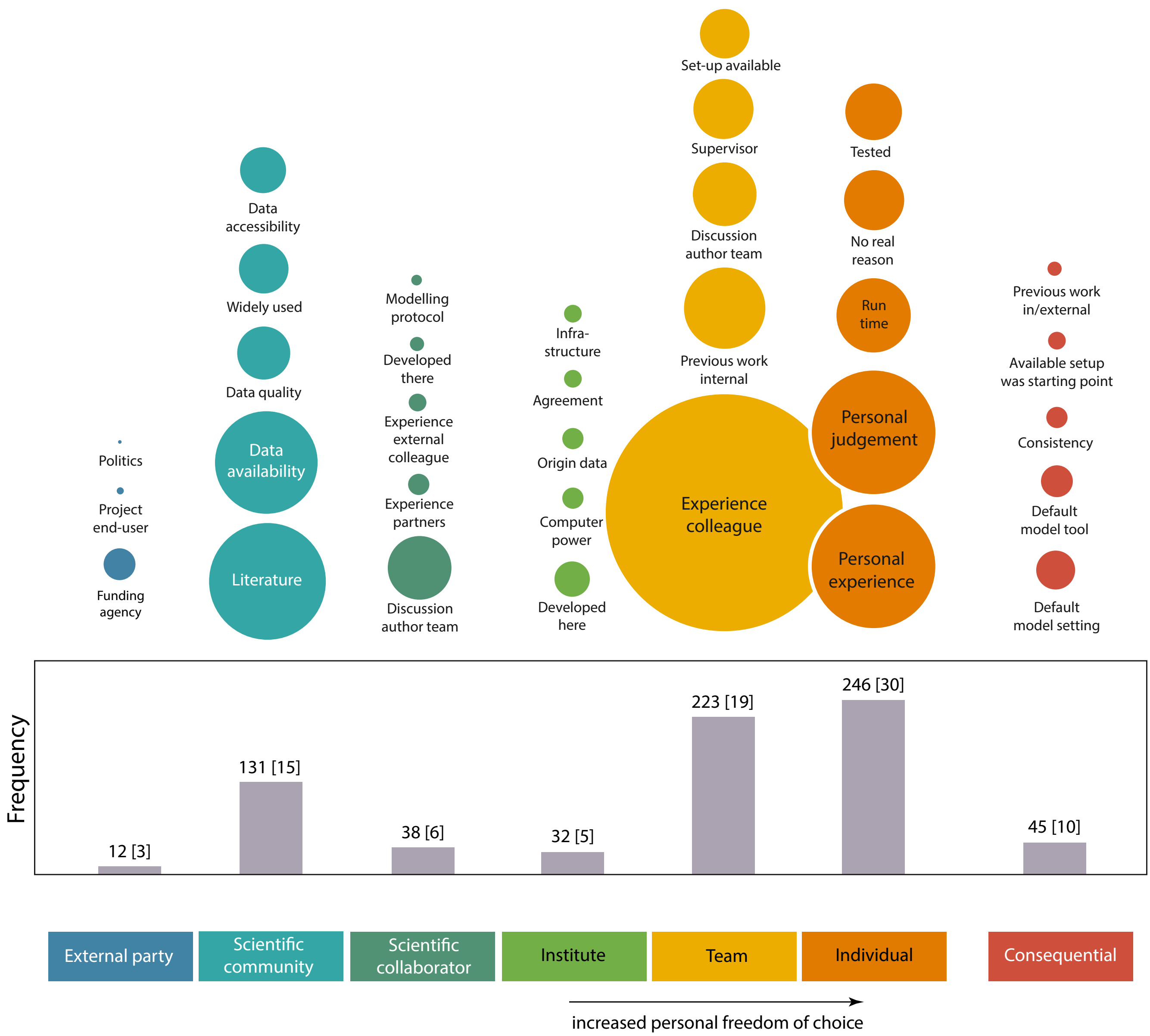

Figure 1. Framework of classified motivations to make modeling decisions. Modeling decisions can be classified as being steered, decided, or prescribed by external parties, the scientific community, collaborators, the institute, the team, or made by the modeler themselves (Individual). Consequential decisions demonstrate a path dependency: certain decisions are prescribed by, or the consequence of earlier decisions. The circles show the five most frequently mentioned motivations per class, with the size proportional to how often this motivation was mentioned in the interviews and their color according to the classes to which they belong. An overview of all motivations can be found in Table 2). The bars indicate how often a reason belonging to a certain class was mentioned in the interviews. The first number above the bar gives the absolute count (for instance, a reason that belongs to the *External party* class was mentioned 12 times across all interviews), the number between the square brackets gives the number of reasons that were assigned to a class (i.e., three different reasons have been related to the class *External party*).

> I: "But why not use this one directly then? Why then still use the [data set]?"
> P3: "Because the [data set] has to be used."
> I: "Because that's what [funding agency] told you?"
> P3: "Well, it was just suggested that we should use it. [..] They say, we have the data but it is incomplete, but use as much as you can, okay? Okay."

The class *Broader scientific community* contains motivations that can be related to the current paradigms in the hydrologic community (e.g., *typically used*, *widely used*, *school knowledge*), the availability of data (e.g., *data* *availability*, *data quality*), and the literature. Fifteen different motivations were assigned to this class (Table 2). From all motivations in this class, literature was most frequently mentioned: 33 times across 12 interviews. For instance, in the following quote:

> "Of course, we went through the whole literature. We looked into all the possible [functions] that we could possibly find, review papers, all that stuff." [P12]

However, 18 out of 33 times literature is mentioned in combination with other motivations, such as colleagues:

> ".. there is an option for what optimization method I want to use, and then I read some papers and I talk with some of the colleagues." [P1]

The class *Scientific collaborator* refers to motivations that come from collaborations within the scientific field but outside the institute where the modeler is working. Six different motivations were assigned to this class (Table 2). Most frequently mentioned was *Discussion author team* (18 times across seven interviews), where a modeling decision was based on a discussion with the authors of the paper (this motivation was also assigned to class *Team* because it depends on the composition of the author team). Related to *Discussion author team* was the motivation *Research stay*: The modeler did a research stay at another institute, which fostered the collaboration and led to a joint publication.

The class *Institute* mainly relates to facilities provided by the institute where the modeler works. Five motivations are assigned to this class (Table 2). Motivations relate for instance to computer infrastructure, computer power (having access to a high-performance cluster), and agreements that the institute might have with certain data-providers:

> P7: "At least, in [country], [data-provider] puts a lot of restrictions, so we can have it now, but it was not the case before."
> I: "You have an agreement now."
> P7: "Yeah, we have an agreement. So data accessibility is also an issue."

The class *Team* represents motivations that are related to colleagues, group leaders and supervisors, and to scripts, model set-up, and code availability within the team. In total, 19 different motivations were assigned to this class (Table 2), and with 223 text labels related to these motivations, this is the second most frequently mentioned class after *Individual*. By far most frequently mentioned across all identified motivations is *Experience colleagues*, 67 times in 13 out of 14 interviews. This motivation appears in many different forms and shapes related to many different modeling decisions; for example, programming language, data use, quality control, and modeling vision. A few examples are provided below:

> "We used [programming language] because, like many people in the team use [programming language] and they have developed packages, so.. " [P6]
> ".. when you're not born a modeler, you need other people's support, and it's much easier if the support is given by people next door, and I had very experienced modelers at that time in our group when I joined it, and they were very keen in bringing support" [P7]
> "Yeah, because he has done the calibration in the department for many years, and he has developed lots of codes to extract river basin information." [P2]

Another interesting motivation belonging to the *Team* class is *Set-up was available already*. A specific model, with a specific set-up, was used simply because it was there already. Many of the decisions related to the configuration of the model, such as the spatial resolution, are then not made by the one doing the study, but by a colleague who made the set-up for a previous study. Therefore, this motivation very much relates to *Experience colleagues*:

> "So I don't know, I don't know why my colleagues did that, but they did it. They just, I mean, we had these two set-ups, and I was working on this model development study, and so for me it was a great opportunity to use both and why they did it… I don't know." [P5]

This last quote was assigned both to *Set-up was available already* and *Not exactly known*.

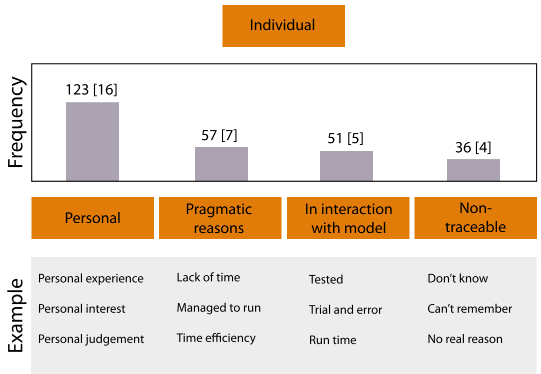

Figure 2. The class *Individual* can be subdivided into four subclasses. Several examples are provided for each subclass. The bars indicate how often a reason belonging to a certain subclass was mentioned in the interviews. The first number above the bar gives the absolute count, the number between the square brackets gives the number of different motivations that were assigned to this subclass.

For the motivations in the *Team* class, it is relevant to be aware that the interviewed modelers all work in a team dedicated to modeling. This can enhance the frequency of relying on experience of colleagues in the modeling process. This is further discussed in Section 4.1.

The class *Individual* includes motivations that are related to the individual modeler. This class contains 30 motivations (Table 2), which can be further subdivided into four subclasses (Figure 2): *Personal*, *Pragmatic reasons*, *In interaction with the model*, and *Nontraceable*. Most frequently occurring are the *Personal* motivations, especially *personal experience* and *personal judgment*. Personal experience is often used in the same way that experience from colleagues is used for modeling decisions:

> "It was basically what I used before, a couple of years ago. So it was more like, I had the experience with something, I already have the code, so it was easy for me to just apply it." [P1]

Expert judgment was very often related to the modeler's perception that something was or was not relevant to consider, such as shown in this quote:

> "I don't think that the results of the study do depend on that choice. I have no proof for that, haha. But that's how it is." [P5]

But expert judgment was sometimes also a motivation to make adaptations, for instance to parameter values:

> I: "Were these numbers really the average from all the data you had or did you do some spatial correction or, again, some expert judgment correction?"
> P13: "There was no spatial correction. Definitely not. It was expert judgment based on available data."

Since the modeler is the one who actually runs the model, there is a specific subclass *In interaction with the model*. This relates to motivations that require that the model is run. For instance, something was tested with the model, with trial and error a certain decision was made, or the model run-time appeared to be too long. Run-time was both classified in the subclass *In interaction with model* and subclass *Pragmatic reasons*. Under pragmatic reasons, there are many other motivations related to time constraints (something was done for time efficiency, or not done due to a lack of time). The last subclass, *Nontraceable*, are answers that indicate that the underlying motivation cannot be traced back. Five percent of all motivations were non-traceable. This reveals that there are some motivations that are not documented in the paper, and that can also not be traced back by directly asking the modeler.

The final class is *Consequential*. This relates to choices that are the consequence of decisions that were made earlier and demonstrates the path dependency in hydrological modeling (Lahtinen et al., 2017). Consequential motivations can again be subdivided into two subclasses. On the one hand, motivations can be rather stringent, such as *only option* and *not compatible*:

> "I cannot say it's justified to use an average value but that's the only option we had." [P12]

On the other hand, motivations can be consequential but not stringent, for instance related to using default options (*default model setting*, *default model tool*, *default tool setting*). It is not strictly necessary to use the default calibration algorithm that comes with a model, but still this default algorithm will often be used, and depends on the model selected earlier.

There was one motivation that could not be classified in this framework, and that is *Vision*. This is a very broad motivation and relates to the modeling vision or modeling philosophy of the modeler. This vision is probably the result of personal experience, personal interest, personal judgment, the team and the supervisor, and the broader scientific community. This motivation will be further discussed in Section 3.4.

Overall, the classes *Team* and *Individual* are by far the most frequently occurring classes off motivations for certain modeling decisions (Figure 1). What the presented framework undeniably shows is that modeling is inherently a

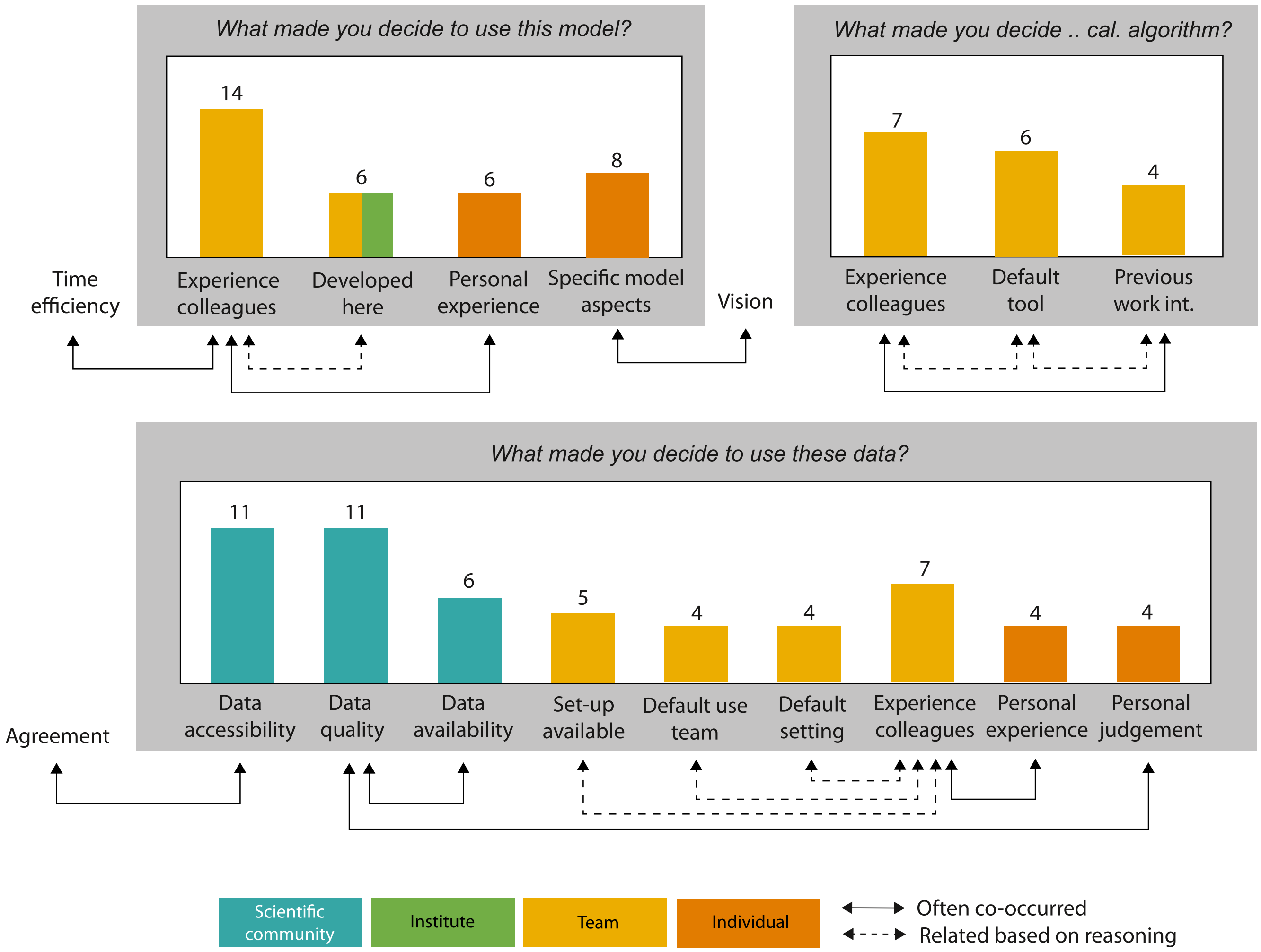

Figure 3. The motivations for three model choices: model selection, data selection, and calibration algorithm selection. The bars indicate the frequency that this motivation was mentioned. Motivations connected with a solid arrow often cooccurred in the same sentence. Motivations connected through a dashed arrow can be connected based on reasoning (see text). Motivations outside the gray panels but connected through a solid line often cooccurred, but not in the context of this specific decision.

social process, where many more factors than the expertise of the modelers themselves play a large role. Although the modeler has personal freedom, modeling is not a solitary activity, but an activity that is influenced by local context. The role of the direct surrounding of the modeler (colleagues and the available facilities and infrastructure) and the social embeddedness of modeling decisions demonstrates that it takes a village to run a model.

3.2. Modeling Decisions

In this section, three modeling decisions are analysed in more detail, to demonstrate how different motivations for a decision can interact and co-occur. The decisions are: model choice ("*What made you decide to use this model?*"), data choice ("*What made you decide to use these data?*") and the choice for a certain calibration algorithm ("*What made you decide to use this calibration algorithm?*"). The first two decisions, model and data choice, were raised in all interviews. The choice of calibration algorithm was only relevant in 11 out of 14 interviews, in the three other studies no calibration was performed.

3.2.1. Model Selection

As demonstrated in the top left panel of Figure 3, *Experience colleagues* was the most frequently mentioned motivation for the decision to use a specific model (in 8 out of 14 interviews). The experience of colleagues with a specific model was often related to personal experience with the model, and if a model was selected because it was developed "here" (i.e., at the institute or in the group where the modeler works), it can be expected that there are colleagues that have experience with this model. Motivations at institute, team, and personal level thus come together. Experience seems a very important criterion to select a model, as was hypothesized by Addor and Melsen (2019). Outside the context of model selection, *Experience colleagues* is often mentioned together with *Time efficiency*: it is efficient to use something that colleagues have experience with. Time efficiency was not explicitly mentioned in the model selection context, but can implicitly be related.

The second most frequently mentioned motivation for model selection was *Specific model aspects* (in 5 out of 14 interviews): The model was selected because it has a certain asset according to the modeler. This asset can for instance be that the model has a real-time data-assimilation framework, or that it has a certain scaling concept that appeals to the modeler. Very often, these specific model aspects were related to more general vision of the modeler on hydrological modeling. An example of a vision is:

> "I think we are stepping into a new era of hydrological modeling, which is good because it was not there, perhaps ten years ago. It will provide new features that the society can benefit from. And another part is very high resolution modeling, which was not available couple of years ago, going up to 1 km." [P2]

This modeler perceives spatial resolution as very important, and therefore selected a model which can run at very high resolution relative to the spatial coverage.

There seems to be a dichotomy between the more pragmatic experience-based motivation and the vision-based or model-aspect motivation to select a model. This is, however, not necessarily the case. First, there is a clear role for recruitment here. Two modelers from different institutes mentioned in the interview that they applied for a job at that institute because they appreciated the modeling approach. Recruitment leads to the alignment of the vision between colleagues so that experience of colleagues and vision become intertwined. The modeling vision can also be developed while working at a certain institute and can become internalized, as described in Section 3.3. The apparent dichotomy between the pragmatic and vision-based motivation thus hides underlying systems, related to recruitment and internalization, that connect these motivations.

Note that *none* of the interviewees directly related the model choice to the research question of the study. One of the interviewees formulated it this way:

> "Yeah, yeah, we have the tools, we have the experience, and then we start asking questions" [P11]

showing that it is not the research question that is driving model selection, but that available tools (models but also methods) and experience, that together represent the expertise of the team or institute, determines which research questions are addressed.

3.2.2. Data Selection

Data requirements differed substantially between the different model studies of the interviewees. Data requirements vary from input forcing data, to land use, soil or other spatial catchment data, to observation data for validation. The most frequently mentioned motivations are that data is used because it is accessible and because it has good quality (Figure 3). But also here, experience plays a role. Data quality was often related to personal experience and personal judgment. Experience plays a role in two different ways: having experience with the data itself, and having experience with processing the data.

> "We have used it [these data] repeatedly and we know that it's quite good." [P7]

This quote refers to having good experience with the data itself, the quality assessment of the data is based on experience. In the following quote, experience in relation to data is relevant for the processing of the data:

> "[…] they are easily available in our research team because we are used to use this kind of data for most applications, so we can process this kind of data very easily." [P8]

Many other team-related motivations were mentioned, such as *Set-up was available already*. This means that a model set-up from a previous study was used. The modeler did not make any decisions related to data use, but

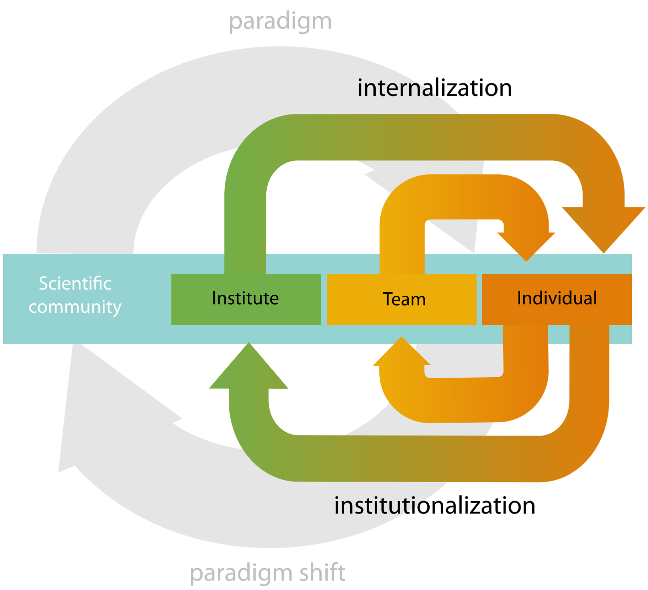

Figure 4. Visualization of the processes of internalization and institutionalization. Modeling decisions imposed by the institute or the research team on the individual, for example, through the available modeling ecosystem, eventually get internalized by the individual and become a personal preference or choice. Vice versa, dependent on the position of the individual, personal preferences and choices can become institutionalized. The preferences of, for instance, the group leader can become standard use in the team. These processes take place in the context of the broader scientific community of which the institute, research team, and individual modeler are all part of. Common methods in the scientific community can become internalized, and strategies of a certain team or institute can become widely used in the scientific community.

adopted all the decisions from the previous set-up, thereby relying on the colleague who made these decisions. The same applies to making use of default model settings and default data that are used within the team.

Often cooccurring with *Accessibility* is *Agreement* (5 out of 14 interviews). This means that there is an agreement between the institute and a certain data-provider about access to the data. Such agreements can determine which data are used.

Similar to model choice, also for data selection two pillars seem to underlay the decision: on the one hand experience, and on the other hand data accessibility, quality, and availability. These pillars are connected: quality assessment can be related to experience, and accessibility can be related to agreements of the institute.

3.2.3. Calibration Algorithm

The decision for the calibration algorithm is much more straight forward than for the other two decisions: all motivations are classified at the team level and can be related (Figure 3). Often co-occurring are *Previous work internal* and *Experience colleagues*. *Previous work internal* in this context indicates that colleagues investigated different calibration algorithms, which led to a preference at the team level. This can be the default implementation of a certain algorithm in the model ecosystem (*Default model tool*).

So while one would be expected to think more about run time, efficiency, risk of ending in local minima et cetera when selecting a calibration algorithm, these kind of trade-offs and evaluations were usually already made earlier by colleagues, and this was not reconsidered in new studies:

> I: “So you also checked other algorithms?”
> P7: “Not for this study, let's say that's part of our old, let's say, base of tests, where we did a lot of comparisons.”

This fits in the perspective that science is incremental, and new studies build on the knowledge gained in earlier studies. From that perspective it is remarkable, however, that this incremental building of knowledge seems to a large extent tight to the research team. The results of the old base of tests on calibration algorithms, referred to in the quote, are for instance only published in an internal report. Another comparison that was done within this team and that has led to the preference for a specific potential evaporation equation, was published in peer-reviewed scientific literature, but this study was not picked up by the other two teams—which can be understood given the large number of studies and publications available on this topic. Besides, much of the created knowledge related to model decisions is specific to the modeling ecosystem of the team that did the investigation, that is, given certain forcing products or certain model settings. The result is that each team does its own analyses and sticks to its own preferences.

3.3. Internalization and Institutionalization

The processes of internalization and institutionalization became apparent when evaluating the motivations and the decision framework. These processes are depicted in Figure 4 and discussed in more detail in this section.

The process of institutionalization in hydrological modeling context starts with an individual that uses a certain method or approach, for instance, because this person has experience with it, for example, from another job. Other team members make use of the experience of this individual, and as such it can become standard use in the team. If it then becomes the default tool in the modeling ecosystem, the method or approach is institutionalized. This process can also work the other way around. An individual uses the default tool that is available in the team. Eventually, this individual starts to defend this tool as if it were the individuals own decision to use it: The method or approach is internalized.

Several cycles of institutionalization and internalization could be recognized by interviewing members of the same team with different positions and experience. An example is the use of a certain downscaling technique:

10.1029/2021WR030600

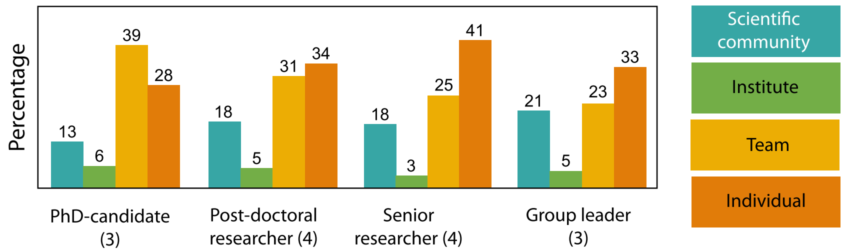

Figure 5. The percentage of motivations across the different classes, split out for the position of the interviewee. Three PhD-candidates, four postdocs, four senior researchers, and three group leaders were interviewed. Only the classes that represent a substantial part of the motivations are displayed.

> P3: "Yes, so, downscaling, how you do that thing? So you can again play with many things, but the [method] was developed here."
> I: "You developed it here?"
> P3: "Yes, we have our own code. I wrote this from my old times in [city 1] and [city 2]. When I came here it was the first thing I did because I needed to prepare the data for [model]."

This code has become part of the modeling ecosystem and as such became institutionalized, as noticed by another interviewee:

> "Most of the preprocessing that is required to run the model was covered by the infrastructure, the scripts and the database. So, the flow is there, and then I have to do nothing but to press a few enters and then that is done." [P1]

In internalization, this default option is defended as being the best option (the use of passive voice in the quote of P2 suggests that internalization is not completely accomplished yet):

> "Yeah, so. It has been decided to use [method] for interpolation. There are other options, but it is assumed that adding the [variable] will be more realistic." [P2]

The cycles of institutionalization and internalization take place at the institute/team/individual level. However, the institute, team and the individual are also part of a scientific community. Internalization of common procedures within the scientific community are often referred to as paradigms. Institutionalization at the level of the scientific community, that is, the procedure of an individual or a team becomes the new scientific community standard, requires a paradigm shift. This is generally harder to achieve than institutionalization of a procedure at the institute/team level. It can even become controversial, Lloyd and Oreskes (2018) discuss how the climate community responded to the proposal of a new method for climate change attribution, from fraction of attributable risk to storylines. The newly proposed method was accepted by the team of the initiator, but was perceived controversial at first by the broader climate community.

The role of internalization and institutionalization could be related to the experience and position of the modeler in the team. In that case, it should be visible in the distribution of motivations across the different categories; more individually motivated and less team-motivated decisions for more experienced modelers. To test this, the percentage of modeling decisions related to a specific class was split out by research position, where position is assumed to represent modeling experience (Figure 5).

Especially the role of team-related decisions decreased with increasing experience—from 39% of the motivations being team-related for a PhD-candidate to 23% team-related motivations for a group leader. The individually motivated decisions increase up to senior researcher, but decrease again for the group leader. The results suggest that there is indeed a relation between position, used as a proxy for experience, and the level of team dependency, which might be the result of institutionalization processes. This should be studied with a larger sample of investigated modelers to confirm this hypothesis.

3.4. Epistemic Versus Contextual Values

The classes described in Section 3.1 were determined in an inductive manner. There are, however, frameworks available for the classification of scientific choices—although not specifically developed for hydrological modeling. A particularly relevant framework is presented by McMullin (1982), and is explored further in this section.

McMullin (1982) acknowledges value-judgment to be an explicit part of science, but makes a distinction between two types of values: epistemic values and contextual values. Epistemic values are supposed to support the truth-like character of science. They relate to the value-judgment of a theory: The theory is assessed based on the epistemology of the modeler. This is the motivation *Vision*, that was not assigned to any class yet. Examples of epistemic values in the interviews were:

> "The main kind of approach in the [..] developing team is to keep things simple" [P1]
> "There is a sentence from Einstein: As simple as possible but not simpler. And I prefer this approach […]" [P10]

These two quotes come from interviewees that perceived a simple model as better than a complex model, where the second quote also implicitly includes a judgment on performance (".. *but not simpler*"; Note that different definitions of epistemic values exist, and that "simplicity" is not necessarily recognized as an epistemic value by McMullin (2009), but is acknowledged as such by Lycan (1985). Simplicity is thus not generally recognized as epistemic value, even though it is a widely applied hydrological modeling philosophy, as also demonstrated by the quotes from the interview). A contrasting epistemic value, that was not encountered in the interviews, is that all processes should be included in the model: the model should be as complete as possible. A modeler following this epistemology might for example, choose to include a certain process in the model, even if it is not exactly understood yet how this process works, because this modeler believes that this is the best way to discover the truth. See also Hrachowitz and Clark (2017) for a discussion on contrasting modeling philosophies.

Contextual values relate to moral, personal, social, political, and cultural values (Reiss & Sprenger, 2017). An example is that gender bias can be introduced in hydrological modeling because of a culture in which men are more involved in the use and development of models, as described in Packett et al. (2020). In this study, the definition of contextual values is widened to any local context in which the model is set-up and run. This also includes pragmatic project related boundaries, such as time constraints.

All 83 motivations were reclassified as being either epistemic or contextual: 74 motivations were classified as contextual, and seven as epistemic. Two motivations (*literature* and *consistency*) were classified as both because it was not directly clear where they belong. The unequal distribution of motivations over the two types of values (Figure 6), combined with the relative frequency of the motivations, demonstrates that contextual values dominate the modeling process. This is not necessarily bad. Without considering feasibility (e.g., run time, computer power) it would be hard to finish a model study. Furthermore, experience, both personal and from colleagues, is a valuable asset that can improve the quality of the modeling process. Experience can help in identifying spurious results and their source, and in fixing bugs in the code:

> "I know [model], like, very good. All 40,000 lines of it" [P5]

Experience with a model can also avoid that models are used for goals for which they were not developed. At the same time model experience also increases the risk of unintentional model use: Experience with a model may cause it to become the default model, so that the question regarding the most appropriate model is no longer asked and the model is applied to questions for which it was not developed.

Besides, experience also plays a role in value-judgment, as argued by McMullin (1982): ".. *let me recall how the skills of epistemic value-judgment are learned. Apprentice scientists learn them not from a method book but from watching others exercise them.*" And this is where experience from colleagues (classified as contextual) starts to intervene with the vision of a modeler (classified as epistemic): Modelers develop a modeling vision over time, inspired by their colleagues and surrounding. As such, modeling vision is also influenced by the processes of internalization and institutionalization as discussed in the previous section. The value-judgment of young modelers will mature over time, and might eventually challenge the practice of colleagues or the scientific community (Kuhn, 1962 (reprint 2012); Polanyi, 1958).

10.1029/2021WR030600

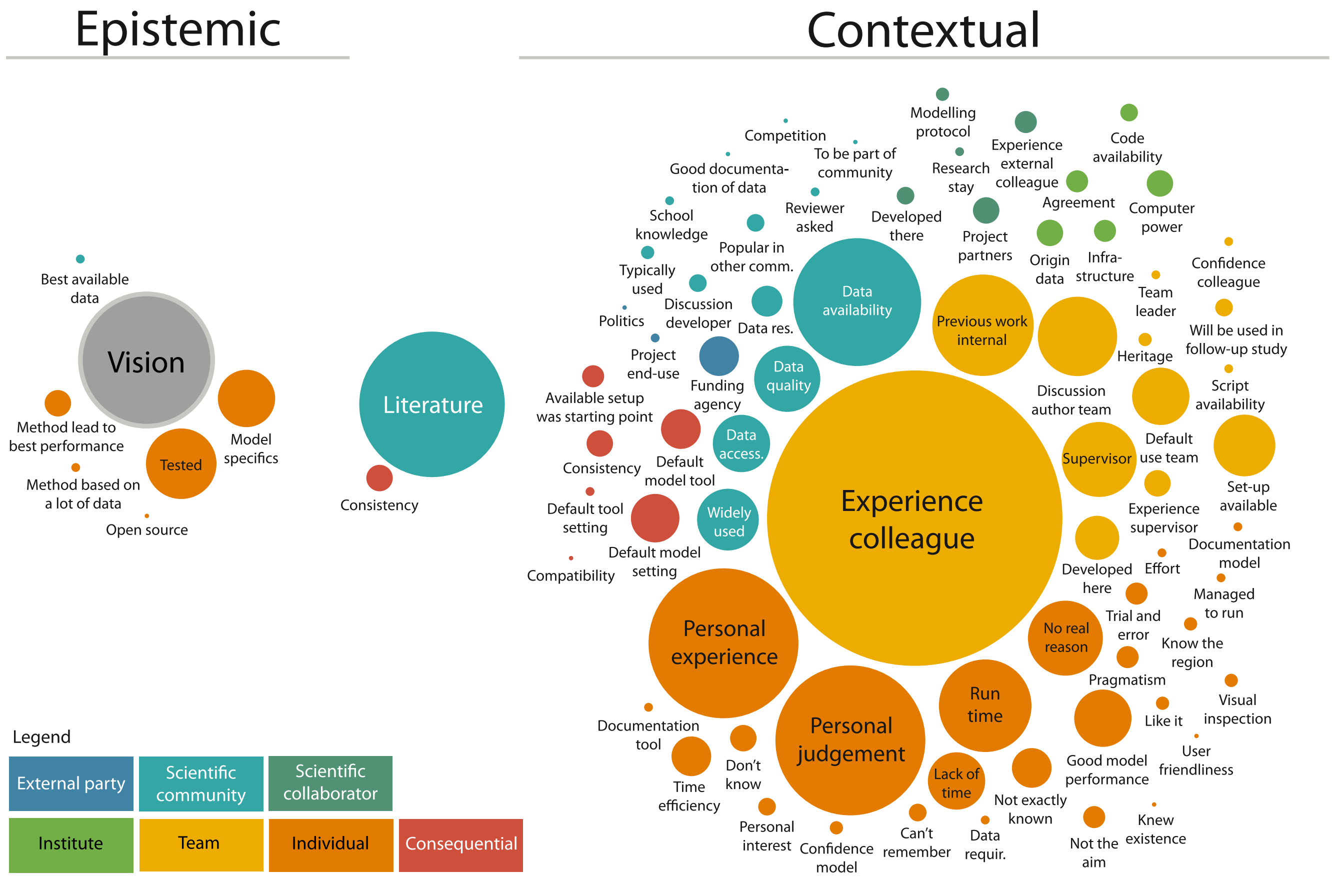

Figure 6. Division of the identified model motivations into epistemic and contextual motivations, as defined and discussed in McMullin (1982): epistemic motivations relate to the truth-like character of science, while contextual motivations depend on values, biases, and local context. The dot size is proportional to the frequency that this motivation was encountered in the interviews. The colors align with the classes presented in Figure 1. The motivation "Vision" did not fit any of the classes, "Literature" and "Consistency" are placed in the middle because no clear choice between epistemic and contextual could be made for these motivations.

It appears that model studies are to a large extent context dependent, rather than epistemically driven—although both are connected through vision development and experience. This does not disqualify the scientific value of model studies; many context dependent decisions are simply necessary in order to be able to run a model. But it is important to acknowledge that these contextual decisions will steer the model results.

4. Discussion

4.1. Representativeness of the Sample

The goal population of the interviews were scientific hydrological modelers. A geographic focus on western Europe was chosen for feasibility and to constrain the total population size. The interviewees were all part of modeling-focused research teams, of which two teams, the ones associated with national research institutes, were rather large (>10 people, although not all with a permanent position). This might have biased the results. Modeling-focused teams at universities might be smaller, and hydrological modelers can also be the only ones in the group or department working on modeling. This gives them less opportunities to rely on, for instance, experience from colleagues.

The average number of times that "*Experience from colleagues*" was mentioned as motivation in the interviews with the modelers from the two national research institutes was 7.0 and 5.6, respectively, while from the three interviewees working at a university, "Experience from colleagues" was on average only mentioned 1.3 times. One of the interviewees who worked at a national research institute at the time of the interview, worked at a small university department before. The modeling study to which the interview was related was conducted while the interviewee was still working at this department. "Experience from colleagues" was only mentioned once during this interview, compared to the 7 times on average over all team members of this interviewee. In contrast, "Personal judgment" was mentioned 6 times by this person, in comparison to 1.4 times for the team average. These examples demonstrate that the team composition and the environment in which the modeler functions influences the weight of the different motivations. There may also be motivations that were not identified within the current sample. Modelers that are part of smaller modeling groups, individual modelers, or modeling groups that are not focused around one specific model(family) are underrepresented in the presented results.

Furthermore, this study took an Eulerian perspective: modelers were interviewed about motivations for modeling decisions, given their current scientific viewpoint. As such, this study provides a snapshot of the current situation of the interviewees, while modeling vision for instance, is typically something that evolves over time. Using a Lagrangian perspective, tracking the development of a modeler over time, could provide more insights into processes such as internalization, institutionalization, and vision development.

4.2. Models as Social Constructs

A common scientific stance in hydrology, and in natural sciences in general, is a logical positivist view point. The idea behind positivism is that knowledge is obtained from observations, with the goal to determine universal laws (Carnap, 1966). This is also how model development is generally perceived: based on observations, we formulate a hypothesis of how the process works (Savenije, 2009), which can be translated into a numerical model. The existence of multiple modeling philosophies and approaches (e.g., Best et al., 2015; Bouaziz et al., 2021; Hrachowitz & Clark, 2017) already challenges the idea of the identification of universal laws. The existence of multiple visions is assigned to the incomplete understanding of processes (Clark et al., 2011), demonstrating epistemological uncertainty. This "border with ignorance" (Funtowicz & Ravetz, 1993) gives the modeler much freedom to define the perceptual model, that is, based on expert judgment (Krueger et al., 2012), which subsequently influences all modeling steps.

Another premise of positivism is that it aims to conduct science in a value-free manner (Reiss & Sprenger, 2017; Weber, 1917 (reprint 1988)): Science is perceived as an objective endeavor, not influenced by preferences or world views of the scientist. The epistemological uncertainty described above already hampers the objectivity: incompletely understood parts of the system (where it is also up to discussion which parts are understood and which parts are not) are relished with the perception of the modeler. But also beyond the perceptual model, objectivity cannot be achieved. The results of this study demonstrate that values and judgment are abundant throughout the scientific modeling process. This cannot be blamed on the modelers, many decisions are simply necessary to run a model and methodological directives are under-determined.

This leads to the evaluation that models are social constructs (Bijker & Law, 1992; Pinch et al., 1987). Models themselves are the product of socially embedded decisions, as demonstrated by this study. This will make model results time and place dependent—other results will be obtained with different colleagues, different data agreements, or different personal experience. But also the value or trust one assigns to models is socially embedded, as reflected in different modeling visions. One interviewee only trusts models that work across scales, while for many other interviewees, this was not a trust criterion. Both in the construction of models and in the value assigned to models *situatedness* is visible: "*The dependence of meaning on the specifics of particular sociohistorical, geographical, and cultural contexts, social and power relations, and philosophical and ideological frameworks, within which the multiple perspectives of social actors are dynamically constructed, negotiated, and contested.*" (Oxford Dictionary). The results of this study therefore stress the need to move away from the value-free ideal of science and acknowledge that models are social constructs (Melsen, Vos, & Boelens, 2018).

4.3. Moving Forward

The most important implication of acknowledging models as social constructs is that social constructs are not neutral. Certain values and world views have materialized in the model (Bijker, 2007; Melsen, Vos, & Boelens, 2018), for example, as a result of personal judgment, experience from colleagues, and modeling vision, which biases the model results. This is not something that can be solved—there is no ultimate vision, or ultimate personal judgment—but it is something that can be made more visible.

One step could be the formalization of modeling practices. Several (extensive) guidelines exist for good modeling practice and quality assurance (e.g., Jakeman et al., 2006; Refsgaard et al., 2005; van Waveren et al., 1999). They for instance require reproducibility, and a substantiation why a certain model was selected. As such, obeying these standards requires that some of the context-driven decisions move toward epistemically driven decisions. However, epistemically driven decisions are also not universal. Saltelli et al. (2020) discuss the biased narratives that models produce in the context of sustainability. They propose six reflexive lenses to stimulate the exploration of different perspectives. Whereas the formalization of good modeling practices can decrease the number of context-driven decisions, reflexive modeling through multiple lenses allows the exploration of different visions.

Within the context of this study, science-driven modeling, exploration of different visions and an estimation of the methodological uncertainty could also be achieved by having multiple modelers or modeling teams working on the same question—although this study showed that research question, model and modeling vision are intertwined. Also a diversity of modeling visions within the same modeling team could contribute to this, for instance by actively recruiting modelers with another modeling vision. Besides the exploration of different visions, whether through taking multiple lenses or by working with several teams on the same question, transparency and formalization of epistemically driven choices, for example, through a formal modeling vision (see e.g., Nearing et al., 2016), would already be helpful in putting the model results into context.

Given the large share of hydrological scientific studies based on models, it is urgent that the hydrological community critically reflects on its own modeling procedures and standards (not for the first time, see Klemeš, 1997). The large role of local context in model studies leads to the conclusion that model results are time and place dependent and therefore each modeling study can only tell a little part of the story. A combination of good modeling practice guidelines that stimulate moving from context-driven to epistemically driven decisions, and a formal modeling vision that underpins the epistemically driven decisions, can help to clarify the context in which the model was developed and run. If we want to capture the actual robustness of model results, we need a diversity of opinions, perspectives, and approaches, which can only be achieved if modeling procedures are more transparent and explicit. In this way, it will be possible to compare model results of studies that were based on different visions.

5. Conclusions

Fourteen modelers from three different institutes were interviewed about a recent hydrological model study they conducted. The goal was to identify the motivations regarding decisions related to the selection, configuration and execution of the model. Across the interviews, 83 different motivations were identified, which could be classified into seven categories: motivations related to external parties, the scientific community, scientific collaborators, the institute where the modeler works, the team where the modeler works, individual decisions, and consequential decisions. Most motivations were related to the team and individual choices. Especially experience from colleagues frequently appeared as a motivation. Many motivations are social embedded and demonstrate that modeling is, in general, not an individual endeavor.

Model decisions and approaches can be institutionalized and internalized: a modeler can introduce a concept to the team and this can become the team's standard (institutionalization), or a modeler can follow the default team approaches and starts defending them as their own choice (internalization). The role of the team decreases with increasing model experience: this might be the result of institutionalization and internalization. The role of the team was also larger for the modelers working at research institutes than those working in a smaller university group.

Several decisions were investigated in more detail. For model selection, two types of motivations were identified: experience (from colleagues or the modelers themselves), and model vision (the model has certain assets that align with the vision of the modeler). Even though this seems a dichotomy at first, both are related through recruitment (people with a certain vision apply to jobs where they can use models that align with their vision) and through internalization: modelers develop an epistemic vision on how modeling should be conducted, inspired by colleagues. Eventually, modelers can develop their own vision and challenge the vision of colleagues. It appeared that model selection is not driven by the research question, but that available tools and experience, representing the expertise of the team, determine which research questions are addressed. It was also shown that contextual values, that is, values related to the circumstances and context in which the model was configured and run, play a major role in modeling studies. The vast majority of the motivations for modeling decisions were related to specific circumstances, for instance which data were available, experience from colleagues, or just for the sake of time. Only a minority of the motivations could be related to the epistemology of the modeler.

The results show that models are not objective tools, but social constructs: they are the product of socially embedded decisions, and also the value or trust one assigns to models is socially embedded as reflected in different modeling visions. This leads to the conclusion that model results are time and place dependent, and that every modeling study can only tell a small part of the complete story. Acknowledging models as social constructs should affect the way that we estimate model uncertainty. It means that we have to account for the context in which the model was developed and configured, and that a diversity of opinions, perspectives, and approaches is necessary to obtain a fair idea of the robustness of model results. This can only be achieved if modeling procedures become more transparent and if the modeling vision is made explicit for each model study.

Data Availability Statement

For the protection of privacy sensitive information, the interview data are unavailable for public release. The interview protocol is attached as Supporting Information S1.

Acknowledgments
I would like to thank the hydrological modelers that allowed me to interview them and shared their views and perspectives. Joost Buitink and Germano Ribeiro Neto, both from Wageningen University, for the test interviews. In alphabetic order (this order does not align with the interviewee numbers mentioned in the text): Vazken Andreassian, Sebastian Apers, Paul Astagneau, Michel Bechtold, Lila Collet, Akash Koppa, Pallav Kumar Shrestha, Gabrielle de Lannoy, Husain Najafi, Gaia Piazzi, Maria-Helena Ramos, Luis Samaniego, Guillaume Thirel, and Stephan Thober. I would also like to thank Charles Perrin for arranging the interviews at his team and for the valuable feedback, and Remko Uijlenhoet for suggestions for this study. Furthermore, I would like to thank Joseph Guillaume and two anonymous reviewers for their constructive and helpful feedback. This study was conducted with a personal Veni grant (nr. 17297, entitled "What about the modeler? The human-factor in constructing Earth and environmental predictions") from the Dutch Research Council.

References

Adams, W. C. (2015). Conducting semi-structured interviews. In *Handbook of practical program evaluation* (pp. 492–505). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119171386.ch19

Addor, N., & Melsen, L. (2019). Legacy, rather than adequacy, drives the selection of hydrological models. *Water Resources Research*, *55*, 378–390. https://doi.org/10.1029/2018WR022958

Babel, L., Vinck, D., & Karssenberg, D. (2019). Decision-making in model construction: Unveiling habits. *Environmental Modelling & Software*, *120*, 104490. https://doi.org/10.1016/j.envsoft.2019.07.015

Beck, M., & Krueger, T. (2016). The epistemic, ethical, and political dimensions of uncertainty in integrated assessment modeling. *WIREs Climate Change*, *7*, 627–645. https://doi.org/10.1002/wcc.415

Best, M., Abramowitz, G., Johnson, H., Pitman, A., Balsamo, G., Boone, A., & Vuichard, N. (2015). The plumbing of land surface models: Benchmarking model performance. *Journal of Hydrometeorology*, *16*, 1425–1442. https://doi.org/10.1175/JHM-D-14-0158.1

Beven, K. J. (2012). *Rainfall-runoff modelling, the Primer. Down to basics: Runoff processes and the modelling process* (2nd ed.,Vol. Ch.1). John Wiley & Sons.

Bijker, W. (2007). *Dams and dikes, thick with politics* (pp. 109–123). The University of Chicago Press Books. https://doi.org/10.1086/512835

Bijker, W., & Law, J. (1992). *Shaping technology/building society. Studies in sociotechnical change*. MIT Press.

Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., et al. (2021). Behind the scenes of streamflow model performance. *Hydrology and Earth System Sciences*, *25*(2), 1069–1095. https://doi.org/10.5194/hess-25-1069-2021

Burt, T., & McDonnell, J. (2015). Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. *Water Resources Research*, *51*, 5919–5928. https://doi.org/10.1002/2014WR016839

Carnap, R. (1966). In M. Gardner, (Ed.), *Philosophical foundations of physics—An introduction to the philosophy of science*. Basic Books Inc.

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., et al. (2015). Virtual laboratories: New opportunities for collaborative water science. *Hydrology and Earth System Sciences*, *19*, 2101–2117. https://doi.org/10.5194/hess-19-2101-2015

Chegwidden, O., Nijssen, B., Rupp, D., Arnold, J., Clark, M., Hamman, J., et al. (2019). How do modeling decisions affect the spread among hydrologic climate change projections? Exploring a large ensemble of simulations across a diversity of hydroclimates. *Earth's Future*, *7*, 623–637. https://doi.org/10.1029/2018EF001047

Clark, M., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*, W09301. https://doi.org/10.1029/2010WR009827

Elsawah, S., & Jakeman, A. (2020). Preface for the special issue on good modelling practices. *Mathematics and Computers in Simulation*, *175*, 1–2. https://doi.org/10.1016/j.matcom.2020.01.007

Funtowicz, S., & Ravetz, J. (1993). Science for the post-normal age. *Futures*, *25*, 739–755.

Guillaume, J., Helgeson, C., Elsawah, S., Jakeman, A., & Kummu, M. (2017). Toward best practice framing of uncertainty in scientific publications: A review of water resources research abstracts. *Water Resources Research*, *53*, 6744–6762. https://doi.org/10.1002/2017WR020609

Gupta, H. V., Clark, M. P., Vrugt, J., Abramowitz, G., & Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*, W08301. https://doi.org/10.1029/2011WR011044

Hamilton, S., Fu, B., Guillaume, J., Pierce, S., & Zare, F. (2019). A framework for characterising and evaluating the effectiveness of environmental modelling. *Environmental Modelling & Software*, *118*, 83–98. https://doi.org/10.1016/j.envsoft.2019.04.008

Holländer, H. M., Bormann, H., Blume, T., Buytaert, W., Chirico, G. B., Exbrayat, J.-F., et al. (2014). Impact of modellers' decisions on hydrological a priori predictions. *Hydrology and Earth System Sciences*, *18*(6), 2065–2085. https://doi.org/10.5194/hess-18-2065-2014

Hrachowitz, M., & Clark, M. (2017). Hess opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth System Sciences*, *21*(8), 3953–3973. https://doi.org/10.5194/hess-21-3953-2017

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*, 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004

Klemeš, V. (1997). Of carts and horses in hydrologic modeling. *Journal of Hydrologic Engineering*, *2*(2), 43–49. https://doi.org/10.1061/(ASCE)1084-0699(1997)2:2(43)

Krueger, T., Page, T., Hubacek, K., Smith, L., & Hiscock, K. (2012). The role of expert opinion in environmental modelling. *Environmental Modelling & Software*, *36*, 4–18. https://doi.org/10.1016/j.envsoft.2012.01.011

Kuhn, T. (1962 (reprint 2012)). *The structure of scientific revolutions* (4th ed.). The University of Chicago Press.

Kyngäs, H. (2020). Inductive content analysis. In *The application of content analysis in nursing science research* (pp. 13–21). Springer. https://doi.org/10.1007/978-3-030-30199-6_2

Lahtinen, T., Guillaume, J., & Hämäläinen, R. (2017). Why pay attention to paths in the practice of environmental modelling? *Environmental Modelling & Software*, *92*, 74–81. https://doi.org/10.1016/j.envsoft.2017.02.019

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Towards an integrated data assimilation framework. *Water Resources Research*, *43*, W07401. https://doi.org/10.1029/2006WR005756

Lloyd, E. A., & Oreskes, N. (2018). Climate change attribution: When is it appropriate to accept new methods? *Earth's Future*, *6*(3), 311–325. https://doi.org/10.1002/2017EF000665

Lycan, W. (1985). Epistemic value. *Synthese*, *64*, 137–164.

Mayer, L. A., Loa, K., Cwik, B., Tuana, N., Keller, K., Gonnerman, C., et al. (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

McMillan, H., Seibert, J., Petersen-Overleir, A., Lang, M., White, P., Snelder, T., et al. (2017). How uncertainty analysis of streamflow data can reduce costs and promote robust decisions in water management applications. *Water Resources Research*, *53*, 5220–5228. https://doi.org/10.1002/2016WR020328

McMullin, E. (1982). Values in science. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *2*, 3–28.

McMullin, E. (2009). In M. Curd, & S. Psillos (Eds.), *The virtues of a good theory—The Routledge companion to philosophy of science*. Routledge.

Melsen, L., Addor, N., Mizukami, N., Newman, A., Torfs, P., Clark, M., et al. (2018). Mapping (dis)agreement in hydrologic projections. *Hydrology and Earth System Sciences*, *22*, 1775–1791. https://doi.org/10.5194/hess-22-1775-2018

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Mendoza, P., & Uijlenhoet, R. (2019). Subjective modeling decisions can significantly impact the simulation of flood and drought events. *Journal of Hydrology*, *568*, 1093–1104. https://doi.org/10.1016/j.jhydrol.2018.11.046

Melsen, L., Vos, J., & Boelens, R. (2018). What is the role of the model in socio-hydrology? Discussion of 'prediction in a socio-hydrological world'. *Hydrological Sciences Journal*, *63*, 1435–1443.

Menard, C., Essery, R., Krinner, G., Arduini, G., Bartlett, P., Boone, A., et al. (2021). Scientific and human errors in a snow model intercomparison. *Bulletin of the American Meteorological Society*, *102*(1), E61–E79. https://doi.org/10.1175/BAMS-D-19-0329.1

Nearing, G., Tian, Y., Gupta, H., Clark, M., Harrison, K., & Weijse, S. (2016). A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, *61*(9), 1666–1678. https://doi.org/10.1080/02626667.2016.1183009

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

Packett, E., Grigg, N., Wu, J., Cuddy, S., Wallbrink, P., & Jakeman, A. (2020). Mainstreaming gender into water management modelling processes. *Environmental Modelling & Software*, *127*, 104683. https://doi.org/10.1016/j.envsoft.2020.104683

Pinch, T., Bijker, W., & Hughes, T. (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT Press.

Polanyi, M. (1958). *Personal Knowledge: Towards a post-critical philosophy*. The University of Chicago Press.

Refsgaard, J., Henriksen, H., Harrar, W., Scholten, H., & Kassahun, A. (2005). Quality assurance in model based water management – Review of existing practice and outline of new approaches. *Environmental Modelling & Software*, *20*(10), 1201–1215. https://doi.org/10.1016/j.envsoft.2004.07.006

Reiss, J., & Sprenger, J. (2017). Scientific objectivity. In E. Zalta (Ed.), *The Stanford Encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity/

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*, W05521. https://doi.org/10.1029/2009WR008328

Saltelli, A., Benini, L., Funtowicz, S., Giampietro, M., Kaiser, M., Reinert, E., et al. (2020). The technique is never neutral. how methodological choices condition the generation of narratives for sustainability. *Environmental Science & Policy*, *106*, 87–98. https://doi.org/10.1016/j.envsci.2020.01.008

Savenije, H. (2009). HESS opinions: The art of hydrology. *Hydrology and Earth System Sciences*, *13*, 157–161. https://doi.org/10.5194/hess-13-157-2009

van Waveren, R., Groot, S., Scholten, H., van Geer, F., Wösten, J., Kroeze, R., et al. (1999). *Vloeiend modelleren in het waterbeheer—Handbook Good Modelling Practice* (Tech. Rep. No. 99-05/99.036). STOWA/RIZA.

Wagener, T., Wheater, H., & Gupta, H. (Eds.), (2004). *Rainfall-runoff modelling in gauged and ungauged catchments*. Imperial College Press.

Weber, M. (1917 (reprint 1988)). Der sinn der 'wertfreiheit' der soziologischen und ökonomischen wissenschaften. *Gesammelte Aufsätze zur Wissenschaftslehre*, *451–502* (I used the interpretation of Reiss and Sprenger (2017)).

Weber, R. (1990). *Basic content analysis (No. 49)*. Sage.