**Legacy, rather than adequacy, drives the selection of hydrological models**

N. Addor[1], L.A. Melsen[2]

[1] Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK
[2] Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, The Netherlands

**Key points**

Text mining of 1500+ peer-reviewed articles enabled us to relate hydrological models to institutes, regions and research topics
We provide evidence of regional preferences in model use across the world and underline the decisive influence of legacy on model selection
We reflect on current tendencies and future model development strategies, and advocate for a broader use of modular modelling frameworks

**Keywords**

Model selection, model evaluation, bibliometric study, text mining, community model, modular modelling frameworks

**Abstract**

The findings of hydrological modelling studies depend on which model was used. Although hydrological model selection is a crucial step, experience suggests that hydrologists tend to stick to the model they have experience with, and rarely switch to competing models, although these models might be more adequate given the study objectives. To gain quantitative insights into model selection, we explored the use of seven rainfall-runoff models (HBV, Topmodel, VIC, mHM, GR4J, PRMS, Sacramento) based on the abstract of 1529 peer-reviewed papers published between 1991 and 2018. We provide quantitative evidence of regional preferences in model use across the world and demonstrate that specific models are consistently preferred by certain institutes. Model attachment is particularly strong. In ~70% of the studies, the model selected can be predicted solely based on the affiliation of the first author. The influence of adequacy on the model selection process is less clear. Our data reveal that each model is used across a wide range of purposes, landscapes, temporal and spatial scales (i.e., as a “model of everything and everywhere”). Model intercomparisons can provide guidance for model selection and improve model adequacy, but they are still rare (because each model must usually be setup individually) and the insights they provide are currently limited (because they are rarely controlled experiments). We suggest that moving from fixed-structure models to modular modelling frameworks (master templates for model generation) can overcome these issues, enable a more collaborative and responsive model development environment, and result in improved model adequacy.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2018WR022958

## 1. Introduction

Every hydrological modelling study involves the selection of a hydrological model. Selection of a particular model implies choosing the perceptual and conceptual model (e.g., Jakeman et al., 2006, Beven 2012). This step is crucial, because the outcomes of the study depend on the model selected (e.g., Holländer et al., 2009; Fenicia et al., 2014; Mendoza et al., 2016; Hauser et al., 2017; Melsen et al., 2018a). Therefore, ideally, the selection of this model should be based on its adequacy for the research question (such as the landscape of the region, the temporal and spatial scale, and the purpose of the study, for instance flood modelling or water resources management). However, experience suggests that hydrological models are usually not primarily selected because of their adequacy, but because of legacy reasons, where legacy involves practicality, convenience, experience and habit.

To gain quantitative insights into the model selection process, we performed a bibliometric study: we used text-mining tools to analyse the content of published peer-reviewed articles. Other hydrological studies have employed bibliometric methods, for instance to stress the relative decline of papers on field hydrology (Burt and McDonnell, 2015), to assess the impact of publications in leading hydrological journals (Clark and Hanson, 2017) and to explore how uncertainty is communicated in hydrological publications (Guillaume et al., 2017). To our knowledge, this is the first bibliometric study exploring the use of hydrological models across the world.

We used text-mining to search the abstracts of 1529 hydrological modelling studies published between 1991 and 2018. These abstracts were selected because they mention one of the seven hydrological models this study focuses on. We extracted the affiliation of the first author and, when available, the country of application of the model and keywords describing the context of the study. This enabled us to link model selection to the purpose of the study and to the affiliation of the first author. The fact that regional preferences in model use exist is well acknowledged among hydrologists. Here, we explore how these regional preferences were established and explore what this reveals about the use and development of hydrological models.

Regional preferences in model use can be approached from a 'geography of science' perspective (Livingstone, 2003). Although scientific endeavors are generally perceived as "universal", there are ample examples from the field of Science and Technology Studies that demonstrate that the way science is conducted is place-bound, related for instance to routine practices in laboratories (e.g. Latour and Woolgar, 1986), cultural values (see e.g. Melsen et al, 2018b for a discussion of the social construction of technology applied to hydrological modelling), and how institutions are organised - Miller and Edwards (2001), for example, describe how different institutionalizations of climate research in the US and the UK eventually lead to different research approaches, and Mahony and Hulme (2018) discuss the locality of science in the perspective of climate models and the implications for decision making. Acknowledging that the way science is conducted is place-bound, two main hypotheses can be developed to explain regional preferences in model use, either based on the landscape (the adequacy hypothesis) or the institute (the legacy hypothesis).

Naylor (2005) argues that "*we could fruitfully consider the ways landscapes, regions and places inform - consciously or not - scientific theories and practices*". This can be applied to hydrological modelling. To take a simple example, the dedication of particular attention to the development of the snow routine in a model developed in Scandinavia indicates that the local landscape inspires the model builder. One of two hypotheses that this study focuses on is that the landscape also inspires model selection (adequacy). Law and Mol (2001) discuss that scientific knowledge has locality, but that transport of knowledge exists. Models are often applied in the place where they were developed, but there is transport of model use to other places. Shapin (1998) and Law and Mol (2001) state that the transport of knowledge requires trust in the source. In the case of models, this means that the institute that developed the model, but also the model itself, should be trustworthy in order for the model to be adopted by other institutes (Shapin, 1994, and Hardwig, 1991). The second hypothesis tested by this study is that the institute determines model selection (legacy).

In summary, to gain a better understanding of regional preferences in model use (i.e., to better understand the drivers of model selection), we used these bibliometric data to address two main questions:

1. What are the regional differences in hydrological model use across the world?
2. Is model selection more driven by legacy (hydrologists use the model they have most experience with) or by adequacy (hydrologists use the model most adequate for their specific research question)?

The bibliometric search we designed to address these two questions is presented in Section 2. Differences in model use across the world are discussed in Section 3.1, while the importance of legacy and adequacy in model selection is explored in Section 3.2 and 3.3, respectively. These findings are then utilised to reflect on tendencies in the development and use of hydrologic models over the last decades (Section 4.1), to discuss possible shortcomings of this development (Section 4.2), and to propose an alternative strategy for model use and development (Section 4.3). Conclusions are provided in Section 5.

## 2. Data and methods: Text-mining of peer-reviewed papers to explore model use across the world

We selected seven frequently-used conceptual hydrologic models: the Hydrologiska Byrans Vattenbalansavdelning model (HBV, Bergström, 1976; Seibert and Vis, 2012), the Variable Infiltration Capacity model (VIC, Liang et al., 1994; Hamman et al, 2018), the Sacramento soil moisture accounting model (SACRAMENTO, Burnash et al., 1973), the Génie Rural 4 model (GR4J, Perrin et al., 2003), the mesoscale hydrological model (mHM, Samaniego et al., 2010), the TOPography based hydrological model (TOPMODEL, Beven and Kirkby, 1979; Beven and Freer, 2001; Metcalfe et al., 2015), and the precipitation Runoff Modelling System (PRMS, Leavesley et al., 1983; Markstrom et al, 2015). The selection of these seven models is based on their popularity and similarity. These conceptual models are comparatively easy to implement and run, hence in theory, hydrologists could switch from one to the other depending on their research question.

We searched the Scopus database for studies using at least one of the seven models we selected. The search was restricted to the title, keywords and abstract, since Scopus does not enable to search the main text (see Text S1; Pautasso, 2014; Franceschini et al, 2016; Franceschini et al, 2017; Calver et al, 2017). In a bibliometric review, both errors of commission (including papers that are not targeted) and errors of omission (missing papers that are targeted) can occur (Borrett et al., 2018). To avoid errors of commission, the papers returned by Scopus were screened using three different methods (see Text S2). Our study is subject to errors of omission because using Scopus meant that we disregarded i) studies relying on one the seven models but do not mentioning the model name in their abstract, title or keywords and ii) non peer-reviewed studies, such as technical reports. To evaluate the representativity of the papers we retrieved, we compared them to reference lists of papers maintained for some models by their developing team (Text S3, Figure S2 and Table S4). This demonstrated that no reference list is complete, thereby highlighting the difficulty to keep track of all the studies published using a specific model. It also showed that retrieval rate (the fraction of papers in each reference list covered by the papers we retrieved) varies substantially from one model to the other (from 6% for PRMS to 55% for mHM). This is partially explained by the different publication methods employed by user groups of different models (see Text S3 for more details).

The paper screening resulted in a total of 1529 papers, which constitute the core material of this study. For each paper we retrieved the title, authors, year of publication, journal name and abstract using R (R Core Team, 2018) and the package Bibliometrix (Aria and Cuccurullo, 2017). Subsequently, the titles and abstracts were searched for keywords describing where the study was conducted, the spatial (e.g., ‘global’) and temporal (e.g., ‘forecast’) scale of the study, specific catchment descriptors (e.g., ‘alpine’) and the goal of the study (e.g., ‘drought’). This process is described in more details in Text S4.

## 3 Results

### 3.1. Regional differences in model use based on the text mining of 1500+ abstracts.

To provide context for this bibliometric study, let us underline that the number of publications per year steadily increased over the research period (Figure S3). It increased by almost a factor 20 between the first five years and the last five years of our study period (1990-1994 and 2013-2017, respectively). This increase in publications is part of a more general trend in academia and in hydrologic science (Burt and McDonnell, 2015). Out of the 1529 abstracts, the models most frequently used are VIC, HBV and TOPMODEL (each accounting for 33%, 23% and 22% of the abstracts, respectively). Publications involving VIC and HBV show a steady increase over time, while the use of Topmodel has, in absolute terms, remained constant but in relative terms has decreased over time (Figure S3).

The data we collected illustrate that clear regional preferences exist in the use of hydrologic models across the world, both in terms of numbers of studies published and regions over which models are deployed. Regions such as North America, Europe, China and Australia host the institutes publishing the large majority of the papers and they are also the regions over which models are most frequently deployed (Figure 1). In contrast, regions such as Africa, the middle East, Central and South America are underrepresented, both in terms of publishing institutes and application of models. Similar geographical biases also prevail in hydrological field studies (Burt and McDonnell, 2015). Further, the large majority of the studies were applied in a region in the affiliation country of the first author, as indicated by the fact that the thickest lines in Figure 1 are circular arrows.

In addition to showing where hydrological modelling studies are conducted, Figure 1 provides quantitative evidence that there are regional preferences in the use of specific models. For instance, VIC and SAC are predominantly used in the USA, while HBV dominates in Scandinavia and in other parts of Europe. These regional model preferences can be well related to the history of model development. VIC and SAC, whose development started and is led by institutes in the USA, account for the large majority of the studies applied in the USA. In contrast, HBV, Topmodel and GR4J, which originate in Europe, are mostly used by institutes in Europe, HBV being particularly popular in Scandinavia and GRJ4 in France. These results are in line with those of Singh and Woolhiser (2002), who underlined the popularity of HBV in Scandinavia and the wide application of Topmodel throughout Europe.

### 3.2 Exploring model legacy by quantifying hydrologists' attachment to their model

To explore the influence of legacy on model selection, we defined "model attachment" as the fraction of papers published by each institute using each model. We attributed each paper to the institute of its first author. Aggregating our results by institute provides enough papers to draw reliable conclusions (aggregating the papers by researcher would have led to too small subsamples), while still preserving some geographic details that would be lost by an aggregation at the country scale (Figure S4). For each institute, we determined which model is used most frequently and refer to this model as the institute "favourite model". In a hypothetical situation, in which there would be no favourite model, the seven models this study focuses on would be used in around 1/7 (~14%) of the publications of any institute. In practice, however, each institute has a favourite model, which is used more frequently than the others. Note that given the nature of our bibliometric data set, favourite model means favourite model among the seven models this study focuses on.

Figure 2a shows how frequently the seven models are used in the publications of ten selected institutes. Three of these institutes used a single model for all the publications we retrieved: MHM for UFZ Leipzig (Germany), Topmodel for Lancaster University (UK), and SAC for NOAA/NWS/NCEP (US). This means that the attachment of these institutes for their respective favourite model is 100%. For seven other institutes, more than one model has been employed and a favourite model can be identified. The attachment of these institutes for their model can be high: for instance 94% for SMHI (HBV), 92% for University of Washington (VIC), and 73% for IRSTEA (GR4J). Statistics over 50 institutes that published at least 5 papers confirm that there is typically a model used more frequently than the other models. The institutes use their favourite model far more than in 14% of their publications that would be expected if there was no favourite model (Figure 2b). We found that the institutes use their favourite model on average for 74% of the publications (Figure 2c). In other words, it is possible to determine for 74% of these studies which model was used solely based on the affiliation of the first author (when including all the institutes, this number increases to 79%, see Figure S5). Figure 2d shows that the institutes attachment for their favourite model remains high over the years. Although there is a slight decline over the last five years, our data show that a model that has been extensively used in the past keeps being used for many years. Note however that the data we collected only enable us to characterise switches among the seven models we selected, but not switches to other models. Figure 2d nevertheless reflects the inertia of modelling practices and the influence of legacy when selecting a hydrological model.

### 3.3. Exploring model adequacy using keywords describing the purpose of each study

Adequacy implies that the hydrological model is selected because it is particularly well suited for the purpose of the study. We related adequacy to the main variable of interest (e.g., flood modelling or water resources management), the landscape, and the temporal and spatial scale of the application (see keywords in Table S4). We analysed the occurrence of these keywords and explored whether some models are used more frequently for some specific purposes (possibly meaning that they are particularly well adapted to those purposes) or if, in contrast, models are used to address a wide range of challenges, without much discrimination based on the purpose of the study.

We presented the results using mosaic plots (Figure 3) and we asked two questions: i) is a specific model used significantly more often for one type of purpose (e.g., used more for flood than for drought modelling or more for regional than for global modelling) and ii) for one type of purpose (such as the exploration of hydrological processes at the regional scale), is one model used significantly more often than the others? To answer the first question, we analysed the mosaic plots one column at a time, and for the for the second question, we analysed the mosaic plot one row at a time. We did not consider combinations of keywords, since only few studies provided keywords for all the categories we considered (note that the number of studies in which one of the categories is described varies, see headers in Figure 3).

Often, the keywords describing the study were used with a comparable frequency for the seven models we selected. A clear example is that the models are equally likely to be used in studies whose abstract mentions the regional scale (top row of Figure 3d). Some differences between models canbe noted when focussing on the variable of interest (Figure 3a): SACRAMENTO is used more often for flood than for drought modelling, and the reverse is true for mHM (note that the scaling used to account for the different number of papers per model artificially boosts mHM importance in this Figure). But overall the models investigated are used to simulate most (when not all) the variables of interest we selected, meaning that the variable of interest isnot a strict discriminator for model selection. The landscape descriptors reveal some patterns, although these patterns do not have much predictive power when it comes to model selection (Figure 3b). For instance, five out of seven models have been applied both in tropical and alpine regions, which implies that landscape descriptors alone cannot explain model selection. Differences between models are even weaker when considering keywords describing the temporal and spatial scale. All models are used across temporal scales, from forecasts to climate projections, except mHM for which less papers are available than for the other models (Figure 3c). Similarly, models are widely applied across spatial scales, there are examples of studies at the catchment, regional, continental and global scale for almost each model, and it is hard to argue that a specific model is preferred at a specific spatial scale (Figure 3d). This result partially stems from the relatively homogenous subsample of rainfall-runoff models we considered here, clearer differences might have appeared if, for instance, more spatially distributed models had been included.

Overall, our text-mining analysis based on about 1500 abstracts provides little evidence that models are selected because of their adequacy for the research question and we could not show that a specific model is particularly popular for a specific purpose. The hypothesis that model selection is driven by legacy is better supported by our results than the hypothesis that model selection is driven by adequacy. We do not claim that the models investigated here are inadequately used. We observe that they tend to be applied across the whole range of landscapes, scales, and variables of interest that we investigated.

**4. Discussion**

Our bibliometric study showed that hydrologists typically keep using the model they are already using. In this discussion part of the paper, we start by formulating hypotheses on why this approach works (Section 4.1), we continue by asking whether this mode of operation leads to missed opportunities (Section 4.2), and finish by reflecting on a strategy to move forward (Section 4.3).

**4.1 Why is model selection based on legacy a system that works and continues to work?**

The dominant tendency of selecting models because of legacy reasons is, in many respects, a system that works. A proof is that, in the papers we retrieved for this study, model selection was strongly influenced by legacy, but this did not prevent these papers from being published in peer-reviewed journals. In fact, their number per year is increasing (Figure S3) and there is little sign that this tendency is changing (Figure 2d). We propose the following reasons to explain why this system works.

Over years of use of a model, hydrologists gain a deep knowledge and understanding of this model. They can in particular quickly set it up in a new environment, identify the origin of suspicious results, as well as adapt and improve its code. Part of this expertise is gained by the modellers as students. HBV is for instance extensively used for teaching in Europe (Seibert and Vis, 2012) and institutes like the University of Washington and Princeton in the USA have trained students over the last decades with VIC and sometimes involved them in model development. Some of these students moved to China, where they kept using VIC, which contributed to its popularity there. Experience is an important asset in science. Training is needed to work with certain laboratory equipment (e.g. Latour and Woolgar, 1986), and the same is true when working with computer models. Laboratory work involves *tinkering* (Knorr Cetina, 1981) and *bricolage* (Latour and Woolgar, 1986), and so does hydrological modelling. Thinkering for example involves recalibrating a model with new parameter intervals because the simulations are not satisfactory in the eye of the expert. The large role of legacy in model selection can thus be justified by the scientific need to have experience with the experimental set-up. Further, sustained model use leads to the development of a model "ecosystem", consisting of tools and data sets, such as a user-friendly interface, built-in parameter estimation tools, support in producing model files, or publicly available forcing and parameter files. This leads to a productive modelling environment, made by a combination of adequacy (modellers become proficient with their model) and legacy (using their model is efficient).

Another reason that makes modellers stick to the model they are already using, is that they perceive it as adequate. This bibliometric study does not enable us to determine how adequate the models are considered by their users, but Fleming (2009) provides interesting insights. He conducted a survey among 45 users of hydrologic models (from academia and the industry), to investigate which models they used, and why (his study is complementary to this study, since it involved practitioners who may not publish often or at all and therefore would not appear in our bibliometric study). The most frequent reason given to explain model selection (28% of the respondents) is that the model "captures all physical hydrologic processes to [the] extent necessary for [their] project". In other words, adequacy is a key criterion for model users and they consider that the model they select is adequate. Fleming (2009) did not ask model users how they came to the conclusion that the model "captures all physical hydrologic processes". It is possible that it is based on their expert opinion (Krueger et al., 2012). Note that Fleming (2009) reported that 21% of the respondents indicated that they select a specific model because it is "in standard use in [their] organization", which highlights the important role of legacy in model selection, and is consistent with our results

Furthermore, besides the advantage of experience with a model discussed above, there are other clear advantages for using generally accepted models within the scientific field. Importantly, these advantages have a clear sociological component. The current functioning of the scientific community stimulates the use of well-established methods in several ways. Both Horrobin (1990) and Calver (2015) discuss how the current review process and focus on paper citations might hamper the uptake of innovative methods, e.g. because reviewers might not be familiar with the new approach or because of the risks involved with trying new approaches. Martin (2000) discusses problems in current grant systems, and describes the conventional approach bias: "*grants are much more likely to support tried-and-true approaches, while challenging, innovative or unorthodox proposals are seldom funded*". An example of the weight of legacy in hydrology is the use of the Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970) to evaluate models. With the advent of automated calibration techniques, models can provide high NSE values in most catchments. Yet, studies relying on more process-oriented metrics (e.g., DeBoer-Euser et al., 2017), stricter evaluation procedures (e.g., Coron et al., 2012; Refsgaard et al., 2013) and multiple data sets (e.g., Rakovec et al, 2016) have demonstrated that the resulting simulations often involve unrealistic features and lack robustness. Despite these caveats, the NSE is still broadly applied, as it allows for comparison to all the other studies using NSE. From a more conceptual perspective, holding on to conventional methods fits within a period of '*normal science*' (Kuhn, 1962). The training of young researchers in the conventional views and frameworks is inherent to normal science, thereby maintaining current practice. This does not imply that no scientific progress is made, but merely that scientific research is highly structured within the current paradigm. Conventional methods are thus valuable, and stimulated in current scientific practice. This might, however, hamper innovation.

### 4.2. Are hydrologists missing opportunities because of the prevalence of legacy in model selection?

There are clear motivations to keep using the same model, as discussed in the previous section. The question, however, is whether hydrologists are missing opportunities because of this mode of operation. The perceived adequacy of their model, which perpetuates its use, is not necessarily well-grounded (Best et al., 2015; Seibert et al., 2018). And even if the model selected based on legacy is adequate, more adequate models might be available. It is, however, generally unclear how significant the gain in adequacy will be when switching to another model, and whether this gain justifies the effort of switching models (Cunderlik et al., 2013). Furthermore, limited guidance exists on how to select a model for a given purpose or landscape (e.g., McDonnell and Woods, 2004; Boorman et al., 2007). There is evidence that parameter tuning alone is often not sufficient and that model structure should be adapted to the environment (e.g., Savenije, 2010; Kavetski and Fenicia, 2011; Hartmann et al, 2014; Fenicia et al, 2014; Gharari et al, 2014). Yet, although studies based on small samples of catchments have demonstrated that specific modelling changes improve model realism, there is a lack of general guidance on the selection of hydrological models.

Such guidance can be provided by model intercomparisons (e.g., Loague and Freeze, 1985; Refsgaard and Knudsen, 1996; Slater et al, 2001; Reed et al., 2004; Smith et al., 2012; Krysanova et al., 2017; Schellekens et al., 2017; Kollet et al., 2017; Krinner et al, 2018), yet with fixed-structure models, setting up and running a large ensemble of hydrological models is difficult and costly (each model must be setup individually, so a large number of research groups is required to run a large ensemble of models). This makes consciously comparing competing models and selecting the most adequate one(s) challenging. Furthermore, insights gained from model intercomparisons are limited, because it is usually impossible to pinpoint which differences between the models explain differences between the simulations (Clark et al., 2011). In other words, models biases can rarely be traced back to specific modelling decisions, so although model intercomparisons tell us *which* model(s) perform best, they rarely tell us *why*.

Furthermore, when hydrologists focus on their own model, they arguably miss opportunities to improve models in a way that benefits the community. In many respects, the current development of hydrological models is siloed (code produced for one model is rarely usable in other models, so improvements made to one model rarely benefit other models) and fragmented (there is little international coordination of development efforts). This type of model development leads to duplicated implementations (think of the number models relying on a similar degree-day snow parameterisation), which absorbs a significant portion of the economic resources dedicated to model development, without necessarily advancing the understanding and simulation of hydrological processes. Finally, this fragmentation makes it difficult to gain an overview of what has been developed, i.e., which parts of the model space have been sampled.

### 4.3. Outlook: moving from flexible models to modular modelling frameworks?

The development and use of hydrological models over the last decades has led to a wide range of models and to a competent user base. The current paradigm is that institutes focus on the development of individual hydrological models and hydrologists tend to have a favourite model which they use to address a wide range of research questions and applied challenges. While this system works and continues to work, it may restrain the efficiency and insights provided by hydrological modelling. An alternative paradigm, which would enable the community to overcome the limitations outlined above, would be to rely more importantly on modular modelling frameworks (MMFs). MMFs are master templates for model generation, which enable users to build models in a pick-and-mix approach, to compare competing hypotheses, and to contribute with code contained in modules that would be easily usable by others in the same framework. Over the last two decades, several MMFs have been developed, spanning a wide range of complexities and offering different levels of user friendliness (e.g., Leavesley et al, 1996; Clark et al., 2008; Savenije, 2010; Fenicia et al., 2011; Niu et al., 2011; Clark et al, 2015). In this study, we consider that a modelling framework is modular if i) for the majority of its main structural elements, several options are available and can be used interchangeably and ii) the architecture of the framework was designed to allow for this modularity and enables the addition of new modules. Although most standard hydrological models offer some degree of structural flexibility (e.g., VIC can be run in energy-balance or water-balance mode and HBV can account or not for the slope aspect), we do not consider them as MMFs based on the two criteria introduced above.

MMF present essential advantages when compared to fixed models, as discussed in particular by the studies referenced in the previous paragraph. Here, we highlight some of these advantages in the framework of our legacy-adequacy assessment. Firstly, MMFs make it comparatively easy to switch between model structures. Hence, they enable hydrologists to run an ensemble of model structures (i.e., to conduct their own model intercomparison experiment), select the model structures most adapted to the purpose of their study and, eventually, provide guidance on model selection which benefits the community (e.g., Staudinger et al, 2011; McMillan et al, 2011; Clark et al., 2011; Coxon et al, 2013). This reduces the role of legacy in the model selection process and increases adequacy. Secondly, MMFs enable the construction of controlled experiments explaining (instead of only describing) differences between models, effectively enabling researchers to trace model uncertainties and biases back to modelling decisions (e.g., Kavetski and Fenicia, 2011). This means that model adequacy is supported by a better understanding of model behaviour. Thirdly, the modularity of MMFs facilitates the incorporation of new insights and knowledge in models, improving the responsiveness of the modelling community to theoretical breakthroughs and thereby increasing the realism of hydrological models (Archfield et al, 2015; Clark et al., 2016).

A fourth argument in favour of the use of MMFs is that they stimulate pluriformity in modelling approaches. When advocating for a pluriform approach in ecological modelling, Jørgensen and Müller (2000) use Gödel's theorem: "the infinite truth can never be condensed in a finite theory". Because of the inability of isolated models to fully capture the complexity of real-word systems, a plurality of legitimate perspectives are necessary to reflect this complexity. One way to embrace pluriformity is via the method of multiple working hypotheses (Chamberlin, 1890; Clark et al., 2011), which is central to MMFs and enables competing theories to co-exist and to be compared. Importantly, MMFs allow for additional working hypotheses to be progressively added (as modules), leading to a more coordinated and systematic sampling of the model space than if fixed-structure models were used. This can help with the realisation that the model choice for a specific part of the system is to a large extent a human enterprise (e.g. Zellmer, 2006 and Peck, 2008).

MMFs have been available for some time, and one may wonder how to accelerate their development and adoption by the community. Because of the decisive role of legacy (in particular, the attachment of modellers to their model), it is unlikely that the entire (or even the majority of the) community will gather behind one model or modelling framework. In fact, there is indecision on the need for a community hydrological model (Weiler and Beven, 2015). An alternative strategy to a community-wide approach is to rely on grassroots initiatives. Grassroot initiatives stimulating the uses of MMFs by making them more user-friendly, providing training opportunities with MMFs, producing guidance for model selection that can be used by the community, and encouraging modellers to contribute with modules, have the potential to shift the current modelling paradigm from fixed-structure models to MMFs. Modules (also referred to as libraries or packages) have been successfully used for decades by the main programming languages (e.g., Fortran, Python). We believe that a similar approach would benefit the hydrological community, and that there would be significant gains to move from model to module development.

## 5. Conclusions

The goal of this study was to investigate regional preferences in hydrologic model use across the world, and to explain these regional preferences by exploring the role of legacy and adequacy in model selection. To this end, we conducted a bibliometric analysis on 1500+ peer-reviewed papers in which one of the seven hydrologic models we focus on (HBV, Topmodel, VIC, mHM, GR4J, PRMS, Sacramento) was applied. The bibliometric analysis revealed clear regional preferences in model use. In 74% of the cases, the applied model could be predicted solely based on the institute of the first author, underlining the role of legacy in model selection. Our analysis of keywords describing the context of the study did not provide clear evidence of the importance of adequacy in model selection. Our results therefore provide stronger support for the hypothesis that legacy, rather than adequacy, drives model selection,. This study also highlights how sociological phenomena shape hydrological modelling practices  and contribute to explain the strong role of legacy. This invites for further exploration of the social practice of science, which, so far, has hardly been explored within Earth sciences (Wainwright, 2012).

The importance of legacy in the model selection process does not mean that adequacy is not considered by modelers. When asked how they selected their model, modellers indicate that its suitability for the purpose of their study is crucial (Fleming, 2009). Model adequacy is enhanced by the expertise and tools modelers develop over the years, which are part of model legacy. This illustrates that untangling the role of adequacy and legacy in model selection is challenging. Our analysis shows that the models considered are used across the whole range of purposes and scales we explored, indicating the high level of trust modellers place in the abilities of their model. We suggest that the tendency to select models based on legacy leads to missed opportunities, in particular because simulations realism can improve when using one of the multitude of models developed by the community, instead of the modeler's usual model. Yet, limited guidance currently exists to support model selection, and it is usually unclear how significant the benefits of switching to another model would be. A key reason for this is that comparing a wide range of competing models for a specific purpose is costly with current hydrological models.

We propose that one way to increase the influence of adequacy in the model selection process is to use modular modelling frameworks (MMFs) more broadly. Several MMFs already exist and have clear advantages when compared to (fixed-structure) traditional models. MMFs facilitate the comparison of a wide range of models (i.e., hydrologists can conduct their own model intercomparison experiments and select the most adequate model structure), they enable controlled modelling experiments (i.e., in which model biases and uncertainties can be traced back to specific modelling decisions), and finally, MMFs open the door to a more collaborative and better coordinated model development environment, in which modules can be used across models and the model space is sampled in a more systematic way. Arguably, switching from traditional fixed-structure models to MMFs represents a paradigm shift in model use and development. Given the inertia of modelling habits, this shift will take time and is most likely to succeed if conducted by grassroots initiatives. We think that the time is right for this shift to happen, and that it will accelerate progress on the understanding and modelling of hydrological processes.

**Acknowledgements**

This research was supported by the Swiss National Science Foundation (grant P2ZHP2_161963). We thank the Editor and the four anonymous Reviewers for their constructive comments, Andy Wood, Jan Seibert and Bart Nijssen for discussing with us their experience and views on the selection of hydrological models, and Lucie Babel for her insights into the sociological aspects of model development. We also acknowledge the feedback of the participants of the second workshop on "Improving the theoretical underpinnings of hydrologic models" organised in Sopron, Hungary, 15-18 April 2018. The

bibliometric data are available online: http://www.hydroshare.org/resource/a74d78bd440444f2a54b2994771354d9

**References**

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W. H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S. and Over, T.: Accelerating advances in continental domain hydrologic, Water Resour. Res., 51, doi:10.1002/2015WR017498, 2015.

Aria, M. and Cuccurullo, C.: Bibliometrix: An R-tool for comprehensive science mapping analysis, J. Informetr., 11(4), 959-95, doi: 10.1016/j.joi.2017.08.007, 2017.

Bergström, S. (1976), Development and application of a conceptual runoff model for Scandinavian catchments, Tech. rep., SMHI Report RHO 7, Norrköping.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. a., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J. ., Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. a., Stevens, L. and Vuichard, N.: The plumbing of land surface models: benchmarking model performance, J. Hydrometeorol., 16, 1425–1442, doi:10.1175/JHM-D-14-0158.1, 2015.

Beven, K. J. and Kirkby, M.J. (1979). A physically based variable contributing area model of basin hydrology. Hyd. Sci. Bull. 24, 43-69

Beven, K. J., Rainfall-Runoff modelling, The Primer - 2nd Edition, Ch.1. Down to Basics: Runoff Processes and the Modelling Process. John Wiley & Sons, 2012.

Beven, K. and Freer, J.: A dynamic TOPMODEL, Hydrol. Process., 15, 1993–2011, doi:10.1002/hyp.252, 2001.

Boorman, D., Williams, R., Hutchins, M., Penning, E., Groot, S., and Icke, J.: A model selection protocol to support the use of models for water management, Hydrol. Earth Syst. Sci., 11(1), 634-646, doi: 10.5194/hess-11-634-2007, 2007.

Burnash, R., R. Ferral, and R. McGuire (1973), A generalized streamflow simulation system - conceptual modeling for digital computers, Tech. rep., U.S. Department of Commerce, National Weather Service and State of California, Department of Water Resources

Burt, T. P. and McDonnell, J. J.: Whither field hydrology? the need for discovery science and outrageous hydrological hypotheses, Water Resour. Res., 51(8), 5919–5928, doi:10.1002/2014WR016839, 2015.

Calver, M.C.: Please don't aim for a highly cited paper. Australian Universities' Review, 57 (1), 45-51, available at: https://files.eric.ed.gov/fulltext/EJ1053513.pdf, 2015.

Calver, M.C., Goldman, B., Hutchings, P.A. and Kingsford, R.T.. Why discrepancies in searching the conservation biology literature matter. Biological Conservation 213, 19- 26, 2017

Chamberlin, T. C. (1890), The method of multiple working hypotheses, Science (old series), 15, 92.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T. and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44(12), doi:10.1029/2007WR006735, 2008.

Clark, M. P., Kavetski, D. and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47(9), doi:10.1029/2010WR009827, 2011.

Clark, M. P., McMillan, H. K., Collins, D. B. G., Kavetski, D. and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 2: Process-based evaluation of model hypotheses, Hydrol. Process., 25(4), 523–543, doi:10.1002/hyp.7902, 2011.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J. and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51, 2498–2514, doi:10.1002/2015WR017198, 2015.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbulluoglu, E. and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, Water Resour. Res., 52(3), 2350–2365, doi:10.1002/2015WR017910, 2016.

Clark, M. P. and Hanson, R. B.: The citation impact of hydrology journals, Water Resour. Res., 53, 4533–4541, doi:10.1002/2017WR021125, 2017.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M. and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water. Resour. Res., 48(5), 1–17, doi:10.1029/2011WR011721, 2012.

Coxon, G., Freer, J., Wagener, T., Odoni, N. A. and Clark, M.: Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, Hydrol. Process., 28(25), 6135–6150, doi:10.1002/hyp.10096, 2013.

Cunderlik, J.M., Fleming, S.W., Jenkinson, R.W., Thiemann, M., Kouwen, N, and Quick, M.: Integrating logistical and technical criteria into a multiteam, competitive watershed model ranking procedure, J. Hydrol. Eng., 18, 641-654, doi:10.1061/(ASCE)HE.1943-5584.0000670, 2013.

De Boer-Euser, T., Bouaziz, L., De Niel, J., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., Savenije, H., Thirel, G. and Willems, P.: Looking beyond general metrics for model comparison - Lessons from an international model intercomparison study, Hydrol. Earth Syst. Sci., 21, 423–440, doi:10.5194/hess-21-423-2017, 2017.

Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47(W11510), doi:10.1029/2010WR010174, 2011.

Fenicia F, Kavetski D, Savenije HHG, Clark MP, Schoups G, Pfister L, Freer J. 2014. Catchment properties, function, and conceptual model representation: Is there a correspondence? Hydrological Processes 28 (4): 2451–2467 DOI: 10.1002/hyp.9726

Fleming, S.W.: An informal survey of watershed model users: preferences, applications, and rationales, Streamline Watershed Manage. Bull. 13(1), 32-35, http://www.forrex.org/sites/default/files/publications/articles/Streamline_Vol13_No1_Art6.pdf, 2009.

Franceschini, F., Maisano, D., Mastrogiacomo, L.: Influence of omitted citations on the bibliometric statistics of the major manufacturing journals. Scientometrics 103, 1083–1122, doi: 10.1007/s11192-015-1583-9, 2015.

Franceschini, F., Maisano, D., Mastrogiacomo, L.: The museum of errors/horrors in Scopus. J. Inf. Secur. 10, 174–182, doi: 10.1016/j.joi.2015.11.006, 2016.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H. and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18(12), 4839–4859, doi:10.5194/hess-18-4839-2014, 2014.

Guillaume, J. H. A., Helgeson, C., Elsawah, S., Jakeman, A. J. and Kummu, M.: Toward best practice framing of uncertainty in scientific publications: A review of Water Resources Research abstracts, Water Resour. Res., 53, 6744–6762, doi:10.1002/2017WR020609, 2017.

Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R. and Mao, Y.: The Variable Infiltration Capacity Model, Version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility, Geosci. Model Dev., 11, 3481–3496, doi:10.5194/gmd-11-3481-2018, 2018.

Hardwig, J.: The role of trust in knowledge, J. Philos. 88 (12), 693-708, doi: 10.2307/2027007, 1991.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J. and Weiler, M.: Karst water resources in a changing world: Review of hydrological modeling approaches, Rev. Geophys., 52, doi:10.1002/2013RG000443, 2014.

Hauser, M., Gudmundsson, L., Orth, R., Jézéquel, A., Haustein, K., Vautard, R., van Oldenborgh, G., Wilcox, L., Seneviratne, S., 2017. Methods and model dependency of extreme event attribution: the 2015 European drought. Earth's Future 5, 1034–1043. doi:10.1002/2017EF000612.

Holländer, H., Blume, T., Bormann, H., Buytaert, W., Chirico, G., Exbrayat, J., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Blöschl, G., Flühler, H., 2009. Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. Hydr. Earth Syst. Sci. 13, 2069–2094. doi:10.5194/hess-13-2069-2009.

Horrobin, D. F.: The philosophical basis of peer review and the suppression of innovation. J. Am. Med. Ass. 263, 1438-1441, doi:10.1001/jama.1990.03440100162024, 1990.

Jakeman, A.J., Letcher, R.A., and Norton, J.P.: Ten iterative steps in development and evaluation of environmental models, Env. Modell. Softw. (21), 602-614, doi: 10.1016/j.envsoft.2006.01.004, 2006.

Jørgensen S. E. and Müller, F. (eds.): Handbook of ecosystem theories and management, Boca Raton FL: Lewis Publishers, 2000.

Kavetski, D. and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, Water Resour. Res., 47(W11511), doi:10.1029/2011WR010748, 2011.

Knorr Cetina, K.D.: The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science, Oxford: Pergamon Press, 1981

Kollet, S., Sulis, M., Maxwell, R. M., Paniconi, C., Putti, M., Bertoldi, G., Coon, E. T., Cordano, E., Endrizzi, S., Kikinzon, E., Mouche, E., Mügler, C., Park, Y. J., Refsgaard, J. C., Stisen, S. and Sudicky, E.: The integrated hydrologic model intercomparison project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks, Water Resour. Res., 53(1), 867–890, doi:10.1002/2016WR019191, 2017.

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., Hall, A., Rott, H., Brutel-Vuilmet, C., Kim, H., Ménard, C. B., Mudryk, L., Thackeray, C., Wang, L., Arduini, G., Balsamo, G., Bartlett, P., Boike, J., Boone, A., Chéruy, F., Colin, J., Cuntz, M., Dai, Y., Decharme, B., Derry, J., Ducharne, A., Dutra, E., Fang, X., Fierz, C., Ghattas, J., Gusev, Y., Haverd, V., Kontu, A., Lafaysse, M., Law, R., Lawrence, D., Li, W., Marke, T., Marks, D., Nasonova, O., Nitta, T., Niwano, M., Pomeroy, J., Raleigh, M. S., Schaedler, G., Semenov, V., Smirnova, T., Stacke, T., Strasser, U., Svenson, S., Turkov, D., Wang, T., Wever, N., Yuan, H. and Zhou, W.: ESM-SnowMIP: Assessing models and quantifying snow-related climate feedbacks, Geosci. Model Dev. Discuss., (July), 1–32, doi:10.5194/gmd-2018-153, 2018.

Krueger, T., Page, T., Hubacek, K., Smith, L., Hiscock, K.:. The role of expert opinion in environmental modelling. Env. Modell. Softw. (36), 4–18. doi:10.1016/j.envsoft.2012.01.011, 2012.

Krysanova, V., Vetter, T., Eisner, S., Huang, S., Pechlivanidis, I., Strauch, M., Gelfan, A., Kumar, R., Aich, V., Arheimer, B., Chamorro, A., van Griensven, A., Kundu, D., Lobanova, A., Mishra, V., Plötner, S., Reinhardt, J., Seidou, O., Wang, X., Wortmann, M., Zeng, X. and Hattermann, F. F.: Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide—a synthesis, Environ. Res. Lett., 12(10), 105002, doi:10.1088/1748-9326/aa8359, 2017.

Kuhn, T.S.: The structure of scientific revolutions, Chicago, IL: University of Chicago Press, 1962 (fourth edition: 2012).

Latour, B. and Woolgar, S.: Laboratory life: The construction of scientific facts, Princeton, NJ: Princeton University Press, 1986 (reprint of 1979).

Law, J. and Mol, A.: Situating technoscience: an inquiry into spatialities, Environment and Planning D: Society and Space (19), 609-621, doi: 10.1068/d243t, 2001.

Leavesley, G.H., Lichty, R.W., Troutman, B.M., and Saindon, L.G., 1983, Precipitation-runoff modeling system—User's manual: U.S. Geol. Surv. Water Resour. Invest. Rep. 83-4238.

Leavesley, G. H., Markstrom, S. L., Brewer, M. S. and Viger, R. J.: The Modular Modeling System (MMS) - The physical process modeling component of a database-centered decision support system for water and power management, Water. Air. Soil Pollut., 90(1–2), 303–311, doi:10.1007/BF00619290, 1996.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99(D7), 14,415–14,458.

Livingstone, D. N.: Putting science in its place: Geographies of scientific knowledge. Chicago, IL: University of Chicago Press, 2003.

Loague, K. M. and Freeze, R. A.: A comparison of rainfall runoff modelling techniques on small upland catchments, Water Resour. Res., 21(2), 229–240, 1985.

Mahony, M., and Hulme, M.: Epistemic geographies of climate change: Science, space and politics. Progress in Human Geography, 42, 395-424, doi: 10.1177/0309132516681485, 2018.

Markstrom, S.L., Regan, R.S., Hay, L.E., Viger, R.J., Webb, R.M.T., Payn, R.A., and LaFontaine, J.H., 2015, PRMS-IV, the precipitation-runoff modeling system, version 4: U.S. Geological Survey Techniques and Methods, book 6, chap. B7, 158 p., https://dx.doi.org/10.3133/tm6B7.

Martin, B.: Research grants: problems and options. Australian Universities' Review 43, 17-22, available at: http://www.bmartin.cc/pubs/00aur.pdf, 2000.

McDonnell, J. J. and Woods, R.: On the need for catchment classification, J. Hydrol., 299, 2–3, doi:10.1016/j.jhydrol.2004.09.003, 2004.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M. and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, Hydrol. Process., 25(4), 511–522, doi:10.1002/hyp.7841, 2011.

Melsen, L., Addor, N., Mizukami, N., Newman, A., Torfs, P., Clark, M., Uijlenhoet, R. and Teuling, R.: Mapping (dis)agreement in hydrologic projections, Hydrol. Earth Syst. Sci., 22, 1775–1791, doi:10.5194/hess-22-1775-2018, 2018a.

Melsen, L., Vos, J., and Boelens, R.: What is the role of the model in socio-hydrology? Discussion of "Prediction in a socio-hydrological world", Hydr. Sci. J., doi: 10.1080/02626667.2018.1499025, 2018b.

Metcalfe, P., Beven, K. and Freer, J.: Dynamic TOPMODEL: A new implementation in R and its sensitivity to time and space steps, Environ. Model. Softw., 72, 155–172, doi:10.1016/j.envsoft.2015.06.010, 2015.

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D. and Rajagopalan, B.: How do hydrologic modeling decisions affect the portrayal of climate change impacts?, Hydrol. Process., 30(7), 1071–1095, doi:10.1002/hyp.10684, 2016.

Miller, C.A. and Edwards, P.N. (eds): Changing the atmosphere: Expert knowledge and environmental governance. Cambridge, MS: MIT university press, 2001.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models. Part I: a discussion of principles, J. Hydrol., 10, 282–290, 1970.

Naylor, S.: Introduction: Historical geographies of science - places, contexts, cartographies. British Journal for the History of Science, 38(1), 1-12, doi: 10.1017/S0007087404006430, 2005.

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M. and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res. Atmos., 116(D12109), doi:10.1029/2010JD015139, 2011.

Pautasso, M.: The jump in network ecology research between 1990 and 1991 is a
Web of Science artefact. Ecol. Modell. 286, 11–12. doi: 10.1016/j.ecolmodel.2014.04.020, 2014.

Peck, S. L.: The hermeneutics of ecological simulation. Biol. Philos, 23, 283-402. doi:10.1007/s10539-008-9109-y, 2008.

Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, J. Hydrol. 279(1-4), 275-289.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.r-project.org/, 2018.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M. and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, J. Hydrometeorol., 17(1), 287–307, doi:10.1175/JHM-D-15-0054.1, 2016.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J. and DMIP Participants: Overall distributed model intercomparison project results, J. Hydrol., 298, 27–60, doi:10.1016/j.jhydrol.2004.03.031, 2004.

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. a., Drews, M., Hamilton, D. P., Jeppesen, E., Kjellström, E., Olesen, J. E., Sonnenborg, T. O., Trolle, D., Willems, P. and Christensen, J. H.: A framework for testing the ability of models to project climate change and its impacts, Clim. Change, 122(1–2), 271–282, doi:10.1007/s10584-013-0990-2, 2013.

Refsgaard, J. C. and Knudsen, J.: Operational validation and intercomparison of different types of hydrological models, Water Resour. Res., 32(7), 2189–2202, doi:10.1029/96WR00896, 1996.

Samaniego L., R. Kumar, and S. Attinger (2010), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res., 46,W05523

Savenije, H. H. G.: HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)," Hydrol. Earth Syst. Sci., 14(12), 2681–2692, doi:10.5194/hess-14-2681-2010, 2010.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W. and Weedon, G. P.: A global water resources ensemble of hydrological models: the earthH2Observe Tier-1 dataset, Earth Syst. Sci. Data, 9, 389–413, doi:10.5194/essd-2016-55, 2017.

Seibert, J.: On TOPMODEL's ability to simulate groundwater dynamics, Regionalization in Hydrology (Proc. Conf. at Braunschweig, March 1997) (ed. by B. Dickkrüger, M.J. Kirkby and U. Schröder), *IAHS Publication* 254: 211-220

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sc., 16(9), 3315–3325, doi:10.5194/hess-16-3315-2012, 2012.

Seibert, J., Vis, M., Lewis, E. and van Meerveld, I.: Upper and lower benchmarks in hydrological modeling, Hydrol. Process., 1–8, doi:10.1002/hyp.11476, 2018.

Shapin, S.: A Social History of Truth: Civility and science in seventeenth-century England, Chicago, IL, University of Chicago press, 1994.

Shapin, S.: Placing the view from nowhere: historical and sociological problems in the location of science, Trans. Inst. Br. Geogr. (23) 5–12, doi: 10.1111/j.0020-2754.1998.00005.x, 1998.

Singh, V. P., and Woolhiser, D.A.: Mathematical modeling of watershed hydrology, J. Hydrol. Eng., 7(4), 270-292, doi: 10.1061/(ASCE)1084-0699(2002)7:4(270), 2002.

Slater, A. G., Schlosser, C. A., Desborough, C. E., Pitman, A. J., Henderson-Sellers, A., Robock, A., Vinnikov, K. Y., Entin, J., Mitchell, K., Chen, F., Boone, A., Etchevers, P., Habets, F., Noilhan, J., Braden, H., Cox, P. M., de Rosnay, P., Dickinson, R. E., Yang, Z.-L., Dai, Y.-J., Zeng, Q., Duan, Q., Koren, V., Schaake, S., Gedney, N., Gusev, Y. M., Nasonova, O. N., Kim, J., Kowalczyk, E. A., Shmakin, A. B., Smirnova, T. G., Verseghy, D., Wetzel, P. and Xue, Y.: The Representation of Snow in Land Surface Schemes: Results from PILPS 2(d), J. Hydrometeorol., 2(1), 7–25, doi:10.1175/1525-7541(2001)002<0007:TROSIL>2.0.CO;2, 2001.

Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N., Anderson, E. A. and Cosgrove, B. A.: The distributed model intercomparison project – Phase 2: Motivation and design of the Oklahoma experiments, J. Hydrol., 418–419, 3–16, doi:10.1016/j.jhydrol.2011.08.055, 2012.

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P. and Tallaksen, L. M.: Comparison of hydrological model structures based on recession and low flow simulations, Hydrol. Earth Syst. Sc., 15(11), 3447–3459, doi:10.5194/hess-15-3447-2011, 2011.

Wainwright, S. P.: Science studies in physical geography: An idea whose time has come? Progress in Physical Geography, 36, 786-812, doi:10.1177/0309133312450997, 2012.

Weiler, M. and Beven, K.: Do we need a Community Hydrological Model?, Water Resour. Res., 51, 1–8, doi:10.1002/2014WR016731, 2015.

Zellmer, A. J., Allen, T. F. H., and Kesseboehmer, K.: The nature of ecological complexity: A protocol for building the narrative. Ecological Complexity, 3, 171-82, doi:10.1016/j.ecocom.2006.06.002, 2006.

# Figures

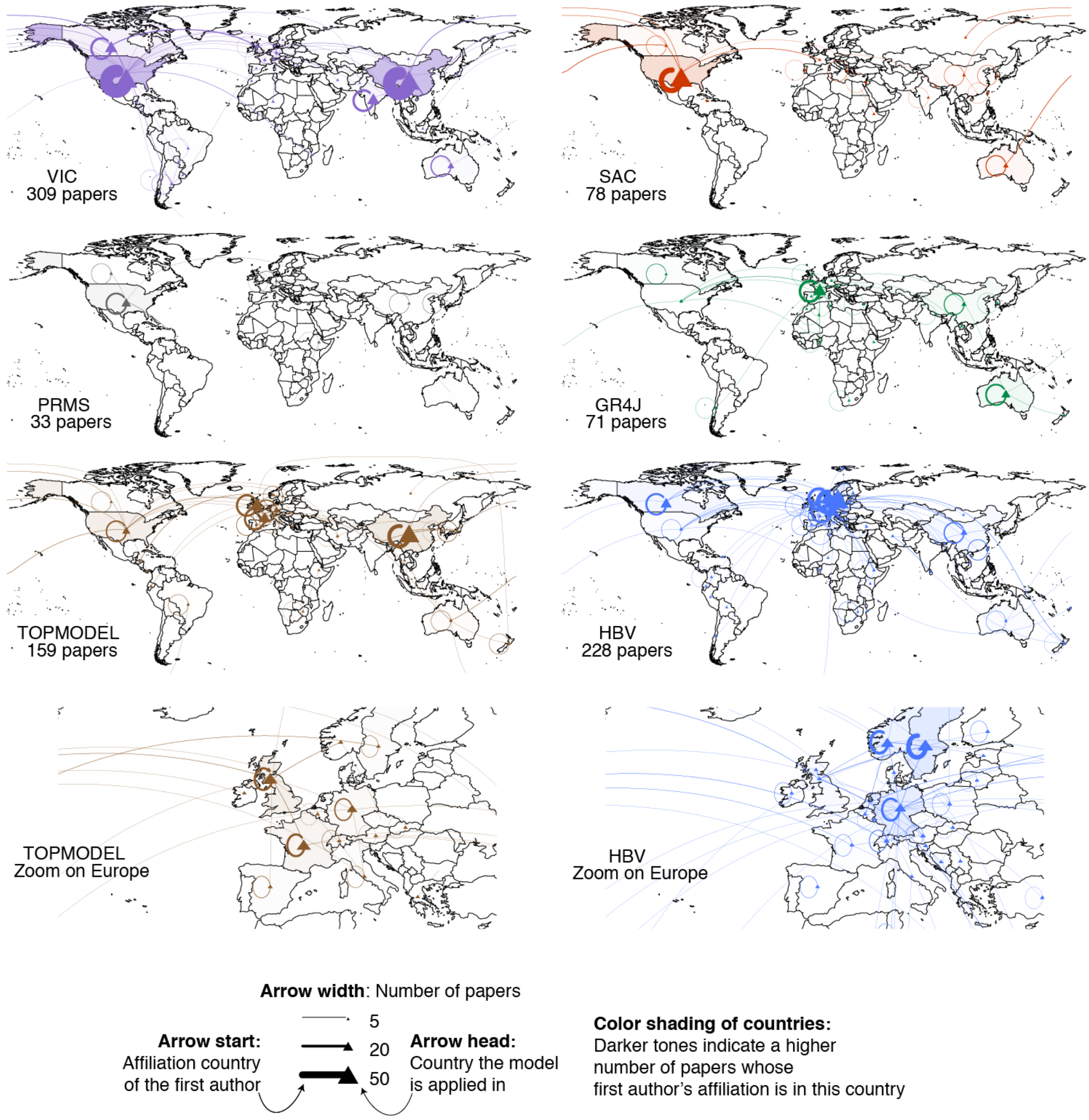


**Figure 1: Maps illustrating where models are predominantly used (see legend). Circular arrows indicate that the affiliation country of the first author is the same as the country the model is applied in. The nature of these maps means that papers for which the country of application could not be determined were excluded. The number of papers is indicated for each model. For space reasons, we do not show the mHM map.**

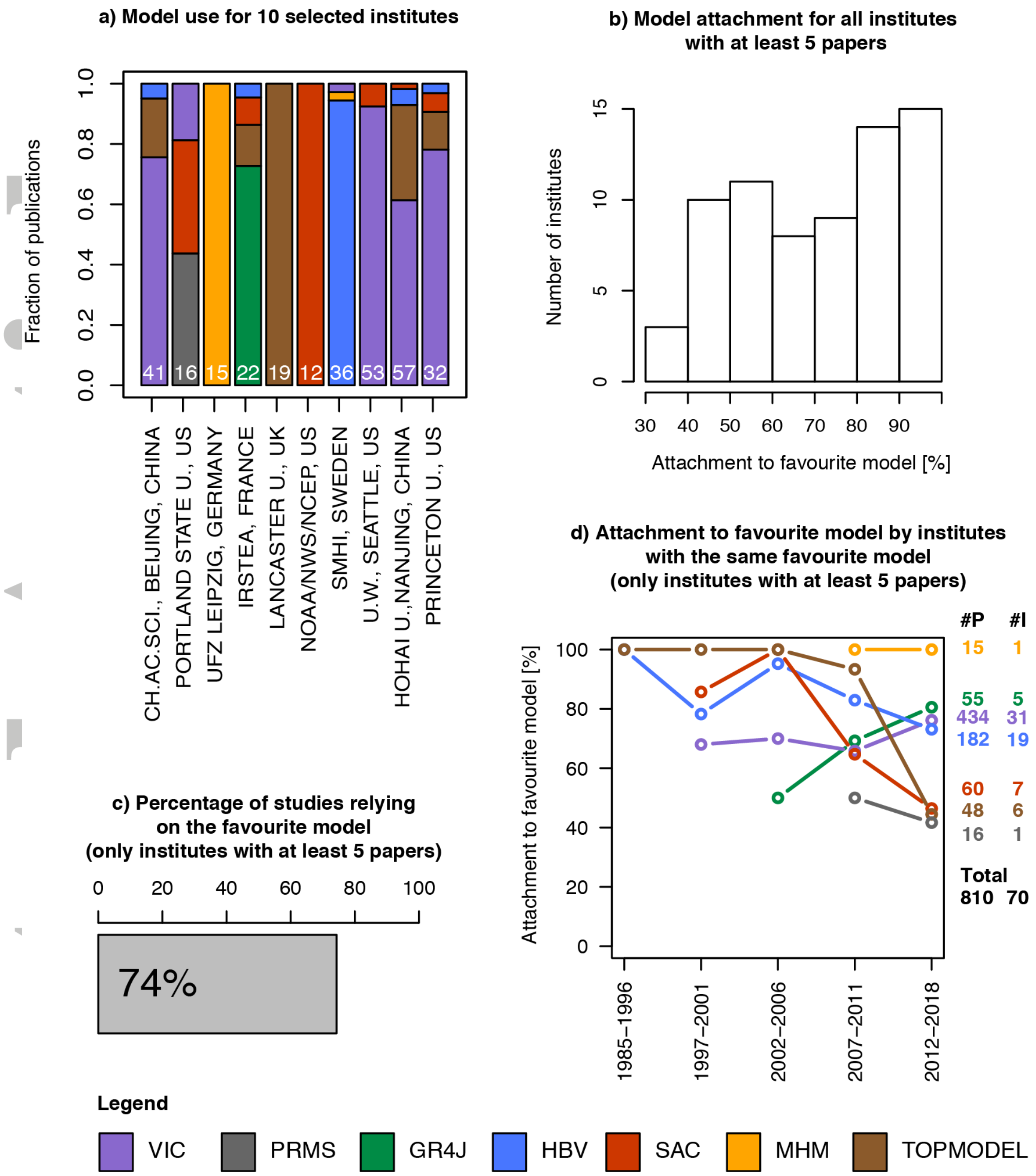


**Figure 2: a) Relative frequency of the seven models in publications from 10 selected institutes, the number of papers per institute is indicated in white. b) Histogram of institutes attachment to their favourite model. c) Percentage of studies relying on the favourite model, i.e., how frequently could the model used be correctly predicted using solely the institute of the first author? d) Variation of model attachment from 1988 to 2017, the institutes are pooled by favourite model, the number of papers and institutes involved in each line are indicated (#P and #I, respectively). Figures 2b to d are based on all the papers from institutes with at least 5 papers (810 papers, 70 institutes).**

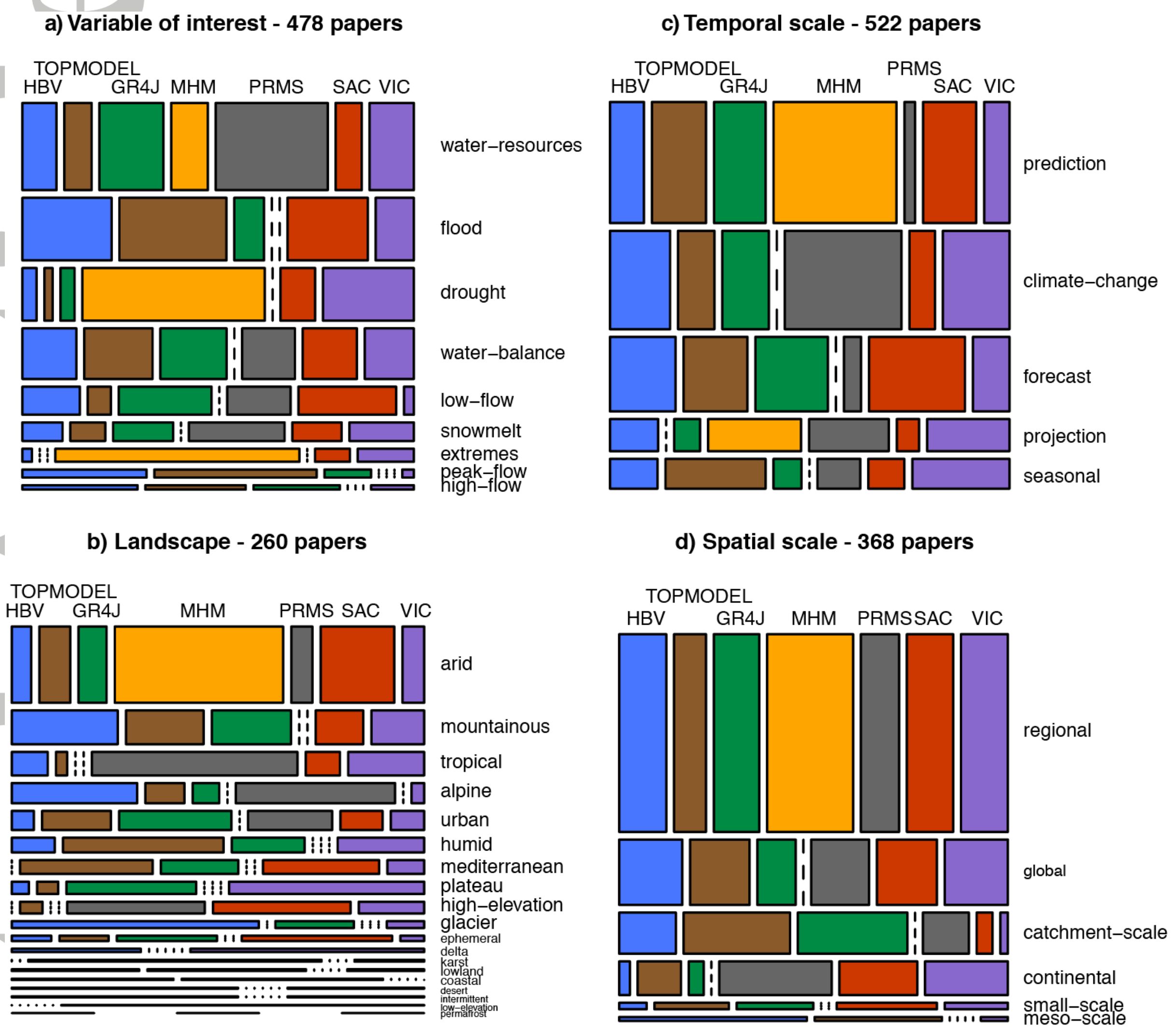


**Figure 3: Mosaic plots showing how frequently pre-defined keywords describing the variable of interest, the landscape, the temporal and spatial scale were employed in the titles, abstracts and keywords. The width of each box indicates how frequently the associated keyword is used for each model (the frequency was scaled by the number of papers for each model, i.e., the plots show the relative frequency if all the models were used in an equal number of papers). The titles indicate the number of papers used for each mosaic.**