Community Workflows to Advance Reproducibility in Hydrologic Modeling: Separating model-agnostic and model-specific configuration steps in applications of large-domain hydrologic models

Wouter Johannes Maria Knoben[1,1,1], Martyn P. Clark[2,2,2], Jerad Bales[3,3,3], Andrew Bennett[4,4,4], S. Gharari[5,5,5], Christopher B Marsh[1,1,1], Bart Nijssen[4,4,4], Alain Pietroniro[6,6,6], Raymond J Spiteri[1,1,1], David Gavin Tarboton[7,7,7], Andrew W Wood[8,8,8], and Guoqiang Tang[1,1]

[1]University of Saskatchewan
[2]University of Saskatchewan Centre for Hydrology, Canmore Coldwater Laboratory
[3]Unknown
[4]University of Washington
[5]University of Saskatchewan Coldwater Laboratory
[6]University of Calgary
[7]Utah State University
[8]National Center for Atmospheric Research (UCAR)

November 30, 2022

Abstract

Despite the proliferation of computer-based research on hydrology and water resources, such research is typically poorly reproducible. Published studies have low reproducibility because of both incomplete availability of the digital artifacts of research and a lack of documentation on workflow processes. This leads to a lack of transparency and efficiency because existing code can neither be checked nor re-used. Given the high-level commonalities between existing process-based hydrological models in terms of their input data and required pre-processing steps, more open sharing of code can lead to large efficiency gains for the modeling community. Here we present a model configuration workflow that provides full reproducibility of the resulting model instantiation in a way that separates the model-agnostic preprocessing of specific datasets from the model-specific requirements that specific models impose on their input files. We use this workflow to create both a continental and a local setup of the Structure for Unifying Multiple Modeling Alternatives (SUMMA) framework connected to the mizuRoute routing model. These examples show how a relatively complex model setup over a large domain can be organized in a reproducible and structured way that has the potential to accelerate hydrologic modeling for the community as a whole. We provide a tentative blueprint of how a community modeling paradigm can be built on top of workflows such as this. We term this initiative the "Community Workflows to Advance Reproducibility in Hydrologic Modeling' (CWARHM; pronounced "swarm').

Community Workflows to Advance Reproducibility in Hydrologic Modeling: Separating model-agnostic and model-specific configuration steps in applications of large-domain hydrologic models

W. J. M. Knoben[1], M. P. Clark[1,2], J. Bales[3], A. Bennett[4], S. Gharari[5], C. B. Marsh[5], B. Nijssen[6], A. Pietroniro[7], R. J. Spiteri[8], G. Tang[1], D. G. Tarboton[9], A. W. Wood[10]

[1]Centre for Hydrology, University of Saskatchewan, Canmore, Alberta, Canada
[2]Department of Geography and Planning, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[3]Consortium of Universities for the Advancement of Hydrologic Science, Inc, Cambridge, Massachusetts, USA
[4]Hydrology & Atmospheric Sciences, University of Arizona, Tucson, Arizona, USA
[5]Centre for Hydrology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[6]Civil and Environmental Engineering, University of Washington, Seattle, Washington, USA
[7]Schulich School of Engineering, Department of Civil Engineering, University of Calgary, Calgary, Alberta, Canada
[8]Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[9]Utah Water Research Laboratory, Utah State University, Logan, Utah, USA
[10]National Center for Atmospheric Research, Boulder, Colorado, USA

Key Points:

- Reproducible, transparent modeling increases confidence in model simulations and requires careful tracking of all model configuration steps
- We show an example of model configuration code applied globally that is traced and shared through a version control system
- Standardizing file formats and sharing of code can increase efficiency and reproducibility of modeling studies

---

Corresponding author: W. J. M. Knoben, `wouter.knoben@usask.ca`

Abstract

Despite the proliferation of computer-based research on hydrology and water resources, such research is typically poorly reproducible. Published studies have low reproducibility due to incomplete availability of data and computer code, and a lack of documentation of workflow processes. This leads to a lack of transparency and efficiency because existing code can neither be quality controlled nor re-used. Given the commonalities between existing process-based hydrological models in terms of their required input data and pre-processing steps, open sharing of code can lead to large efficiency gains for the modeling community. Here we present a model configuration workflow that provides full reproducibility of the resulting model instantiations in a way that separates the model-agnostic preprocessing of specific datasets from the model-specific requirements that models impose on their input files. We use this workflow to create large-domain (global, continental) and local configurations of the Structure for Unifying Multiple Modeling Alternatives (SUMMA) hydrologic model connected to the mizuRoute routing model. These examples show how a relatively complex model setup over a large domain can be organized in a reproducible and structured way that has the potential to accelerate advances in hydrologic modeling for the community as a whole. We provide a tentative blueprint of how community modeling initiatives can be built on top of workflows such as this. We term our workflow the "Community Workflows to Advance Reproducibility in Hydrologic Modeling" (CWARHM; pronounced "swarm").

1 Introduction

Confidence in published findings depends on the reproducibility of the experiments and analyses that support these findings. In computational Earth System sciences research, reproducibility requires knowledge of the computer code and data that underpin a given manuscript. Such computer code can range from a few lines of code that are used to turn data into figures or compute certain statistical properties of the data, to modern process-based hydrologic models that can contain many thousands of lines of code. Despite encouraging progress in journal policies (Blöschl et al., 2014; Clark, Luce, et al., 2021), it is still difficult to reproduce published findings in the hydrologic sciences (Hutton et al., 2016; Stagge et al., 2019). Stagge et al. (2019) estimate that results may only be reproducible for between 0.6% to 6.8% of nearly 2000 peer-reviewed manuscripts published in six hydrology and water resources journals, due to a lack of sufficiently clearly described methods and a lack of the necessary input data or processing code.

In complex process-based hydrologic model applications, one additional barrier to reproducibility is the effort required to configure the model. It is not uncommon to hear claims that in such modeling studies 80% of overall effort is spent on configuring the model for a specific use case, and only 20% of overall effort is spent on using the model to answer research questions (e.g., Table 2.8 in Miles, 2014). Model configuration efforts are spent on assembling appropriate data sources for meteorological forcing data and geospatial parameter fields, wrangling these data into the specific format required by the model, defining appropriate model settings, and specifying the required computational infrastructure (e.g., finding the right collection of software libraries, installing or compiling the model, creating the required scripts to run the model). Additional time costs arise from dealing with the subjectivity in defining appropriate computational sub-domains (such as where to draw the boundaries for Hydrologic Response Units (Flügel, 1995)), interpreting soil and land cover maps, aggregating geospatial data into some form of representative value for a computational unit, and the associated iterative model configuration and testing steps. This model configuration process is typically poorly documented and extremely time-consuming. In short, the reproducibility problem for process-based hydrologic modeling occurs in part because of the lack of efficiency in model configuration tasks.

Reproducibility of computational science can be improved by following certain recommended best practices for open, accessible, and reproducible science (e.g., Gil et al., 2016; Hutton et al., 2016; Sandve et al., 2013; Stodden & Miguez, 2013). Most focus is currently on advancing the FAIR principles, which state that data, code, and methods must be Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). Reproducibility requires FAIR data, but also includes sharing details about hardware, software versions, and data versions (Añel, 2017; Bast, 2019; Hut et al., 2017; Sandve et al., 2013). The environmental modeling community is interacting with these prescribed best practices in multiple ways. Choi et al. (2021) identify three ongoing main thrusts aimed at making computational environmental science more open, reusable, and reproducible. First, data and models are increasingly openly available online through services as GitHub, Hydroshare, and institutional repositories. Second, computational environments are increasingly recorded and standardized through container applications (e.g., Docker, Singularity) or in self-documenting notebooks. Third, Application Programming Interfaces (APIs) such as the pySUMMA API (Choi, 2020; Choi et al., 2021) make interacting with complex models or data increasingly easier. In practice however, most progress in FAIR science is arguably on Accessibility, whereas the other aspects of FAIR have received less attention.

A key issue is that little attention is devoted to efficient reproducibility of the full modeling workflow, which includes data acquisition, data preprocessing, model installation, model runs and post-processing of simulations. Efficiency is promoted in a general sense through freely shared code and packages that perform specific tasks in the modeling chain (for example, see Slater et al., 2019, for an overview of R packages that can be used to populate a modeling workflow), and with model-specific tools such as VIC-ASSIST (Wi et al., 2017). Dedicated efforts to ensure end-to-end reproducibility of modeling studies are less common. Exceptions are Leonard and Duffy (2013, 2014, 2016), who provide an in-depth description of a web-based interface for data preprocessing and visualization of simulations from the PIHM model, geographically constrained to the United States; Havens et al. (2019, 2020), who provide an end-to-end workflow for setting up, running, and analyzing a physics-based snow model; Vorobevskii et al. (2020); Vorobevskii (2022), who develop an R package that sets up a simple hydrologic model anywhere on the planet for a given domain discretization shapefile provided by the user; and Coon and Shuai (2022), who provide a Python-based tool to configure watershed models across the United States. Compared to sharing a model's input and output data (which would also enable a study to be reproduced), sharing complete workflows can be more efficient in terms of required storage space. A workflow also provides a transparent record of all modeling decisions and enables a more broadly defined form of reproducibility in which a study can be repeated for a different region, a different data set, or a different version of the same model to see if the original conclusions still hold.

The examples mentioned in the previous paragraph show that it is possible to document workflows for a specific model (or, perhaps more accurately, for a specific version of a model). A further challenge is in designing workflows in such a way that parts of a workflow that configures Model A can be re-used in a workflow that configures Model B. We refer to such a design as separating the model-agnostic and model-specific parts of model configuration (see also Miles, 2014; Miles & Band, 2015, for an example of this concept using EcoHydroLib for general data preprocessing and RHESSysWorfklows for creating model-specific input files applied to small watersheds across the CONUS and Australia). In the case of process-based hydrologic modeling, models such as VIC (Hamman et al., 2018; Liang et al., 1994), MESH (Pietroniro et al., 2007), SUMMA (Clark et al., 2015a, 2015b; Clark, Zolfaghari, et al., 2021) and SVS (Husain et al., 2016) can be different in how they discretize the modeling domain, the physical processes they include, and the equations used to describe a given process. However, at their core, these models are designed to solve the same general water and energy conservation equations (Clark, Zolfaghari, et al., 2021).

Consequently, the data requirements for a myriad of extant hydrologic models will vary in the specifics, but are similar in a general sense. In particular, process-based hydrologic models have similar needs for meteorological forcing data and geospatial parameter fields. Preprocessing of these similar data requirements does not need to rely on specifics of the models themselves. For example, in the case of satellite-based MODIS land cover data, model-agnostic steps are (1) downloading the source data, (2) stitching the source data together into a coherent global map, (3) projecting this map into the Coordinate Reference System of interest, (4) subsetting from the global data only the domain of interest, and (5) mapping the resulting data in pixels onto model elements. Model-specific steps would be to convert the resulting information (i.e., which pixels/land classes are present per model element) to the specific format a model requires (e.g., storing the most common land class per model element as a value in a netCDF file which the model reads during initialization), and, if necessary, perform some form of data transformation to connect land class data to model parameter values or settings (e.g., by defining a lookup table that contains parameter values for each land cover type). Community-wide efficiency gains are possible if workflows distinguish between model-agnostic and model-specific steps and enable straightforward re-use of the workflow for model-agnostic steps (see also Essawy et al., 2016; Gichamo et al., 2020, who make this argument in the context of web-based model configuration tools).

The previous discussion leads us to conclude that the hydrologic modeling community can substantially improve how it shares model configuration code across modeling groups. The key issue is that model physics code is increasingly distributed under open-source licenses but the code that creates the necessary model inputs is typically neither well-documented nor available without contacting the model developers. To move towards a culture of community Earth System modeling, we define three distinct steps:

1. For a given model, model configuration code should be publicly available and divided into model-agnostic and model-specific steps;
2. The configuration workflows of multiple different models, ideally using different data sets, should be integrated into a proof-of-concept of a generalized model configuration workflow;
3. A community-wide collaborative effort should refine the proof-of-concept into a flexible model configuration framework.

The purpose of this paper is to introduce an open-source model configuration workflow that enables full reproducibility of a process-based hydrologic model setup for any location on the planet, with the workflow code divided into model-agnostic and model-specific parts. In other words, we perform the first of the three steps outlined above. This advances our immediate goal of using this model configuration for a variety of projects by reducing the time commitment needed to create model configurations for different domains and by increasing confidence in the modeling outcomes due to increased transparency and the possibility to reproduce results. Our broader goal is to foster a community modeling culture within the Earth System sciences.

The workflow described in this manuscript contributes to this goal in two separate ways. First, our code is openly accessible and therefore reusable by others who wish to use all or part of it for their own experiments. Second, the documented lack of reproducible hydrologic science (e.g., Stagge et al., 2019) suggests that there are barriers within the hydrologic community to adopt more reproducible science. By providing a full example of how a reproducible modeling study can be designed, we intend to lower at least some of these barriers. A model-agnostic workflow approach, as proposed here, would also conform directly to ISO 9001 requirements for quality assurance and quality control systems for software development, as the World Meteorological Organization (WMO) describes in its guidance to WMO members on implementing a quality management system for national meteorological and hydrological services (World Meteorological Organization, 2017).

The remainder of this paper is organized as follows. In Section 2, we outline several high-level design considerations for reproducible modeling workflows and describe how we implemented these principles in an example of such a workflow. The example workflow uses open-source input data with global coverage, an open-source, spatially distributed, physics-based hydrologic modeling framework (SUMMA; Clark et al., 2015a, 2015b; Clark, Zolfaghari, et al., 2021), and an open-source network routing model (mizuRoute; Mizukami et al., 2016, 2021) to generate hydrologic simulations across multiple spatial scales. Technical details about the models and a step-by-step description of the workflow code are given in Appendix A. In Section 3, we present three test cases, covering large-domain (global, continental) and local-scale model configurations to show that a single workflow can be used to configure experiments that vary in terms of spatial and temporal resolution and coverage. In Section 4, we reflect on the current state of reproducibility in large-domain hydrologic modeling, with particular focus on why existing efforts have seen only limited uptake and outline a path forward.

2 Increasing efficiency and reproducibility in Earth System modeling

2.1 Workflow design considerations

The reproducibility of modeling studies can be improved through openly published workflows that track all decisions made during model configuration. We propose four general guidelines for such model configuration workflows in the Earth System sciences. These guidelines are informed by existing efforts to promote reproducibility and efficiency in large-domain modeling efforts, and by our own experience with creating such large-domain model configurations for process-based hydrologic models. We consider challenges for novice and advanced modelers. Briefly, our recommendations are as follows:

1. **Separate model-agnostic and model-specific tasks**. The steps in the workflow must remain model-agnostic for as much of the workflow as possible and provide outputs in standardized, commonly used data formats. This increases the potential utility of the code base for use in different projects and for users of different models.
2. **Clarity for modelers**. The workflow must be easily accessible and usable in its default form. A clear structure of the code accompanied by accurate documentation and in-line comments increase the ease-of-use for novice and advanced modelers alike.
3. **Modularity encourages use beyond the original application**. Customization of the workflow must be possible and easy. This makes it possible to adapt, improve, or change specific parts of the workflow to access new data sets, use new processing algorithms, or target different models.
4. **Traceability is key**. Every outcome of each step in the workflow must be accompanied by metadata that describe the configuration code that generated the outcome. This guarantees that, even if changes are made to the model configuration code, any workflow outcome can still be traced back to its original settings.

In Section 2.2 we discuss an example of a model configuration workflow based on these design considerations. In Section 2.3 we first provide a general description of model configuration steps and then expand on each of the four points outlined above.

2.2 An example workflow for large-domain hydrologic modeling

2.2.1 Workflow description

Based on the design considerations listed in Section 2.1, we created a model configuration workflow for the Structure for Unifying Multiple Modeling Alternatives (SUMMA; Clark et al., 2015a, 2015b; Clark, Zolfaghari, et al., 2021) hydrologic model and the mizuRoute routing model (Mizukami et al., 2016, 2021). Briefly, SUMMA is a process-based, spatially-distributed hydrologic model that can be used to simulate the water and energy balance for given locations in space. mizuRoute is a vector-based routing model that can be used to route runoff from a hydrologic or land surface model through a river network. Detailed descriptions of both models can be found in Section A1. We selected both models for their flexible nature, computational capacity to model very large domains and availability of local expertise. Implementing configuration code for specific models (i.e., SUMMA and mizuRoute) in a generalized workflow, as we describe in this paper, is the first step on a possible path towards a community modeling culture that we outline in the Introduction.

Figure 1 provides a high level overview of our workflow in five key steps:

1. Workflow preparation, where workflow settings are defined and the necessary folder structures are generated;
2. Model-agnostic preprocessing, accomplishing data preparation steps that do not rely on any characteristics of the models being used. Data resulting from this step can thus be used for multiple different models;
3. Remapping of prepared data onto model elements. This step is listed as optional because not all models need this step;
4. Model-specific preprocessing to create model input files based on the prepared data sources, and generate model simulations;
5. Analysis and visualization to summarize model simulations into statistics and figures.

Progressively more detailed overviews of model-agnostic and model-specific tasks can be found in Appendix A2 and Figures A2, A3 and A4. Despite the seemingly large number of model-specific tasks in those figures, the time costs (in terms of code development) are larger for the model-agnostic tasks. The design considerations presented in this section and our implementation of them as described in Section 2.3 are comparable to existing efforts in the field of eco-hydrology involving the EcoHydroLib, RHESSysWorkflows and HydroTerre tools (Miles, 2014; Miles & Band, 2015; Leonard et al., 2019; Choi, 2021), suggesting that this is a logical way to organize modeling workflows.

2.2.2 Workflow scope

The workflow scope deliberately excludes spatial discretization and parameter estimation (Figure 2). The scope of our workflow implementation assumes that the user has access to a basin discretization stored as an ESRI shapefile that defines the area of interest as discrete modeling elements (e.g., grid cells, sub-basins). Such a discretization may be derived from digital elevation models (see e.g. TauDEM or the geospatialtools code base, Sazib, 2016; Tesfa et al., 2011; Chaney & Fisher, 2021), or obtained from existing basin discretization products, such as HydroBASINS (Lehner & Grill, 2013) or the MERIT Hydro basin delineation (Lin et al., 2019). Moreover, the workflow does not currently include fine-tuning of model parameter values through calibration or estimation from auxiliary data sources. These calibration methods require selecting from a wide variety of calibration algorithms, each with their own strengths and weaknesses (e.g., Arsenault et al., 2014), and an even wider variety of objective functions that express the (mis)match between a model's simulations and observations of hydrologic states and fluxes (e.g., Murphy, 1988; Clark, Vogel, et al., 2021; Gupta et al., 2008; McMillan, 2021; Mizukami et al., 2019; Nash & Sutcliffe, 1970; Olden & Poff, 2003; Pushpalatha et al., 2012), relying on a variety of further choices related to spatial scaling (e.g., Samaniego et al., 2010), regionalization (e.g., Bock et al., 2015) and regularization of the calibration problem (e.g., Doherty & Skahill, 2006). These model calibration choices are not easily standardized and require auxiliary data in the form of observations that are not readily available globally. The modular nature of our workflow implementation allows methods for basin discretization and parameter estimation to be integrated easily into our existing code base, but doing so is planned for future work in an attempt to keep the scope of this first workflow example manageable.

2.2.3 Workflow execution

We present this workflow as a collection of Bash and Python scripts, stored inside a folder structure that clearly indicates the appropriate order in which the scripts should be executed (see Section 4.2.3 for a discussion of the choice to use scripts instead of other options). The latest version of the workflow is available through GitHub: `https://github.com/CH-Earth/CWARHM`. The GitHub repository also contains further documentation that helps a user set up the required computational environment and provides succinct explanations of the purpose of various scripts, decisions and assumptions in cases where such explanations are necessary. Lastly, the repository contains the basin discretization used for our third test case that divides the upper part of the Bow River basin (Alberta, Canada) into discrete modeling elements, so that users have immediate access to all the materials needed to implement our workflow.

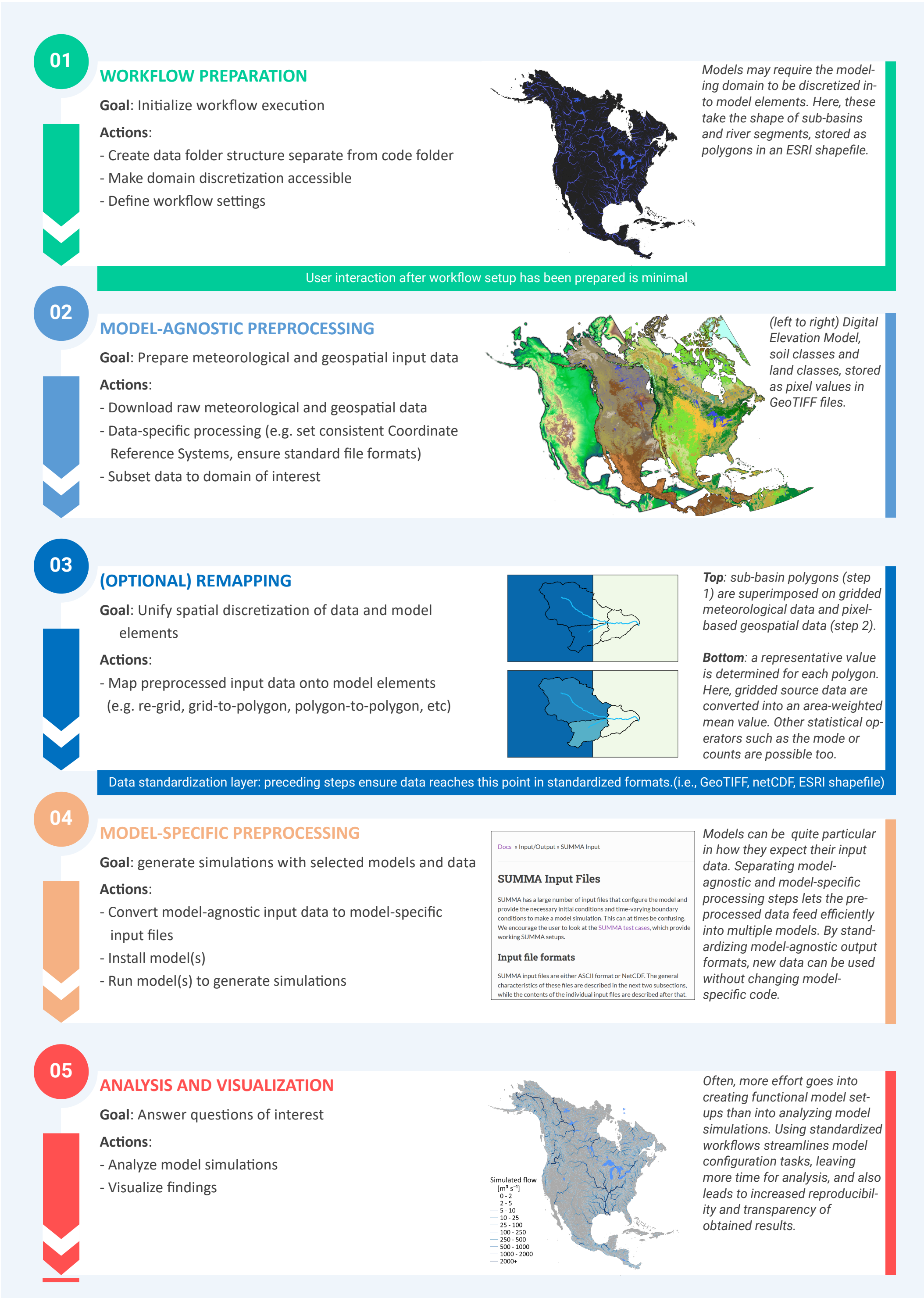

Figure 1. High-level overview of a workflow that separates model-agnostic and model-specific tasks. Model-agnostic tasks are shown in blue and model-specific tasks are shown in orange and red. A similarly high-level but more technical flowchart of such a workflow, using SUMMA (a process-based hydrologic model) and mizuRoute (a routing model) as example models, can be found in Figure A2. Technical details of our implementation of model-agnostic and model-specific processing steps can be found in Figures A3 and A4 respectively.

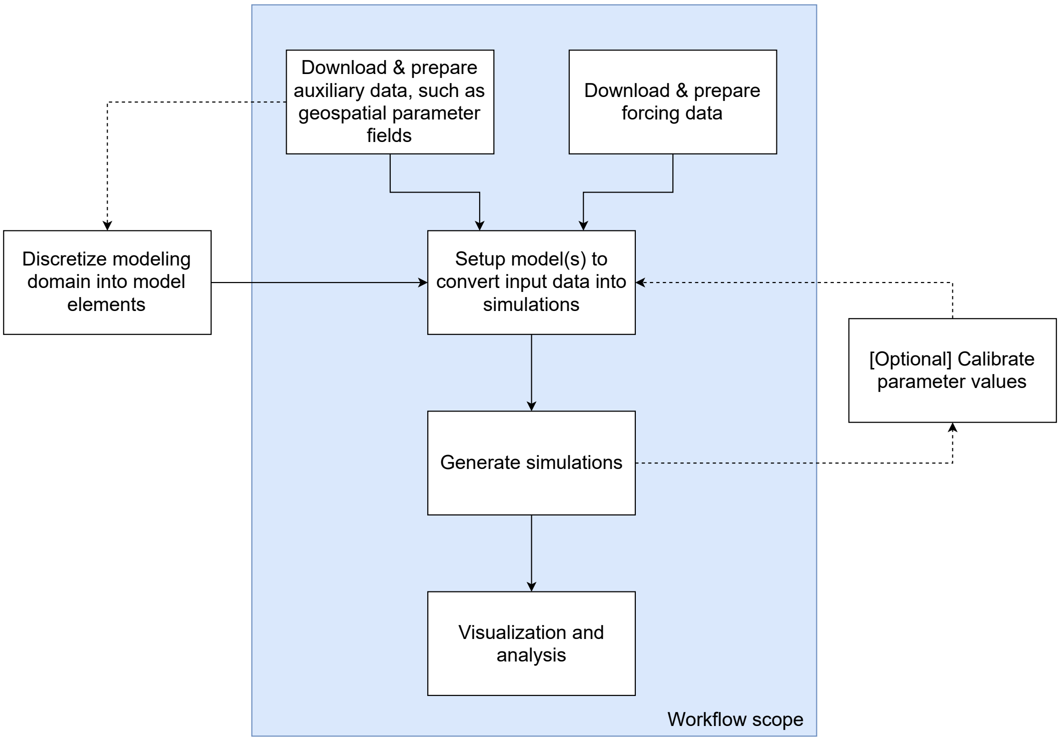

Figure 2. Schematic overview of a typical modeling workflow, with the scope of the example workflow described in this paper shown by the colored box. Dashed lines indicate potential connections between elements (such as geospatial parameter fields informing basin discretization, and parameter calibration feeding back into the model setup step where parameters for a new run are defined) that are not yet included as part of our workflow.

2.3 Implementation of workflow design recommendations

2.3.1 Separate model-agnostic and model-specific tasks

Our first design principle recommends separating model-agnostic and model-specific tasks. Model-agnostic tasks (shown in blue in Figure 1; light grey in Figure A2 and Figure A3) are those tasks that are the same regardless of the model being used, under the assumption that the model requires a given data input at all. In our workflow implementation these tasks include the downloading of meteorological forcing data and geospatial parameter fields (i.e., a digital elevation model (DEM), soil classes and vegetation classes), in some cases clipping raw datasets to the domain of interest and mapping of these data onto model elements such as grid cells or catchments. Fully model-agnostic outputs in this example are netCDF (.nc) files of meteorological forcing data (i.e., gridded hourly data at 0.25° latitude/longitude resolution) and GeoTIFF (.tif) files of various geospatial parameter fields.

Model-specific tasks (shown in orange and red in Figure 1; dark grey in Figure A2 and Figure A4) involve installing the chosen models, transforming the pre-processed data into the specific format the model requires, and running the models. In our workflow implementation this involves finding the mean elevation, mode land class and mode soil class per model element and exporting certain information about the modeling elements (area, latitude and longitude location, slope of the river network, etc.) into the netCDF files our models expect.

Due to the complex nature of existing models and their long histories of development, certain tasks cannot be cleanly separated into model-agnostic and model-specific tasks. The mapping of prepared forcing data and geospatial parameter fields onto model elements (shown in dark blue in Figure 1; intermediate grey shade in Figure A2 and Figure A3) is an example of such a task. Certain models run on the same spatial resolution as the forcing and/or geospatial data grid, or are able to ingest gridded data in their native alignment and internally map these onto the required model discretization. In our case, this remapping must be done outside the models. In the case of forcing data, the model-agnostic output of meteorological forcing files are mapped onto the model elements (catchments in this case), resulting in catchment-averaged model forcing. Temperature time series are further modified with catchment-specific lapse rates to account for elevation differences between the forcing grid and model elements. In the case of parameter fields, intersections between the model-agnostic GeoTIFF files and the shapefile of the modeling domain are generated. These intersections show how often each elevation level, soil class, and land class occurs in each model element. These processes cannot be called truly model-agnostic because some models do not require them, but neither are they fully model-specific. To ensure maximum usability for different models, workflows must therefore be as modular as possible so that modelers can mix and match from available code to suit the particularities of their chosen model (i.e., our third design principle, described later).

2.3.2 General layout and workflow control

Our second design principle prescribes an intuitive interface for hydrologic modelers. We recognize two elements here: first, the code and data structure must be clear and easy to understand. Second, interacting with the workflow must be straightforward. Our example implementation strives to achieve both of these goals through a clean separation of code and data and the use of a single configuration file (hereafter referred to as a "control file") that outlines high-level workflow decisions such as file paths, spatial and temporal extent of the experiment, and details about the shapefiles that contain the domain discretization. Using configuration or control files is common practice in software design applications (see e.g. Sen Gupta et al., 2015) and avoids the need to introduce hardcoded elements such as file paths and variable values in the code itself.

In a typical application of our example workflow, the user first creates a local copy of the code provided on our GitHub repository. We refer to this local code as the "code directory". The user would then specify a path in the control file that specifies where workflow data (such as forcing and parameter data downloads, model input files and model simulations) will be stored. The workflow is set up to read this path from the control file, create the specified folder structure and store all data for a given modeling domain in the user-specified data folder (referred to as the "data directory"). This allows a clean separation between the workflow code itself and the data downloaded and preprocessed by the workflow code (Figure 3). The workflow's default settings ensure that the data directory is populated with folders and subfolders with descriptive names, making navigation of the generated data clear.

Table 1 contains a subset of the information that is stored in the control file that defines the workflow settings for a model configuration for the Bow River at Banff, Canada (see Section 3 for a description of this test case). The control file contains the high-level information needed by the workflow, such as the name of the user's shapefiles, the names of required attributes in each shapefile, the spatial extent of the modeling domain, the years for which forcing data should be downloaded, and file paths and names for all required data. The workflow scripts read information from the control file as needed. Keeping all information in one place enables a user to quickly generate model configurations for multiple domains, without needing to scour all individual scripts for hardcoded file paths, domain extents, etc. For example, changing the simulation period for a given do-

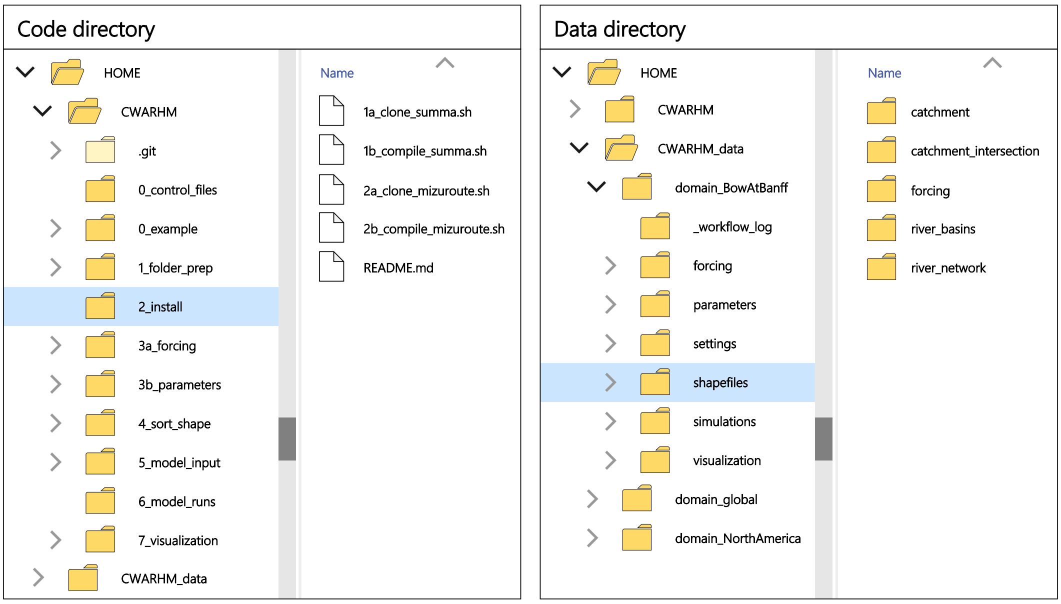

Figure 3. Example of separated code and data directories. The code directory (left) contains the scripts as available on the repository's GitHub page. The data directory (right) contains the forcing data, parameter data, setting files, shapefiles and model simulations that are used and generated by the workflow code.

main requires changing two values in the control file, after which selected code can be re-run to download and preprocess the necessary forcing data and run new simulations. To configure our chosen models for a new domain (assuming that no changes to the model or desired data sets are introduced), a user only needs to provide a new domain discretization file and update in the control file the name of the domain (so that a new data folder can be generated), the names of the discretization files, and the bounding box of the new domain. The workflow can then be fully re-run to create a model configuration for the new domain, without any changes being made to the workflow scripts themselves.

2.3.3 Flexibility at each step of model setup

Our third design principle recognizes that process-based models are complex entities and that the setup procedures for any given model are model- or even experiment-specific. Not all models will need to go through the same configuration steps, nor will every model experiment need the settings as defined in our example workflow. Our example workflow (Figure 1; details in Section A2) therefore aims to encourage adaptation beyond our original application through modularity and documentation.

First, we have chosen to present the workflow as a collection of scripts (i.e., the workflow code is stored in simple text files that can be executed from the command line) rather than a Python package, R library, executable module or similar, so that the user has straightforward access to the workflow code. This presentation simplifies adapting the code to different models or experiments by lowering the skill threshold needed to make adaptations to our code base, and is likely closer to the ways in which model configuration is currently often done. Second, the workflow separates model setup into numerous small tasks (see Figures A2 -A4) and saves all intermediate results to files. This modularity makes it straightforward to branch out from our chosen defaults at any given step in the modeling workflow. Third, for this iteration of our workflow, we have chosen to move

Table 1. Example of part of a workflow control file, showing settings for the Bow at Banff test case (see Section 3; actual control file available on the GitHub repository - see the Section "Open Research" at the end of this manuscript). These control files are simple text files containing three columns. The "Setting" column contains specific strings that each script in the repository looks for to identify which line in the control file contains the information the script needs. The "value" column contains the actual information, such as file paths, names of shapefiles and shapefile attributes, etc. Descriptions of each field are included for the user's benefit but not used by the setup scripts. The benefit of collecting all information and settings in a single file is that it avoids hard-coding this information in the workflow itself, making it straightforward to apply the same workflow for a new experiment by simply updating the control file.

Setting	Value	Description
Modeling domain settings		
root_path	/user/CWARHM_data	Root folder where data will be stored.
domain_name	BowAtBanff	Used as part of the root folder name for the prepared data.
Settings of user-provided catchment shapefile		
catchment_shp_path	default	If 'default', uses 'root_path/domain_[name]/shapefiles/catchment'.
catchment_shp_name	bow_dist_elev_zone.shp	Name of the catchment shapefile. Requires extension '.shp'.
catchment_shp_gruid	GRU_ID	Name of the GRU ID column (can be any numeric value, HRU's within a single GRU have the same GRU ID).
catchment_shp_hruid	HRU_ID	Name of the HRU ID column (consecutive from 1 to total number of HRUs, must be unique).
catchment_shp_area	HRU_area	Name of the catchment area column. Area must be in units [m^2].
catchment_shp_lat	center_lat	Name of the latitude column. Should be a value representative for the HRU.
catchment_shp_lon	center_lon	Name of the longitude column. Should be a value representative for the HRU.
Forcing settings		
forcing_raw_time	2008,2013	Years to download: Jan-[from],Dec-[to].
forcing_raw_space	51.7/-116.5/50.9/-115.5	Bounding box of the shapefile: lat_max/lon_min/lat_min/lon_max. Will be converted to ERA5 download coordinates in script. Order and use of '/' to separate values is mandatory.
forcing_time_step_size	3600	Size of the forcing time step in [s].

high-level decisions into the control file and leave various modeling decisions as assumptions in the workflow scripts. We have spent considerable effort on documenting any such assumptions (see Section A2) to let advanced users make targeted changes to the workflow code. Examples of these decisions include the number of soil layers used across the modeling domain, values for the initial model states, and default routing parameters. In future versions of our workflow, such decisions may be moved to a dedicated experiment-control file.

2.3.4 Code provenance

Our fourth design principle relates to traceability. The decision to separate code and data directories potentially introduces a disconnect between code and data, and situations may arise where it is no longer clear which version of a given piece of code generated a particular piece of data. This can happen in cases where the workflow code is updated after having already been used to create (part of) a model configuration. Although the changes to the workflow code can be tracked through version control systems such as Git, it is much more difficult to trace which version of the code generated the data. Every script in our example workflow therefore places both a log file and a copy of its code in the data sub-directory on which it operates. This ensures that, even if a user makes changes to the code directory, a record exists in the data directory of the specific code used to generate the files in that data directory. Copies of the model settings are stored in their simulation data directories by default so that simulation provenance can be traced as well.

3 Test cases

The test cases described in this section use the SUMMA and mizuRoute models. We refer the reader to Section A1 for details about both models and definitions of certain model-specific terms, such as Grouped Response Units (GRUs) and Hydrological Response Units (HRUs). For all test cases, meteorological input data are obtained from the ERA5 data set (Hersbach et al., 2020), elevation data are obtained from the MERIT Hydro data set (Yamazaki et al., 2019), land use data are obtained from the MODIS MCD12Q1 data set (Friedl & Sulla-Menashe, 2019), and soil data are obtained from the Soilgrids 250m data set (Hengl et al., 2017). Detailed descriptions of the input data can be found in Section A2.

3.1 Global model configuration

This first test case simulates hydrologic processes across planet Earth to illustrate the large-domain applicability of our approach. The global domain (excluding Greenland and Antarctica) is divided into 2,939,385 sub-basins or Grouped Response Units (GRUs; median GRU size is 36 km^2; mean size is 45 km^2) derived from the global MERIT basins data set (Lin et al., 2019). Simulations are run for a single month (1979-01-01 to 1979-01-31) at a 15-minute temporal resolution. Figure 4 shows summary statistics of several simulated variables. By design, we ran these simulations without a model spin-up period so that we might confirm our models function in regions where under typical conditions after model spin-up we would not expect to see much hydrologic activity (e.g., extremely water-limited regions). The value of this test case is to demonstrate that the workflow is applicable anywhere on the planet, and that the size of the model domain does not provide an insurmountable barrier to open and reproducible hydrologic modeling. The workflow documents every decision made during model configuration and enables repeatable simulations of this model domain with only a fraction of the original effort needed.

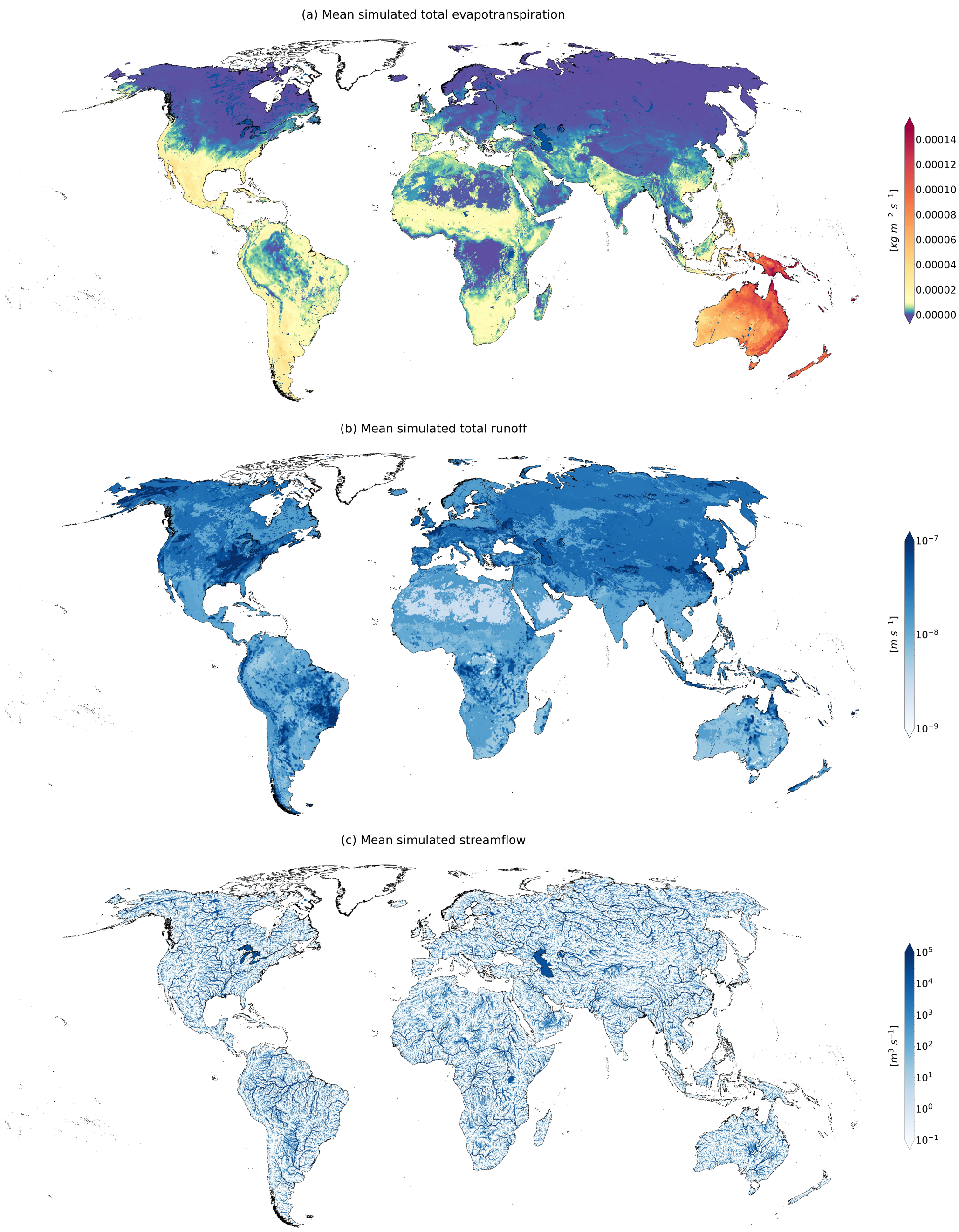

Figure 4. Overview of global simulations. SUMMA does not perform any computations for GRUs that are classified as being mostly open water, and mizuRoute was run without using its option to simulate routing through lakes and reservoirs. Lake delineations of lakes > 100 km^2 are obtained from the HydroLAKES data set (Messager et al., 2016) and used to mask open-water GRUs in this figure. Model setup uses default parameter values, and results are for illustrative purposes only. (a) Mean simulated total evapotranspiration, calculated as the sum of transpiration, canopy evaporation and soil evaporation. Note that the color scale has been designed to show global variability and local variability in Oceania simultaneously. (b) Mean runoff, calculated as the sum of surface runoff, downward drainage from the soil column, and lateral flow from the soil column. (c) Mean streamflow as determined by mizuRoute's Impulse Routing Function approach.

3.2 Continental model configuration

This second test case uses 40 years of hourly forcing data to simulate hydrologic processes over the North American continent and illustrates the combined large-domain and multi-decadal applicability of our approach. The continental domain is divided into 517,315 sub-basins (median GRU size is 33 km^2; mean size is 40 km^2) derived from the global MERIT basins data set (Lin et al., 2019). Simulations are run from 1979-01-01 to 2019-12-31, again at a 15-minute temporal resolution. Figure 5 shows summary statistics of several simulated variables: as expected, snow accumulation tends to be higher in mountainous and higher-latitude locations; total soil water values are lower in the arid regions of the central and western US and Canada and northern Mexico; evapotranspiration rates fluctuate according to available energy (i.e., by latitude) and water; and large river networks are clearly visible as a result of accumulation of upstream river flow. These results are outputs from a model run with default process parametrizations and parameter values, and improvements to either or both will likely improve local model accuracy. However, the visible large-scale patterns appear hydrologically sensible and give us confidence that this initial model configuration is a solid basis for further model improvement and development. The modular nature of our workflow enables improvements to any single part of it without needing to change any other parts of the model configuration code, which contributes to increased efficiency in model improvement and use.

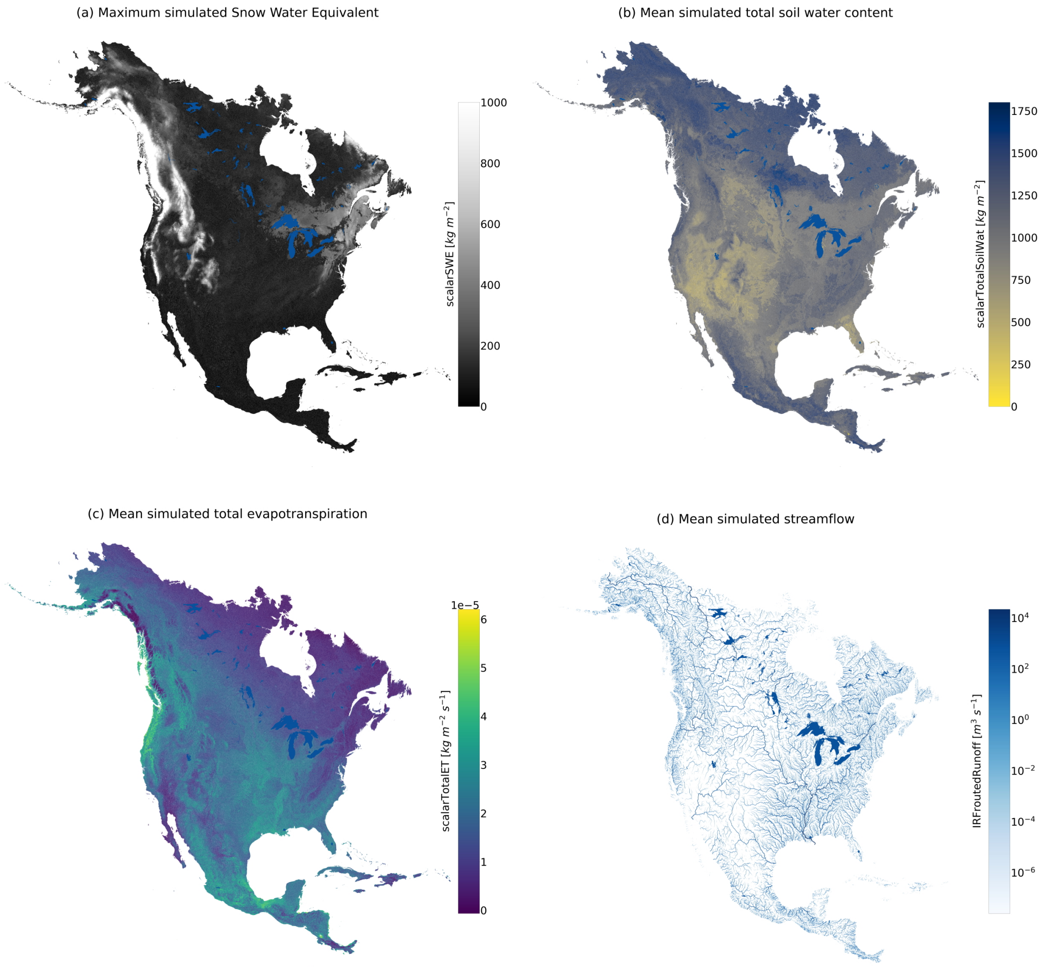

Figure 5. Overview of large-domain multi-decadal simulations. SUMMA does not perform any computations for GRUs that are classified as being mostly open water, and mizuRoute was run without using its option to simulate routing through lakes and reservoirs. Lake delineations of lakes $> 1{,}000$ km^2 are obtained from the HydroLAKES data set (Messager et al., 2016) and used to mask open-water GRUs in this figure. Model setup uses default parameter values, and results are for illustrative purposes only. (a) Maximum simulated Snow Water Equivalent per GRU is capped at 1,000 [kg m^{-2}] for visualization purposes. (b) Mean simulated total soil water content, which includes both liquid and solid water in the soil profile. (c) Mean simulated evapotranspiration, defined as the sum of evaporation from the soil profile and the canopy, and transpiration by vegetation. (d) Mean streamflow as determined by mizuRoute's Impulse Routing Function approach.

3.3 Local model configuration

The modeling domain in the global and continental test cases is discretized into sub-basins (Grouped Response Units, GRUs, in SUMMA terminology) of roughly equal area. SUMMA uses a flexible spatial discretization approach that allows GRUs to be sub-divided in as many Hydrologic Response Units (HRUs) as the modeler thinks practical and relevant. These HRUs can be used, for example, to represent different elevation zones, differences in soil or land use, differences in topography, or a combination of several of these elements (see Section A1 for a more detailed explanation). As a more localized test case, we created a subset of the MERIT basins data set (Lin et al., 2019) that covers the Bow River from the Continental Divide to the town of Banff, Alberta, Canada. We then sub-divided each MERIT sub-basin (i.e. each GRU) into multiple HRUs based on 500m elevation increments (Figure 6a), created a new control file for this new domain, and re-ran the workflow code. No changes were necessary to any of the workflow scripts because the scripts obtain all the required information from the updated control file and the code is generalized to handle both the large-domain case, where GRUs are not sub-divided into HRUs, and this local case, where HRUs are used. Note that this local test case could be for any basin on the planet.

This third test case uses hourly forcing data from 2008-01-01 to 2013-12-31 (again run at a 15-minute sub-step resolution). Temperature lapse rates are applied to the forcing data for each individual HRU, meaning that the hydrometeorological conditions are somewhat different in each HRU despite the forcing grid cells being relatively large compared to the delineated catchments (see Figure A7). Figure 6b shows that simulated Snow Water Equivalent (SWE) varies per HRU and accumulated streamflow varies per stream segment. These figures provide a rudimentary test of the generated model setup for a location for which we have clear expectations about how the simulations should look (see also a cautionary note on the use of global data products in Appendix B). As may be expected, more snow accumulates at higher elevations, whereas the valley bottoms have a lower snowpack due to warmer air temperatures but larger flows due to their larger accumulated upstream area. As with the global and continental simulations, this local test case is fully reproducible and all model configuration decisions are stored as part of the workflow. This local test case also shows that different model configurations (in terms of spatial discretization in GRUs and HRUs) can be generated by the same model-specific workflow code.

4 Discussion

4.1 To what extent does our workflow fulfill reproducibility requirements?

Best practices for open and reusable computational science can be briefly summarized as follows (e.g., Gil et al., 2016; Hutton et al., 2016; Stodden & Miguez, 2013): data must be available and accessible, code and methods must be available and accessible, active development on issues with data, code, and methods must be possible, and licensing of data and code should be as permissive as possible. These requirements are formalized in the FAIR principles (Findable, Accessible, Interoperable, Reusable; Wilkinson et al., 2016) but by themselves are not enough to guarantee reproducibility of computational science (e.g., Añel, 2017; Bast, 2019; Hut et al., 2017). To be fully reproducible, details about hardware, software versions, and data versions also need to be recorded and shared (e.g., Choi et al., 2021; Chuah et al., 2020; Essawy et al., 2020). Such practices require a certain time investments but the benefits are clear: the resulting science is more transparent, can be more easily reproduced, and follow-up work will be more efficient because less time is spent on mundane tasks such as data preparation.

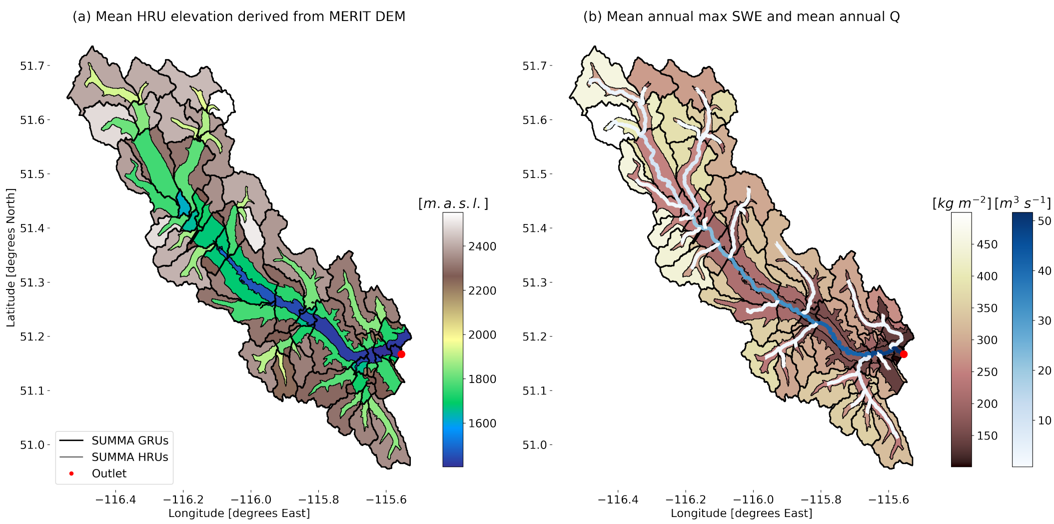

Figure 6. Overview of local simulations. Model setup uses default parameter values, and results are for illustrative purposes only. (a) Mean HRU elevation as derived from MERIT Hydro DEM. (b) Mean of maximum SWE per water year shown for each HRU, and mean annual streamflow shown for each river segment. Only data from complete water years is included.

Sandve et al. (2013) outline ten rules for reproducible computational science in the field of Computational Biology, and these are also applicable to Earth System modeling. Our workflow follows nine of these guiding principles:

(1) Our workflow stores copies of the scripts that generate data together with the data itself, which allows a researcher to track how a given result was produced;
(2) Our workflow contains no manual data manipulation: all changes to the data are done in scripts and can be traced;
(3) An exact version of all software used is tracked, partly as installable Python environments and partly on the workflow repository for command line utilities;
(4) All scripts are version controlled through Git;
(5) Our workflow is modular and stores intermediate results in individual folders to aid in debugging of setups and to allow easy diversion from our workflow;
(6) All data that may support analysis and figures are systematically stored in a logical folder structure;
(7) Our chosen model structure is flexible in prescribing outputs, removing a need to modify the model source code to display specific results;
(8) Our visualization code keeps a precise record of which results file contains the data that underpin a given figure and thus a record exists of which data support a given textual statement about the analysis;
(9) The workflow code is publicly accessible.

Their tenth principle, keeping accurate note of the seeds that underpin any element of randomness in the analysis, does not apply here. Sandve et al. (2013) also recommend sharing access to simulation results. This can be done through repositories such as HydroShare or Pangaea but may be infeasible in the case of large-domain Earth System modeling. For example, storing all input and output data of our continental test case takes approximately 13 TB.

Internal tests on different hardware and by different researchers indicate that our workflow effectively implemented these principles for open and reproducible science in practice: the workflow can be used to generate identical model inputs and outputs by specifying exact library, package, and model versions. Some caveats apply, however. Although it is possible to trace model source code versions through Git commit IDs, such IDs can obviously not account for local code modifications that are not tracked through Git. Good "computational lab hygiene" is needed to ensure consistency between what is reported to have been done and what has in fact been done. Further, not all data sets that underlie our model setups have Digital Object Identifiers assigned to specific versions of the dataset. Given the size of the data sets involved, sharing the data itself is infeasible, and some care must be taken to precisely track when data were downloaded as a means of making the use of data without DOIs traceable. Last, reproducibility is ensured through specifying exact versions of packages and libraries but many of these packages and libraries are undergoing rapid development and new versions are released frequently. There is a potential issue for reproducibility if older software versions for one reason or another are no longer available (though for fully open-source software this should theoretically not happen). New versions of specific software may however become incorporated into a new version of a workflow if they provide some needed functionality. To ensure backward compatibility, such new workflow versions must therefore also be assigned a new DOI so that any specific workflow version can be tracked and re-used when needed.

4.2 Towards community modeling

4.2.1 Short-term benefits of using workflows

This paper introduces a modular model configuration workflow that separates model-agnostic and model-specific configuration steps. The two main benefits of approaching environmental modeling from this angle are clear: configuring multiple modeling experiments becomes much more efficient, and results are reproducible, because all model configuration decisions can be traced. These benefits address two problems that currently affect Earth System modeling. First, creating a typical model configuration is both difficult and time consuming, and it is possible that model configuration tasks do not receive the attention they deserve. Code may not be checked as thoroughly as may be necessary because bugs may not be readily apparent, and any time spent on model configuration is consequently not spent on writing journal articles or meeting report deadlines. Configuring models can be more efficient if model configuration code is freely and openly shared. This enables time that is currently spent on creating model configurations to instead be spent on in-depth analysis, improving the model representation of real-world processes, and fixing any bugs that may be found in the configuration code or the model source code. If bugs are found, tracing the experiments that are affected by these bugs is possible, and it will be clear which studies need to be corrected. Openly shared model configuration code therefore has the potential to increase the robustness of model simulations and accelerate advances in modeling capabilities. Second, by publishing workflow code alongside a manuscript, the provenance of scientific results remains traceable (see e.g. Hutton et al., 2016; Melsen et al., 2017). This can increase confidence in model results. It also enables more effective follow-up studies because all decisions that underpin the original study can be found in the public domain.

4.2.2 Long-term vision for community workflows

We see workflows such as the one presented in this paper as the first step towards a community-wide modeling framework. Figure 7 illustrates an example of such a framework using the workflow code presented in this paper as examples of each framework layer (see also Miles, 2014). In addition to a division between model-specific and model-agnostic tasks, we envision a framework that distinguishes between data-specific and data-agnostic

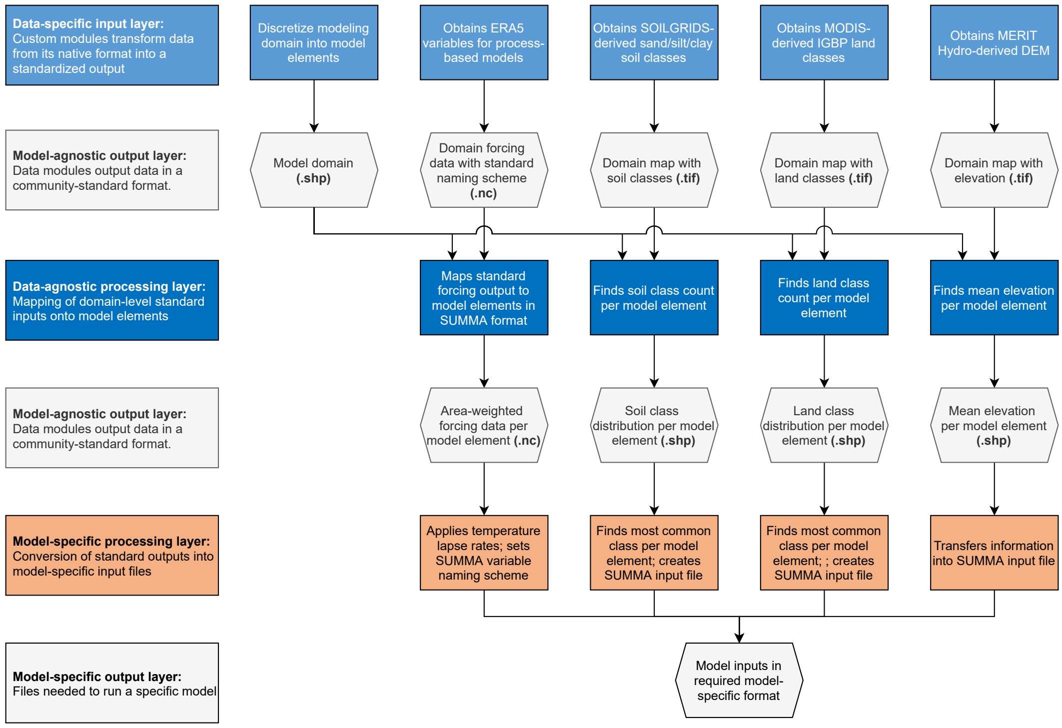

Figure 7. Schematic overview of a generalized community modeling framework, populated with examples from our SUMMA setup configuration workflow. Key to this modular approach is community-wide agreement on the formats used in each model-agnostic standardization layer. Such standards enable a modular approach to model-configuration, where existing modules can be seamlessly replaced, as long as they are designed to read and output data in the agreed-upon formats.

preprocessing steps. Processing layers would be separated by standardization layers that prescribe the output format for the preceding processing layer and consequently the input format for the following processing layer (see also Miles, 2014, where the use of common file types is discussed as a recommended approach between the model-agnostic and model-specific parts of workflows). Community-wide agreement on the formats used in standardization layers will promote efficient interoperability of different data-specific processing modules, possibly as part of broader work on international hydrologic standards (e.g., HY_Features, Blodgett & Dornblut, 2018). Using our workflow as an example, we have created data-specific processing modules for ERA5 meteorological data, SOILGRIDS-derived soil classes, MODIS-derived land classes, and a MERIT Hydro-derived DEM. These modules generate data in standardized formats (in this case, netCDF4 forcing data and GeoTIFF spatial maps) that in turn feed into the data-agnostic remapping layer. This layer generates further model-agnostic data in netCDF4 and Shapefile formats that are then transformed into SUMMA's inputs through a model-specific processing layer.

Our currently defined model-agnostic tasks are of course still implicitly SUMMA-centric (i.e., we have completed those tasks because they generate the data that SUMMA requires), though in principle the outputs can immediately be used by other models. The modular nature of our workflow makes adding new datasets and processing steps as straightforward as writing new data-specific and data-agnostic routines and inserting them in a further unchanged workflow. Changing to a different model requires writing a new model-specific interface layer, but existing data processing scripts can remain untouched (again, assuming that the new model has data needs that can be met by already existing data processing scripts). This means that the workflow can be tailored to a specific model or experiment in a fraction of the time needed to create the model configuration from scratch. The modified workflow can then be published alongside the new modeling results to keep those results traceable.

It is of course possible that our attempt to separate model tasks into model-agnostic and model-specific parts is not equally applicable to different models that are currently in use. In such cases, we hope that providing a tangible example of how model configuration code can be organized and shared in a structured way will nevertheless inspire others to create their own workflows. Modifying our workflow or adapting it for different purposes in such ways is the second step we anticipate as needed to move towards a community modeling paradigm. By creating new or modifying existing workflows for new experiments and models, the required structure of a generalized model setup workflow may become apparent. As a third and final step, this generalized workflow can be formalized into a community-driven modeling framework that enhances efficiency and transparency in Earth System modeling.

To initiate the process of creating a community-driven modeling framework, our workflow is available as open-source code: `https://github.com/CH-Earth/CWARHM` (last access: 2022-07-27). We have chosen a permissive license (GNU GPLv3) that allows others to freely use and modify our code under the conditions that the modified code base is published under the same license, with attribution of its source and a list of changes. We envision a gradual process in which our repository is modified by others (either piecemeal or by incorporating our entire code base in a new repository as, for example, a Git submodule), increasingly more data-specific and models-specific processing capabilities are made public, and appropriate formats for standard file formats become apparent. Deciding if and how to integrate these different elements into a single modeling framework is a decision the community will need to make in due course.

4.2.3 Where do workflows stand in the existing reproducibility landscape in hydrologic modeling?

We approach the workflow problem from a catchment modeling perspective within the wider Earth System modeling community (see the definitions of different communities in Archfield et al., 2015). Calls for more efficient, transparent, and shareable model configuration approaches are not new in the catchment modeling community (see e.g. Blair et al., 2019; Famiglietti et al., 2011; Hutton et al., 2016; Tarboton et al., 2009; Weiler & Beven, 2015) and considerable progress along these lines has been made. For example, Sen Gupta et al. (2015) standardize model inputs and outputs to efficiently couple a snow accumulation and melt routine with an existing open source modeling framework; Ecohydrolib (Miles & Band, 2015; Miles, 2014; Miles et al., 2022) is a Python API that automatically preprocesses ecohydrologic parameter fields and forms the basis of a model configuration workflow for the RHESSys model; Bandaragoda et al. (2019) develop a general interface for building and coupling multiple models, using the Landlab toolkit (Hobley et al., 2017; Barnhart et al., 2020); Gan et al. (2020) integrate a web-based hydrologic model service with a data sharing system to promote reproducible workflows; HydroDS (Gichamo et al., 2020; Dash & Tarboton, 2022) is a web-based service that can be used to prepare input data for modeling; Bennett et al. (2018, 2020) create a tool to estimate hourly forcing input for physics-based models from commonly available daily data; Bavay et al. (2022) describe a tool that can be used to effectively create a Graphical User Interface for a given model; Essawy et al. (2016) provide an example of how containerization (storing a full computational environment into a software container) enhances reproducibility; and Kurtzer et al. (2017, 2021) develop a means of saving and transferring software and computing environments on and between High Performance Computing clusters. Put together, most if not all elements for fully reproducible, easy-to-use, computational hydrology already exist. So far, however, uptake of these tools is regrettably not widespread.

We speculate that uptake of existing tools is somewhat low for multiple reasons. First, these tools are typically provided as self-contained packages where some form of interface exists between the user and the source code. Such packages tend to be easy to use for their intended purpose but take time to understand and do not necessarily provide much flexibility to deviate from their intended purpose. Layering additional functions on top of an existing package or modifying a package's source code is certainly possible, but can be outside the comfort zone of many users. Second, several model-configuration tools are provided as web-based services. This can be appealing because, for example, data can be pre-downloaded to speed up model configuration and model simulations can be easily shared. The advantage of such approaches is that they can be combined with some form of server-side data transformations (e.g., subsetting or averaging), which minimizes data transfers. Storing the inputs for and outputs of large-domain simulations can, however, be cumbersome, and keeping pre-downloaded data up-to-date and sufficient for all user needs takes sustained, long-term effort. A further complication is that it is regrettably common that such web-based services require some form of manual interaction with the webpage, limiting opportunities to automate data acquisition tasks. Third, the lack of community agreement on standard data formats means that developers of new tools typically decide to have their tool output data in a format relevant to their own application, which may not be a format that is widely used by others. It is cumbersome for developers to have their tools ingest multiple different data formats and such functionality is therefore somewhat rare. Community-wide agreement on a set of standard data formats, such as proposed in Figure 7, will make it easier for developers to know which data formats their tools must be able to ingest and produce to guarantee seamless interaction with other existing tools.

In short, some of the existing tools may be overdesigned or unsuitable for where the majority of the community currently stands. Such tools are typically designed by a small group of people, using a proof of concept or test case that is directly applicable to the developers' own work. Developers can make educated guesses about how their tool can be made more general beyond their proof of concept, as we had to do here. Actually extending these proof of concepts typically relies on the original developers having both the motivation and opportunity to implement functionality for others (e.g., incorporating new data sets or including model-specific layers for other models) or on new developers being willing to first understand the existing package or web-service and then modifying it.

Our approach to provide a tangible example of how to structure model configuration tasks is different. First, our use of scripts that allow a user to immediately access the workflow code is likely much more similar to how many models are currently configured than if we had wrapped our workflow code in some form of user interface (such as a Python package, R library, or web interface). This lowers the barrier to trying our approach. Second, our use of standardization layers that require intermediate files to be in commonly used data formats (GeoTIFF, netCDF, ESRI shapefiles) makes it easy to adapt small parts of our workflow without needing to change any upstream or downstream configuration tasks. Third, there are clear and immediate benefits of adopting a workflow approach of the type proposed in this paper that are unrelated to how widely (or not) this approach is adopted: creating new configurations for the models used in such workflows will be more efficient and the resulting science will stand on a firmer foundation than closed-source results. Should our approach become more widely adopted, then the path to a community modeling framework builds itself: as more examples of model configuration workflows become available, our preliminary sketch of a community modeling framework in Figure 7 can be refined or redrawn. The best approach to design, build, and maintain such a community framework can be decided in due course, and appropriate funding may be sought when needed. Advancing the paradigm of community modeling requires active participation of the community. By providing an example of a community modeling workflow, we hope to encourage uptake, modification and adaptation of such community approaches.

4.3 Future work

We outlined three steps to move toward a culture of community modeling in the Earth sciences in Section 1:

1. For a given model, model configuration code should be publicly available and divided into model-agnostic and model-specific steps;
2. The configuration workflows of multiple different models, ideally using different data sets, should be integrated into a proof-of-concept of a generalized model configuration workflow;
3. A community-wide collaborative effort should refine the proof-of-concept into a flexible model configuration framework.

This manuscript provides an example of the first step in this list, by showing how configuration code for a single model can be implemented in a more general framework. Ongoing work focuses on the second step, by integrating multiple different models such as MESH (Pietroniro et al., 2007) and HYPE (Lindström et al., 2010; Arheimer et al., 2020) into our workflow by adding the necessary processing code for these models. This work is nearing completion, and both models have successfully been able to re-use the model-agnostic part of the code base described in this paper, suggesting that a 'bottom-up' kind of approach to community modeling is feasible.

New processing code naturally involves writing new model-specific routines that convert existing pre-processed data into the specific formats each new model needs. Inclusion of additional models also necessitates certain new model-agnostic processing routines. For example, whereas SUMMA works on the assumption that a single computational element has a single (possibly dominant) land cover type (but allows spatially flexible configurations so that each different land cover type can be assigned its own computational element), MESH lets the user specify a histogram of land cover types within each grid cell. Our current implementation of model-agnostic land cover remapping therefore still follows the implicit assumption that the required processing output is a single land cover class per model element. A new routine is needed that returns the histogram of land classes per model element that MESH requires. Examples such as these show that a modular approach to a generalized community modeling framework as described in Section 4.2.2 and Figure 7, where new processing modules can be inserted without requiring changes to existing upstream and downstream routines, is a likely path forward on the road to community modeling.

5 Conclusions

This paper describes a code base that provides a general and extensible solution to configure hydrologic models. Specifically, the paper provides a tool that can be used to create reproducible configurations of the Structure for Unifying Multiple Modeling Alternatives (SUMMA, a process-based hydrologic model) and mizuRoute (a vector-based routing model). We consider this the implementation of a single model in a general framework that separates model-agnostic and model-specific configuration tasks. Such a separation of tasks makes inclusion of new models in this framework relatively straightforward because most of the data pre-processing code can remain unchanged and only model-specific code for the new model needs to be added.

The critical component of this framework are standardization layers, which prescribe the details of the file formats that must come out of the preceding processing layer and form the input of the following processing layer. By standardizing inputs and outputs, the code that forms the processing layers only needs to concern itself with these prescribed formats. Changing specific processing modules to, for example, pre-process a different data set, perform a different way of mapping data onto model elements, or prepare input files for a different model, can therefore happen in isolation from the remainder of the workflow as long as the new processing code accepts and returns data in the prescribed formats. We show examples of this approach with global and multi-decadal continental SUMMA and mizuRoute simulations, and with a local SUMMA configuration that uses a more complex spatial discretization than the global and continental simulations use.

Future work will involve adding model-specific code for multiple additional models and any needed data-specific preprocessing modules. We have termed this initiative "Community Workflows to Advance Reproducibility in Hydrologic Modeling" (CWARHM; "swarm") and we encourage others to be part of this model-agnostic workflow initiative. The configuration code for the SUMMA and mizuRoute setup shown in this manuscript is available on GitHub: `https://github.com/CH-Earth/CWARHM` (last access: 2022-07-27).

Appendix A Workflow description

This section describes in detail our example of a model setup workflow that follows the design principles outlined in Section 2. The workflow code, model code, software requirements, and data are fully open-source to follow the FAIR principles. The workflow is written in Python and Bash, using input data with global coverage, a spatially distributed, physics-based hydrologic modeling framework designed to isolate individual modeling decisions (Clark et al., 2015a, 2015b; Clark, Zolfaghari, et al., 2021), and a network routing model (Mizukami et al., 2016, 2021) that connects the individual hydrologic model elements through a river network. This example workflow can be used to generate a basic SUMMA and mizuRoute setup anywhere on the globe and is designed such that the model-agnostic parts of the code can easily feed into other modeling chains.

Part of the code in this repository is adapted from or inspired by work performed at the National Centre for Atmospheric Research and the University of Washington.

A1 Models

This section provides a brief overview of SUMMA (Clark et al., 2015a, 2015b; Clark, Zolfaghari, et al., 2021) and mizuRoute (Mizukami et al., 2016, 2021) to the extent relevant to understand our workflow. We refer the reader to the original papers that describe each model for further details. We selected both models for their flexible nature, computational capacity to model very large domains, and availability of local expertise. Both models are written in Fortran, and their source code needs to be compiled before the models can be used.

A11 Structure for Unifying Multiple Modeling Alternatives (SUMMA)

SUMMA is a process-based modeling framework designed to isolate specific modeling decisions and evaluate competing alternatives for each decision, with the ability to do so across multiple spatial and temporal configurations. SUMMA solves a general set of mass and energy conservation equations (Clark et al., 2015a; Clark, Zolfaghari, et al., 2021) and includes multiple alternative flux parametrizations (Clark et al., 2015b). It separates the equations that describe the model physics from the numerical methods used to solve these equations, allowing the use of state-of-the-art numerical solving techniques

(Clark, Zolfaghari, et al., 2021). SUMMA is available as Free and Open Source Software (FOSS) and under active development (see `https://www.github.com/CH-Earth/summa`).

SUMMA organizes model elements into Grouped Response Units (GRUs) that can each be further subdivided into multiple Hydrologic Response Units (HRUs). This enables flexible spatial discretization of modeling domains. For example, point-scale studies are possible by defining the domain as a single GRU that contains exactly one HRU (GRU area can be an arbitrary value because all fluxes and states are calculated per unit area; see e.g. Clark et al., 2015b). It is equally possible to mimic grid-based model setups such as commonly used in land-surface modeling schemes by defining each GRU to be equivalent to a grid cell and optionally using the HRUs to account for sub-grid variability (e.g. mimicking the tiled grid approach of traditional VIC and MESH setups; Liang et al., 1994; Pietroniro et al., 2007). Finally, GRUs can represent the (sub-)catchments of a given river system with HRUs being areas of similar hydrologic behavior within each GRU. Such model configurations can use GRUs and HRUs of irregular shape, which has several advantages over grid-based setups (see e.g. Gharari et al., 2020). Most importantly, such spatial configurations can accurately follow the actual topography of the modeling domain, and this makes model results easier to visualize and interpret. SUMMA is configured with irregularly shaped computational elements in the test cases presented in this paper.

A12 mizuRoute

mizuRoute is a vector-based river routing model specifically designed for large-domain applications such as modeling of hydrologic processes across a continental domain. It organizes the routing domain into Hydrologic Response Units (HRUs; i.e., catchments) and stream segments that meander through the HRUs and provide connections between them (Mizukami et al., 2016, 2021). It can process inputs from hydrologic models with both grid- and vector-based setups and provides different options for channel routing: Kinematic Wave Tracking (KWT) and Impulse Response Function (IRF). For a given stream segment, the IRF method constructs a set of unique Unit Hydrographs (UH) for each upstream segment which is used to route runoff from each upstream reach independently. In other words, the routed runoff in a given stream segment is a simple sum of the UH runoff generated in all upstream segments. The KWT method instead tracks channel runoff as kinematic waves moving through the stream network with their own celerity. mizuRoute is available as FOSS and under active development (see `https://github.com/ESCOMP/mizuRoute`), with a particular focus on improving its representation of lakes and reservoirs (Gharari et al., 2022; Vanderkelen et al., 2022).

A13 Note on definitions

SUMMA distinguishes between Grouped Response Units (GRUs) and Hydrological Response Units (HRUs). SUMMA's main modeling element is the GRU, which can be sub-divided into an arbitrary number of HRUs. SUMMA can handle GRUs and HRUs of any shape (e.g., points, grid cells, catchments) and these terms therefore refer to model elements of arbitrary shape and size. In this workflow, we use mizuRoute to route runoff between SUMMA's GRUs. Potentially confusingly, mizuRoute refers to all routing basins as HRUs only and does not use the term GRU. As a result, what SUMMA calls GRUs are referred to as HRUs by mizuRoute. For consistency with both sets of documentation, we use their own terminology for model elements where possible. Figure A1 shows a graphical example of the differences in terminology.

A2 Workflow description

This section briefly describes each step shown in the workflow diagram (Figure 1 in the main manuscript, with further technical details in Figures A2, A3 and A4). Fig-

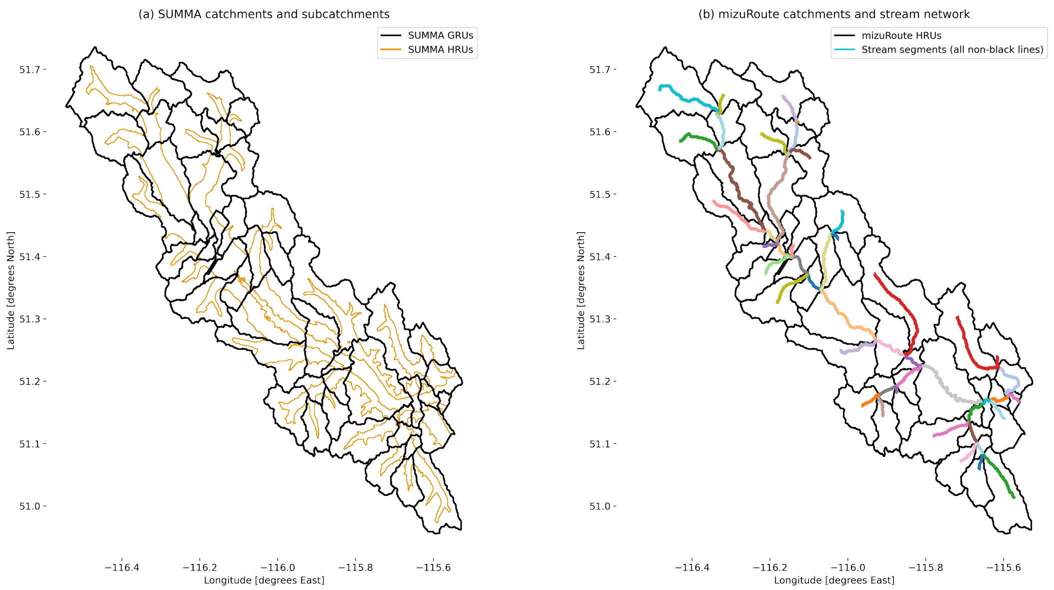

Figure A1. Catchment of the Bow River at Banff (Alberta, Canada) discretized into (a) SUMMA and (b) mizuRoute model elements, showing associated terminology. SUMMA HRUs in (a) represent different elevation bands within each SUMMA GRU. A SUMMA GRU always contains at least one SUMMA HRU. There is no upper limit to the number of HRUs a single SUMMA GRU can be divided into. A single SUMMA HRU is never part of more than one SUMMA GRU. In our example, SUMMA GRUs are identical to mizuRoute HRUs. mizuRoute stream segments are shown in different colors to emphasize that in this case each mizuRoute HRU maps 1:1 onto a single stream segment; only a single color is shown in the legend for brevity, but all non-black lines are stream segments.

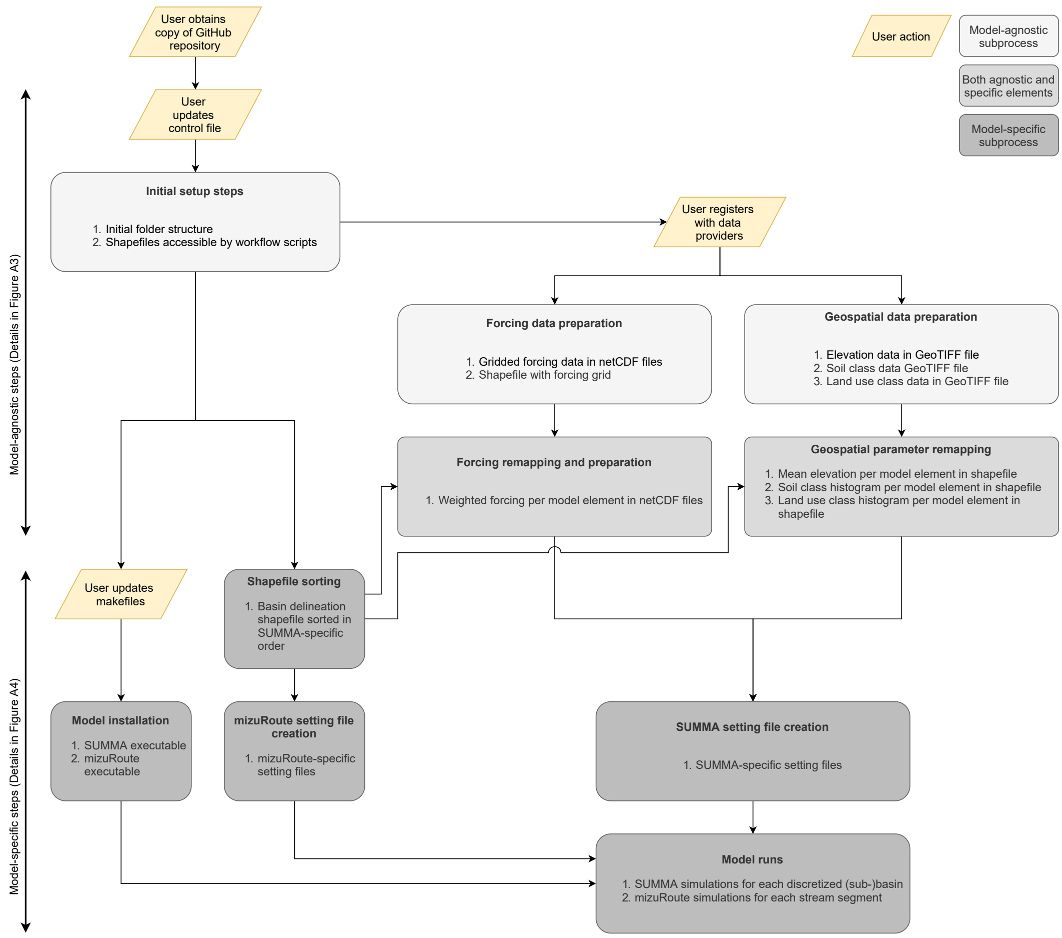

Figure A2. High-level overview of model configuration steps, using SUMMA (a process-based hydrologic model) and mizuRoute (a routing model) as example models. Configuration tasks are separated into model-agnostic and model-specific tasks (details in Figure A3 and Figure A4 respectively). Each rounded box specifies the outcomes of that configuration task as a numbered list.

ures are generated using the test case configured for the Bow River catchment located in Alberta, Canada (see Figure A1 for an overview of this domain). This test case covers a geographically small area (approximately 2200 km^2) and uses a more complex model setup (SUMMA GRUs subdivided into multiple HRUs) than the continental test case (where SUMMA GRUs contain exactly one HRU each), making it the best choice to visualize model setup procedures. Italicized phrases in this section indicate folders, scripts, or variables as found in the GitHub repository. To start, a user would download or clone the complete GitHub repository. The following sections provide more detail about the scripts found within the GitHub repository. Although our workflow requires only limited user interaction to generate a model configuration for a new domain, we do make certain assumptions about this model configuration which users should be aware of. These assumptions are specified in each subsection.

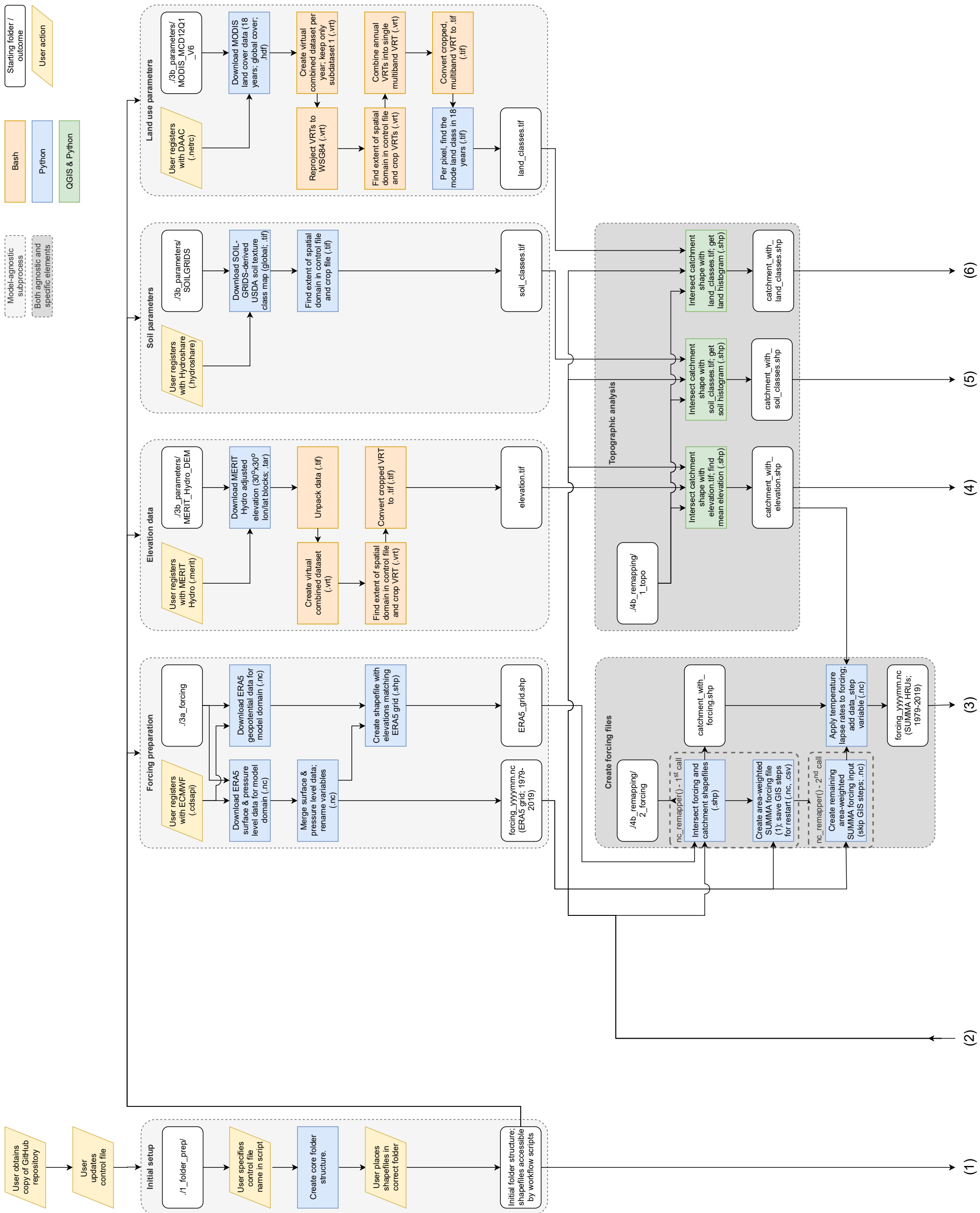

Figure A3. Model-agnostic configuration steps. Each rectangular block corresponds to a specific model setup task and is accompanied by a specific script with Python or Bash code, stored in a GitHub repository. Rounded rectangles indicate starting points of specific sub-tasks (mainly showing which folder in the repository contains certain parts of the workflow) and the outcomes of each sub-process. Parallelograms indicate actions the user must perform. Numbers show connections with the model-specific configuration tasks in Figure A4.

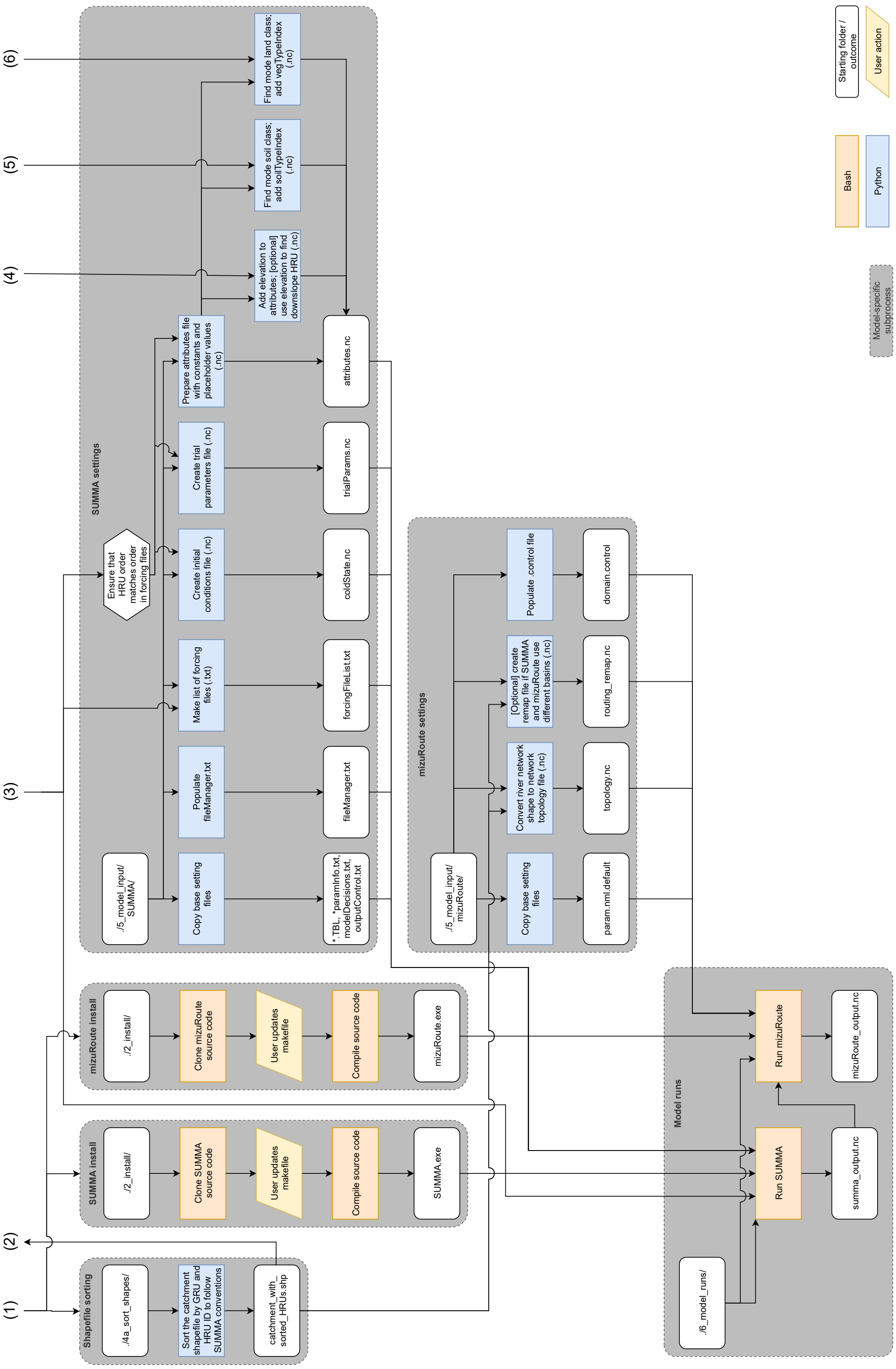

Figure A4. Model-specific configuration steps. Each rectangular block corresponds to a specific model setup task and is accompanied by a specific script with Python or Bash code, stored in a GitHub repository. Rounded rectangles indicate starting points of specific sub-tasks (mainly showing which folder in the repository contains certain parts of the workflow) and the outcomes of each sub-process. Parallelograms indicate actions the user must perform. The hexagon indicates an aspect of SUMMA's input requirements (i.e., not an action or script) and is shown to clarify why creating the forcing files is on the critical path towards creating the other necessary model configuration files. Numbers show connections with the model-agnostic configuration tasks in Figure A3.

A21 Workflow setup and folder structure

This section describes the steps "User updates control file" and the steps contained in the box "Initial setup" (Figure A3).

A21.1 Control files Control files are the main way for a user to interact with the workflow. They contain high-level information such as file paths, file names, variable names, and specification of the spatial and temporal extent of the modeling domain (see also Sen Gupta et al., 2015) A new control file needs to be created by the user for each new domain. As an example, the control file for the *Bow_at_Banff* test case is included as part of the Github repository, in the folder *./CWARHM/0_control_files.* The READMEs of each sub-folder on the GitHub repository contain a list of the settings in the control file on which the scripts in that sub-folder rely.

A21.2 Folder preparation The workflow separates generated data from the code used to generate the data. The script in the folder *./CWARHM/1_folder_prep* generates a basic data folder structure in a location of the user's choosing (see Figure 3b). This basic folder structure generates a main data folder with a subdirectory for the current domain. In this domain folder, it further generates a dedicated folder where the user can place their shapefiles that delineate the SUMMA catchments (hydrologic model GRUs and HRUs), mizuRoute catchments (routing model HRUs), and mizuRoute river network. This is the only script in the workflow that needs to be manually modified if a setup for a new domain is generated. A user will need to modify the variable *sourceFile* so that it points to the control file for the current domain. In our example, this is set to *control_Bow_at_Banff.txt.* The script then copies the contents of this control file into a new file called *control_active.txt*, which is the file every other workflow script will search for. The variable *sourceFile* needs to be updated when a control file for a new domain is used. Note that the contents of the file *control_active.txt* determine which folders and files the other workflow scripts operate on.

A21.3 Domain shapefiles With a basic folder structure in place, the user can now move their prepared shapefiles into the newly generated folders (assuming the control file uses 'default' values for these shapefile paths). Briefly, the shapefiles should contain: geometries that delineate the hydrologic model GRUs and HRUs, the routing model HRUs, and the routing model river network in a regular latitude/longitude projection (in other words, in the Coordinate Reference System defined by EPSG:4326; `https://epsg.io/4326` [last access, 2021-10-11]). Each shapefile needs to specify certain properties of the model domain, such as identifiers for each GRU, HRU, and stream segment; HRU area and centroid location, stream segment slope and length; and the stream segment ID into which a given HRU drains.

Detailed requirements for the shapefiles are provided in the README in *./CWARHM/1_folder_prep.* Example shapefiles for the *Bow_at_Banff* test case are part of the repository and can be found in the subfolders of *./CWARHM/0_example.*

A22 Model-agnostic workflow elements

This section provides details about the model-agnostic elements of the workflow (shown in light grey in Figure A3). For convenience, this section is organized to follow the four model-agnostic sub processes: pre-processing of forcing, elevation, soil, and land use data.

A22.1 Pre-processing of forcing data Our chosen forcing product is the ERA5 reanalysis data set (Copernicus Climate Change Service (C3S), 2017; Hersbach et al., 2020) provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 data are available as hourly data for the period 1979 to present minus 5 days, at a 31 km spatial grid that covers the Earth's surface or at a re-gridded 0.25° x 0.25° latitude/longitude resolution. ERA5 data preparation includes two-way interactions between atmosphere, land surface and ocean surface components. The ERA5 model setup includes different atmospheric layers and ERA5 data are available at 137 different pressure levels (i.e., heights above the surface), as well as at the surface. The lowest atmospheric level is L137, at geopotential and geometric altitude 10 m (i.e., 10 m above the land surface). To limit the influence of ECMWF's land model on our required forcing variables (simulating the land surface response is SUMMA's role after all), we obtain air temperature, wind speed, and specific humidity at the lowest pressure level (Hersbach et al., 2017)instead of at the land surface. Precipitation, downward shortwave radiation, downward longwave radiation, and air pressure are unaffected by the land model coupling and can be downloaded at the surface level (Hersbach et al., 2018).

Surface and pressure level data are stored in two different data archives and are accessed in different ways. Download scripts for each separate archive are found in folder *./CWARHM/3a_forcing/1a_download_forcing.* These scripts access the C3S Climate Data Store (CDS) using the user's credentials (instructions on how to obtain and store credentials can be found in the README in the download folder) and download the necessary data in monthly blocks of hourly data at a regular 0.25 x 0.25° latitude/longitude resolution. The spatial and temporal extents of the domain are taken from the control file. As per the ERA5 documentation, ERA5 data should be seen as point data, even though standard visualization approaches typically show this kind of data as an interpolated grid. In our example workflow, we make the simple assumption that each ERA5 point contains forcing data that are representative for the grid of size 0.25° x 0.25° of which the grid point is the centroid. The workflow code automatically finds which ERA5 grid points to download based on the catchment bounding box specified in the control file (Figure A5). Once downloaded, the code in *./CWARHM/3a_forcing/2_merge_forcing* can be used to merge the surface and pressure level downloads into a single netCDF file, which is used for further processing. During this merging process, the ERA5 variable names are also changed to more descriptive ones.

Gridded forcing data does not map directly onto irregular model elements such as HRUs. Code in *./CWARHM/3a_forcing/3_create_shapefile* generates a shapefile for the forcing data that outlines the forcing grid (dotted red lines in Figure A5), which is later used to find the relative contribution of each forcing grid cell to the forcing of each HRU. The elevation of each ERA5 grid point is added to this shapefile. Elevation data is later used to apply temperature lapse rates based on the difference in elevation of the ERA5 data and mean HRU elevations. As per the ERA5 documentation, the elevation of each ERA5 data point is found by dividing the geopotential [m^2 s^{-2}] of each point (downloaded through scripts in *./CWARHM/3a_forcing/1b_download_geopotential*) by the gravitational acceleration [m s^{-2}].

Key assumptions in this part of the workflow are (1) that the user has access to the Copernicus Data Store. Instructions on how to obtain access are given in the README in folder *./CWARHM/3a_forcing.* (2) We consider that using forcing data that are the result of interaction between the atmospheric and land surface model components is undesirable and hence somewhat limit this interaction by downloading certain variables at the lowest pressure level instead. (3) ERA5 data points are assumed to be representative of grids of size 0.25° x 0.25°. (4) Gravitational acceleration is assumed to be constant at $g = 9.80665$ [m s^{-2}] (Tiesinga et al., 2019), although in reality this value would vary depending on latitude and altitude. (5) ERA5 variable names are changed to more descriptive ones that are also the names SUMMA expects these variables to have.

A22.2 Pre-processing of geospatial parameter fields Three different types of geospatial data are required for our example model setup. A Digital Elevation Model (DEM) provides the elevation of each HRU and is both a SUMMA input and required to apply temperature lapse rates as a preprocessing step. Maps of soil classes and vegetation classes are needed to utilize parameter lookup tables. These tables specify values for multiple

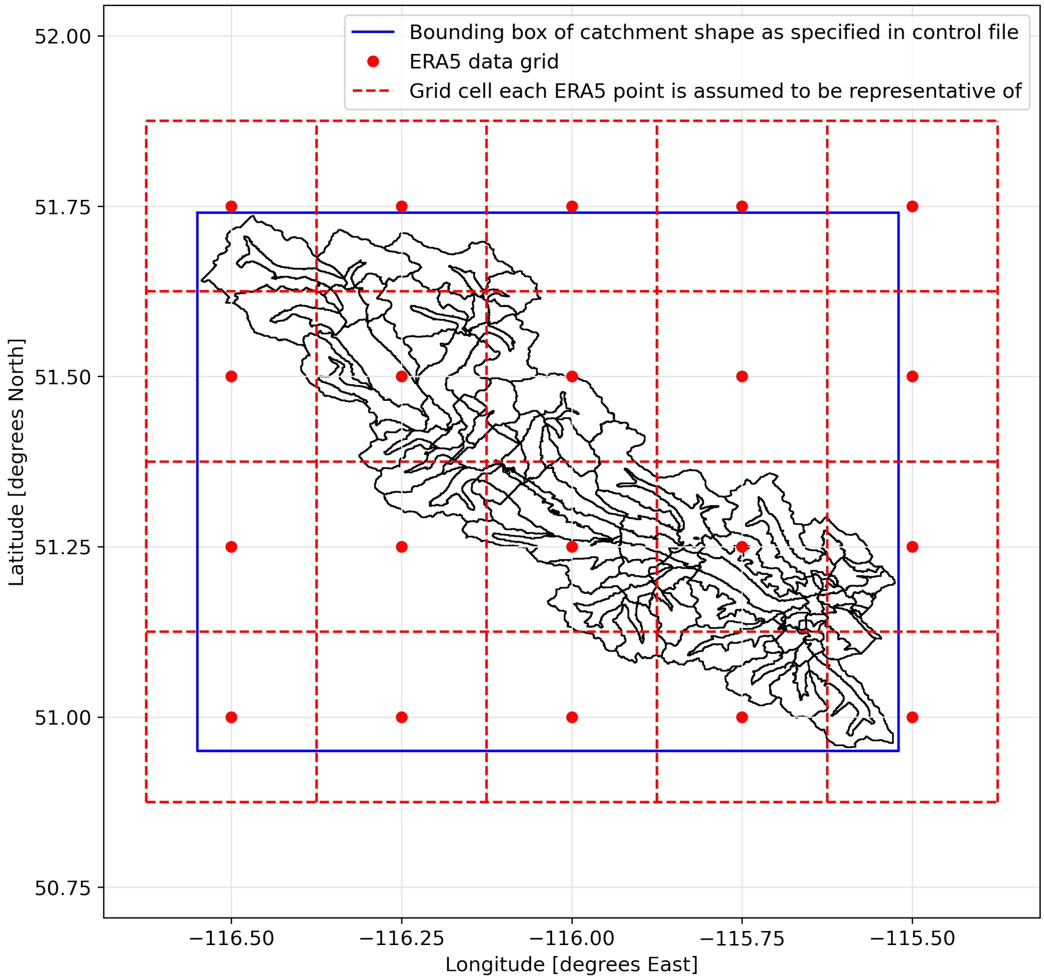

Figure A5. Overview of ERA5 data points, catchment and bounding box and how ERA5 data is assumed to overlap the catchment for the Bow at Banff test case.

parameters for a variety of soil and land classes. By knowing the soil or land class for a given HRU, SUMMA uses the predefined parameter values for those classes.

Digital Elevation Model We use the hydrologically adjusted elevations that are part of the MERIT Hydro dataset (Yamazaki et al., 2019) to determine HRU elevations. The MERIT Hydro hydrography maps cover the area between 90° North and 60° South at a spatial resolution of 3 arc-seconds. They are derived from the MERIT DEM (Yamazaki et al., 2017), which itself is the result of extensive error correction of the SRTM3 (Farr et al., 2007) and AW3D-30m (Tadono et al., 2016) DEMs. Scripts can be found in the subdirectories of *./CWARHM/3b_parameters/MERIT_Hydro_DEM.*

MERIT Hydro data are provided as compressed data packages that cover 30° x 30° areas. Based on the spatial extent of the domain, as given in the control file, the required 30° areas are downloaded in compressed format. Data are then uncompressed so that the individual GeoTIFF files are accessible. These files are first combined into a Virtual Dataset (VRT), from which the exact modeling domain is extracted into a new VRT. The VRT with the extracted subdomain is then converted into a single GeoTIFF file that contains the DEM for the modeling domain. A key assumption is that the user has access to the MERIT Hydro data. Instructions on how to obtain access are given in the README in folder *./CWARHM/3b_parameters/MERIT_Hydro_DEM.*

Vegetation classes We use MODIS MCD12Q1_V6 data (Friedl and Sulla-Menashe, 2019) to determine land cover classes at the HRU level. MODIS MCD12Q1 data are available for the years 2001 to 2018 at a 500 m resolution. The data set contains land cover classes for multiple different land cover classification schemes. Each data layer is the result of supervised classification of MODIS reflectance data (Friedl & Sulla-Menashe, 2019). Scripts can be found in the subdirectories of *./CWARHM/3b_parameters/MODIS_MCD12Q1_V6.*

MODIS MCD12Q1 data is provided as multiple Hierarchical Data Format (HDF) files that each cover a part of the planet's surface at a given time. The source data files are in a sinusoidal projection and of irregular shape which makes it difficult to extract a specific region. Therefore, the workflow downloads all available individual HDF files for each year (i.e., global coverage). The individual files for each data year are combined into one Virtual Dataset (VRT) per year for easier processing. Only the data layer of interest, the International Geosphere Biosphere Programme (IGBP) land cover classification, is included in the VRT. The VRT is reprojected from its original sinusoidal projection into a regular latitude/longitude grid (EPSG:4326) from which the modeling domain is extracted. The annual VRTs are then combined into a single multi-band VRT, which is then converted to a multi-band GeoTIFF file. The MODIS documentation advises against using the data of an individual year due to data uncertainty (Sulla-Menashe & Friedl, 2018). Therefore, the mode land class between 2001 and 2018 is identified as the most likely class for each pixel and stored as a new GeoTIFF file.

Key assumptions are (1) that the user has access to NASA's Earth Data website. Instructions on how to obtain access are given in the README in folder *./CWARHM/3b_parameters/MODIS_M* (2) Our example uses the IGBP land cover classification data, which is one of multiple options available.

Soil classes Our example uses a global map of soil texture classes (Knoben, 2021) derived from the SoilGrids 250m dataset (Hengl et al., 2017) to specify representative soil classes at the HRU level. The SoilGrids data are provided at a 250 m resolution and at seven standard depths (up to 2 m depth). Data are the result of a combination of approximately 150,000 observed soil profiles, 158 remote sensing-based soil covariates, and multiple machine learning methods. SoilGrids maps of sand, silt, and clay percentages were converted to a soil texture map for each depth using the soil texture class boundaries of Benham et al. (2009). For each 250 m map point, the mode soil class of the seven soil layers was selected as a representative value for the soil column as a whole, resulting in a single global map of soil texture classes. The pre-processing code needed to create this map (data download, data merge into a coherent map, conversion from percentages to soil texture, finding the mode of each soil column) is accessible as part of the data resource (Knoben, 2021). Scripts can be found in the subdirectories of *./CWARHM/3b_parameters/SOILGRIDS*.

The global soil texture class map is provided at the same horizontal resolution as the underlying SoilGrids data. The workflow first downloads a map with global coverage. The spatial extent of the modeling domain is extracted based on the bounding box specified in the control file and stored as a new GeoTIFF file.

Key assumptions are (1) that the user has access to Hydroshare. Instructions on how to obtain access are given in the README in folder *./CWARHM/3b_parameters/SOILGRIDS and* (2) the global soil map used assumes that mode soil class in each soil column can be considered as the representative soil class for the entire soil column and that the soil properties (such as saturated conductivity and pore volume) for the mode class are representative of the properties of the column. This approach ignores the existence of layered soil profiles and the differences in water movement this can cause (e.g., Vanderborght et al., 2005). This also assumes that the most common class contains the layers that are most hydrologically active and relevant for modeling purposes.

A23 Mapping of data to model elements

This section provides details about the mapping of preprocessed forcing data onto model elements (shown in the intermediate grey shade in Figure A3). This process cannot be called truly model-agnostic because whether it is needed depends on the model in question: some models are able to ingest the pre-processed data directly.

A23.1 Geospatial parameter fields In our example, geospatial data in the form of GeoTIFF files containing the DEM, land classes, and vegetation classes cannot be ingested by the hydrologic model directly. The data must be mapped onto the model elements (HRUs) as delineated in the catchment's shapefile. These procedures use the open-source QGIS project (QGIS Development Team, 2021) to provide the necessary Python functions (*./CWARHM/4b_remapping/1_topo*). Key assumptions are (1) that MERIT Hydrologically Adjusted Elevation data need to be aggregated into mean elevation values per model element, whereas (2) soil and vegetation classes need to be aggregated into histograms that summarize the distribution of values per model element.

A23.2 Forcing data Figure A6 shows the original gridded air temperature values on an arbitrary day and the HRU-averaged values on that same day that are obtained by mapping the gridded forcing data onto the model elements. For each model element, the relative overlap with each ERA5 grid cell determines the weight with which that forcing grid cell contributes to the HRU-averaged value. This procedure is applied to all seven forcing variables and all time steps to generate HRU-averaged forcing (*./CWARHM/4b_remapping/2_forcing*).

We then apply a constant environmental lapse rate of 0.0065 $K{\cdot}m^{-1}$ (Wallace & Hobbs, 2006, p. 421) to the HRU-averaged air temperature data to account for any differences between ERA5 data point elevation and mean HRU elevation (Figure A7). To avoid excessive data access, the SUMMA-specific variable *data_step* (which specifies the temporal resolution of the forcing data in [s]) is added to each forcing file at the same time as lapse rates are applied.

Key assumptions are that (1) a temporally and spatially constant lapse rate can be used. This is common in gridded analysis but typically not locally accurate (Minder et al., 2010). Local lapse rates may be very different from this assumed value, especially in complex terrain and at seasonal or hourly time scales (Cullen & Marshall, 2011; Minder et al., 2010) . Regionally and temporally variable lapse rates are a possible way to improve this part of the workflow (e.g., Dutra et al., 2020) but doing so is beyond the

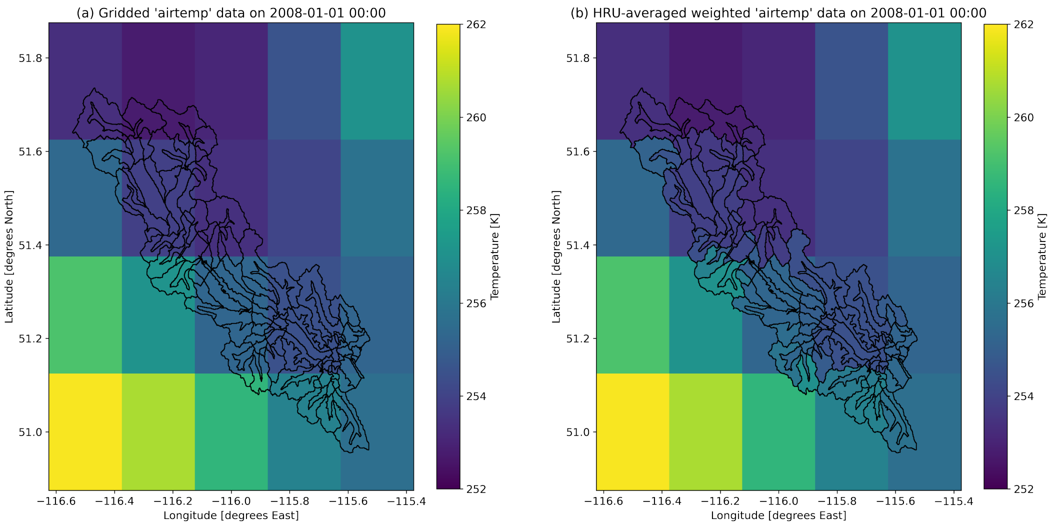

Figure A6. (a) Original gridded air temperature data as found in the ERA5 data. (b) HRU-averaged air temperature obtained as a weighted average of the relative contributions of each ERA5 grid cell to each HRU. Temperatures shown outside the catchment boundaries are the original gridded values.

scope of this study. (2) The influence of slope and aspect on radiation fluxes is currently not accounted for in forcing data preparation.

A24 Model-specific workflow elements

This section provides details model-specific steps of the workflow (shown in dark grey in Figure 2). These steps form the interface between preprocessed data and models.

A24.1 SUMMA and mizuRoute installation The source code for both SUMMA and mizuRoute can be obtained through GitHub (see Section A1). Scripts in *./CWARHM/2_install* provide code to download the latest version of both models to a local machine. Both models are written in Fortran and need to be compiled to create executables. The exact commands and settings needed will vary between different computational environments. The workflow contains examples of model compile code for a specific High Performance Computing environment.

Key assumptions are as follows. (1) The user has determined the appropriate settings to compile both models on their own computational infrastructure and made the necessary changes to our provided example code. (2) Both scripts assume that the "develop" branch of each model is the version of interest. (3) A Linux or MacOS environment is recommended because compiling the SUMMA and mizuRoute source code requires a netCDF-Fortran library to be installed locally and this library is not supported on Windows yet. A basic alternative that avoids compiling the source code is to install pySUMMA and mizuRoute through Conda, but this provides pre-compiled executables only. Access to the source code is not possible and updates present on GitHub may not immediately appear in the pySUMMA Conda distribution.

A24.2 Shapefile sorting to ensure expected order of model elements SUMMA makes certain assumptions about GRU and HRU order in its input files. These are: (1) GRUs and HRUs are in the same order if the forcing files and all SUMMA input files that con-

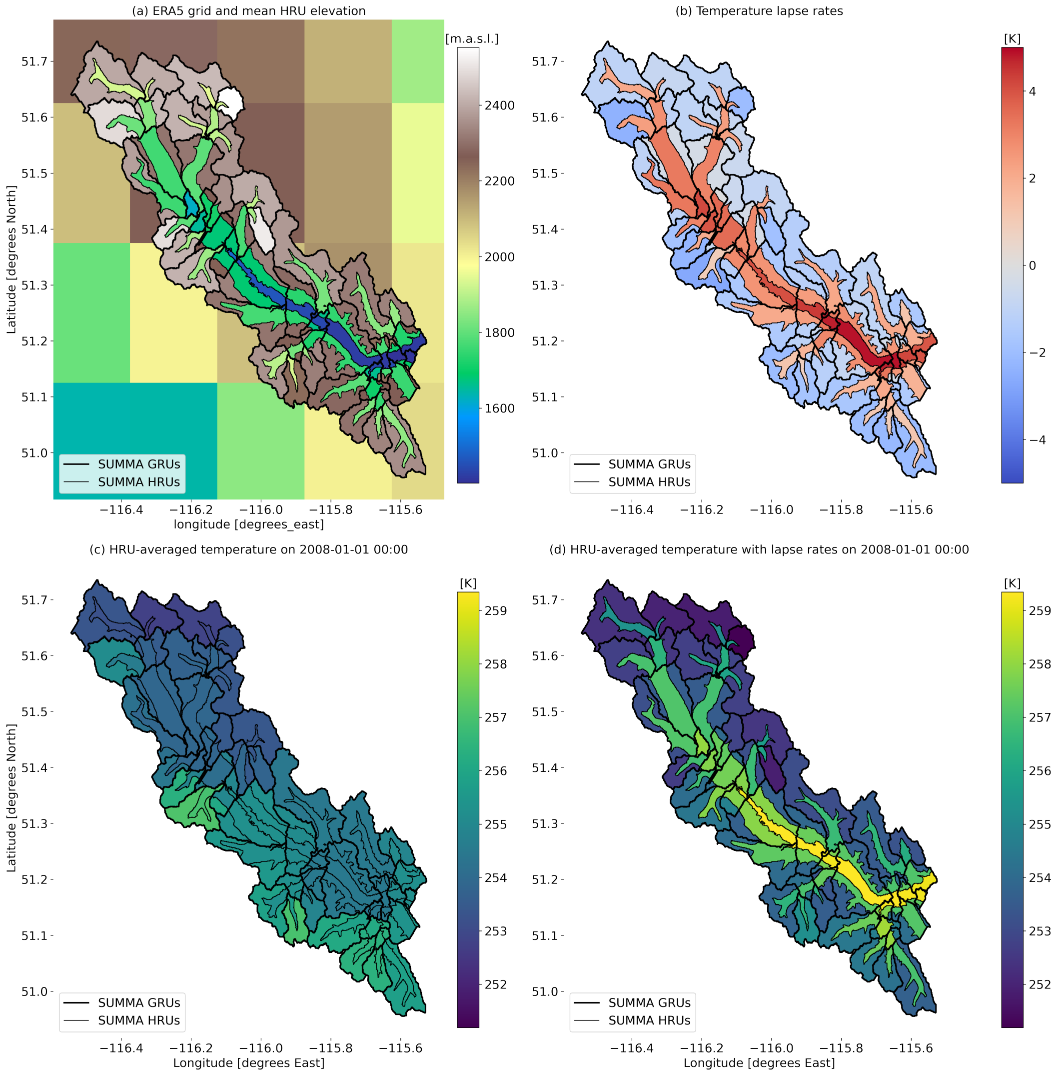

Figure A7. (a) HRU-averaged elevation derived from MERIT Hydro adjusted elevations data. ERA5 grid point elevation calculated from geopotential data and a spatially constant gravitational acceleration value, visualized as grid cells. (b) Temperature lapse values based on a constant lapse rate and a weighted difference between ERA5 grid point elevation and HRU mean elevation. (c) Air temperature data before lapse rates are applied. (d) Air temperature data after lapse rates are applied.

tain information at the GRU and HRU level; and (2) HRUs inside a given GRU are found at subsequent indices in each NetCDF file. Note that these requirements do not specify anything about the values of the GRU and HRU IDs and only focus on the order in which the IDs appear in files. The code in *./CWARHM/4a_sort_shape* sorts the shapefile that contains the catchment delineation into GRUs and HRUs before this shapefile is used by other scripts. This is more efficient than postponing this sorting until the SUMMA input files are generated. A key assumption is that computational efficiency is an important consideration and therefore this model-specific requirement should be run before the (model-agnostic) remapping is performed.

A24.3 SUMMA input files SUMMA requires several different configuration files: (1) default parameter values at the GRU and HRU level; (2) *lookup tables* with predefined soil and vegetation parameters for different soil and land classes; (3) a *model decisions* file that specifies which modeling decisions (e.g., the type of numerical solver) and flux parametrizations to use; (4) an *output control* file that specifies which internal model variables to write as model output, at which temporal resolution to do so and which, if any, summary statistics to provide; (5) a *file manager* file that specifies the file paths to all model inputs and outputs as well as the time period for the simulation; (6) a *forcing file list* file that specifies the names of all meteorological forcing files to use; (7) a *trial parameter* file that can be used to overrule any parameter value specified in the default parameter files and in the lookup tables that can be helpful to quickly test different parameter values during e.g., calibration; (8) an *initial conditions* file that specifies the model states at the beginning of the first time step; and (9) an *attribute* file that contains topographic information such as elevation, soil type, and land use type at the HRU level.

In our example setup, files with default parameter values, lookup tables, model decisions, and requested outputs are provided as part of the repository. These files do not require any information from the preprocessing steps for forcing data and geospatial parameter fields and can therefore simply be copied into the new SUMMA settings directory. The file manager and forcing file list are populated with information available in the workflow control file. The workflow generates a trial parameter file that, for our test cases, specifies a required value for only one parameter. This parameter controls the time resolution of SUMMA's simulations and is here specified as 900 seconds (i.e., four times smaller than the 1-hourly forcing data resolution) to improve numerical convergence of the model equations. The initial conditions file serves a dual purpose: it specifies the model states at the start of the simulation and the vertical discretization of the soil domain into discrete layers. In this example, SUMMA is initialized with eight soil layers of increasing thickness (0.025 m for the top layer, 1.50 m for the bottom layer), without any snow or ice present, with some soil and groundwater liquid water storage and at a constant temperature of the soil, and canopy domains of 10°C. The attributes file is populated with data from the user's shapefiles (GRU and HRU IDs, HRU-to-GRU mapping, longitude and latitude, HRU area) and from the geospatial preprocessing steps. Figure A8 shows the original geospatial parameter fields that are the outcomes of our model-agnostic preprocessing steps and how these are converted into model-specific values for SUMMA's attributes file. All scripts are available in the subdirectories of *./CWARHM/5_model_input/SUMMA*.

Key assumptions are (1) that the HRU and GRU default parameter files, model decisions and lookup table files are assumed to be sensible choices for the domain of interest. In particular, the choice of ROSETTA lookup table for soil properties (NCAR Research Applications Laboratory | RAL, 2021; U.S. Department of Agriculture: Agricultural Research Service (USDA ARS), 2021) and the modified IGBP table for vegetation properties (NCAR Research Applications Laboratory | RAL, 2021) inform how the geospatial data is preprocessed (i.e., which geospatial data sets are used and how they are transformed). (2) Vertical discretization of domain is currently set at eight soil layers with increasing thickness with depth. (3) Initial conditions are dry and warm, and there is no snow and ice present in the domain. (4) Model decisions relying on *contourLength* and *tan_slope* attributes are not supported (currently this is the baseflow model decision *qbaseTopmodel*, as well as certain radiation calculations that account for slope inclination). Attribute variable *downHRUindex* is only used if decision *qbaseTopmodel* is active and is therefore set to zero.

A24.4 mizuRoute input files mizuRoute requires several configuration files: (1) a *default parameter* file that has values for its different routing schemes; (2) a *network topology* file that contains a description of the river network and its properties; (3) optionally, a *remapping* file that shows how output from a hydrologic model should be mapped onto mizuRoute's routing network; and (4) a *mizuRoute.control* file that specifies the necessary file paths and routing settings. In our example setup, a default routing parameter file is provided as part of the repository. This file does not require any information from the preprocessing steps for forcing data and geospatial parameter fields and can therefore simply be copied into the new mizuRoute settings directory. The network topology file contains a description of the routing basins and their associated stream segments. It specifies which basins and segments exist, which segment each basin drains into and physical properties of the domain such as drainage area, segment length and segment slope. The optional remapping file only needs to be used in cases where the hydrologic model operates on model elements that do not map directly onto mizuRoute's routing basins. In such a case the remapping file specifies the weight with which each hydrologic catchment contributes flow to each routing basin. The mizuRoute.control file is populated with information available in the workflow control file. All scripts are available in the subdirectories of *./CWARHM/5_model_input/mizuRoute.*

Key assumptions are (1) that the provided routing parameter values are appropriate for the domain and (2) Hillslope routing (i.e., routing between different SUMMA HRUs inside a given SUMMA GRU) is performed by SUMMA. mizuRoute is configured to do the river network routing between different SUMMA GRUs.

A24.5 Model runs Model runs use the compiled SUMMA and mizuRoute executables to perform simulations using the inputs and settings defined in their respective configuration files (*./CWARHM/6_model_runs*). As part of the model run scripts, model configuration files are copied into the simulations output directories. This ensures traceability of the simulations by keeping a record of the settings used to generate the simulations.

A25 Post-processing

Post-processing of model results in this example is limited to the code needed to generate the modeling domain figure in this manuscript (*./CWARHM/7_visualization*). Further visualization code may be added over time, as such code is created for specific experiments.

Appendix B Note on data accuracy

Our example workflow uses ERA5 forcing data, MERIT Hydro DEM, SOILGRIDS-derived soil texture classes, and MODIS IGBP land classes for their global coverage. This enables global applications of the workflow. Such global datasets are based on a combination of observations and geospatial data processing methods to estimate data values for locations where no observations are available. These approaches may need to sacrifice local information content for global coverage and are not always able to utilize the most accurate local data available.

ERA5 is a reanalysis product from a data assimilating numerical weather prediction model. ERA5 precipitation estimates compare favorably to other global products at a daily resolution (Beck et al., 2019) but are typically not as accurate as local gauge

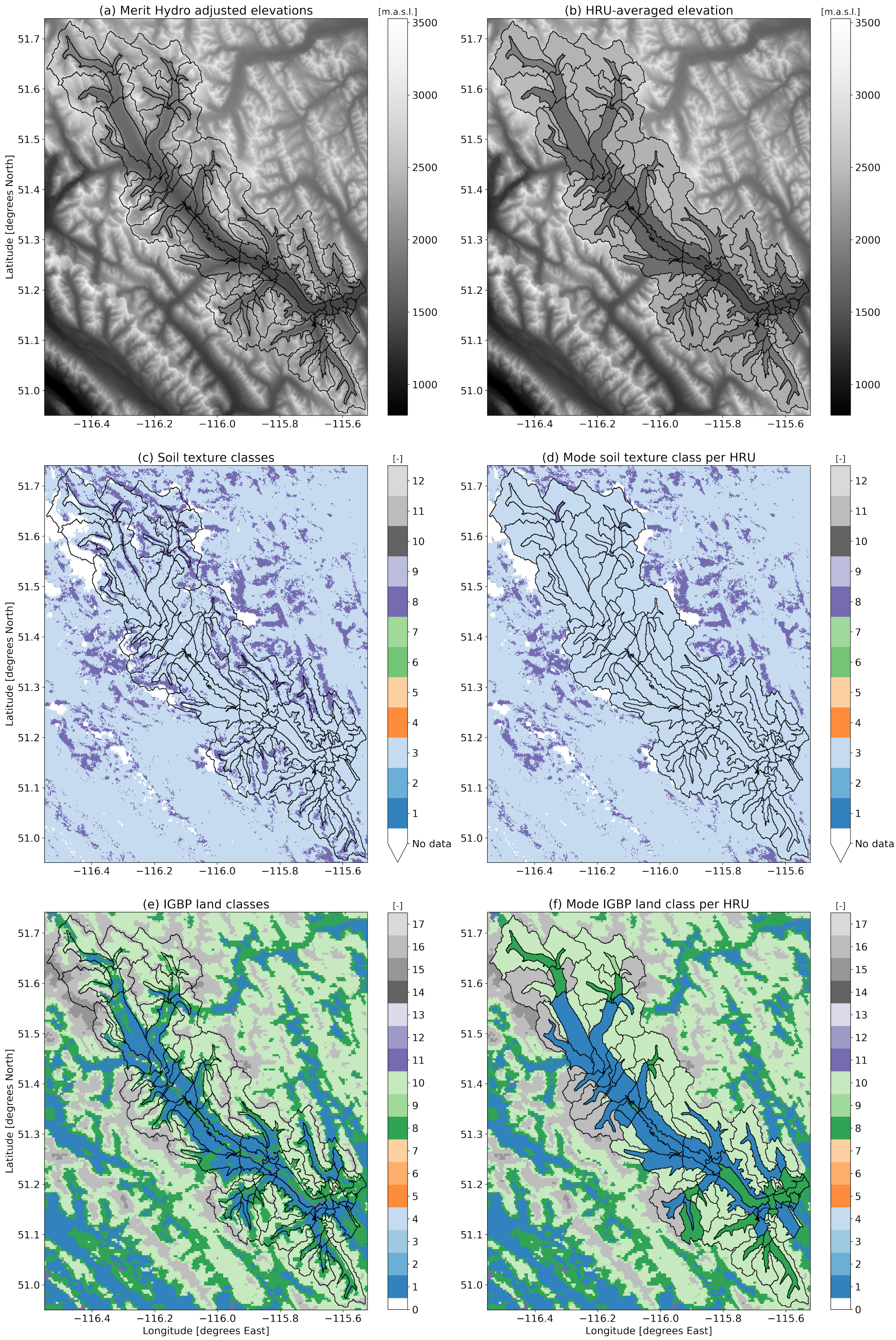

Figure A8. Mapping of geospatial parameter fields onto model elements. (a, b) MERIT Hydro adjusted elevations DEM source data and the mean elevation per HRU. (c, d) Soil texture classes derived from SOILGRIDS sand, silt and clay percentages and the most common class per HRU. (e, f) IGBP land classes from MODIS data and the most common class per HRU.

or radar-based observations, especially in regions with complex topography (e.g., Amjad et al., 2020; Q. Jiang et al., 2021; Tang et al., 2020; Xu et al., 2019). H. Jiang et al. (2020) show a similar reduced accuracy of ERA5 compared to station observations for direct and diffuse solar radiation estimates. Less is known about the accuracy of the remaining ERA5 forcing variables used in our workflow, and it is possible that the relatively coarse resolution of ERA5 data means that these variables may not be as accurate as local products.

The MERIT Hydro hydrologically adjusted elevation dataset (Yamazaki et al., 2019) is based on the MERIT DEM (Yamazaki et al., 2017), which itself is the result of applying an error-removal algorithm to existing space-borne DEMs. It is available globally at approximately 90 m spatial resolution. The MERIT Hydro data represent an advance over earlier products such as HydroSheds (Lehner et al., 2008), especially at higher latitudes, but some uncertainty in the produced hydrography data remains in regions with low topographic variation, with endorheic basins, with seasonally varying connectivity, and with channel bifurcations. The MERIT Hydro hydrologically adjusted elevations are a modification of the MERIT DEM that satisfies the condition "downstream pixels are not higher than upstream pixels". This procedure relies on a combination of correctly identifying endorheic basins, connections between sub-basins, and adjusting pixel elevations to create continuous flow paths. It is unknown to what extent this procedure affects the mean catchment elevation we derive from the hydrologically adjusted elevation. It is plausible that mean catchment elevations derived from this data will be less accurate in regions with rapidly varying topography, where catchment slopes are steep compared to the MERIT Hydro resolution.

The SoilGrids database uses observations of approximately 150,000 soil profiles, pseudo-observations that encode expert knowledge in a similar way to actual observed soil profiles, and machine learning to provide global estimates of various soil properties at a 250 m resolution. Ten-fold cross-validation of the resulting sand, silt, and clay percentage data used in our workflow shows that this approach explains approximately 75% of the variation in these soil properties. There is no systematic over or under prediction of these properties, but large differences between estimates and observations exist nonetheless in certain cases (Hengl et al., 2017).

MODIS MCD12Q1_v6 data uses a combination of random forests, bias and error correction based on ancillary data, and a hidden Markov Model approach to convert pre-processed satellite reflectance imagery into land cover classification categories. Ten-fold cross-validation of the resulting classification indicates that the IGBP classes used in our workflow are accurate in approximately two-thirds of cases. Misclassifications tend to occur in regions that contain substantial land cover variability at scales smaller than the 500 m MODIS resolution is provided at and along climatic gradients where the cover type changes gradually (Sulla-Menashe et al., 2019).

We therefore recommend that users replace our chosen global data products with more appropriate local data if such data are available and the project scope lies within the data domain. Due to the modular nature of the workflow, this replacement requires only minimal changes to the model configuration code. In terms of Figure 7, incorporating a different data set would require a new data-specific pre-processing module for which our existing workflow can serve as a guide. We emphasize that this workflow is intended to provide a baseline configuration upon which a user can improve. Our workflow does not contain any elements that compare the resulting simulations to observations to ascertain the quality of these simulations. A model setup generated through this workflow should thus not be assumed to be fit for a given purpose, unless shown to be so by the user's own model evaluation procedures.

Open Research

The latest version of the workflow code presented in this study is available on https://github.com/CH-Earth/CWARHM with the specific version used to generate Figures 4, 5, 6, A1, A5-A8 via https://dx.doi.org/10.5281/zenodo.7134868 (Knoben, Marsh, & Tang, 2022) accessible under GNU GPL v3.0.

The SUMMA (Clark et al., 2015a, 2015b; Clark, Zolfaghari, et al., 2021) versions used for simulations in this paper can be identified by Git commit ID edd328c8c2e7b81c3b222d4c7d2544769036fd4 (global domain excluding North America) and Git commit ID 3d17543db618cb5b9c7600d6d0de658943056c93 (North America domain and Bow at Banff domain). Source code accessible on https://github.com/CH-Earth/summa under the GNU GPLv3 license.

The mizuRoute (Mizukami et al., 2016, 2021) version used for simulations in this paper can be identified by Git commit ID 137820620f624f84f8cdb1d4e9884b8222a3f3df (global domain excluding North America), Git commit ID c2de53d242fc41b94c48119d23b78da1f35719ee (North America domain) and Git commit ID d43066b56a7361f3d4a9c7b07264d7d52a9686f1 (Bow at Banff domain). Source code accessible on https://github.com/ESCOMP/mizuRoute under the GNU GPLv3 license.

The single level ERA5 data (Hersbach et al., 2018) used as meteorological model input data are available at the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) via https://dx.doi.org/10.24381/cds.adbb2d47 under the *Licence to use Copernicus Products* (https://cds.climate.copernicus.eu/api/v2/terms/static/licence-to-use-copernicus-products.pdf; last access 2021-11-04).

The pressure level ERA5 data (Hersbach et al., 2017) used as meteorological model input data are available at the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) via MARS request (no DOI) under *Licence to use Copernicus Products* (https://cds.climate.copernicus.eu/api/v2/terms/static/licence-to-use-copernicus-products.pdf; last access 2021-11-04). Data downloaded on 2021-04-17 for the Bow at Banff test case; between 2020-11-14 and 2020-12-23 for the North America test case; and on 2021-06-20 for the global test case.

The MERIT Hydro Hydrologically Adjusted Elevations (Yamazaki et al., 2019) used as Digital Elevation Model to determine mean catchment elevations is available at http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/ (last webpage access on 2021-11-04) as version v1.0.1 (no DOI available; data downloaded on 2021-04-17 for the Bow at Banff test case; on 2021-05-15 for the North America test case; between 2022-06-03 and 2022-07-02 for the global test case), accessible under CC-BY-NC 4.0 or ODbL 1.0.

The MODIS MCD12Q1 V6 data (Friedl & Sulla-Menashe, 2019; Sulla-Menashe & Friedl, 2018; Sulla-Menashe et al., 2019) used to find a representative IGBP land cover class for each model element is available at the NASA EOSDIS Land Processes DAAC via https://dx.doi.org/10.5067/MODIS/MCD12Q1.006, with no restrictions on reuse, sale or redistribution.

The Global USDA-NRCS soil texture class map (Knoben, 2021) derived from the Soilgrids250m data set (Hengl et al., 2017) and used to find a representative USGS soil type class for each model element is available as a Hydroshare resource via https://dx.doi.org/10.4211/hs.1361509511e44adfba814f6950c6e742, under ODbL v1.0.

The shapefiles that contain the catchment delineations for all test cases are derived from the MERIT Hydro basins data set (Lin et al., 2019), which is originally made available for research purposes on http://hydrology.princeton.edu/data/mpan/MERIT_Basins/. The basin discretization and river network files for the Bow at Banff test case are a subset of the original files, with the original basins further discretized into elevation bands. The Bow at Banff shapefiles are provided as part of the workflow repository. For the global test case the original MERIT Hydro basin and hillslope files were merged into a single shapefile per continent, as were the separate river network files. For the continental test case the original MERIT Hydro basin and hillslope files were merged into a single shapefile per continent, and updated to correct any invalid geometries in basin polygons and to separate coastal hillslope polygons into two separate polygons if the original polygon was intersected by a river segment. The separate river network files were merged into a single file as well. The shapefiles that contain the catchment delineation and river network for the global and continental test cases are available as a Hydroshare resource (Knoben, Clark, et al., 2022) via `https://dx.doi.org/10.4211/hs.46d980a71d2c4365aa290dc1bfdac823`, under CC BY-NC-SA.

Acknowledgements

We are thankful for the general comments for workflow improvement given by Dave Casson, Hongli Liu, Jim Freer and Hannah Burdett. We are also thankful for the efforts of Reza Zolfaghari and Kyle Klenk who assisted in understanding and solving specific numerical issues, to Louise Arnal for feedback on the workflow schematics, and to Ala Bahrami, Dayal Wijayarathne and Tricia Stadnyk for letting us share their progress on incorporating the MESH and HYPE models into our workflow concept. Finally, we gratefully acknowledge the effort of the reviewers whose comments helped us clarify the message in this paper.

Wouter Knoben, Martyn Clark, Shervan Gharari, Chris Marsh, Ray Spiteri and Guoqiang Tang were supported by the Global Water Futures program, University of Saskatchewan. The work of David Tarboton was supported by the US National Science Foundation under collaborative grants ICER 1928369 and OAC 1835569. The work of Jerad Bales was supported by NSF grant EAR 1849458. NCAR co-author contributions were supported by the US Army Corps of Engineers Responses to Climate Change program and the US Bureau of Reclamation under Cooperative Agreement R16AC00039.

References

Amjad, M., Yilmaz, M. T., Yucel, I., & Yilmaz, K. K. (2020, May). Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography. *Journal of Hydrology*, *584*, 124707. Retrieved 2021-06-17, from `https://linkinghub.elsevier.com/retrieve/pii/S0022169420301670` doi: 10.1016/j.jhydrol.2020.124707

Archfield, S. A., Clark, M., Arheimer, B., Hay, L. E., McMillan, H., Kiang, J. E., ... Over, T. (2015, December). Accelerating advances in continental domain hydrologic modeling. *Water Resources Research*, *51*(12), 10078–10091. Retrieved 2021-06-22, from `http://doi.wiley.com/10.1002/2015WR017498` doi: 10.1002/2015WR017498

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2020, February). Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrology and Earth System Sciences*, *24*(2), 535–559. Retrieved 2022-06-02, from `https://hess.copernicus.org/articles/24/535/2020/` doi: 10.5194/hess-24-535-2020

Arsenault, R., Poulin, A., Côté, P., & Brissette, F. (2014, July). Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration. *Journal of Hydrologic Engineering*, *19*(7), 1374–1384. Retrieved from `http://ascelibrary.org/doi/10.1061/%28ASCE%29HE.1943-5584.0000938` (ISBN: 10.1061/(ASCE)HE.1943-5584.0000938) doi: 10.1061/(ASCE)HE.1943-5584.0000938

Añel, J. A. (2017, March). Comment on "Most computational hydrology is not reproducible, so is it really science?" by Christopher Hutton et al.: NECESSARY CONDITIONS FOR REPRODUCIBILITY. *Water Resources Research*, *53*(3), 2572–2574. Retrieved 2021-02-17, from http://doi.wiley.com/10.1002/2016WR020190 doi: 10.1002/2016WR020190

Bandaragoda, C., Castronova, A., Istanbulluoglu, E., Strauch, R., Nudurupati, S., Phuong, J., … Wang, S. (2019, October). Enabling Collaborative Numerical Modeling in Earth Sciences using Knowledge Infrastructure. *Environmental Modelling & Software*, *120*, 104424. Retrieved 2021-06-19, from https://linkinghub.elsevier.com/retrieve/pii/S1364815219301562 doi: 10.1016/j.envsoft.2019.03.020

Barnhart, K. R., Hutton, E. W. H., Tucker, G. E., Gasparini, N. M., Istanbulluoglu, E., Hobley, D. E. J., … Bandaragoda, C. (2020). Short communication: Landlab v2.0: A software package for earth surface dynamics. *Earth Surface Dynamics*, *8*(2), 379–397. Retrieved from https://esurf.copernicus.org/articles/8/379/2020/ doi: 10.5194/esurf-8-379-2020

Bast, R. (2019, August). A FAIRer future. *Nature Physics*, *15*(8), 728–730. Retrieved 2021-02-17, from http://www.nature.com/articles/s41567-019-0624-3 doi: 10.1038/s41567-019-0624-3

Bavay, M., Reisecker, M., Egger, T., & Korhammer, D. (2022, January). Inishell 2.0: semantically driven automatic GUI generation for scientific models. *Geoscientific Model Development*, *15*(2), 365–378. Retrieved 2022-07-26, from https://gmd.copernicus.org/articles/15/365/2022/ doi: 10.5194/gmd-15-365-2022

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., … Wood, E. F. (2019, January). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, *23*(1), 207–224. Retrieved 2021-06-17, from https://hess.copernicus.org/articles/23/207/2019/ doi: 10.5194/hess-23-207-2019

Benham, E., Ahrens, R. J., & Nettleton, D. W. (2009). *MO5 Soil Technical Note-16 - Clarification of Textural Class Boundaries* (Tech. Rep.). Lincoln, Nebraska: National Soil Survey Center, USDA-NRCS.

Bennett, A., Hamman, J., & Nijssen, B. (2020, March). MetSim: A Python package for estimation and disaggregation of meteorological data. *Journal of Open Source Software*, *5*(47), 2042. Retrieved 2021-06-19, from https://joss.theoj.org/papers/10.21105/joss.02042 doi: 10.21105/joss.02042

Bennett, A., Hamman, J., Nijssen, B., Clark, E., & Andreadis, K. (2018, September). *Uw-Hydro/Metsim: Triangular Precipitation Reference.* Zenodo. Retrieved 2022-07-13, from https://zenodo.org/record/1418546 doi: 10.5281/ZENODO.1418546

Blair, G. S., Beven, K., Lamb, R., Bassett, R., Cauwenberghs, K., Hankin, B., … Towe, R. (2019, December). Models of everywhere revisited: A technological perspective. *Environmental Modelling & Software*, *122*, 104521. Retrieved 2021-06-19, from https://linkinghub.elsevier.com/retrieve/pii/S1364815218303773 doi: 10.1016/j.envsoft.2019.104521

Blodgett, D., & Dornblut, I. (2018, January). *OGC® WaterML 2: Part 3 - Surface Hydrology Features (HY_features) - Conceptual Model.* Retrieved 2021-12-02, from http://docs.opengeospatial.org/is/14-111r6/14-111r6.html#_catchment_area

Blöschl, G., Bárdossy, A., Koutsoyiannis, D., Kundzewicz, Z. W., Littlewood, I., Montanari, A., & Savenije, H. (2014, April). On the future of journal publications in hydrology. *Water Resources Research*, *50*(4), 2795–2797. Retrieved 2021-02-17, from http://doi.wiley.com/10.1002/2014WR015613 doi: 10.1002/2014WR015613

Bock, a. R., Hay, L. E., McCabe, G. J., Markstrom, S. L., & Atkinson, R. D.

(2015). Parameter regionalization of a monthly water balance model for the conterminous United States. *Hydrology and Earth System Sciences Discussions*, *12*(9), 10023–10066. Retrieved from http://www.hydrol-earth-syst-sci-discuss.net/12/10023/2015/hessd-12-10023-2015.html doi: 10.5194/hessd-12-10023-2015

Chaney, N., & Fisher, C. (2021, November). *chaneyn/geospatialtools: geospatialtools_nov2021.* Zenodo. Retrieved 2021-11-12, from https://zenodo.org/record/5676031 doi: 10.5281/ZENODO.5676031

Choi, Y.-D. (2020). *Toward open and reproducible environmental modeling by integrating online data repositories, computational environments, and model application programming interfaces.* HydroShare. Retrieved 2020-07-13, from http://www.hydroshare.org/resource/33cfb9622a354442b2b0a869b15ea6b0

Choi, Y. D. (2021). *Advancing Reproducibility in Environmental Modeling: Integration of Open Repositories, Process Containerizations, and Seamless Workflows* (Doctoral dissertation, University of Virginia). doi: 10.18130/G9TG-0C69

Choi, Y.-D., Goodall, J. L., Sadler, J. M., Castronova, A. M., Bennett, A., Li, Z., … Tarboton, D. G. (2021, January). Toward open and reproducible environmental modeling by integrating online data repositories, computational environments, and model Application Programming Interfaces. *Environmental Modelling & Software*, *135*, 104888. Retrieved 2021-01-13, from https://linkinghub.elsevier.com/retrieve/pii/S1364815220309452 doi: 10.1016/j.envsoft.2020.104888

Chuah, J., Deeds, M., Malik, T., Choi, Y., & Goodall, J. L. (2020). Documenting Computing Environments for Reproducible Experiments. In *Parallel Computing: Technology Trends* (Edited by I. Foster, G. R. Joubert, L. Kučera, W. E. Nagel and F. Peters ed., Vol. 36, pp. 756–765). Amsterdam, Netherlands: IOS Press. Retrieved from http://doi.org/10.3233/APC200106

Clark, M. P., Luce, C. H., AghaKouchak, A., Berghuijs, W., David, C. H., Duan, Q., … Tyler, S. W. (2021, April). Open Science: Open Data, Open Models, …and Open Publications? *Water Resources Research*, *57*(4). Retrieved 2021-11-01, from https://onlinelibrary.wiley.com/doi/10.1029/2020WR029480 doi: 10.1029/2020WR029480

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., … Rasmussen, R. M. (2015a, April). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, *51*(4), 2498–2514. Retrieved 2021-02-03, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017198 doi: 10.1002/2015WR017198

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., … Marks, D. G. (2015b, April). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resources Research*, *51*(4), 2515–2542. Retrieved 2021-02-03, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017200 doi: 10.1002/2015WR017200

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., … Papalexiou, S. M. (2021, September). The Abuse of Popular Performance Metrics in Hydrologic Modeling. *Water Resources Research*, *57*(9). Retrieved 2021-11-24, from https://onlinelibrary.wiley.com/doi/10.1029/2020WR029001 doi: 10.1029/2020WR029001

Clark, M. P., Zolfaghari, R., Green, K. R., Trim, S., Knoben, W. J. M., Bennett, A., … Spiteri, R. J. (2021, April). The numerical implementation of land models: Problem formulation and laugh tests. *Journal of Hydrometeorology*. Retrieved 2021-04-20, from https://journals.ametsoc.org/view/journals/hydr/aop/JHM-D-20-0175.1/JHM-D-20-0175.1.xml doi: 10.1175/JHM-D-20-0175.1

Coon, E. T., & Shuai, P. (2022, November). Watershed Workflow: A toolset for parameterizing data-intensive, integrated hydrologic models. *Environmental Modelling & Software*, *157*, 105502. Retrieved 2022-10-05, from https://linkinghub.elsevier.com/retrieve/pii/S1364815222002031 doi: 10.1016/j.envsoft.2022.105502

Copernicus Climate Change Service (C3S). (2017). *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.* Retrieved 2020-03-26, from https://cds.climate.copernicus.eu/cdsapp#!/home

Cullen, R. M., & Marshall, S. J. (2011, September). Mesoscale Temperature Patterns in the Rocky Mountains and Foothills Region of Southern Alberta. *Atmosphere-Ocean*, *49*(3), 189–205. Retrieved 2021-03-17, from http://www.tandfonline.com/doi/abs/10.1080/07055900.2011.592130 doi: 10.1080/07055900.2011.592130

Dash, P., & Tarboton, D. (2022). *Hydro-DS.* Retrieved 2022-07-13, from https://github.com/CI-WATER/Hydro-DS/

Doherty, J., & Skahill, B. E. (2006, August). An advanced regularization methodology for use in watershed model calibration. *Journal of Hydrology*, *327*(3-4), 564–577. Retrieved 2022-07-28, from https://linkinghub.elsevier.com/retrieve/pii/S0022169405006542 doi: 10.1016/j.jhydrol.2005.11.058

Dutra, E., Muñoz-Sabater, J., Boussetta, S., Komori, T., Hirahara, S., & Balsamo, G. (2020, May). Environmental Lapse Rate for High-Resolution Land Surface Downscaling: An Application to ERA5. *Earth and Space Science*, *7*(5). Retrieved 2021-03-17, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019EA000984 doi: 10.1029/2019EA000984

Essawy, B. T., Goodall, J. L., Voce, D., Morsy, M. M., Sadler, J. M., Choi, Y. D., . . . Malik, T. (2020, December). A taxonomy for reproducible and replicable research in environmental modelling. *Environmental Modelling & Software*, *134*, 104753. Retrieved 2021-06-19, from https://linkinghub.elsevier.com/retrieve/pii/S1364815219311612 doi: 10.1016/j.envsoft.2020.104753

Essawy, B. T., Goodall, J. L., Xu, H., Rajasekar, A., Myers, J. D., Kugler, T. A., . . . Moore, R. W. (2016, April). Server-side workflow execution using data grid technology for reproducible analyses of data-intensive hydrologic systems. *Earth and Space Science*, *3*(4), 163–175. Retrieved 2021-11-23, from https://onlinelibrary.wiley.com/doi/10.1002/2015EA000139 doi: 10.1002/2015EA000139

Famiglietti, J. S., Murdoch, L., Lakshmi, V., Arrigo, J., & Hooper, R. (2011). *Establishing a Framework for Community Modeling in Hydrologic Science* (Tech. Rep.). Beckman Center of the National Academies, University of California at Irvine.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., . . . Alsdorf, D. (2007, May). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, *45*(2), RG2004. Retrieved 2021-06-12, from http://doi.wiley.com/10.1029/2005RG000183 doi: 10.1029/2005RG000183

Flügel, W.-A. (1995, April). Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrological Processes*, *9*(3-4), 423–436. Retrieved 2022-07-26, from https://onlinelibrary.wiley.com/doi/10.1002/hyp.3360090313 doi: 10.1002/hyp.3360090313

Friedl, M., & Sulla-Menashe, D. (2019). *MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006.* NASA EOSDIS Land Processes DAAC. Retrieved 2021-06-12, from https://lpdaac.usgs.gov/products/mcd12q1v006/ (Type: dataset) doi: 10.5067/MODIS/MCD12Q1.006

Gan, T., Tarboton, D. G., Dash, P., Gichamo, T. Z., & Horsburgh, J. S. (2020, August). Integrating hydrologic modeling web services with online data sharing to prepare, store, and execute hydrologic models. *Environmental Modelling & Software*, *130*, 104731. Retrieved 2021-06-19, from https://linkinghub.elsevier.com/retrieve/pii/S1364815219311661 doi: 10.1016/j.envsoft.2020.104731

Gharari, S., Clark, M. P., Mizukami, N., Knoben, W. J. M., Wong, J. S., & Pietroniro, A. (2020). Flexible vector-based spatial configurations in land models. *Hydrology and Earth System Sciences*, *24*, 5953–5971. doi: https://doi.org/10.5194/hess-24-5953-2020

Gharari, S., Vanderkelen, I., Tefs, A., Mizukami, N., Stadnyk, T. A., Lawrence, D., & Clark, M. P. (2022, March). *A Flexible Multi-Scale Framework to Simulate Lakes and Reservoirs in Earth System Models* (preprint). Hydrology. Retrieved 2022-07-25, from http://www.essoar.org/doi/10.1002/essoar.10510902.1 doi: 10.1002/essoar.10510902.1

Gichamo, T. Z., Sazib, N. S., Tarboton, D. G., & Dash, P. (2020, March). HydroDS: Data services in support of physically based, distributed hydrological models. *Environmental Modelling & Software*, *125*, 104623. Retrieved 2021-06-19, from https://linkinghub.elsevier.com/retrieve/pii/S1364815217308009 doi: 10.1016/j.envsoft.2020.104623

Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., . . . Yu, X. (2016, October). Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science*, *3*(10), 388–415. Retrieved 2021-11-24, from https://onlinelibrary.wiley.com/doi/10.1002/2015EA000136 doi: 10.1002/2015EA000136

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations : elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *3813*(22), 3802–3813. doi: https://doi.org/10.1002/hyp.6989

Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R., & Mao, Y. (2018, August). The Variable Infiltration Capacity model version 5 (VIC-5): infrastructure improvements for new applications and reproducibility. *Geoscientific Model Development*, *11*(8), 3481–3496. Retrieved 2021-04-30, from https://gmd.copernicus.org/articles/11/3481/2018/ doi: 10.5194/gmd-11-3481-2018

Havens, S., Marks, D., Sandusky, M., Hedrick, A., Johnson, M., Robertson, M., & Trujillo, E. (2020, November). Automated Water Supply Model (AWSM): Streamlining and standardizing application of a physically based snow model for water resources and reproducible science. *Computers & Geosciences*, *144*, 104571. Retrieved 2021-11-10, from https://linkinghub.elsevier.com/retrieve/pii/S0098300420305598 doi: 10.1016/j.cageo.2020.104571

Havens, S., Marks, D., Sandusky, M., Johnson, M., Robertson, M., Hedrick, A., & Kormos, P. (2019, September). *Automated Water Supply Model (AWSM).* Zenodo. Retrieved 2022-07-13, from https://zenodo.org/record/3455762 doi: 10.5281/ZENODO.3455762

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., . . . Kempen, B. (2017, February). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. Retrieved 2021-06-11, from https://dx.plos.org/10.1371/journal.pone.0169748 doi: 10.1371/journal.pone.0169748

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., . . . Thépaut, J.-N. (2018). *ERA5 hourly data on single levels from 1979 to present.* doi: 10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . Thépaut, J.-N. (2017). *Complete ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.*

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. (2020, July). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. Retrieved 2021-11-23, from https://onlinelibrary.wiley.com/doi/10.1002/qj.3803 doi: 10.1002/qj.3803

Hobley, D. E. J., Adams, J. M., Nudurupati, S. S., Hutton, E. W. H., Gasparini, N. M., Istanbulluoglu, E., & Tucker, G. E. (2017). Creative computing with landlab: an open-source toolkit for building, coupling, and exploring two-dimensional numerical models of earth-surface dynamics. *Earth Surface Dynamics*, *5*(1), 21–46. doi: 10.5194/esurf-5-21-2017

Husain, S. Z., Alavi, N., Bélair, S., Carrera, M., Zhang, S., Fortin, V., ... Gauthier, N. (2016, August). The Multibudget Soil, Vegetation, and Snow (SVS) Scheme for Land Surface Parameterization: Offline Warm Season Evaluation. *Journal of Hydrometeorology*, *17*(8), 2293–2313. Retrieved 2021-11-24, from http://journals.ametsoc.org/doi/10.1175/JHM-D-15-0228.1 doi: 10.1175/JHM-D-15-0228.1

Hut, R. W., van de Giesen, N. C., & Drost, N. (2017, May). Comment on "Most computational hydrology is not reproducible, so is it really science?" by Christopher Hutton et al.: Let hydrologists learn the latest computer science by working with Research Software Engineers (RSEs) and not reinvent the waterwheel ourselves. *Water Resources Research*, *53*(5), 4524–4526. Retrieved 2021-02-17, from http://doi.wiley.com/10.1002/2017WR020665 doi: 10.1002/2017WR020665

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016, October). Most computational hydrology is not reproducible, so is it really science?: REPRODUCIBLE COMPUTATIONAL HYDROLOGY. *Water Resources Research*, *52*(10), 7548–7555. Retrieved 2021-02-17, from http://doi.wiley.com/10.1002/2016WR019285 doi: 10.1002/2016WR019285

Jiang, H., Yang, Y., Bai, Y., & Wang, H. (2020, January). Evaluation of the Total, Direct, and Diffuse Solar Radiations From the ERA5 Reanalysis Data in China. *IEEE Geoscience and Remote Sensing Letters*, *17*(1), 47–51. Retrieved 2021-06-17, from https://ieeexplore.ieee.org/document/8725564/ doi: 10.1109/LGRS.2019.2916410

Jiang, Q., Li, W., Fan, Z., He, X., Sun, W., Chen, S., ... Wang, J. (2021, April). Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. *Journal of Hydrology*, *595*, 125660. Retrieved 2021-06-17, from https://linkinghub.elsevier.com/retrieve/pii/S0022169420311215 doi: 10.1016/j.jhydrol.2020.125660

Knoben, W. J. M. (2021). *Global USDA-NRCS soil texture class map* (Tech. Rep.). Retrieved 2021-06-11, from http://www.hydroshare.org/resource/1361509511e44adfba814f6950c6e742 (Type: dataset) doi: 10.4211/hs.1361509511e44adfba814f6950c6e742

Knoben, W. J. M., Clark, M. P., Gharari, S., & Tang, G. (2022). *Community Workflows to Advance Reproducibility in Hydrologic Modeling: Separating model-agnostic and model-specific configuration steps in applications of large-domain hydrologic models - Continental basin discretizations and river networks* (Tech. Rep.). Retrieved 2022-07-28, from http://www.hydroshare.org/resource/46d980a71d2c4365aa290dc1bfdac823 (Type: dataset) doi: 10.4211/hs.46d980a71d2c4365aa290dc1bfdac823

Knoben, W. J. M., Marsh, C. B., & Tang, G. (2022, October). *CH-Earth/CWARHM: CWARHM workflow updates during peer review.* Zenodo. Retrieved 2022-10-02, from https://zenodo.org/record/7134868 doi: 10.5281/ZENODO.7134868

Kurtzer, G. M., Cclerget, Bauer, M., Kaneshiro, I., Trudgian, D., & Godlove, D. (2021, April). *hpcng/singularity: Singularity 3.7.3.* Zenodo. Retrieved 2022-07-13, from https://zenodo.org/record/1310023 doi: 10.5281/ZENODO.1310023

Kurtzer, G. M., Sochat, V., & Bauer, M. W. (2017, May). Singularity: Scientific containers for mobility of compute. *PLOS ONE*, *12*(5), e0177459. Retrieved 2021-11-11, from https://dx.plos.org/10.1371/journal.pone.0177459 doi: 10.1371/journal.pone.0177459

Lehner, B., & Grill, G. (2013, July). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems: GLOBAL RIVER HYDROGRAPHY AND NETWORK ROUTING. *Hydrological Processes*, *27*(15), 2171–2186. Retrieved 2021-11-10, from https://onlinelibrary.wiley.com/doi/10.1002/hyp.9740 doi: 10.1002/hyp.9740

Lehner, B., Verdin, K., & Jarvis, A. (2008). New Global Hydrography Derived From Spaceborne Elevation Data. *Eos, Transactions American Geophysical Union*, *89*(10), 93. Retrieved 2021-06-18, from http://doi.wiley.com/10.1029/2008EO100001 doi: 10.1029/2008EO100001

Leonard, L., & Duffy, C. (2016, April). Visualization workflows for level-12 HUC scales: Towards an expert system for watershed analysis in a distributed computing environment. *Environmental Modelling & Software*, *78*, 163–178. Retrieved 2021-02-17, from https://linkinghub.elsevier.com/retrieve/pii/S1364815216300019 doi: 10.1016/j.envsoft.2016.01.001

Leonard, L., & Duffy, C. J. (2013, December). Essential Terrestrial Variable data workflows for distributed water resources modeling. *Environmental Modelling & Software*, *50*, 85–96. Retrieved 2021-02-17, from https://linkinghub.elsevier.com/retrieve/pii/S1364815213001941 doi: 10.1016/j.envsoft.2013.09.003

Leonard, L., & Duffy, C. J. (2014, November). Automating data-model workflows at a level 12 HUC scale: Watershed modeling in a distributed computing environment. *Environmental Modelling & Software*, *61*, 174–190. Retrieved 2021-02-17, from https://linkinghub.elsevier.com/retrieve/pii/S1364815214002230 doi: 10.1016/j.envsoft.2014.07.015

Leonard, L., Miles, B., Heidari, B., Lin, L., Castronova, A. M., Minsker, B., . . . Band, L. E. (2019, January). Development of a participatory Green Infrastructure design, visualization and evaluation system in a cloud supported jupyter notebook computing environment. *Environmental Modelling & Software*, *111*, 121–133. Retrieved 2022-10-11, from https://linkinghub.elsevier.com/retrieve/pii/S1364815217308290 doi: 10.1016/j.envsoft.2018.10.003

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*, 14415–14428.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., . . . Wood, E. F. (2019, August). Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resources Research*, *55*(8), 6499–6516. Retrieved 2021-02-03, from https://onlinelibrary.wiley.com/doi/10.1029/2019WR025287 doi: 10.1029/2019WR025287

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., & Arheimer, B. (2010, June). Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research*, *41*(3-4), 295–319. Retrieved 2022-06-02, from https://iwaponline.com/hr/article/41/3-4/295/822/Development-and-testing-of-the-HYPE-Hydrological doi: 10.2166/nh.2010.007

McMillan, H. K. (2021, January). A review of hydrologic signatures and their applications. *WIREs Water*, *8*(1). Retrieved 2021-04-27, from https://onlinelibrary.wiley.com/doi/10.1002/wat2.1499 doi: 10.1002/wat2.1499

Melsen, L. A., Torfs, P. J. J. F., Uijlenhoet, R., & Teuling, A. J. (2017, March). Comment on “Most computational hydrology is not reproducible, so is it really science?” by Christopher Hutton et al.: REPRODUCING COMPUTATIONAL STUDIES. *Water Resources Research*, *53*(3), 2568–2569. Retrieved 2021-06-18, from http://doi.wiley.com/10.1002/2016WR020208 doi: 10.1002/2016WR020208

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016, December). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, *7*(1), 13603. Retrieved 2021-11-03, from http://www.nature.com/articles/ncomms13603 doi: 10.1038/ncomms13603

Miles, B. (2014). *SMALL-SCALE RESIDENTIAL STORMWATER MANAGEMENT IN URBANIZED WATERSHEDS: A GEOINFORMATICS-DRIVEN ECOHYDROLOGY MODELING APPROACH* (Unpublished doctoral dissertation). A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill, Chapel Hill 2014.

Miles, B., & Band, L. E. (2015). Ecohydrology Models without Borders?: Using Geospatial Web Services in EcohydroLib Workflows in the United States and Australia. In R. Denzer, R. M. Argent, G. Schimak, & J. Hřebíček (Eds.), *Environmental Software Systems. Infrastructures, Services and Applications* (Vol. 448, pp. 311–320). Cham: Springer International Publishing. Retrieved 2021-06-20, from http://link.springer.com/10.1007/978-3-319-15994-2_31 (Series Title: IFIP Advances in Information and Communication Technology) doi: 10.1007/978-3-319-15994-2_31

Miles, B., Leonard, L., & Blodgett, D. (2022). *EcohydroLib.* Retrieved 2022-07-13, from https://github.com/selimnairb/EcohydroLib

Minder, J. R., Mote, P. W., & Lundquist, J. D. (2010, July). Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. *Journal of Geophysical Research*, *115*(D14), D14122. Retrieved 2021-03-17, from http://doi.wiley.com/10.1029/2009JD013493 doi: 10.1029/2009JD013493

Mizukami, N., Clark, M. P., Gharari, S., Kluzek, E., Pan, M., Lin, P., … Yamazaki, D. (2021, June). A Vector-Based River Routing Model for Earth System Models: Parallelization and Global Applications. *Journal of Advances in Modeling Earth Systems*, *13*(6). Retrieved 2021-11-02, from https://onlinelibrary.wiley.com/doi/10.1029/2020MS002434 doi: 10.1029/2020MS002434

Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., … Brekke, L. D. (2016, June). mizuRoute version 1: a river network routing tool for a continental domain water resources applications. *Geoscientific Model Development*, *9*(6), 2223–2238. Retrieved 2021-02-03, from https://gmd.copernicus.org/articles/9/2223/2016/ doi: 10.5194/gmd-9-2223-2016

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019, June). On the choice of calibration metrics for “high-flow” estimation using hydrologic models. *Hydrology and Earth System Sciences*, *23*(6), 2601–2614. Retrieved from https://www.hydrol-earth-syst-sci.net/23/2601/2019/ doi: 10.5194/hess-23-2601-2019

Murphy, A. H. (1988). Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *MONTHLY WEATHER REVIEW*, *116*, 2417–2424. doi: https://doi.org/10.1175/1520-0493(1988)116%3C2417:SSBOTM%3E2.0.CO;2

Nash, J., & Sutcliffe, J. (1970, April). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/

0022169470902556 doi: 10.1016/0022-1694(70)90255-6

NCAR Research Applications Laboratory | RAL. (2021). *Noah-Multiparameterization Land Surface Model (Noah-MP LSM).* Retrieved 2021-11-15, from https://ral.ucar.edu/solutions/products/noah-multiparameterization-land-surface-model-noah-mp-lsm

Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, *19*(2), 101–121. (ISBN: 15351459 (ISSN)) doi: 10.1002/rra.700

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., ... Pellerin, P. (2007, May). Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences*, *11*(4), 1279–1294. Retrieved 2021-04-20, from https://hess.copernicus.org/articles/11/1279/2007/ doi: 10.5194/hess-11-1279-2007

Pushpalatha, R., Perrin, C., Moine, N. L., & Andréassian, V. (2012, February). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, *420-421*, 171–182. (Publisher: Elsevier B.V.) doi: 10.1016/j.jhydrol.2011.11.055

QGIS Development Team. (2021). *QGIS Geographic Information System.* QGIS Association. Retrieved from https://www.qgis.org/

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5), 1–25. doi: 10.1029/2008WR007327

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013, October). Ten Simple Rules for Reproducible Computational Research. *PLoS Computational Biology*, *9*(10), e1003285. Retrieved 2021-07-15, from https://dx.plos.org/10.1371/journal.pcbi.1003285 doi: 10.1371/journal.pcbi.1003285

Sazib, N. S. (2016). *Physically Based Modeling of the Impacts of Climate Change on Streamflow Regime* (Doctoral dissertation, A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill, Utah State University). doi: 10.26076/7BBE-E62D

Sen Gupta, A., Tarboton, D., Hummel, P., Brown, M., & Habib, S. (2015, June). Integration of an energy balance snowmelt model into an open source modeling framework. *Environmental Modelling & Software*, *68*, 205–218. Retrieved 2021-06-19, from https://linkinghub.elsevier.com/retrieve/pii/S1364815215000663 doi: 10.1016/j.envsoft.2015.02.017

Slater, L. J., Thirel, G., Harrigan, S., Delaigue, O., Hurley, A., Khouakhi, A., ... Smith, K. (2019, July). Using R in hydrology: a review of recent developments and future directions. *Hydrology and Earth System Sciences*, *23*(7), 2939–2963. Retrieved 2021-02-17, from https://hess.copernicus.org/articles/23/2939/2019/ doi: 10.5194/hess-23-2939-2019

Stagge, J. H., Rosenberg, D. E., Abdallah, A. M., Akbar, H., Attallah, N. A., & James, R. (2019, March). Assessing data availability and research reproducibility in hydrology and water resources. *Scientific Data*, *6*(1), 190030. Retrieved 2021-02-17, from http://www.nature.com/articles/sdata201930 doi: 10.1038/sdata.2019.30

Stodden, V., & Miguez, S. (2013). Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. *SSRN Electronic Journal*. Retrieved 2021-02-17, from http://www.ssrn.com/abstract=2322276 doi: 10.2139/ssrn.2322276

Sulla-Menashe, D., & Friedl, M. A. (2018). *User Guide to Collection 6 MODIS Land Cover (MCD12Q1 andMCD12C1) Product* (User Guide). Retrieved from https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., & Friedl, M. A. (2019, March). Hierarchical mapping of annual global land cover 2001 to present: The MODIS

Collection 6 Land Cover product. *Remote Sensing of Environment*, *222*, 183–194. Retrieved 2021-06-17, from `https://linkinghub.elsevier.com/retrieve/pii/S0034425718305686` doi: 10.1016/j.rse.2018.12.013

Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., & Iwamoto, H. (2016, June). GENERATION OF THE 30 M-MESH GLOBAL DIGITAL SURFACE MODEL BY ALOS PRISM. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B4*, 157–162. Retrieved 2021-06-12, from `http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B4/157/2016/isprs-archives-XLI-B4-157-2016.pdf` doi: 10.5194/isprsarchives-XLI-B4-157-2016

Tang, G., Clark, M. P., Papalexiou, S. M., Ma, Z., & Hong, Y. (2020, April). Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sensing of Environment*, *240*, 111697. Retrieved 2021-06-17, from `https://linkinghub.elsevier.com/retrieve/pii/S0034425720300663` doi: 10.1016/j.rse.2020.111697

Tarboton, D. G., Horsburgh, J. S., Maidment, D. R., Whiteaker, T., Zaslavsky, I., Piasecki, M., . . . Whitenack, T. (2009). Development of a community hydrologic information system. In (pp. 988–994). Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation.

Tesfa, T. K., Tarboton, D. G., Watson, D. W., Schreuders, K. A., Baker, M. E., & Wallace, R. M. (2011, December). Extraction of hydrological proximity measures from DEMs using parallel processing. *Environmental Modelling & Software*, *26*(12), 1696–1709. Retrieved 2021-11-23, from `https://linkinghub.elsevier.com/retrieve/pii/S1364815211001794` doi: 10.1016/j.envsoft.2011.07.018

Tiesinga, E., Mohr, P. J., Newell, D. B., & Taylor, B. N. (2019). *The 2018 CODATA Recommended Values of the Fundamental Physical Constants.* Gaithersburg, MD 20899: National Institute of Standards and Technology. Retrieved 2021-10-11, from `http://physics.nist.gov/constants`

U.S. Department of Agriculture: Agricultural Research Service (USDA ARS). (2021). *ROSETTA Class Average Hydraulic Parameters.* Retrieved 2021-11-15, from `https://www.ars.usda.gov/pacific-west-area/riverside-ca/agricultural-water-efficiency-and-salinity-research-unit/docs/model/rosetta-class-average-hydraulic-parameters/`

Vanderborght, J., Kasteel, R., Herbst, M., Javaux, M., Thiéry, D., Vanclooster, M., . . . Vereecken, H. (2005). A Set of Analytical Benchmarks to Test Numerical Models of Flow and Transport in Soils. *Vadose Zone Journal*, *4*(1), 206. Retrieved 2021-06-12, from `https://www.soils.org/publications/vzj/abstracts/4/1/0206` doi: 10.2136/vzj2005.0206

Vanderkelen, I., Gharari, S., Mizukami, N., Clark, M. P., Lawrence, D. M., Swenson, S., . . . Thiery, W. (2022, June). Evaluating a reservoir parametrization in the vector-based global routing model mizuRoute (v2.0.1) for Earth system model coupling. *Geoscientific Model Development*, *15*(10), 4163–4192. Retrieved 2022-07-25, from `https://gmd.copernicus.org/articles/15/4163/2022/` doi: 10.5194/gmd-15-4163-2022

Vorobevskii, I. (2022, May). *hydrovorobey/Global_brook90: zonedo.* Zenodo. Retrieved 2022-07-13, from `https://zenodo.org/record/6535132` doi: 10.5281/ZENODO.6535132

Vorobevskii, I., Kronenberg, R., & Bernhofer, C. (2020, July). Global BROOK90 R Package: An Automatic Framework to Simulate the Water Balance at Any Location. *Water*, *12*(7), 2037. Retrieved 2021-04-25, from `https://www.mdpi.com/2073-4441/12/7/2037` doi: 10.3390/w12072037

Wallace, J. M., & Hobbs, P. V. (2006). *Atmospheric science: an introductory sur-*

vey (2nd ed ed.) (No. v. 92). Amsterdam ; Boston: Elsevier Academic Press. (OCLC: ocm62421169)

Weiler, M., & Beven, K. (2015, September). Do we need a Community Hydrological Model?: DO WE NEED A COMMUNITY HYDROLOGICAL MODEL? *Water Resources Research*, *51*(9), 7777–7784. Retrieved 2021-06-19, from http://doi.wiley.com/10.1002/2014WR016731 doi: 10.1002/2014WR016731

Wi, S., Ray, P., Demaria, E. M., Steinschneider, S., & Brown, C. (2017, December). A user-friendly software package for VIC hydrologic model development. *Environmental Modelling & Software*, *98*, 35–53. Retrieved 2022-10-14, from https://linkinghub.elsevier.com/retrieve/pii/S1364815216308131 doi: 10.1016/j.envsoft.2017.09.006

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016, December). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. Retrieved 2021-02-17, from http://www.nature.com/articles/sdata201618 doi: 10.1038/sdata.2016.18

World Meteorological Organization. (2017). *Guide to the Implementation of Quality Management Systems for National Meteorological and Hydrological Services and Other Relevant Service Providers* (Tech. Rep. No. WMO-No. 1100).

Xu, X., Frey, S. K., Boluwade, A., Erler, A. R., Khader, O., Lapen, D. R., & Sudicky, E. (2019, August). Evaluation of variability among different precipitation products in the Northern Great Plains. *Journal of Hydrology: Regional Studies*, *24*, 100608. Retrieved 2021-06-17, from https://linkinghub.elsevier.com/retrieve/pii/S2214581818303410 doi: 10.1016/j.ejrh.2019.100608

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019, June). MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resources Research*, *55*(6), 5053–5073. Retrieved 2021-02-03, from https://onlinelibrary.wiley.com/doi/10.1029/2019WR024873 doi: 10.1029/2019WR024873

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., ... Bates, P. D. (2017, June). A high-accuracy map of global terrain elevations: Accurate Global Terrain Elevation map. *Geophysical Research Letters*, *44*(11), 5844–5853. Retrieved 2021-06-12, from http://doi.wiley.com/10.1002/2017GL072874 doi: 10.1002/2017GL072874