



eWaterCycle: a hyper-resolution global hydrological model for river discharge forecasts made from open source pre-existing components

Rolf Hut¹, Niels Drost², Maarten van Meersbergen², Edwin Sutanudjaja³, Marc Bierkens^{3,4}, and Nick van de Giesen¹

¹Chair of Water Resources Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

²Netherlands eScience Center, Amsterdam, The Netherlands

³Department of Physical Geography, Utrecht University, Utrecht, The Netherlands

⁴Deltares, Utrecht, The Netherlands

Correspondence to: Rolf Hut (r.w.hut@tudelft.nl)

Abstract. eWaterCycle is an open source hyperresolution (10km x 10km) global hydrological forecasting framework that runs an ensemble of hydrological models. Forced with a weather forecast ensemble, it predicts river discharge and river discharge uncertainty nine days ahead. Daily satellite soil moisture observations are assimilated into the state of the model ensemble using an Ensemble Kalman Filter. We demonstrate that it is feasible to build such a system using pre-existing, open source, components that communicate through standard interfaces. The PCRGLOBWB2.0 (van Beek et al., 2011; Sutanudjaja et al., 2014) model is used to model hydrology globally, forced with GFS (Kanamitsu, 1989; Kanamitsu et al., 1991; Moorthi et al., 2001) weather forecast. The operational soil moisture product from the HSAF (Drusch et al., 2009; De Rosnay et al., 2011) service is assimilated into the model ensemble using OpenDA (Velzen et al., 2016), a data assimilation framework. Output of the model ensemble is presented in a Cesium (Analytical Graphics, 2011) based visualization. All communication between framework components is through standard file types (NetCDF)(Rew and Davis, 1990) and services (Web Map Service) (de La Beaujardiere, 2006). Communication between model and data assimilation framework is through the Basic Model Interface (BMI)(Peckham et al., 2013). The forecasts is available at forecast.ewatercycle.org. By using standard open interfaces, the different components of the model can be replaced with relative ease, facilitating future model comparison studies without the need of extensive Computer Science support. This makes eWaterCycle, in addition to an operational forecasting model, a testbed environment where the impact of different model structures, input sources and / or data assimilation schemes can easily be studied. Setup instructions to run the eWaterCycle project on local hardware are provided, allowing the hydrological community to build on this open source framework.

1 Introduction

The need for a “hyper-resolution” global hydrological model has been expressed by Wood *et al* Wood et al. (2011) as a “grand challenge” in hydrology. Global modeling of hydrology has improved in resolution to grids as small as 0.5° by 0.5° (Bierkens et al., 2015). The review by Bierkens (2015) of state of the art hydrological global models shows that global hydrological



models are used for climate sensitivity analysis (Nijssen et al., 2001), and for long term global change analysis (Prudhomme et al., 2014), but rarely for operational forecasting of river discharge, with Alfieri et al. (2013) a notable exemption.

On basin or continental scale, many hydrological ensemble prediction systems (HEPS) do exist (Thiemig et al., 2015; Kauffeldt et al., 2016; Emerton et al., 2016). Most HEPS take an ensemble of weather forecasts to force an ensemble of hydrological models. Data assimilation of observed states of the hydrological model is not a common feature of either global, or continental HEPS (Thiemig et al., 2015). For optimal estimation of hydrological variables, it is important to assimilate relevant observations (Lopez Lopez et al., 2015). Recent work by Kurtz et al. (2016) has shown that it is feasible to run very large scale, global HEPS on high performing computing infrastructure (super computers). This does, as Kurtz et al. (2016) states, require detailed knowledge on how to couple the data assimilation scheme to the model. Close collaboration between hydrologists and computer scientists is, therefore, required.

The grand challenge as laid out by Wood et al. (2011) specifically argues that focus of the hydrological community should be on combining, or coupling, models from different subfields: surface and subsurface hydrology, land and atmosphere, water quantity and water quality. To achieve this, different models should be coupled using open interfaces, allowing researchers from different Earth science disciplines to work together and with computer scientists without having to know intimately each other's field. Different open interfaces to communicate between model(components) have recently been developed (Donchyts et al., 2010; Peckham et al., 2013). It should be possible that these models are written in different programming languages. For these coupled models to have societal value within the context of an HEPS, they should be run as an ensemble forecast, and use data assimilation. Data assimilation scheme and models should communicate through open interfaces, decoupling the need for computer scientists and geoscientists to have intimate knowledge of each other's field, while still allowing geoscientists to reap the benefit of advances in the computer sciences.

In this paper we demonstrate that it is technically feasible to run a global hydrological ensemble forecasting system on high performance infrastructure without deep coupling. The resulting forecasts are available on forecast.ewatercycle.org. In our "eWaterCycle" system the hydrological model is coupled to the data assimilation scheme using an open interface. All interactions between the components: scheduler, preprocessor, data assimilation scheme, hydrological model and visualization are either through standard file formats or through open interfaces. The source code for all components of the eWaterCycle model is available in open source GitHub repositories, allowing anyone with access to the proper hardware to setup and run their own global hydrological ensemble prediction system and, more importantly, to improve upon the system provided and make steps, as a community, to achieve the goals set out in the grand challenge.

2 eWaterCycle architecture

The eWaterCycle system consists of 5 components:

- orchestrator: Cylc
- pre- and post- processor: CDO



- data assimilation: OpenDA
- cross-language model coupling: Thrift
- model: PCR-GLOBWB 2.0
- visualization: Cesium and ncWMS

5 Furthermore, two input data sources are used: the GFS weather forecast and the EUMETSAT Soil moisture product. Figure 1 shows an overview of how the components and input sources are connected by the interfaces. All interfaces are either standard file formats (NetCDF and GRIB) or open interfaces (BMI and WMS). Tables 1 and 2 give an overview of all the components, input sources and interfaces, including where they can be found online. In the sections below every interface, input source and component will be described.

10 3 Input data sources

The eWaterCycle system has two main input sources that change daily: The GFS weather prediction and the EUMETSAT soil moisture product. Below we describe how these inputs are processed. The model does require more inputs, such as a Digital Elevation Model (DEM), but these do not change between model runs (see section 4.5 for details)

3.1 GFS

15 The U.S. National Center for Environmental Prediction (NCEP) runs a deterministic weather prediction, the Global Forecast System (GFS)(Kanamitsu, 1989; Kanamitsu et al., 1991; Moorthi et al., 2001), as well as an ensemble of weather predictions, the Global Ensemble Forecast System (GEFS)(Toth et al., 2006; Zhu, 2005). The GFS has a higher spatial resolution (0.25° versus 1°) and is run for a longer forecasting time (16 days versus 8 days). The GEFS is run with 21 ensemble members. Other Numerical Weather Predictions are available, such as those provided by the European Centre for Medium-Range Weather
20 Forecasts (ECMWF). eWatercycle uses the NCEP model because the data is available for anyone to download for free. A particular interesting feature of GFS and GEFS is the NOMAD online filter that can be used to extract and download only the variables of interest (Rutledge et al., 2006). For the PCR-GLOBWB 2.0 model these variables are precipitation and temperature. In eWaterCycle we create an input ensemble that is a hybrid of GFS and GEFS by superimposing the deviation of the GEFS ensemble mean on GFS. The script that downloads the data and produces the input ensemble for the PCRglobWB 2.0 model
25 is discussed in section 4.2.

3.2 ECWMF SM-DAS-2

The SM-DAS-2 product by ECWMF (Drusch et al., 2009; De Rosnay et al., 2011), available through the HSAF service, was chosen to be assimilated in eWaterCycle. In (Albergel et al., 2012) the data from SM-DAS-2 were compared to ground based



measurements of soil moisture. SM-DAS-2 is technically not 100% an observation, but rather the result of assimilating observations from the ASCAT satellite into the ECWMF land-surface model. From a computational point of view, SM-DAS-2 provides a sort of “worst case” scenario for eWaterCycle, as it is a large global high resolution observation dataset. By assimilating the SM-DAS-2 product into the PCRGlobWB 2.0 model it is shown that assimilating large datasets into an hydrological ensemble prediction system using open source components is feasible.

4 Components

The five components that make up the eWaterCycle system are described below. For each component the history and original purpose is briefly sketched followed by a description of the function the component has within the eWatercycle system.

4.1 Orchestration: Cylc

Cylc is a workflow management system designed especially for running cyclic tasks such as our forecast, where each next forecast relies on results from the previous forecast (Oliver, 2008). Cylc tracks which tasks are able to run and automatically submits these to a cluster or supercomputer for processing. How then to perform a task is not up to Cylc. Instead user-provided commands or scripts are run. This makes it easy to integrate Cylc into an existing system, as most existing scripts can be used as is. Cylc was originally developed at the New Zealand National Institute of Water and Atmospheric Research (NIWA), and is now used at the UK Met Office and others.

In the eWaterCycle system, we use Cylc to automatically run our forecast, from downloading necessary weather data, to uploading our final result to the web application server. See Figure 3 for a graphical representation of our workflow by the Cylc monitoring application.

4.2 Pre- and Post- processor: CDO

The Climate Data Operators (CDO, 2015) are a set of command line operators that work on gridded data in NetCDF or GRIB files (see section 5.1 below). The CDO includes basic operations such as adding similar variables from different files with identical grids, but also more complex operators to interpolate gridded data to a new grid using different interpolation schemes (Kriging, nearest neighbor, bilinear, etc.). Multiple CDO operators can be linked into a single command, which will be executed efficiently in memory without saving the intermediate results. Multiple CDO commands can be combined together in bash-scripts, but integration in Python and other computing languages is also possible.

In eWaterCycle, CDO is used for pre-processing the weather forecast data into NetCDF files that the PCR-GLOBWB model (see section 4.5) uses as input. Furthermore, CDO is used to gather the outputs of the ensemble members, which each produce a single NetCDF file, into a single file that includes ensemble mean and standard deviation as input for the visualization (see section 4.6). The bash scripts used in eWaterCycle are available on GitHub, see table 2.



4.3 Data Assimilation: OpenDA

Data assimilation (DA), the science of incorporating state or output observations in continuously running forecasting models, is often built as an add-on to an existing model. However, implementing DA techniques both correctly and fast, in terms of computing time required, is no trivial matter (Kurtz et al., 2016). Also, research into improved DA schemes is hindered by this practice, as the improved scheme needs to be implemented for each model on which it needs to be tested. To solve this problem, OpenDA was developed (Velzen et al., 2016). OpenDA is a generic model independent data assimilation framework. Once an existing model has been coupled to OpenDA, it immediately has DA capabilities.

In the eWaterCycle system, we use OpenDA to assimilate soil moisture observations into our model using an Ensemble Kalman filter (EnKF) (Evensen, 2003). OpenDA creates 20 ensemble members of our model. The initial state of each member (at $T = -24$ hours) is carried over from the previous forecast run. Each member is forced using input from each GEFS/GFS ensemble member as described in section 3.1. At $T=0$ the observations are assimilated into the model. As the amount of observations is large (about 50.000 pixels with observations), we use the sequential EnKF implementation as present in the OpenDA framework. This assimilates the observations one by one. This uses less memory, at the cost of an increased runtime and the additional assumption that all observations are independent. Having to use this workflow highlights that EnKF has scaling issues when large ensembles or large numbers of observations, such as satellite data, are assimilated. Although efficient memory variants of the EnKF exist (Hut et al., 2015), these have not yet been implemented in standard DA frameworks such as OpenDA. After the DA step, OpenDA runs our model for the remainder of the forcing to create a forecast. The state at $T = 0$ after DA is also kept for the next forecast.

We have extended OpenDA with a model independent coupling based on Thrift and BMI. This *BMI Bridge* is capable of running any BMI compliant model, and can run multiple ensemble members in parallel on multiple machines. This version of OpenDA is currently available as part of the eWaterCycle repository on Github, see table 1. See sections 4.4 and 5.2 for more details.

Figure 2 gives an overview of the time dependencies in the eWaterCycle system. It shows how each day, observations are assimilated in the forecast for that day. The resulting assimilated state is the start ($T = 0$) for the next model run, resulting in a new forecast.

4.4 Model Coupling: Thrift

Apache Thrift is a system for creating cross-language services (Slee et al., 2007). Thrift was originally developed by Facebook to couple services implemented using different programming languages, but is now hosted as an open source project by the Apache Foundation.

Thrift couples services by the use of a Remote Procedure Call (RPC) model. A programmer defines an interface, called service, in a language neutral fashion using a Interface Definition Language (IDL) file. From this file, Thrift then generates a client and server application for this service in a range of different programming languages. A service could, for instance, be implemented in Go, while a client is implemented in C++.



In the eWaterCycle system we use Thrift to couple our model (see Section 4.5) to OpenDA. Our model has been implemented in Python while OpenDA has been implemented in Java. Thrift allows us to bridge this language gap with very little effort. Furthermore, instead of implementing model specific functions, we use the BMI standard (see Section 5.2). This, in effect, creates a model independent bridge capable of coupling any model to OpenDA, written in any of the over 20 programming languages supported by Thrift.

4.5 Global Hydrological Model: PCR-GLOBWB 2.0

The global hydrological model PCR-GLOBWB (van Beek et al. (2011); Sutanudjaja et al. (2014), Sutanudjaja et al., in prep.) simulates continuous fields of fluxes and stocks in various water storage components, such as soil moisture, surface water, snow and groundwater on spatial and temporal scales, with spatial resolution of 5 arcminutes, or approximately 10 km at the equator.

PCR-GLOBWB computes, for each cell, the vertical water exchanges between soil layers and between top soil layer and atmosphere. Processes represented in the model are rainfall and snowmelt, percolation and capillary rise, as well as evaporation and transpiration fluxes. Under the soil layer is a groundwater store. Groundwater is fed by net groundwater recharge. The process of groundwater feeding baseflow is modelled as a linear reservoir. This behaviour is exempt from direct influence of evaporation and transpiration fluxes. Capillary rise from the active groundwater store depends on the simulated groundwater storage, the soil moisture deficit and the unsaturated hydraulic conductivity. Fluxes are computed for various land cover types by including sub-grid variations. Differences in topography, vegetation phenology, and soil properties are included. PCR-GLOBWB includes a physically-based scheme for infiltration and runoff, resulting in direct runoff, interflow, as well as groundwater baseflow and recharge. River discharge is calculated by accumulating and routing all runoff components along the drainage network. For more details about the PCR-GLOBWB model, including its parameterization, we refer to its technical reports and other relevant papers (Van Beek and Bierkens, 2009; Van Beek, 2008; Sutanudjaja et al., 2011, 2014). PCR-GLOBWB is used and described in other papers such as the recent Tangdamrongsub et al. (2016a, b).

Over time, several features of PCR-GLOBWB have been implemented since its first application (van Beek et al., 2011). One new feature concerns two-way interactions between water availability and abstraction, including simulating variations in irrigation and other water demands, such as livestock, industrial and domestic water demands, and allocation of available surface water and groundwater resources (see de Graaf et al. (2014); Wada and Bierkens (2014); Erkens and Sutanudjaja (2015)). Combined, these features determine how the demands are allocated to the available water resources and where the return flows of unconsumed water go. The new features lead to the simulation of essential feedbacks, particularly in irrigated areas where water demand is large. Together with many other new developments, these advanced and integrated water demand and allocation modules are being consolidated in the new version of PCR-GLOBWB 2.0. This new version fully integrates global hydrology and water resource models and will be implemented at a spatial resolution of 5 arc minutes (Sutanudjaja et al., in prep.).

In this research, an alpha release of the 5 arc minutes version of PCR-GLOBWB 2.0 is used. In this alpha version all calculations are executed but irrigation and other human influences are situated within single pixels, as was the case in the



original one degree resolution model. Considering within pixel human influence is unrealistic at 5 arc minutes resolution and greatly reduced the accuracy of the simulated discharge. Furthermore, the irrigation routine requires historical data which is available when running PCR-GLOBWB for historical analyses. We like to stress that the focus of this paper is not on the accuracy of the prediction but rather on the numerical and computational feasibility to create and run a hyper-resolution global hydrological model for river discharge forecasts made from pre-existing open source components.

Some other currently ongoing developments regarding PCR-GLOBWB include a coupling to lateral groundwater flow modelling for simulating spatio-temporal groundwater head dynamics ((Sutanudjaja et al., 2011, 2014; de Graaf et al., 2014; Sutanudjaja et al., 2014)). PCR-GLOBWB can be extended with an external surface water and flood routing scheme (Winsemius et al., 2013) and a physically-based freshwater surface temperature model (van Beek et al., 2012).

10 4.6 Visualization: Cesium and ncWMS

ncWMS is a server side application capable of reading any NetCDF file written using the Climate and Forecasting (CF) conventions (Blower et al., 2013), and making the data available as a Web Map Service(WMS). ncWMS automatically determines available variables in a file, and creates maps colored according to map data and a user selected color scale. See Section 5.3 for more information on WMS.

15 Cesium is a Javascript 3D virtual Globe library (Analytical Graphics, 2011). It uses WebGL for rendering, which makes it very fast, and it is capable of displaying a wide variety of data types such as vectors, 3D models, and 2D maps.

To visualize the result of the forecast we have created Cesium-ncWMS, a web application based on ncWMS and Cesium. The forecast results are automatically uploaded to our web server running ncWMS. In turn, the web application can be used to change the settings for color maps and displayed data. The server uses the settings provided by the web application, together with the data in NetCDF to provide WMS image tiles, time series data and legend graphics to the Cesium-NcWMS web application. The user can simultaneously zoom in to the very high resolution forecast results anywhere on the world, and get time series data for any point on the globe. See figure 4 for a screenshot.

5 Interfaces

25 Interfaces are the glue between the components of the eWaterCycle system, which contains two data interfaces (BMI, WMS) and two data file formats or data abstractions (GRIB, NetCDF). Below the reasons for choosing the interfaces are explained.

5.1 GRIB and NetCDF

Ideally different components should exchange data through interfaces if at all possible, reducing the need to write and read large files. However, many components still use files, both as inputs or outputs, as the main way of communicating with other components. In eWaterCycle two filetypes are used to share data between components: GRIB and NetCDF.

30 GRIB (GRIdded Binary or General Regularly-distributed Information in Binary form) is a data format used mostly to store weather data. The format was standardized by the World Meteorological Organization (WMO) (Dey and others, 2007). It has



been designed to store fields of gridded variables and has been the de-facto standard fileformat for numerical weather prediction data for decades.

The Network Common Data Form (NetCDF) is not so much a fileformat, as it is a data abstraction (Rew and Davis, 1990). NetCDF data can live on a server, which can be queried, or in a file, which can also be queried. In NetCDF, metadata are stored together with the data. The implementation of NetCDF makes it easy to query a NetCDF file and only receive the part of the file that is of interest. Variables, including their units, have to be precisely documented. Different fields of science have set standards on how to store the units relevant to their field. For geosciences, these are encoded in the Climate and Forecasting (CF) conventions (Eaton et al., 2003). The NetCDF software project is hosted by the University Corporation for Atmospheric Research (UCAR).

10 5.2 BMI

BMI (Basic Model Interface) is a standard for interfacing models in geosciences (Peckham et al., 2013). As the name implies, BMI asks very little of a model. BMI is explicitly designed to be easy to implement by modellers. BMI requires a model to implement a small number of functions, most usually already available in some form in a model. Examples include a function to initialize the model, get and set the state of variables, run the model one or more steps, and get some metadata on the model, such as what variables are available, what units do these have, etc. Bindings for different languages, including Fortran, Python, Java, C, and C++, are available. All in all, implementing a BMI interface for a model should take no more than a day or so, far less than most other interfaces and frameworks.

In the eWaterCycle project, we use BMI as the interface between OpenDA and PCR-GlobWB. As these are written in different languages, we use Thrift to couple the two. The use of BMI also allows reuse of components in other projects. The OpenDA-BMI model bridge is already being re-used in the OpenStreams project (Weerts 2016, in preperation).

5.3 WMS

Web Map Service (WMS) is a standard for retrieving map based image data from a remote service on the Internet (de La Beaujardiere, 2006). WMS allows a user to specify which part of the map is needed, which variables and styles need to be displayed, etc. WMS is implemented using a wide variety of backends, ranging from simple files to large database systems. WMS is widely supported in client applications, including both desktop applications, and 2D and 3D web applications.

In the eWaterCycle project we use WMS as the intermediary protocol between our visualization and the data server. As a proof of concept, we also created a second visualization making use of WMS. This was eWaterLeaf (Drost, 2016), a simple 2D map visualization showing the ensemble mean of the forecast of discharge based on Leaflet (Agafonkin, 2015). Both visualizations are re-usable in other projects. Our Cesium-ncWMS visualization makes use of some extensions to WMS present in ncWMS, for instance to support a logarithmic value scale. To use Cesium-ncWMS with a different data server would therefore require some work.



6 eWaterCycle hardware requirements

Although all software used and created in the eWaterCycle is available as open source, running the model also requires capable hardware. As the eWaterCycle forecast runs 20 ensemble members simultaneously, total storage, memory, and cpu requirements are dominated by these. In the setup used in the eWaterCycle system, our model produces around 5Gb of output, as well as 2Gb of state files needed for the next iteration of the forecast. For the entire ensemble, on the order of 150Gb storage is required per forecast. This can be reduced by producing less variables in the output, or increased by keeping more intermediate data for verifying the correctness of the model and data assimilation.

As for memory and cpu requirements, each model instance requires 20Gb of memory, with short peaks of about 40 Gb. We have implemented the capability to spread the ensemble members over multiple machines. For computing the live forecast we use 3 machines (nodes) of the Dutch Cartesius supercomputer with 256Gb memory each.

The website platform we use has more modest requirements. We use a t2.medium instance in the Amazon EC2 Cloud to host the website and data server. This has 2 (virtual) cores and 4Gb of memory. When under heavy load, EC2 will automatically start another instance of the service, and load balance requests between the two. The website can be visited at forecast.ewatercycle.org (van Meersbergen et al., 2015).

7 Preliminary validation and opportunities for further development

The focus of this paper is on the development and technology of the eWaterCycle system and not on its accuracy. A thorough statistical evaluation and error attribution of the output of eWaterCycle will be the focus of a different paper. A brief preliminary validation is provided here to give the readership a first glance at the output of eWaterCycle and its general validity.

Four discharge stations from the Global Runoff Database (GRDB) as collected by the Global Runoff Data Center (GRDC in the Bundesanstalt für Gewässerkunde, 2015) were selected. The selection was based on the criteria data availability and global coverage. Only stations for which daily discharge measurements were available for a significant part of 2015 were selected. Geographical coverage was optimized by choosing rivers on four continents, Europe, North America, South America and Africa. No Asian rivers met our criteria. See table 3 for meta-data on the selected stations.

For the locations of the four stations, the discharge as predicted by eWatercycle has been received from the eWaterCycle archive for every forecast made in the period 2015-04-01, when we first started eWaterCycle, until 2015-12-31, the last day for which discharge is available in the GRDC database at the time of writing. From each of these forecasts, the first and the last forecast day is selected. The four plots in figure 5 show the measured discharge at the stations, as well as the predicted ensemble for the first day and the ninth of the forecast of eWaterCycle. The Nash-Sutcliffe coefficient between the ensemble mean and the observed discharges for both the first day forecast as well as the ninth day forecast are provided in table 3.

Figure 5 and table 3 clearly show that the current alpha version of the model does not forecast river discharge too well, as was to be expected from the description of the model in section 4.5. Especially rivers with large human influence, such as the Colorado, perform rather poorly. The forecasts for the larger rivers (Zambezi and Colorado) slowly converge towards realistic discharge values, indicating that a longer spin-up is required for PCR-GLOBWB 2.0. Figure 5 and table 3 clearly show that for



these four rivers the output of eWaterCycle has room for improvement. Future work will upgrade the model from the current used alpha version. The modular nature of eWaterCycle, as explained in this paper, will facilitate updating the model without having to re-build the entire system.

8 Conclusions

5 We demonstrated that it is feasible to build a hydrological ensemble prediction system (HEPS) based on previously available open source components communicating through open interfaces. The eWaterCycle system runs an ensemble of the PCR-GlobWB 2.0 model. Open DA is used to assimilate satellite observations of soil moisture into the model state. The forecasts are communicated via a visualization based on Cesium and is available online at forecast.ewatercycle.org. By building eWaterCycle from pre-existing open source components and running it on existing high performance hardware we prove that HEPS
10 need not necessarily be built from scratch, as is currently usually the case (Emerton et al., 2016; Kurtz et al., 2016). By separating the different components that make up a HEPS and communicating through standard interfaces it is easier to focus development and innovations on particular components without having to redesign the entire system.

This paper focused on presenting the components of the eWaterCycle system, and made the case that it is currently doable for anyone with access to high performance computing power to setup their own version of eWaterCycle. This paper did
15 intentionally not focus on the accuracy of the predictions that the eWaterCycle system makes, which will be the focus of a future paper in a hydrological journal.

Although all components used in eWaterCycle are open source, not all components initially used open interfaces. Within eWaterCycle a BMI interface was added to Open DA to separate the data assimilation software from the model software. The BMI-thrift bridge, developed in eWaterCycle and first presented in this paper, makes it possible to easily switch models, even
20 if those models are written in different programming languages.

As shown by the software installation overview provided in Appendix A. eWaterCycle can be installed by anyone following the steps and with access to high performance computing. However, we acknowledge that although we have proven that a HEPS can be build from open source components by anyone, it still requires at least some affiliation with computer science. In future work we will focus on improving the eWaterCycle system in such a way that hydrologists without any high performance
25 computing experience can easily install the eWaterCycle system and run their own model within it. An expert in observations should be able to upload relevant observations to be assimilated into eWaterCycle without having to understand Computer Science needed to run OpenDA. Similarly, a data assimilation expert should be able to improve the data assimilation algorithm without having to worry about where exactly on the supercomputer the state is stored in the model. We aim to realize this by continuing to make use of standard interfaces. This will help to generate hydrological forecasts based on the best available data
30 assimilation schemes, the best available hydrological models and the best available observations.



9 Code availability

All code is made available for the public. See table 2 for the locations and DOIs of the different components of the eWaterCycle system.

10 Author contributions

- 5 N. Drost and E. Sutanudjaja adapted the PCRGLOBWB model from version 1.0 to 2.0. N. Drost and R. Hut built the central eWaterCycle workflow. N. Drost and R. Hut, together with A. Kockx from Deltares, implemented OpenDA. M. van Meersbergen built the visualisation on top of Cesium. R. Hut and N. Drost wrote this article. N. van de Giesen and M. Bierkens were research PIs, contributed to discussions, conceptualizations, and article write up.

Appendix A: Software Installation Overview

- 10 This section provides a high-level overview of all the software required to run the eWaterCycle forecast. It is deliberately scarce on details, as these get outdated quickly. A more detailed installation guide is included with the eWaterCycle main forecast software (Drost and Hut, 2016).

A1 Generic Prerequisites

- 15 eWaterCycle is a system build from pre-existing open source components. To run the eWaterCycle system, each of these components will have to be installed on your local hardware. First, all pre-requisite software and libraries need to be installed. Most are available as standard packages in Linux distributions. Some need to be downloaded from the website of the software

List of prerequisites:

- Python (<https://www.python.org>)
- Java (<https://java.com/en/download/>)
- 20 – GDAL (<http://www.gdal.org/>)
- CDO (<https://code.zmaw.de/projects/cdo/wiki/cdo>)
- NCL (<http://www.ncl.ucar.edu/>)
- NCO (<http://nco.sourceforge.net>)
- PCRaster (<http://pcraster.geo.uu.nl>)
- 25 – Cylc (<https://cylc.github.io/cylc/>)



A2 PCRGlobWB-2.0

eWaterCycle uses a eWaterCycle specific version of PCRGlobWB, with added support for the OpenDA BMI-Thrift-Bridge. This can be downloaded from the eWaterCycle GitHub page the <https://github.com/ewatercycle>.

In addition to the model, a dataset containing parameters of the model is also required. The most recent version can be
5 obtained from author Sutanudjaja.

A3 OpenDA

eWaterCycle makes use of OpenDA as a data assimilation framework. As part of the eWaterCycle project we contributed the new generic BMI model interface to OpenDA, and this is now part of the standard OpenDA codebase. See the OpenDA website <http://openda.org> for more information and installation instructions.

10 Since eWaterCycle makes use of newly developed features in OpenDA, at the time of publishing this article, it requires the development version of OpenDA. For more details see the installation guide included with the eWaterCycle software.

A4 Initial conditions

Like most forecasts, the eWaterCycle system requires some state from the previous forecast to run. The original initial conditions for the eWaterCycle forecast were created by using output of the (deterministic) PCRGlobWB model after running this
15 for a few years using archived GFS forecast data, and including spinup.

To start running the eWaterCycle system, the state of the last few forecasts (both high and low resolution) can be downloaded from our website.

A5 Running the forecast

The eWaterCycle forecast is ran using Cylc. We created a so-called Suite that is as self contained as possible. Once all the
20 needed software and data is available, running the forecast should be relatively straightforward by running the provided suite. The location of the data and software used, as well as on which machines to run the model on, can be set using a simple settings file.

Running the high resolution model requires a lot of memory and cpu power. The main bottleneck for running larger models is memory. Thus, the suite also includes configurations for a reduced resolution version of the model, already requiring about
25 25Gb memory. This reduced version should run on a modest workstation.

Acknowledgements. This research is part of the eWaterCycle project. This work is in part funded by the research programme of the Netherlands Organisation for Scientific Research (NWO) under project number MP-272-13. This work is in part funded by the Netherlands eScience Center (www.nllesc.nl).

The authors would like to thank Arno Kockx and Steff Hummel from Deltares for their support in working with OpenDA.

30 The eWaterCycle system is run on the Cartesius Supercomputer that is operated by SURFSara.

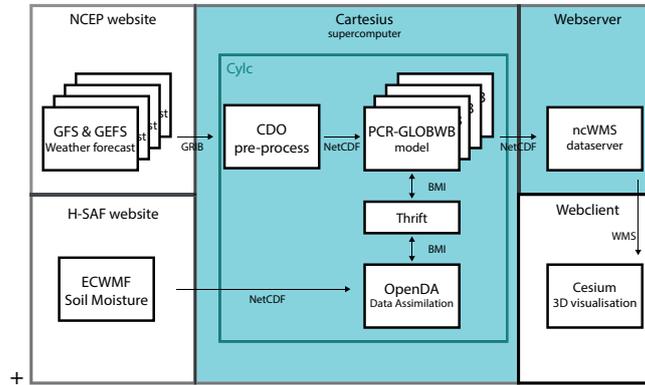


Figure 1. Overview of the components, inputs and interfaces used in the “eWaterCycle” system. All components are open source. See table 1 for an overview of the repositories where the components can be found. The two input sources are freely available online. All interfaces between the components are either standard file formats or standard interfaces. The background boxes indicate the hardware on which the components are running.

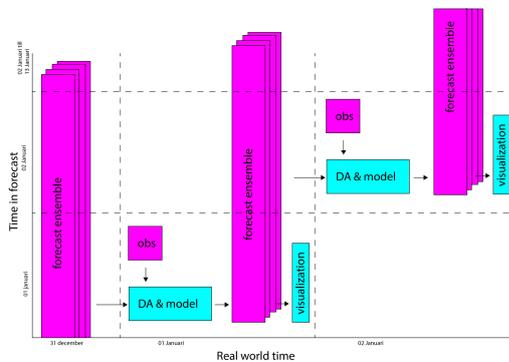


Figure 2. Graphical illustration of daily cycle of running the eWaterCycle forecast. Each day, "Yesterday's" forecast is combined with new observations to estimate the best starting conditions for a new forecast run.

References

Agafonkin, V.: Leaflet — an open-source JavaScript library for interactive maps, <http://leafletjs.com/>, 2015.

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., and Wagner, W.: Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, *Remote Sensing of Environment*, 118, 215–226, doi:10.1016/j.rse.2011.11.017, <http://www.sciencedirect.com/science/article/pii/S0034425711004147>, 2012.

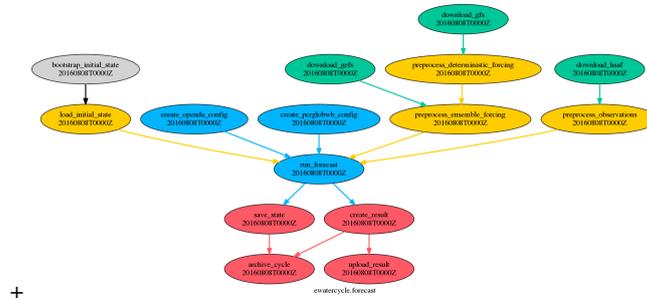


Figure 3. Graph of the Cylc suite that runs the eWaterCycle system. Each node in the graph represents a command that Cylc executes: a script, program. Each edge in the graph represents a dependency between parts of the system. The entire graph represents a single execution of the forecasting system.

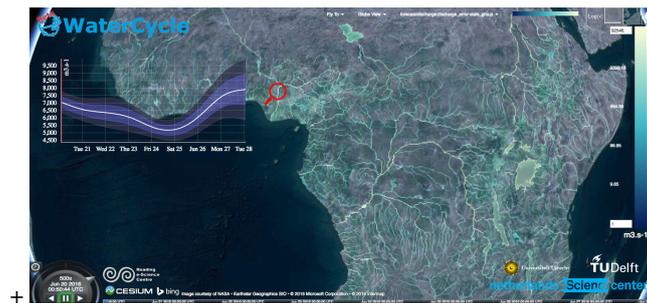


Figure 4. Screenshot of the eWaterCycle Forecast Website. This visualization was created with Cesium-NcWMS, and shows the forecast for a certain day on the map as well as a graph showing the forecasted value over time for a single point (the graph on the top left). The graph also shows the uncertainty of the forecast in the form of a plume, where the width of the plume corresponds to the standard deviation of the forecast ensemble.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, *Hydrol. Earth Syst. Sci.*, 17, 1161–1175, doi:10.5194/hess-17-1161-2013, <http://www.hydrol-earth-syst-sci.net/17/1161/2013/>, 2013.

Analytical Graphics, I.: Cesium - WebGL Virtual Globe and Map Engine, <https://cesiumjs.org/>, 2011.

- 5 Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, *Water Resources Research*, 51, 4923–4947, doi:10.1002/2015WR017173, <http://onlinelibrary.wiley.com/doi/10.1002/2015WR017173/abstract>, 2015.
- Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, *Hydrological Processes*, 29, 310–320, doi:10.1002/hyp.10391, <http://onlinelibrary.wiley.com/doi/10.1002/hyp.10391/abstract>, 2015.
- 10 Blower, J. D., Gemmill, A. L., Griffiths, G. H., Haines, K., Santokhee, A., and Yang, X.: A Web Map Service implementation for the visualization of multidimensional gridded environmental data, *Environmental Modelling & Software*, 47, 218–224, doi:10.1016/j.envsoft.2013.04.002, <http://www.sciencedirect.com/science/article/pii/S1364815213000947>, 2013.

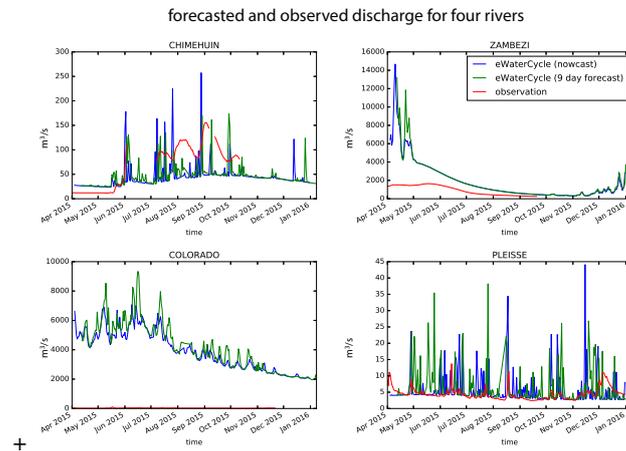


Figure 5. Measured and predicted discharge for four rivers. Both the nowcast and the 9-day forecast are shown. Nash Sutcliff coefficients and station metadata are provided in table 3. As was to be expected from a model under development that does not correctly include withdrawal of water by humans, the model under performs at this moment. In a future paper we will replace the current model with an improved version and present an analyses of the predictive performance of eWaterCycle based on all available river discharge observations worldwide.

CDO: Climate Data Operators, <http://www.mpimet.mpg.de/cdo>, 2015.

de Graaf, I. E. M., van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Dynamic attribution of global water demand to surface water and groundwater resources: Effects of abstractions and return flows on river discharges, *Advances in Water Resources*, 64, 21–33, doi:10.1016/j.advwatres.2013.12.002, <http://www.sciencedirect.com/science/article/pii/S0309170813002467>, 2014.

5 de La Beaujardiere, J.: OpenGIS® web map server implementation specification, Open Geospatial Consortium Inc., OGC, pp. 06–042, 2006.
De Rosnay, P., Drusch, M., Balsamo, G., Isaksen, L., and Albergel, C.: Extended Kalman Filter soil moisture analysis in the IFS, *ECMWF Spring Newsletter*, 127, 12–16, 2011.

Dey, C. and others: Guide to the WMO table driven code form used for the representation and exchange of regularly spaced data in binary form: FM 92 GRIB, Tech. rep., WMO Tech. Rep., 98 pp.[Available online at http://www.wmo.int/pages/prog/www/WMOCodes/Guides/-GRIB/GRIB2_062006.pdf], 2007.

Donchyts, G., Hummel, S., Vaneček, S., Groos, J., Harper, A., Knapen, R., Gregersen, J., Schade, P., Antonello, A., and Gijsbers, P.: OpenMI 2.0—What’s new, in: *Proceedings of 2010 International Congress on Environmental Modelling and Software*, Ottawa, Canada, 2010.

Drost, N.: eWaterleaf: Demo Release, <http://zenodo.org/record/60030>, DOI: 10.5281/zenodo.60030, 2016.

Drost, N. and Hut, R.: ewatercycle: First release candidate, <https://zenodo.org/record/60114>, DOI: 10.5281/zenodo.60114, 2016.

15 Drost, N. and kockx, A.: bmi: First release, <http://zenodo.org/record/60032>, DOI: 10.5281/zenodo.60032, 2016.

Drost, N. and van Meersbergen, M.: Cesium-NcWMS: eWaterCycle specific, <https://zenodo.org/record/60031>, DOI: 10.5281/zenodo.60031, 2016.

Drost, N., Kocks, A., Hummel, S., and Hut, R.: eWaterCycle patched version of OpenDA, <http://zenodo.org/record/60077>, DOI: 10.5281/zenodo.60077, 2016.



- Drusch, M., Scipal, K., de Rosnay, P., Balsamo, G., Andersson, E., Bougeault, P., and Viterbo, P.: Towards a Kalman Filter based soil moisture analysis system for the operational ECMWF Integrated Forecast System, *Geophysical Research Letters*, 36, L10401, doi:10.1029/2009GL037716, <http://onlinelibrary.wiley.com/doi/10.1029/2009GL037716/abstract>, 2009.
- Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., and others: NetCDF Climate and Forecast (CF) metadata conventions, Version, 2003.
- Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., Baugh, C. A., and Cloke, H. L.: Continental and global scale flood forecasting systems, *Wiley Interdisciplinary Reviews: Water*, 3, 391–418, doi:10.1002/wat2.1137, <http://onlinelibrary.wiley.com/doi/10.1002/wat2.1137/abstract>, 2016.
- Erkens, G. and Sutanudjaja, E. H.: Towards a global land subsidence map, *Proceedings of the International Association of Hydrological Sciences*, 372, 83–87, doi:10.5194/piahs-372-83-2015, <http://www.proc-iahs.net/372/83/2015/>, 2015.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343–367, doi:10.1007/s10236-003-0036-9, <http://link.springer.com/article/10.1007/s10236-003-0036-9>, 2003.
- GRDC in the Bundesanstalt für Gewässerkunde, 56068 Koblenz, G.: Global River Discharge Measurements, <http://grdc.bafg.de>, 2015.
- Hut, R., Amisigo, B. A., Steele-Dunne, S., and van de Giesen, N.: Reduction of Used Memory Ensemble Kalman Filtering (RumEnKF): A data assimilation scheme for memory intensive, high performance computing, *Advances in Water Resources*, 86, Part B, 273–283, doi:10.1016/j.advwatres.2015.09.007, <http://www.sciencedirect.com/science/article/pii/S0309170815002080>, 2015.
- Kanamitsu, M.: Description of the NMC Global Data Assimilation and Forecast System, *Weather and Forecasting*, 4, 335–342, doi:10.1175/1520-0434(1989)004<0335:DOTNGD>2.0.CO;2, [http://journals.ametsoc.org/doi/abs/10.1175/1520-0434\(1989\)004%3C0335:DOTNGD%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0434(1989)004%3C0335:DOTNGD%3E2.0.CO%3B2), 1989.
- Kanamitsu, M., Alpert, J., Campana, K., Caplan, P., Deaven, D., Iredell, M., Katz, B., Pan, H.-L., Sela, J., and White, G.: Recent Changes Implemented into the Global Forecast System at NMC, *Weather and Forecasting*, 6, 425–435, doi:10.1175/1520-0434(1991)006<0425:RCIITG>2.0.CO;2, [http://journals.ametsoc.org/doi/abs/10.1175/1520-0434\(1991\)006%3C0425:RCIITG%3E2.0.CO%3B2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0434(1991)006%3C0425:RCIITG%3E2.0.CO%3B2), 1991.
- Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., and Thielen, J.: Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level, *Environmental Modelling & Software*, 75, 68–76, doi:10.1016/j.envsoft.2015.09.009, <http://www.sciencedirect.com/science/article/pii/S1364815215300529>, 2016.
- Kurtz, W., He, G., Kollet, S. J., Maxwell, R. M., Vereecken, H., and Hendricks Franssen, H.-J.: TerrSysMP-PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface–subsurface model, *Geosci. Model Dev.*, 9, 1341–1360, doi:10.5194/gmd-9-1341-2016, <http://www.geosci-model-dev.net/9/1341/2016/>, 2016.
- Lopez Lopez, P., Wanders, N., Schellekens, J., Renzullo, L. J., Sutanudjaja, E. H., and Bierkens, M. F. P.: Improved large-scale hydrological modelling through the assimilation of streamflow and downscaled satellite soil moisture observations, *Hydrology and Earth System Sciences Discussions*, 12, 10559–10601, doi:10.5194/hessd-12-10559-2015, <http://www.hydrol-earth-syst-sci-discuss.net/12/10559/2015/>, 2015.
- Moorthi, S., Pan, H.-L., and Caplan, P.: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system, US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology, Program and Plans Division, <http://www.emc.ncep.noaa.gov/gmb/moorthi/gam.html>, 2001.
- Nijssen, B., O'Donnell, G. M., Hamlet, A. F., and Lettenmaier, D. P.: Hydrologic Sensitivity of Global Rivers to Climate Change, *Climatic Change*, 50, 143–175, doi:10.1023/A:1010616428763, <http://link.springer.com/article/10.1023/A%3A1010616428763>, 2001.



- Oliver, H.: Cylc, <https://cylc.github.io/cylc/>, 2008.
- Peckham, S. D., Hutton, E. W. H., and Norris, B.: A component-based approach to integrated modeling in the geosciences: The design of CSDMS, *Computers & Geosciences*, 53, 3–12, doi:10.1016/j.cageo.2012.04.002, <http://www.sciencedirect.com/science/article/pii/S0098300412001252>, 2013.
- 5 Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, *Proceedings of the National Academy of Sciences*, 111, 3262–3267, doi:10.1073/pnas.1222473110, <http://www.pnas.org/content/111/9/3262>, 2014.
- Rew, R. and Davis, G.: NetCDF: an interface for scientific data access, *IEEE Computer Graphics and Applications*, 10, 76–82, doi:10.1109/38.56302, 1990.
- 10 Rutledge, G. K., Alpert, J., and Ebisuzaki, W.: NOMADS: A Climate and Weather Model Archive at the National Oceanic and Atmospheric Administration, *Bulletin of the American Meteorological Society*, 87, 327–341, doi:10.1175/BAMS-87-3-327, <http://journals.ametsoc.org/doi/abs/10.1175/BAMS-87-3-327>, 2006.
- Slee, M., Agarwal, A., and Kwiatkowski, M.: Thrift: Scalable cross-language services implementation, Facebook White Paper, 5, 2007.
- 15 Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin, *Hydrol. Earth Syst. Sci.*, 15, 2913–2935, doi:10.5194/hess-15-2913-2011, <http://www.hydrol-earth-syst-sci.net/15/2913/2011/>, 2011.
- Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data, *Water Resources Research*, 50, 687–705, doi:10.1002/2013WR013807, <http://onlinelibrary.wiley.com/doi/10.1002/2013WR013807/abstract>, 2014.
- 20 Sutanudjaja, E. H., Drost, N., Kockx, A., and Buitinck, L.: PCR-GLOBWB_model: eWaterCycle Development Version, https://github.com/eWaterCycle/PCR-GLOBWB_model/tree/v2.0_alpha_ewatercycle, DOI: 10.5281/zenodo.60774, 2016a.
- Sutanudjaja, E. H., Van Beek, R., Wada, Y., Bosman, J., Drost, N., de Graaf, I., de Jong, C., Lopez Lopez, P., Pessenteiner, s., Schmitz, O., Straatsma, M., Wanders, N., Wisser, D., and Bierkens, Marc F P, B.: PCR-GLOBWB version v2.1.0_alpha, https://github.com/UU-Hydro/PCR-GLOBWB_model/tree/v2.1.0_alpha, DOI: 10.5281/zenodo.60764, 2016b.
- 25 Tangdamrongsub, N., Ditmar, P. G., Steele-Dunne, S. C., Gunter, B. C., and Sutanudjaja, E. H.: Assessing total water storage and identifying flood events over Tonlé Sap basin in Cambodia using GRACE and MODIS satellite observations combined with hydrological models, *Remote Sensing of Environment*, 181, 162–173, doi:10.1016/j.rse.2016.03.030, <http://www.sciencedirect.com/science/article/pii/S0034425716301171>, 2016a.
- 30 Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., Sutanudjaja, E. H., Sun, Y., Xia, T., and Wang, Z.: Improving estimates of water resources in a semi-arid region by assimilating GRACE data into the PCR-GLOBWB hydrological model, *Hydrology and Earth System Sciences Discussions*, pp. 1–43, doi:10.5194/hess-2016-354, <http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-354/>, 2016b.
- Thiemig, V., Bisselink, B., Pappenberger, F., and Thielen, J.: A pan-African medium-range ensemble flood forecast system, *Hydrol. Earth Syst. Sci.*, 19, 3365–3385, doi:10.5194/hess-19-3365-2015, <http://www.hydrol-earth-syst-sci.net/19/3365/2015/>, 2015.
- 35 Toth, Z., Talagrand, O., and Zhu, Y.: The attributes of forecast systems: a general framework for the evaluation and calibration of weather forecasts, *Predictability of Weather and Climate*, 584, 595, 2006.



- Van Beek, L.: Forcing PCR-GLOBWB with CRU data, Report Department of Physical Geography, Utrecht University, Utrecht, The Netherlands, 2008.
- Van Beek, L. and Bierkens, M. F.: The global hydrological model PCR-GLOBWB: conceptualization, parameterization and verification, Utrecht University, Utrecht, The Netherlands, 2009.
- 5 van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, *Water Resources Research*, 47, W07 517, doi:10.1029/2010WR009791, <http://onlinelibrary.wiley.com/doi/10.1029/2010WR009791/abstract>, 2011.
- van Beek, L. P. H., Eikelboom, T., van Vliet, M. T. H., and Bierkens, M. F. P.: A physically based model of global freshwater surface temperature, *Water Resources Research*, 48, W09 530, doi:10.1029/2012WR011819, <http://onlinelibrary.wiley.com/doi/10.1029/2012WR011819/abstract>, 2012.
- 10 van Meersbergen, M., Drost, N., Sutanudjaja, E. H., and Hut, R.: ewatercycle forecast website, <http://forecast.ewatercycle.org/>, 2015.
- Velzen, N. v., Altaf, M. U., and Verlaan, M.: OpenDA-NEMO framework for ocean data assimilation, *Ocean Dynamics*, 66, 691–702, doi:10.1007/s10236-016-0945-z, <http://link.springer.com/article/10.1007/s10236-016-0945-z>, 2016.
- Wada, Y. and Bierkens, M. F. P.: Sustainability of global water use: past reconstruction and future projections, *Environmental Research Letters*, 9, 104 003, doi:10.1088/1748-9326/9/10/104003, <http://stacks.iop.org/1748-9326/9/i=10/a=104003>, 2014.
- 15 Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., and Bouwman, A.: A framework for global river flood risk assessments, *Hydrol. Earth Syst. Sci.*, 17, 1871–1892, doi:10.5194/hess-17-1871-2013, <http://www.hydrol-earth-syst-sci.net/17/1871/2013/>, 2013.
- Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water, *Water Resources Research*, 47, W05 301, doi:10.1029/2010WR010090, <http://onlinelibrary.wiley.com/doi/10.1029/2010WR010090/abstract>, 2011.
- 20 Zhu, Y.: Ensemble forecast: A new approach to uncertainty and predictability, *Advances in Atmospheric Sciences*, 22, 781–788, doi:10.1007/BF02918678, <http://link.springer.com/article/10.1007/BF02918678>, 2005.



Table 1. Online location of Components, inputs and interface of the eWaterCycle system

Type	Name	online location	citation
component	Cylc	github.com/cylc/cylc	Oliver (2008)
component	CDO	code.zmaw.de/projects/cdo	CDO (2015)
component	OpenDA	OpenDA.org	Velzen et al. (2016)
component	Thrift	thrift.apache.org	Slee et al. (2007)
component	PCRGlobWB 2.0	github.com/UU-Hydro/PCR-GLOBWB_model	van Beek et al. (2011); Sutanudjaja et al. (2014, 2016b)
component	Cesium	cesiumjs.org	(Analytical Graphics, 2011)
Input	GFS Weather forecast	nomads.ncep.noaa.gov/	Kanamitsu (1989); Kanamitsu et al. (1991); Moorthi et al. (2001)
Input	EUMETSAT Soil Moisture	hsaf.meteoam.it/soil-moisture.php	Drusch et al. (2009); De Rosnay et al. (2011)
Interface	Basic Model Interface (BMI)	csdms.colorado.edu/wiki/BMI_Description	Peckham et al. (2013)
Interface	NetCDF	www.unidata.ucar.edu/software/netcdf/	Rew and Davis (1990)
Interface	GRIB	www.wmo.int/pages/prog/www/WMOCodes.html	Dey and others (2007)
Interface	WMS	www.opengeospatial.org/standards/wms	de La Beaujardiere (2006)

Table 2. Online location of eWaterCycle specific implementation of components

Name	online location	Digital Object Identifier (DOI)
Scheduler implemented in Cylc	github.com/eWaterCycle/eWaterCycle	Drost and Hut (2016)
Pre- and postprocessor using CDO	github.com/eWaterCycle/eWaterCycle	Drost and Hut (2016)
eWatercycle specific version of OpenDA	github.com/eWaterCycle/eWaterCycle	Drost et al. (2016)
Generic coupling bridge based on Thrift	contributed to OpenDA	N/A
PCRGlobWB 2.0	github.com/eWaterCycle	Sutanudjaja et al. (2016a)
Web Visualization based on Cesium and ncWMS	github.com/eWaterCycle/Cesium-NcWMS	Drost and van Meersbergen (2016)
BMI binding for eWaterCycle	github.com/eWaterCycle/bmi	Drost and kockx (2016)

Table 3. Metadata of discharge measurements stations from GRDC used for validation and Nash Sutcliff coefficients between forecasts and observations

station ID	River Name	latitude	longitude	Nash Sutcliff nowcast	Nash Sutcliff end forecast
3275750	Chimehuin	-39.80	-71.22	-0.19	0.02
1291100	Zambezi	-17.48	24.30	-26.16	-22.50
4152050	Colorado (United States)	32.73	-114.63	-169252	-196125
6340365	Pleisse	51.20	12.39	-7.63	-11.33