

# UCLA

## UCLA Previously Published Works

### Title

The Regional Hydrologic Extremes Assessment System: A software framework for hydrologic modeling and data assimilation

### Permalink

<https://escholarship.org/uc/item/1892r8xd>

### Journal

PLOS ONE, 12(5)

### ISSN

1932-6203

### Authors

Andreadis, Konstantinos M  
Das, Narendra  
Stampoulis, Dimitrios  
[et al.](#)

### Publication Date

2017

### DOI

10.1371/journal.pone.0176506

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

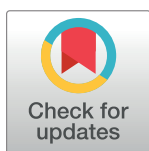
RESEARCH ARTICLE

# The Regional Hydrologic Extremes Assessment System: A software framework for hydrologic modeling and data assimilation

Konstantinos M. Andreadis<sup>1\*</sup>, Narendra Das<sup>1</sup>, Dimitrios Stampoulis<sup>1</sup>, Amor Ines<sup>2,3</sup>, Joshua B. Fisher<sup>1</sup>, Stephanie Granger<sup>1</sup>, Jessie Kawata<sup>1</sup>, Eunjin Han<sup>4</sup>, Ali Behrangi<sup>1</sup>

**1** Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States of America, **2** Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, United States of America, **3** Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI, United States of America, **4** International Research Institute of Climate and Society, Columbia University, Palisades, NY, United States of America

\* [kandread@jpl.nasa.gov](mailto:kandread@jpl.nasa.gov)



## OPEN ACCESS

**Citation:** Andreadis KM, Das N, Stampoulis D, Ines A, Fisher JB, Granger S, et al. (2017) The Regional Hydrologic Extremes Assessment System: A software framework for hydrologic modeling and data assimilation. PLoS ONE 12(5): e0176506. <https://doi.org/10.1371/journal.pone.0176506>

**Editor:** Guy J-P. Schumann, Bristol University/ Remote Sensing Solutions Inc., UNITED STATES

**Received:** February 15, 2017

**Accepted:** April 11, 2017

**Published:** May 18, 2017

**Copyright:** © 2017 Andreadis et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The software code and minimal datasets used for the case studies presented here are available at <https://github.com/nasa/RHEAS>.

**Funding:** The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Support was provided by the Jet Propulsion Laboratory's Research & Technology Development program

## Abstract

The Regional Hydrologic Extremes Assessment System (RHEAS) is a prototype software framework for hydrologic modeling and data assimilation that automates the deployment of water resources nowcasting and forecasting applications. A spatially-enabled database is a key component of the software that can ingest a suite of satellite and model datasets while facilitating the interfacing with Geographic Information System (GIS) applications. The datasets ingested are obtained from numerous space-borne sensors and represent multiple components of the water cycle. The object-oriented design of the software allows for modularity and extensibility, showcased here with the coupling of the core hydrologic model with a crop growth model. RHEAS can exploit multi-threading to scale with increasing number of processors, while the database allows delivery of data products and associated uncertainty through a variety of GIS platforms. A set of three example implementations of RHEAS in the United States and Kenya are described to demonstrate the different features of the system in real-world applications.

## Introduction

Water resources management is a major challenge globally, involving tradeoffs between multiple objectives (e.g., water supply, agriculture, hydropower, ecology) and coordination with a heterogeneous set of stakeholders. Consequently, decision-making in the context of water resources management requires that agencies and practitioners have accurate information on water and energy conditions with as much lead time as possible. Such information is often derived from datasets being offered to end users by data producers, i.e. a producer-driven process [1], rather than users running their own product-generating systems. Despite increased efforts for improved interaction between science organizations (i.e. data producers) and end-users [2], adoption of information for management decisions will be accelerated by direct

(JBF), and by NASA's SERVIR program NNH11ZDA001N-SERVIR (SG).

**Competing interests:** The authors have declared that no competing interests exist.

interaction with and customization of the information-producing system [3]. The aforementioned customization would be rather expensive (in terms of resources) if the developers themselves made modifications, which are likely specific to each end-user. In many cases though, the end-users do not have the technical expertise nor the time or resources to easily implement these beneficial customizations or interact with the information system extensively, thus making the need for a system that can be interacted with in a relatively simple and straightforward fashion more important.

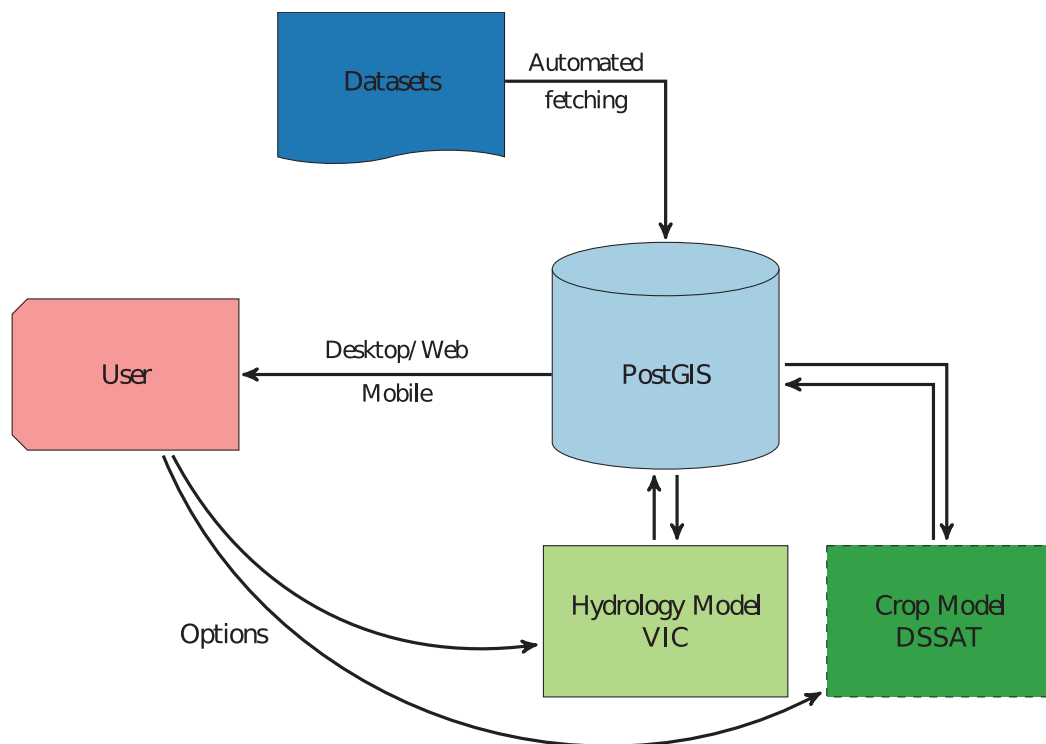
Most likely, a water resources information system would require a model that can simulate the hydrologic response at different spatial and temporal scales. Implementation of such a modeling system is hindered by the dearth of observations that can help calibrate and validate its predictions [4]. In data-poor regions these necessary observation datasets can primarily be obtained from satellite observations, which albeit adds to the difficulty of managing ever-increasing data volumes it also offers the opportunity to better constrain hydrologic models through data assimilation [5].

Here, we present a prototype software framework (the Regional Hydrologic Extremes Assessment System, RHEAS) that automates the ingestion of diverse datasets (both observational and model-based), and the deployment of a hydrologic model incorporating data assimilation and facilitating the coupling with other earth science models. This integrated system is primarily geared for easy implementation and customization requiring relatively little input from end users. Numerous modeling systems that integrate different models and observations have been developed in the past and were used as a stepping stone during the design of RHEAS. The difficulty of discovering, downloading and extracting datasets (either observational or model-based) has been previously recognized [6]. As a result, existing hydrologic software have included capabilities of downloading a predetermined set of datasets (e.g. [7]), while other software have leveraged standardized web services for downloading (e.g. [8]) as well as additionally implementing data discovery options (e.g. [9]).

An important aspect of a hydrologic modeling system is the internal representation of the data (input and output), with a variety of standardized file formats existing (e.g. Network Common Data Form, NetCDF or Hierarchical Data Format, HDF). In addition to files, data can be stored within databases (e.g. PostgreSQL) that can represent spatial and geographic objects resulting in model-independent data management. There is an advantage in using a model-agnostic storage option for the data, since that would facilitate their transferability across models [10] and external visualization and analysis software (e.g. [11]). Although some hydrologic modeling software incorporate GIS components internally (e.g. [12]), RHEAS enables the interfacing with GIS software creating a system that can capture, store, analyze and visualize geospatial data.

Data assimilation methods have been used in hydrologic research and applications to merge heterogeneous observations and models in order to improve model predictions [13]. Such algorithms have been incorporated within hydrologic modeling software, either for specific models (e.g. [14–16]) or as generic frameworks (e.g. [17–20]). RHEAS employs a number of assimilation algorithms by using an object-oriented and modular design, similar to software such as the Land Information System (LIS, [21]). By designing software as modular components and abstracting their functionality, code reuse is maximized and the flexibility on the modeling tools used is facilitated.

A detailed description of the RHEAS software is given in the following section, including its architecture, different components and operating modes. A set of three example applications of the developed software framework are presented in the Application examples section, while the software status and some potential future directions are discussed in the final section of the paper.



**Fig 1. Software architecture.** Simplified schematic of the RHEAS software architecture.

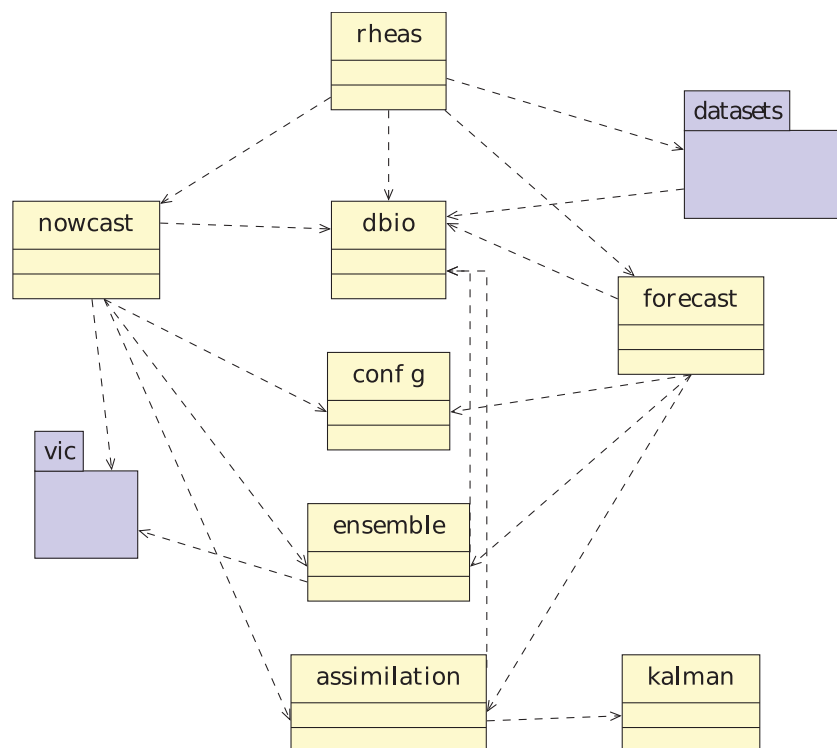
<https://doi.org/10.1371/journal.pone.0176506.g001>

## Software description

RHEAS is a modular software framework that has been developed at the NASA Jet Propulsion Laboratory (JPL) aiming at facilitating the deployment of water resources simulations and the assimilation of remote sensing observations. At the core of the system lies a hydrologic model, the Variable Infiltration Capacity model, that can be run both in nowcasting (i.e. estimation of the current time period) and forecasting (i.e. estimation for future time periods) modes. The nowcast simulation periods can be arbitrarily long, while forecast simulations depend on the length of the meteorological forecasts. In particular, seasonal forecasts will range between 1 and 6 months while long-term forecasts (e.g. climate projections) can range from 5 to 100 years. A suite of datasets from multiple sources are utilized by the system to either force or assimilate observations into the hydrologic model. Data assimilation can constrain hydrologic simulations leading to improved model states and/or parameterizations, and is explicitly incorporated within RHEAS.

## System architecture

Fig 1 shows a schematic of the RHEAS software architecture and its major components. The datasets that are used to perform the model simulations as well as the model outputs are stored within a GIS-enabled relational database (PostGIS), facilitating a model-agnostic dataset format. The latter design choice had several advantages compared to other hydrologic modeling systems: (i) system modularity since the hydrologic model (or any other model that can be added to RHEAS) needs to only interface against the database and not any other model's internal data formats; (ii) GIS functionality that allows spatial operations, complex queries and



**Fig 2. UML component diagram.** Unified Modeling Language (UML) diagram describing the components (i.e. modules) within RHEAS.

<https://doi.org/10.1371/journal.pone.0176506.g002>

analytics on the stored datasets; (iii) ability to serve data through well established technologies (either web, desktop or mobile). The hydrologic model (Variable Infiltration Capacity, VIC) is the primary modeling component within RHEAS, while other models can be coupled (in this case a crop model, DSSAT) extending the system's applicability. All models contained in RHEAS retrieve their input and store their output in the PostGIS database, which can serve these data to users via different interfaces. The datasets that are not produced by the RHEAS models, including satellite observations and model data that are used to generate inputs or constraints for the models, are automatically fetched from various sources and ingested into the PostGIS database.

We designed the RHEAS software following a hybrid approach that combined modular and object-oriented programming. The functionality of the software was broken down into a set of components: (i) configuration, (ii) database operations (I/O and processing), (iii) model simulations, and (iv) data assimilation. Functions associated with each component were encapsulated either within a module or a class if both attributes and methods were needed to describe a component. For example, observations required both attributes (e.g. spatial resolution) and functionality (e.g. retrieve observed variable) and therefore were represented as objects. A simplified UML component diagram is shown in Fig 2 with the main module *rheas* importing the database (*dbio*), simulation (*nowcast* and *forecast*), and *data-sets* modules. Users can define and perform simulations in RHEAS through a text-based configuration file with parsing functions described in the *config* module. Most modules need access to the database functionality, hence they import the *dbio* module. In addition,

**Table 1. Structure of PostGIS tables representing each RHEAS dataset.**

|                                                                                                                   |
|-------------------------------------------------------------------------------------------------------------------|
| <code>rid</code> : integer /* unique identifier for each date and raster tile */                                  |
| <code>fdate</code> : date /* date representation */                                                               |
| <code>rast</code> : raster /* georeferenced data */                                                               |
| <a href="https://doi.org/10.1371/journal.pone.0176506.t001">https://doi.org/10.1371/journal.pone.0176506.t001</a> |

the simulation modules require the classes defined in the `vic` package to run the hydrologic model, the `ensemble` module to perform stochastic simulations, and the `assimilation` module to assimilate satellite observations.

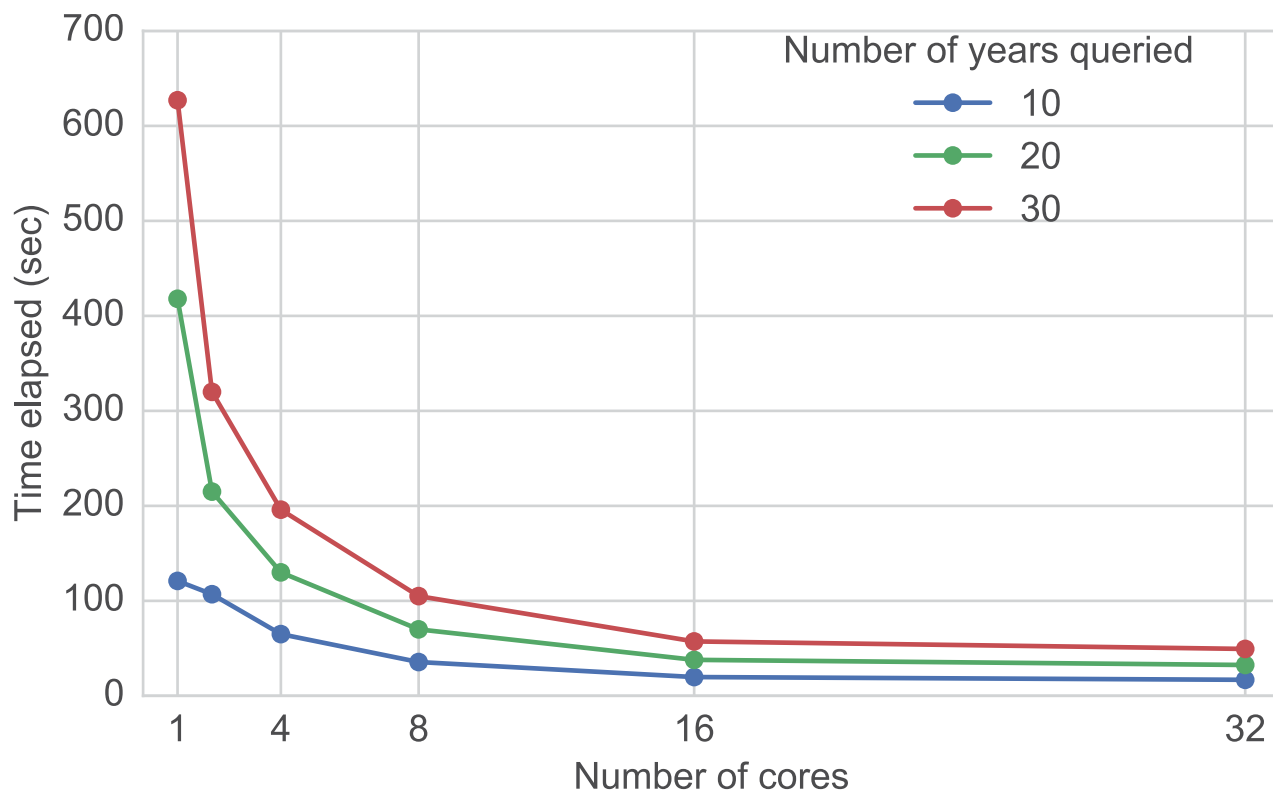
## Dataset ingestion and storage

A number of earth science datasets are available to be used in RHEAS representing many hydro-meteorological variables (e.g. precipitation), with each of being defined as a class within the `datasets` package (Fig 2). The PostGIS database, where the RHEAS datasets are stored, is a spatial extension to the widely-used PostgreSQL object-relational database system [22]. The datasets that are organized in tables (i.e. relations) have both a spatial and temporal dimension, with each time snapshot of the data being stored as a row with columns representing the date, a unique identifier key, the actual data and its geographical information. Table 1 shows the structure of each dataset table along with the type of its column. The `raster` type is represented as a binary blob within PostGIS, and can have multiple bands of georeferenced pixel values.

PostgreSQL allows for querying the data by applying various operators (e.g. union, intersection) with arbitrary constraints. PostGIS expands these capabilities to allow processing and analytic functions (e.g. classification, statistics), map algebra, map reprojections, and spatial resampling to coarser or finer resolutions. If the data are represented as rasters, they are split into tiles (by the PostGIS software) with each row containing a specific raster tile for each date to improve performance. Additional optimizations were implemented and tested in the database to improve query performance. Each dataset ingested in the database has different spatial resolutions, and therefore needs to be resampled to the model resolution when generating its inputs. Since the model has a set of pre-defined spatial resolutions, caching can be employed to significantly reduce query times. Resampled data can be cached as new tables, as the result of a query precomputed from operating on the original raster table. Furthermore, each database table can be clustered, i.e. its rows reordered, based on the spatial proximity of each tile allowing for quicker access to time series of the same or nearby regions.

Although PostgreSQL does not support multi-threaded queries explicitly, RHEAS utilizes database connection pools (i.e. a group of independent database connections) to execute queries in parallel. Fig 3 shows the performance of three queries with increasing number of processors used. Each query corresponds to a different period length retrieved from the database (10, 20, and 30 years) with the spatial domain being a basin of 3,013 pixels. The database performance actually scales very well as the number of processors increase from 1 to 16, with the scaling being almost linear. The single-core performance is actually 2 times slower than the dual-core query time due to the overhead introduced by the multi-threaded execution.

The dataset tables are grouped into schemas, with each schema representing a variable. For example, all precipitation-related datasets are contained within the `precip` schema. A list of the available datasets is shown in Table 2 along with the temporal and spatial resolution of each. Some of these data products have near-global coverage, while others have a regional focus (e.g. RFE2 over Africa or PRISM over the continental United States). RHEAS can



**Fig 3. Database performance.** Example database performance with multiple processors querying different number of years (10, 20, and 30).

<https://doi.org/10.1371/journal.pone.0176506.g003>

automatically fetch each of these datasets and ingest them in the PostGIS database. A separate module was developed for each dataset, although function names for each module are identical (e.g. `download`) allowing for a common interface for the dataset functionality. Data providers use different standards to provide the datasets, both in terms of web services (e.g. OpenDAP, FTP) and file formats (e.g. NetCDF, HDF). Python decorators, a metaprogramming technique, can be used to transform the dataset module's functions during runtime and facilitate code reuse. This allows defining a dataset module using only a Uniform Resource Locator (URL) address and specifications of the URL and file types. As an example, the CHIRPS rainfall dataset that is provided as a set of Geotiff files at a web repository is fetched by simply defining the function in Table 3 with 5 lines of Python code. The `http` decorator dynamically adds the functionality of retrieving the files defined with the `url` variable, while the `geotiff` decorator specifies how data can be extracted from the retrieved file.

The user can define which datasets should be ingested by creating a RHEAS configuration file that only optionally requires a bounding box, start and end dates along with the names of the datasets. If the optional arguments are not provided, RHEAS will query the database for the latest date that had been downloaded and update the dataset to today's date. Table 4 shows an example configuration that can be used to download CHIRPS and NCEP data into the RHEAS database. The configuration file follows the INI format and is composed of sections and pairs of key/values. The domain section defines the geographical area, while each dataset has its own section and consequently its own parameters (e.g. period to download). Using this mechanism, RHEAS significantly simplifies retrieving of satellite and model datasets

**Table 2. List of data products available in the RHEAS database.**

| Variable                 | Product   | Spatial Resolution | Temporal Resolution | Period available | Reference |
|--------------------------|-----------|--------------------|---------------------|------------------|-----------|
| Precipitation            | TRMM      | 0.25°              | Daily               | 1998-Present     | [23]      |
|                          | GPM       | 0.1°               | Daily               | 2014-Present     | [24]      |
|                          | PRISM     | 4 km               | Daily               | 1981-Present     | [25]      |
|                          | CMORPH    | 8 km               | Daily               | 1998-Present     | [26]      |
|                          | CHIRPS    | 5 km               | Daily               | 1981-Present     | [27]      |
|                          | Princeton | 0.25°              | Daily               | 1948-Present     | [28]      |
| Temperature              | RFE2      | 25 km              | Daily               | 2000-Present     | [29]      |
|                          | NCEP      | 1.875°             | Daily               | 1948-Present     | [30]      |
|                          | PRISM     | 4 km               | Daily               | 1981-Present     | [25]      |
| Wind speed               | Princeton | 0.25°              | Daily               | 1948-Present     | [28]      |
|                          | NCEP      | 1.875°             | Daily               | 1948-Present     | [30]      |
| Soil moisture            | Princeton | 0.25°              | Daily               | 1948-Present     | [28]      |
|                          | AMSR-E    | 25 km              | Daily               | 2002-2011        | [31]      |
|                          | SMOS      | 40 km              | Daily               | 2009-Present     | [32]      |
| Evapotranspiration       | SMAP      | 36 km              | Daily               | 2015-Present     | [33]      |
| Snow cover fraction      | MOD16     | 1 km               | 8 days              | 2000-2013        | [34]      |
| Water storage            | MODSCAG   | 500 m              | Daily               | 2000-Present     | [35]      |
|                          | MOD10     | 500 m              | Daily               | 2000-Present     | [36]      |
| Leaf Area Index          | GRACE     | 300 km             | Monthly             | 2002-Present     | [37]      |
| Meteorological forecasts | MCD15     | 1 km               | 8 days              | 2002-Present     | [38]      |
|                          | IRI       | 2.5°               | Monthly             | 2000-Present     | [39]      |
|                          | NMME      | 0.5°               | Daily               | 2000-Present     | [40]      |

<https://doi.org/10.1371/journal.pone.0176506.t002>

**Table 3. Code snippet (URL shortened) defining the fetch function for the CHIRPS dataset module.**

```
@geotiff
@http
def fetch(args):
    url = "http://ftp.chg.ucsb.edu/.../chirps-v2.0.0:04d.1:02d.2:02d.tif.gz"
    return url, args
```

<https://doi.org/10.1371/journal.pone.0176506.t003>

**Table 4. Example configuration for downloading multiple datasets using RHEAS.**

```
[domain]
minlat: -10
maxlat: 5
minlon: -20
maxlon: -12
[ncep]
[chirps]
startdate: 2000-1-1
enddate: 2000-2-1
```

<https://doi.org/10.1371/journal.pone.0176506.t004>

(including batch downloading) and automatically ingesting them in the PostGIS database for further processing by the user.

## Hydrologic modeling

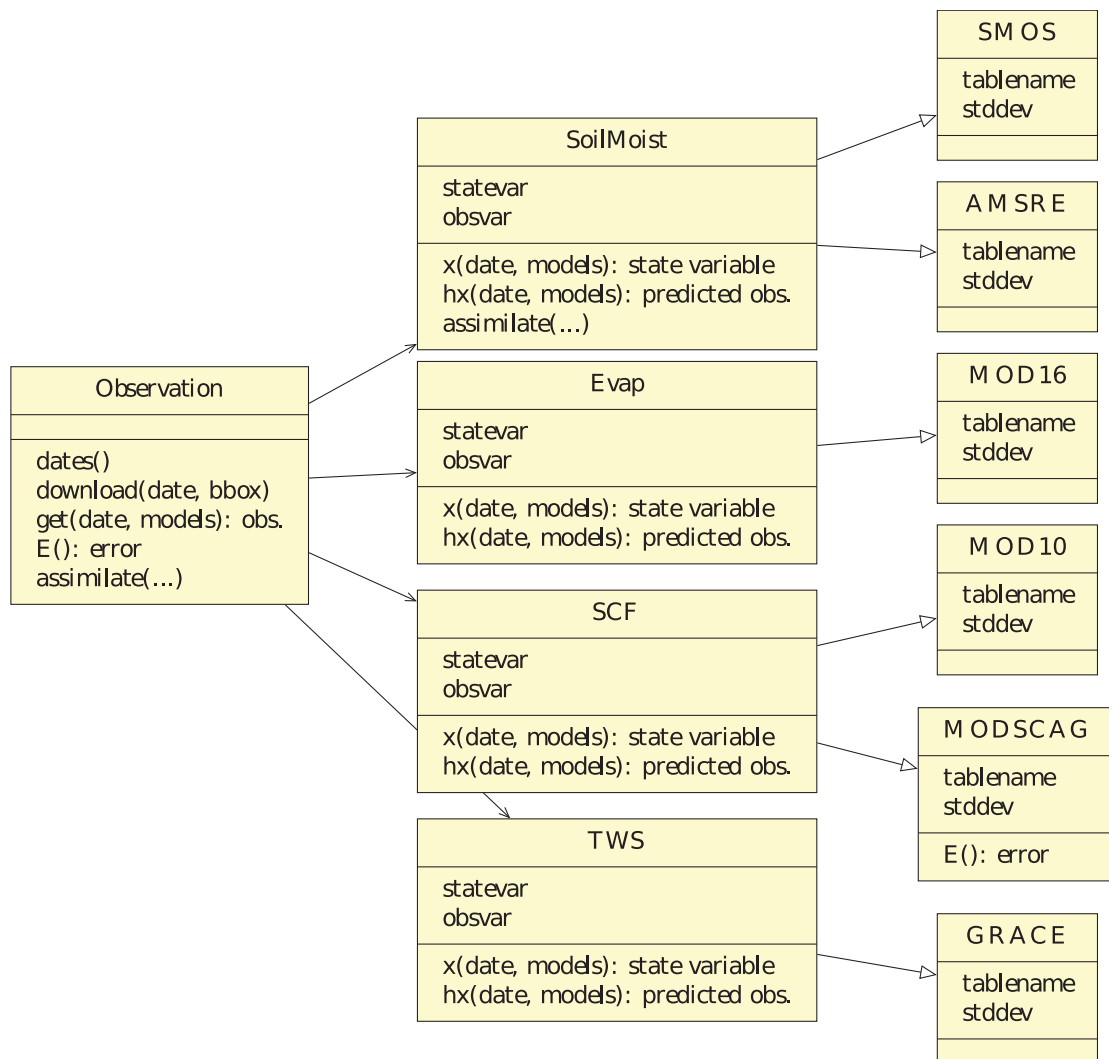
The hydrology model deployed in RHEAS is the Variable Infiltration Capacity (VIC) model [41]. VIC solves the energy and water balance over a gridded domain including a soil-vegetation-atmosphere scheme that models how moisture and energy fluxes between land and atmosphere are controlled by vegetation and soil. Numerous studies have utilized VIC to simulate the hydrology of large river basins continentally and globally (e.g. [42]), making it a good choice for the RHEAS software. The input requirements for VIC include meteorological data that force the model, and information on soil properties, elevation, and land cover. Although multiple sources exist for providing the information on land cover, soils etc., RHEAS has a set of datasets that are utilized to run VIC simulations at varying spatial resolutions ( $1^\circ$ ,  $1/2^\circ$ ,  $1/4^\circ$  globally, and additionally  $1/8^\circ$  and  $1/16^\circ$  over the continental U.S.). Topography information used to partition each model grid cell into elevation zones is derived from the GTOPO30 global digital elevation model, which has a spatial resolution of 30 arc-seconds ( $\sim 1$  km). Land cover information can be readily obtained from satellite datasets, such as the Moderate resolution Imaging Spectroradiometer (MODIS) global product that is generated at a 500-m spatial resolution [43]. Finally, VIC requires information on soil properties which are adapted from global and regional implementations of the VIC model [44, 45].

The VIC model is implemented as a class that contains the functionality of preparing the necessary input files (meteorological forcings, soil and land cover information), running the model executable, and ingesting the model output into the PostGIS database. The model class is encapsulated in the `vic` package, which also contains modules for saving/loading model state files (used during data assimilation) and parsing the user-provided configuration into model options.

## Data assimilation

Data assimilation allows for the optimal merging of model predictions and observations by statistically taking into account the errors in both. Earth Science observations can be ingested into RHEAS using a variety of data assimilation algorithms, with the default being the Ensemble Kalman Filter (EnKF). Additional assimilation algorithms included in RHEAS are the square root EnKF (SREnKF), and the Local Ensemble Transform Kalman Filter (LETKF). The EnKF [46] is a variant of the standard Kalman Filter optimal estimation algorithm and has been widely used in hydrology [13]. The SREnKF is similar to the EnKF, but avoids sampling errors introduced in the standard algorithm resulting in improved state estimates [47]. The LETKF is similar to the other ensemble filters, but performs the analysis (i.e. state update) independently for each model grid point, uses only observations that may affect specific grid point (i.e. localization), and offers algorithmic improvements that enhance the efficiency of the assimilation [48]. The assimilation algorithms within RHEAS have been implemented as abstract classes, i.e. can theoretically work with any model (assuming that they have the ability to restart simulations from a previously saved state), and utilize existing linear algebra libraries [49].

All data assimilation algorithms require an estimate of the model uncertainty, and in the case of the aforementioned techniques that uncertainty is captured by representing the hydrologic variable to be estimated stochastically with an ensemble. Ensembles of model simulations can be generated either by perturbing model forcings and/or parameters (e.g. [19]), or sampling appropriately from climatology. The generation of the ensemble can be controlled by the



**Fig 4. UML observation-class diagram.** UML diagram of classes representing observational datasets that are assimilated (the SMAP class is omitted for visualization purposes).

<https://doi.org/10.1371/journal.pone.0176506.g004>

user (via the configuration file), with all the aforementioned options implemented within RHEAS. When an observation becomes available, the model state is updated leading to an optimal estimate (in terms of least squares). When RHEAS is in nowcast mode the simulation proceeds until the next observations become available, whereas in forecast mode observations are assimilated up to the forecast initialization date after which the model(s) run “free”.

In order to streamline the assimilation of multi-sensor observations, RHEAS is using an object-oriented software design maximizing code reuse [50]. Fig 4 shows a UML diagram of the software classes that represent the observational datasets and their inter-relationships. An abstract class type (Observation) contains the functionality to download the different data products and query the database for the latest data available, retrieve the observation vector for a specific date, generate the observation errors, and perform the data assimilation. The Observation abstract data type is implemented as a set of parent classes for each observation variable

and encapsulates parameters and functions specific to that variable. For example, the soil moisture class (SoilMoist in Fig 4) defines the state (total-column soil moisture) and the observed variable (top-layer soil moisture), and functions for estimating the predicted measurements and deriving the state ensemble. Additionally, if observation-specific methods for data assimilation exist, such as bias correction for soil moisture [51], they are added to the level of this abstract class and are transparently available to each observation sub-class. The latter sub-classes encapsulate parameters specific to a data product (table name in the database, a standard deviation for its error) and inherit their functionality from the parent abstract classes although these can be overridden. For example, a different function that generates the observation error can be defined for the MODSCAG class (Fig 4). Alternatively, uncertainty in the observations can be defined by the user in the RHEAS configuration file (stored locally on the user's computer) by providing the name of a probability distribution (PDF) and a set of parameters. Most current approaches to specifying uncertain parameters in statistical software use markup language representations, either through structured formats such as XML (EXtensible Markup Language) and JSON (JavaScript Object Notation) or an actual Application Programming Interface (API), with a prominent example being UncertML [52]. A slightly different approach was taken in RHEAS, where dynamic module loading from the Scipy library [53] was used to provide the function that samples the PDF to generate the observation errors. Scipy provides a large number of probability distributions, with RHEAS having fallback functions in cases where the user either defines an unavailable distribution or does not provide enough parameters for it. Furthermore, additional assimilation algorithms such as the Particle Filter [54] can be implemented within RHEAS, by utilizing the Observation classes and the EnKF class as a template.

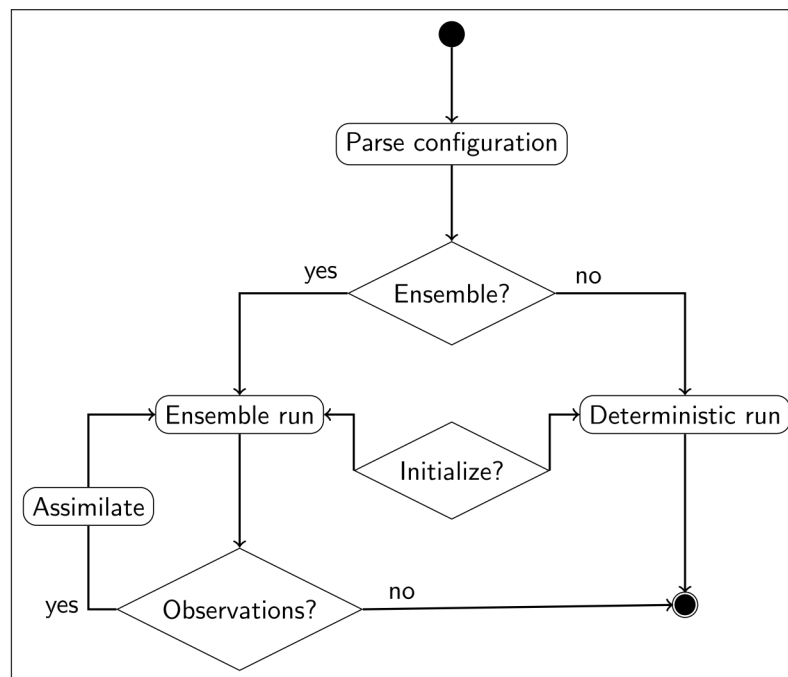
## Model coupling

The modular architecture of RHEAS and the ability to access data in a model-agnostic manner (via the PostGIS database) allow the one-way (i.e. offline) coupling with other environmental models by simply developing an interface against the database itself rather than the hydrologic model. As an example, an agricultural model has been coupled within the RHEAS framework enhancing the software's applicability. The crop model included in RHEAS is based on the Decision Support System for Agro-technology Transfer (DSSAT) modeling system [55]. DSSAT is a process-based model that simulates the growth, development and yield of a crop under given management practices and soil properties (e.g. fertility, water holding capacity). Additional input to DSSAT includes time series of weather variables (rainfall, air temperature, and net solar radiation) that are used to drive the soil hydrology physical model component within DSSAT. The latter interacts with the crop model component of DSSAT simulating the plant's phenology, morphology, and yield.

The DSSAT model implementation used within RHEAS is a modified version of the baseline model that can stop and restart at arbitrary times, whereas crop models generally run continuously from sowing until maturity, failure or harvest by design [56]. This modification was necessary (not just for DSSAT but any model that is coupled within RHEAS) in order to facilitate data assimilation of soil moisture and LAI observations during different phases of crop growth. Moreover, it has been adapted to be deployable over a gridded domain in contrast to the original DSSAT version that is point-based.

## Simulation modes and configuration

Each RHEAS simulation begins with the parsing of a user-provided configuration file that is populated with various simulation parameters. The configuration file follows the INI format,



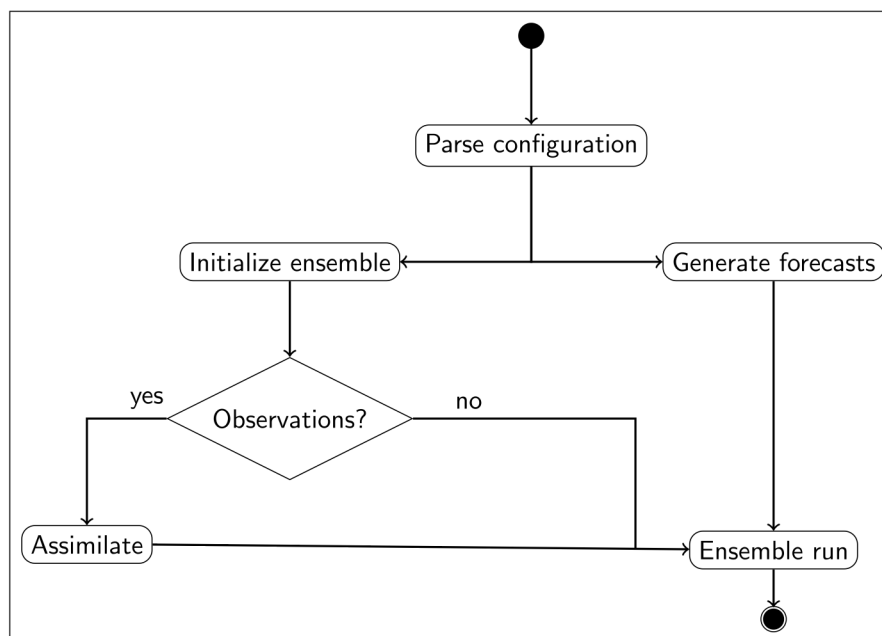
**Fig 5. Nowcast diagram.** Sequence diagram for the nowcasting mode of RHEAS.

<https://doi.org/10.1371/journal.pone.0176506.g005>

with each section corresponding to the type of simulation (nowcast or forecast) and the model used (VIC and/or DSSAT). At a minimum the simulation configuration requires the type of model to be used, the period of simulation, a vector GIS file that defines the model domain, the spatial resolution, and a name for the simulation. Similarly, the VIC model configuration requires the source of meteorological data, and a set of output variables that are written to the RHEAS database. Additional options can be set by the user through the configuration file, although defaults have been preset in order to simplify the deployment of model simulations for non-expert users.

Nowcast simulations (Fig 5) can either be performed deterministically (i.e. single model realization) or stochastically (i.e. ensemble of models). Depending on user input the models can be initialized from a saved state to ensure proper spin-up, which can itself consist of multi-year simulations. If observations are available during the simulation period, they can be assimilated into the model sequentially. The update frequency can be set by the user, since RHEAS supports keywords such as “weekly” and “monthly”. Moreover, data assimilation can only be performed with a stochastic simulation since the assimilation algorithms implemented within RHEAS require an ensemble to describe the model uncertainty.

Similar to the nowcasting mode, forecast simulations (Fig 6) commence with parsing the user-provided configuration file. The hydrologic model requires meteorological forecast data that need to be disaggregated spatially and/or temporally to match the model’s spatial and temporal resolution. The forecast methods implemented in RHEAS are probabilistic, with an ensemble of models spun up to the forecast initialization date. If observations are available on that date, they can be assimilated into the model and the forecast simulation is then launched. The simulation period depends on the duration of the meteorological forecast (e.g. 3 months



**Fig 6. Forecast diagram.** Sequence diagram for the forecasting mode of RHEAS.

<https://doi.org/10.1371/journal.pone.0176506.g006>

for seasonal forecasts), with the entire model ensemble being saved into the database at the end. The meteorological forecasts can be generated either by resampling from climatology or from an atmospheric model.

Whenever a stochastic simulation is performed, an ensemble of model output is saved in the RHEAS database. Ensemble simulations are performed with multi-threaded processing, i.e. each ensemble member is run by a different CPU core, hence accelerating the simulation time. A byproduct of ensemble simulations is the derivation of an uncertainty estimate for each output variable, which can be expressed as the ensemble's standard deviation. Although more sophisticated techniques of estimating uncertainty exist such as Bayesian (excluding Kalman Filters) methods [57], their implementation were beyond the scope of the initial release of RHEAS but could be added potentially to enhance the prediction system. Based on basic decision theory, uncertainty can be considered as a representation of a set of possible states or outcomes with a known probability of occurrence [58]. A decision-maker can choose from a set of possible alternative actions that correspond to each outcome, making even the simple uncertainty estimate from RHEAS potentially useful.

Table 5 shows an example configuration file for a nowcast simulation that assimilates SMOS soil moisture and GRACE water storage observations. The output variables, soil moisture and evaporation, will be written in the database under the schema `testing` as raster tables `soil_moist` and `evap` respectively with each row corresponding to each model time step (daily in this case). Since an ensemble simulation is performed using this configuration, a column specifying the ensemble member number will be added to the aforementioned output tables while additional tables containing rasters of the standard deviation for each variable will be added.

**Table 5. Example RHEAS configuration file for a nowcast simulation.**

|                           |
|---------------------------|
| [nowcast]                 |
| name: testing             |
| startdate: 2012-3-1       |
| enddate: 2012-9-30        |
| basin: domain.shp         |
| resolution: 12km          |
| ensemble size: 10         |
| [vic]                     |
| precip: prism             |
| temperature: prism        |
| wind: ncep                |
| save: soil_moist, evap    |
| observations: smos, grace |

<https://doi.org/10.1371/journal.pone.0176506.t005>

## Application examples

Here, the RHEAS framework was implemented in three case study areas in order to demonstrate and evaluate its capabilities before potentially being deployed operationally, with example results described below. Table 6 shows the execution times for each of the case-study simulations.

### Drought nowcasting in California

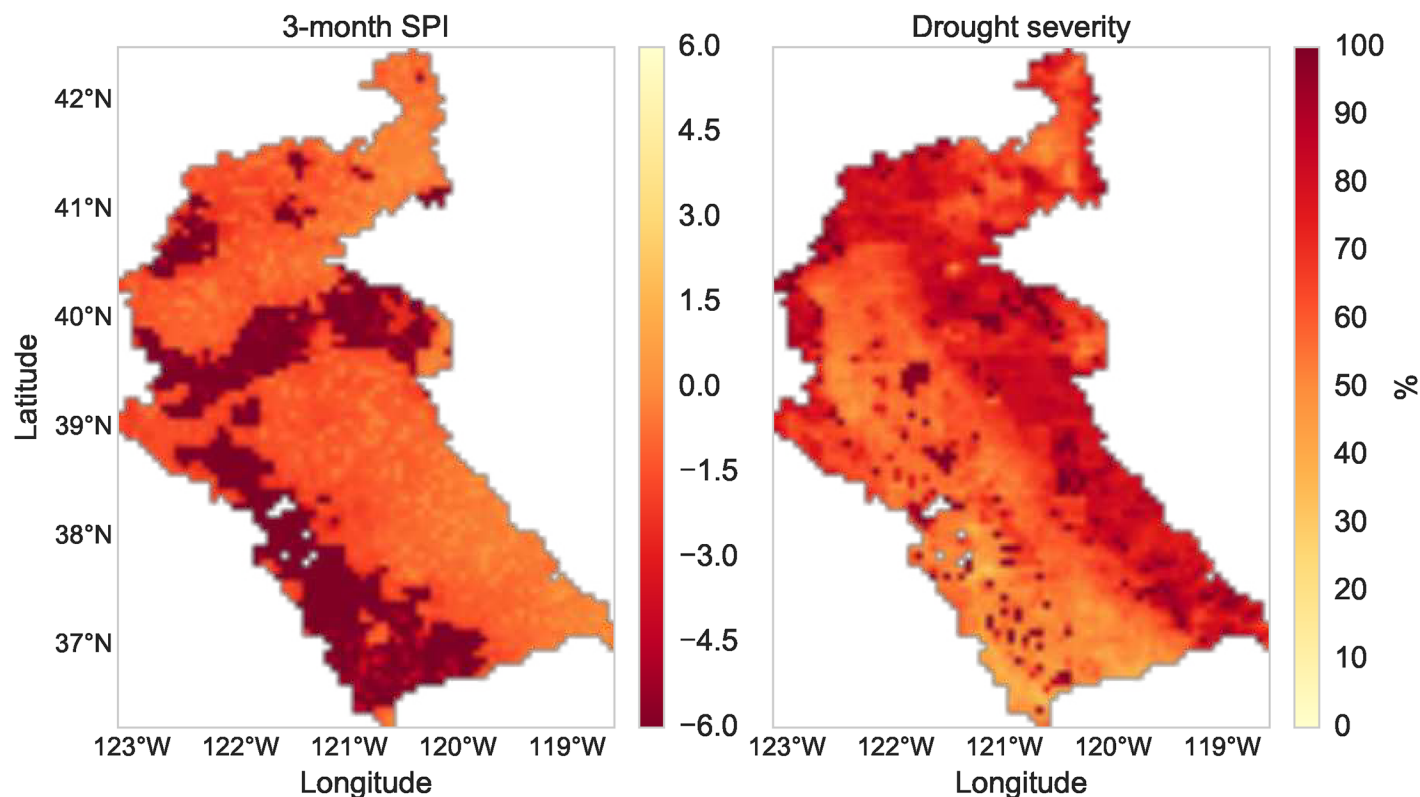
The Sacramento and San Joaquin river basins are located in California and cover most of the Central Valley, which is one of the most productive agricultural areas in the United States. Water resources in the region have been adversely affected by a severe drought that began in the winter of 2011 [59]. Consequently, accurate information on drought characteristics such as the ones produced by RHEAS become very important for water resources managers and practitioners. Fig 7 shows two example maps of drought indicators over the basin on July 2014. The left map shows the 3-month Standardized Precipitation Index (SPI), which is based on the probability of seasonal precipitation and reflects short-term moisture conditions [60]. Agricultural drought severity (Fig 7, right) is derived from the root zone soil moisture expressed as a percentile of the 1981-2010 climatology using the methodology of [61].

The simulations performed to produce these maps used the PRISM dataset to derive precipitation and air temperature forcings for the VIC model, while NLDAS was used to derive wind speed. In addition to the drought data products generated from RHEAS, uncertainty estimates for all simulated hydrologic variables were available. Fig 8 shows a map of uncertainty

**Table 6. Execution times and domain size for each of the case-study simulations (using a 3-GHz 8-core Intel Xeon E5 processor).**

| Case study                    | Domain size (pixels) | Execution time (hrs) |
|-------------------------------|----------------------|----------------------|
| California drought nowcasting | 753                  | 0.84                 |
| Colorado flow forecasting     | 3,402                | 3.37                 |
| Kenya crop yield nowcasting   | 761                  | 1.37                 |

<https://doi.org/10.1371/journal.pone.0176506.t006>



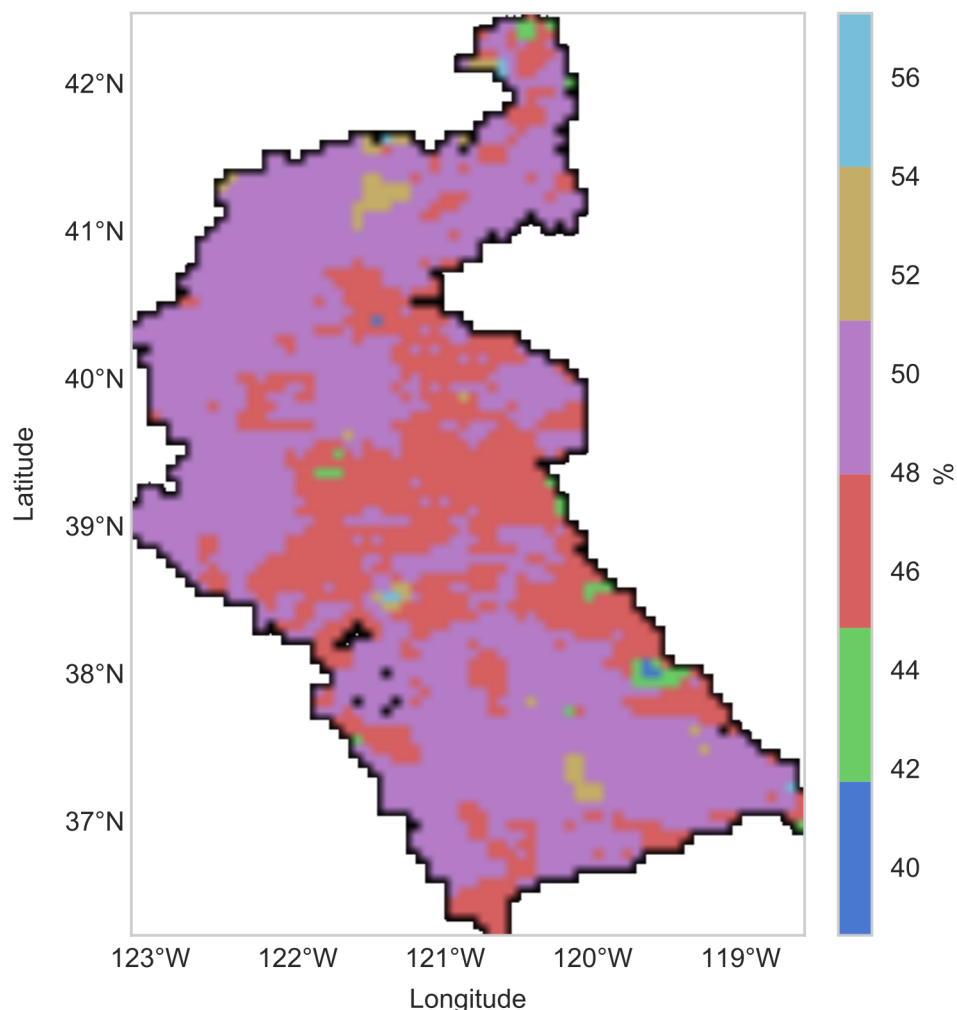
**Fig 7. Drought data products.** Maps of the 3-month Standardized Precipitation Index (left) and agricultural drought severity (right) on 1 July 2014 over the Sacramento/San Joaquin basin.

<https://doi.org/10.1371/journal.pone.0176506.g007>

(expressed as the percentage standard deviation of a 5-member ensemble) in soil moisture over the basin at the end of August 2014.

### Flow forecasting in the Upper Colorado River

The Upper Colorado River plays a very important role for the water resources of the western United States [62], having a rather diverse intra-basin physiography (e.g. elevations range between about 1,000 to more than 4,000 m). Snow controls the timing and magnitude of peak runoff in the basin [63], and therefore has significant implications for water supply management. Observations of snow cover can potentially improve the estimation of streamflow and its forecast skill. In order to test that hypothesis, seasonal forecasts of streamflow were generated using RHEAS at the Colorado River Basin Forecast Center's forecast points. MODIS snow cover observations were assimilated during forecast initialization (1 April 2009), and ESP was used to simulate hindcasts of streamflow. Streamflow was generated by using the off-line VIC river routing model [64] with inputs from the RHEAS simulations. Fig 9 shows time series of streamflow forecasts along with the actually observed streamflow at Taylor River near Almont, Colorado (ALTC2 station). In contrast to open-loop forecasts, assimilated forecasts ingest MODSCAG observations during the initialization date. The forecast ensemble means show that the assimilation of the snow cover observations improved the streamflow forecast skill after mid-May. The improved snow water equivalent (SWE) estimation during forecast



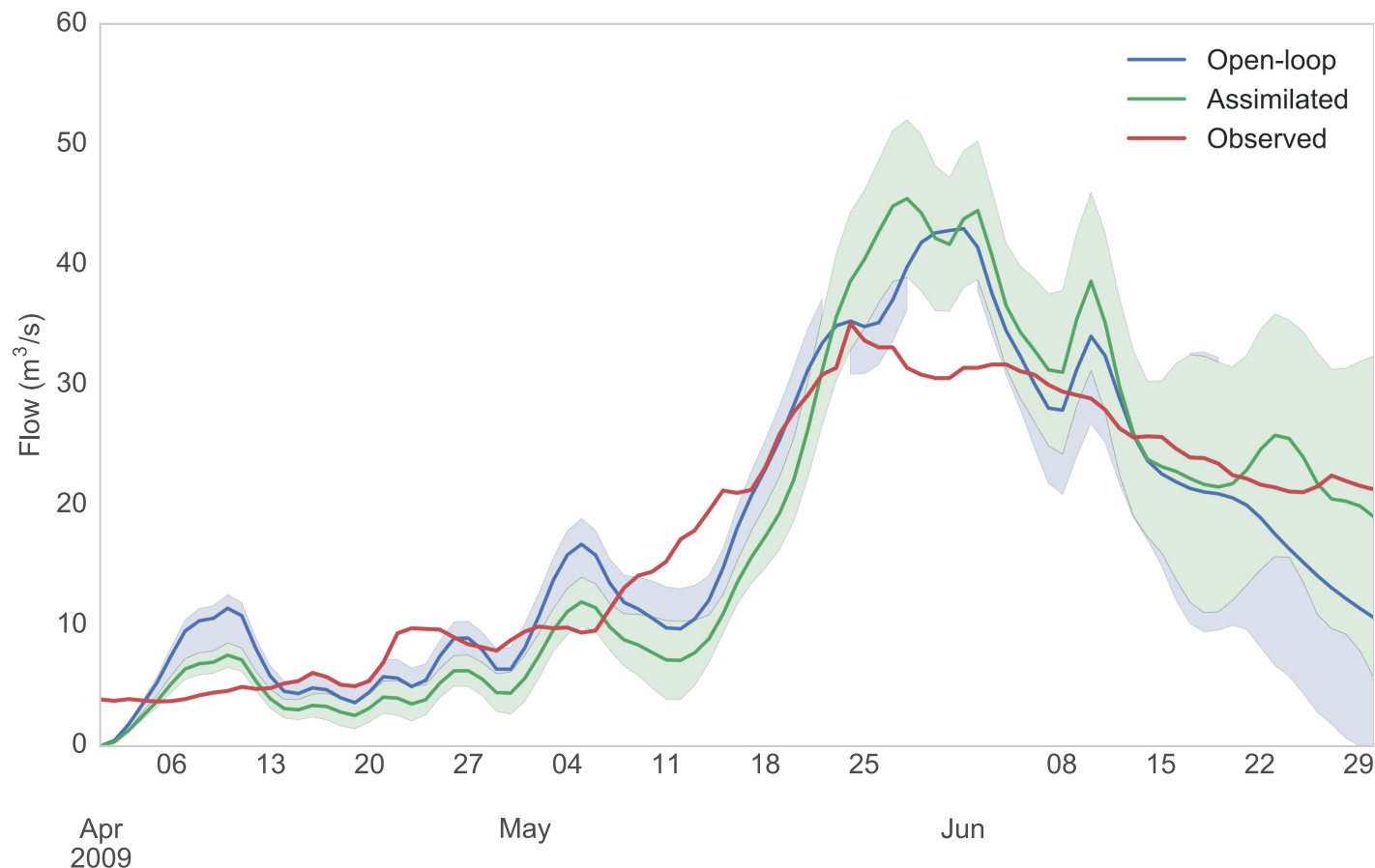
**Fig 8. Uncertainty map.** Map of uncertainty (derived from ensemble  $1\sigma$ ) in soil moisture over Sacramento/San Joaquin river basin on 31 August 2014.

<https://doi.org/10.1371/journal.pone.0176506.g008>

initialization manifests as improvements in streamflow when snowmelt begins to have a significant contribution to the basin's outflow. Additionally, the ensemble spread from both simulations is shown in Fig 9, in the form of the 25th and 75 percentile bounds, with the assimilated forecast range being smaller than the open-loop one suggesting a reduction in uncertainty.

### Crop growth nowcasting in Kenya

RHEAS has been implemented over several countries in the East Africa region, with the goal of producing hydro-agricultural nowcasts and forecasts that can eventually be used to inform decision-making by farmers. Fig 10 shows an example map of simulated maize yield after the earlier planting season in 2011 over Kenya. The yield estimates have been spatially aggregated at the county level, although the finest scale of the simulated output can be defined by a user-provided GIS vector file. The DSSAT model was driven by soil moisture, net solar radiation that were derived from the coupled VIC simulation, as well as LAI, rainfall and air temperature

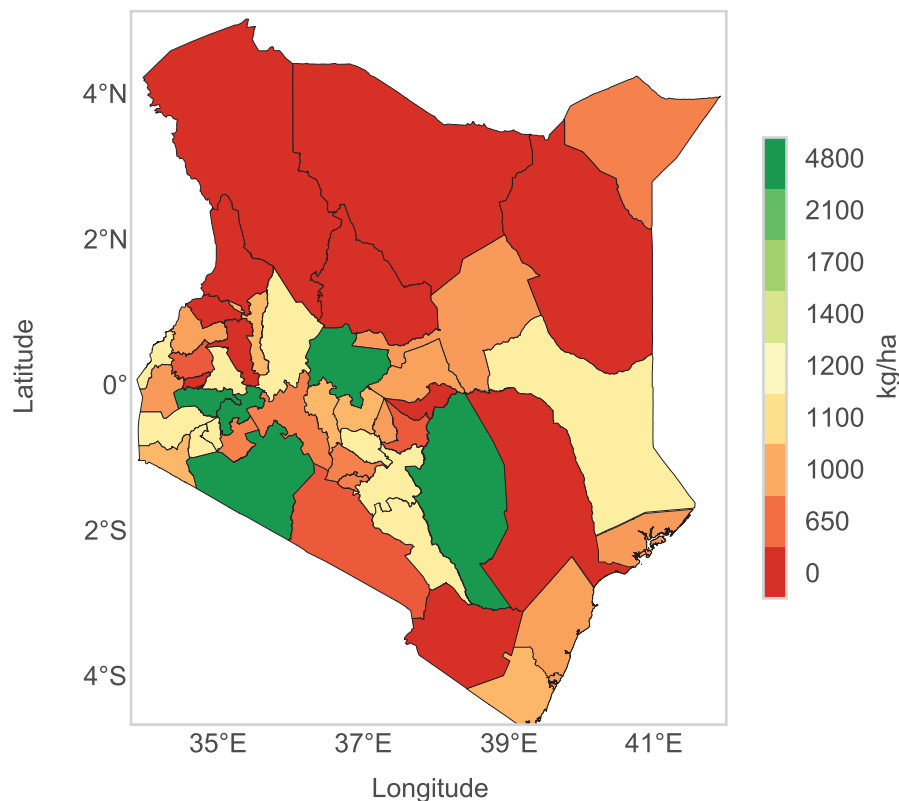


**Fig 9. Streamflow forecast plot.** Time series of forecasted (open-loop and assimilated) and observed streamflow at Taylor River with forecasts initialized on 1 April 2009. Forecasts are bounded by the 25th and 75th percentile of the ensemble.

<https://doi.org/10.1371/journal.pone.0176506.g009>

derived from MODIS observations and the CHIRPS and NCEP datasets respectively. A first-order estimate of planting dates were obtained from a global,  $1/2^\circ$  resolution crop calendar dataset [65]. These estimates have been ingested in the RHEAS database and were used for this simulation, while maize cultivar information (3 varieties) were taken from the Agricultural Model Intercomparison Project [66]. No irrigation and low fertilization were assumed for this simulation, since most of Kenya's agriculture is rain-fed [67] while getting specific information on fertilizer application can be difficult. A drought occurred in East Africa during 2011 affecting agricultural productivity, with counties in Kenya having low yield (Fig 10). Although county-specific yield data were not available, a crude comparison with the national scale with FAOSTAT reported a yield of 1,584 kg/ha for maize while RHEAS simulated a yield of 1,364 kg/ha.

Detailed validation data are difficult to obtain, but yield observations for maize were available over the Nzoia River basin during 2000–2006. The basin has an area of 17,392 km<sup>2</sup>, and a RHEAS simulation at a 25km resolution produced yield estimates during the earlier growing season of the same period as the observations. Fig 11 shows the comparison of the simulated and the observed yields, and with the exception of 2003 when RHEAS significantly overestimated yield (4.51 tons/ha versus 2.42 tons/ha), the model shows reasonably good agreement.



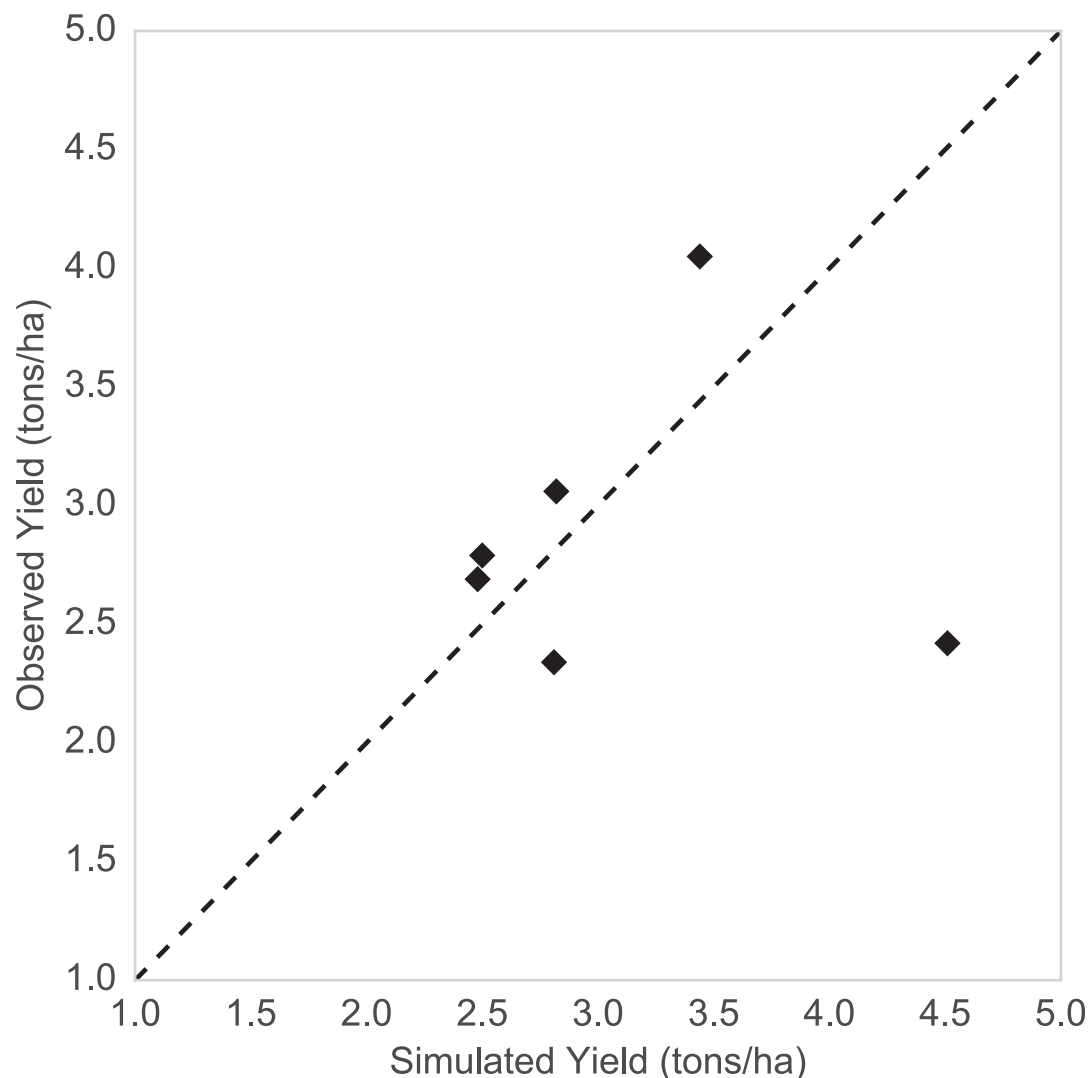
**Fig 10. Crop yield map.** Map of maize yield over Kenya in 2011 (first planting season) disaggregated to county level.

<https://doi.org/10.1371/journal.pone.0176506.g010>

## Summary and future directions

A software framework, RHEAS, that facilitates the deployment of water resources simulations through the ingestion and assimilation of a variety of datasets (including satellite observations) was presented along with three case studies. RHEAS has a spatially-enabled PostgreSQL/PostGIS database at the core with various datasets, both satellite and model-based, being ingested automatically. Apart from the study area, simulation period, and names of input and assimilated datasets, the user does not need to specify any additional parameters, although a programming API (in Python) allows the further customization of the system. Nowcast and forecast simulations can be performed with or without assimilating satellite observations, with the latter representing most of the water cycle components. The RHEAS database allows for the different modeling components to interface against it, simplifying the coupling of a crop growth model. Moreover, the GIS features of the database facilitate the dissemination of the RHEAS data products to diverse platforms (desktop, web and mobile). The output data products include an exhaustive set of hydrologic variables, with each having an uncertainty estimated associated with it.

Compared to similar modeling systems that either require extensive configuration or scripting a solution custom to a specific end-user, RHEAS allows the implementation of a nowcast and/or forecasting system with minimal inputs from user automating and abstracting many of the details away. Although the case studies presented did not include extensive validation, they showcased the ability of RHEAS to generate a suite of data products in relatively diverse



**Fig 11. Crop yield validation.** Comparison of simulated and observed maize yields over the Nzoia River basin during the earlier growing season of 2000–2006.

<https://doi.org/10.1371/journal.pone.0176506.g011>

contexts. The spatial scales that RHEAS is applicable are governed by the hydrology model at its core, VIC, which has been implemented at resolutions ranging from  $1/16^\circ$  to  $1^\circ$ . Therefore the minimum spatial scales for the current version of RHEAS should be on the order of 5 km. Nonetheless, specific regional applications may require further calibration (currently achievable with external software tools) of the VIC model parameters and validation against either in-situ or satellite measurements.

The modular architecture and design of RHEAS could allow for potential modifications in the future that could enhance its applicability. In the context of decision-making, examples could include the use of a soil moisture change product to provide outlook on the crop growth potential, or the drought onset/duration product to plan for food storage. Apart from coupling other types of models (currently the crop growth model DSSAT is available), such as a hydrodynamic model making the system suitable for flood forecasting applications at even sub-hourly time steps. In addition, other hydrologic models can be added within RHEAS to

supplement VIC creating a multi-model ensemble, which could improve predictability. Extension of the RHEAS software framework with new models will require the development of modules that implement I/O between the model and the PostGIS database, and execution of the model physics (including the preparation of its custom input files). Currently there are plans to couple additional agricultural models into the framework representing different crops (e.g rice, cacao) but we anticipate that code contributions from the open source community will further enhance the framework's applicability.

## Acknowledgments

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Support was provided by the Jet Propulsion Laboratory's Research & Technology Development program, and by NASA's SERVIR program. The software has an open-source license and is available for download at <https://github.com/nasa/RHEAS>.

## Author Contributions

**Conceptualization:** KMA.

**Data curation:** KMA DS JBF EH.

**Formal analysis:** KMA ND AI.

**Funding acquisition:** JBF SG.

**Investigation:** KMA.

**Methodology:** KMA ND AI.

**Project administration:** SG JBF KMA.

**Resources:** KMA SG.

**Software:** KMA ND AI.

**Supervision:** KMA.

**Validation:** KMA SG.

**Visualization:** KMA.

**Writing – original draft:** KMA.

**Writing – review & editing:** KMA ND AI SG JBF JK DS EH AB.

## References

1. Horsburgh JS, Tarboton DG, Piasecki M, Maidment DR, Zaslavsky I, Valentine D, et al. An integrated system for publishing environmental observations data. *Environ Model Softw*. 2009; 24(8):879–888. <https://doi.org/10.1016/j.envsoft.2009.01.002>
2. Cash DW, Clark WC, Alcock F, Dickson NM, Eckley N, Guston DH, et al. Knowledge systems for sustainable development. *Proc Natl Acad Sci U S A*. 2003; 100(14):8086–8091. <https://doi.org/10.1073/pnas.1231332100> PMID: 12777623
3. Dilling L, Lemos MC. Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy. *Glob Environ Chang*. 2011; 21(2):680–689. <https://doi.org/10.1016/j.gloenvcha.2010.11.006>

4. Kirchner JW. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour Res.* 2006; 42(3):1–5. <https://doi.org/10.1029/2005WR004362>
5. Liu Y, Gupta HV. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resour Res.* 2007; 43(7). <https://doi.org/10.1029/2006WR005756>
6. Beran B, Piasecki M. Engineering new paths to water data. *Comput Geosci.* 2009; 35(4):753–760. <https://doi.org/10.1016/j.cageo.2008.02.017>
7. Horsburgh JS, Reeder SL. Data visualization and analysis within a Hydrologic Information System: Integrating with the R statistical computing environment. *Environ Model Softw.* 2014; 52:51–61. <https://doi.org/10.1016/j.envsoft.2013.10.016>
8. Castronova AM, Goodall JL, Ercan MB. Integrated modeling within a hydrologic information system: An OpenMI based approach. *Environ Model Softw.* 2013; 39:263–273. <https://doi.org/10.1016/j.envsoft.2012.02.011>
9. Ames DP, Horsburgh JS, Cao Y, Kadlec J, Whiteaker T, Valentine D. HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environ Model Softw.* 2012; 37:146–156. <https://doi.org/10.1016/j.envsoft.2012.03.013>
10. Holzworth DP, Snow V, Janssen S, Athanasiadis IN, Donatelli M, Hoogenboom G, et al. Agricultural production systems modelling and software: Current status and future prospects. *Environ Model Softw.* 2015; 72:276–286. <https://doi.org/10.1016/j.envsoft.2014.12.013>
11. Brooking C, Hunter J. Providing online access to hydrological model simulations through interactive geospatial animations. *Environ Model Softw.* 2013; 43:163–168. <https://doi.org/10.1016/j.envsoft.2013.01.011>
12. Viviroli D, Zappa M, Gurtz J, Weingartner R. An introduction to the hydrological modelling system PRE-VAH and its pre- and post-processing-tools. *Environ Model Softw.* 2009; 24(10):1209–1222. <https://doi.org/10.1016/j.envsoft.2009.04.001>
13. Reichle RH. Data assimilation methods in the Earth sciences. *Adv Water Resour.* 2008; 31(11):1411–1418. <https://doi.org/10.1016/j.advwatres.2008.01.001>
14. Liu S, Shao Y, Yang C, Lin Z, Li M. Improved regional hydrologic modelling by assimilation of stream-flow data into a regional hydrologic model. *Environ Model Softw.* 2012; 31:141–149. <https://doi.org/10.1016/j.envsoft.2011.12.005>
15. Rudd AC, Rouilstone I, Eyre JR. A simple column model to explore anticipated problems in variational assimilation of satellite observations. *Environ Model Softw.* 2012; 27–28:23–39. <https://doi.org/10.1016/j.envsoft.2011.10.001>
16. Giannaros TM, Kotroni V, Lagouvardos K. WRF-LTNGDA: A lightning data assimilation technique implemented in the WRF model for improving precipitation forecasts. *Environ Model Softw.* 2016; 76:54–68. <https://doi.org/10.1016/j.envsoft.2015.11.017>
17. Karssen D, Schmitz O, Salamon P, de Jong K, Bierkens MFP. A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environ Model Softw.* 2010; 25(4):489–502. <https://doi.org/10.1016/j.envsoft.2009.10.004>
18. Werner M, Schellekens J, Gijsbers P, van Dijk M, van den Akker O, Heynert K. The Delft-FEWS flow forecasting system. *Environ Model Softw.* 2013; 40:65–77. <https://doi.org/10.1016/j.envsoft.2012.07.010>
19. Ridler ME, Van Velzen N, Hummel S, Sandholt I, Falk AK, Heemink A, et al. Data assimilation framework: Linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI). *Environ Model Softw.* 2014; 57:76–89. <https://doi.org/10.1016/j.envsoft.2014.02.008>
20. Browne PA, Wilson S. A simple method for integrating a complex model into an ensemble data assimilation system using MPI. *Environ Model Softw.* 2015; 68:122–128. <https://doi.org/10.1016/j.envsoft.2015.02.003>
21. Kumar SV, Peters-Lidard CD, Tian Y, Houser PR, Geiger J, Olden S, et al. Land information system: An interoperable framework for high resolution land surface modeling. *Environ Model Softw.* 2006; 21(10):1402–1415. <https://doi.org/10.1016/j.envsoft.2005.07.004>
22. Obe R, Hsu L. *PostGIS in Action*. 2nd ed. Greenwich, CT, USA: Manning Publications Co.; 2015.
23. Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, et al. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *J Hydrometeorol.* 2007; 8(1):38–55. <https://doi.org/10.1175/JHM560.1>
24. Hou AY, Kakar RK, Neeck S, Azarbarzin AA, Kummerow CD, Kojima M, et al. The Global Precipitation Measurement (GPM) Mission. *Bull Am Meteorol Soc.* 2014; 95(5):711–722. <https://doi.org/10.1175/BAMS-D-13-00164.1>

25. Daly C, Halbleib M, Smith JL, Gibson WP, Doggett MK, Taylor GH, et al. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int J Climatol*. 2008; 28(15):2031–2064. <https://doi.org/10.1002/joc.1688>
26. Joyce RJ, Janowiak JE, Arkin Pa, Xie P. CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *J Hydrometeorol*. 2004; 5(3):487–503. [https://doi.org/10.1175/1525-7541\(2004\)005%3C0487:CAMTPG%3E2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005%3C0487:CAMTPG%3E2.0.CO;2)
27. Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci Data*. 2015; 2. <https://doi.org/10.1038/sdata.2015.66>
28. Sheffield J, Goteti G, Wood EF. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J Clim*. 2006; 19(13):3088–3111. <https://doi.org/10.1175/JCLI3790.1>
29. Xie P, Arkin Pa. Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J Clim*. 1996; 9:840–858. [https://doi.org/10.1175/1520-0442\(1996\)009%3C0840:AOGMPU%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009%3C0840:AOGMPU%3E2.0.CO;2)
30. Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo JJ, Fiorino M, et al. NCEP-DOE AMIP-II reanalysis (R-2). *Bull Am Meteorol Soc*. 2002; 83(November):1631–1643+1559. <https://doi.org/10.1175/BAMS-83-11-1631>
31. Njoku EG, Jackson TJ, Lakshmi V, Chan TK, Nghiem SV. Soil moisture retrieval from AMSR-E. *IEEE Trans Geosci Remote Sens*. 2003; 41(2 PART 1):215–228. <https://doi.org/10.1109/TGRS.2002.808243>
32. Barré HMJP, Duesmann B, Kerr YH. SMOS: The mission and the system. *IEEE Trans Geosci Remote Sens*. 2008; 46(3):587–593. <https://doi.org/10.1109/TGRS.2008.916264>
33. Entekhabi D, Njoku EG, O'Neill PE, Kellogg KH, Crow WT, Edelstein WN, et al. The soil moisture active passive (SMAP) mission. *Proc IEEE*. 2010; 98(5):704–716. <https://doi.org/10.1109/JPROC.2010.2043918>
34. Mu Q, Zhao M, Running SW. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens Environ*. 2011; 115(8):1781–1800. <https://doi.org/10.1016/j.rse.2011.02.019>
35. Painter TH, Rittger K, McKenzie C, Slaughter P, Davis RE, Dozier J. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sens Environ*. 2009; 113(4):868–879. <https://doi.org/10.1016/j.rse.2009.01.001>
36. Salomonson VV, Appel I. Development of the aqua MODIS NDSI fractional snow cover algorithm and validation results. *IEEE Trans Geosci Remote Sens*. 2006; 44(7):1747–1756. <https://doi.org/10.1109/TGRS.2006.876029>
37. Landerer FW, Swenson SC. Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resour Res*. 2012; 48(4):1–11. <https://doi.org/10.1029/2011WR011453>
38. Yang W, Shabanov NV, Huang D, Wang W, Dickinson RE, Nemani RR, et al. Analysis of leaf area index products from combination of MODIS Terra and Aqua data. *Remote Sens Environ*. 2006; 104(3):297–312. <https://doi.org/10.1016/j.rse.2006.04.016>
39. Barnston AG, Li S, Mason SJ, Dewitt DG, Goddard L, Gong X. Verification of the first 11 years of IRI's seasonal climate forecasts. *J Appl Meteorol Climatol*. 2010; 49(3):493–520. <https://doi.org/10.1175/2009JAMC2325.1>
40. Kirtman BP, Min D, Infanti JM, Kinter JL, Paolino DA, Zhang Q, et al. The North American Multi-Model Ensemble (NMME): Phase-1 Seasonal to Interannual Prediction, Phase-2 Toward Developing Intra-Seasonal Prediction. *Bull Am Meteorol Soc*. 2013.
41. Liang X, Lettenmaier DP, Wood EF, Burges SJ. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J Geophys Res*. 1994; 99(D7):14415. <https://doi.org/10.1029/94JD00483>
42. Sheffield J, Andreadis KM, Wood EF, Lettenmaier DP. Global and continental drought in the second half of the twentieth century: Severity-area-duration analysis and temporal variability of large-scale events. *J Clim*. 2009; 22(8):1962–1981. <https://doi.org/10.1175/2008JCLI2722.1>
43. Friedl Ma, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, et al. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens Environ*. 2010; 114(1):168–182. <https://doi.org/10.1016/j.rse.2009.08.016>
44. Voisin N, Wood AW, Lettenmaier DP. Evaluation of Precipitation Products for Global Hydrological Prediction. *J Hydrometeorol*. 2008; 9(3):388–407. <https://doi.org/10.1175/2007JHM938.1>
45. Xia Y, Mitchell K, Ek M, Sheffield J, Cosgrove B, Wood E, et al. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2

- (NLDAS-2): 1. Intercomparison and application of model products. *J Geophys Res.* 2012; 117:D03109. <https://doi.org/10.1029/2011JD016051>
46. Evensen G. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dyn.* 2003; 53(4):343–367. <https://doi.org/10.1007/s10236-003-0036-9>
47. Evensen G. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.* 2004; 54(6):539–560. <https://doi.org/10.1007/s10236-004-0099-2>
48. Hunt BR, Kostelich EJ, Szunyogh I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Phys D Nonlinear Phenom.* 2007; 230(1-2):112–126. <https://doi.org/10.1016/j.physd.2006.11.008>
49. Van Der Walt S, Colbert SC, Varoquaux G. The NumPy array: A structure for efficient numerical computation. *Comput Sci Eng.* 2011; 13(2):22–30. <https://doi.org/10.1109/MCSE.2011.37>
50. David O, Ascough JC, Lloyd W, Green TR, Rojas KW, Leavesley GH, et al. A software engineering perspective on environmental modeling framework design: The Object Modeling System. *Environ Model Softw.* 2013; 39:201–213. <https://doi.org/10.1016/j.envsoft.2012.03.006>
51. Kumar SV, Reichle RH, Harrison KW, Peters-Lidard CD, Yatheendradas S, Santanello Ja. A comparison of methods for a priori bias correction in soil moisture data assimilation. *Water Resour Res.* 2012; 48(3):1–16. <https://doi.org/10.1029/2010WR010261>
52. Williams M, Cornford D, Bastin L, Ingram B. Exchanging Uncertainty: Interoperable Geostatistics? In: *Quantitative Geology and Geostatistics.* Springer Science; 2010. p. 321–331.
53. Jones E, Oliphant T, Peterson P, et al.. SciPy: Open source scientific tools for Python; 2001. Available from: <http://www.scipy.org/>
54. Salamon P, Feyen L. Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *J Hydrol.* 2009; 376(3-4):428–442. <https://doi.org/10.1016/j.jhydrol.2009.07.051>
55. Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt La, et al. The DSSAT cropping system model. *Eur J Agron.* 2003; 18(3-4):235–265. [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7)
56. Ines AVM, Das NN, Hansen JW, Njoku EG. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens Environ.* 2013; 138:149–164. <https://doi.org/10.1016/j.rse.2013.07.018>
57. Lu D, Ye M, Hill MC, Poeter EP, Curtis GP. A computer program for uncertainty analysis integrating regression and Bayesian methods. *Environ Model Softw.* 2014; 60:45–56. <https://doi.org/10.1016/j.envsoft.2014.06.002>
58. Morgan MG, Henrion M, Small M. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*; 1990.
59. Seager R, Hoerling M, Schubert S, Wang H, Lyon B, Kumar A, et al. Causes of the 2011–14 California Drought. *J Climate.* 2015; 28(18):6997–7024. <https://doi.org/10.1175/JCLI-D-14-00860.1>
60. Keyantash J, Dracup Ja. The quantification of drought: An evaluation of drought indices. *Bull Am Meteorol Soc.* 2002; 83(8):1167–1180. [https://doi.org/10.1175/1520-0477\(2002\)083%3C1191:TQODAE%3E2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083%3C1191:TQODAE%3E2.3.CO;2)
61. Andreadis KM, Clark Ea, Wood AW, Hamlet AF, Lettenmaier DP. Twentieth-Century Drought in the Conterminous United States. *J Hydrometeorol.* 2005; 6(6):985–1001. <https://doi.org/10.1175/JHM450.1>
62. Woodhouse Ca, Gray ST, Meko DM. Updated streamflow reconstructions for the Upper Colorado River Basin. *Water Resour Res.* 2006; 42(5):1–16. <https://doi.org/10.1029/2005WR004455>
63. Painter TH, Skiles SM, Deems JS, Bryant AC, Landry CC. Dust radiative forcing in snow of the Upper Colorado River Basin: 1. A 6 year record of energy balance, radiation, and dust concentrations. *Water Resour Res.* 2012; 48(7):1–14. <https://doi.org/10.1029/2012WR011985>
64. Lohmann D, Raschke E, Nijssen B, Lettenmaier DP. Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrol Sci J.* 1998; 43(1):131–141. <https://doi.org/10.1080/02626669809492107>
65. Sacks WJ, Deryng D, Foley Ja, Ramankutty N. Crop planting dates: An analysis of global patterns. *Glob Ecol Biogeogr.* 2010; 19(5):607–620.
66. Rosenzweig C, Jones JW, Hatfield JL, Ruane aC, Boote KJ, Thorburn P, et al. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agric For Meteorol.* 2013; 170:166–182. <https://doi.org/10.1016/j.agrformet.2012.09.011>
67. Portmann FT, Siebert S, Döll P. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochem Cycles.* 2010; 24(1):1–24. <https://doi.org/10.1029/2008GB003435>