Detailed Overview of the Multimodel Multiproduct Streamflow Forecasting Platform

Tirthankar Roy[1, 2], Juan B. Valdés[2], Aleix Serrat-Capdevila[3], Matej Durcik[2, 4],
Eleonora M.C. Demaria[5], Rodrigo Valdés-Pineda[2] and Hoshin V. Gupta[2]

1. *Civil and Environmental Engineering, University of Nebraska-Lincoln*
2. *Hydrology and Atmospheric Sciences, University of Arizona*
3. *Water Global Practice, The World Bank*
4. *Biosphere 2, University of Arizona*
5. *Pima County Flood Control District, Tucson*

Abstract

We present a detailed overview of the Multi-model Multi-product Streamflow Forecasting (MMSF) Platform, which has been developed recently at the University of Arizona under the NASA SERVIR Program, to ease its operational implementation. The platform is based on the use of multiple hydrologic models, satellite-based precipitation products, advanced bias correction schemes, model calibration, and probabilistic model averaging, with the goal of improving forecast accuracy and better-characterizing forecast uncertainties, especially in poorly gauged basins. This paper includes a brief description of the platform, followed by all the relevant information a user would need to implement the platform on any new river basin.

Keywords

MMSF Platform, Streamflow Monitoring, Streamflow Forecasting, Satellite Precipitation, Operational Forecasting.

1. Introduction

1.1. MMSF Platform

Multimodel **M**ultiproduct **S**treamflow **F**orecasting (**MMSF**; Roy et al., 2017a, 2017b) is a new streamflow monitoring and forecasting platform that integrates multiple hydrologic models and satellite-based precipitation products (SPPs). The platform follows successive steps of precipitation bias correction, hydrologic simulations using calibrated models, streamflow bias correction, probabilistic forecast representation, and probabilistic model averaging, to improve the forecast accuracy and better characterize forecast uncertainty. Numerical precipitation forecasts (NPFs) can be added to the SSPs in the platform to generate streamflow forecasts with lead time.

The current version (v1.0) of the platform consists of three hydrologic models: VIC (Liang et al., 1994), HYMOD (Boyle et al., 2000), and HBV-EDU (Aghakouchak and Habib, 2010), four satellite-based precipitation products: TMPA-RT (Huffman et al., 2007), PERSIANN-CCS (Hong et al., 2004), CMORPH (Joyce et al., 2004), and CHIRPS (Funk et al., 2015). Being modular in nature, it allows new (existing) models or products to be added (removed). The real-time version of the platform (MMSF-RT) operates based on eight different modules executing specific tasks, as given in **Figure 1**. Module 1 is for the initial setup. Module 2 deals with the input data. Bias correction and hydrologic model input generation are carried out in Module 3. Module 4 includes the models. Module 5, 6, and 7 are for streamflow bias correction, probabilistic forecast generation, and probabilistic forecast merging. Finally, Module 8 is for final forecast generation, web visualization, and data delivery.

Streamflow forecasts are generated on a daily basis and made available through our research website (www.swaat.arizona.edu). Each day, the platform generates lumped and distributed precipitation plots for the basins and probabilistic streamflow forecasts with mean and uncertainty bounds. Furthermore, it also helps visualize the results through an interactive streamflow plot. We also make the simulation results freely available for research and/or academic use. The current version of the platform is set up on the servers at the University of Arizona, USA and the Regional Center for Mapping of Resources for Development (RCMRD), Kenya. Ongoing research in the group is focusing on assessing the forecasting skill of Global Forecast System (GFS), which is meant to increase the lead time of the streamflow forecasts (Valdés-Pineda et al. *under prep*).

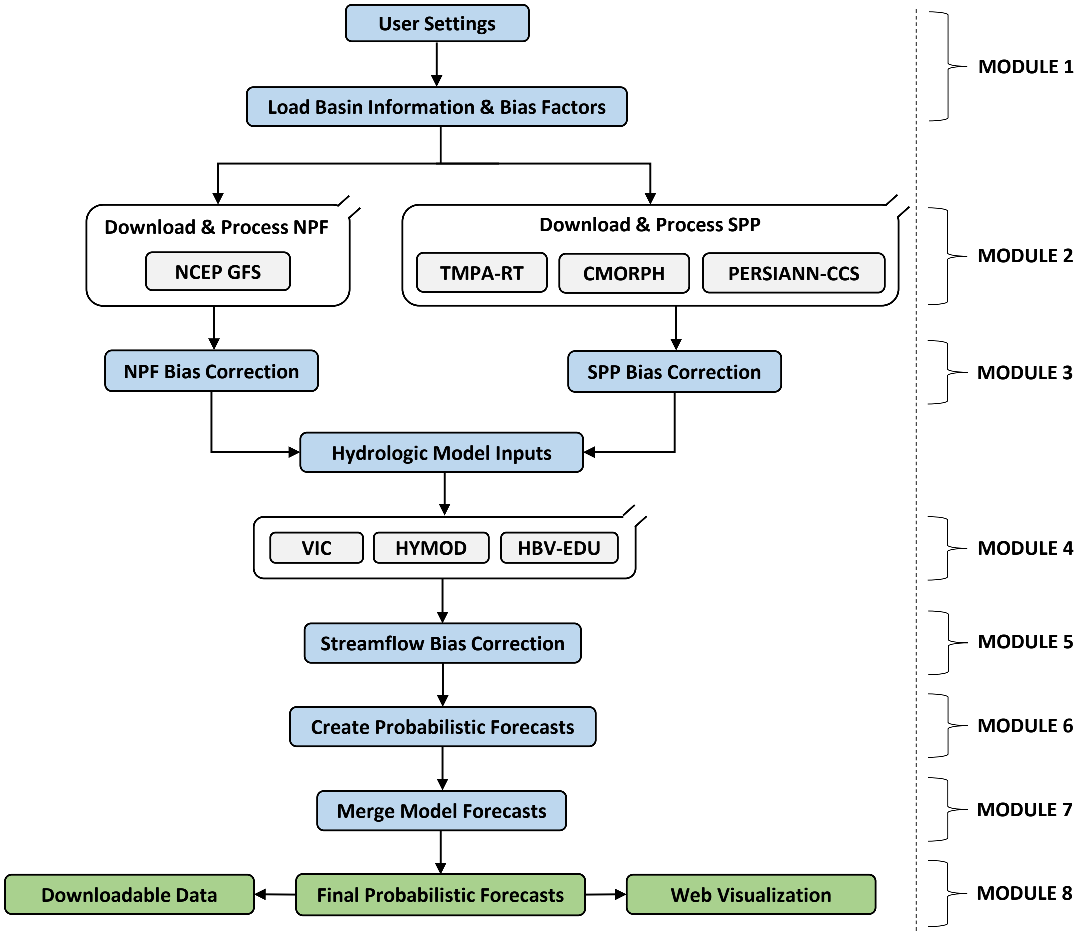

Figure 1: Working modules in the MMSF-RT platform (Roy et al., 2017b).

1.2. MMSF-Basic

In order to simplify the transferability of the platform to any new basin, we have developed a lighter version of the original MMSF Platform, called MMSF-Basic. In this particular version, steps involving the bias correction of streamflows (Module 5 in Figure 1) and probabilistic forecasts characterization (Module 6 Figure 1) are omitted. For basins with observed streamflow records, multimodel streamflow forecasts are merged deterministically using weights calculated from the variance of the corresponding errors during calibration. For basins without streamflow observations, similar to the MMSF platform, MMSF-Basic calculates the arithmetic average of the multimodel forecasts to produce the merged forecasts, and prediction intervals are calculated on a daily basis using the multimodel forecasts either from their empirical distribution or a Gaussian fit. Thus, MMSF-Basic is based on six modules as shown in **Figure 2**. The platform is set up for any single unit catchment with one outlet point. The flow diagram indicating different steps involved in real-time operation of the MMSF-Basic platform is presented in **Figure 3**.

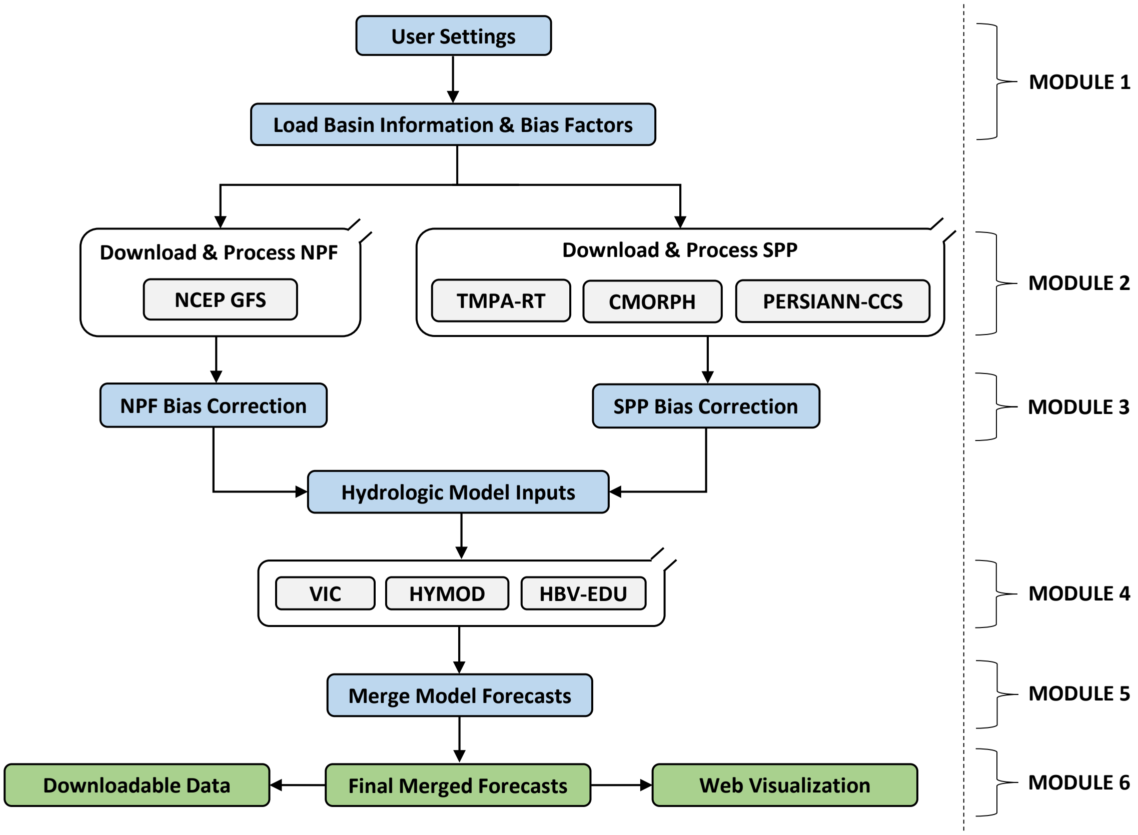

Figure 2: Working modules in the MMSF-Basic Platform.

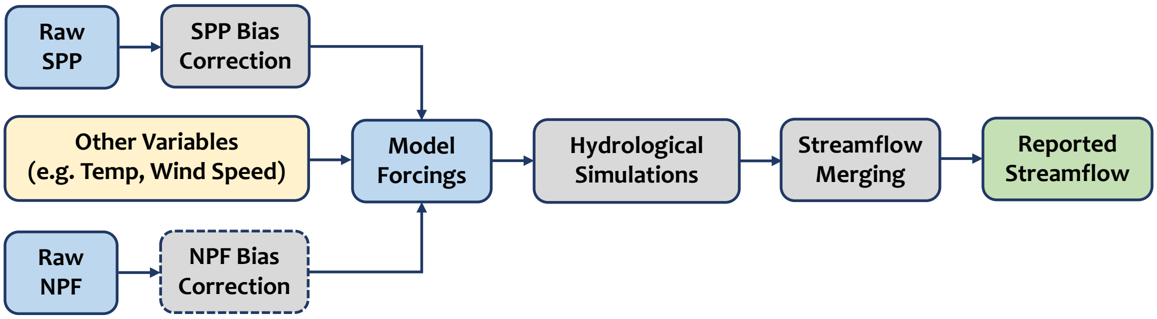

Figure 3: Flow diagram indicating different steps involved in the MMSF-Basic real-time monitoring and forecasting platform. Note that the NPF bias correction (dashed box) is not implemented in the current version of the platform.

1.3. Technical Details

The MMSF platform is originally written in MATLAB. In order to install it on a new computer system, MATLAB compiler needs to be installed first. For more information on MATLAB licensing, the readers can contact Mathworks directly (www.mathworks.com). A compiled version of the platform is also made available for computer systems that do not have MATLAB installed, which requires free MATLAB Compiler Runtime (MCR) to be installed in the system. For a general note on MATLAB compilation steps, please refer to the supplementary materials of Roy et al. (2017b).

The platform runs primarily on Linux servers, requiring roughly 2GB memory (RAM) and 1GB disk space. No additional software packages need to be installed. The program size is around 100MB. The platform is free for academic and/or research use at this point.

We have also made available a Python-based version of the platform with the same configuration and nomenclature as the MATLAB-based version (Credit: Forest Carter, School of Geography and Development, University of Arizona).

1.4. Organization of the Paper

This paper provides detailed guidelines about the installation of the MMSF platform. Section 2 presents the nomenclature used for the development of the platform and Section 3 provides a conceptual summary of the platform. Section 4 describes the toolbox. Section 5 discusses the technical issues and their solutions for implementing the platform on a new computer system. Section 6 presents the steps involved in implementing the platform for a new river basin. Section 7 summarizes a case study published recently regarding the implementation of the platform on the Mara River basin in Africa.

2. Nomenclature

The toolbox uses short but self-explanatory names for the underlying functions and variables. **Table 1** presents the meaning of different suffixes and prefixes used for the variables. A similar naming style can be followed while adding a new model or precipitation product to the platform. Please refer to **Section 3** for function names.

Table 1: Variable names and their interpretations.

Notation	Location	Meaning	Variables
‘P’	Prefix	Precipitation	PF, PbcF, PdC, PhC, PbcC, PtotC, PdT, PhT, PbcT, PtotT, PdP, PhP, PbcP, PtotP
‘Q’	Prefix	Streamflow	QsimC, QsimT, QsimP, Qprob, Qmerg
‘prob’	Prefix	Probabilistic	Qprob (Not Applicable for MMSF-Basic)
‘merg’	Prefix	Merged	Qmerg
‘ci’	Prefix	Confidence Interval	Qci
‘C’	Suffix	CMORPH	PdC, PhC, PbcC, PtotC, QsimC
‘T’	Suffix	TMPA-RT	PdT, PhT, PbcT, PtotT, QsimT
‘P’	Suffix	PERSIANN-CCS	PdP, PhP, PbcP, PtotP, QsimP
‘F’	Suffix	Forecasts (GFS)	PF, PbcF
‘h’	Middle	Hourly/3-Hourly	PhC, PhT, PhP
‘d’	Middle	Daily	PdC, PdT, PdP
‘rw’	Middle	Raw	PrwF
‘bc’	Middle	Bias-corrected	PbcF, PbcC, PbcT, PbcP
‘sim’	Middle	Simulated	QsimC, QsimT, QsimP
‘tot’	Middle	Total	QtotC, QtotT, QtotP
‘lump’	Middle	Lumped	PlumpC, PlumpT, PlumpP, PlumpF
‘dist’	Middle	Distributed	PdistC, PdistT, PdistP, PdistF

3. Conceptual Summary

The MMSF platform is developed to ensure more accurate real-time streamflow monitoring and short-term forecasting capabilities in the poorly gauged basins, accompanied by some realistic representation of uncertainties in the forecasts. The overall goal was to improve water management practices in the basins by utilizing satellite-based datasets and cutting-edge modeling tools. While accurate forecasts are needed to better manage water resources, information about the uncertainties is equally important, since it shows how confident we can be with our forecasts.

The MMSF platform is based on the use of satellite precipitation data. Satellites measure certain aspects of the atmosphere and use that information to estimate precipitation based on different retrieval algorithms. Each product is different in its own way, and it is hard to select the "best" product from the lot since not all of them demonstrate superior performance from all aspects. For example, one precipitation product may be very good for capturing the high precipitation events, whereas the other might be good for capturing the long-term mean. Therefore, it makes more sense to use all the information available at our disposal and try to combine them in a meaningful way, which is exactly what the MMSF platform does. It uses multiple precipitation products and hydrological models and combines the outcomes using different statistical techniques. It also uses statistical techniques to remove the errors in the precipitation products and streamflow outputs (bias correction) and adjust the parameters of the model such that they are able to reproduce streamflow similar to the ones observed (model calibration).

The current version (v1.0) of the platform is fully automated and scheduled to run at a particular time during the day. During each daily run, it downloads precipitation data from the respective servers and removes their errors statistically. These products are then used in multiple hydrologic models, which are pre-calibrated using in-situ streamflow observations from the pilot basins. Once the models produce streamflow forecasts, the latter is further bias-corrected to remove any additional error. Then, the deterministic forecasts are converted to probabilistic forecasts statistically. Finally, the probabilistic forecasts are merged using some statistical model averaging techniques.

The satellite precipitation products are available in the near-real-time. Thus, the primary outputs from the MMSF platform are the near-real-time probabilistic streamflow forecasts. In order to generate forecasts with some lead time, numerical precipitation forecasts need to be appended to the satellite precipitation products and the combined products are then used as inputs to the hydrologic models. The website presents all the daily results in the form of plots and downloadable data.

4. Toolbox Description

4.1. Toolbox Main Functions

The toolbox includes different function files for its eight modules (**Table 2**), each performing some specific tasks for producing the daily forecasts.

Table 2: Function files in the MMSF-RT toolbox adopted from the supplementary materials of Roy et al. (2017b).

Modules	Scripts	Description
Main script	mmsf.m	Main script for MMSF-RT daily runs.
Module 1: Initial setup	setdir.m	Read all associated directory paths.
	basininfo.m	Read basin coordinates and areas.
	timeinfo.m	Read simulation time span and starting date.
Module 2: Precipitation downloading and processing	dlgfs.m	Download GFS forecasts.
	dlspp.m	Download SPPs.
	regrid.m	Regrid to consistent spatial resolution.
	createlump.m	Lump daily SPPs.
	lumping.m	Lump 10-day GFS (3D Matrix).
	dlcmorph.m*	Sub-function of dlspp.m to download CMORPH data.
	dlpccs.m*	Sub-function of dlspp.m to download PERSIANN-CCS data.
	dltmpa.m*	Sub-function of dlspp.m to download TMPA-RT data.
	ascpccs.m*	Sub-function of dlpccs.m to extract ASCII PERSIAN-CCS data.
	bincmorph.m*	Sub-function of dlcmorph.m to extract binary CMOPRH data.
	creategrid.m*	Sub-function of dlpccs.m to create spatial grids.
Module 3: SPP bias correction	bcgfs.m	Bias-correct GFS.
	bcspp.m	Bias-correct SPPs.
	saveprcp.m	Save precipitation for visualization and data publication.
	loadbf.m*	Sub-function of bcspp.m to load bias factors for any given day.
Module 4: Hydrologic model simulation	runhbv.m	Run HBV-EDU with bias-corrected SPPs and GFS forecasts.
	runhymod.m	Run HYMOD with bias-corrected SPPs and GFS forecasts.
	runvic.m	Run VIC-3L with bias-corrected SPPs and GFS forecasts.
	simhbv.m*	Sub-function of runhbv.m to simulate HBV-EDU model.
	simhymod.m*	Sub-function of runhymod.m to simulate HYMOD model.
Module 5: Streamflow bias correction	bcflow.m	Bias-correct simulated streamflows from all three models. (Not Applicable for MMSF-Basic)
Module 6: Probabilistic forecast representation	probflow.m	Create probabilistic streamflow forecasts. (Not Applicable for MMSF-Basic)
Module 7: Forecast merging	mergegaged.m	Merge bias-corrected streamflow forecasts in gaged basins.
	mergeungaged.m	Merge bias-corrected streamflow forecasts in ungaged basins.
Module 8:	plotaxis.m	Set plot properties for customizing web visualization.
	plotlump.m	Plot lumped precipitation using SPPs and GFS forecasts.

Visualization and data publication	plotdist.m	Plot daily contour maps of all distributed SPPs.
	plotind.m	Plot individual forecasts time series.
	plotmerged.m	Plot merged forecasts time series with confidence bounds.
	savecsv.m	Save relevant information for web visualization and data publication.

** Sub-functions that run within the main functions of the envelope script 'mmsf.m'.*

4.2. Additional Toolbox Files

In addition to the function files presented in **Table 2**, The MMSF toolbox also relies upon some additional files for its daily run, which are included in **Table 3**.

Table 3: MMSF-RT additional supporting files.

File Name	**Function**
bzip2	Used for unzipping CMORPH
bunzip2	Used for unzipping CMORPH
bzip2.dll	Dynamic Link Library used in bzip2.exe
gzip	Used for unzipping TMPA-RT and PERSIANN-CCS
lumped_models	A folder with MATLAB codes or a compiled program to support the lumped model runs (HYMOD and HBV-EDU). Original codes are written by Prof. Hoshin Gupta, Department of Hydrology and Atmospheric Sciences, University of Arizona.

4.3. Toolbox Installation

The toolbox can be installed either with MATLAB scripts or with compiled MATLAB executable programs. There isn't any strict requirement on how to organize the setup directories. It is totally up to the users. Here we provide an example.

The data files and the scripts can be organized as described below:

1. mmsf
2. user_files
3. updated_files
4. supporting_files
5. external_models

The 'mmsf' is the main directory with all the source codes and associated programs. In the compiled version of the platform, it is the main executable file. The 'user_files' directory includes the ASCII text files which can be modified by the users without the need for programming (See **Table 4** for details). The files that are updated daily are stored in the 'updated_files' directory. These include daily precipitation and streamflow files. The 'supporting_files' directory stores the files that are created once and not changed daily. It has three sub-directories:

a. 'basin_files': Stores different basin specific files, such as shapefiles, mask files, etc.
b. 'bias_files': Stores bias factor files calculated offline.
c. 'model_files': Stores all model-related files, such as calibrated parameters, lumped and distributed CHIRPS climatologies, climatologies of maximum, minimum and average temperature, wind speed, potential evapotranspiration, model-product combination weights, etc.

The 'external_models' directory contains models that are coupled with the platform without any changes in their source codes. Currently, the platform includes VIC. The configuration of the VIC model coupled with the platform is similar to the original model.

Note that the soil parameter file, vegetation parameter file, and vegetation library (located at 'external_models/vic/param/vic/') need to be updated before running the platform for a new basin unless some common parameters and libraries are used unanimously for all basins. Also, the directory paths and simulation periods need to be updated in VIC Global Parameter File (located at 'external_models/vic/src_vic/') and Routing Input File (located at 'external_models/vic/src_rout/') in order for the model to run on a new system.

5. Setting up for a New Computing System

The MMSF platform is set up for the Linux operating system, however, it can also be run in Windows provided two main issues are addressed:

1. The executable files used in extracting the precipitation products should be compatible with Windows.
2. The VIC model is originally setup in Linux environment. It can also be run in Windows using some Linux emulator such as Cygwin.

Nevertheless, we recommend that the platform be installed directly on a Linux server. If the compiled version of the platform is implemented, it needs to be ensured that the version of the Linux operating system is the same in the source (where the scrips are compiled using MATLAB compiler) and the destination (where the platform is being set up) systems.

6. Setting up for a New River Basin

There are some offline preprocessing steps involved before setting up the platform for a new river basin. In this section, we present the required basin-specific changes in detail.

6.1. Change User Inputs

The ‘user_inputs’ directory within the main MMSF-RT directory contains several ASCII text files that need to be modified before running the platform for a new basin (**Table 4**).

Table 4: User input files within the MMSF-RT platform.

Files	**Description**
basinfile.asc	Includes basin specific information. The first line is the basin area in sq. km and the remaining four lines are for longitude west, longitude east, latitude north, and latitude south coordinates in decimal degrees, respectively.
pathfile.asc	Includes the new paths. ‘maindir’ is the main MMSF directory, ‘pubdir’ is the public directory, ‘vicforc’ is where VIC model forcings are created, ‘vicsrc’ is the directory containing VIC source codes and global parameter file, ‘routsrc’ is the directory with routing source codes and input file, ‘routres’ is where the routed outputs are stored, ‘savedir’ is where the ‘data’ directory is located and finally, ‘gfsdir’ is where the GSF data are downloaded.
startdate.asc	Includes information regarding the starting date of forecasts. Note that the platform can run in hindcasting mode to calculate the previous forecasts, however, sometimes the storage locations and formats of the precipitation products are changed by the developers, which require modifying the associated downloading scripts.
mysh.sh	Shell script for the Crontab scheduler. Update the directory for a new basin.

6.2. Create Initialization Files

In order to create the initialization files, go to ‘…/mmsf/preprocessing/’.

- If MATLAB is preinstalled, run “createvars.m” in MATLAB. Otherwise:
- For Linux, type the following command on the terminal:

  ```
  ‘./run_createvars.sh mcr_path’
  ```

  where mcr_path is the directory where the MCR is installed.

- For Windows double click on “createvars.exe”.

Once the initialization files are created (**Table 5**), copy them to the main MMSF directory. Note that these files will have zeros initially, and depending on the starting date they will start filling up the actual values. These are the files that get updated daily, and so they are not compiled along with the source codes in the compiled version of the platform.

Table 5: Initialization files within the MMSF-RT platform.

Files	**Description**
PlumpC	Lumped CMORPH
PlumpT	Lumped TMPA-RT
PlumpP	Lumped PERSIANN-CCS
PdistC	Distributed CMORPH
PdistT	Distributed TMPA-RT
PdistP	Distributed PERSIANN-CCS
Qavg	Average Streamflow
Qci	Streamflow Confidence Intervals

6.3. Setup Folder Structure

The input and output files of the MMSF-RT platform are saved at different levels as shown in **Figure 4**. There are three main categories, 'inputs', 'outputs', and 'web'. Within the 'input' directory, again, there are two sub-categories, 'fixed' and 'variable'. The fixed inputs are the ones that need to be prepared offline before running the platform, and once they are ready, they do not need to be modified. They can also be compiled along with the source codes in the compiled version of the platform. A detailed description of the fixed input files is given in **Section 5.4**. The variable inputs are the different types of precipitation data that are being stored on a daily basis, which include both raw and bias-corrected values on different time scales (hourly, 3-hourly, and daily).

The output folder stores the daily streamflow forecasts. In 'forecast_daily' folder, 9-day ahead forecasts for each day are stored for future analysis, whereas, in the 'forecast_total' folder, all the forecasts and also the lumped precipitation values are stored in a file called 'webvis.csv'. Note that this .csv file is the main output from the platform, which is then used for data visualization and downloading options available on the website. The 'web' directory contains a copy of the 'webvis.csv' file.

For a new basin, clean the content within the 'data' folder within the 'preprocessing' folder without deleting the sub-folders, then place it on the desired location, and then update the directory address in 'pathfile.asc' (see description in **Section 5.1**).

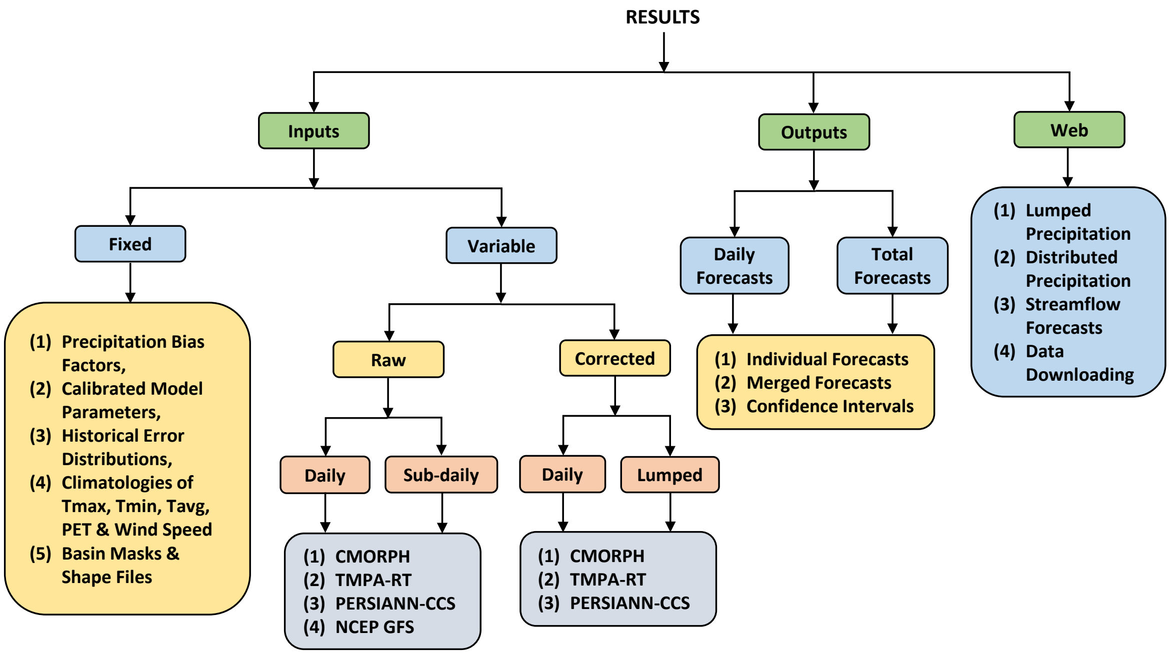

Figure 4: Folder structure within the MMSF-RT platform. Source:(Roy et al., 2017b).

6.4. Create MMSF Supporting Files

The MMSF platform uses some supporting model files that need to be created once for every new basin (**Table 6**). Note that these files can be created using any programming language or even MS Excel for the most part. These files are located in 'data/input/fixed/' directory, and can be divided into three different categories: 'basin_files', 'bias_files', 'model_files'.

Table 6: MMSF Platform supporting files.

Category	**Name**	**Description**
Basin Files	maskgfs.asc	Mask for the GFS data (0.5°).
	maskspp.asc	Mask for the satellite precipitation products (0.05°).
Bias Files	bfcmorph.xls	Excel file with monthly bias factor matrices for CMORPH in each sheet.
	bftmpa.xls	Excel file with monthly bias factor matrices for TMPA-RT in each sheet.
	bfpccs.xls	Excel file with monthly bias factor matrices for PERSIANN-CCS in each sheet.
Model Files	chirpsclimdist.mat	Distributed CHIRPS precipitation climatology used for 10-day ahead forecasting when the GFS data are not available. It is a 3D matrix with latitude coordinates (rows), longitude coordinates (columns), and time (third dimension).

	chirpsclimlump.asc	The lumped version of CHIRPS climatology for the basin. Use ‘maskspp.asc’ to create this file. This serves the same purpose as the distributed version, but for lumped models.
	calpar.xls	Excel file with calibrated lumped model parameters. Each sheet is for a particular model-product combination. Note that the calibrated VIC parameter file has a different format and is located within the ‘vic’ directory.
	Tmax.asc	Lumped maximum temperature climatology (°C) for the basin extracted from any available source.
	Tmin.asc	Lumped minimum temperature climatology (°C) for the basin extracted from any available source.
	Tavg.asc	Lumped average temperature climatology (°C) for the basin calculated by taking the average of Tmax and Tmin.
	wspd.asc	Lumped average wind speed climatology (m/s) for the basin. It can be extracted from any available source.
	PET.asc	Potential Evapotranspiration climatology (mm) calculated by the Hargreaves Equation using the Tmax and Tmin climatologies and the latitude coordinate of the center of the basin.
	TempPET	Mean monthly temperature (°C) and PET (mm) placed in two adjacent columns.
	month.asc	Supporting file for HBV-EDU which includes the indices of the months (Jan=1, Feb=2 … Dec=12) in a single column for each day in the entire simulation period starting from the date mentioned in ‘startdate.asc’ and including the extra day in the leap years.
	errorvar.asc	Error variances associated with each model-product combination. Calculate the variances form the calibration period (simulated flow–observed flow) and store the values in this file. These variances are then used for calculating the weights for multi-model merging.
	errortab.xls (Not Applicable for MMSF-Basic)	Excel file with historical error time series (simulated flow–observed flow). Each sheet contains one column with streamflow errors corresponding to one particular model-product combination. Follow the same order of these combinations both in ‘errortab.xls’ and ‘errorvar.asc’. For example, if the error associated with HYMOD and CMORPH is put in the first sheet of ‘errortab.xls’, put the variance of the same error distribution on the first position of ‘errorvar.asc’, and so on.

6.5. Create Bias Factors

The bias factors for precipitation bias correction need to be calculated once before running the platform for a given basin. There is one bias factor matrix for each precipitation product. Within that bias factor file (data structure) there are 12 monthly bias factor matrices. The factors are calculated on a monthly level using an approach similar to the linear scaling (Roy et al., 2017a). The real-time MMSF platform loads the corresponding bias factors based on the given day and corrects the daily precipitation values in real-time.

Following are the steps involved in creating the bias factors for the basin:

1. Download raingauge data for the basin.
2. Download CHIRPS data for the basin.
3. Download other SPPs for the basin.
4. Regrid SPPs to a resolution consistent with CHIRPS (0.05°).
5. Find out CHIRPS grid cells that contain the rain gauges. If the gauge is located exactly on the boundary of the adjacent cells, take the spatial average of the corresponding cells.
6. Find out a common time frame between CHIRPS and raingauges.
7. Remove CHIRPS data for the days when the raingauge data are missing so that there is a correspondence between the two datasets.
8. For each raingauge, calculate the long term bias factor by diving the temporal mean of the raingauge with the temporal mean of the corresponding CHIRPS grid cell (or spatially averaged cells wherever applicable).
9. Calculate an average bias factor for the entire basin by taking the average of the bias factors for all the raingauges.
10. Multiply the daily CHIRPS values with the average bias factor, thereby creating the adjusted CHIRPS data.
11. Using the adjusted CHIRPS data created in the previous step, calculate monthly bias factors for each grid cells for all the SPPs. To do this first, find a common time frame between the adjusted CHIRPS and the SPPs, and then calculate the factors for each grid cell by diving the month-averages of the adjusted CHIRPS by the month-averages of the SPPs. For example, the bias factor for January would be a ratio between the adjusted CHIRPS Jan mean and the SPP Jan mean, and so on.
12. Create the data structure for each SPP (e.g. bias_CMORPH etc.), containing all the monthly bias factor matrices.

6.6. Hydrologic Models

Three different hydrologic models are being used in the current version of the MMSF platform: HYMOD, HBV-EDU, and VIC.

6.6.1. HYMOD

HYMOD (Boyle et al., 2000) is a conceptual hydrologic model with two different modules: vertical nonlinear soil moisture accounting (SMA) based on Moore (1985) rainfall excess concept and horizontal linear routing (ROUT) based on the leaky linear reservoir and Nash Cascade. Mean daily precipitation and potential evaporation data are used to drive the model, which produces daily estimates of actual ET and streamflow as outputs. HYMOD has total six parameters as given in **Table 7**. Note that the same order of parameters is maintained in the 'calpar.xls' file used within the MMSF platform.

6.6.2. HBV-EDU

HBV-EDU (Aghakouchak and Habib, 2010) is a conceptual hydrologic model with four basic modules: snowmelt and snow accumulation, soil moisture and effective precipitation, evapotranspiration, and runoff response. The model can run on daily or monthly time steps using time series of precipitation and temperature, long-term estimates of mean monthly temperature, and potential evapotranspiration as input variables. Total runoff is simulated by the model as a sum of three components: surface runoff, interflow, and baseflow. There are four state variables in the model: soil moisture level, snow level, storage in the upper tank responsible for quick-flow, and storage in the lower tank responsible for slow-flow or baseflow. The model parameters are presented in **Table 7**.

6.6.3. VIC

VIC (Liang et al., 1994) is a semi-distributed physically-based hydrologic model that carries out calculations individually for each grid cell within the basin. The model is run within the MMSF platform in water balance mode, using precipitation, maximum temperature, minimum temperature, and wind speed as inputs. It has three soil layers and a thin canopy layer on top. Outflows calculated from each grid cell are routed first to the channel and then from the channel to the basin outlet (Lohmann et al., 1996, 1998). The model parameters that are calibrated for the MMSF platform (more sensitive) are given in **Table 7**.

Note that all VIC model related files are located inside the 'vic' directory within the main MMSF directory. Both VIC model (written in C) and its routing component (written in FOTRAN) need to be compiled on a new computing system. The compilation is straightforward. Make sure that the 'gfortran' compiler is installed on the Linux server for FOTRAN code compilation. On the Linux terminal window, change the current directory to the source directory of VIC and type 'make'. Do the same for the routing model. Once the compilations are successful, the Linux executables for VIC (vicNl) and routing model (rout) will be created within the corresponding directories. For more information, visit the following website: vic.readthedocs.io/en/master/.

6.7. Model Calibration

Hydrologic models used in the MMSF platform need to be calibrated before setting up the platform for any new basin unless some nominal parameters are used for all the basins. The calibration is independent of the daily running of the platform. In our case, we used the SCE-UA optimization algorithm (Duan et al., 1992, 1993) for calibrating the three models. Calibration of the lumped models (HYMOD and HBV-EDU) is faster but the calibration of the distributed VIC model is computationally demanding. In our case, we coupled the VIC model within the MATLAB-based SCE-UA code and set it run on the server, which took several days to complete (around 9 days for 2 complexes and 25 loops and 20 days for 5 complexes and 25 loops). We recommend the use of high-performance computing facility, whenever available, for calibrating the VIC model. For the lumped models, in addition to the MATLAB scripts, we have also developed a compiled version of the calibration routine that runs on Linux environment.

Table 7: Hydrologic model parameters.

Model	**Parameter**	**Description** [**L**: Length | **θ**: Temperature | **T**: Time]
HBV-EDU	DD	Degree day factor [$L\theta^{-1}T^{-1}$]
	FC	Field capacity [L]
	β	Shape coefficient for calculating effective precipitation [-]
	C	Parameter to derive adjusted PET [θ^{-1}]
	K_0	Upper tank overflow response function [T^{-1}]
	L	Upper tank overflow threshold [L]
	K_1	Upper tank outlet response function [T^{-1}]
	K_2	Lower tank outlet response function [T^{-1}]
	K_p	Upper to lower tank percolation coefficient [T^{-1}]
	PWP	Permanent wilting point [L]
HYMOD	H_{uz}	Height of soil moisture accounting tank [L]
	B	Distribution function shape parameter [-]
	α	Quick-slow split parameter [-]
	N_q	Number of quickflow routing tanks [-]
	K_s	Slow flow routing tanks rate parameter [-]
	K_q	Quick flow routing tanks rate parameter [-]
VIC	b	Parameter defining the shape of the variable infiltration capacity curve [-]
	Ds	Fraction of Ds_{max} where non-linear baseflow begins [-]
	Ds_{max}	Maximum baseflow from the lowest soil layer [LT^{-1}]
	Ws	Fraction of the maximum soil moisture of the lowest soil layer where non-linear baseflow occurs [-]
	D_2	Depth of the 2nd layer [L]
	D_3	Depth of the 3rd layer [L]

6. Example Case Study for the Mara River Basin

6.1. The Mara River Basin

The MMSF platform has been applied in multiple pilot basins in Africa under the NASA SERVIR Project for daily streamflow monitoring and forecasting. Here we summarize a case study from Roy et al. (2017a, 2017b) for the Mara River basin located in East Africa (**Figure 5**). The basin is close to the equator and covers parts of Kenya (65%) and Tanzania (35%), with a total area of around 13,500 km^2. The Mara River meets several tributaries to reach to the mouth at the Musoma Bay, Lake Victoria. The basin is divided into six main sub-basins: Nyangores, Amala, Lemek, Talek, Sand, and Mid-Mara.

Precipitation in the basin shows bimodal distribution as a result of the yearly oscillation of the Inter Tropical Convergence Zone (ITCZ). The primary rainy season is during March to May, while October to December receive moderate rainfall. The mean annual precipitation varies within the range of 600 to 1500 mm (McClain et al., 2014).

The total population in the basin is around one million, with crop farming, livestock rearing, tourism, and wildlife sanctuary being the primary socioeconomic activities (Mati et al., 2005). The basin is a popular tourist attraction because it encompasses Masai-Mara National Reserve (Kenya) and Serengeti National Park along with game reserves (in Tanzania). The area has experienced large deforestation recently, where almost one-third of the basin was transformed into farmland and tree plantation by 2000 (Mati et al., 2005). The Mara River plays a crucial role in the basin, being the only perennial river in the region (McClain et al., 2014).

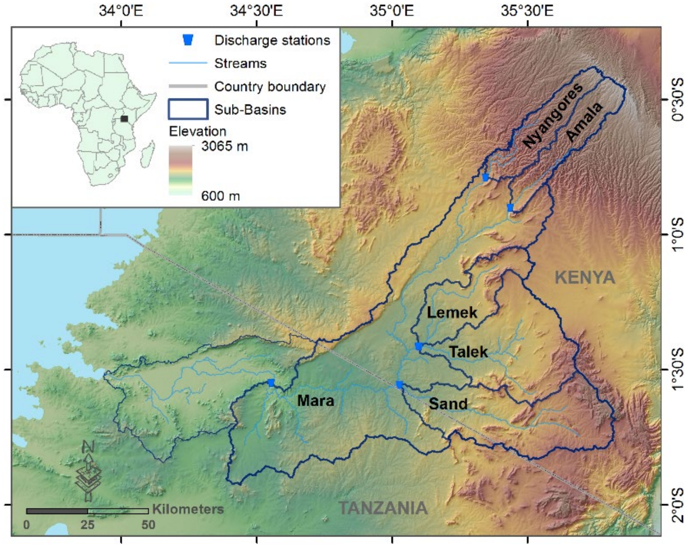

Figure 5: Mara River basin and the sub-basins covering parts of Kenya and Tanzania. Figure adapted from Roy et al. (2018).

6.2. Building Blocks of the MMSF Platform

In this section, we present an example case study covering the implementation of the MMSF platform for real-time streamflow monitoring and forecasting in the Mara River basin of East Africa. The platform is based on the successive steps of precipitation bias correction, calibrated hydrologic model simulations, bias correction of simulated streamflows, probabilistic forecast representation, and probabilistic forecast merging. Calculation of bias factors and calibration of hydrologic model parameters are done offline. On any given day, new satellite precipitation data are first downloaded and then bias-corrected before being used as inputs to the hydrologic models. Each model is run separately with individual precipitation products. Once the model simulations are done, they are bias-corrected and superimposed with historical errors to derive probabilistic forecasts. Finally, the probabilistic forecasts are merged using some probabilistic model averaging technique (e.g. Inverse Error Variance Averaging).

Figure 6a shows an example of precipitation bias correction. In this case, a two-step linear scaling-based bias correction scheme has been implemented on the CMOPRH data using CHIRPS as reference. As can be seen, the corrected precipitation follow the 45 deg line more closely while plotted against the reference, which implies that the bias correction has been effective. Streamflow bias correction results are presented in **Figure 6b**. The bias correction scheme significantly improves the streamflow forecasts, as evident from the sorted values of the bias-corrected and reference forecasts, which closely follow the 45 deg line.

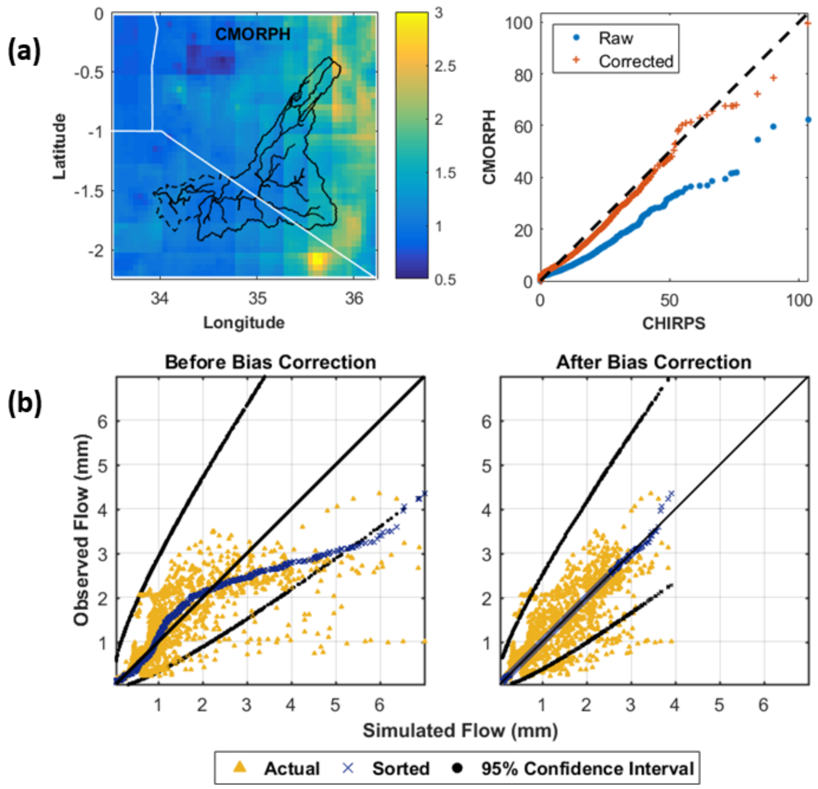

Figure 6: (a) Bias correction of CMORPH precipitation using CHIRPS as the reference data. (b) Bias correction of HBV-EDU streamflow forecasts. Bias-corrected TMPA-RT precipitation is used as one of the inputs to the model. Adapted from: Roy et al., 2017b.

Figure 7 shows the merged forecasts created using the IVA method along with the observations and the confidence intervals. As can be seen, the merged forecast mean closely follows the observations. The confidence intervals also effectively contains the observations.

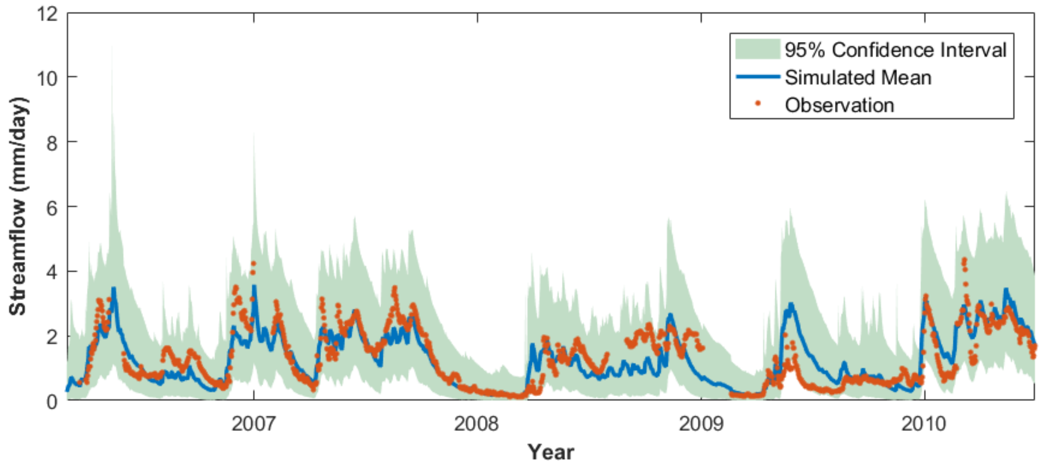

Figure 7: Merged streamflow forecasts from the MMSF platform (Station: Bomet Bridge, Mara River basin, East Africa). Merging is carried out using the IVA technique.

8. Summary and Concluding Remarks

This paper provides detailed information about the operational implementation of the MMSF platform for real-time streamflow monitoring and forecasting using satellite precipitation products. MMSF platform uses multiple hydrologic models and precipitation products to better characterize input and model structural uncertainties. The paper includes details about how to implement the platform in a new basin or on a new computing system. The main scripts, as well as the supporting files, are discussed in detail so that the user has a clear idea about different components of the platform.

MMSF is a state-of-the-art system for real-time streamflow monitoring and forecasting. The platform reinforces the need to produce uncertainty bounds on streamflow forecasts so that the lack of our knowledge about inputs, physical and modeling systems is correctly propagated to the decision-makers. It characterizes forecast uncertainty in a better way by taking into account the historical error distributions, and therefore, unlike ensemble-based forecasting systems, the information, in this case, is not limited solely to the ensemble members. To note, this is the first stable version of the platform (v1.0) and it will be upgraded in the future as necessary. There are several scopes for future research, for example, (1) how to use other satellite-derived variables to improve the performance of the platform, (2) how to improve statistical postprocessing to reduce the forecast uncertainties, (3) how to implement the platform on larger scales, (4) how to improve forecasting of peaks, etc. Any updates made hereafter will be posted on our research website. The readers are encouraged to contact the developers for acquiring the latest version of the platform

Acknowledgment

The main support for this work was provided by the NASA-USAID SERVIR Program through award number 11-SERVIR11-58. International Center for Integrated Water Resources Management (ICIWaRM-UNESCO) provided partial support for initial prototype development. Additional support was available from the UNL Research Council Faculty Seed Grant. The conversion of scripts from MATLAB to Python in MMSF-Basic was carried out by Forest Carter, School of Geography and Development, University of Arizona. The need for this paper became evident during a training workshop on the operational implementation of the MMSF Platform at RCMRD, Kenya. Faith Mitheu and Steven Otieno from RCMRD provided valuable feedbacks from an application point of view. The entire SERVIR Applied Sciences Team (including Ashutosh Limaye, Begum Rabeya Rushi, Emily Adams, and many others) provided valuable feedback throughout the course of this work.

References

Aghakouchak, A. and Habib, E.: Application of a Conceptual Hydrologic Model in Teaching Hydrologic Processes, Int. J. Eng. Educ., 26(4), 1–11, 2010.

Boyle, D. P., Gupta, H. V. and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resour. Res., 36(12), 3663–3674, doi:10.1029/2000WR900207, 2000.

Duan, Q., Sorooshian, S. and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28(4), 1015–1031, doi:10.1029/91WR02985, 1992.

Duan, Q. Y., Gupta, V. K. and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, J. Optim. Theory Appl., 76(3), 501–521, 1993.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Sci. Data, 2, 150066, doi:10.1038/sdata.2015.66, 2015.

Hong, Y., Hsu, K.-L., Sorooshian, S. and Gao, X.: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System, J. Appl. Meteorol., 43(12), 1834–1853, doi:10.1175/JAM2173.1, 2004.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P. and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, J. Hydrometeorol., 8(1), 38–55, doi:10.1175/JHM560.1, 2007.

Joyce, R. J., Janowiak, J. E., Arkin, P. a. and Xie, P.: CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution, J. Hydrometeorol., 5(3), 487–503, doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2, 2004.

Liang, X., Lettenmaier, D. P., Wood, E. F. and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99(D7), 14415, doi:10.1029/94JD00483, 1994.

Lohmann, D., Nolte-Holube, R. and Raschke, E.: A large-scale horizontal routing model to be coupled to land surface parametrization schemes, Tellus A, 48(5), 708–721, doi:10.1034/j.1600-0870.1996.t01-3-00009.x, 1996.

Lohmann, D., Rashke, E., Nijssen, B. and Lettenmaier, D. P.: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, Hydrol. Sci. J., 43(1), 131–141, doi:10.1080/02626669809492107, 1998.

Mati, B. M., Mutie, S., Home, P., Mtalo, F. and Gadain, H.: Land use changes in the transboundary mara basin: A threat to pristine wildlife sanctuaries in East Africa. In A Paper presentation at the: 8th International River Symposium, Brisbane, Australia (pp. 6-9)., in 8th International River Symposium, Brisbane, Australia. [online] Available from: http://vicres.net/downloads/LandUseChanges.pdf, 2005.

McClain, M. E., Subalusky, A. L., Anderson, E. P., Dessu, S. B., Melesse, A. M., Ndomba, P. M., Mtamba, J. O. D., Tamatamah, R. A. and Mligo, C.: Comparing flow regime, channel hydraulics, and biological communities to infer flow–ecology relationships in the Mara River of Kenya and Tanzania, Hydrol. Sci. J., 59(3–4), 801–819, doi:10.1080/02626667.2013.853121, 2014.

Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, Hydrol. Sci. J., 30(2), 273–297, doi:10.1080/02626668509490989, 1985.

Roy, T., Serrat-Capdevila, A., Gupta, H. and Valdes, J.: A platform for probabilistic Multimodel and Multiproduct Streamflow Forecasting, Water Resour. Res., 53(1), 376–399, doi:10.1002/2016WR019752, 2017a.

Roy, T., Serrat-Capdevila, A., Valdes, J., Durcik, M. and Gupta, H.: Design and implementation of an operational multimodel multiproduct real-time probabilistic streamflow forecasting platform, J. Hydroinformatics, 19(6), 911–919, doi:10.2166/hydro.2017.111, 2017b.

Roy, T., Valdés, J. B., Lyon, B., Demaria, E. M. C., Serrat-Capdevila, A., Gupta, H. V., Valdés-Pineda, R. and Durcik, M.: Assessing hydrological impacts of short-term climate change in the Mara River basin of East Africa, J. Hydrol., 566, 818–829, doi:10.1016/j.jhydrol.2018.08.051, 2018.

Valdés-Pineda et al. 2017. (Will be updated during the proofreading stage).