HEPEX intercomparison of streamflow post-processors

Post-Processing hydrologic model simulations (Phase 1)

Executive summary

Aim: evaluation of streamflow post-processors for hydrologic simulations (MOPEX data from several locations in the eastern USA)

Item [1]	Extended run options	Priority run I "restricted choice"	Priority run II "free choice"
Dataset [2]	Hydrologic simulations with calibrated MOPEX models	Hydrologic simulations with calibrated MOPEX models	Hydrologic simulations with calibrated MOPEX models
Data period [3]	01/01/1962 to 31/12/1997	01/01/1962 to 31/12/1997	01/01/1962 to 31/12/1997
Models	sac, gr4j, isba, vic, noah, swb, swap	sac	Free choice (describe)
Catchments (USGS IDs). See: http://waterdata.usgs.gov/usa/nwis/uv?03451500 (change ID as needed)	03451500, 08167500, 01643000, 05455500, 03054500, 07186000, 01608500, 01668000, 07378500, 08172000, 03179000, 03364000	03451500 08167500 01643000 05455500 (others, time permitting)	03451500 08167500 01643000 05455500 (others, time permitting)
Flow accumulation volumes (time step)	1. Daily 2. Free choice [4]	Daily	Daily
Cross-validation approach	1. Leave-one-year-out cross-validation [5] 2. 50/50 split-sample validation [6] 3. Dependent validation [7]	Leave-one-year-out cross-validation	Leave-one-year-out cross-validation
Prior streamflow observations or simulations as auxiliary predictors	1. No prior streamflow 2. Any number of prior streamflow observations or simulations [8]	No prior streamflow	Free choice (describe)
Other auxiliary predictors	1. Precipitation amount	None	Free choice (describe)
Stratified estimation	1. Time (e.g. season) 2. Flow amount 3. Time and flow amount	None	Free choice (describe)
Lead-time[9]	N/A (hydrologic simulations)	N/A (hydrologic simulations)	N/A (hydrologic simulations)
Output [10]	N=100 equally likely ensemble members[11]	N=100 equally likely ensemble members	N=100 equally likely ensemble members

[1] Data is available for download at: ftp.hydro.washington.edu/pub/nathalie/HEPEX/ (anonymous ftp). See also the Annex: "File format and scenario coding"

[2] Schaake JC, Cong S, Duan Q. 2006. The U.S. MOPEX data set. *IAHS Publication* 307: 9-28.
Schaake JC, Duan Q, Andreassian V, Franks S, Hall A, Leavesley G. 2006. The model parameter estimation experiment (MOPEX). *Journal of Hydrology* 320: 1-2. DOI: 0.1016/j.jhydrol.2005.07.054.

[3] This full data period can be used as data used for the hydrologic model spin-up and missing data periods have already been removed

[4] By way of example, data with the 5-day accumulations is provided for download and post-processors can be applied directly to this 5-day dataset

[5] Separate estimation of the post-processor for each possible combination of n-1 calendar years of data in an n-calendar-year dataset, with independent prediction for the remaining one calendar year (the n*1, independently predicted, one-calendar-year datasets, represent the full independent validation period)

[6] Use alternate years for calibration, namely years {1962+2i; i=0,…,17}, and validation, namely years {1963+2i; i=0,…,17}

[7] The full period of data is used for calibration and validation

[8] Fewer days or averages of several days may be used

[9] Predictions will be restricted to the valid time of each simulated streamflow; i.e. the ensembles will comprise one valid time only (the prediction time, which corresponds to the valid time of the simulated streamflow) and a "forecast lead time" of zero

[10] Output data should be provided in the format given in the Annex: "File format and scenario coding"

[11] See the Reference document on how to generate equally like members

HEPEX intercomparison of streamflow post-processors

Post-Processing hydrologic model simulations (Phase 1)

ANNEX: File format and scenario coding

1. Experimental data (data files to be downloaded)

- **Summary:** Streamflow discharge observations and hydrologic model simulation data come from the Model Parameter Estimation Experiment (MOPEX). Data are provided for 12 unregulated river basins in the southeastern part of the United States. Daily discharge simulations were generated by 7 different hydrological models. Simulations were done with calibrated model parameters. The simulation period covers 36 years, starting on January 1, 1962 and ending on December 31, 1997. The period used for model warm-up (1960-1961) was removed. Simulation data for all of the models is complete for the full 36-year period. Therefore, the full 36-year period can be used for post-processing.
- **Location to download the data: ftp.hydro.washington.edu/pub/nathalie/HEPEX/ (anonymous ftp)**
- **File to download:** all data are in the zipped file **HEPEX_Post-processing_data.zip**
- **Inside "HEPEX_Post-processing_data.zip":** you will find 12 data files, one for each river basin. These files are named **"basinid_cal.zip"**, where "basinid" is the 8-digit ID particular for each basin and "cal" refers to the use of models with calibrated parameters.
- **Inside each "basinid_cal.zip",** two files can be found, one for each type of flow accumulation volumes (time step) to be used in the post-processing:
  - **"basinid_cal_1D.dat":** contains daily streamflow data
  - **"basinid_cal_5D.dat":** contains accumulated 5-day streamflow data
- **Format of the *.dat files:**
  - By rows: there is a separate data record for each time step of the 36-year period.
  - By columns: each file contains the model simulations for the 7 models used in the experiment and the corresponding streamflow observations

Col 1: date
Col 2: observed precipitation (mm)
Col 3: observed streamflow discharge (mm/1 day or mm/5 days) *
Col 4-11: simulated streamflow discharge for the 7 hydrologic models (mm/1 day or mm/5 days)

* The one-day and five-day accumulations are in mm/day and mm/5 days, respectively, and the dates represent the valid times ending on the specified dates (i.e. for the previous five days in the case of the five-day accumulations).

Example of file format for basin 01608500:

- For 1 day accumulations:

01608500_cal_1D.dat - Bloc-notes

Fichier Edition Format Affichage ?

DATE	PRECIP	OBS_Q	swb	gr4j	isba	noah	sac	swap	vic(mm/day)
19620101	0.60	0.47	0.48	0.95	0.32	0.29	0.36	0.51	0.36
19620102	0.19	0.47	0.47	0.84	0.15	0.25	0.32	0.46	0.29
19620103	0.06	0.44	0.47	0.75	0.39	0.28	0.30	0.37	0.28
19620104	0.00	0.41	0.87	0.67	0.15	0.30	0.37	1.13	0.36
19620105	0.00	0.44	1.06	0.61	0.14	0.32	0.98	1.81	0.33
19620106	11.28	0.53	3.23	1.60	1.60	1.17	2.24	1.37	1.35
19620107	15.00	2.79	6.08	4.46	1.79	1.73	4.83	8.80	2.54
19620108	0.03	4.08	3.63	3.43	1.25	1.21	4.98	4.46	1.88
19620109	1.15	2.45	0.71	2.22	1.60	0.76	2.49	1.81	1.43
19620110	0.60	1.64	0.70	1.69	1.33	0.53	1.19	1.66	1.02
19620111	0.03	1.10	0.69	1.35	0.97	0.37	0.68	1.83	0.61
19620112	0.07	0.82	0.68	1.12	0.67	0.28	0.50	1.32	0.41
19620113	0.00	0.76	0.67	0.96	0.56	0.23	0.42	0.93	0.39
19620114	0.00	0.67	0.67	0.84	0.28	0.22	0.38	0.81	0.37
19620115	0.10	0.71	0.69	0.74	0.15	0.23	0.38	0.71	0.35
19620116	11.67	0.82	1.32	1.86	1.60	0.86	0.78	3.70	1.47
19620117	0.00	0.85	1.30	1.90	0.14	0.90	1.32	2.21	1.15
19620118	0.00	0.67	0.71	1.29	0.14	0.32	1.09	1.07	0.93
19620119	0.02	0.58	0.70	1.07	0.14	0.25	0.60	0.89	0.77
19620120	3.44	0.56	0.69	1.15	0.56	0.28	0.42	1.03	0.78

(...)

- For 5-day accumulations:

01608500_cal_5D.dat - Bloc-notes

Fichier Edition Format Affichage ?

DATE	PRECIP	OBS_Q	swb	gr4j	isba	noah	sac	swap	vic(mm/5 day)
19620105	0.85	2.23	3.35	3.82	1.15	1.44	2.33	4.28	1.62
19620110	28.06	11.49	14.35	13.40	7.57	5.40	15.73	18.10	8.22
19620115	0.20	4.06	3.40	5.01	2.63	1.33	2.36	5.60	2.13
19620120	15.13	3.48	4.72	7.27	2.58	2.61	4.21	8.90	5.10
19620125	10.46	3.14	9.67	6.30	1.83	3.56	8.61	6.83	4.68
19620130	5.47	3.63	4.78	4.69	1.64	3.14	4.98	4.10	4.12
19620204	10.31	2.78	5.27	4.54	2.07	2.16	2.76	4.14	4.27
19620209	2.15	3.61	4.86	3.38	1.00	2.78	4.90	3.10	4.07
19620214	15.71	2.47	4.29	4.70	2.76	1.85	1.50	1.48	3.47
19620219	7.93	2.12	5.17	4.01	1.78	2.64	2.80	1.75	6.77
19620224	24.13	4.89	12.24	5.60	4.09	4.35	11.21	11.39	8.02
19620301	36.67	29.80	22.15	21.81	17.78	12.81	25.17	29.10	21.59
19620306	34.54	10.73	6.44	15.11	14.83	8.15	6.53	9.89	15.56
19620311	20.02	5.40	5.79	25.20	18.59	11.71	1.74	4.31	14.32
19620316	17.98	26.25	17.99	13.72	26.14	24.28	15.62	34.04	33.61
19620321	21.95	20.30	16.70	6.84	15.67	22.57	13.49	19.46	18.81
19620326	14.96	37.57	21.17	13.97	18.36	32.89	31.48	22.48	23.69
19620331	2.80	10.35	6.59	4.04	7.39	19.48	4.63	5.22	10.77
19620405	13.93	7.61	9.54	4.53	7.33	14.23	9.20	4.76	10.70
19620410	31.15	9.13	15.48	7.69	13.02	14.40	13.48	10.21	14.15

(...)

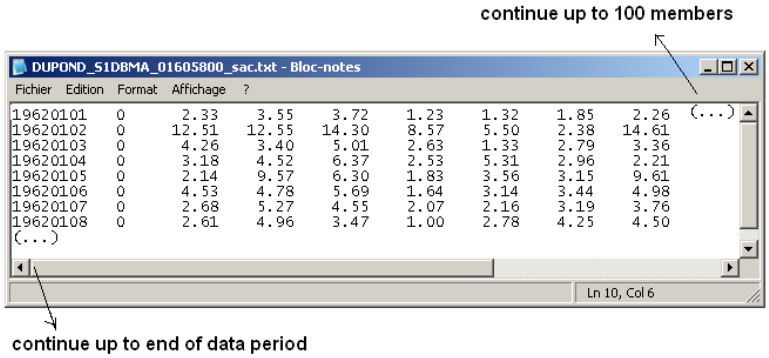
2. Results of post-processing (output data to be uploaded by 1st November 2012)

- **Summary:** The application of a post-processor technique should result in 100 equally likely ensemble members for each of the 12 river basins in the experimental dataset, and over the validation periods adopted in the scenario used by the participant.
- **Location to upload the data: ftp.hydro.washington.edu/incoming/nathalie/TEMP (anonymous ftp).** Note that, following upload, it may not be possible to edit/delete the uploaded files. See the contact information below for any questions or issues that arise during the upload. **Also, please email the organizers once your files have been uploaded.**
- **File to upload:** for each scenario performed, all output data generated (ensemble predictions) and its corresponding information form should be uploaded in a unique zipped file. The name of this zipped file should follow the convention below:
  - **File name**: **"IDparticipant_IDscenario.zip"**, where
    - **IDparticipant** is a 6-character ID assigned by participants to identify their submission (participants choice)
    - **IDscenario** is a 6-character ID assigned by participants to identify their scenario (the same Scenario ID indicated in the information form)
  - **Inside "IDparticipant_IDscenario.zip"**, there should be the information form (Word or PDF file) and all the *.txt files corresponding to the output of the post-processor.
- **File name for the output of the post-processor:**
  - **File name**: **"IDparticipant_IDscenario_IDbasin_IDmodel.txt"**, where
    - **IDparticipant** is the a 6-character ID assigned by participants to identify their submission (participants choice) (same above)
    - **IDscenario** is the 6-character ID assigned by participants to identify their scenario (the same Scenario ID indicated in the information form) (same above)
    - **IDbasin** is the 8-character basin ID number (see below)
    - **IDmodel** is the 3-character model ID (see below)
- **Data format of each *.txt output file**
  - The files should be in ASCII format, with one ASCII file for each basin and model considered in a given scenario (i.e. do not mix multiple locations or models in a single file).
  - By rows, a separate data record for each time step of the prediction period. By columns, the 100 members of the ensemble prediction for the corresponding time step. No header should be added.
  - In each row, each data record should have the following field structure, with the fields being either tab or space delimited:

Col 1: date (in the format yyyymmdd). This is the date for which the predictions apply
Col 2: 0 *
Col 3 to 102: one data value for each ensemble member (100 in total) (mm/1 day or mm/5 days) **

* column 2 should be all zeros to indicate the forecast lead time (= "0" for hydrological simulations).
** use -999 to mark any missing data.

Example file format submitted for one scenario: basin 01605800; model sac; daily time step:



3. ID of the river basins and complementary information:

IDbasin	Long	Lat	Area (sq.mi.)	Flood flow (mm/day)	Station Name
01608500	-78.6544	39.4469	1471	12.9849262	S BRANCH POTOMAC RIVER NR SPRINGFIELD, WV
01643000	-77.3661	39.3869	817	15.6225659	MONOCACY R AT JUG BRIDGE NR FREDERICK, MD
01668000	-77.5181	38.3222	1596	41.1607531	RAPPAHANNOCK RIVER NR FREDERICKSBURG, VA
03054500	-80.0403	39.1500	916	26.4065827	TYGART VALLEY RIVER AT PHILIPPI, WV
03179000	-81.0106	37.5439	394	13.3657759	BLUESTONE RIVER NR PIPESTEM, WV
03364000	-85.9256	39.2000	1707	8.78444131	EAST FORK WHITE RIVER AT COLUMBUS, IND.
03451500	-82.5786	35.6092	945	20.2125148	FRENCH BROAD RIVER AT ASHEVILLE, N. C.
05455500	-91.7156	41.4664	573	6.69810912	ENGLISH RIVER AT KALONA, IA
07186000	-94.5661	37.2456	1164	13.6491437	SPRING RIVER NEAR WACO, MO
07378500	-90.9903	30.46389	1280	17.2933884	AMITE RIVER NEAR DENHAM SPRINGS, LA
08167500	-98.3833	29.8603	1315	29.1185620	GUADALUPE RIVER NR SPRING BRANCH, TX
08172000	-97.6506	29.6661	838	6.07112566	SAN MARCOS RIVER AT LULING, TX

4. ID of the models:

IDModel	Note
cmb	If a combination (e.g. average) of models is used (details should be given in the information form)
sac	Sacramento model
grj	GR4J model
isb	ISBA model
vic	VIC model
noa	NOAH model
swb	Simple Water Balance model
swa	SWAP model

Contact:

Any questions or suggestions should be addressed to:

Dr. James Brown (james.brown@hydrosolved.com), or
Dr. Nathalie Voisin (Nathalie.Voisin@pnnl.gov), or
Dr. Maria-Helena Ramos (maria-helena.ramos@irstea.fr).

Thank you for participating in the
HEPEX intercomparison of streamflow post-processors!