

RAPID I/O files

By Cédric H. David (cedric.david@jpl.nasa.gov),
25 Oct 2010, updated 21 Jan 2015, 20 Dec 2016

Introduction

This document gives a brief explanation of the files used as input and output (I/O) in RAPID. RAPID can be used in two modes: regular simulation and optimization of model parameters. The regular simulation computes river flow based on runoff. The optimization procedure allows finding best model parameters based on a comparison between model-computed stream flow and measured flow at gaging stations, while varying model parameters.

A case-study of the San Antonio and Guadalupe Basins is used here as an example. Similar files are necessary to use RAPID in other domains. For the San Antonio and Guadalupe Basins study, the river network has 5,175 river reaches and RAPID is run for four years (2004-2007).

About file types

Most RAPID I/O files are coma-separated-variable (.csv) files. These files can be easily opened in a text editor, or using a spreadsheet program. The two largest files of RAPID I/O are netCDF (.nc) files. These can be read with Fortran (if the netCDF library is installed) as done in RAPID, but there are many other ways of reading netCDF files (see <http://www.unidata.ucar.edu/software/netcdf/>). Regular ASCII files were previously used in RAPID instead of netCDF, but the reading and writing performance of netCDF is much better for large scientific datasets.

Input files required for both regular simulation and for optimization of model parameters

rapid_connect_file (for ex: rapid_connect_San_Guad.csv)

This file provides information on the river network and its connectivity. Multiple fields are available, all of same size (the number of river reaches in domain), each entry corresponds to one river reach. The first field is a list of unique IDs of all river reaches in a basin. The second field is the ID of the unique reach located downstream in the river network. The third field is the number of upstream reaches, with a maximum set at runtime. The remaining fields are the IDs of the upstream reaches. All fields have no units. The relative ordering of river reaches in the first field is unimportant, but all other fields must correspond to the first field.

In the example given, this file is based on the NHDPlus dataset (particularly the NHDFlowline Value Added Attribute table).

Vlat_file (for ex: m3_riv_San_Guad_2004_2007_cst.nc)

This file contains the inflow of water (as a volume in m³) coming from outside of the river network (land and aquifers) and going into the river network, for each river reach and for each 3-hourly time step. This

is usually computed based on the outputs of a land surface model and/or a groundwater model (using surface and sub-surface runoff computations). This volume is stored in a netCDF file with two dimensions: ID and time. The ordering of reach IDs and the size are the same as in `rapid_connect_file`. In the example given, this file is based on outputs from the Noah-MP model.

riv_bas_id_file (for ex: `basin_id_San_Guad_hydroseq.csv`)

This file has the list of IDs (no units) of river reaches used when the model runs. It can either be the entire domain defined in `rapid_connect_file`, or a smaller basin within the domain. Any order can be used for this list of IDs. However, the ordering can influence the speed at which equations are solved. Upstream-to-downstream ordering provides the best performance.

In the example given, the entire domain is used and the ordering is going from upstream to downstream.

for_tot_id_file (for ex: `forcingtot_id_dam_springs.csv`)

This file has the list of IDs where the modeler can use observations instead of modeled flow for given locations. Forcing can be used at all, some or none of these locations, as chosen by the modeler. If forcing is used, the river network connectivity is broken downstream of the corresponding observed river reach (to avoid double-counting of flow) and the observed flow is used as upstream inflow for the downstream reach. The flow at the observed reach remains unchanged to allow for comparison.

Qfor_file (for ex: `Qfor_dam_springs_2004_2007.csv`)

This file has the daily observed flow rates (in m^3/s) corresponding to `forcingtot_id_file`. The ordering and size are the same as in `forcingtot_id_file`.

for_use_id_file (for ex: `forcinguse_id_dam_springs.csv`)

This is the list of observed river IDs where the modeler wants to use the forcing option (observations at a given reach replace the inflow for the downstream reach). Any order can be used here.

Input files required for regular simulation only

k_file (for ex: `k_San_Guad_2004_3.csv`)

This is a vector of parameters k used in the Muskingum method, one value for each river reach. These values can be computed outside of RAPID, or an outcome of the optimization procedure. The parameter k is a time (in seconds). The ordering is the same as `rapid_connect_file`.

x_file (for ex: `x_San_Guad_2004_3.csv`)

This is a vector of parameters x used in the Muskingum method, one value for each river reach. These values can be computed outside of RAPID, or an outcome of the optimization procedure. The parameter x is dimensionless. The ordering is the same as `rapid_connect_file`.

Input files required for optimization of model parameters only

obs_tot_id_file (for ex: `gage_id_San_Guad_2004_2007_full.csv`)

This is the list of reaches where observations are available.

Qobs_file (for ex: Qobs_San_Guad_2004_2007_full.csv)

This file has the daily observed flow rates (in m^3/s) corresponding to gage_id_file. The ordering and size are the same as in gage_id_file.

obs_use_id_file (for ex: gage_id_San_Guad_2004_2007_full.csv)

This is the list of reaches where observations are used during the optimization procedure.

kfac_file (for ex: kfac_San_Guad_1km_hour.csv)

This is a vector of first guesses for the parameters k (in seconds), used as the initial value during the optimization procedure. The ordering is the same as rapid_connect_file.

Output files for regular simulation

Qout_file (for ex: Qout_San_Guad_2004_2007.nc)

This is a netCDF file. m^3/s . This flow rate is stored in a netCDF file with two dimensions: reach ID and time. The ordering is the same as basin_id_file.

Output files for optimization

The optimization procedure does not generate outputs inside a file. All outputs are printed directly on the terminal (standard output, also known as stdout), as the program runs. However, one may want to save the information by redirecting standard output into a file, using the > operator.

Summary table

File name	Used in regular simulation	Used in optimization procedure	Number of fields	Units	Requirement on number of entries per field	Requirement on ordering	Time variant?	Temporal resolution
rapid_connect_file	yes	yes	varies	none	none	none	no	
Vlat_file	yes	yes	1	m^3	same as rapid_connect_file	same as rapid_connect_file	yes	3-hourly
riv_bas_id_file	yes	yes	1	none	smaller than rapid_connect_file	none (upstream-to-downstream helps performance)	no	
for_tot_id_file	optional	optional	1	none	none	none	no	
Qfor_file	optional	optional	1	m^3/s	same as for_tot_id_file	same as for_tot_id_file	yes	daily
for_use_id_file	optional	optional	1	none	smaller than for_tot_id_file	none	no	
k_file	yes	no	1	s	same as rapid_connect_file	same as rapid_connect_file	no	
x_file	yes	no	1	none	same as rapid_connect_file	same as rapid_connect_file	no	
obs_tot_id_file	no	yes	1	none	none	none	no	
Qobs_file	no	yes	1	m^3/s	same as obs_tot_id_file	same as obs_tot_id_file	yes	daily
obs_use_id_file	no	yes	1	none	smaller than obs_tot_id_file	none	no	
kfac_file	no	yes	1	s	same as rapid_connect_file	same as rapid_connect_file	no	
Qout_nc_file	yes	no	1	m^3/s	same as riv_bas_id_file (automatic in RAPID)	same as riv_bas_id_file (automatic in RAPID)	yes	3-hourly

Sizes in example given

Example file name	Number of entries per field	Number of time steps
rapid_connect_San_Guad.csv	5175	
m3_riv_San_Guad_2004_2007_cst.nc	5175	1460x8
basin_id_San_Guad_hydroseq.csv	5175	
forcingtot_id_dam_springs.csv	3	
Qfor_dam_springs_2004_2007.csv	3	1460
forcinguse_id_dam_springs.csv	3	
k_San_Guad_2004_3.csv	5175	
x_San_Guad_2004_3.csv	5175	
gage_id_San_Guad_2004_2007_full.csv	36	
Qobs_San_Guad_2004_2007_full.csv	36	1460
gage_id_San_Guad_2004_2007_full.csv	36	
kfac_San_Guad_1km_hour.csv	5175	
Qout_San_Guad_2004_2007.nc	5175	1460x8

Further information

RAPID website: <http://rapid-hub.org/>

RAPID source code: <https://github.com/c-h-david/rapid/>