How to prepare RAPID inputs from Version 1 of NHDPlus

By Cédric H. David (cedric.david@jpl.nasa.gov),
13 Jun 2014, updated 21 Jan 2015

Goal

Learn how to create some of the files that are necessary to run RAPID from Version 1 of the enhanced National Hydrographic Dataset (NHDPlus v1) using ArcGIS.

Requirements

The ArcGIS software, and basic knowledge on how to use it. An Internet connection.

Introduction

RAPID needs to know where river reaches are and how they are connected before it can route water through them. To do so, RAPID uses two routing parameters (Muskingum k and x) for each river reach. Also, the automated calibration of RAPID uses an estimate of parameter k which is necessary when running in optimization mode. In RAPID simulations, these files all need to use the same sorting for river reaches and are traditionally named as follows (where "Domain" is the name of the river basin considered):

- rapid_connect_Domain.csv.
- k_Domain.csv.
- x_Domain.csv.
- kfac_Domain.csv.

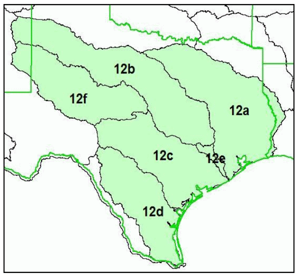
Finally, because RAPID can run in a subset of the entire domain where data is available, a list of river reaches in this subset is needed. The sorting of this file is independent from that the previous files, and an upstream-to-downstream sorting is valuable here although not mandatory.

- riv_bas_id_Domain.csv

The purpose of this tutorial is to build all the aforementioned files. Focus is here made on the combination of the Guadalupe and San Antonio Basins in Texas which were both used in the first RAPID publication (David et al., 2011, Journal of Hydrometeorology).

Download NHDPlus data for the Guadalupe and San Antonio Basins

Data corresponding to Version 1 of NHDPlus are available at http://www.horizon-systems.com/NHDPlus/NHDPlusV1_data.php. The Guadalupe and San Antonio Basins are located in the Texas Gulf Coast Hydrologic Region which is Region 12 in NHDPlus:



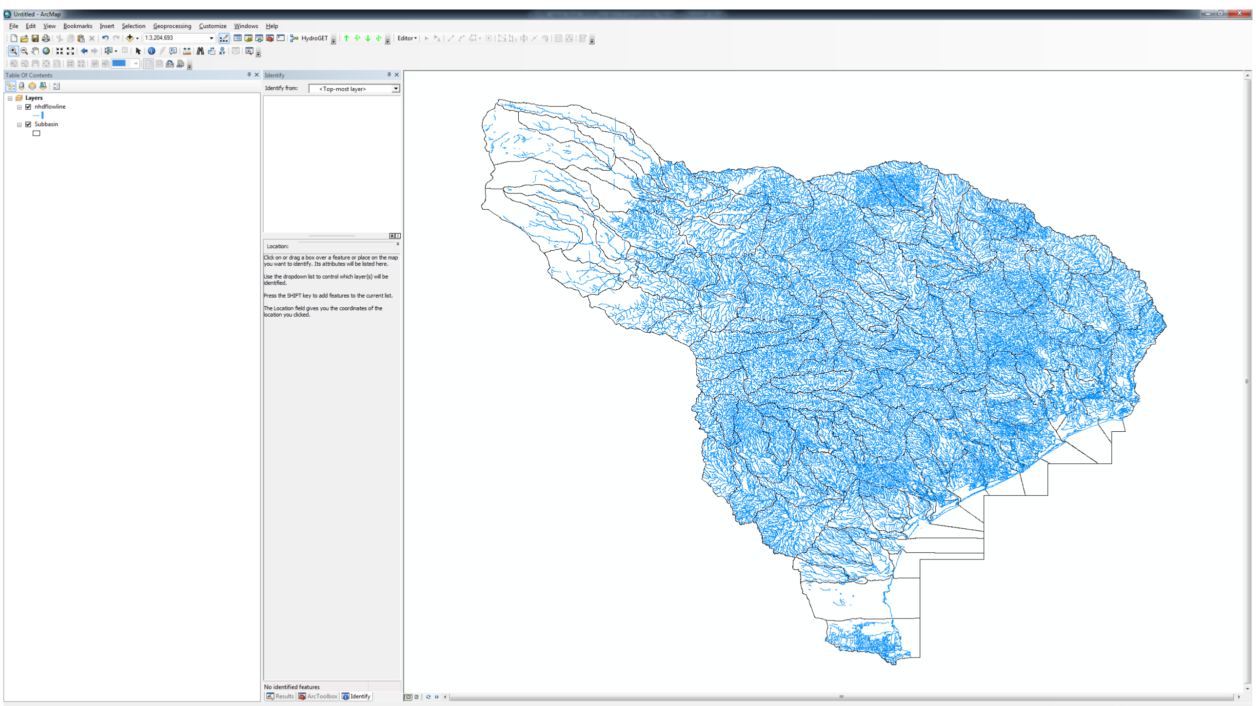
The files necessary for this tutorial can be downloaded by clicking on the following link in the aforementioned webpage:

Region 12, Version 01_01, National Hydrography Dataset	NHDPlus12V01_02_NHD	Shapefile and DBF

Of interest here will be three files included in the archive above: nhdflowline.shp, Subbasin.shp and NHDFlowlineVAA.dbf.

Select the study domain

Open the files nhdflowline.shp and Subbasin.shp in ArcGIS.



The Guadalupe and San Antonio Basins are identified by their 6-digit hydrologic unit codes from the classic USGS classification (121002 and 121003, respectively). They include the following features of subbasin.shp (8-digit HUCs):

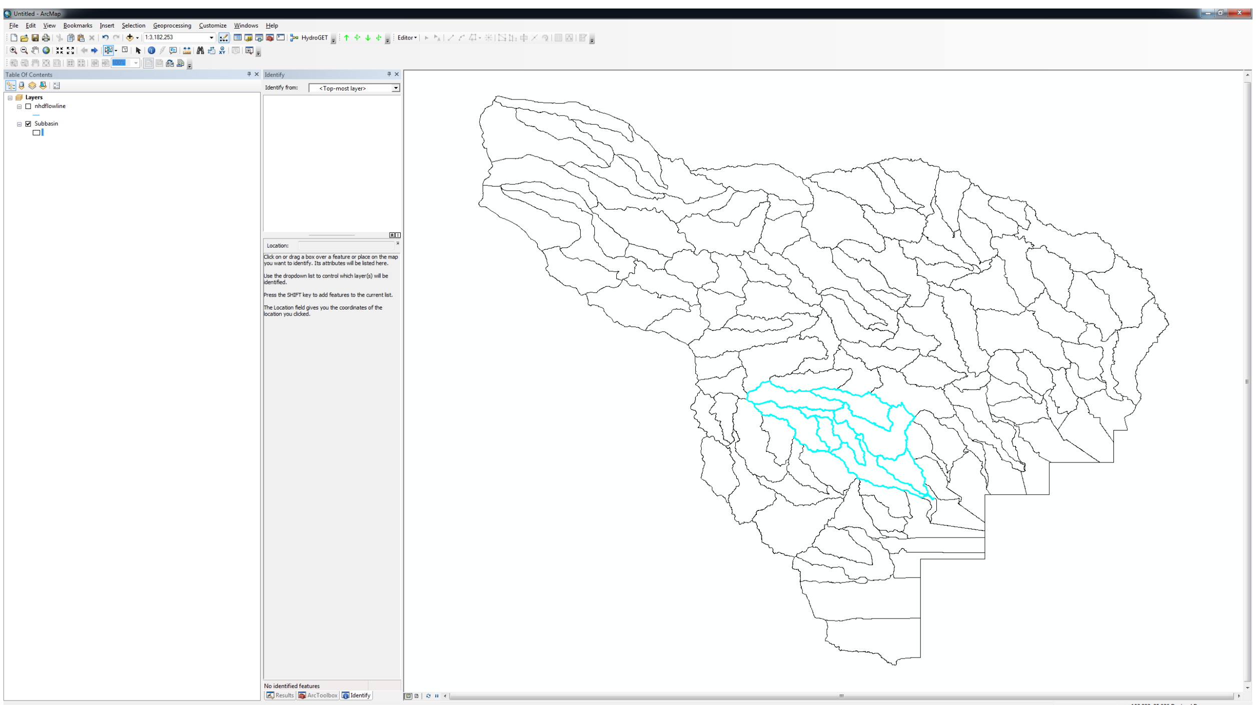
Table

Subbasin

FID	Shape	OBJECTID	SOURCE	HUC_8	HU_8_NAME	SHAPE_LENG	SHAPE_AREA
88	Polygon	1	1	12100201	Upper Guadalupe. Texas.	4.008095	0.349434
89	Polygon	1	1	12100202	Middle Guadalupe. Texas.	5.065719	0.517164
90	Polygon	1	1	12100203	San Marcos. Texas.	3.84642	0.330373
91	Polygon	1	1	12100204	"12100301	3.027367	0.25007
92	Polygon	1	1	12100301		1.932038	0.123643
93	Polygon	1	1	12100302	Medina. Texas.	4.067297	0.325432
94	Polygon	1	1	12100303	Lower San Antonio. Texas.	5.276658	0.352218
95	Polygon	1	1	12100304	Cibolo. Texas.	3.990906	0.205831

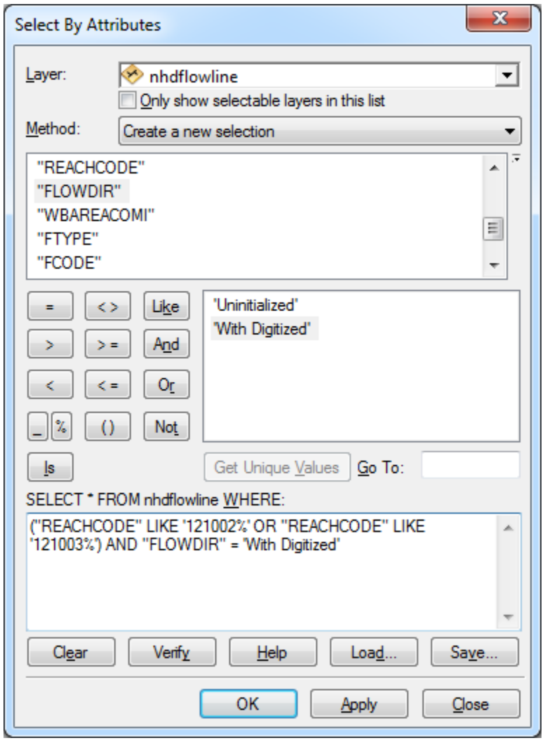
1 (8 out of 122 Selected)

Subbasin

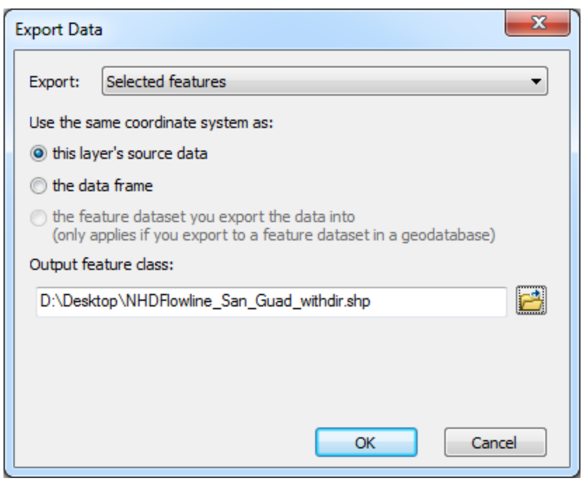


RAPID can only simulate the river reaches where flow direction is known and hence the corresponding reaches need be determined. This can be done in NHDPlus by selecting features based on "FLOWDIR"='With Digitized'. The river reaches located in the Guadalupe and San Antonio Basins can be selected by using the first 6 digits of their REACHCODE with a statement similar to: "REACHCODE" LIKE '121002%'.

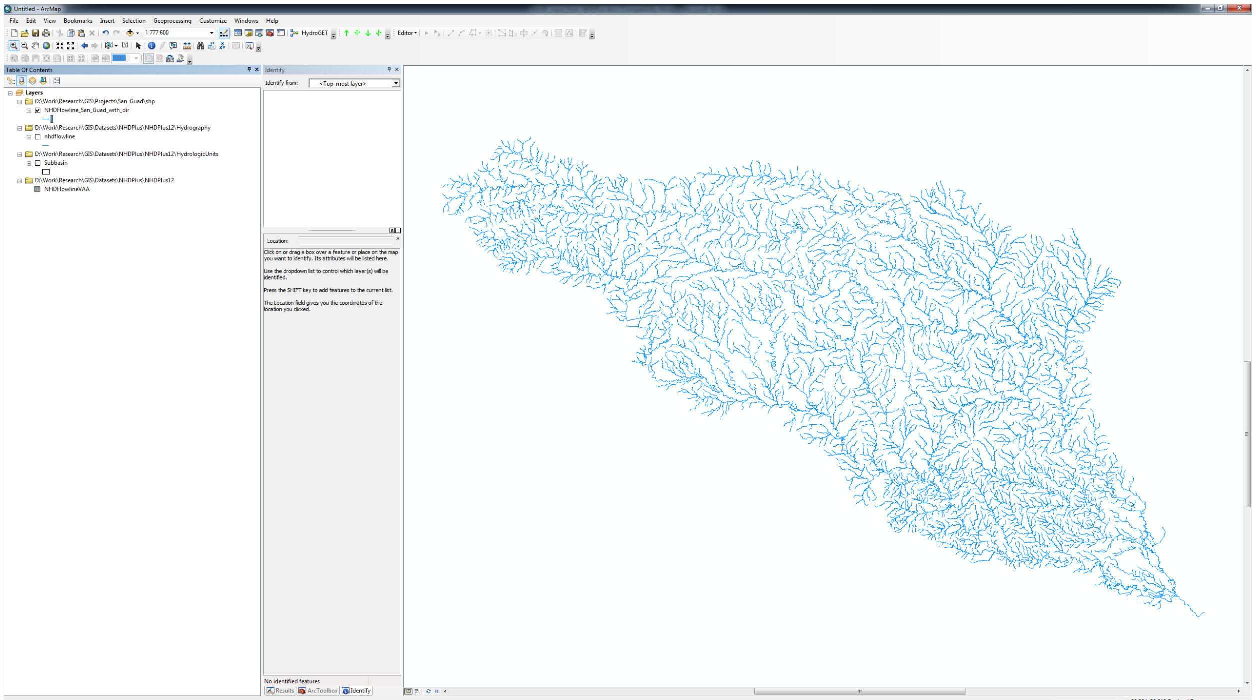
Use the Select by Attributes tool to determine the reaches located in the Guadalupe and San Antonio Basins that have a known flow direction. The corresponding statement is ("REACHCODE" LIKE '121002%' OR "REACHCODE" LIKE '121003%') AND "FLOWDIR" = 'With Digitized'.



Now export the corresponding reaches in a file called NHDFlowline_San_Guad_with_dir.shp.

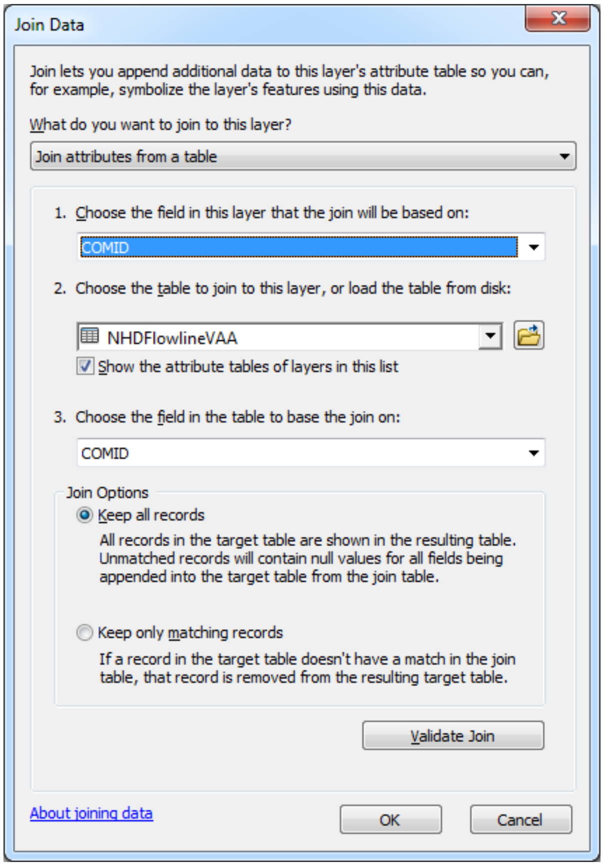


The file NHDFlowline_San_Guad_with_dir.shp should contain only 5,175 of the 74,615 features originally available in nhdflowline.shp.

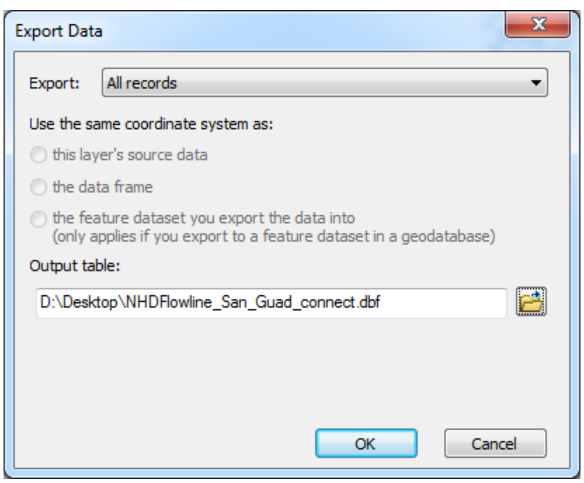


Create a table with information related to the river reaches of interest

Create a join between NHDFlowline_San_Guad_with_dir.shp and NHDFlowlineVAA.dbf (available as part of the NHDPlus archive downloaded earlier in this tutorial).



Now export this table as NHDFlowline_San_Guad_connect.dbf:



And delete all fields except for:

Table

NHDFlowline_San_Guad_connect

OID	COMID	LENGTHKM	FROMNODE	TONODE	HYDROSEQ	DIVERGENCE
0	161957	2.362	2775300047	2775300316	2776000255	0
1	161957	4.556	2775300002	2775300317	2776000246	0
2	161957	6.666	2775300003	2775300317	2776000247	0
3	161957	1.264	2775300368	2775300318	2776000253	0
4	161957	0.314	2775300318	2775300319	2776000252	0
5	161958	0.475	2775300830	2775300319	2776000256	0
6	161958	1.979	2775300317	2775300320	2776000245	0
7	161958	0.652	2775300320	2775300322	2776000244	0
8	161958	1.393	2775300321	2775300320	2776000248	0
9	161958	5.786	2775300013	2775300321	2776000249	0
10	161959	4.965	2775300014	2775300322	2776000243	0
11	161959	0.335	2775300322	2775300324	2776000242	0
12	161959	2.336	2775300319	2775300323	2776000251	0
13	161959	1.734	2775300323	2775300321	2776000250	0
14	161959	6.382	2775300020	2775300324	2776000241	0
15	161960	3.516	2775300324	2775300325	2776000240	0
16	161960	7.504	2775300023	2775300325	2776000239	0
17	161960	6.707	2775300024	2775300326	2776000237	0
18	161960	1.34	2775300325	2775300326	2776000238	0
19	161960	2.187	2775300326	2775300327	2776000236	0
20	161961	1.356	2775300362	2775300328	2776000231	0
21	161961	1.635	2775300327	2775300332	2776000228	0
22	161961	2.923	2775300033	2775300329	2776000225	0
23	161961	5.826	2775300328	2775300327	2776000229	0
24	161961	0.139	2775300331	2775300330	2776000234	0
25	161962	3.098	2775300332	2775300329	2776000226	0
26	161962	2.217	2775300039	2775300331	2776000235	0
27	161962	3.821	2775300040	2775300332	2776000227	0
28	161962	5.753	2775300041	2775300333	2776000116	0

1 (0 out of 5175 Selected)

NHDFlowline_San_Guad_connect

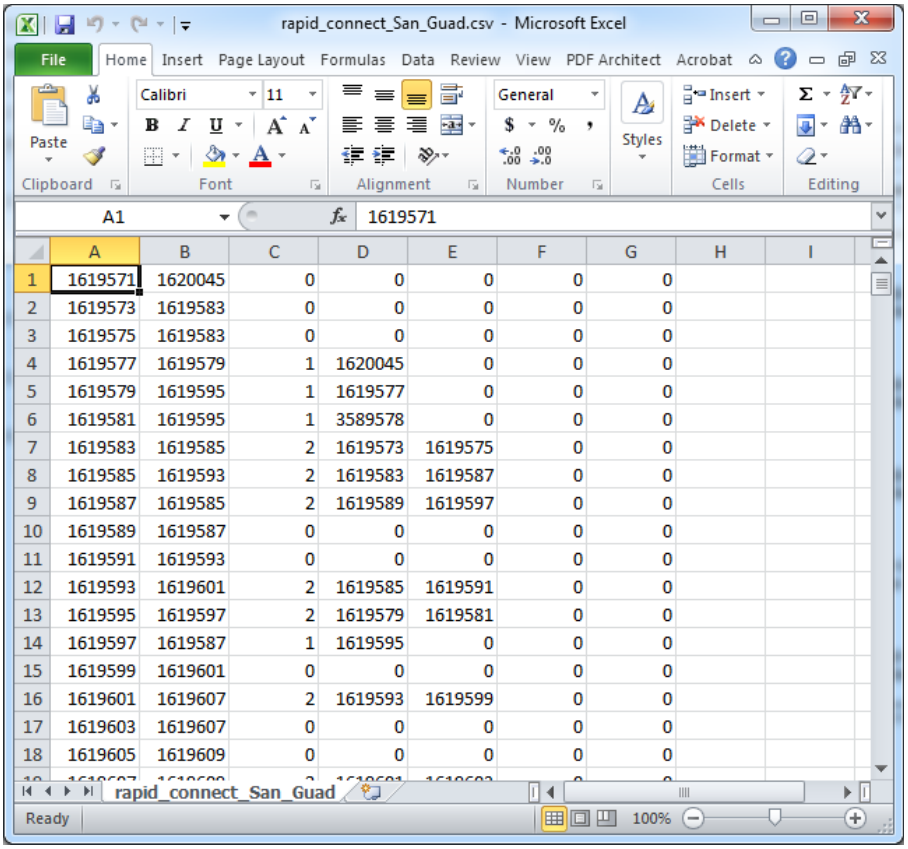
Create the RAPID connectivity file

The number of rows in the RAPID connectivity file corresponds to the number of river reaches in the river network (here 5,175). Note that there is no header row. The number of columns is 3+max_up where max_up is the maximum number of upstream reaches for any reach in the river network. The columns are as follows:

- RIV_ID. The unique integer identifying a given river reach.
- DOWN_ID. The unique integer identifying the unique reach located directly downstream of RIV_ID.
- NB_UP. The number of river reaches located directly upstream of RIV_ID.
- UP1. The unique integer of a reach (if it exists) that is located directly upstream of RIV_ID.
- UP2. The unique integer of a reach (if it exists) that is located directly upstream of RIV_ID and that differs from UP1.
- UP3. The unique integer of a reach (if it exists) that is located directly upstream of RIV_ID and that differs from UP1 and UP2.
- UP4. The unique integer of a reach (if it exists) that is located directly upstream of RIV_ID and that differs from UP1, UP2 and UP3.
- etc.

Note that **any sorting can be used for sorting RIV_ID values in the first column**. However, because **the sorting needs be consistent among many RAPID input files**, RIV_ID values **will be sorted here in increasing order**.

A Python script that uses the file NHDFlowline_San_Guad_connect.dbf as input can allow creating the RAPID connectivity file rapid_connect_San_Guad.csv. One can do so by knowing that a given river reach i flows into reach j if the TONODE attribute of the former is the FROMNODE attribute of the latter. Note that this connectivity between upstream and downstream shall only be kept if the downstream reach has a DIVERGENCE attribute value of 1 or 0. This is because a DIVERGENCE attribute value of 2 indicates that a reach is part of a minor divergence and **RAPID can currently only accommodates one unique downstream reach for each river reach**. However, the DIVERGENCE attribute value of the upstream element is not to be checked otherwise risking that reaches that are part of minor divergences do not flow anywhere.



	A	B	C	D	E	F	G
1	1619571	1620045	0	0	0	0	0
2	1619573	1619583	0	0	0	0	0
3	1619575	1619583	0	0	0	0	0
4	1619577	1619579	1	1620045	0	0	0
5	1619579	1619595	1	1619577	0	0	0
6	1619581	1619595	1	3589578	0	0	0
7	1619583	1619585	2	1619573	1619575	0	0
8	1619585	1619593	2	1619583	1619587	0	0
9	1619587	1619585	2	1619589	1619597	0	0
10	1619589	1619587	0	0	0	0	0
11	1619591	1619593	0	0	0	0	0
12	1619593	1619601	2	1619585	1619591	0	0
13	1619595	1619597	2	1619579	1619581	0	0
14	1619597	1619587	1	1619595	0	0	0
15	1619599	1619601	0	0	0	0	0
16	1619601	1619607	2	1619593	1619599	0	0
17	1619603	1619607	0	0	0	0	0
18	1619605	1619609	0	0	0	0	0

Create the parameter files

The number of rows in all RAPID parameter files corresponds to the number of river reaches in the river network (here 5,175) and these files only contain a unique column. A Python script that uses the file NHDFlowline_San_Guad_connect.dbf as input can allow creating all RAPID parameter files as follows:

- kfac_San_Guad.csv contains a time (s) which is computed based on the river reach length using a flow wave celerity of 1 km/h. This can be done using $kfac_j$ = LENGTHKM*3600.
- k_San_Guad.csv contains a time (s) which is computed based on the river reach length. There are two ways to compute k:
  - Based on a wave celerity c (m/s) and the river reach length: k_j = LENGTHKM*1000/c. A default celerity of c=0.8 m/s can be used.
  - Based on kfac_San_Guad.csv through the use of a multiplying factor $\lambda_k > 0$: $k_j = \lambda_k * kfac_j$. A default value of $\lambda_k = 0.35$ can be used.
- x_San_Guad.csv contains a non-dimensional number varying between 0 and 0.5. While a different value can be picked for each river reach, many existing RAPID applications use a unique value for x throughout the entire river network. This can be done by selecting $\lambda_x \in [0,5]$ and applying: $x_j = \lambda_x * 0.1$. The value $\lambda_x = 3$ can be used as default.

Create subset file

The number of rows in the subset file corresponds to the size of the river network included in the larger river network that will be used to run RAPID. Here, this means that the subset file can have anywhere between 1 and 5,175 (included) river reaches. While any sorting can be used for the subset file, computing time in RAPID can be greatly decreased if some upstream-to-downstream sorting is used. Such can be done in NHDPlus by sorting the HYDROSEQ attribute value in decreasing order.

In this example the subset file will include all 5,175 rivers of the Guadalupe and San Antonio Basins. The file riv_bas_id_San_Guad.csv can be created from NHDFlowline_San_Guad_connect.dbf by sorting RIV_ID such that the corresponding HYDROSEQ attribute value is in decreasing order.

Further information

RAPID website: http://rapid-hub.org/
RAPID source code: https://github.com/c-h-david/rapid/