This file is a quick guide for the computations done in the article 2018WR022675RR submitted to WRR journal.

The computations are divided in four part codes:

Part I- Data preparation

Part II- Computation of complexity for each basin

Part III- Post processes

Part IV- Graphical outputs

Part I:

In this part, data from MOPEX format is converted MATLAB format and the basins with less data gaps is chosen.

To start, the "Run_Readmpx2018.m" file should be run in MATLAB. The other scripts and functions would be used within this file. All MOPEX raw data format are in "RawMopexData" folder and the output files of this part (which is the input files for Part II) would be saved in "DataInputNew" folder.

If the raw input files were not available, please download all files with "*.dly" file extension accessible from the address below:

ftp://hydrology.nws.noaa.gov/pub/gcip/mopex/US_Data/Us_438_Daily/

Part II- The complexity and prediction of test data for each basin with the robust nearest neighbor algorithm are computed in FORTRAN. I used "Intel® Parallel Studio XE 2016" in "Visual Studio" environment as compiler. This algorithm was originally developed in:

Pande, Saket; Mac McKee, and Luis A. Bastidas "Complexity-based robust hydrologic prediction" Water resources research 45.10 (2009).

The results of that calculation are included in this script. It calculates robust prediction of runoff based on k-nearest neighbor algorithm presented in the second appendix of above article

In the folder "BasinRes", there are 413*5 txt files with a name for example like this pattern: "predictRBST_cat_12_4_1". This means the file is related to the results of the catchment ID number (this is our number, this number will be converted to the catchment ID, in order to identify it in the GIS file to show it on the map) is 12; there are 4 years for calibration (which is assumed constant for all the basins) and the file represents repeat number one out of 5 repeats (each repeat includes 730 time steps or 2 years as query). Each file has 730 rows and 6 columns. These 730 rows are our query points in the output space. Columns are: column 1 is “predicted runoff”, column 2 is “GENERAILIZED ERROR”, column 3 is “b-beta chosen”, column 4 is “input dimension chosen”, column 5 is “complexity for the query” and column 6 is “OBSERVED runoff”.

Important Note:

This step is the most time consuming part of computation. Thus, I divided it the running process of 413 into 5 separate compile and run of the FORTRAN routine. First, 1st-99th basin, second 100th-199th basin and so on. Here for example the 400th-413th basin calculations are available. As a result, the catchment number in output file names in this fashion is from 1-99 for the 4 first folder and 1-14 in the last one and should be renamed to 1-413 to be used for the rest of calculations.

The input files (from Part I) are in “DAT” folder.

Part III- Post processes on data such as choosing the indices for considering its complexity in the test data (by depth function). MATLAB and R

To run this part, the main program is “Run413.m” file which is ready to run.

The depth calculations have been done in R environment using ‘depth’ package which can be found here:

https://cran.r-project.org/web/packages/depth/depth.pdf

All results of three depth methods calculations were saved in three files: “DepthLiu.mat”, “DepthOja.mat” and “DepthTukey.mat”.

The appropriate threshold should be chosen according to the depth method in “D_Threshold” global variable in line 14 of the “Run413” file.

“LoadresultsFiles413.m” produces complexity calculation results. Instead, for avoiding the repeated calculations, “resuts413.mat” file contains these results.

“ReadCharecteristics.m” file, all basin characteristics are being load or calculated.

The results of all above steps are saved in a “X.txt” file.

Part IV- Graphical outputs and comparing complexity with basin properties are done.

The X.txt file includes a matrix with 7 columns: ‘Porosity’, ‘PSI’,’Mean Slope’, ‘GNDVI’, ‘P/PE’,’WRC’,’ slope of fdc’,’Complexity’.

Heat maps and figures are calculated in this part.