Assessing Water Budget Sensitivity to Precipitation Forcing Errors in Potomac River Basin Using VIC Model

Reetam Majumder[1], Redwan Walid[2], Jianyu Zheng[3]
RAs: Carlos Barajas[1], Pei Guo[2], Chamara Rajapakshe[3]
Mentors: Aryya Gangopadhyay[2], Matthias K. Gobbert[1], Jianwu Wang[2], Zhibo Zhang[3]
Clients: Kel Markert[5], Amita Mehta[4], Nagaraj K. Neerchal[1]

[1]Department of Mathematics and Statistics, UMBC
[2]Department of Information Systems, UMBC
[3]Department of Physics, UMBC
[4]Joint Center for Earth Systems Technology, UMBC
[5]University of Alabama Huntsville / NASA-SERVIR

Cybertraining: Big Data + High Performance Computing + Atmospheric Sciences
University of Maryland, Baltimore County, Spring 2019
cybertraining.umbc.edu

Acknowledgments: NSF, NASA-SEVRIR, UMBC, HPCF, CIRC, JCET

Outline

Motivation

- Ensemble-based data assimilation techniques are often used to estimate the components of water and energy balance
- VIC is an open source model that can simulate land surface-atmosphere fluxes of moisture and energy
- Precipitation forecasts is one of key inputs for the model, and hence VIC can be used to test the effect of precipitation uncertainties on water budget components
- An understanding of how sensitive the model is to precipitation uncertainties in turn can be used to compare precipitation models

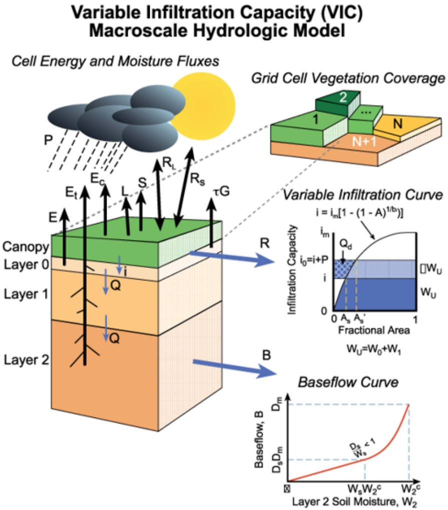
Problem Definition

- We want to run VIC for sub-basins within the Potomac river basin, and also integrate over the entire basin for April to September 2017
- We have two main goals:
    1. Analyze variations in monthly water budget components over sub-basins within the Potomac basin, and also integrated over the entire basin for April to September 2017. The intra-seasonal and seasonal variability can tell us contributions of various sub-basins to the river flow
    2. Conduct multiple model simulations by varying precipitation input based on error estimates in IMERG precipitation data and/or from various state-space models used to fit and predict precipitation, followed by an analysis of the water budget components to see how precipitation errors impact water budget in the basin

Outline

Model Description

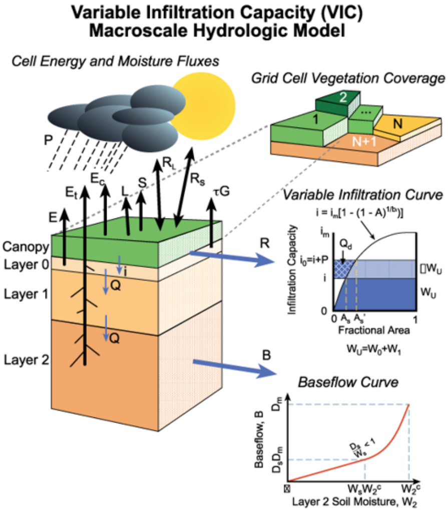
- Large-scale, semi distributed hydrological model
- Developed by Liang et al. [1994]
- Physically based model to be coupled with global circulation model (GCM) simulations
- Simulates land surface-atmospheric fluxes of moisture and energy
- Considered a research model; open source



https://arset.gsfc.nasa.gov/water/webinars/VIC18
Image from Open Access VIC Documentation

Model Features

- Each grid cell is simulated independently
- Only water entering cell is from atmosphere (precipitation)
- Can represent sub-grid vegetation/land cover
- Routing of streamflow is performed independently using a separate model typically the Lohmann et al. [1996; 1998] routing model



https://arset.gsfc.nasa.gov/water/webinars/VIC18
Image from Open Access VIC Documentation

Input data for VIC

- All datasets are downloaded for $76^\circ - 80^\circ W$, $37^\circ - 41^\circ N$, between Jan 2016 and Dec 2018
- VIC requires the following input data at each grid point:

Variable	Source	Temporal Resolution
Precipitation	GPM IMERG	Daily
Surface air temperature	MERRA-2	Daily
Surface wind speed	MERRA-2	Daily
Land cover type	MODIS	Yearly
Leaf Area Index	MODIS	Monthly
Surface Albedo	MODIS	Monthly
Initial soil moisture conditions	Precipitation climatology -	Static
Soil Characteristics	HWSD	Static
Annual Precipitation	IMERG	Static
Elevation	SRTM	Static

Output from VIC

- Our outputs of interest are the monthly water budget components:
  - Evapotranspiration
  - Surface runoff
  - Baseflow
- The model uses data from Jan 2016 to Mar 2017 for spin up, and daily water budget predictions are made for April through September 2017

Data Preparation Tools

- The various data sources must be collated and rearranged in a format such that at each grid point, the entire time series of all the inputs are available
- Additionally, a set of parameter files (global, soil, precipitation etc.) will provide VIC instructions on how to read in the data and what to do with it
- Pre-processing and post-processing scripts written in Python by Kel Markert are available on github. Some packages required are Python2 specific
- An HDF to GeoTIFF conversion tool required to process MODIS data has also been installed on taki; it re-projects MODIS data into WGS84 format
- Finally, the stable version of VIC written in C is also installed in the Team 1 directory

Outline

MODIS data reprojection using HEG

- MODIS data is available in sinusoidal grids ($10^\circ \times 10^\circ$ tile) in HDF file format, which needs to be converted into geographical (WGS84) projection
- HEG has a GUI but it is practically impossible to do the conversion for all the files that way. So we first generated a parameter file for each of Land Cover, LAI and SW Albedo from the GUI
- The parameter file generated for Land Cover can be used for all Land Cover files, and so on
- Using a bash script to loop over all files, HEG can be run from the command line using the parameter files we generated

  ```
  resample -p MyParameter.prm
  ```

  For longer runs this can be submitted as batch jobs via slurm

Other data aggregation steps

- Leaf Area Index (LAI) and Shortwave (SW) Albedo (from MODIS) are converted to monthly format - one value for each month for each grid point. Since they don't vary a lot, we can use monthly averaged data instead of daily data:
  - LAI has a native temporal resolution of 8 days; we use weighted averages to get monthly values
  - SW Albedo is available daily and is averaged to get monthly data
- Northward (v) and Eastward (u) wind component variables (from MERRA-2) are combined as wind speed by $\sqrt{u^2 + v^2}$ for each spatio-temporal component of the data

Preprocessing for VIC compatible inputs

This is an outline of the major data preprocessing steps required before VIC can be run:

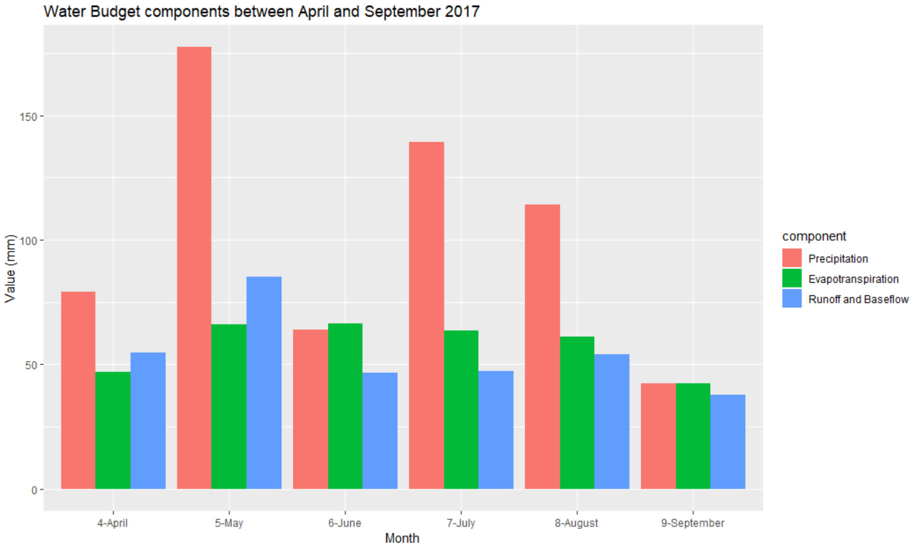
1. Create a grid template of $0.1^\circ \times 0.1^\circ$ resolution for the Potomac basin shapefile
2. Align all raster datasets to the grid template - SRTM, Elevation and Slope, IGBP land cover, soil and precipitation
3. Format input files to create parameter files which would be read by VIC:
   - Snow parameter file from on SRTM, which contains information on the number and structure of snow bands
   - Soil parameter file based on precipitation, elevation, slope and soil characteristics
   - Vegetation parameters and vegetation library based on IGBP LC and MODIS datasets
   - Meteorological forcings from MODIS, GPM IMERG and MERRA-2

Outline

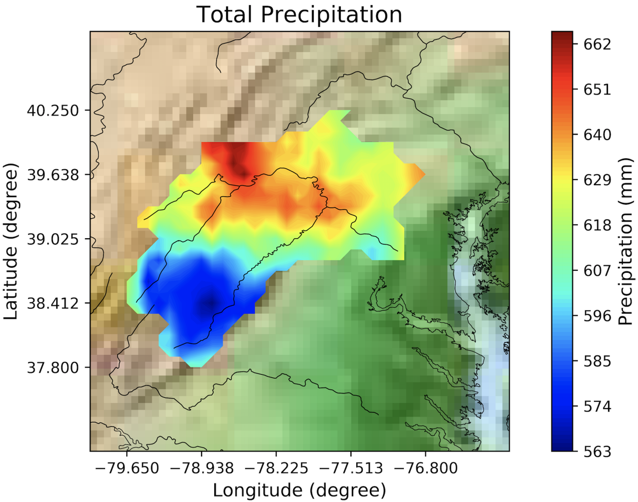
Outputs of interest

- Our primary interest is to compare input precipitation against the water balance between April and September 2017.
- Water Balance is measured as $Precipitation - Evapotranspiration - Runoff - Baseflow$
- Evapotranspiration, runoff and baseflow are available as daily output for each grid point from the model
- The water balance can then be converted into volume based on the area of the basin and sub-basins, and further analysis can be done on it
- However, since we have not had time to calibrate VIC, the exact values cannot be interpreted accurately and we have not pursued a volumetric analysis

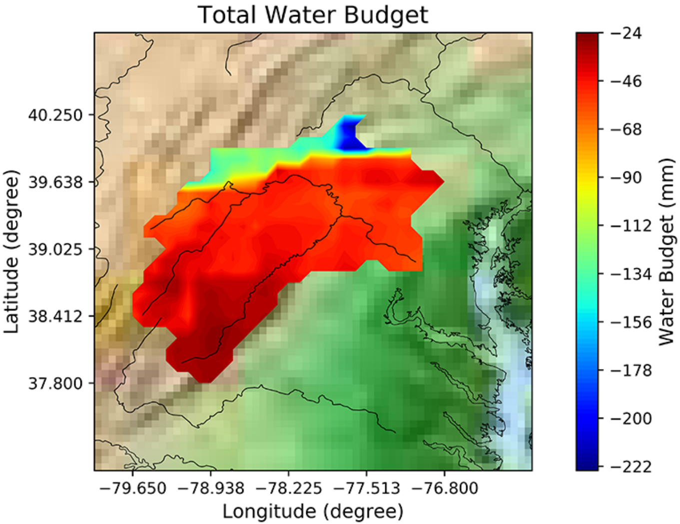
Water budget components between April and September 2017



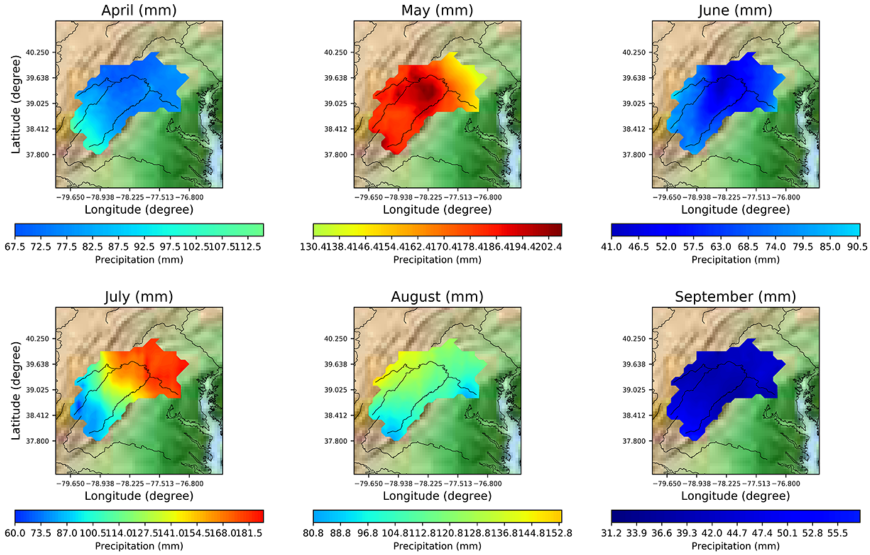
Total Precipitation between April and September 2017



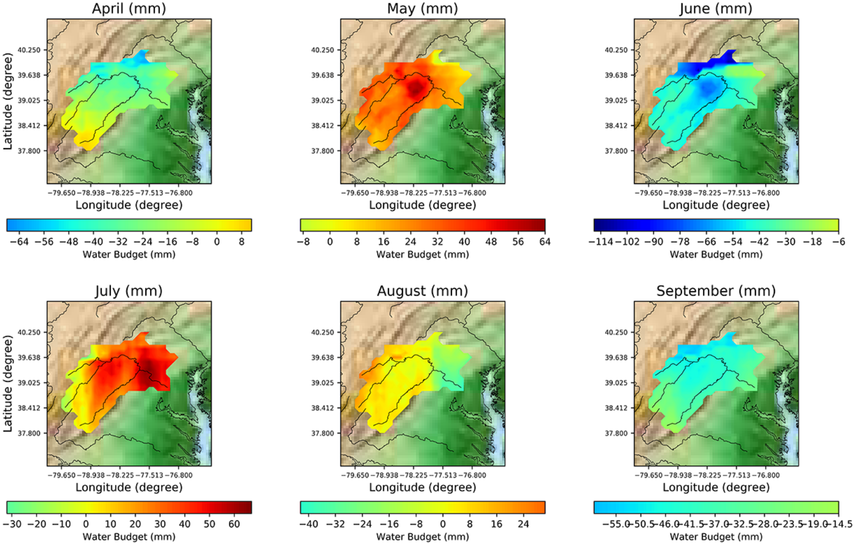
Water balance between April and September 2017



Monthly precipitation in the Potomac Basin



Monthly water balance in the Potomac Basic



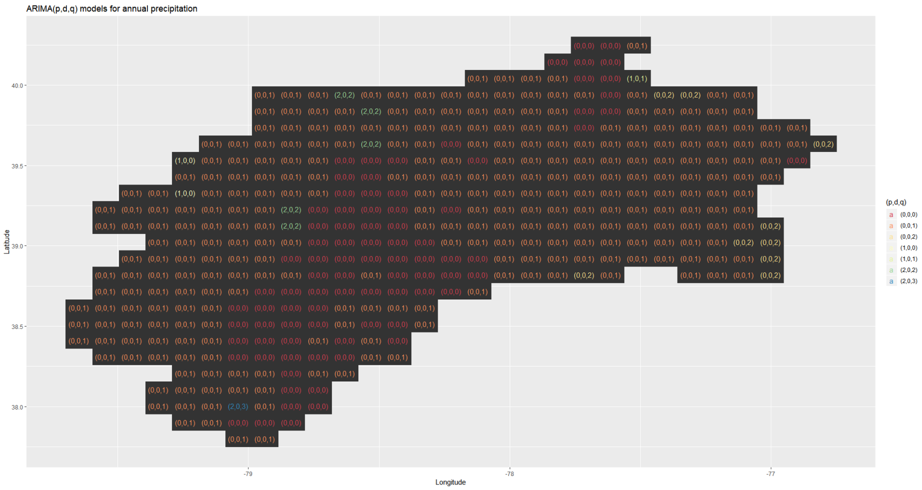
Outline

Spatio-temporal features of the data

- Our data consists of 387 grid points from Jan 2016 to Sep 2017
- Temporally, we found that for over 90% of all grid points there was either no temporal structure or can have an MA(1) model fitted to it; i.e every day's data is correlated with the innovation/error component of the previous day
- For each grid point, we assume the data to be independent and identical in distribution
- This assumption will not be valid if our data stretches back a few more years
- Spatially, we plotted correlograms and looked at the correlations between adjoining data points; for the majority of the data, this value was not high
- Due to time and data constraints we were unable to explore avenues like variogram analysis or looking at directional spatial dependence
- For the purposes of this analysis, we have assumed the data to have no spatial structure

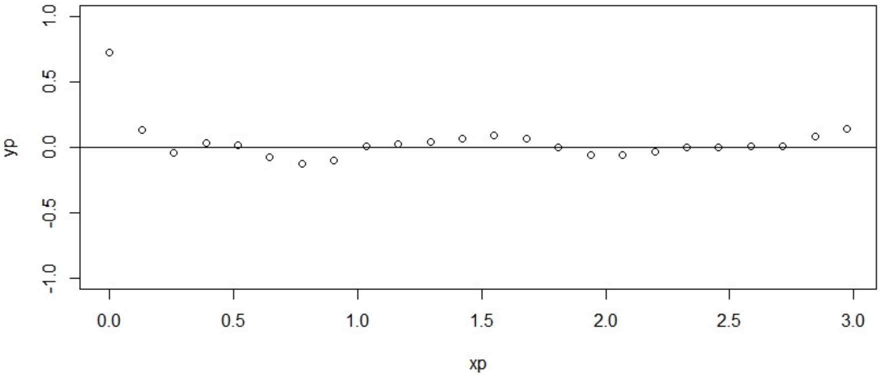
Temporal analysis

Figure: ARIMA(p,d,q) models for annual precipitation



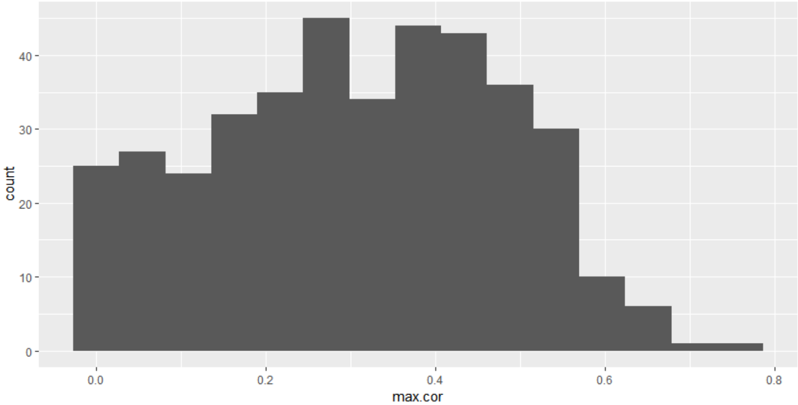
Spatial analysis

- Our main interest is in the second point of the correlogram which corresponds approxiately to the 0.1° distance between grid points
- This is the correlogram for Jan 29 2016 across all grid points. We will get the same value for each day and see how it behaves throughout our dataset



Spatial analysis

Figure: Histogram of spatial correlation between grid points

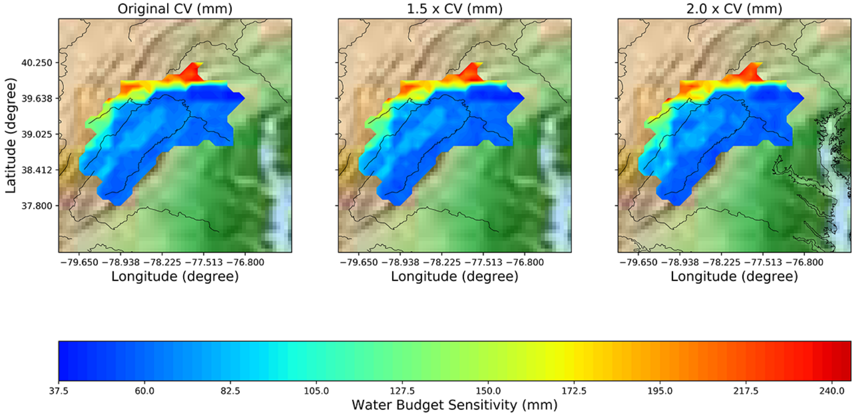


- The truncation on the left corresponds to the days without precipitation

Simulation Algorithm to generate resampling distribution

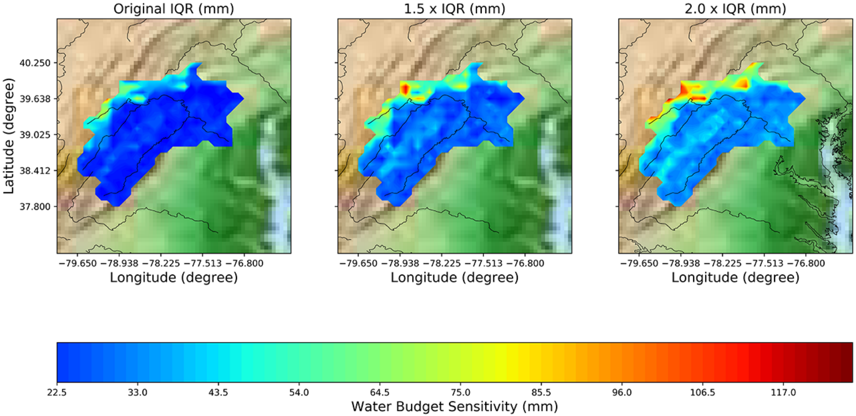
- For each grid point in our data,
  - Estimate the proportion of zeros for the cell as the sample proportion, $\hat{p}_0$
  - Fit a Gamma distribution to the positive section of the data, parameterized by its mean $\hat{\mu}_0$ and coefficient of variation $\hat{cv}_0$
  - Generate $k = 100$ i.i.d. samples each of size 639 based on the above parameters
  - Run the model with each sample, and estimate the water budget based on the output the Standard Error at each point, as well as the Interquartile Range
- Plot summary statistics and visualize the resampling distribution of the water balance between April and September 2017
- Inflate the Coefficient of Variation by 50% and 100%. Plot its effect on the variability in the water balance measured in terms of Standard Error and IQR
- Inflating Coefficient of Variation instead of the bias allows us to measure change in variability while avoiding bias to our estimates

Standard errors (SE) of the resampling distribution of total water balance



- We can see that the northern area shows the highest variability in its water balance
- We don't see a lot of difference when inflating precipitation errors, but that might be because of extreme values in the data

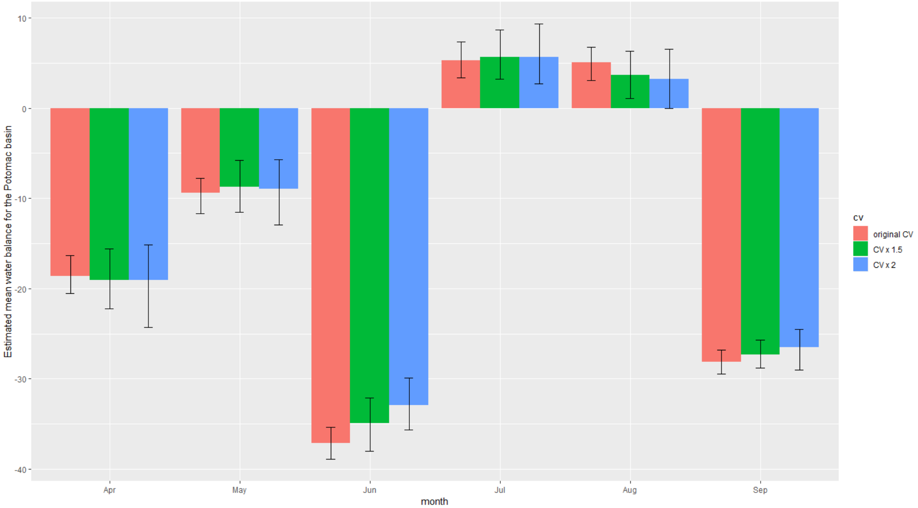
Interquartile range (IQR) of the resampling distribution of total water balance



- With the IQR plot, we see clearer increases in variability in water balance as we inflate the CV
- The northern region still shows the highest variability

Monthly summary statistics of the resampling distribution of water balance

Figure: Mean and 95% empirical confidence interval for monthly water budget



Monthly summary statistics of the resampling distribution of water balance

- While there is some bias present in our estimates, it follows the patterns we observed in our data
- Our main interest however is in the variability of the estimates

CV	Apr	May	Jun	Jul	Aug	Sep
Original	4.25	3.93	3.54	3.98	3.68	2.65
CV x 1.5	6.69	5.71	5.92	5.44	5.27	3.10
CV x 2	9.17	7.28	5.77	6.64	6.64	4.48

Table: Lengths of the empirical 95% Confidence intervals

- We have already seen that some parts of the basin are more affected by the inflated CV than others
- We also see that April and May estimates are most sensitive to variability in the data
- Running a calibrated model for longer duration would allow us to do further inference on this

Outline

Summary

- We were able to run a full scale hydrological model on taki. This required the installation of multiple tools, getting data from different sources and aligning them all. We streamlined the process and have put a framework in place that allows scalability
- We estimated the water budget for the Potomac river basin using VIC for the months of April to September 2017. For a calibrated model, the output leads to further analysis and policy decisions
- We have also visualized which areas' and which months' water budget are most sensitive to precipitation forcing errors
- We started exploring the spatial modeling of precipitation data for use in hydrology forecasts

Next Steps

- Use more data for spin-up; extend the water budget predictions to a longer duration
- Calibrate VIC for better output
- Explore fitting spatio temporal models to the data
- Use the available precipitation errors from GPM-IMERG to simulate better resampling distributions

This project was an exercise in methodology and some preliminary analysis. The framework required for inference is in place. Better estimation tools - model calibration and data quality checks on the physical front, spatio-temporal model fitting on the statistical front, and parallelization from an HPC perspective are the next steps, and each aspect can be extended into further research topics.

Thank You!