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

Introduction

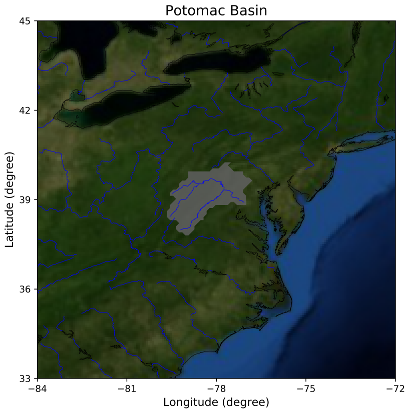


Figure: Extent of the Potomac river basin indicated by the gray shape; rivers are represented by blue lines.

- The Potomac river basin, also known as the Potomac watershed, is located on the East Coast of the USA across West Virginia, Virginia, Pennsylvania, Maryland, and the District of Columbia
- According to the Interstate Commission on the Potomac River Basin, about 600 million gallons per day (mgd) is used for water supply, of which 500 mgd is for the Washington area
- About 1.6 billion gallons is used daily for power plant cooling and other industrial uses. Most of it is returned to the streams,
- Sufficient quantities of clean water are essential for human needs and also for ecosystem health

Introduction

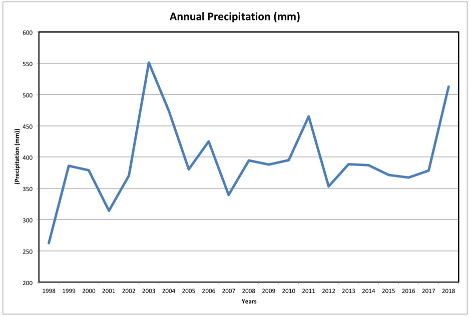


Figure: Basin-averaged annual precipitation time series from NASA's Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) from 1998 to 2018. Substantial year-to-year variations are seen, with the mean around 394mm

- Climate variability also plays an important role in the region's water supply alongside increasing demand
- In particular precipitation, the main source of water, varies inter-annually and makes it challenging to plan for water allocation within the basin
- Understanding seasonal to inter-annual variations in water availability within the basin due to climate variability is important for planning water resources management

Introduction

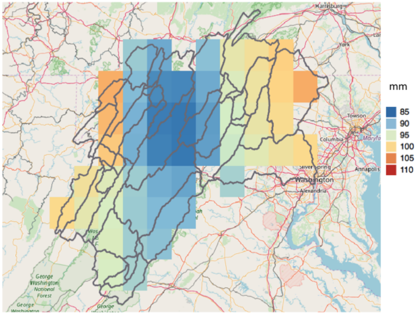


Figure: Map of the Potomac river basin showing mean precipitation from 1998 to 2018 based on TMPA.

- The mean precipitation map from TMPA shows spatial pattern of precipitation over the basin
- The central basin receives relatively lower precipitation compared to the eastern and western portions of the basin.
- In this study, we focus on assessing the spatial and temporal variability of water availability in the Potomac Basin using a hydrologic model
- Understanding seasonal to inter-annual variations in water availability within the basin due to climate variability is important for planning water resources management

Problem Definition

The Water Budget equation

The water budget equation for a river basin, considering that there is no surface, sub-surface, or groundwater net inflow/outflow in the watershed, and that surface runoff and baseflow contribute to discharge, is expressed as

$$\Delta S = Pr - ER - RO - \text{Baseflow}$$

- **ΔS** is the change in water storage in the basin surface (snow amount, soil moisture) and sub-surface (root zone moisture, groundwater) water storage components,
- **Pr** is Precipitation,
- **ET** is Evapotranspiration, a combination of evaporation from bare soil and wet canopy, and transpiration from each vegetation type,
- **RO** is the Surface Runoff, a function of soil wetness and soil infiltration capacity, and
- **Baseflow**, or the Sub-surface Runoff, is the portion of the precipitation that is sustained between precipitation events, and is a function of surface and sub-surface soil moisture.

The available water over the basin, referred to as the water budget or the water balance, can be estimated by subtracting ET, surface and sub-surface runoff from the precipitation.

Problem Definition

- We use Variable Infiltration Capacity (VIC), a hydrologic model, to estimate the water budget components
- VIC uses precipitation, surface temperature and wind data as daily inputs; evapotranspiration, runoff and baseflow are some of its outputs
- In this study we use precipitation estimates from the Global Precipitation Measurement(GPM) mission; specifically, we use Integrated Multi-satellitE Retrievals for GPM (IMERG)data derived by merging passive microwave measurements from a constellation of satellites,calibrated with those from the GPM Core Observatory, and adjusted with measurements from global rain gauge networks
- We aim to assess how intra-seasonal variability of precipitation within the basin affects water budget estimates based on this
- Further, using a sensitivity analysis, we want to quantify how the errors in the IMERG precipitation influence water budget components, both spatially and temporally

Problem Definition

Sensitivity Analysis

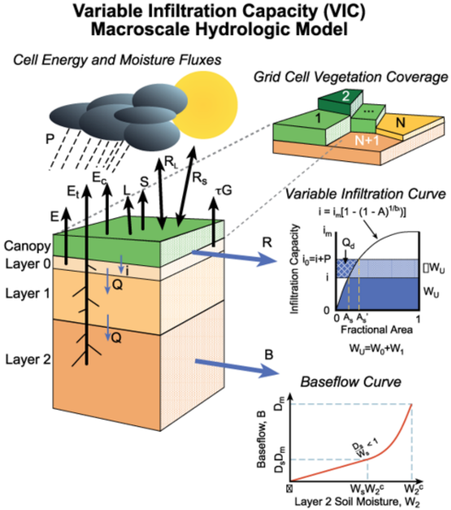
Sensitivity Analysis (SA) can be defined as
"The study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input".

- A few common purposes, or settings, are:
  - **Factor Prioritization** aims to rank input factors by their relative contribution to the output variability
  - **Factor Fixing** identifies if any factor has a negligible effect on output variability
  - **Mapping** aims to determine the region of input variability that significantly alters output, producing extreme values as an example.
- We conduct a simple form of SA by varying (perturbing) precipitation from a nominal value based on its distribution using a simulation model, and assess the impacts of the simulation results via visual inspection.

Outline

Model Description

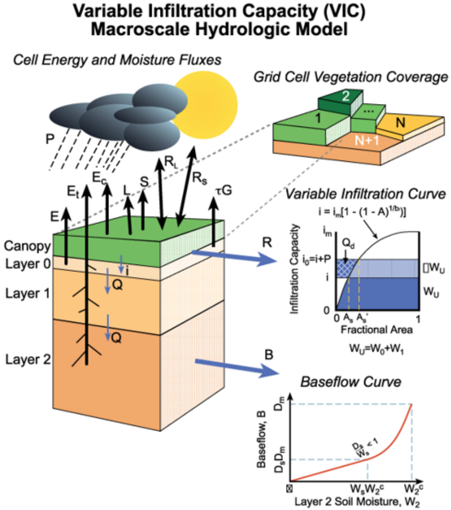
- Semi-distributed, macro-scale hydrological model
- Can be used to understand hydrological processes in almost real-time
- Developed by Liang et al. as a research model around 1994
- Physically based model to be coupled with global circulation model (GCM) simulations
- Can be run in either a water balance or a water-and-energy balance mode
- Requires a minimum one year of spin-up time



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

Model Assumptions

- Water and surface energy budgets are computed independently for each grid cell; variations at a sub-grid level are modeled statistically
- Water is assumed to enter a grid cell from atmosphere (precipitation) only
- It is assumed that there is no channel flow, sub-surface flow, or recharge to soil from the river
- It is assumed that groundwater flow is relatively small compared to surface and sub-surface flow
- These assumptions are generally valid for grid areas larger than 3 km and smaller than 2°
- VIC results are post-processed separately to simulate stream flow



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 2017
- 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

- The model uses data from Jan 2016 to Mar 2017 for spin up, and daily water budget components are analyzed for April through September 2017
- There are a total of 387 grid points/cells in our dataset, and 639 days of data for each grid cell

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

Projection and Aggregation

- 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

Projection and Aggregation

- 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 and Final steps

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

Runtimes

- VIC took under 2 minutes to run for our use case; we expect it to scale with the duration and the number of grid points in our data
- The major time bottleneck was the pre-processing part of the data; The MODIS pre-processing took around 1 hour and the meteorological forcings took around 2 hours to be generated
- The remaining data pre-processing (usually involving the parameter files and raster creation) each took the order of seconds

Outline

Outputs of interest

- Our main focus is on assessing the variability of the water budget components between April and September 2017
- 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
- We looked at these components both for each month, and for each grid point

Water budget components between April and September 2017

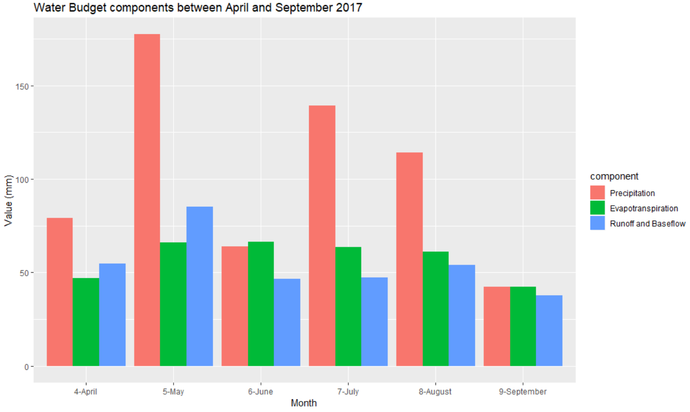
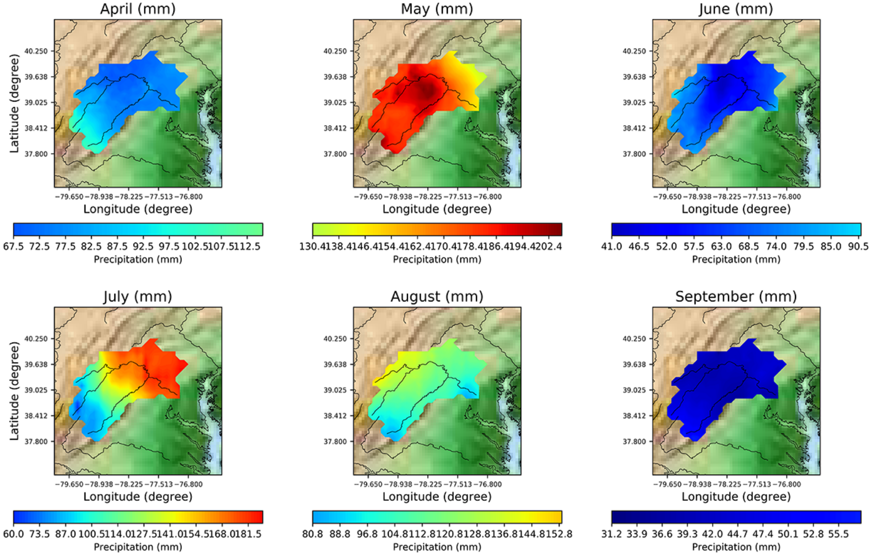


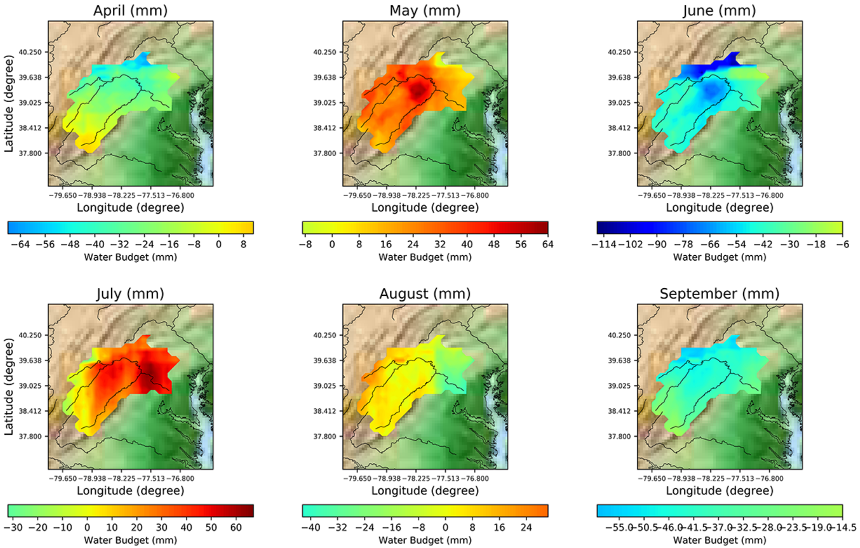
Figure: Basin-averaged monthly mean precipitation (red), evapotranspiration (green) and runoff and base flow (blue) between April and September 2017.

- Precipitation in May, July, and August is much higher than that for April, June, and September during 2017
- Runoff and baseflow are highest in May but lower in July and August; ET is lower than precipitation in April, May and July
- ET in the basin depends on vegetation cover and surface energy fluxes while runoff depends on vegetation cover and soil moisture conditions
- Precipitation has a higher intra-seasonal variation compared to ET and runoff which leads to variation in water balance

Monthly precipitation in the Potomac Basin



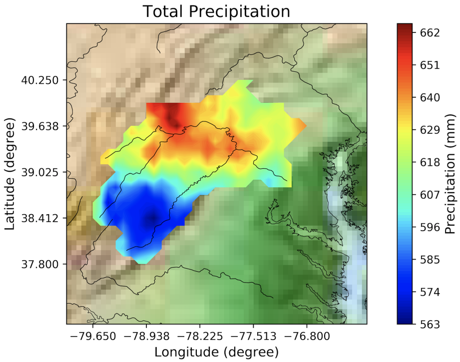
Monthly water balance in the Potomac Basic

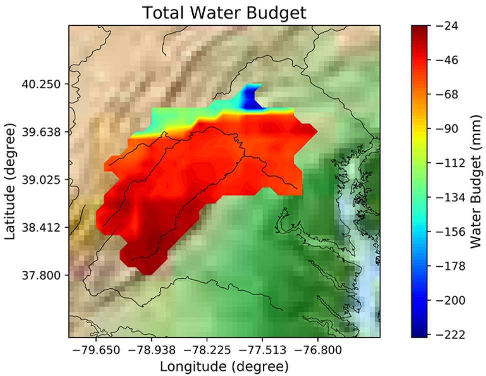


Monthly spatial distributions of precipitation and water balance

- The central and northern parts of the basin appears to have received larger amounts of precipitation in May, July, and August; positive water balance is observed in May and July
- The negative water balance in April, June, and September is consistent with the lower rate of precipitation in those months while ET and runoff values remain comparable or even higher compared to other months
- The distribution of water balance has a similar pattern with the spatial distribution of rainfall, especially the one in April and May.
- During June to September, we believe that increase in vegetation would increase ET within the basin, resulting in higher loss of water to the basin

Total Precipitation and Water Balance Between April and September 2017





- A higher rate of precipitation in the western-central region of Potomac basin and a lower value in the southwest corner is seen
- There is also water deficit across the entire basin, including a visible band at the northern part of the Potomac basin, where much larger negative values can be seen
- We will be examining ET, Runoff, terrain, and snow conditions to explain the presence of this band

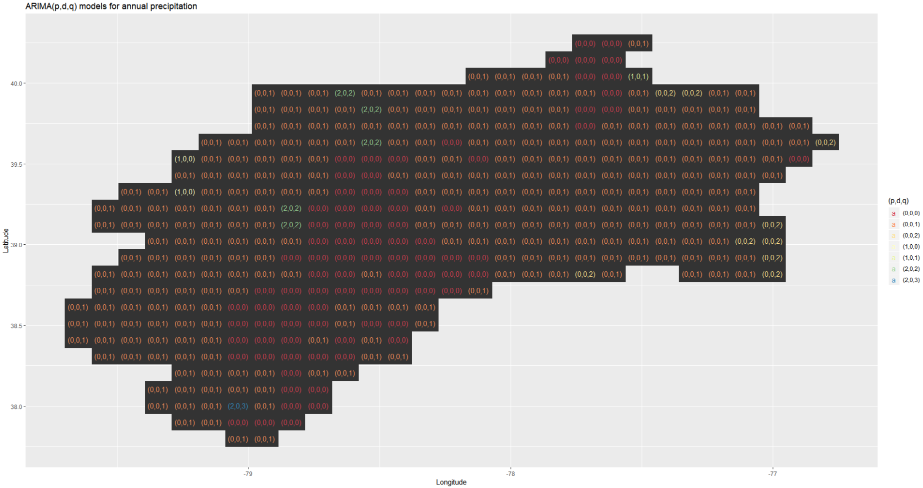
Outline

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

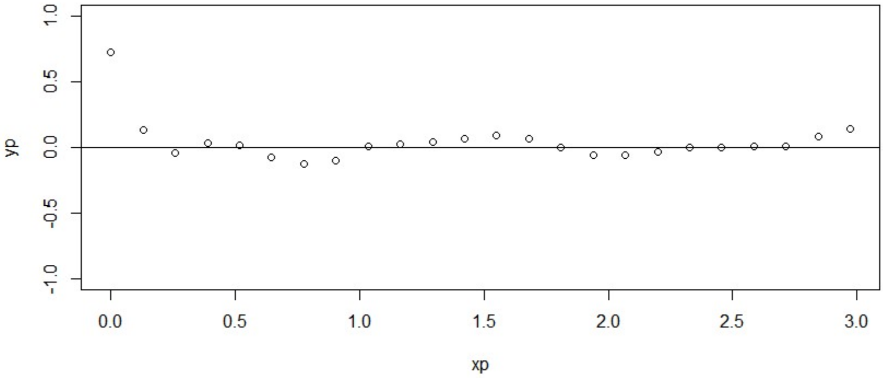
Temporal analysis

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



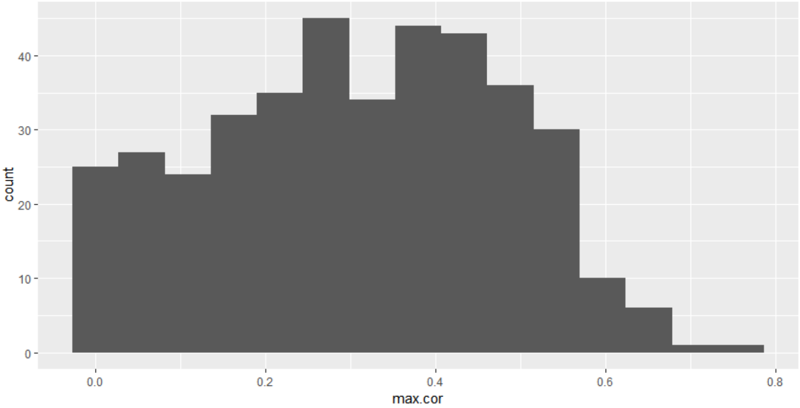
Spatial analysis

- 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

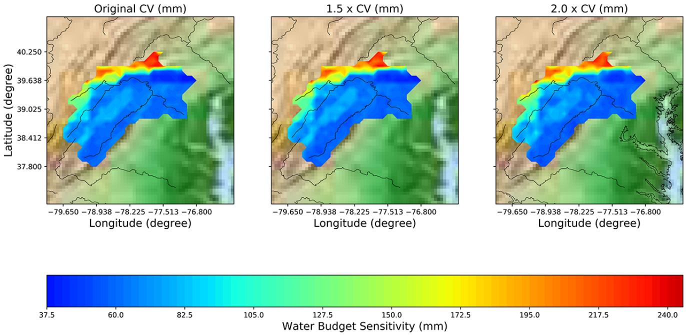


- The truncation on the left corresponds to the days without precipitation

Simulation Algorithm to generate resampling distribution

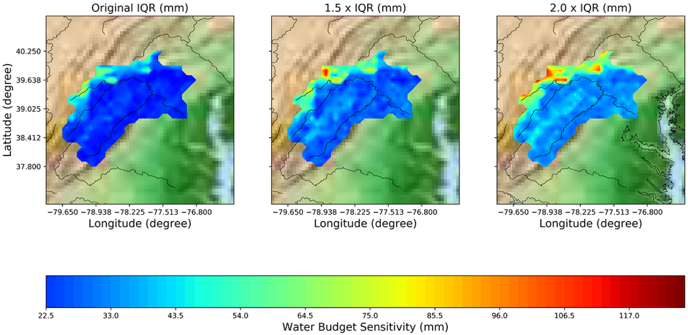
- For each grid point in our data,
  - Estimate the proportion of zeros for the cell as the sample proportion, $\hat{p}_0$
  - Fit a Gamma distribution to the positive section of the data, parameterized by its mean $\hat{\mu}_0$ and coefficient of variation $c\hat{v}_0$
  - Generate $k = 100$ i.i.d. 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 as well as a diagonal channel in the southwest area are most sensitive to forcing errors
- We don't see a lot of difference when inflating precipitation errors, but that might be because of extreme values in the data
- To resolve this, we also look at output IQR instead of Standard Deviation as a measure of sensitivity to precipitation forcing errors

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



- The northern region still shows the highest variability; changes in the southwest channel are also more pronounced
- While it might seem obvious, note that precipitation data from different sources are often assimilated to come up with a single estimate (our IMERG dataset is one such case), a process which can lead to an underestimation or overestimation of its dispersion
- VIC will carry over these features in its output ensembles, which is motivation to choose a precipitation input that best reflects the systemic variability of the process

Monthly summary statistics of the resampling distribution of water balance

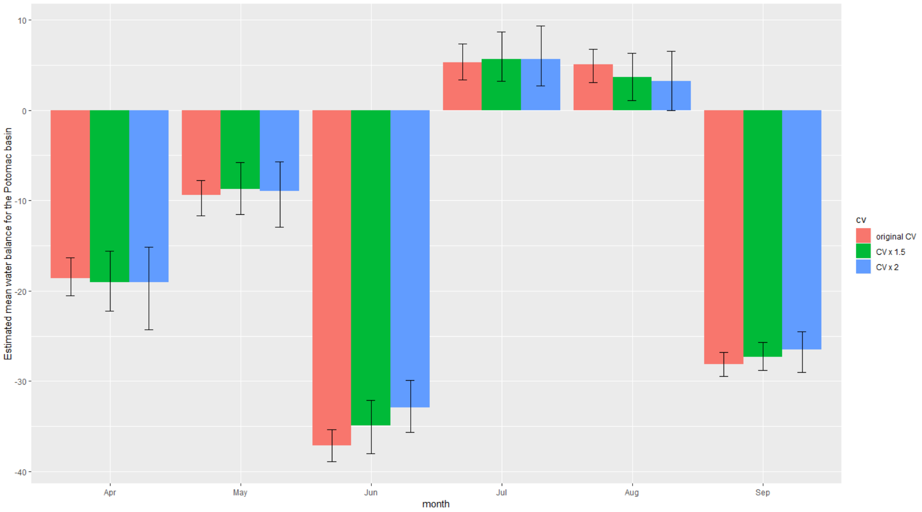


Figure: Mean and 95% empirical confidence interval for monthly water budget

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!