**A Study on Parsimonious Models in Catchments Generating Saturation Excess Runoff**

**Federico Antolini1, Stefany Baron2, Brittany Lopez Barreto3 and Ishrat Jahan Dollan4**

1Federico Antolini; *federico-antolini@uiowa.edu*

2Stefany Baron; *stefbaron@vt.edu*

3Brittany Lopez Barreto, *blopezbarreto@ucmerced.edu*

4Ishrat Jahan Dollan; *idollan@gmu.edu,*

**Academic Advisors:** Eric Tate, *University of Iowa,* Randel Dymond, *Virginia Polytechnic and State University*; Erin Hestir, *University of California, Merced*, Viviana Maggioni, *George Mason University*

**Summer Institute Theme Advisors**: Dr. Fred Ogden, National Water Center, *fred.ogden@noaa.gov;* Dr. Hilary McMillan *,* San Diego State University, *hmcmillan@sdsu.edu*

**Abstract:** Understanding the process of runoff generation with physically interpretable parameters is currently an active area of research.The objective of this report is to compare the performance of TOPMODEL against a new model formulation that is based on TOPMODEL’s underlying assumptions, but has fewer parameters and a stronger focus on soil properties. The existing version of TOPMODEL is a conceptually based rainfall-runoff model that uses watershed topography for predicting subsurface flow and overland flow. Though it is widely used, TOPMODEL has limitation in performance in regions where saturation overland flow occurs. An alternative saturation excess model based on the fundamental concepts of TOPMODEL was developed by our team in Python and was compared with the TOPMODEL package in R in two study areas, Hubbard Brook, New Hampshire and Sleepers River, Vermont. The soil properties incorporated into the alternate model are hydraulic conductivity (Ksat), porosity, and specific yield. We found that TOPMODEL streamflow predictions are in general closer to observed values than the alternate model. TOPMODEL underpredicts most of peak flows, whereas the alternate model consistently overpredicts them. While saturation excess is the dominant runoff generation mechanism in the study areas, it is not sufficient to explain the rainfall-runoff transformation.

**1. Motivation**

Hydrologic models are used for predicting stream flows in areas of interest during storm events. The ability to do this accurately is becoming increasingly important for the management of water resources. With a better understanding of the fundamental physical equations that describe the interactions between soil and water, new modeling approaches like that of the alternative model described here, may be incorporated into the National Water Model (NWM) to generate runoff.

Parsimonious models are those that can make predictions with the use of minimal parameters. These are ideal because fewer parameters result in easier calibration and less uncertainties. The relationship between soil moisture and runoff should not be underestimated since soil can affect both subsurface and overland flows, and eventually streamflow at the outlet. TOPMODEL is used to model the dynamics of contributing saturated areas but saturation excess overland flow routine is not computationally efficient for small catchments. Here we describe an alternate model that uses the same assumptions as in TOPMODEL, however it incorporates saturation excess overland flow considering local topography, soil properties e.g. soil depth, hydraulic conductivity (Ksat) at ground surface, soil porosity, and specific yield. The aim of this alternate simulation strategy is to obtain the most accurate runoff estimates in accordance with the water balance using the least number of model parameters.

**2. Objectives and Scope**

The alternate formulation relies on a parsimonious rainfall-runoff model concept, based on TOPMODEL assumptions. If successful, it is hoped that this alternative formulation can one day be incorporated into the National Water Model (NWM). The alternate model uses higher resolution topography and simple subsurface flow equations in a computationally efficient fashion to model saturation excess overland flow. This study provides insight on the role of saturation excess runoff generation and suggests areas where the National Water Center (NWC) could potentially apply the alternate model. Moreover, this study is to support reproducibility in the field of hydrology, so all the codes and datasets will be available to access through our Github repository provided at the end of the report.

**3. Background and Previous Studies**

Hydrological processes in a catchment are dynamic and heterogeneous. Because of the lack of measurements of state variables and catchment attributes, it is preferred that a minimal number of parameters which represent hydrological connectivity and catchment response are used. A detailed process conceptualization would need more calibrated parameters which increases model uncertainty.

TOPMODEL simulates hydrological fluxes that rely on topographic information (catchment area, local slope and topographic wetness index). [3] considers TOPMODEL as a set of conceptual tools to reproduce the dynamics of both surface and subsurface contributing behavior in semi-distributed way. Topographic wetness index (TWI) represents the tendency of developing saturated conditions in the catchment. Areas with higher TWI values represents greater runoff generation. This index is derived from the upstream area above a pixel that drains through the unit contour at the pixel and tanß is the slope of the ground surface at the location. A simple approximate relationship between catchment storage deficit along with lateral transmissivity is developed within the framework of TOPMODEL. These relationship attributes form the basic physical equations to simulate the response of a catchment. Because of its semi-distributed functional framework, it bridges the gap between the complex distributed process models and simple lumped concepts. R’s TOPMODEL package converts the topographic effects into a distribution of classes while simulating runoff in the outlet.

TOPMODEL relies on three basic assumptions [3]: (1) steady state condition for a saturated zone, (2) local hydraulic gradient of a saturated zone approximated by the local surface topographic slope, tan β, (3) an exponential decay function of transmissivity with depth or deficit. TOPMODEL represents three layers of the soil column (root, unsaturated and saturated zones) as three interconnected reservoirs. The parameters TOPMODEL deals with imclude initial subsurface flow, rate of transmissivity decay in soil profile, root zone storage capacity, and others which are explained fully in Buytaert 2018.

Past studies introduced fractional saturated areas in a catchment correlated with the soil moisture [4]. Conceptually, when precipitation falls (ignoring snowmelt) over those saturated grids, it converts into surface runoff. In absence of precipitation, the groundwater table naturally lowers, water in the soil flows laterally down to an outlet cell which allows subsurface flow out to a stream. Lateral flow is only modeled in the saturated part of the soil since unsaturated flow takes more time to drain. This study proposes a simple relationship between saturated areas and soil water content by introducing a saturation excess scheme. This simplification allows to change the nature of non-linear reservoir for a small-scale catchments (less than 10 km2) than the discretized framework of distributed hydrological models. Flow routing is not computationally efficient in hourly time step for small catchments. The simplification would replace the routing scheme with the volume of overland flow of the alternate formulation in hourly time step.

The alternate formulation focuses on saturation excess overland flow and considers the fundamental assumptions of TOPMODEL. The water balance connected to fractional saturation in a catchment is defined by this formulation. Therefore, the formulation produces functional relationship between soil moisture storage and saturated fraction in a catchment. Flow routing mechanism can be avoided at the outlet. Additionally, the alternate formulation can run without the topographic index to build the functional relationship. As more rain falls onto the watershed, the relationship generates simulation of saturation excess runoff. Therefore, a characteristics diagram of saturated areas and volumetric content in soil can replace the topographic index of hydrological similarity in catchment. The diagram differentiates the amount of overland flow and the vertical flow in a non-linear reservoir considering a shallow groundwater table.

**4. Methodology**

*4.1 Overland Runoff Formulation*

TOPMODEL performance is compared to a proposed alternative rainfall-runoff model that incorporates soil parameters. The incorporation of the soil parameters in the alternative model is necessary in order to combine the effects of saturation excess on runoff on a landscape.

*4.2 Study Sites*

*4.2.1 Hubbard Brook*

The Hubbard Brook watershed in North Woodstock, New Hampshire has been chosen as one of the study sites because it is one of the most comprehensive ecosystems studied in the world [5]. Hubbard Brook forest has nine individual subcatchments, as shown in Figure 1. Subcatchment #7 served as our preliminary study site that was chosen for its hydrologic characteristics such as gradual changes in average temperature and precipitation.

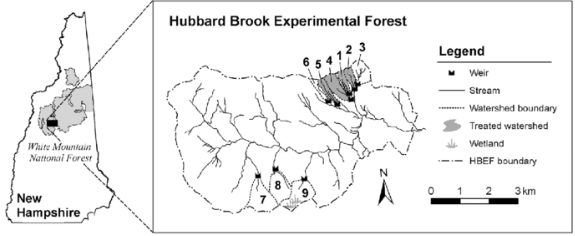
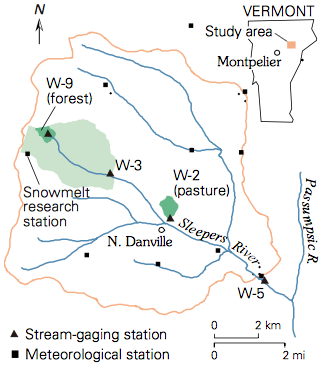


Figure 1: Hubbard Brook Watershed [6]

*4.2.2 Sleepers River*

The Sleepers River Research Watershed (SRRW) in Vermont has been an active hydrologic research site since 1959 and was the setting where Dunne and Black (1970) determined the controls of saturation-excess overland flow (SOF) on streamflow generation shown in Figure 2.

Glaciers that covered New England thousands of years ago have shaped the present landscape and have impacted its hydrological patterns. A majority of the watershed is covered by 1-4 meters of glacial till, which leads to high buffered streamflow due to the weathering of the calcite within the till. The study area has a humid continental climate with an average temperature of 6 degrees Celsius and average annual precipitation of 1.1 meters where approximately 20-30% is from snow [7].

Figure 2: Sleepers River Watershed

*4.3 TOPMODEL Package in R*

The subcatchments described in section 4.2 were first modeled using TOPMODEL. The TOPMODEL function in R requires an input of five variables. The first is a set of ten parameters which are described by Buytaert 2018. The second is a TWI data frame which represents the distribution of TWI values for the Digital Elevation Model (DEM) of the subcatchment. The third variable is a delay dataframe which represents channel routing through the watershed. The fourth and fifth variables are precipitation and potential evapotranspiration values respectively. These last two variables should be vectors of equal length and for the same timesteps. For a more thorough explanation of these five input variables, see Buytaert 2018.

The TOPMODEL package in R comes with functions that can calculate the TWI and delay dataframes using only the DEM of the subcatchment. Precipitation values were obtained from the Analysis of Record for Calibration (AORC) [8] and converted to meters of precipitation per hour. Flow data was obtained from USGS and converted to cubic meters per hour per unit area. Evapotranspiration (ET) data was not available and, therefore, it is estimated using water balance equation. The water balance method was used to find that the average rate of ET for subcatchment W-3 was approximately 3 millimeter per day or 0.003 meter per day. Using this information, an ET vector was created that ranged from 0.002 to 0.005 meter per day with colder months containing lower values and hotter months containing higher values. For future modeling, it is recommended that a watershed with ET data be used so that there is less uncertainty.

Parameters of TOPMODEL were estimated based on the range of values found in literature for subcatchments W-3 of Sleepers River and subcatchment#7 of Hubbard Brook. A team of Fellows of the National Water Center Innovators Program: Summer Institute 2019 form Hydroinformatics theme then created 22000 parameter sets Saltelli sampled for Sobol sensitivity analysis using the parameter ranges that had been provided to them. A full explanation of this sensitivity analysis can be found in their paper, *A Visualization Workflow for Quantifying Parameter Sensitivities to Uncertainties for Hydrologic Models*. Through this sensitivity analysis it was learned that the parameters of initial subsurface flow (qso), rate of decline of transmissivity in the soil profile (m), and maximum root zone storage deficit (srmax) are the most sensitive. Numerous experiments with different values of each of these parameters were performed to evaluate how each would affect simulated flow. One interesting conclusion drawn from this analysis is that different months of the year need to be calibrated with different values of qso. Therefore, TOPMODEL is not recommended for modeling a time period of several months. This paper focuses only on a period of 6 weeks for Sleepers River (April 30 – June 12th, 2017) and a period of 2 weeks for Hubbard Brook (June 24th – July 7th, 2014). These date ranges were chosen so that snow melt would not have to be accounted for.

*4.4 Alternate Formulation*

Our alternate model considers the assumptions of TOPMODEL to characterize the hydrologic behavior of a watershed and applies a synthetic relationship to the rainfall-runoff transformation. The three main assumptions that are addressed in this alternative model are that in a given watershed a) the soil is highly porous, and thus the contribution of infiltration excess to overland flow is negligible, b) the thickness of the unconfined aquifer is limited to a few meters (<5) and is spatially uniform across the watershed, c) the groundwater table has the same slope as the ground surface, which implies that groundwater moves in the same direction as surface runoff.

Part 1 of the model consists of characterizing the watershed with a relationship between the proportion of saturated area and the total water content in the soil, or equivalently the soil water deficit. Because saturation excess is generated by those parts of the watershed that are susceptible to becoming completely saturated and hydrologically connected to the outlet, one can estimate the amount of runoff based on the estimated water content in the soil at the watershed level. To accomplish this, the algorithm uses a DEM to identify flow direction based on prevalent slope in the D8 scheme [9]. The model starts from an ideal condition of complete saturation of the whole watershed and simulates the natural lowering of groundwater and the subsurface flow toward the outlet. At each time step it identifies the saturated cells, calculates the percentage of saturated area across the total area of the watershed and records the overall amount of water stored in the aquifer, both in unsaturated and saturated layers. The pseudocode in Figure 3 summarizes the steps of this algorithm.

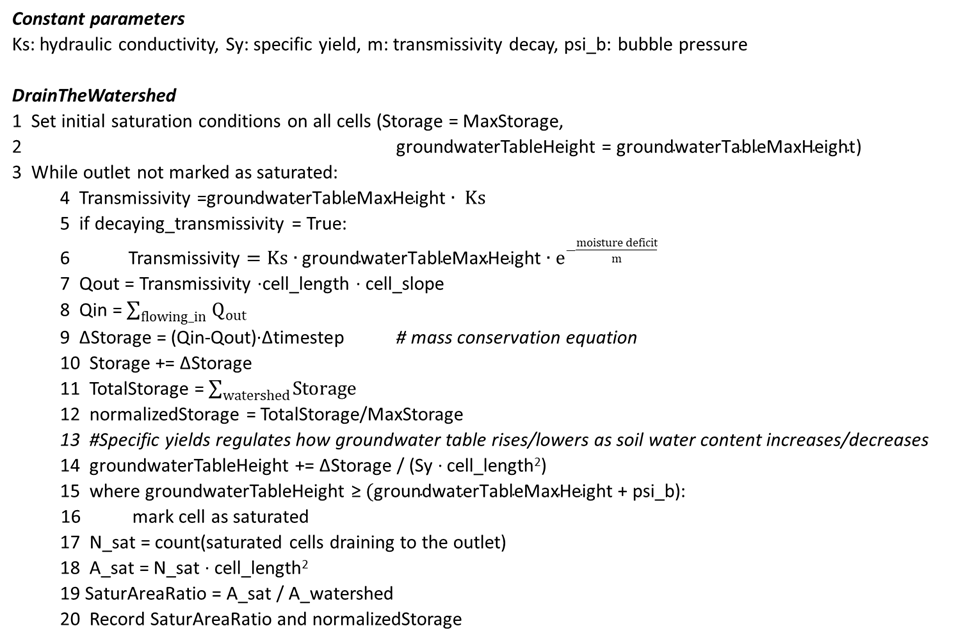


Figure 3: Pseudo code for Part 1 of the alternate model (Drain the Aquifer). The relationship is built at each timestep from recorded values of Saturated Area Ratio and normalized Storage.

The above relationship is sufficient to characterize the relevant processes occurring in the watershed. Part 2 of the model is, in fact, a lump type of model. Differently from Part 1, Part 2 moves away from a distributed model and treats the watershed as one reservoir behaving according to the relationship from Part 1. The model takes precipitation and evapotranspiration as input and applies the mass balance equation to determine subsurface and overland flow at the outlet of the watershed. Pseudocode in Figure 4 summarizes rainfall-runoff transformation of Part 2. The key concept is that soil water storage and surface runoff, which by hypothesis (c) is generated only from saturation excess, are strictly related according to the relationship from Part 1.

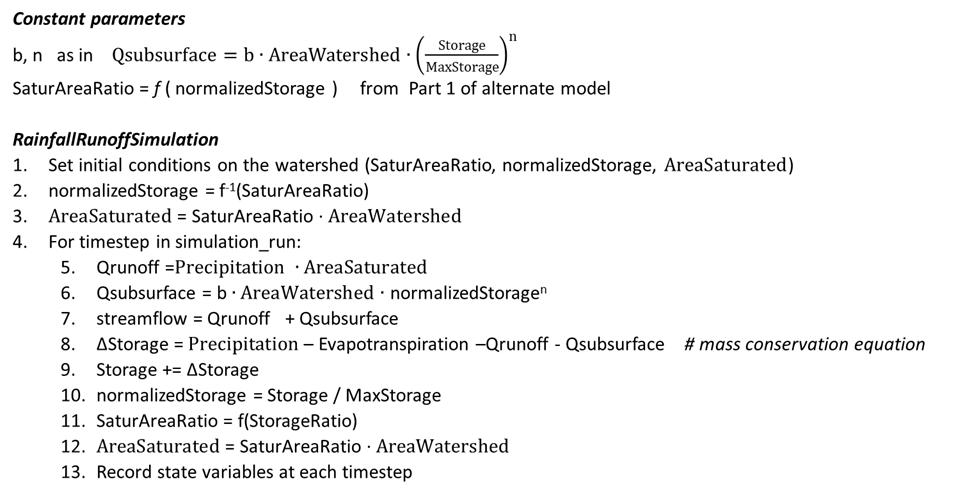


Figure 4: Pseudo code for Part 2 of the alternate model (Rainfall-Runoff Transformation)

**5. Results**

*5.1 Hubbard Brook #7 subcatchment*

Figure 5 shows observed and modeled streamflow for a period of June 24th through July 7, 2014. TOPMODEL does a good job of capturing the size of the peak flow for the storm event that occurs during this time period. The time to peak for the modeled, however, is a few hours delayed from the observed. The alternate model looks very responsive to precipitation inputs, and correspondingly generates sharp flow peaks. The main problem is that this model simulates more peaks than are actually observed. This might be caused by an incorrect modeling of soil storage, i.e. the soil can hold more water than what the model quantifies. However, there are significant values of rainfall which are not followed by observed flow peaks. This poses the question of how reliable precipitation forcing data are in the context of a small catchment. While in a large watershed errors in precipitation values might even out and overall mitigate, similar errors in a small catchment may lead to significant errors in streamflow prediction without possibility of amendment.

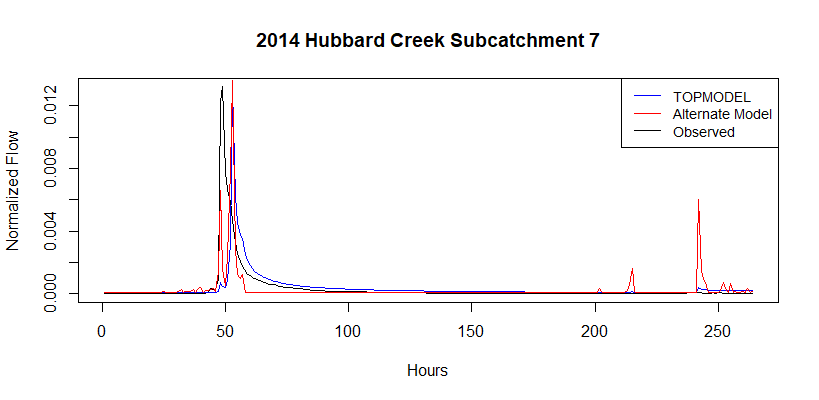


Figure 5: TOPMODEL and alternate model streamflow simulation in HB#7 for the period June 24th to July 7th, 2014.

*5.2 Sleepers River*

Figure 6 shows observed and modeled streamflow from April 30th to June 12, 2017. Visually from the graphs, TOPMODEL does not seem to perform as well here as it did in Hubbard Brook. The model overpredicts most of the peak flows and does not do a good job of capturing the duration of peak flow events. The alternate model performs fairly well in terms of recognizing flow peaks. Observed and modeled peaks align, and the alternate model does not generate more peaks than are observed. Peaks are generally overestimated and shorter in time. The model responds immediately to precipitation inputs and the absence of a delay function determines the sharpness of peaks, especially on recession limbs. One limitation of the alternate model is the absence of a delay function, which is normally introduced in lumped models to account for travel time of water from the most hydrologically distant areas down to the outlet (Beven 2011). This might be introduced in the future for a more accurate representation of events.

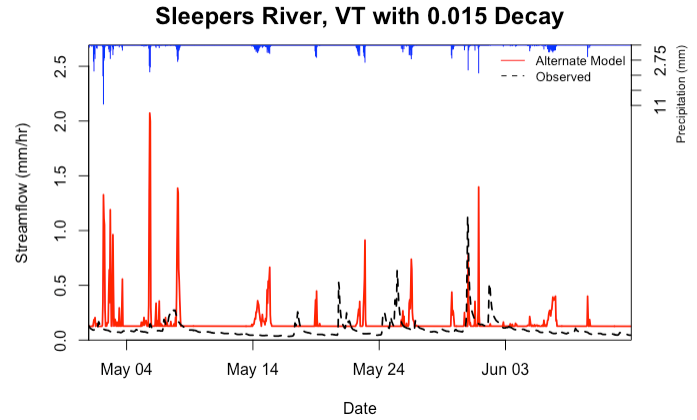
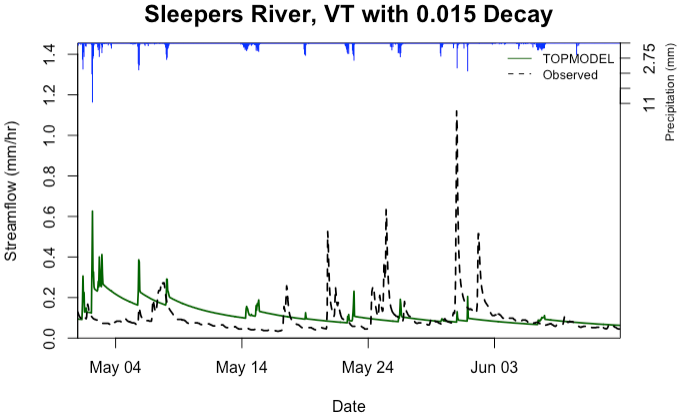
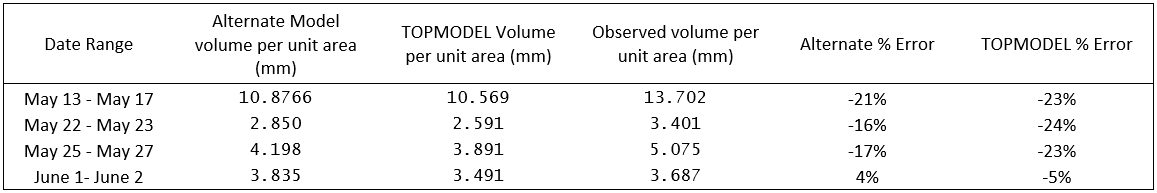


Figure 6: TOPMODEL and alternate model streamflow simulation in Sleepers River for the period May 1st to June 12th, 2017.

Our alternate method is qualitatively valid as it is able to predict when a peak in discharge occurs but is not able to correctly quantify the magnitude of peaks. Nash-Sutcliffe Efficiency or peak average error are typical ways to assess the performance of a hydrologic model [10]. Here, a visual comparison of modeled and observed streamflow is sufficient to identify the systematic discrepancies at each peak. Many factors play a role in the overprediction, including the effective correspondence of model hypotheses to reality, the lack of a proper calibration of the model, the quality of rainfall data, the amount and variability of evapotranspiration along the season.

To compare the performance of TOPMODEL with our alternate model, the volume per unit area was computed for the top 5 peak events that occurred between May 1st and June 12th, 2017. We defined an event to include the timestep of the peak plus the 3 timesteps before and the 3 timesteps after. Since we are using a timestep of one hour in this study, each of the peak events described in Table 1 are of 7-hour duration. Volumes per unit area were computed by integrating the flow per unit area over the 7-hour duration of each peak. Table 1 shows that TOPMODEL performed better than our alternate model 4 out 5 times. TOPMODEL accounts for several things that our alternate model does not account for (such as infiltration excess and the delay function) and these could be reasons why TOPMODEL continues to outperform our alternative model. These are things to consider as we do future work on our alternate model formulation.

Table 1: Volume per unit area of the top five peak events for Sleepers River during April 30 - June 12, 2017



**6. Conclusion**

The need of a hydrologic model that is accurate and computationally efficient simultaneously has been a motivation for the creation of a plethora of models [2]. On the other hand, conceptual models condense hydrologic concepts in a simple and ready-to-apply form, although they often overlook the physics beneath the process. Our study attempts to conjugate these two general trends and presents a model that follows the steps of TOPMODEL as a conceptual model, but also takes advantage of high-resolution data and explicitly introduces physics.

We focused our efforts on the saturation excess, a mechanism of runoff generation which is dominant in some regions of the United States. The groundwater table rises and declines quickly, and the terrain gets frequently saturated in those regions. The model first creates a relationship between soil moisture and saturated areas, then applies a simple rainfall-runoff routine using that relationship. The results are compared to TOPMODEL simulation and observed data to highlight differences in model performances, weaknesses and potential for application in regions needed. Given the simplicity of the alternative model used, discrepancies between model results and observations are not surprising. Some improvements and more testing can be made in the future, such as refining model calibration, introducing a delay function based on watershed size and shape, applying the analysis to larger watersheds in the context of scaling and arid regions, and possibly implementing other hydrologic routines, such as snowmelt and infiltration excess. TOPMODEL is best known to work in high precipitation areas, so it would be worthwhile to investigate whether the alternative model or TOPMODEL would excel in a different climate than tested in this report.

From the beginning of project formulation based on gaps addressed at the onset of the project creation Aquaholic Anonymous was focused on the achievable form of the project. We started from an ambitious project idea and then we narrowed it down to a very specific research field considering stakeholder’s need and time bound. Despite having different expertise and backgrounds we focused on one goal and planned accordingly. Different ideas from each of us allowed room for creativity with extensive literature review. We tried to understand the existing gap and know our stakeholders. We collaborated for our own project tracking board on GitHub.

Overall, the project has been an extremely interesting experience and while we cannot recommend the adoption of the current version of our alternate model, we believe that further investigation can substantially improve it. With the current tendency being the integration of multiple approaches to apply in different landscapes, our contribution can be implemented in such a model, simplifying the analysis and abating computational resources.

**Supplementary Materials:**

Our Github repository is found at <https://github.com/brittbarreto/Aquaholics_Anonymous>, where data sets and scripts used from R’s TOPMODEL and our alternative model’s Python scripts are tracked..

**References**

[1] K. J. Beven and M. J. Kirkby, “A physically based, variable contributing area model of basin hydrology,” *Hydrol. Sci. Bull.*, 1979.

[2] G. K. Devia, B. P. Ganasri, and G. S. Dwarakish, “A Review on Hydrological Models,” *Aquat. Procedia*, 2015.

[3] K. J. Beven, R. Lamb, P. F. Quinn, R. Romanowicz, and J. E. Freer, “Topmodel,” in *Computer Models of Watershed Hydrology*, 1995.

[4] G. Y. Niu, Z. L. Yang, R. E. Dickinson, and L. E. Gulden, “A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models,” *J. Geophys. Res. Atmos.*, 2005.

[5] USDA US Forest Service, “Hubbard Brook Experimental Forest.” .

[6] K. M. Cawley, J. Campbell, M. Zwilling, and R. Jaffé, “Evaluation of forest disturbance legacy effects on dissolved organic matter characteristics in streams at the Hubbard Brook Experimental Forest, New Hampshire,” *Aquat. Sci.*, 2014.

[7] J. B. Stanley, “Sleepers River, Vermont A Water, Energy, and Biogeochemical Budgets Program Site.”

[8] X. Kitzmiller, D. H.; Wu, W.; Zhang, Z.; Patrick, N.; Tan, “The Analysis of Record for Calibration: A High-Resolution Precipitation and Surface Weather Dataset for the United States.” .

[9] S. K. Jenson and J. O. Domingue, “Extracting topographic structure from digital elevation data for geographic information system analysis,” *Photogramm. Eng. Remote Sensing*, 1988.

[10] R. H. McCuen, Z. Knight, and A. G. Cutter, “Evaluation of the Nash–Sutcliffe Efficiency Index,” *J. Hydrol. Eng.*, 2006.