Introduction to hydrologic model calibration

Matin Rahnamay Naeini (rahnamam@uci.edu)

University of California, Irvine

Background

Conceptual Rainfall-Runoff (CRR) models are important tools for simulating or forecasting runoff. These models (f) can parametrically represent as:

$$\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}$$

where $\mathbf{Y}$ is model output, $\boldsymbol{\epsilon}$ is the model error, $\mathbf{X}$ is the model input, and $\boldsymbol{\theta}$ is the parameter of the model (Rahnamay Naeini, et al. 2019). The performance of these models depends on their parameters ($\boldsymbol{\theta}$). Some of these parameters can be measured, however in many cases these parameters are either very expensive to measure or doesn't have a specific physical meaning. In these cases, the parameters are determined through a calibration process (Yapo, Gupta and Sorooshian 1996). In general, the model calibration is the process of finding the parameters of a model given the available (observed) data. In this activity, we learn about the concept of model calibration using a simple hydrologic model.

The hydrologic model in this activity is a simple bucket model. Considering watershed as a bucket, the input to the bucket is precipitation, and the output from this bucket is the leak from the bottom and the spill from the top. Figure 1 shows the schematic of the bucket model.

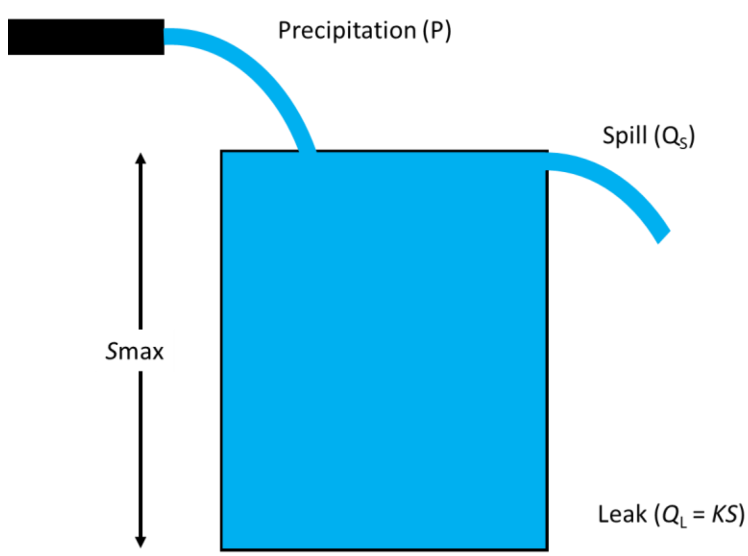

Figure 1 Conceptual bucket model schematic

Other input and outputs are assumed to be negligible. The leak from the bucket resembles the baseflow and the spill is the direct runoff. In this simple bucket model, leakage is function of storage. The parameters of the bucket model are S_{max} and K which are maximum bucket storage (mm) and leakage recession rate (1/day), respectively. Function for this bucket model is provided (BucketModel.m). The function inputs are precipitation, S_{max}, and K and the function outputs are runoff and storage, respectively.

Activity:

Step 1: In this step, generate a synthetic precipitation data. Use a random generator to generate daily precipitation data for 100 days. Gamma distribution is widely used as a suitable distribution to represent precipitation depth (Here, we recommend 10 and 2 for the shape and scale parameters). Assume the generated data is the actual precipitation over your watershed. In reality, this precipitation data is the actual depth of precipitation over watershed.

Step 2: Precipitation depth can be measured in different ways including rain-gauge, radar and remote sensing. Assuming precipitation data is collected by a rain-gauge in the watershed and our previous experiments showed that error in the collected data has Gaussian distribution with mean of 0 and standard deviation of 1 mm, perturb the actual precipitation generated in step 1 with this error to have observed precipitation data. Plot these two precipitation datasets. This perturbed precipitation data represents measured precipitation. In reality, it is not possible to measure the actual precipitation error, since all our measurements have some errors.

Step 3: Specify the bucket model parameters K and $S_{max.}$ K can be any value between 0-1 and S_{max} can have any values. However, based on the recommended parameters for precipitation generator, S_{max} is recommended to be between 60 – 120 mm.

Step 4: Use parameters from previous step and the actual precipitation data (from Step 1) as input to the Bucket model to generate runoff data. We assume this runoff data is the perfect observed data without any error, which doesn’t happen in reality.

Step 5: Since runoff gauges have errors, perturb the actual runoff data to generate observed runoff data. Assume that the runoff gauge error has Gaussian distribution with mean of 0 and standard deviation of 2mm. Plot the actual and observed runoff data.

Step 6: Now we have the observed and actual rainfall and runoff data, we can calibrate model. In reality, we don’t know the actual rainfall and runoff data and we can only use the observed data, which is not error free. The first step in calibration process is to define an objective function. The objective function measures the performance of the model with respect to the available observation. One of the most common objective functions in the field of hydrology is the Root Mean Squared Error (RMSE). RMSE measure the difference between the observed and simulated data (Runoff here) as follows:

$$\text{RMSE} = \sqrt{\frac{(Q_{obs} - Q_{Sim})^2}{N}},$$

Where Q_{obs} is the observed data (runoff), Q_{Sim} is the simulated data (runoff) or model output, and N is the number of data points. Since observed rainfall and runoff are observed data, RMSE value is function of the model parameters (S_{max} and K). In this step, define a function which calculates the RMSE value for the runoff data. The inputs to the RMSE function should be the bucket model parameters.

Step 7: One of the approaches to calibrate a model is to manually change the parameters and observe the RMSE value. Since we know the true model parameters (which are used for generating the synthetic runoff data), try different values for the parameters and see how the RMSE value changes. Does the true value have the lowest RMSE? Is it possible to find a lower RMSE value?

Step 8: Manual calibration is not the best way to find the parameters, especially when we don't have much information about the true parameters and the model has many parameters. In this case, automatic calibration can be used to find the parameters of the model. MATLAB has several built-in optimization functions. These functions can find the minimum value for a (an objective) function by changing the input parameters of that function. One of these functions is **fminsearch**. This function gets the objective function handle and an initial guess for the parameter set and employ Nelder-Mead simplex downhill search method to find the optimum values for function. Use the RMSE function handle from the previous step and an initial guess close to the actual parameters as input for **fminsearch** function to calibrate the model parameter. Are the parameters the same as the true parameters? Is the RMSE value lower than the RMSE achieved in the previous step?

Step 9: In the previous step, select an initial guess far from the true parameters. Is the RMSE value the same as previous step?

Step 10: Assuming that we know the true observed rainfall and runoff data. Use this dataset to calibrate the model with **fminsearch** function and an initial guess close to the true parameters. What is the value of RMSE? Is the parameter set the same as the true parameter set?

Step 11: Repeat this activity with larger dataset (for instance 1000 days) and different parameters for the model.

Step 12: Plot the simulated runoff for different parameter sets and observed runoff and explain why model calibration is necessary? How error in observed data affected parameters of our model?

References

Rahnamay Naeini, Matin, Bita Analui, Hoshin V Gupta, Qingyun Duan, and Soroosh Sorooshian. 2019. "Three decades of the Shuffled Complex Evolution (SCE-UA) optimization algorithm: Review and applications." *Scientia Iranica* 2015-2031.

Yapo, Patrice O, Hoshin V Gupta, and Soroosh Sorooshian. 1996. "Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data." *Journal of Hydrology* 23-48.