Conceptual model of the Hudson River Basin

David Haasnoot and Anne Versleijen
River Basin Hydrology and Water Management

Introduction

The Hudson River Basin is located in the North East of the United States, which is around 35,000 km^2. The lower part of the basin is an estuary. As the river in the lower part is influenced by the sea, only the Northern region is taken into account with an area of 21,000 km^2. The region is showed in Figure 1. The area consists mainly of forest and cropland with some urban areas.

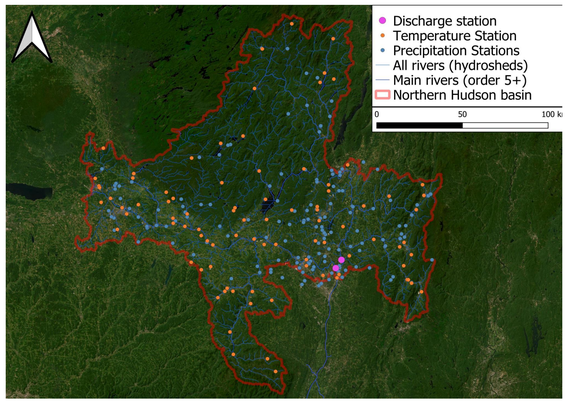

Figure 1. Considered upperpart of the Hudson River Basin. With dots, the different stations are depicted.

Objective

The aim is to describe the different elements of the water cycle within this region. To do so, data and information obtained from remote sensing techniques are used to develop a model.

HBV model

To model the water cycle in the basin, the HBV model is used. It is used both as lumped and distributed model. In the lumped model, the area is taken as homogenous. In the distributed model, the area is divided into wetland, hillslope and plateau areas.

The HBV model consists of various elements which describe different processes in the water cycle. If the air temperature is below -0.5 °C, the precipitation will be stored as snow. Otherwise it will intercept on vegetation until the storage capacity is reached. The excess (P_e) will flow into either the unsaturated rootzone or the fast reservoir. In case of melting snow, this flow will be added to P_e. From the unsaturated rootzone storage, water will leave as transpiration or percolate into the soil.

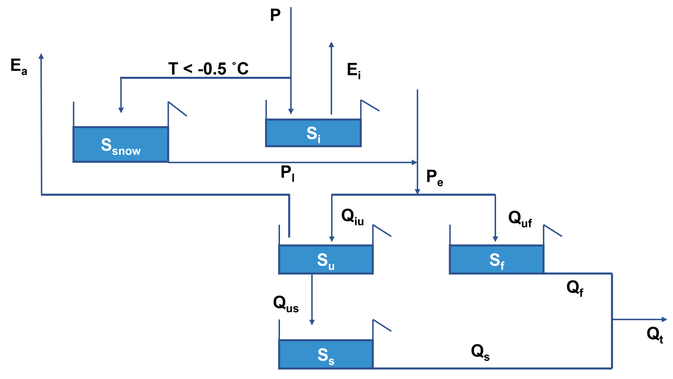

Figure 2. HBV model which is used to model the water cycle in the Hudson River Basin.

Data

For the precipitation data, station data is used which matches ERA5 well. The station data is averaged over the area by using Thiessen polygons. For evaporation data obtained from Global Land Evaporation Amsterdam Model (GLEAM) is used. Discharge station data is retrieved from USGS, the discharge of two stations in the lower part are added together.

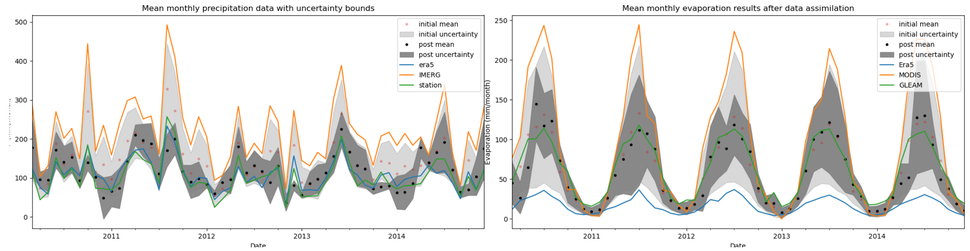

Figure 3. Data assimilation of precipitation (left) and evaporation (right) data.

Results

The lumped HBV model is calibrated for 8 different parameters. The data is divided into sets of 4 years, each time the parameters are calibrated and then tested over the 4 following years. The Nash–Sutcliffe model efficiency coefficient (NSE) is calculated for each calibration, the closer to 1, the higher the predictive skill. The first 1.5 year are taken as spinoff and are not taken into account in calculating the NSE. For each calibration the same range of parameters is used. As can be seen in Figure 4, the NSE is decreasing over the years. Also the NSE of the test data is in lower in most of the cases, with the largest difference in the years 1980-1983 and 1984-1987.

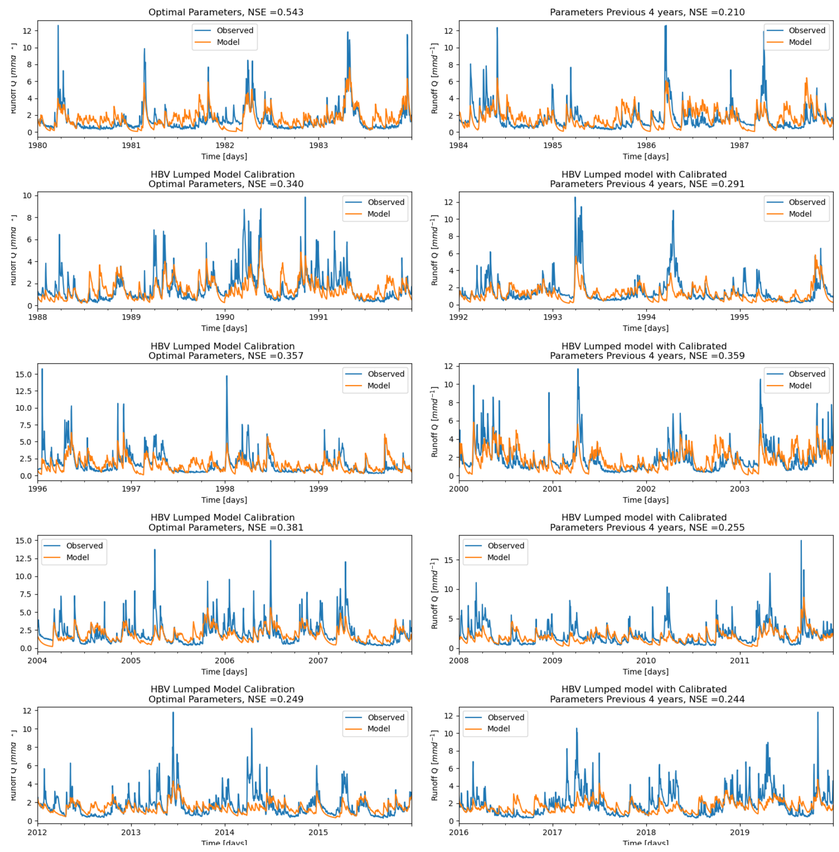

Figure 4. Results of lumped HBV model with highest NSE value.

Conclusion

The modelled runoff has some uncertainties due to the uncertainty in data, but also in the chosen parameters. The 8 model parameters are calibrated using Monte Carlo, trying all different combinations is very time consuming. Almost all parameters show a uniform distribution over the range against the NSE. So a parameter is easily guessed too low. In practice these should be higher, and thus the other parameters will fit much incorrectly.
As can be seen, during the summer months the runoff is overestimated. Next to uncertainty, the occurrence of reservoirs could explain part of the overestimation. In figure 5 the are reservoirs shown which are classified in QGIS. Reservoirs are not taken into account in the model. Using a parameter in the model which reduces the fast flows in the summer, could result in a better fit.

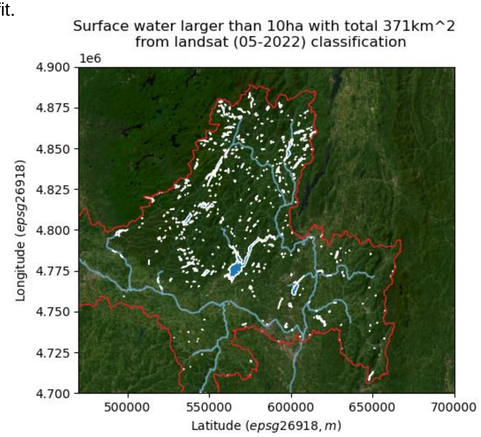

Figure 5.Surface water areas classified by training in QGIS.

Besides this, the distributed model underestimates the discharge during snowfall. The data is taken as average throughout the basin, but in practice it is a heterogenic area. For the distributed model, 21 parameters are needed. This leads to too much uncertainty to use the model for predictions.

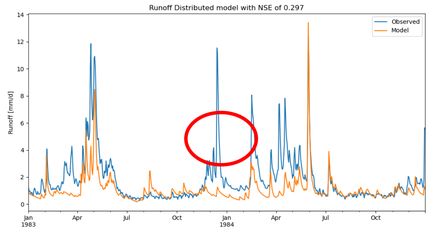

Figure 6. Distributed model underestimating during winter months.

Furthermore, both models have better outcomes when the same parameter range is chosen for the first 20 years based on the NSE. When run several times, the models of the past 20 years always score lower than data used from 1980 till 2000. The uncertainty is large, but it is a remarkable observation.

To calculate both the NSE and the NSE log are used. When using NSE, the runoff is calibrated more towards the peaks, while using the log NSE the baseflow and timing is better modelled.

To improve modelling the peaks more data could be used. Further research could be done by executing multi-objective calibration. This will result in a model which both captures peaks and the base flows.

In this project, the model with highest NSE is chosen, while in practice this could not be the most useful model. Further research could be done by using generalized likelihood uncertainty estimation. Here, the NSE values are weighted and so multiple solutions are taken into account.

All data and code available on GitHub