Student: Timur Sabitov

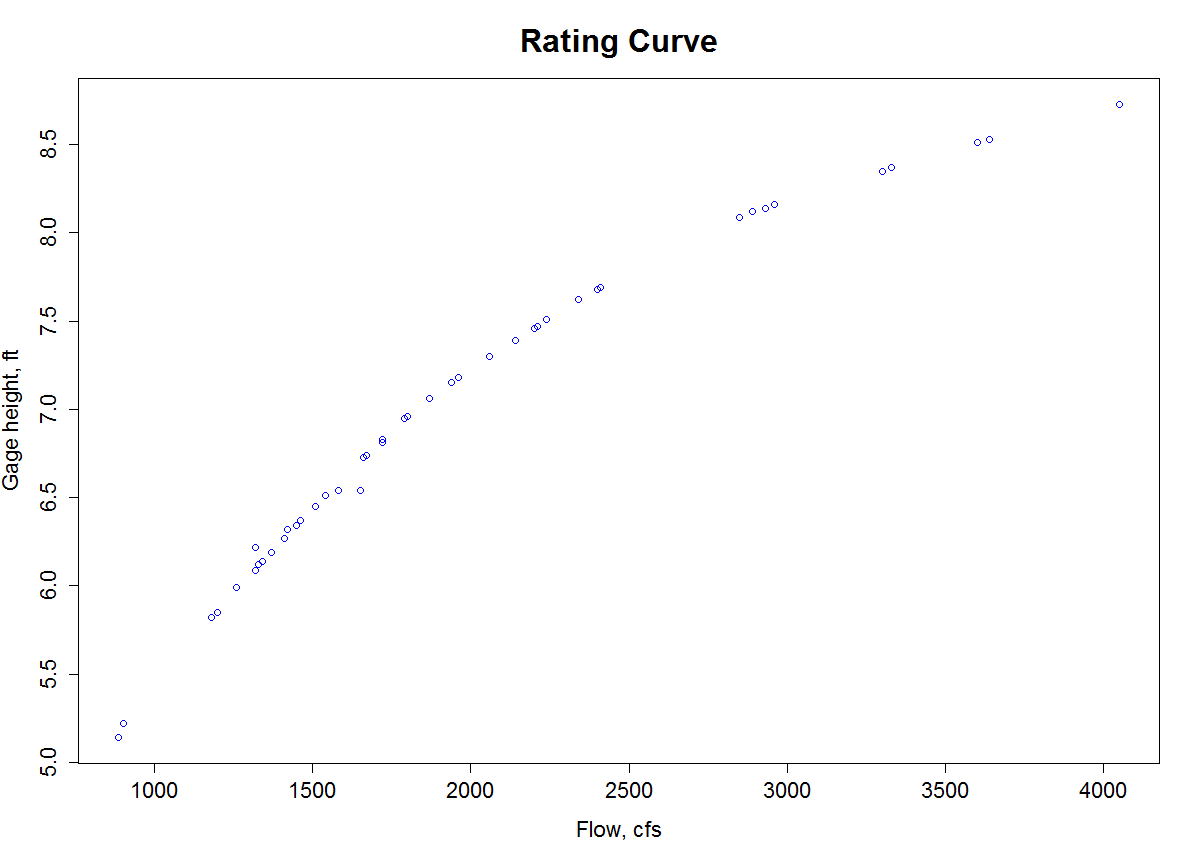
Subject: Hydrologic Modelling

Instructor: Chuck Kroll

**Final Exam**

1. Developing rating curve and the model to estimate peak flow as a function of water level at gage.
2. From the available data on peak flows and gage height measurements at Onondaga Creek at Spencer St. it is possible to see that in 1993 there were no measurements of gage height. To estimate this value all other flows and gage height measurements were extracted and OLS regression performed. With the help of the developed model, we estimated gage height at the missing day as a function of the peak flow at that day.

Rating curve was developed where peak flows at gage are the function of gage height.



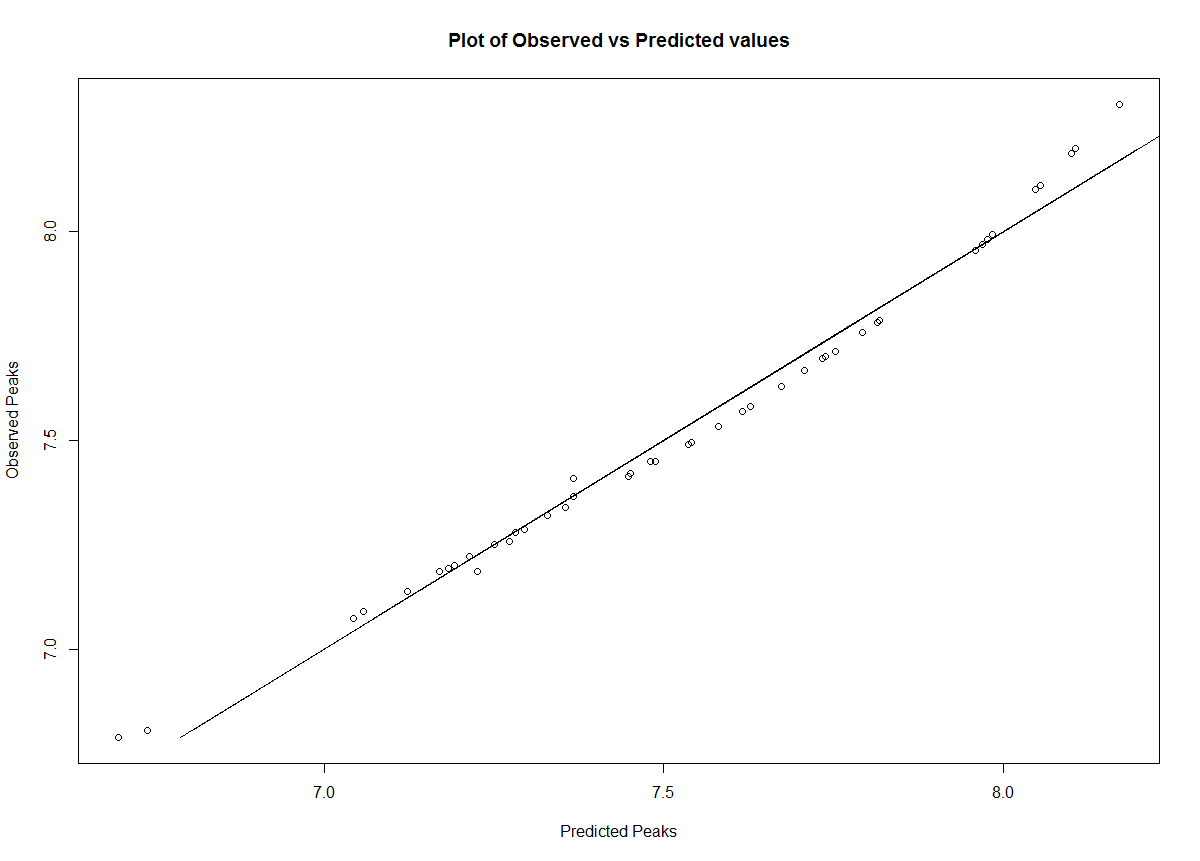
Secondly, we developed a model to estimate peak flows at the gage as a function of gage height and tested model assumptions. We employed OLS regression model. The final model results look like:

Qpeak=exp(α+β\*lg(x))

Where x – is the gage height; α is intercept = 2.141, β is slope = 2.783

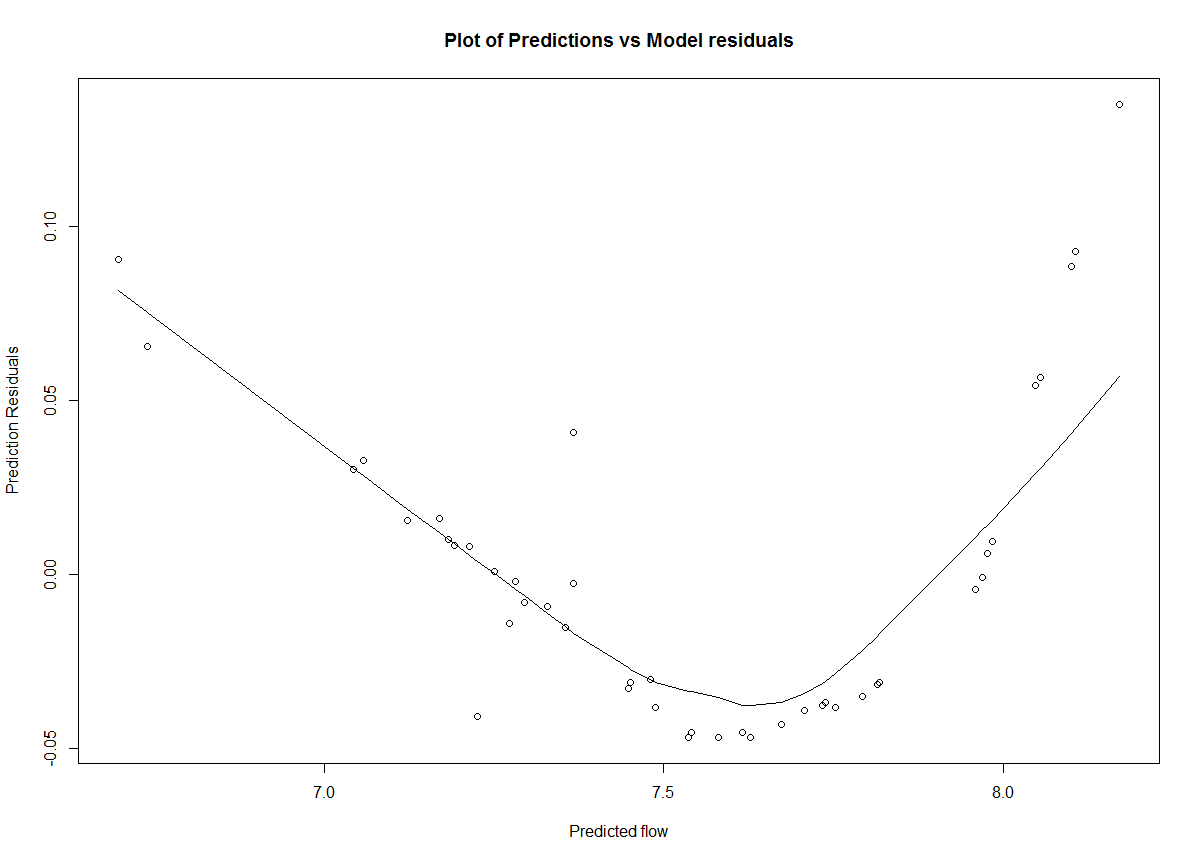
Test of model assumptions:

1. Lack of linearity

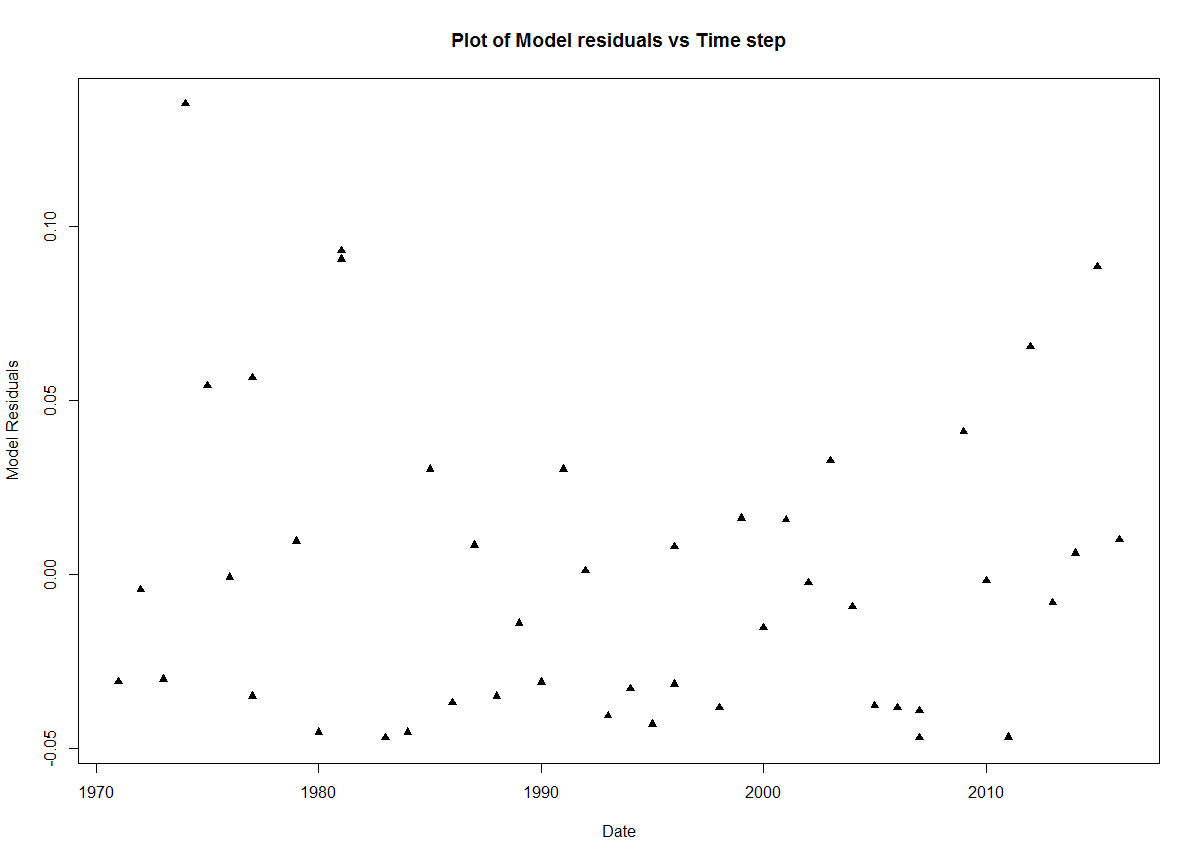


There was no lack of linearity between model predictions and observed peak flows.

However, the plot of predictions residuals vs predicted flows depict bowed shape, which means that there Is an issue with predictions and this is important in the cases where we are extrapolating values. And the assumption of linearity is violated for model residuals.

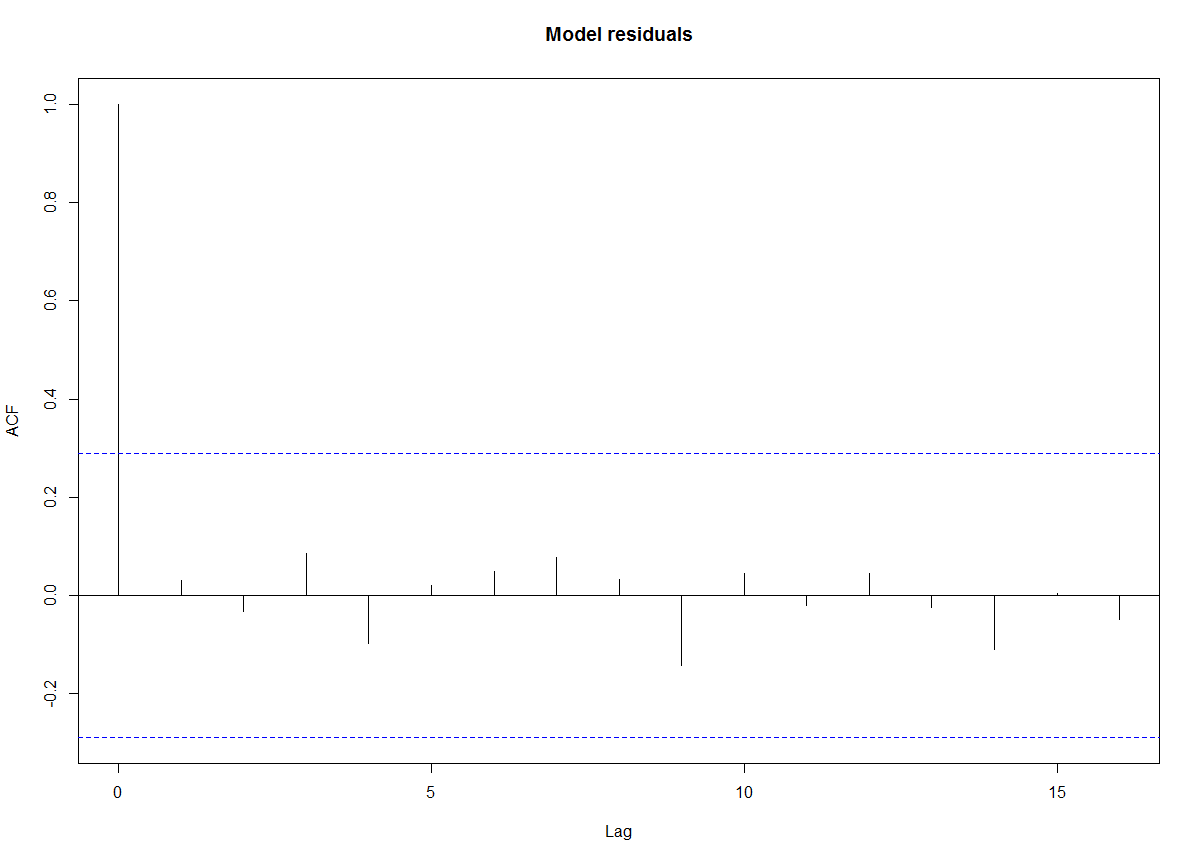


1. Non-constant variance (heteroscedasticity)



From the plot of residuals vs time step, it could be seen that there is an effect of heteroscedasticity, and it means that there is too much of weight on the variable that has large variance, and the assumption of constant variance is violated.

1. Residuals are not independent.



Autocorrelation plot of residuals demonstrated that there is no dependency between lag time and residuals. This assumption is not violated and residuals are independent.

1. Residuals are not normally distributed.

Our test statistics (T.S.) for N – 46 , R(cor) - 0.939, however our rejection region (α) at 5% significance is below 0.978 . As our T.S. < α we conclude that residuals are not normally distributed, and the assumption of normal distribution is violated.

**Model assessment** reveals that there are violations related to our estimates and our final model is not representative. The model could be improved by introducing other variables as drainage area, precipitation, snowmelt, the frequency of storm events.

1. We used the developed model to estimate peak flow at the gage height of 10.5 ft. As a result, peak flow at that gage height following our model is equal to **5914.3** cfs. Extrapolation of values as this one is not recommended.
2. Model to estimate streamflow as a function of gage height is not as helpful as it seems to be. As the channel bed and sides are highly dynamic as well as the velocity of the flow in the stream. These two conditions are highly effecting the amount of the flow flowing through the channel and the relationship to the stream gage height. If one of the sides of channel or riverbed changes, that would affect the height of the water in the streamflow following Archimedes' principle as emerged volume would affect total volume of the flow and therefore the same amount of streamflow could have a different velocity or stream gage height before and after a certain event.
3. Questions

a)

1. Sum of Square Errors (SSE) and Total Sum of Squares (SST).
2. Variance in the depended variable that is described by the independent variable.

b)

AI models decrease the number of degrees of freedom and risk of equifinality in the real-time flood forecasting systems, however, it requires continuous training data (ensembles) for the stream. While physical models for hydrologic prediction does not necessary require training data they could be used for streamflow prediction.

c)

As the Penman method combines energy balance required to support evaporation and adds a term of mass transfer - empirically described diffusion, therefore, it considered as a combination method.

d)

Results of researchers demonstrated a trend of decrease of evaporation from the pan as an indicator of a decrease in total evaporation, and this is contradicting with their own arguments. In general properties of the pan are different from the surrounding environment and there might be more water to evaporate than moisture in the soil. The decrease in the pan evapotranspiration indicates that there is a moisture in the surrounding environment that could be potentially evaporated, thus overall evaporation from the same area still exist and could be more than condition when we have very high pan evaporation that is not always consistent with a soil moisture content.

e)

Because of the decrease in the concentration, there is less area under the curve that determines the concentration of tracer in the flow and our estimates of the flow are upwardly biased (overestimated).

f)

Plotting position determines cumulative distribution function (cdf) and assigns a certain probability of being equal to or less to the value. Plotting position is based on the expected quantiles (observations).

1. Estimation Q100 and Q710

a)

Final Q100 - 8331.74 cfs

Final Q710 – 0.0034 cfs

During the estimation of Q710, rounding of values by 2 decimals produced 0 flow for one of the years, same with LP3 function, as rounding in the function results in the estimates of 0 flow.

b)

|  |  |  |  |
| --- | --- | --- | --- |
| Est. | HW – 2 (at site) | HW -4 (regression | Final (lumped model) |
| Q100 | 26391.09 cfs | 32703.62 cfs | 8331.74 cfs |
| Q710 | 39.83 cfs | 30.52 cfs | 0.0034 cfs |

It is clearly seen that our model estimates are lower. Based on comparison with real values I would suggest that estimates from HW-4 from the regression model are closer to the real values of at site estimates and therefore are more reliable. Our model poorly performs in both estimations of Q710 low flows and Q100.This might be due to the assumed values of unsaturated storage and antecedent moisture content, field capacity of the soil moisture content and value of Kb. Moreover, we used only last 19 years of data instead of 20 as in HW – 2 and HW – 4 and this also could be reflected in final results. High value at the year 20 could impact overall data performance towards higher estimates of Q100 only.