University of Colorado
Department of Civil, Environmental and Architectural Engineering
Advanced Data Analysis Techniques (Statistical Learning Techniques for Engineering and Science)
CVEN 6833
Homework 1
Due: 10/08/2021

Topics: Surface Fitting – *Linear, GLM, Local Polynomials, Linear and Local Polynomial GLMs, GAM, Spatial Models, Hierarchical and Bayesian methods*

Please present your work neatly. Organization of R-commands, functions will fetch 15% of points. Data and sample code from
http://civil.colorado.edu/~balajir/CVEN6833/HWs/HW-1

1. Derive the link function, Fischer score and Information matrix for Poisson distribution. As a bonus do the same for Gamma distribution.

2. Summer (Jun-Sep) and Winter (Dec-Mar) extreme precipitation (3-day max) at ~73 locations across the Southwestern U.S (see figure below) for the period 1964 – 2018 are available at

https://drive.google.com/drive/folders/1uuuZ_9Yi2G9OozeHpLlpWVspD_tXAYUg

In the folders /Winter_temporal and /Summer_temporal

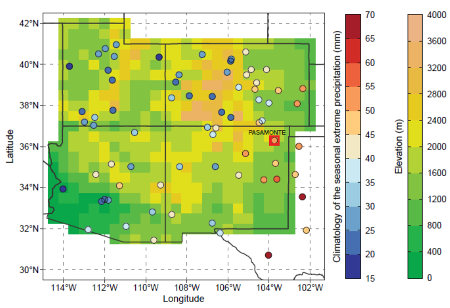

Figure 1. Climatology of the extreme seasonal precipitation for 73 precipitation stations and 0.5-degree elevation grid in (m) of the study area. The red square corresponds to the site of interest, considered in Section 4.

Compute the summer and winter average extreme precipitation at each location – thus, obtaining one value per location. The latitude, longitude and elevation of the locations can be found in the file Precipitation_data.txt. The columns are longitude, latitude, elevation (meters) and summer extreme precipitation average (millimeters)
- The file Elevation_grid_1deg.txt provides information of longitude, latitude and elevation on a high-resolution DEM grid covering this region.

(a) Display the average precipitation – for summer and winter separately - along with the topography as spatial map.

(b) For the winter extreme precipitation, you are interested in obtaining the underlying precipitation surface over this region from these sparse spatial observations – as this will be used in a variety of applications (e.g., flood plain management, natural hazard mitigation etc.). To this end, the first objective is to perform the following:

i. Fit a *best* linear regression model (use one of the objective functions - GCV, AIC, BIC or PRESS; you can also try a couple of them to see any differences). This entails fitting the model with all possible combinations of covariates (Latitude, Longitude and Elevation) and selecting the model with the minimum objective function.

ii. Show the scatterplot of observed and modeled precipitation along with the 1:1 line.

iii. Perform ANOVA (i.e. model significance) and model diagnostics (i.e., check the assumptions of the residuals – Normality, independence, homoskedasticity).

iv. Compute drop-one cross-validated estimates from the best model and scatterplot them against the observed values with the 1:1 line. This and the scatterplot in (ii) above, is to visually see how the model performs in a fitting and cross-validated mode.

v. Drop 10% of observations, fit the model (i.e., the 'best' model from i. above) to the rest of the data and predict the dropped points. Compute RMSE and correlation and show them as boxplots.

vi. Spatially map the model estimates and the standard error from the *best model* on the high-resolution grid

vii. Briefly discuss what you find [bullet points are fine]

3. Repeat 2b. by fitting a ***GLM*** with appropriate link function.

4. Repeat 2b. with ***Local polynomial method***.

5. Repeat 2b with Local Polynomial method but using the appropriate link function (i.e., '***Local GLM***').
[For the Local Polynomial approach the 'best model' involves fitting the best subset of predictors and the smoothing parameter, alpha. You can also compare the GCV from these four different methods.]
Briefly discuss the results from the local polynomial approach and compare them to linear regression.

6. Repeat 2b by fitting a ***Generalized Additive Model*** (GAM) and compare with the GAM fitted in a local polynomial framework.

7. Estimate the spatial surface using ***Kriging***
   i. Fit a variogram
   ii. Repeat iv – vi of problem 2b.

8. Repeat 7. with a ***Hierarchical Spatial Model. [Fit a best linear model as in problem 2 and perform Kriging on the residuals]***

9. Repeat 8. with a ***Bayesian Hierarchical Spatial Model*** (see Verdin et al. 2015, for tips)
[In the Bayesian models, plot the posterior historgram/PDF of the parameters; spatial maps of posterior mean and standard error]

10. I want you to reflect on the suite of analyses performed above on the spatial precipitation data – in particular, the relative performance of the methods, their advantages and disadvantages and potential application of these methods on a problem/data set of your interest. Keep this short and crisp (bullet points are fine).