University of Colorado
Department of Civil, Environmental and Architectural Engineering
Advanced Data Analysis Techniques
CVEN 6833

Homework Set 3 Date :04/05/2025
Due by:05/06/2025

Topics: Parametric/Nonparametric Time Series, Hidden Markov Model, Wavelet Spectral Analysis, Extreme Value Time Series and Copulas

Please present your work neatly. Organization of R-commands, functions will fetch 15% of points. Data, commands etc. at
http://civil.colorado.edu/~balajir/CVEN6833/HWs/HW-3/

K-NN Conditional Ensemble Forecast/Simulation

1. Use K-NN bootstrapping approach to simulate/forecast all India Summer (June-Sep) Monsoon Rainfall (AISMR) based on suite of predictors, also called, 'feature vectors' based on tropical Sea Surface Temperatures during the summer season for the period 1871 - 2016
The predictors to consider are NINO12, NINO3, NINO4, WPI, TROPGRAD and IOD. You can experiment with all or subset of these variables.

The steps for each lead time are:
(i) For each year, *t,* using the feature vector in that year, $\boldsymbol{x}_t$ *obtain* K-nearest neighbors from the historical feature vectors (of course, excluding the year, *t*)
(ii) Select one of the K neighbors using a weight metric. The selected neighbor corresponds to a historical year and with it the associated spring streamflow.
(iii) Repeat steps (ii) say 100 times, to obtain *ensemble simulation* of streamflow for each year t. Compute the mean or median of the ensemble to get a single value.
(iv) Repeat steps (I) – (iii) for all the years and similarly for the two lead times.
(v) Plot the historic flow vs ensemble mean along with the 1:1 line for visual comparison. Compute skill scores for each lead time - correlation between the historic and the mean of the ensemble forecast rainfall, also compute RPSS. Comment on what you find.
(vi)***Bonus:*** For the 2000 – 2016 period Use rainfall data pre-2000 for K-NN resampling. Plot the ensemble forecast and historic rainfall

Conditional Forecast/Simulation - Copulas

2. (i) Repeat problem 1 using Copula Regression – 'gcmr' package in R. There will not be an ensemble forecast, but a mean value.
(ii) Fit a Copula to the spring streamflow and April 1 SWE.

Support Vector Machines

3. Fit a Support Vector Machine (SVM) model for AISMR using predictors. Obtain estimates of flow from the models and compare their performance with historic flows.

Modeling Nonstationary Time Series - HMM & Forecast and simulation

4. Another way to model/simulate a time series is using HMM. For the all India summer monsoon rainfall time fit a HMM and make the forecast. The steps are as follows:

- Fit a best HMM model for the rainfall
- Fit a best GLM (mostly logistic regression) to the state sequence as a function of predictors– i.e., NINO12, NINO4, WPI, IOD and TROPGRAD
  $S\_t = f(S\_t\text{-}1, \boldsymbol{x}_t)$
- Using this best GLM, for each year, *t,* based on the predictor vector obtain the probabilities of the *states* (i.e. the distribution of the HMM)
- Using these state probabilities, simulate flow from the corresponding *state* PDFs – to obtain an ensemble

- Plot the historic vs ensemble mean rainfall along with the 1:1 line for visual comparison. Compute skill scores – RPSS and correlation between the historic flow and the mean of the ensemble forecast.

5. Compare and comment on the results and methods (pros/cons/utility) employed in problems 1 ~ 4.

Multivariate Nonstationary Extreme Value Analysis (EVA)

6. Implement the multivariate EVA in Bracken et al. (2017). The data and codes from this paper are available here

In this the three variables are maximum reservoir elevation, maximum SWE and maximum flow.

- Fit a nonstationary EVA model for each of the variable
- Fit a Copula to capture the dependence between the variables
- For each year obtain the ensemble of 2-year and 100-year return level of the reservoir level
- Compare with and without the Copula

Bonus: Bracken et al. (2017) implemented a Bayesian version (with and without the Copula). The codes can be found at the link above. You can implement their code or write your own.

Singular Spectrum Analysis – Diagnostics & Forecasting

7. For the AISMR perform SSA and makepredictions from it. The steps are as follows:

Diagnostics

a. Select a window size of about 10-20 years (feel free to experiment with the window size); create the Toeplitz matrix and perform SSA.
b. Plot the Eigen spectrum and identify the dominant modes.
c. Reconstruct the dominant modes (i.e. Reconstructed Components - RCs) and plot them. Infer from them the dominant periodicities.
d. Sum the leading modes and *plot them along with the original time series*. This will show the 'filtering' capability of SSA. Feel free to play with the number of RCs.
e. Plot the dominant modes and show their corresponding wavelet spectra.

Prediction

f. Use the 'feature vector' from problem 1 to simulate/project the leading modes and use them to make projections/simulations. Specifically Predict the recent period 2000 - 2016
g. The steps are - apply the SSA to data for the pre-2000; make a prediction for 2001 and repeat for each year through 2016.
h. *Plot the observed and predicted values; compute the median correlation.*