GLY 6932/6862 – Quantitative Methods in Earth Sciences

Assignment 4 – Time Series Analyses: Auto- and Cross-Correlation

The detailed description of the assignment is provided below. As for the deliverables, as before, use MATLAB's publishing capabilities to generate a well-documented html file to present your code and plot, and email me a link to that url. Please use the following subject for the email in which you send me this assignment: 'GLY6862 – Assignment 4 – *lastname*', where '*lastname'* is of course, your last name. This will make it much easier for me to keep track of your assignments. Thanks.

Part A Description:

This portion of the assignment is a "warm-up" for which you've been prepped with some code presented in class. The goal here is to write some clean code to conduct an autocorrelation analysis. Shown below is a table of monthly runoff values for a stream in Kentucky (data set freely available from publisher of: "Statistics and Data Analysis in Geology" by John C. Davis. http://www.kgs.ku.edu/Mathgeo/Books/Stat/index.html). The data are also provided in a single column text file on our course website (`cavecreek.txt`), but the table is shown here so that you can verify the structure of the data in the text file. Examine this table and look for seasonality in the data set to help verify that your time series plot (instructions below) is making sense.

Table 4–20. Monthly runoff of Cave Creek, Kentucky, given in hundredths of inches. Data are collected during a "water year" beginning in October of the preceding calendar year and ending in September, in order to avoid breaking the sequence in the middle of the winter (after Haan, 1977).

Water year	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.
1953	2	5	19	240	86	416	147	354	31	18	7	1
1954	0	2	4	54	22	40	139	35	8	7	6	14
1955	2	4	30	73	463	579	59	197	55	24	28	3
1956	4	6	13	59	637	469	192	28	32	64	38	8
1957	7	10	172	308	325	103	392	68	24	6	5	2
1958	3	106	432	200	221	117	235	236	19	369	170	12
1959	6	9	17	270	195	112	102	24	24	5	4	2
1960	3	36	269	219	313	291	68	19	364	138	14	30
1961	12	52	79	204	295	532	476	414	159	48	18	4
1962	2	6	76	346	401	508	330	79	96	30	8	7
1963	39	141	124	150	146	548	52	25	14	29	11	3
1964	1	4	3	87	173	788	45	21	11	8	2	16
1965	15	7	347	276	230	449	146	31	8	5	1	2
1966	4	2	2	48	281	79	202	332	25	14	41	11
1967	7	119	357	97	161	466	50	476	33	14	15	7
1968	9	38	271	135	98	425	238	199	91	29	75	16
1969	14	22	112	278	216	73	237	74	40	27	66	17
1970	7	25	91	130	389	291	568	206	38	14	6	27

Part A Tasks:

Write a script (.m file) that does the following:

1. Loads the Cave Creek Runoff data set (`cavecreek.txt`), available from the course website, which contain a single column of monthly runoff values (in hundredths of an inch) from Cave Creek in Kentucky.
2. Constructs a vector of date numbers (serial) that correspond to the monthly mean data values for Oct. 1952-Sept. 1970. Be smart here – how do you deal with strange things like February and leap years?
3. Calls a function (`ser_corr_fcn.m`), provided to you, which conducts an autocorrelation analysis and reports back a vector of lags and a vector of correlation coefficients corresponding to the various lags.
4. Plots up the data and the autocorrelation analysis in a visually digestible form.

Part B Description:

This portion of the assignment will provide you an opportunity to practice your data "integration" skills – working with multiple different data sets – and illustrate the value of cross correlation analysis. For this assignment, we will work with monthly river discharge data and two well-established climate indices to test a hypothesis presented in a well-known and much discussed paper:

Milliman, J.D., and Syvitski, J.P.M. (1992) Geomorphic/tectonic control of sediment discharge to the ocean: The importance of small mountainous rivers. Journal of Geology; v.100(5), 525-544.

The fundamental hypothesis developed in their paper is that the freshwater and sediment discharges of small mountainous rivers are more susceptible to episodic events because the small catchment size and lack of floodplains in mountainous rivers prevent much storage of water and sediment. This hypothesis was further tested for California rivers and presented in the following abstract at GSA in 2003:

INFLUENCES OF PDO AND ENSO ON DISCHARGE FROM COASTAL CALIFORNIA RIVERS

FARNSWORTH, Katherine L. and MILLIMAN, John D., School of Marine Science/VIMS, College of Williams & Mary, Gloucester Pt, VA 230622003 Seattle Annual Meeting (November 2–5, 2003) Paper No. 209-11

Streamflow is largely dependent on regional climate, which in turn is related to larger scale phenomena occurring in the oceans and the atmosphere. Much work has been done recently to improve our understanding of the El Niño-Southern Oscillation (ENSO) and its relationship to climate and streamflow throughout the United States. More recently the Pacific Decadal Oscillation (PDO) has been characterized as influencing continental climate of North America, with the highest correlation in the Pacific Northwest (Mantua et al., 1997; Latif and Barnett, 1994). PDO events persist for 20 – 30 years, with evidence of reversals in the oscillation occurring around 1925, 1947, 1977, and possibly in 1999. Here we show that the signature of both ENSO and PDO are evident in annual discharge from coastal California Rivers. For example the Salinas River of central California has significantly higher

```
annual discharge during warm phases of ENSO (El Niño) than during non-El Niño
years. This is due to the high likelihood of strong winter storms in central
California during El Niño years. During warm phases of the PDO there are also
significant increases in annual discharge (cool phases have significant
decreases). High flow years during the warm phase of PDO are all El Niño
classified years, while El Niño years during cool-phase PDO produce
significantly lower discharge. This is likely due to the interaction between
the PDO and ENSO, which favors El Niño like conditions during warm PDO phases
(Mantua et al, 1997).
```

Your task is to design an analysis of the San Lorenzo River discharge data set to support or refute the observations (presented above) for other coastal California rivers. To do so, you will need to consider numerical indices of the El Niño/Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO).

Part B Data:

The monthly data from the San Lorenzo River near Big Trees CA (USGS Stn. 11160500) is available on the course website, [but originally came from: http://waterdata.usgs.gov/nwis/nwisman/?site_no=11160500&agency_cd=USGS]

The time series for the Multivariate ENSO Index (MEI) is available on the course website, [but originally came from: http://www.cdc.noaa.gov/people/klaus.wolter/MEI/mei.html]

The time series for the PDO Index Monthly Values is available on the course website, [but originally came from: http://jisao.washington.edu/pdo/]

Part B Tasks:

1. Create an m-file that imports the monthly streamflow data downloaded in the more standard USGS tab-delimited file format, and import the MEI and PDO indices.

2. Plot the data sets in a format of your choosing to allow for visual comparison.

3. Perform some type of cross-correlation between the data series to test the conclusions in the Farnsworth and Milliman abstract. You will have to write a script to perform the cross-correlation analysis and for this, you should the function provided to conduct the autocorrelation analysis in Part A (ser_corr_fcn.m).

Helpful Hints:

- The method by which you do the cross-correlation is up to you, but you need to keep in mind that there are likely to be monthly to annual lags in the response of river discharge related to the MEI.

- Consider a simple way of grouping discharges related to ENSO and the two main regimes of the PDO (e.g. El Nino years versus La Nina years versus La Nada years, etc.).

- You must show that there is or is not a statistically significant correlation between discharge and ENSO/PDO. To help you with this revisit our class notes where we looked up the critical value for r, the correlation coefficient.

- Beware of OUTLIERS; they can significantly alter your correlations. Some researchers use quantitative methods to exclude outliers. For example, they exclude observations that are outside the range of ±3 standard deviations around the group mean. If you chose this approach, you may want to consider the mean of the log(discharge) and not simply the mean of discharge (why?).