Time Series HW 5

Allison Theobold, Andrea Mack, and Kenny Flagg

October 6, 2016

Groups of up to 3 allowed.
Read Vincent and Mekis's 2006 paper and answer the following questions. I am also posting Zhang et al. (2000) for the reference on the method of trend estimation used.
When you get to Table 3, it is helpful to add a vertical linear between the first "positive" column and the second "negative" column to understand what is presented.
For the following questions, ignore the "national trends" part of it except for possibly the last question. There is plenty to discuss without that piece of the paper.

1. *What is/are the research questions in Vincent and Mekis? You can be a little bit general here – just a couple of sentences on what they are trying to explore.*

   The goal of the author's work is to present trends in various indices of "daily and extreme temperatures and precipitation in Canada for the periods, 1950 - 2003 and 1900 - 2003." They extend the analysis of *Bonsal* and *Zhang* with use of the latest versions of homogenized daily temperatures and adjusted daily precipitation measurements. They have updated the indices based on the definitions by ETCCDMI, so that their analysis of Canadian trends is comparable with the rest of the world.

2. *How many response variables are they analyzing? Explain one temperature and one precipitation based response that they are using in detail.*

   They are analyzing a total of 18 response variables (8 temperature indices and 10 precipitation indices).

   - One temperature index is the **standard deviation of Tmean**, which they include so they can study trends in temperature variability. This is the standard deviation of the daily mean temperature anomaly, where the mean anomaly is reckoned as the difference between the mean temperature for each day and the average of all available daily mean temperatures for 1961 to 1990, and then the standard deviation is computed for each year. Thus this is a measure of the variation in "overall" temperature within the year. I say "overall" to contrast this with the standard deviation of Tmin or Tmax, which should be larger because the latter are based only on the most extreme temperatures of each day; the standard deviation of Tmean should primarily pick up variation across months and seasons.

   - One precipitation index is the **simple day intensity index of P**, defined as the total annual precipitation (inches of rain plus water-equivalent-inches of snowfall) divided by the number of days with measurable levels of precipitation. This measures the daily average amount of precipitation for the days that it precipitated that year. It should be noted that the number of days with precipitation is another of the indices they are analyzing. It would seem more appropriate to model several of their indices simultaneously as a multivariate response, but of course that type of model is very complicated and difficult to fit. We at least should be aware of the dependency among responses and not be surprised when sites that have a trend in the number of days with precipitation also have a trend in the simple day intensity index of P.

3. *How are they assessing trends? You will need to read pages 402-405 in Zhang et al. (2000) to get all the details. I just want the general discussion of the method – do this in a few sentences. Then note their reason(s) for using this method?*

   To assess the trends, they are using the same method as *Zhang, et al.*, modeling each climate index with both a linear time trend, serial correlation between each time point and the one previous to it. They fit the model as follows:

   - They first estimate ρ, the lag 1 autocorrelation, using the original data (Y_t and Y_{t-1})
   - They then de-correlate the data ($W_t = Y_t - \rho(Y_{t-1})$)
   - Then using the de-correlated data (W_t) they estimate the linear trend, β, of the climate variable with time. This is done by taking the median of all the estimated slopes of two pairs in the de-correlated data, $\{(W_i, W_j) : i, j = 1, \ldots, n\}$.
   - Using the estimated linear trend, β, they then de-trend the original data to obtain a new estimate of ρ. Thus, they find $Z_t = Y_t - \beta(t)$, and estimate the lag 1 autocorrelation, ρ, using these de-trended Z_t and Z_{t-1}.
   - This process is repeated until the consecutive estimates of β and ρ are "close enough". It typically takes 3-12 iterations.

4. *I think I successfully extracted their counts from Table 3 into the provided Table3.csv file but you should check my results. Use that information to provide the total number of tests they considered for the shorter and then longer series analyses and the total number of each of those that were "significant". What is the minimum and maximum number of tests performed across the different response variables for each length of series?*

   As far as I can tell, you successfully extracted the authors' counts. In the longer series, the minimum number of tests was 59, done on diurnal temperature range (a temperature measurement) while the maximum number of tests was 76, done on the annual snowfall precipitation (a precipitation measurement). In the shorter series, the minimum number of tests was 148 on summer days (a temperature measurement) while the maximum number of test was 239 on highest 5 day precipitation amount (a precipitation measurement).

   The total number of tests done on temperature indicies was 1849 and the total number of tests done on the precipitation indicies was 2880. The total number of tests done on the shorter time period, 1950-2003 was 3441 and the total number of tests done on the longer time period, 1900-2003 was 1288. The total number of significant tests found in the shorter time period, 1950-2003 was 919 and the total number of significant tests found in the longer time period, 1900-2003 was 580.

	Longer	Shorter
Total Tests	1288	3441
Total Significant	580	919

5. *Again focused on Table 3, if all tests are performed at the 5% level, overall how many results should they have expected to be "significant" within the shorter and then longer trends being considered? How does that compare to their total number of "significant" results on each? Do the same comparison for each response variable in the table (total expected for that variable and number observed for shorter and then longer series). Are there any response variables where the difference in observed "significant" results and expected is such that you don't think there really is much overall evidence on that variable?*

   At the 95% confidence level, we cannot say that exactly 95% of the tests done in a study will detect actual differences for a fixed number of tests. But in the long run, we expect 95% of tests to come to correct conclustions. By chance, we can estimate that 5% of the tests will falsely find significant results. For the tests done in the shorter time frame, that is about 172.05 and in the longer time frame, that is about 64.4 that we expect to wrongly yield significant results.

	Estimated FDR	Total Significant
Longer	1224	580
Shorter	3269	919

Table 1: Shorter time series

Total Tests	Total Signif	Expected Signif	Percent Signif	Index	Type
159	49	8	31	Frost days	Temp
167	41	8	25	Cold days	Temp
167	71	8	43	Cold nights	Temp
148	30	7	20	Summer days	Temp
167	38	8	23	Warm days	Temp
167	58	8	35	Warm nights	Temp
155	36	8	23	Diurnal temp range	Temp
162	19	8	12	Std dev of Tmean	Temp
221	44	11	20	Annual snowfall	Precip
209	48	10	23	Snow to total precip ratio	Precip
208	107	10	51	Days with precip	Precip
214	146	11	68	Days with rain	Precip
208	66	10	32	Simple intensity of P	Precip
214	71	11	33	Simple intensity of R	Precip
208	36	10	17	Max consecutive dry days	Precip
239	16	12	7	Highest 5-day precip	Precip
214	16	11	7	Very wet days	Precip
214	27	11	13	Heavy P days	Precip

Table 2: Longer time series

Total Tests	Total Signif	Expected Signif	Percent Signif	Index	Type
65	40	3	25	Frost days	Temp
75	32	4	19	Cold days	Temp
74	62	4	37	Cold nights	Temp
62	13	3	9	Summer days	Temp
75	18	4	11	Warm days	Temp
74	40	4	24	Warm nights	Temp
59	34	3	22	Diurnal temp range	Temp
73	16	4	10	Std dev of Tmean	Temp
76	22	4	10	Annual snowfall	Precip
72	24	4	11	Snow to total precip ratio	Precip
72	52	4	25	Days with precip	Precip
73	54	4	25	Days with rain	Precip
72	42	4	20	Simple intensity of P	Precip
73	41	4	19	Simple intensity of R	Precip
72	47	4	23	Max consecutive dry days	Precip
75	10	4	4	Highest 5-day precip	Precip
73	12	4	6	Very wet days	Precip
73	21	4	10	Heavy P days	Precip

The lower the number of significant tests, the more I am concerned. Across all indices for both temperature and precipitation measurements, by chance we should expect 7-12 significant results. Particularly in the *Very Wet Days Index* for precipitation, we see that 7% of the tests were significant, which is getting very close to the 5% that we expect to be significant by chance when setting the 95% confidence level. Corrections for multiple testing should have been used.

6. *Consider one panel of one of the figures that displays the trend and test results on a map of Canada. Note which figure you are discussing.*

   (a) *What is being displayed and what does that suggest about changes on this metric? Is there any issue with their display choices?*

   Figure 4c shows the estimated change in mean number of days with precipitation from 1951 to 2003 at each weather station in their dataset. This is computed as the estimated slope of the linear trend multiplied by 54 years. The points are colored green where the trend is increasing and orange where the trend is decreasing. Small Xs mark locations where the slope has p-value > 0.05; filled circles represent locations where they deemed the trend "statistically significant" and the size of the circle corresponds to the magnitude of the slope. Locations in the center of the country tend to have small and negative trends. Locations elsewhere generally have large green circles indicating increases in the number of days with precipitation, although a couple locations on the east coast have decreasing trends. Some locations in British Columbia have much larger increases than elsewhere, as much of that province is covered by a few big green spots.

It is interesting to compare these trends to the trends in the simple day intensity index of P (Figure 4e). Locations with positive trends in number of days with precipitation tend to have negative trends in intensity, suggesting that precipitation is becoming more spread out through the year.

The figure is set up so that "significant" trends are easier to see than weaker trends. The big circles overlap so it is hard to see individual locations. In particular, it is impossible to see Xs that are covered up by large circles of the same color. The presentation could be dramatically improved by replacing the Xs with circles of a different color and overlaying them on top of the plot. This figure also misrepresents the central part of the country because it looks like there is less going on in the center than in the south and along the coasts (where there are large swaths of color), when really there is just less information available about the central region. I would use points of the same size and a continuous color scale. This alternative is not perfect because it is difficult to see small differences in color, but it would be a more honest portrayal of the amount of information available.

(b) *Does it appear that there is some sort of spatial structure to the trend results? What does this suggest about the tests on that variable?*

There absolutely does appear to be spatial structure. Locations near each other have circles of similar size and color, meaning that they have similar estimated trends. This suggests positive spatial autocorrelation. Analogously to temporal autocorrelation, ignoring positive spatial correlation leads to underestimating the standard errors, meaning that their p-values are too small and so their tests are more liberal than they intended.

7. *Provide at least two things you might have done differently if you were involved in this paper. This can be in terms of methods used, the way things were interpreted, or graphical presentation of results. Assume that the data set is more or less what you have to work with and you want to address the same research questions.*

- I would have implemented model checking, to see if the AR(1) used to model the different climate variables would produce data similar to that which was observed.
- Additionally, I would recommend fitting a similar model using Bayesian methods. We fit a similar model in 506, using JAGS, which would provide the researchers with the entire posterior distributions for the seasonal effects, τ_i, as well as the posterior distribution of ρ.
- The author's justification for what months and years were omitted due to missing values was sound, but I would be interested to see if imputation of the missing data would result in slightly different results.

References

Vincent, L. and Mekis, E. (2006) Changes in Daily and Extreme Temperature and Precipitation Indices for Canada over the Twentieth Century. *Atmosphere-Ocean*, 44(2): 177-193.

Zhang, X., Vincent, L., Hogg, W., and Niitsoo, A. (2000) Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-Ocean*, 38(3): 395-429.

R Code

4.
```
tb3 <- read.csv("Table3.csv")
tb3_temp <- tb3[1:8,c(1,3:8)]
temp_tests <- colSums(tb3_temp[,2:7])
tb3_precip <- tb3[10:19,c(1,3:8)]
precip_tests <- colSums(tb3_precip[,-1])

temp_tot <- sum(temp_tests)
precip_tot <- sum(precip_tests)

cat("The total number of tests done on temperature indicies was ", temp_tot,
    " and the total number of tests done on the precipitation indicies was ",
    precip_tot, ".\n", sep = "")

tb3_shorter <- tb3[-9,c(1,3:5)]
tb3_longer <- tb3[-9,c(1,6:8)]

shorter_tot <- sum(colSums(tb3_shorter[,-1]))
longer_tot <- sum(colSums(tb3_longer[,-1]))

cat("The total number of tests done on the shorter time period, 1950-2003 was ",
    shorter_tot, " and the total number of tests done on the longer time period, 1900-2003 was ",
    longer_tot, ".\n", sep = "")


shorter_sig <- shorter_tot - sum(tb3_shorter[,3])
longer_sig <- longer_tot - sum(tb3_longer[,3])

cat("The total number of significant tests found in the shorter time period, 1950-2003 was ",
    shorter_sig,
    " and the total number of significant tests found in the longer time period, 1900-2003 was ",
    longer_sig, ".", sep = "")

tb.out <- data.frame(matrix(c(longer_tot, shorter_tot, longer_sig, shorter_sig),
                            nrow=2, ncol = 2, byrow = TRUE))
rownames(tb.out) <- c("Total Tests", "Total Significant")
colnames(tb.out) <- c("Longer", "Shorter")

print(xtable(tb.out, digits = c(0,0,0)))
```

5.
```
cat("At the 95\\% confidence level, we cannot say that exactly 95\\% of the tests
done in a study will detect actual differences for a fixed number of tests. But in
the long run, we expect 95\\% of tests to come to correct conclustions. By chance,
we can estimate that 5\\% of the tests will falsely find significant results. For
the tests done in the shorter time frame, that is about", 0.05*shorter_tot,
"and in the longer time frame, that is about", 0.05*longer_tot,
"that we expect to wrongly yield significant results.")

tb.sig <- data.frame(matrix(c(0.95*longer_tot, longer_sig, 0.95*shorter_tot, shorter_sig),
                            nrow = 2, ncol = 2, byrow = TRUE))
colnames(tb.sig) <- c("Estimated FDR", "Total Significant")
rownames(tb.sig) <- c("Longer", "Shorter")
print(xtable(tb.sig, digits = c(0,0,0)))
```

```
shorter_var.tot <- apply(tb3_shorter[,c(2:4)], 1,sum)
longer_var.tot <- apply(tb3_longer[,c(2:4)], 1,sum)

shorter_var.sig <- apply(tb3_shorter[,c(2,4)], 1,sum)
longer_var.sig <- apply(tb3_longer[,c(2,4)], 1,sum)

shorter_var.prop <- shorter_var.sig/shorter_var.tot*100
longer_var.prop <- longer_var.sig/shorter_var.tot*100

shorter_var.expected <- shorter_var.tot*0.05
longer_var.expected <- longer_var.tot*0.05

shorter_tb.all <- data.frame(cbind(shorter_var.tot, shorter_var.sig,
                                   shorter_var.expected, shorter_var.prop))
shorter_tb.all$index <- c(paste(tb3_shorter[,1]))
shorter_tb.all$type <- c(rep("Temp", 8), rep("Precip", 10))

colnames(shorter_tb.all) <- c("Total Tests", "Total Signif", "Expected Signif",
                              "Percent Signif", "Index", "Type")
longer_tb.all <- data.frame(cbind(longer_var.tot, longer_var.sig,
                                  longer_var.expected, longer_var.prop))
longer_tb.all$index <- c(paste(tb3_longer[,1]))
longer_tb.all$type <- c(rep("Temp", 8), rep("Precip", 10))

colnames(shorter_tb.all) <- c("Total Tests", "Total Signif", "Expected Signif",
                              "Percent Signif", "Index", "Type")
colnames(longer_tb.all) <- c("Total Tests", "Total Signif", "Expected Signif",
                             "Percent Signif", "Index", "Type")
print(xtable(shorter_tb.all, digits = c(rep(0,7)),
             caption = "Shorter time series"), include.rownames = FALSE)
print(xtable(longer_tb.all, digits = c(rep(0,7)),
      caption = "Longer time series"), include.rownames = FALSE)
```