Water Cycle: Lessons using data skills

Royal Geographical Society with IBG

Advancing geography and geographical learning

Lesson 2: The implications of changes to the global water balance - growing season length in the USA

Lesson Objective

- To describe and discuss the potential implications of a shifting pattern in growing season length in the United States of America
- To use a Mann-Whitney *U* test to determine whether a statistically significant differences exists between the length of the growing season in the Eastern states of America as compared to the Western states of America
- To use a Spearman's Rank Correlation Coefficient to assess the degree of association between the growing season length and frost timing

Setting the Scene

Changes to the global water balance and therefore changes to the supply of water have the potential to severely affect human and economic wellbeing. For example, due to changes in hydrological parameters such as temperature and precipitation, the length of the growing season for crops in the USA has changed in almost every state. States in the Southwest (e.g. Arizona and California) have seen the most dramatic increase. In contrast, the growing season has actually become shorter in a few south eastern states. This has implications for the local economy and local food security as well as the global economy in terms of export of food products from the USA. This lesson is focused on beginning to explore this problem.

The Data

Changes in growing season length have been calculated using temperature data from 750 weather stations throughout the contiguous 48 states. These data were compiled by the National Oceanic and Atmospheric Administration's National Centres for Environmental Information. Growing season length and the timing of spring and fall frosts were averaged across the nation, then compared with long-term average numbers (1895–2015) to determine how each year differed from the long-term average.

1) **Changes to Growing Season Length**

Task

Open the Microsoft Excel Growing Seasons data file. In the Length of Growing Season tab there are three columns: 1) year 2) deviation from average length of growing season and 3) long term average.
Using the data plot a line graph of the year and annual temperature anomaly. You can do this by hand or use excel (it's much quicker!)
Your graph should look like the example below in Figure 1.

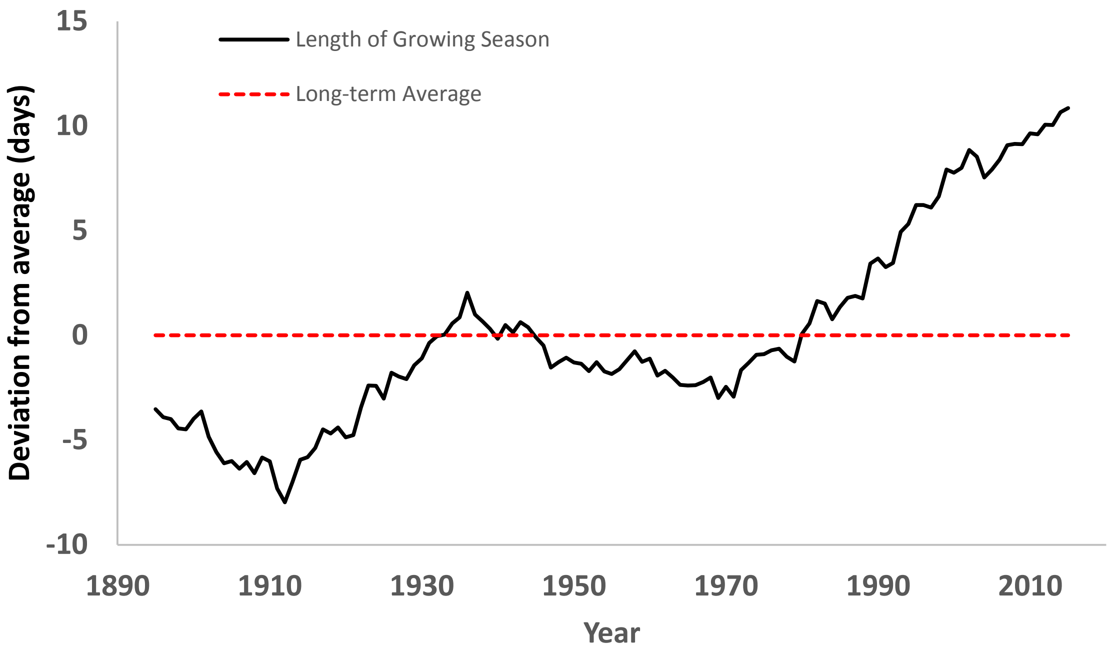

Figure 1: The length of the growing season in the contiguous 48 states compared with a long-term average. For each year, the line represents the number of days shorter or longer than average. Choosing a different long-term average for comparison would not change the shape of the data over time.

Look at Figure 1 and discuss the findings. You might want to think about:

- Is the deviation from the long term average of growing season length consistent through time or is there a specific period of time during which growing season length seems to be changing more rapidly?
- What factors might be driving the growing season trends? For example could this be related to broader changes in other parameters such as changes to precipitation?

2) East vs West – are changes to growing season length spatially consistent

Figure 1 shows that the average length of the growing season in the contiguous 48 states has increased by nearly two weeks since the beginning of the 20th century. A particularly large and steady increase occurred over the last 30 years (see Figure 1). However data also shows that this growth is not spatially consistent across the USA. The following tasks allow you to assess if there is a marked difference in the response of Eastern states as compared to Western states as highlighted by Figure 2.

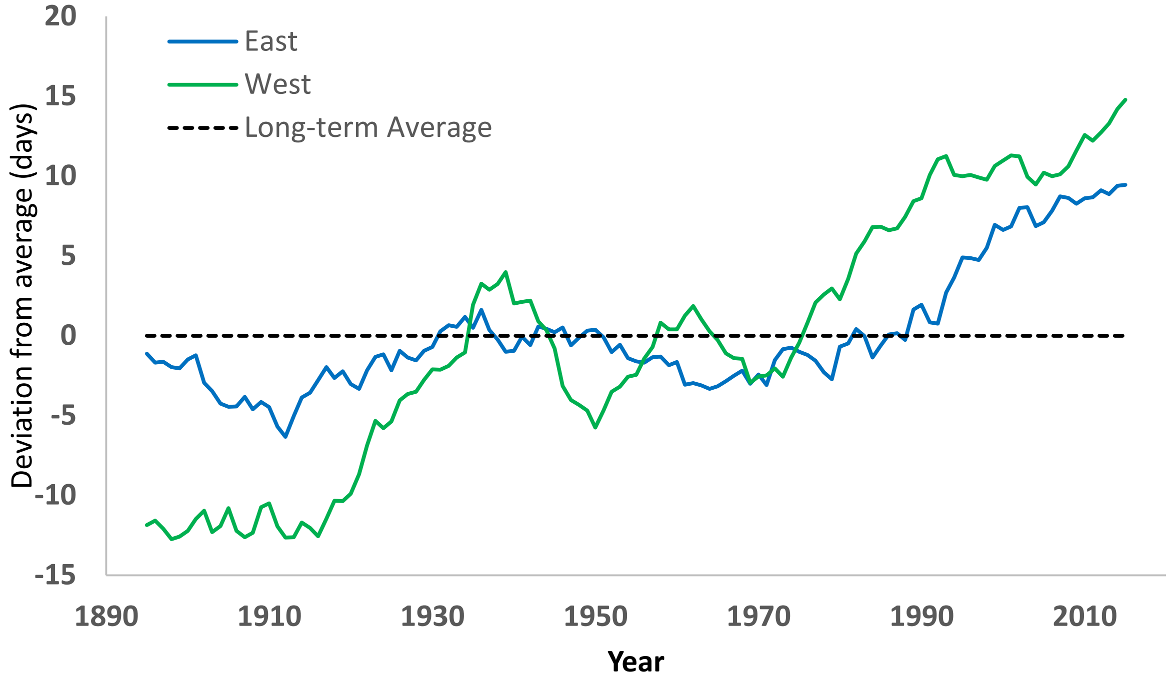

Figure 2: The length of the growing season in the contiguous 48 states compared with a long-term average for Eastern and Western states in the USA. For each year, the line represents the number of days shorter or longer than average. Choosing a different long-term average for comparison would not change the shape of the data over time.

Task

Use the Mann-Whitney *U* test to compare the data in the East and West data tab for the period 1995 to 2015. State your null hypothesis (H_o) and alternative Hypothesis (H_1).

Manually calculate the U-test using the following steps:

(i) Convert the two columns of length of growing season data into two new columns of ranks based on the COMBINED SAMPLE:
(ii) Sum the ranks in each column
(iii) Calculate the sample size (m) for the LARGER of the two sums of the ranks (R_m)
(iv) Calculate the sample size (n) for the SMALLER of the two sums of the ranks (R_n)
(v) Using R_m, m and n and the equation in box 1, calculate the U statistic
(vi) Refer to a table of critical values for U for your sample size. Decided whether to reject or accept your null hypothesis

** TIP** you would reject your null hypothesis if *U* is less than or equal to the critical value.

$$U = mn + \frac{m(m+1)}{2} - \sum R_m$$

Box 1: The Mann-Whitney *U* test. Where *m* and *n* are the sample sizes of the two groups of data and R_m is the smaller of the two sums of the ranks

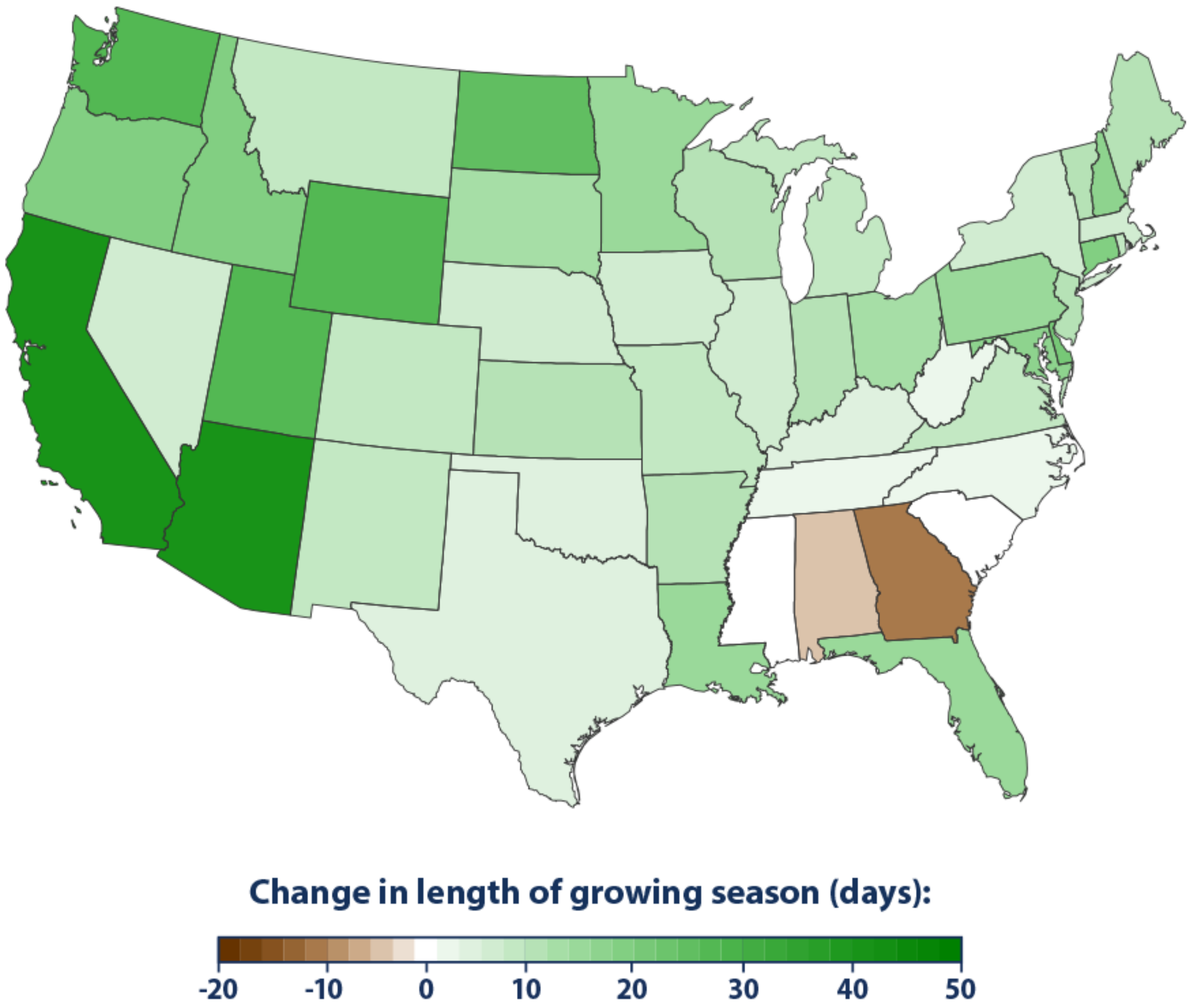

Figure 3: This map shows the total change in length of the growing season from 1895 to 2015 for each of the contiguous 48 states.

Looking at Figures 2 and 3 and the results you generated from the Mann-Whitney *U* test discuss the findings. You might want to think about:

- Is there a statistical difference between the means of length of growing season in the eastern and western states? How significant is the relationship?
- Looking at figure 3 explain the spatial pattern – think about how the topography of the Western states might influence the change in temperature and precipitation patterns which are partly driving the response of the growing season length.

Take it Further

Using excel repeat the process on the full data set from 1985 – 2015. You might want to think about:

- How the length of the data record affects the results. Is there still a statistical difference between the eastern and western growing seasons when the Mann-Whitney *U* test is calculated over the entire data period?

3) **Correlation between frosts and growing season lengths?**

In recent years, the final spring frost has been occurring earlier than at any point since 1895, and the first autumn frost has been arriving later. Since 1980, the last spring frost has occurred an average of three days earlier than the long-term average, and the first fall frost has occurred about three days later. This change to the timing of the frosts is believed to be linked to the change in the length of the growing season; this data will help you explore if there is a relationship between frost timing and growing season lengths.

Task
Using the data between 1980 and 2015 calculate the Spearman's Rank Correlation Coefficient to assess the degree of association between the frosts and growing season length. State your null hypothesis (H_o) and alternative Hypothesis (H_1).

i. Manually calculate the Spearman's Rank Correlation Coefficient using the following steps:
ii. Rank the growing season length data from highest (rank 1) to lowest (rank n).
iii. Rank the spring frost data from highest (rank 1) to lowest (rank n).
iv. Substitute the ranked data into the formula in Box 2 and calculate the Spearman's Rank Correlation Coefficient.
v. Repeat steps i - iv for the autumn frost data.

$$r_s = 1 - \left(\frac{6 \sum_{i=1}^{n} d^2}{\left(n^3 - n\right)} \right)$$

Box 2: The Spearman's Rank Correlation Coefficient. Where r_s is the Spearman's Rank Correlation Coefficient, *d* is the difference in the rank between the two data sets and *n* is the number of paired observations.

The Spearman's Rank Correlation Coefficient on its own is not a measure of significance and it is not possible to judge whether the value is close enough to 1 to be statistically significant. A *Student's T Test* can be performed to test the statistical significance of the data.

Task

Calculate the *Student's T Test* for the two Spearman's Rank Correlation Coefficients calculated previously using the formula in box 3. Compare the calculate value of t with the tabulated value in Student's t tables at the 95% confidence level and at n-2 degrees of freedom and then reject or accept the null hypothesis you made previously.

$$t = r_s x \sqrt{\frac{n-2}{1-r_s^{\,2}}}$$

Box 3: *The Student's T Test*. Where t is the Student's t statistic, r_s is the Spearman's Rank Correlation Coefficient and *n* is the sample size.

Looking at results you generated from the Spearman's Rank Correlation Coefficient and the Students T Test discuss the findings. As a point of note the Spearman's Rank Correlation Coefficient ranges between 0 and ± 1 with a zero indicating no correlation and 1 indicating a perfect correlation. You might want to think about:

- Is there a correlation between the length of the growth season and the date of the spring frost? How strong is this correlation? Is it significant?
- Is there a correlation between the length of the growth season and the date of the autumn frost? How strong is this correlation? Is it significant?
- Which has a stronger correlation? Why do you think this might be?

Plenary

Return to the main lesson question. Ask the students to discuss how:

- Do you think the trend in increasing growing season and the earlier date of the first frosts will continue into the future or do you think there will become a point where the change in these parameters will level off? Why?
- What potential other implications are there for the changes to the dates of the first and last frost?