

Name _____ Period _____ Date _____

Activity: Measuring the Continental Drift between Europe and North America

Scientists have been taking VLBI measurements for around 25 years to determine the distance between the Westford, Massachusetts radio telescope and the one in Wettzell, Germany. These measurements show that the distance between North America and Europe, about 6000 km, is drifting away at a steady rate of about *17 mm per year*. This means the data is accurate to 1 part in a billion! How do scientists obtain the continental drift from the data?

Understanding the data

The data can be accessed on the website,

<http://www.haystack.mit.edu/edu/pcr/downloads/>, or on a CD ROM from Haystack Observatory. Select "Geodesy.xls". You will need Excel to open the data. The data format is shown below:

MJD	Length	Uncert.	stn1	stn2	Epoch	Year	Month	Day
45655.31453	5998325.312	0.0168	WESTFORD	WETTZELL	1983.881	1983	11	17

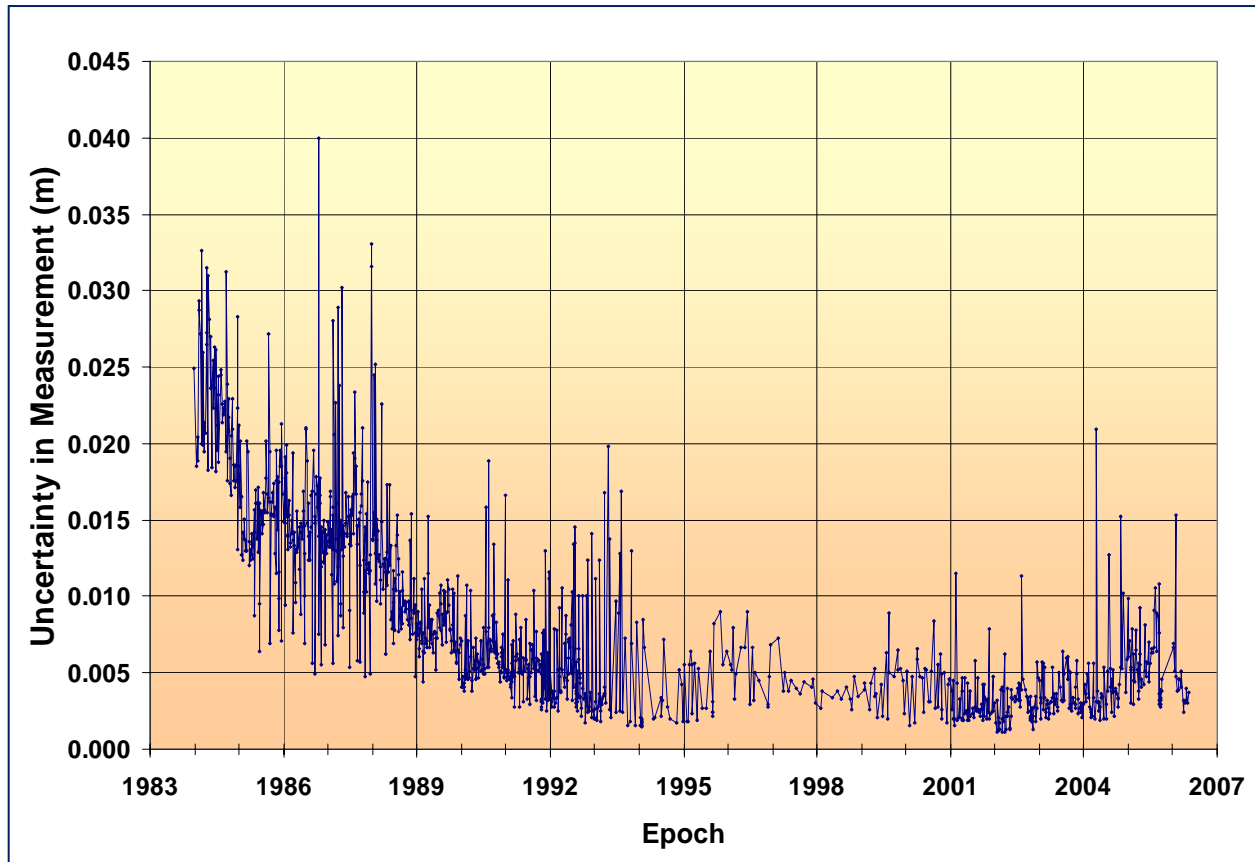
MJD stands for Modified Julian Day, a method of dating that has been used for centuries by astronomers and geophysicists in order to have an unambiguous dating system based on continuing day counts.

Analyzing the Uncertainties in the Data

Scientific measurements have uncertainties and error. They are always present but are reduced with better instrumentation and experimental procedures. In order to see this, examine the measurement uncertainties as a function of time.

- Using Excel, make a graph of Uncertainty (column 3) on the y-axis vs. Epoch (column 6) on the x-axis. Your graph should be similar to the one on the next page. More detailed instructions can be found in the [Appendix](#) under "[Plotting the Uncertainty Data](#)"

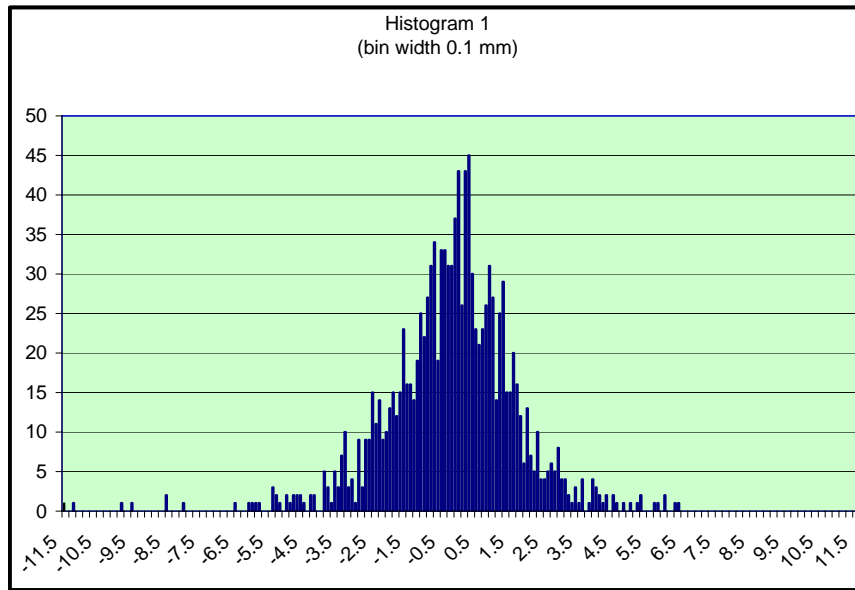
You can pick out some of the outliers because of their large deviations from average values that make up a year.



1. Answer the following questions concerning the above uncertainties plot.

a. Explain the trend in uncertainty from 1984 to 1994.

b. Describe the trend in uncertainty from 1994 to 2006.



2. From the shape of the histogram, can you conclude that is the distribution of the data is random? Explain why or why not.

Identifying the Outliers in the Measurements

If the data is randomly distributed, scientists make use of the normal distribution where outliers are “three sigma”, approximately three standard deviations from the mean value. Look at the Uncertainty vs. Epoch graph, and find the data points that have large deviations from the average uncertainties. For example, data point 404 has an uncertainty 0.1098.

MJD	Length	Uncert.	stn1	stn2	Epoch	Year	Month	Day
47494.22441	5998325.398	0.1098	WESTFORD	WETTZELL	1988.914	1988	11	29

The standard deviation of all of the uncertainty data around is about 0.0080. An outlier would be any data point with uncertainty larger than $3 \times 0.0080 = 0.024$. Since point 404 is much larger, it can be statistically eliminated.

Next step is to identify and eliminate all outliers. There are two methods:

- i) Eliminate the outliers from the Uncertainty graph by selecting those which are obviously bad data points. **Warning! Be careful because the data improves after 1984 and you would end up eliminating most of those years.**
- ii) A more reliable and detailed method of finding outliers can be found in the [Appendix](#) under "[Calculating Outliers](#)". Using this method you should find 35 outliers. Write the outlier data point and its uncertainty in the table below.

Show your instructor your table.

	Data Point	Uncertainty		Data Point	Uncertainty
1			21		
2			22		
3			23		
4			24		
5			25		
6			26		
7			27		
8			28		
9			29		
10			30		
11			31		
12			32		
13			33		
14			34		
15			35		
16			36		
17			37		
18			38		
19			39		
20			40		

3. Why is it important to identify and eliminate outliers in a data set?

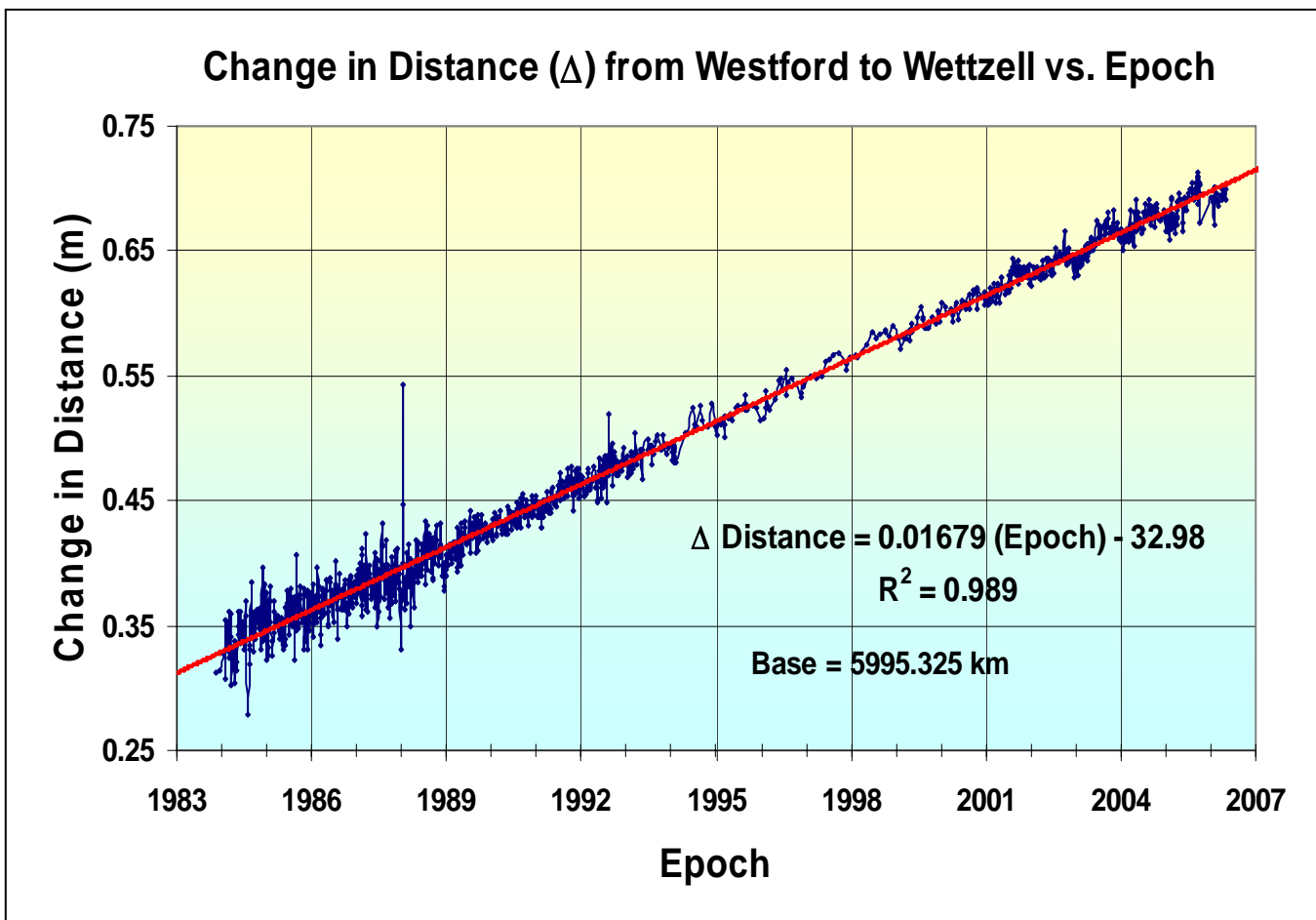
Finding the Continental Drift

Once the outliers are eliminated, the continental drift can be found using the following procedure;

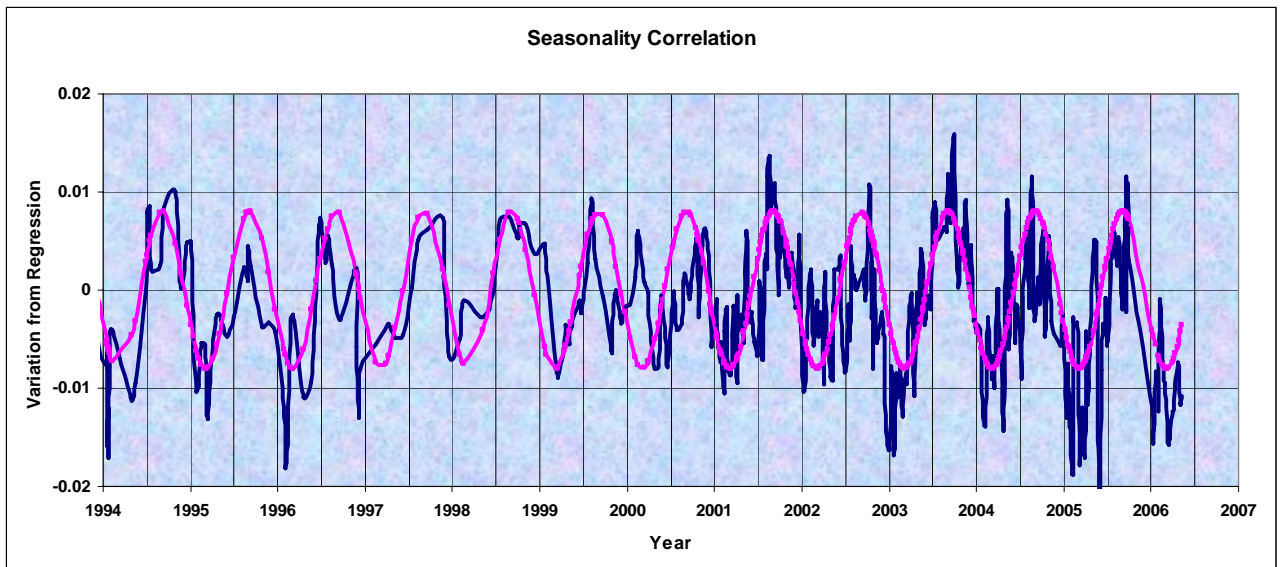
- Go to Excel.
- Click on the chart icon on top menu.
- Go to the "Standard Types" tab and choose "XY Scatter." Select the chart in the upper top left corner. Click "Next >"
- Go to the "Series" tab, not the "Data Range" tab.
- X values:
 - ❖ Click on the red icon to the right.
 - ❖ Type: "=Sheet1!\$F\$2:\$F\$1204".
 - ❖ Click on the red icon again.
- Y values:
 - ❖ Click on the red icon to the right.
 - ❖ Type: "=Sheet1!\$G\$2:\$G\$1204".
 - ❖ Click on the red icon again.
 - ❖ Click "Next >"
- Label the graph and the axes.
 - ❖ Chart title: Change in Distance from Westford to Wettzell vs. Epoch
 - ❖ Value (x) axis: Epoch
 - ❖ Value (y) axis: Change in distance (meters)
 - ❖ Click "Next >"
- Choose the location of the graph
 - ❖ "As object in: Sheet 1," and click "Finish."
- Right click on any data point.
 - ❖ Go to format data series.
 - ❖ Adjust the size of the marker to 2 points.
 - ❖ Click "OK."

- Double left click on the horizontal (x) axis and go to the scale tab.
 - ❖ Minimum: 1983
 - ❖ Maximum: 2007
 - ❖ Major unit: 3
 - ❖ Minor unit 1
- Add a trendline and equation by right clicking on a data point.
 - ❖ Select "Trendline."
 - ❖ Click on the linear regression box.
 - ❖ Then select the "Options" tab and select show equation on graph and show R2 value. Click "OK."
 - ❖ Drag the equation and R2 value to a convenient location.

You may want to do other options to make your graph look nicer. Your graph should look similar to this:



7. For the advanced student: A closer look at the measurements from 1994 to 2006 shows predictable long range trend.



- Examine the trend. Do you agree it shows a pattern? Explain your reasoning.
- Give a reasonable explanation for this trend.
- Why might this trend not be seen prior to 1994?
- How would you make the trend adjustment to the data? Don't do it, just show how it would be done.

Appendix

1. Plotting the Uncertainty Data

- Go to the website: <http://www.haystack.mit.edu/edu/pcr/downloads/>
- Download the geodesy data by selecting "Geodesy.xls."
- Click on the "Chart Wizard" on top menu.
- Go to the Standard types tab and select "XY (Scatter)". Click the graph at the second row left corner.
- Click "Next >".
- Go to the "Series" tab, not the "Data Range" tab.
- Click on "Add".
- X values:
 - Click on the red icon to the bottom right.
 - Type: "=Sheet1!\$F\$2:\$F\$1239". Click on the red icon again. Or drag over the F column to include all the data.
- Y values:
 - Click on the bottom red icon to the right.
 - Type: "=Sheet1!\$C\$2:\$C\$1239". Click on the red icon again. Or drag over the C column to include all the data.
- Click "Next >".
- Label the graph and the axes.
 - ❖ Chart title: Uncertainties
 - ❖ Value (x) axis: Epoch
 - ❖ Value (y) axis: Uncertainty in Measurement (m)
- Click "Next >".
- Choose the location of the graph: As object in: "Sheet 1"
- Click "Finish".
- Right click on any data point.
- Go to format data series.
- Adjust the size of the marker to 4 points.
- Click "OK."

Spend some time making the graph more functional. For example,

- Adjust the years from 1983 to 2007 with major grid of 3.
- Set the y-axis format to a minimum of 0.
- Add gridlines.
- Add color for the background.
-

2. Making a Histogram

- To make a histogram, you will first have to get the changes in distance between the two telescopes down to the mm level.
- Go to cell J1, type in "Change in distance."
- In J2, type " $=B2- \$B\2 ".
- Press "Enter."
- Highlight cell J2 then drag the column down to J1239.
- Press "Ctrl-D" (fill down). This will subtract B2 from all the data points in the column.
- Go to cell K1, type in "Random" for random error.
- In K2, type " $=1000*(J2-(0.01679*F2-33.2931))$ ". This will convert the differences to mm.
- Highlight cell K2 then drag the column down to J1239.
- Press "Ctrl-D" (fill down).
- Go to cell M1, type "Bin number"
- In M2, type "-40", in M3 "-39". Highlight cells M2 and M3 and release the mouse. Now place the cursor down to the lower right hand corner until a "+" appears. Drag down to until a 40 occurs (you will see the numbers on the right).
- On the top menu select "Tools" → "Data Analysis" → "Histogram."
- For "Input Range," enter " $\$K\$2:\$K\1239 ."
- For "Bin Range," enter " $\$M\$2:\$M\82 ."
- Click on "Output Range" and enter " $\$O\2 "
- Check off "Chart Output" and click "Ok."

3. Calculating Outliers

- Pick to the first empty cell (C 1240) after the last data value in column C.
- In it, type: " $=stdev(C2:C1239)$ ".
- Press "Enter" to calculate the standard deviation.
- Record your value _____.
- In order to reject bad data points, you will have to find the values of 3 times the standard deviation for the uncertainties. Pick the next cell below.
- Type: " $=3*(stdev(C2:C1239))$ " OR type " $=3*\langle \text{click on C1240} \rangle$ "
- Press "Enter" to calculate this value.
- Record your value which is 3 x standard deviation _____.

Why do this? Because any value outside this range can be assumed to be an outlier.

4. Eliminating outliers

Verify that you have deleted all the data that has a value greater than 3 standard deviations by finding the maximum uncertainty value

- Pick the cell C1243 (column C, row 1243)
- Type: "=max(C2:C1239)" This will give you the maximum uncertainty value in the data set.
- Record the maximum uncertainty value _____.
- If the maximum uncertainty is **greater** than the value of 3 standard deviations previously found (0.024), then the measurement is considered an outlier and it must be eliminated.
 - ❖ Highlight all the uncertainties on column C.
 - ❖ Go to Edit on the menu at the top of the Excel Spreadsheet and click "Find."
 - ❖ Enter the maximum uncertainty value.
 - ❖ Locate this value on the spreadsheet.
 - ❖ Highlight the entire row containing this value.
 - ❖ Right click on the highlighted row and delete.
- **NOTE:** The maximum uncertainty value will continuously change as outliers are being eliminated.
- Repeat this same process over and over until the maximum value is less than the value for 3 standard deviations (0.024).