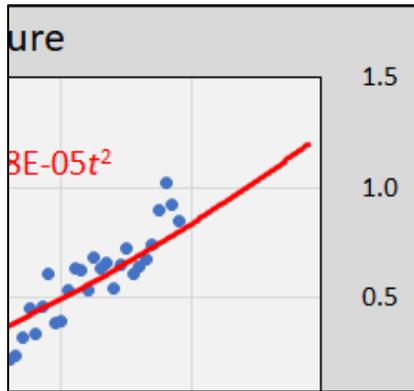# Analysis of 7 post-1900 global time series with Excel

Lee De Cola

LDECOLA@COMCAST.NET

Detail of a time series plot

Communication about accelerating changes in Earth's environment is the responsibility of everyone with a basic understanding of data. This hands-on workshop uses Microsoft Excel to visualize and analyze 7 yearly biogeophysical global indicators gathered from authoritative sources into a single database. You will: 1) be introduced to post-1900 global change measurements in their SI ('metric') units, 2) practice effective visualization of time series data, and 3) learn the fundamentals of forecasting based on regression and other models. The exercise focuses on WHAT IS MEASURED and not 'what causes what,' 'what is to be done,' or 'who is to blame?' The workshop can be done on your own, in a group, or with a demonstrator.

PREREQUISITES: A computer and familiarity with Excel.

1. The data

1.1. Visit the project website

Go to the Seven World Series (7WS) project page **ldecola.net/projects/global** and examine the table displaying metadata for each of the time series, e.g. for **surface temperature**:

1. **NAME**, a short nickname (e.g. **TEMP**),
2. **UNITS**, an abbreviated name for the measurement (**celsius**) and used in the database,
3. **SERIES**, a short scientific name for the series (surface temperature),
4. **MEASUREMENT**, the units spelled out in SI units notation (anomaly degrees Celsius),
5. **SOURCE** website of the time series (click on the link to visit the NASA Goddard Institute for Space Studies).

Keep the project webpage open for reference.

? This symbol invites you to address challenges on your own after the workshop.

1.2. Download the data

Below the metadata is a link to the data – **click** on it:
ldecola.net/projects/global/global.csv
to download the spreadsheet, then **click** on the file, which should open in Excel. (IOS computers open the data as a text file, which Excel should be able to read. Any other data app – I use R – can also read this dataset.)

Save the file as an Excel spreadsheet **global.xlxs** on the desktop (or wherever you want store it for future analysis).

1.3. Improve visibility of the worksheet

Make the Excel window large enough so that you can see its contents at all times. For greater visibility use the **Zoom slider** at lower right to magnify the window by **clicking** on the '+' sign; I like 150% so I click 5 times.

The worksheet is organized into rows and columns, here identified by numbers on the left and capital letters at the top, respectively. The first value in each row is **year** and the first value in each column is an abbreviation for the units of each time series – to remind us of exactly what each series measures!

Left-click on row header 1 to highlight the column headings then click **Home tab > Font: Bold** and **Alignment: Align Right**.

To keep that top header row always visible click on the **View tab > Window > Freeze Panes > Freeze Top Row**. Scroll down in the worksheet to see what happens.

In what follows, I may refer to a **series** by 1) the letter Y, 2) the **row** of a worksheet by its year t (e.g. $t = 1950$) and 3) an **individual value** as Y_t, so for **TEMP** ($Y_{1950} = -0.18$). This is standard mathematical notation.

1.4. Inspect the dataset

Spend a moment examining the dataset. How many columns are there and what's being measured (see metadata)? How many rows and what do they represent?

Scroll up and down, compare the first and last rows; are they very different? (You can freeze the top 2 rows and then compare 1901 to 2018, etc.)

? Scroll to the year of your birth, inspect the value of **POP**, and compare it to the latest values. (You could freeze your birthyear.)

? What are the actual units of **SOLAR**? of **year**? Later, learn what the numbers represent in the International System of Units (SI). How much does **SOLAR** vary relative to its mean?

An easy way to get insights to a series is to **Left-click** on its column (e.g. **celsius**) and inspect statistics on the status bar at the bottom. Look at the other columns as well. Confirm that **year** does indeed span 1901-2018.

? What can you tell about the change in temperature from the Min and Max values?

2. Plot surface temperature (**celsius** v. **year**)

2.1. Make a time series plot

The way to visualize a time series is with a plot in the Y versus t space. **Left-click** on column A (**year**), and **Ctrl-left-click** on column C (**celsius**) to highlight both columns. Then **Left-click on > Insert > Charts > Recommended Charts > Scatter** (the 2nd example) **> OK**.

You should get a scatter plot with **year** on the horizontal X-axis and (anomaly degrees) **celsius** on the vertical Y-axis. If necessary, make the plot big enough to easily read, then click & drag the plot so that its ratio of width to height is about 3:2.

Before examining the behavior of temperature over time, the axes should be fixed. To improve the value axis, **Left-click the left vertical axis > Right-click > Format Axis > Axis Options > Axis Options (the Bar Chart symbol) > Bounds > Minimum: –1** and **Maximum: 1.5.** To fix the horizontal axis select **Horizontal axis crosses > Axis value: −1.0 > Enter**.

Finally, shift the value axis to the right, near the present – the time that we're interested in! **Left-click the horizontal year axis > Format Axis > Axis Options > Vertical axis crosses: Maximum axis value**.

❓ To make the fonts of the axis labels larger, Right-click on the Vertical axis > Font > Size: 12 and Font color: black.

❓ To better format the values: **Format Axis > Axis Options > Number > Category > Custom > Format Code: Custom:**
  ' `0.0;-0.0`' (don't use the single quotes but DO use the leading space) **> Add**.

2.2. Examine the plot

Close your eyes, take 3 deep breaths, and meditate on your plot: the 'behavior' of the data over time and perhaps in the future. The time dimension shows the measurements spanning about 1900 to 2020, and the value dimension Y (as we confirmed earlier) ranges from a minimum of about – 0.5 to 1.0 °C.

❓ If you have access to a printer, Right-click the plot > File > Print Settings: Print Selected Chart then click Print. Mark it up with any comments you may have. Draw a line through the data and estimate its slope.

Global surface temperature is clearly increasing over the period, with an overall change from about $Y_t = -0.4$ at the beginning of the 20th century to about $Y_t = 0.8$ around 2010, for a slope of about 1.2°C per 100 years = .012°C per year (or 0.012 °C yr^{-1} as you may see the units shown in scientific publications).

The pattern moreover is generally accelerated (concave upward) but with some interesting details, e.g. in the 1940s around the time of World War 2.

❓ What is the *current* temperature change per decade in American units?

❓ Speculate about what temperature might be in 2040, how fast it's increasing, etc. Visit the GISS website to find out more about 'anomaly' means.

❓ Why did TEMP appear to drop in the 1940s?

2.3. Add annotation

If you wanted to use this plot in a paper or presentation you should add some more annotation, which the online metadata provides. To add axes labels **Left-click the plot > Chart Elements ('+' sign) > check Axis Titles**.

Use the online metadata table to do the following. Give the plot the `TEMP` SERIES description 'surface temperature,') give the Y-axis the label MEASUREMENT units ('anomaly degrees Celsius'), and label the X-axis with the SOURCE ('NASA Goddard Institute for Space Studies').

I suggest leaving the horizontal bounds for the time axis as is; in fact I like the way the points hang in time showing when the data series begins and ends.

❓ Do earlier observations exist; was temperature 'stable' before 1900; what will the future bring?

❓ You could use a lighter color to contrast the plot area from its frame, make the various text elements larger, etc. See the fragmentary example at the beginning of this handout. Make all the text black rather than dark gray.

❓ I think the markers are a bit big so that some are on top of others; Right-click any data point > Format Data Series > Fill & Line > Marker Options > Border > No line.

❓ Would another chart design (e.g. 'Line' or even 'Bar') be clearer – or more confusing? Experiment: but if you're going to be plotting more time series from the 7WS project they should have comparable designs.

3. Forecasting with trendlines

3.1. Fit a linear regression model

Forecasting is a branch of data science that fits statistical models to time series measurement to understand how a given phenomenon changes and what values it might take in the future: next year, decade, century. If a model fits past data well then it's likely to forecast well – time will tell! Excel provides a rich selection of models you can use interactively to fit mathematical equations to the data.

Right-click on any data point > Add Trendline and a Format Trendline menu opens to right with **Trendline Options > Linear** regression selected and faintly drawn as a dotted line through the datapoints.

Make the trendline more visible: **Right-click the line > Format Trendline > Fill & Line (paint bucket) > Line > Solid line > Color: Red; Width: 2; Dash type: Solid**. (You can choose whatever symbology you want, but I like things simple and bold.)

The line should approximate your 'eyeball' trend estimate from above. Notice how the line 'fits' the data by ignoring its transitory ups and downs to capture the trend. Make the analysis statistically explicitly **Left-click on the trendline > Trendline Options (the bars) > check** both **Display Equation** and **Display R-squared on chart**.

Left-click the linear regression statistics box and move it away from the trendline, then **Triple-click** the statistics text and select a bigger **font size**, e.g. 14 pt. and **color: red** (or whatever color you chose for the line itself).

For a report it is customary to arrange the regression coefficients with the Y-intercept first and also to round the statistics as follows:

$$Y_t = -18.1 + .0093t$$
$$R^2 = .828$$

The reported slope of the line .009 °C yr⁻¹ validates the eyeball estimate above. But see how the linear trend underestimates such behavior as growth post-1960 and especially since 2000. The R^2 also shows that (to use the language of colloquial statistics) 'about 83% of the variance in global temperature is explained' by the linear regression model, but we can do better and still satisfy Occam.

? Look at some statistics references to learn more about regression coefficients, R-squared, variance, etc. But the picture is clear; global surface temperature is increasing – and the rate of increase is itself increasing (i.e. accelerating)!

3.2. Fit a polynomial model

Right-click on the trendline and explore other models, which may give better predictions (and work for other series), but for now select **Polynomial > order: 2**. You may have to delete the linear statistics box and re-select the Equation and R-squared, to refresh the statistics:

$$Y_t = 280 - 0.296t + 8 \times 10^{-5}t^2$$
$$R^2 = .882$$

The R^2 statistic has improved and the concave-upward parabola gives a more satisfying visual fit, reflecting accelerated change over the period – temperature is increasing *and* the increments are increasing!

NOTE: The t^2 quadratic coefficient is very small because `year` is a large and arguably arbitrarily valued predictor.

> **?** Use **Trendline Options** to **Forecast** the model e.g. 10 or 20 years and consider whether that's likely or not.

3.3. Discussion of the model

The regression model provides a great deal of information. For example, examine the recent strong positive residuals (the difference between the points and the line); is that a new trend? And what happened during World War 2? Most importantly notice that over the 120 years temperature was generally accelerating, from close to 0 at the beginning of the 20th Century to .04°C yr^{-1} at the end of the series.

Another regression approach would be to analyze the temperature data say for the period 1960-2018 using a linear or polynomial degree-2 model. If you work on your own you'll see that this piecewise approach will be useful with the logarithm of the EMISSIONS series.

> **?** Try higher-degree polynomials; consider how much (or little) explanatory power is added to R^2 by another term.

Save your spreadsheet!

4. Next steps…

4.1. Forecast the other time series.

Choose another of the 7 series and reproduce the steps above, beginning with the plot instructions. Compare results for linear and quadratic models in a table showing such statistics as R^2, coefficients, forecasts, etc.

Try some of the other trendline models; for accelerating series, $\log(Y_t)$ might give meaningful results, but you may have to add a constant to get it to work.

Moving average is a useful way to remove 'noise' in time series and perhaps to detect cyclicality.

Make a pocket 'data card' showing the current values of the 7WS, their growth rates, and 2040 predicted forecasts. Memorize the current values – cite them at parties, of which you're guaranteed to be the life!

In you're working in a group use modulo arithmetic to assign researchers to series. Take the last 2 digits of your birth year (e.g. 1995) and divide by 7, then add 1 to the remainder to chose another of the 7WS time series. E.g. (95 mod 7) + 1 = 4 + 1 = 5 so you'd analyze series #5 (EMISSIONS). If you got 2 then pick something besides TEMP.

4.2. Advanced analysis

I generally use R for advanced time series analysis, but Excel has Data > Data Analysis > Exponential Smoothing that you might try.

Visit the sources in the metadata table for other analytical ideas (loess smoothing, autoregression, piecewise regression, etc.)

The 7 World Series are a useful start, but there are many other datasets that could be added: ice cover, methane emissions, clouds, global energy consumption, sea temperature – the list is endless, but most are showing accelerated change.

4.3. Keep in touch!

I welcome feedback and especially corrections and suggestions, – and would like to see this project assisted and even carried on by other researchers. Finally, I'm willing to give this workshop via Zoom to any interested group, and in person within biking, kayaking, or public transport distance of Reston, Virginia.