Discover secrets
your spreadsheet
can’t tell you

An Oil Industry Application

Correlating Well Logs
Using Time Series Modelling

Spreadsheets are essential business tools which you should never give up. Spreadsheets are excellent for tracking numbers, creating basic graphs, simple reports and calculating basic mathematics. They are indispensable for “what if ?” analysis and many financial tasks , such as budgeting.

When you need to go beyond simple summaries and basic row-and-column maths; when you want more insight into your data, you’re ready for a stats package. Together , a spreadsheet and an application like SPSS 7.0 for Windows or SYSTAT 5.0 for Windows give you the tools necessary to **make better, more informed decisions**.

“For more statistical tools than you’ll get from a spreadsheet such as Excel, with-out the learning curve, SPSS for Windows is a great buy.”

-Michael Burgard PC/Computing

Geologic measurements in sequences

Electric well log data are characterised by their position along a single line or drillhole. That is, they form a sequence, and the position at which a data point occurs within the sequence is important. Techniques for examining data having a positional characteristic , such as a resistivity trace, are considered part of the field of *time series analysis* which is a branch of statistics. Time and space in this situation can be considered interchangeably. A time series is simply a collection of observations made sequentially in time.

Geoscientists are concerned not only with analysis of data in sequences, but also with the comparison of two or more sequences. An obvious example is statigraphic correlation, either of measured sections or electric well logs. Although it falls within the framework of geophysics, most interpretation of the logs is carried out by people who consider themselves as petroleum engineers or geologists rather than geophysicists.

Why do it?

A Geoscientist’s motive for numerical correlation may be a simple desire for speed, as in the production of geologic cross-sections from digitised logs stored in a database. Alternatively, the analyst may be faced with a correlation problem where the recognition of equivalency is beyond his or her ability. Subtle degrees of similarity, too slight for unaided detection, may provide the clues that will allow the geoscientist to **make a decision** where none is otherwise possible. Numerical methods allow the geoscientist to consider many variables simultaneously, a powerful extension of his/her pattern recognition facilities.

Finally, because of the absolute objectivity of the statistical software, mathematical correlation represents a challenge to the interpreter. If a geoscientist’s correlation disagrees with that established by SPSS (which is the software that we use in this white paper) the geoscientist must determine the reason for the discrepancy. Such forced scrutiny may reveal complexities or biases not apparent during the initial examination.

Cross-correlation functions

The cross correlation function computes correlations between two variables (or series as statisticians refer to them) in which the values of one series (one drill hole's well log) are displaced by any specified number of time periods. This displacement can be both forwards or backwards and hence allows an assessment of whether one series (well log) leads or lags another. **<u>By inspecting the CCF between two series you can often see the lag at which they are most highly correlated.</u>**

Stationarity

"The Data Editor looks like a spreadsheet but gives you flexibility to edit data values and labels cases"

-Manatee Root Directory

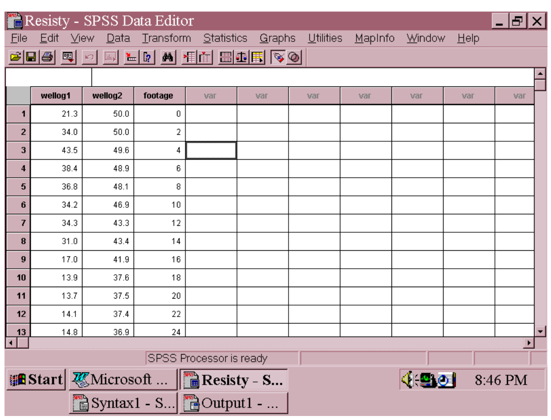
CCF plots can only be applied to series that are stationary. A series is said to be stationary if its mean and variance remain about the same over the length of the series.

Considerations to keep in mind

The example data used in this paper are from mid-North America, where Paleozoic sedimentation apparently was uniform over wide areas (Davis, 1973). The sequence of alternating limestones and shales may be difficult to correlate from well to well in areas of poor control, because the stratigraphic section appears to repeat again and again, and it is difficult to choose between the many possible correlations.

Cross-correlation may help to resolve such correlational problems, but should be **interpreted with caution**. In some instances , pretreating the data by filtering ,such as a moving average, may emphasize underlying characteristics and aid in cross-correlation. Remember, however that filtering reduces the independence of the data points and lowers the degrees of freedom by an undetermined amount. Interpret significance tests from smoothed data with great caution.

The Resistivity Data From The Two Wells



The first column *wellog1* contains the digitised electric resistivity log of an oil well through a Carboniferous section in the American state of Iowa. The second column *wellog2* contains the digitised electric resistivity log of an oil well through a Carboniferous stratigraphic section in NE Kansas. The third column records distances down the well at 2 foot spacing.

Product Tutorial

Below we request individual down hole plots for both wells , using the SPSS graphic user interface (GUI) after having already launched SPSS 7.0 for Windows from the Windows 95 desktop. We will “point and click” our way through the analysis steps.

1. Opening the SPSS data file *resisty.sav*
In order to analyse data, you must have data to run. You can use the main menu File command to open an existing SPSS file.

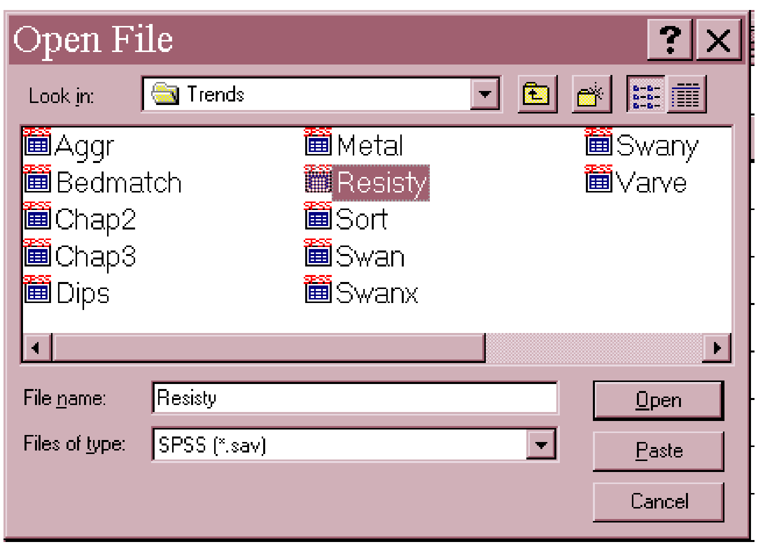
Click **File** in the main menu
Click **Open** in the File menu

SPSS displays the Open File dialog box which lets you specify the file name, directory (folder), and type of file you want to open. By default the dialog box displays SPSS files with the extension .SAV from the folder SPSS in the files list box entitled ‘Look In :’. You can change the specifications to read different types of files from different folders.

“Easily the slickest Windows-based statistics package... new users will love how smoothly the Windows dressing works.”

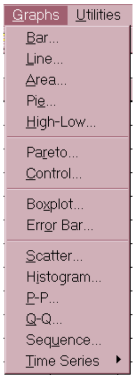
- InfoWorld

Double click *resisty.sav* in the ‘Look in:’ list box or
Click once on *resisty.sav* in the ‘Look in:’ list box and
Click the Open command button once to open the file.



SPSS puts the contents of the file in the Data Editor window, which was shown at the base of the previous page. You can make changes to the individual cell entries in the file by highlighting a cell and changing a data value (just like Excel). You can also add new data to the file in the Data Editor window. You use the mouse or the arrow keys to move around the file to examine the different variables or the contents of particular cells.

2. Making a down hole resistivity plot (a time series plot)



The Graphs menu contains a list of general statistical graphs. We will choose **Sequence** in order to plot a trace of the resistivity profile down the well.

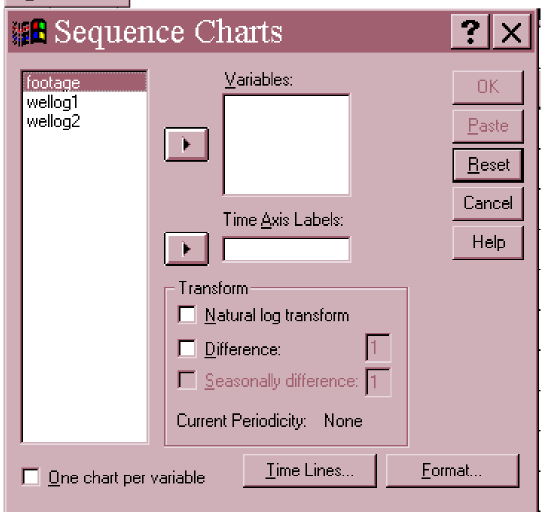
From the menu choose:

Graphs
Sequence...

This opens the Sequence Charts dialog box , as shown in the screen capture below.

"SPSS' graphing capability is impressive."

- InfoWorld



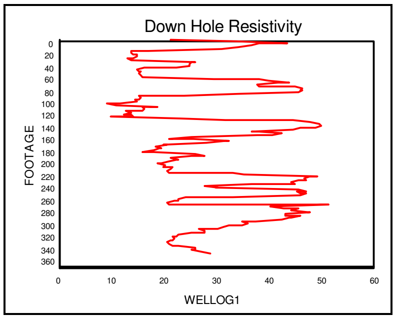
The numeric variables in the file are displayed in the source variables list on the left hand side of the dialog. Sequence charts require a data file structure in which cases (rows) are sorted in a sequential order.

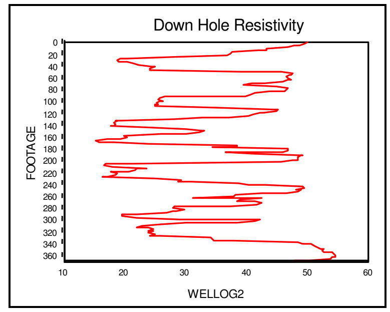
By default, the time axis (which we are using for distance down the hole) is simply labeled 1 to *n*, or it is labeled with the values of the variable *footage* by clicking this field name and then the button pointing to the text box labeled 'Time Axis Labels:'

There are check boxes for optionally log transforming the variable or differencing it. Differencing is used to impart stationarity in data whose values "drift" off scale. A stationary series is a necessary precursor to doing CCF plots, but we do not have to worry about this with the resistivity data for the two wells because they are stationary in their raw state (i.e. both mean and variance are roughly constant).

Since we wish to make one plot for each well we will click the check box labeled One chart per variable before clicking on **OK** to make the graphs.

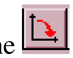
3. The Graphs





"SPSS provides a wealth of charting methods and styles."

- S.M.U.G. News

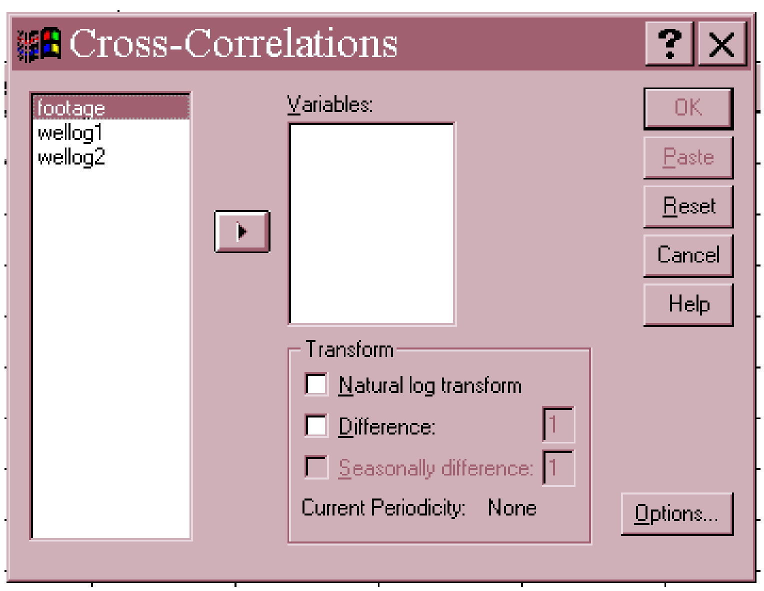
By default the sequence charts have distance along the horizontal axis, like a soil traverse. This was changed to a more logical vertical "down the hole" orientation by clicking on the  "90 degree" tool bar "icon" after double-clicking on the graph. The axis labels were centred and a title added as well.

4. Making the CCF Plot

The Cross-Correlations procedure calculates cross-correlation coefficients. It is simple to use. From the menus choose:

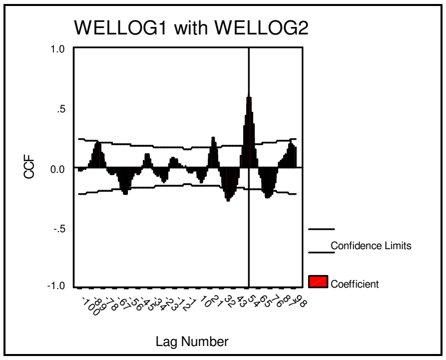
Graphs
  Time Series
    Cross-Correlations...

This opens the Cross-Correlations dialog box, as shown below.



Select both wellog1 and wellog2 and move them into the Variables list and click on **OK**. The resulting plot is shown on the next page.

Cross-Correlation Plot Of Well Logs 1 and 2



There is a large positive correlation of 0.603 at lag 57. The numerical value of the coefficients are displayed, along with a low-resolution CCF plot in the SPSS output window. The positive lag indicates that the first series (wellog1) *leads* the second series (wellog2). Thus we have to "slide" wellog1 "past" wellog2 by 57 intervals to get a "perfect match" and to correlate the two logs.

> **"SPSS for Windows distinguishes itself with smooth operation, great data manipulation and superb documenta-tion."**
>
> ***- InfoWorld***

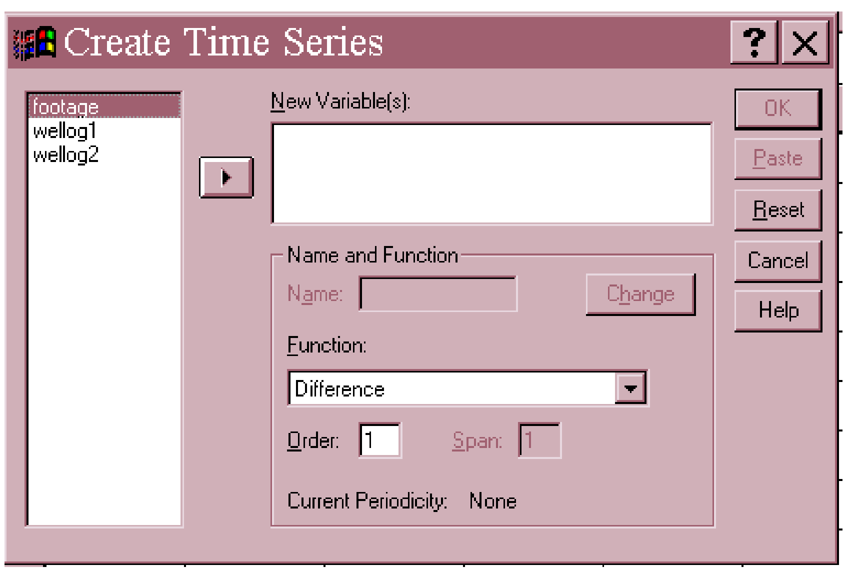
Statisticians would say that that wellog1 is a *leading indicator* of wellog2 57 intervals later.

5. Sliding one log past the other

We do this by adding a new column to our SPSS spreadsheet by lagging the log for well 1 by 57 intervals so that each value of *wellog1* is associated with the value of *wellog2* from 57 intervals before it. **SPSS performs this operation easily**. From the menus choose:

Transform
  Create Time Series...

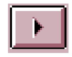
The Create Time Series dialog box is shown below.



How to fill in the Create Time Series dialog box

"SPSS for Windows provides enhanced statistical procedures while enabling non-special-ists to han-dle statistics on a PC."

- PC Magazine

Select *wellog1* and press  . SPSS then generates the following "equation" which appears in the New Variable(s) list:

wellog1_1 = DIFF (wellog1,1)

If you clicked on **OK**, this expression would create a new variable named *wellog1_1*, containing the differences for series *wellog1* and we don't want that. SPSS chooses differencing by default, since this is one of the most common time series transformations. To use other transformations, you use the controls in the Name and Function group:

• Highlight the contents of the Name text box (*wellog1_1*) and type a name that you want to replace it, such as *lead57in*, since the new series is going to be a leading indicator with a lag of 57 intervals.

🡻 Choose the Lag function from the Function drop-down list box. **Since wellog1 leads wellog2, a lagged copy of wellog1 will be correlated with that series.**

• The Order text box shows a value of 1. Highlight this and type 57 to lag the value of wellog1 by 57 intervals.

• Click on the 

command button. The New Variables list should now contain:

lead57in = LAG (wellog1,57)

• Click on **OK** to create the new time series.

If you go to the SPSS spreadsheet (the Data Editor) , you will see a new column containing the new variable *lead57in*. The first 57 observations will have a period, representing a missing value, since the file lacks information about the *wellog1* prior to observation 1. Other observations will equal the value of *wellog1* 57 intervals higher.

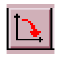
We are now ready to do **the statistical equivalent of putting the two logs on a light table** by plotting a sequence chart with *wellog2* and *lead57in* on the one chart.

To do this we now return to the Sequence Chart dialog, which you will recall we reached last time by choosing from the menu bar:

Graphs
  Sequence...

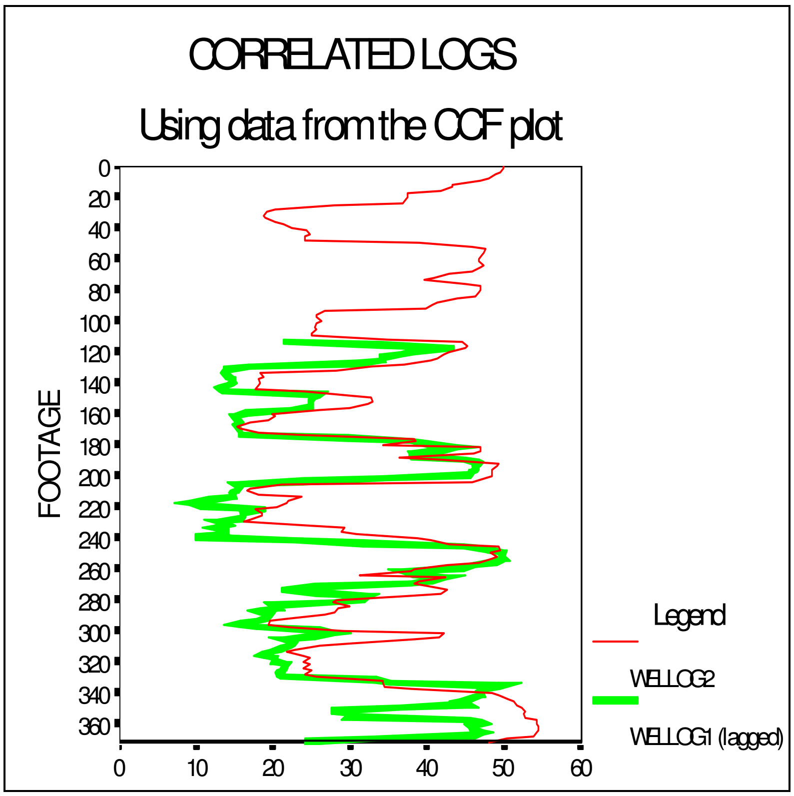
The numeric variables in the file are still displayed in the source variables list on the left hand side of the dialog, but this time there is the newly made variable called *lead57in* there as well.

Move the variables *lead57in* and *wellog2* to the Variables list by clicking on the variable *lead57in* and holding down the Ctrl key and clicking on *wellog2*. Now click the  button and **both** variables move across into the Variables list. Footage again should move into the 'Time Series Labels' text box. This time we leave the check box empty; the one which is labeled 'One chart per variable'. **This ensures that both logs plot on the one graph**.

We again use the  button to rotate the plot so that footage is the vertical axis. The final plot is shown below with the log for wellog1 in **bolder** heavier weight to make it stand out from the log for wellog2.

"...you don't need to be a computer whiz to use SPSS."

- Windows Magazine



Summary

Statistical techniques

Statistical software such as SPSS is the perfect complement to your spreadsheet. Spreadsheets are great for everyday tasks, such as tracking budget numbers and creating simple summary reports and graphs.

help you better use resources and give credibility to your ideas

However, there are times when you need more information from your data and you need to perform in-depth analysis. At these times, you need SPSS. Since SPSS was designed for in-depth analysis, **you get better information from your data.**

SPSS is the right choice to take your analysis to the next level. SPSS connects to your data regardless of where or how it is stored. SPSS can **uncover hidden patterns** and trends that rarely emerge using spreadsheet row-and-column maths. SPSS gives you great looking graphs and reports so you can easily communicate the results of your analysis. Together, SPSS and your spreadsheet can take your business data and translate in into meaningful information so you can make fully informed decisions - and make that **next oil discovery !**

Acknowledgements

The example data and case study discussed in this paper come from the first and second editions (1973,1986) of *'Statistics and Data Analysis In Geology'* written by Dr. John Davis of the Kansas Geological Survey in the USA.

Both John Davis and Garry Edser are members of the *International Association for Mathematical Geology*, which is dedicated to development and application of quantitative methods in the Earth Sciences. The IAMG is affiliated with the International Union of Geological Sciences (IUGS) and the International Statistical Institute (ISI). At the end of 1995 the IAMG had 555 members worldwide , of whom 30 come from Australia. Garry Edser has been a member since 1989. Garry has also been a member of the Australasian Institute of Mining and Metallurgy since 1978.