Can a Moving Average of the Data Be Deceptive?

Taking a closer look at this approach to data sets.

By Dean V. Neubauer

Q **Sometimes I want to take a moving average of my data to reduce the noise I am seeing, but can I actually fool myself?**

A A moving average, or rolling average, of a set of data taken in time sequence is formed by taking the average of the first *k* observations, and then subsequently dropping the first (oldest) value and adding the most recent (newest) value to form the next average. It is probably the simplest and most common procedure for smoothing data to reduce its noise level.

Consider a set of observations in time order $x_1, x_2, ..., x_n$ that can be partitioned into consecutive moving averages of span size *k* in this manner.

$$x_1, x_2, \cdots, x_k, x_{k+1}, \cdots, x_n$$

breaks up into moving windows of *k* points

$$\{x_1, \cdots, x_k\}, \{x_2, \cdots, x_{k+1}\}, \cdots, \{x_{n-k+1}, \cdots, x_n\}$$

averaging each window of *k* points yields $n - k + 1$ moving averages

$$\{\bar{x}_1\}, \{\bar{x}_2\}, \cdots, \{\bar{x}_{n-k+1}\}$$

The purpose of the moving average is to simply smooth out periodic fluctuations in a time series data set. If the data contain a cycle of constant period and amplitude, the moving average can remove it. Let's start with a simple example. The following data represent the output of a process for Monday through Friday over a period of a month, where commas separate days and semicolons separate weeks:

Week	M	Tu	W	Th	F
1	3,	2,	1,	1,	2;
2	3,	2,	1,	1,	2;
3	3,	2,	1,	1,	2;
4	3,	2,	1,	1,	2;

Here, we find that use of $k = 5$ days (one week) for our moving average results in a first week average of 1.8. Dropping the first Monday value of 3 and adding the second Monday value of 3 produces 1.8 for the next moving average. Table 1 shows all the moving averages for this dataset. Since the cycle does not change from week to week, and the length of time used for the moving average span (5) is the same as the length of the cycle of 5 days (one week), the moving average remains a constant value. Hence, the cyclic effect has been removed.

So, how then can a moving average be deceptive at times? Obviously, the typical dataset would not be expected to exhibit the consistent nature seen in our simple data set. We will often see additional variation added to any cyclic effects. So, it is really only necessary to address the random component of this variation. In other words, calculating a moving average should give a result for which one part is a moving average of the random variation component of the overall variation.

This brings us to the problem with moving averages. Successive moving averages of *k* values will have $k - 1$ values in common. This means that they will be positively correlated and the degree of correlation will increase as *k* gets larger. In other words, the larger *k* becomes, consecutive moving averages will have more data in common and be more positively correlated, and the data will look smoother (less variable). However, this sequence of positively correlated values will form an oscillation pattern. This implies that a series of moving averages that removes cyclic variation will still show oscillation due to the presence of random variation. We would not want to be deceived by this oscillation pattern as representing a real phenomenon.

Let's look at an example to illustrate this scenario.

EXAMPLE

Suppose that we generate 200 random normal numbers with a mean of 0 and a standard deviation of 1. Statisticians refer to such random numbers as belonging to a standard normal distribution. We will form moving averages of $k = 5$, $k = 10$, and $k = 25$. Figure 1a shows a plot of the individual random values. Figures 1b and 1c show the moving averages of $k = 5$ and $k = 10$.

Note that even though the individual data represent random values with no apparent trends seen in Figure 1a, these two figures exhibit what appears to be some fluctuation, which can be deceiving if the user attempts to explain the pattern. Of course, people that are good at thinking up explanations have no trouble "interpreting" them. By the time we extend the moving average to $k = 25$ points (Figure 1d), the data has become so smoothed that there is little to see. As you would expect, the plot becomes flatter the larger k becomes as the positive correlation among moving averages takes over.

SUMMARY

It should be stressed that the data in this last example came from a stable "system" representing random variation with no outside influences. Not all datasets are like this in practice as they contain some amount of non-normal variation. Thus, the use of moving averages on any dataset requires that the user consider how much they need to "smooth" the data, i.e., what an appropriate value of k should be used. A good rule of thumb is that if you know that you want to smooth out the type of variation seen within k points, then use a moving average of k points. If you are dealing with a process, then consider the process dynamics. If the process doesn't change much within k points, i.e., it is essentially random variation, then use a moving average of k points.

Don't let the moving averages deceive you into seeing something that is not really there! For more information on basic control charting methods, including exponentially weighted moving average charts, see E2587, Practice for Use of Control Charts in Statistical Process Control and Manual 7, *Manual on Presentation of Data and Control Chart Analysis*, 8th edition.

Table 1 — Computation of Moving Averages for Daily Process Data

Week	Day	Value	Moving Average of $k = 5$
1	M	3	
1	Tu	2	
1	W	1	
1	Th	1	
1	F	2	Avg{3,2,1,1,2} = 1.8
2	M	3	Avg{2,1,1,2,3} = 1.8
2	Tu	2	Avg{1,1,2,3,2} = 1.8
2	W	1	Avg{1,2,3,2,1} = 1.8
2	Th	1	Avg{2,3,2,1,1} = 1.8
2	F	2	Avg{3,2,1,1,2} = 1.8
3	M	3	Avg{2,1,1,2,3} = 1.8
3	Tu	2	Avg{1,1,2,3,2} = 1.8
3	W	1	Avg{1,2,3,2,1} = 1.8
3	Th	1	Avg{2,3,2,1,1} = 1.8
3	F	2	Avg{3,2,1,1,2} = 1.8
4	M	3	Avg{2,1,1,2,3} = 1.8
4	Tu	2	Avg{1,1,2,3,2} = 1.8
4	W	1	Avg{1,2,3,2,1} = 1.8
4	Th	1	Avg{2,3,2,1,1} = 1.8
4	F	2	Avg{3,2,1,1,2} = 1.8

Figure 1a — Individual Random Data

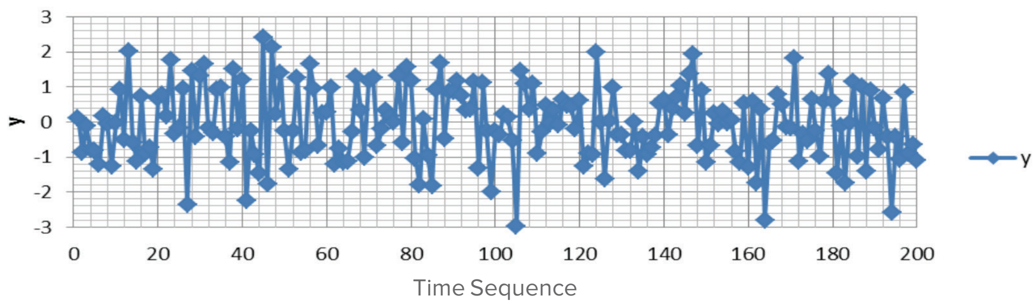


Figure 1b — Moving Averages of $k = 5$

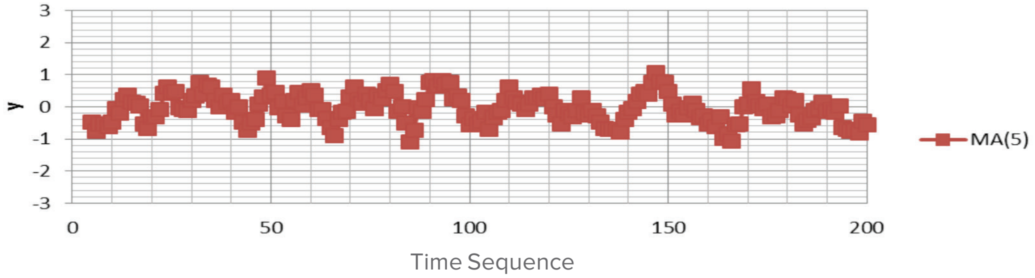


Figure 1c — Moving Averages of $k = 10$

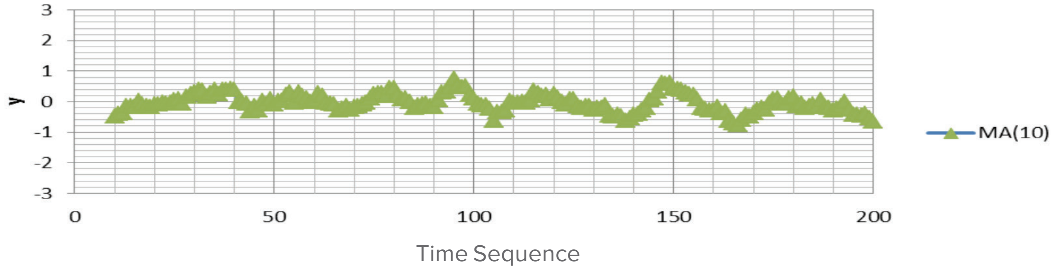
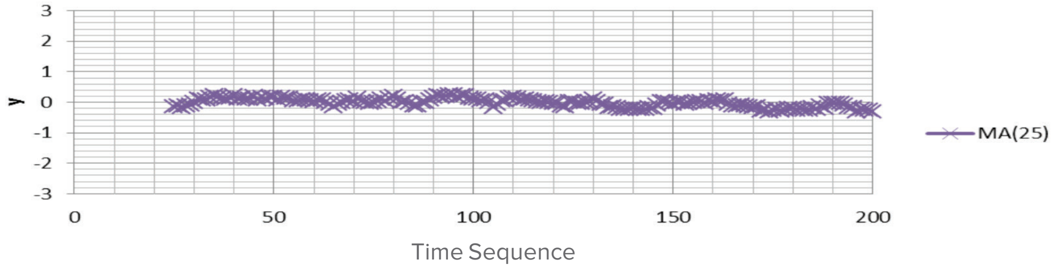


Figure 1d — Moving Averages of $k = 25$



Dean V. Neubauer, Corning Inc., Corning, New York, is an ASTM International fellow, chairman of E11.90.03 on Publications and coordinator of the Data Points column; he is immediate past chairman of Committee E11 on Quality and Statistics.

Get more tips for ASTM standards development
www.astm.org/sn-tips

Data Points archive
www.astm.org/standardization-news/datapoints