QUALITY ENGINEERING
2016, VOL. 28, NO. 1, 143–145
http://dx.doi.org/10.1080/08982112.2015.1100475

Discussion of "Statistical process monitoring of time-dependent data"

M. Schoonhoven and R. Goedhart

Institute for Business and Industrial Statistics, Department of Operations Management, University of Amsterdam, Amsterdam, The Netherlands

First of all, we would like to thank the authors for their work. The topic is highly relevant in practice, and the proposed solution yields great opportunities. In this discussion we provide some remarks and suggestions related to the proposal.

Statistical process monitoring is used to monitor a wide variety of processes. Common methods for monitoring processes typically assume that process data are not autocorrelated. However, in practice this assumption is often violated, in which case the traditional methods are often not appropriate. For this reason, new statistical methodologies are required that are applicable in this situation. The authors aim to develop a direction of monitoring such time-dependent data. As they indicate,

> In this contribution we will focus on the analysis of time-dependent processes since this is the scenario most often encountered in practice, due to high sampling systems and the natural behavior of many real-life applications. (p. 127)

Various examples of time-dependent data are given in the article. Two specific examples of it are given by

> Contemporary processes are typically highly automated, with in-line sensor technologies that produce vast amounts of data in a short period of time being the common situation. The result is the availability of large process streams that often display autocorrelation because of the fast sampling schemes relative to the process dynamics. (p. 127)

And more concrete (temperatures used to monitor the bearings of a rotating machine):

> Because of the variable load of the machine, fluctuating ambient temperature and wind speed, the temperatures vary widely as function of time. Given those fluctuations which are observable but uncontrollable and unpredictable, it is expected that temperature will rise when problems occur with the bearings. This is typically seen as short increases due to temporal blocking, and those are most often much smaller than the observed temperature variation under normal operation. (p. 138)

Although these both appear to be problems of time-dependent data, we feel that there is an essential difference between the two in the definition of what is out of control. In the first example, we are dealing with time-dependent *measurements*. In the given example, the displayed autocorrelation is present "because of the fast sampling schemes relative to the process dynamics." This autocorrelation is irrelevant to what is considered out of control. When there is a special cause, it is likely that multiple consecutive signals will be obtained due to fast sampling schemes. However, these signals indicate the same special cause. Similarly, in the absence of such a special cause, it is likely that consecutive measurements will not lead to a signal. This, however, does not change the meaning of "out of control." Thus, it does not seem necessary to develop new statistical process monitoring (SPM) methods for this situation.

In the second example we are dealing with a time-dependent *process*. Consider the mentioned example of the machine bearings' temperature. The major difference lies in the sense of what out of control means. Here, there are no real absolute boundaries available. The variation under normal operating conditions is larger than the change in temperature caused by temporal blocking. For this reason, the structure of autocorrelation (caused by the weather, location, etc.) is extremely relevant. For such problems, new SPM methods such as the one proposed are desired.

In our view, a clear distinction should be made between these two mentioned types of time-

CONTACT M. Schoonhoven m.schoonhoven@uva.nl Institute for Business and Industrial Statistics, Department of Operations Management, University of Amsterdam, Plantage Muidergracht 12, Amsterdam 1018 TV, The Netherlands.

dependency. This is because both types have a different interpretation of the meaning of out of control. The definition of when a process is actually out of control is perhaps the most important to consider in SPM.

Time-dependent measurements

In the case that we are dealing with time-dependent measurements but the definition of out of control is not time-dependent, the interpretation of out of control is rather straightforward. The process mean and standard deviation can be monitored to evaluate the process.

For the time-dependency of measurements, the authors mention two different possibilities. The first possibility is when the sampling schemes are too fast relative to the process dynamics. What this means is that the data contain measurements that are taken quickly after each other, so that a signal is likely to be followed by another signal (indicating that the same special cause is still present). On the other hand, when no special cause is present, it is likely that this will be the case as well for the next measurement, because the time between two measurements is small.

The other possibility of time-dependent data in this situation is when the measurements are correlated due to sensor aging. Sensors can deteriorate over time, potentially leading to a bias or inefficiency in the measurements. Although this does not give information about the underlying process, unreliable measurements of the process are a problem as well.

When a signal occurs, this can have two possible causes: either because the monitored statistic is actually out of bound or because the measurement system has become too inaccurate. Although the process might actually still be in control in the latter case, unreliable measurements are still a problem for the process monitoring and therefore should be investigated as well. Thus, in our view, the cause of a signal does not change the consequences of it.

In order to assess the performance of various monitoring methods for autocorrelated data, the authors consider the false detection rate (FDR). As they mention, the autocorrelation in the data brings a lot of variation to the FDR.

> Although the observed false detection rate of the Hotelling's T^2 statistics is generally within expectation, we see that the dispersion of the FDR values increases as the autocorrelation increases. This is a direct result of the inherent dynamics on the Hotelling's T^2 statstics, since it increases the probability of having consecutive measurements with similar values.

Indeed, a higher autocorrelation increases the probability of having consecutive measurements with similar values. However, the question is whether this is actually a problem at all. In the case that the sampling scheme is too fast relative to the process dynamics, it seems obvious that a signal has a high probability of being followed by another signal. However, we are investigating the same special cause. Because the sampling scheme is relatively fast, the signals will also occur quickly after another, until the underlying problem (e.g., sensor or process) is solved. The variability in the FDR does not seem very relevant. Either something is going on (whether sensor or process), in which case multiple consecutive signals are likely, or nothing is going on, leading to a high probability of no signals.

Time-dependent underlying process

In the case that we are dealing with a time-dependent underlying process, the situation becomes a lot more complicated. In this case, out of control no longer means a deviation of the mean or standard deviation but of a nonstationary model as a whole. A good example is indicated earlier, with the machine temperatures. There, the actual temperature is not relevant because it depends on multiple variables, but instead a deviation from the "normal" model is of interest.

As is shown in the article, traditional methods often have problems dealing with this type of nonstationary data. The authors therefore propose a method based on cointegration. Although the proposed approach seems suitable, it requires more detailed information on the monitoring procedure. It is not imminently clear when a process is considered to be out of control by this method, and the performance of such a model has not been assessed yet. This performance might also vary depending on the degree of nonstationarity. In addition, this method would only be applicable when there is evidence of cointegration. An alternative method is required when this is not the case.

Although it is very interesting (and, in the end, the goal) to apply the method to practical situations such as the machine temperatures, its performance should be assessed based on simulations. After defining the monitoring procedure, simulations similar to those in Section 2 should be performed to determine the FDR for this procedure.

For the performance it is also interesting to evaluate the detection capabilities, because this is in the end the final goal of SPM. As mentioned in the example, the change in temperature due to temporal blocking is often much smaller than the normal variation. It is thus very relevant to what degree this method is capable of detecting this small change. The performance should be compared against existing methods for different circumstances, to address potential benefits and downsides of the proposed method.

Concluding remarks

Although we do recognize the potential of the proposed method, we want to emphasize that the use of it is not recommended for any type of autocorrelated data. Whether or not to apply this method is heavily dependent on the definition of out of control, which varies depending on the situation. If the autocorrelation of the data is caused, for example, by fast sampling schemes relative to the process dynamics, we recommend against using this method, because the autocorrelation is irrelevant for what is out of control. The same holds for changes in sensor accuracy, because this is a problem in its own as well and should be considered as a special cause.

If the underlying process is such that the out-of-control situation heavily depends on the autocorrelated structure of the data, such as is the case for the machine bearings' temperature, the proposed method seems to be very promising. It would be of great interest to evaluate the method further and assess it based on its performance in different environments. With a clear structural design, the proposed method could become a very valuable addition to the existing SPM methodology.

About the authors

Marit Schoonhoven obtained her PhD in statistics in 2011 at the University of Amsterdam, The Netherlands. She is associate professor at the Department of Operations Management and senior consultant at the Institute for Business and Industrial Statistics (IBIS UvA) of the University of Amsterdam, The Netherlands. Her current research interests include control charting techniques and operations management methods.

R. Goedhart is a consultant of statistics at IBIS UvA and a PhD student at the Department of Operations Management at the University of Amsterdam, The Netherlands. He is working on his PhD on estimation effects and corrections in Statistical Process Monitoring.

References

De Ketelaere, B., T. Rato, E. Schmitt, and M. Hubert. 2016. Statistical process monitoring of time-dependent data. *Quality Engineering* 28(1):127–142.