Rejoinder

Peihua Qiu
Department of Biostatistics, University of Florida
2004 Mowry Road, Gainesville, FL 32610

I was invited to give a plenary speech at the 2017 Stu Hunter Research Conference in March 2017 at Copenhagen, Denmark. The program organizers suggested that I could talk about either statistical image processing or statistical process control (SPC), because image data had become a major data source in engineering applications and SPC had made several significant progresses in the past 10-15 years. I am fortunate to have made some contributions to both image processing (Qiu 2005) and SPC (Qiu 2014) in the past 25 years, and online monitoring of image data streams has become a major tool for quality control and management in manufacturing and other industries. Therefore, I decided to present some of my personal perspectives on the research about SPC of image data, a topic combining both image processing and SPC. Since jump regression analysis (JRA) is a basic tool for image processing and it could be a useful tool for solving some recent SPC problems as well, it was also covered in the presentation. After the presentation was summarized in the main paper titled "Jump regression, image processing, and quality control" (Qiu 2018), Drs. Alistair Forbes, Wei Jiang and Wenpo Huang provided some constructive comments about the paper. I am grateful for their time and effort in reading my paper carefully and in stimulating live discussions at the conference and here. Next, I will provide my thoughts about some of their comments given in their discussions.

Applications in Metrology and Business

In the main paper Qiu (2018), I mentioned that sequential online monitoring of image data streams had broad applications, and the applications I discussed focused mainly on manufacturing industries, satellite monitoring of earth surface, and magnetic resonance imaging (MRI) that I had some research experience. All discussants agree that image monitoring is a fundamental problem in many research and application areas. Dr. Forbes works at the UK's National Metrology Institute. He told us in his discussion that images were commonly used in metrological research and applications. One interesting problem he introduced was related to geometric evaluation of an artefact in question. In order to improve accuracy, we often use different imaging approaches (e.g., X-ray, sensor array, rotary table) to acquire different 2-D images, and then reconstruct the 3-D artefact geometry from the observed 2-D images. In computer science, image reconstruction is an active research area (e.g., Herman 1980, Schultz et al. 2014). Usually, a sequence of 2-D images are taken from different angles of the 3-D object, and then a 3-D image is reconstructed from the 2-D images using the Fourier and inverse Fourier transformations. To improve the quality of the reconstructed 3-D image, we need to combine information from the 2-D images taken by different imaging techniques. Because the images taken by different imaging approaches may not match with each other well geometrically, they need to be registered properly in advance using image registration methods. See Figure 7 in Qiu (2018) and the related discussion. Also, as pointed out by Dr. Forbes in his discussion, 2-D images taken by different imaging techniques have different data contamination and spatial correlation properties. They need to be preprocessed before data combination or comparison. See the next part for a related discussion about proper handling of correlated image data. Dr. Forbes pointed out that "uncertainty evaluation or quality assessment" of image data was also required in these applications. In statistical terminologies, I believe this is related to noise or other data contamination level, which should be properly estimated for each image type in order to make the data combination or comparison reliable.

Drs. Jiang and Huang discussed some interesting applications of process online monitoring in business, related to customers' business behaviors and activities. In those applications, customers' behaviors and activities change over time. The observed data are often temporally correlated and contain outliers (i.e., isolated and unusually large or small observations) or change points (i.e., systematic changes in customers' activities). Because customers' behaviors and activities usually last for a long time or even for ever, multiple outliers and change points are possible. All these features pose a great challenge for us to handle the related online process monitoring problems in a reasonable way, because traditional SPC charts usually detect the first change point only in online process monitoring. Related to the dynamic nature of customers' business behaviors and activities, longitudinal data analysis tools in statistics should be useful (e.g., Xiang et al. 2013), where the time-varying behaviors can be modeled by parametric or nonparametric longitudinal models. In the literature of process monitoring or shift detection, people usually use parametric time series models for describing temporal data correlation. In reality, such models can hardly describe the true data correlation well, because the temporal data correlation is often related to the complicated impact of many confounding risk factors (e.g., the culture, weather, and economy) on people's business behaviors and activities. Therefore, the more flexible temporal data correlation structure allowed by some recent longitudinal data analysis methods should describe the true correlation structure better. To sequentially monitor processes with time-varying in-control (IC) process distributions, we have developed a new methodology called dynamic screening system (DySS) in recent several years (Qiu and Xiang 2014, 2015, Li and Qiu 2016, 2017, Qiu et al. 2017). Some versions of the DySS method can monitor dynamic processes with temporally correlated observations, and the temporal correlation structure is allowed to be nonparametric. However, the current DySS method can only detect the first change point in online process monitoring, and its performance could be impacted negatively by outliers. Therefore, it requires much future research to generalize the DySS method for detecting multiple change points and modify the current DySS method so that the modified methods are robust to outliers. To this end, the JRA methods should be useful because they can handle multiple jump points in the process mean function and its derivatives, and some JRA methods are robust to outliers (cf., Qiu 2005, Chapter 3).

Spatial Correlation Modeling

Dr. Forbes pointed out that in some applications the change in the correlation among different data streams might be our interest. He mentioned the application to monitor the structure health of a bridge under forced degradation. In this application, different sensors are usually attached to the bridge to measure different aspects of the structure health of the bridge, including tilt, strain and acceleration. In this example, the correlation among readings from different sensors might be more relevant, compared to observation streams from individual sensors, because data from individual sensors are usually quite stable over time. Generally speaking, the multivariate DySS methods mentioned in the previous part should be able to detect a correlation change in this problem. However, those multivariate DySS methods were designed mainly for detecting changes in the mean function. They may not be as sensitive to a correlation change as we want them to be. In the SPC literature, we could not find much discussion about monitoring of correlation among multiple components of a multivariate process, which requires some future research effort.

Drs. Jiang and Huang pointed out that proper modeling of spatial correlation was important in monitoring spatial data. They brought to our attention two important applications: one for spatio-temporal disease surveillance and the other for wafer surface monitoring. They agreed with me that the commonly used principal component analysis (PCA) tool would generally be inappropriate to use in such applications, because the PCA approach does not take into account the spatial information in the observed data. As discussed in the main paper Qiu (2018), the Markov random field (MRF) framework is popular in the image processing literature to accommodate spatial correlation in the observed image intensities. However, the assumed Markovian properties of this approach are hard to verify in practice, and the computation involved is generally intensive. In recent years, statisticians have made a great effort to develop methods for modeling correlated spatial data. Methods based on regression modeling include the one using temporal basis functions (cf., Lindström et al. 2015), which can be implemented in the *R* package *SpatioTemporal*, the function estimation methods using B-splines (e.g., Choi et al. 2013), and the ANOVA-type method (Heuvelink and Griffith 2010). Other alternative approaches include the ones in the Gaussian process framework (e.g., Diggle et al. 2013), and the Bayesian dynamic modeling framework (e.g., Finley et al. 2015). Some of these methods impose restrictive assumptions on the observed data and on their spatial correlation structure. Users should check the validity of these assumptions properly before actually using them.

Concluding Remarks

As mentioned in the main paper, I believe that SPC will find more and more applications in this big data era, because many big data involve data streams and SPC is a power tool for sequentially monitoring data streams. However, different new applications raise many new challenges that traditional SPC methods cannot handle properly, which include time-varying (or dynamic) IC process distributions, detection of multiple change points, complicated spatial or spatio-temporal data correlation, integration of data from different sources, and so forth. To address these challenges and the new features of the related process monitoring problems, we should develop new SPC methodologies or modify the traditional SPC methods so that the new features can be accommodated adequately. To this end, some methods or ideas in JRA should be helpful. I am glad that Drs. Forbes, Jiang and Huang agreed with me on these perspectives. I really appreciate their stimulating critics and comments. Hope my main paper and the follow-up discussions can promote future research in sequential monitoring of image data and in big data analysis as well.

Acknowledgments: This research is supported in part by an NSF grant.

References

Choi, I.K., Li, B., and Wang, X. (2013), "Nonparametric estimation of spatial and space-time covariance function," *Journal of Agricultural, Biological, and Environmental Statistics*, **18**, 611–630.

Diggle, P.J., Moraga, P., Rowlingson, B., and Taylor, B.M. (2013), "Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm," *Statistical Science*, **28**, 542–563.

Finley, A.O., Banerjee, S., and Gamerman, D. (2015), "spBayes for large univariate and multivariate point-referenced spatio-temporal data models," *Journal of Statistical Software*, **63(13)**, 1–28.

Herman, G.T. (1980), *Image Reconstruction from Projections: The fundamentals of Computerized Tomography*, New York: Academic Press.

Heuvelink, G.B.M., and Griffith, D.A. (2010), "Space time geostatistics for geography: a case study of radiation monitoring across parts of Germany," *Geographical Analysis*, **42**, 161–179.

Li, J., and Qiu, P. (2016), "Nonparametric dynamic screening system for monitoring correlated longitudinal data," *IIE Transactions*, **48**, 772–786.

Li, J., and Qiu, P. (2017), "Construction of an efficient multivariate dynamic screening system," *Quality and Reliability Engineering International*, in press.

Lindström, J., Szpiro, A., Sampson, P.D., Bergen, S., and Sheppard, L. (2015), "SpatioTemporal: an R package for spatio-temporal modelling of air-pollution," https://cran.r-project.org/web/packages/SpatioTemporal/index.html.

Qiu, P. (2005), *Image Processing and Jump Regression Analysis*, John Wiley & Sons: New York.

Qiu, P. (2014), *Introduction to Statistical Process Control,* Boca Raton, FL: Chapman Hall/CRC.

Qiu, P. (2018), "Jump regression, image processing and quality control (with discussions)," *Quality Engineering*, in press.

Qiu, P., and Xiang, D. (2014), "Univariate dynamic screening system: an approach for identifying individuals with irregular longitudinal behavior," *Technometrics*, **56**, 248–260.

Qiu, P., and Xiang, D. (2015), "Surveillance of cardiovascular diseases using a multivariate dynamic screening system," *Statistics in Medicine*, **34**, 2204–2221.

Qiu, P., Zi, X., and Zou, C. (2017), "Nonparametric Dynamic Curve Monitoring," *Technometrics*, in press.

Xiang, D., Qiu, P., and Pu, X. (2013), "Local polynomial regression analysis of multivariate longitudinal data," *Statistica Sinica*, **23**, 769–789.

Schultz, G., Gallichan, D., Reisert, M., Hennig, J., and Zaitsev, M. (2014), "MR image reconstruction from generalized projections," *Magnetic Resonance in Medicine*, **72**, 546–557.