# Technometrics, ISSN 0040-1706
# Volume 64, number 4 (2022)

**Statistical Modeling and Monitoring of Geometrical Deviations in Complex Shapes With Application to Additive Manufacturing** P. 437-456

Riccardo Scimone, Tommaso Taormina, Bianca Maria Colosimo, Marco Grasso, Alessandra Menafoglio & Piercesare Secchi


**Abstract**

The industrial development of new production processes like additive manufacturing (AM) is making available novel types of complex shapes that go beyond traditionally manufactured geometries and 2.5D free-form surfaces. New challenges must be faced to characterize, model and monitor the natural variability of such complex shapes, since previously proposed methods based on parametric models are not applicable. The present study proposes a methodology that applies to complex shapes represented in the form of triangulated meshes, which is the current standard for AM data format. The method combines a novel bi-directional way to model the deviation between the reconstructed geometry (e.g., via X-ray computed tomography) and the nominal geometry (i.e., the originating 3D model) with a profile monitoring approach for the detection of out-of-control shapes. A paradigmatic example consisting of an egg-shaped trabecular shell representative of real parts produced via AM is used to illustrate the methodology and to test its effectiveness in detecting real geometrical distortions.


**Analyzing Nonparametric Part-to-Part Variation in Surface Point Cloud Data** P. 457-474

Anh Tuan Bui & Daniel W. Apley


**Abstract**

Surface point cloud data from three-dimensional optical scanners provide rich information about the surface geometry of scanned parts and potential variation in the surfaces from part-to-part. It is challenging, however, to make full use of these data for statistical process control purposes to identify sources of variation that manifest in a more complex nonparametric manner than variation in some prespecified set of geometric features of each part. We develop a framework for identifying nonparametric variation patterns that uses dissimilarity representation of the data and dissimilarity-based manifold learning, which helps discover a low-dimensional implicit manifold parameterization of the variation. Visualizing how the parts change as the manifold parameters are varied helps build an understanding of the physical characteristic of the variation. We also discuss using the nominal surface of parts when it is accessible to improve the computational expense and visualization aspects of the framework. Our approaches clearly reveal the nature of the variation patterns in a real cylindrical-part machining example and a simulated square head bolt example.


**Temporal Characterization and Filtering of Sensor Data to Support Anomaly Detection** P. 475-486

Michael J. Grosskopf, Kary Myers, Earl Lawrence & Derek Bingham


**Abstract**

We present an approach for characterizing complex temporal behavior in the sensor measurements of a system in order to support detection of anomalies in that system. We first characterize typical behavior by extending a hidden Markov model-based approach to time series alignment. We then use a trace of that learned behavior to develop a particle filter that enables efficient estimation of the filtering distribution on the state space. This produces filtered residuals that can then be used in an anomaly detection framework. Our motivating example is the daily behavior of a building's heating, ventilation, and air conditioning (HVAC) system, using sensor measurements that arrive every minute and induce a state space with 15,120 states. We provide an end-to-end demonstration of our approach showing improved performance of anomaly detection after application of alignment and filtering compared to the unaligned data. The proposed model is implemented as a computationally efficient R package alignts (align time series) built with R and Fortran 95 with OpenMP support.

**Transparent Sequential Learning for Statistical Process Control of Serially Correlated Data**

P. 487-501

Peihua Qiu & Xiulin Xie

**Abstract**

Machine learning methods have been widely used in different applications, including process control and monitoring. For handling statistical process control (SPC) problems, conventional supervised machine learning methods (e.g., artificial neural networks and support vector machines) would have some difficulties. For instance, a training dataset containing both in-control and out-of-control (OC) process observations is required by a supervised machine learning method, but it is rarely available in SPC applications. Furthermore, many machine learning methods work like black boxes. It is often difficult to interpret their learning mechanisms and the resulting decision rules in the context of an application. In the SPC literature, there have been some existing discussions on how to handle the lack of OC observations in the training data, using the one-class classification, artificial contrast, real-time contrast, and some other novel ideas. However, these approaches have their own limitations to handle SPC problems. In this article, we extend the self-starting process monitoring idea that has been employed widely in modern SPC research to a general learning framework for monitoring processes with serially correlated data. Under the new framework, process characteristics to learn are well specified in advance, and process learning is sequential in the sense that the learned process characteristics keep being updated during process monitoring. The learned process characteristics are then incorporated into a control chart for detecting process distributional shift based on all available data by the current observation time. Numerical studies show that process monitoring based on the new learning framework is more reliable and effective than some representative existing machine learning SPC approaches.

**Adaptive Partially Observed Sequential Change Detection and Isolation**

P. 502-512

Xinyu Zhao, Jiuyun Hu, Yajun Mei & Hao Yan1

**Abstract**

High-dimensional data has become popular due to the easy accessibility of sensors in modern industrial applications. However, one specific challenge is that it is often not easy to obtain complete measurements due to limited sensing powers and resource constraints. Furthermore, distinct failure patterns may exist in the systems, and it is necessary to identify the true failure pattern. This work focuses on the online adaptive monitoring of high-dimensional data in resource-constrained environments with multiple potential failure modes. To achieve this, we propose to apply the Shiryaev–Roberts procedure on the failure mode level and use the multi-arm bandit to balance the exploration and exploitation. We further discuss the theoretical property of the proposed algorithm to show that the proposed method can correctly isolate the failure mode. Finally, extensive simulations and two case studies demonstrate that the change point detection performance and the failure mode isolation accuracy can be greatly improved.

## TSEC: A Framework for Online Experimentation under Experimental Constraints

P. 513-523

Simon Mak, Yuanshuo Zhao, Lavonne Hoang & C. F. Jeff Wu

**Abstract**

Thompson sampling is a popular algorithm for tackling multi-armed bandit problems, and has been applied in a wide range of applications, from website design to portfolio optimization. In such applications, however, the number of choices (or arms) $N$ can be large, and the data needed to make adaptive decisions require expensive experimentation. One is then faced with the constraint of experimenting on only a small subset of $K \ll N$ arms within each time period, which poses a problem for traditional Thompson sampling. We propose a new Thompson Sampling under Experimental Constraints (TSEC) method, which addresses this so-called "arm budget constraint." TSEC makes use of a Bayesian interaction model with effect hierarchy priors, to model correlations between rewards on different arms. This fitted model is then integrated within Thompson sampling, to jointly identify a good subset of arms for experimentation and to allocate resources over these arms. We demonstrate the effectiveness of TSEC in two applications with arm budget constraints. The first is a simulated website optimization study, where TSEC shows considerable improvements over industry benchmarks. The second is a portfolio optimization application on industry-based exchange-traded funds, where TSEC provides more consistent and greater wealth accumulation over standard investment strategies.

## Novelty and Primacy: A Long-Term Estimator for Online Experiments

P. 524-534

Soheil Sadeghi, Somit Gupta, Stefan Gramatovici, Jiannan Lu, Hao Ai & Ruhan Zhang

**Abstract**

Online experiments are the gold standard for evaluating impact on user experience and accelerating innovation in software. However, since experiments are typically limited in duration, observed treatment effects are not always stable, sometimes revealing increasing or decreasing patterns over time. There are multiple causes for a treatment effect to change over time. In this article, we focus on a particular cause, user-learning, which is primarily associated with novelty or primacy. Novelty describes the desire to use new technology that tends to diminish over time. Primacy describes the growing engagement with technology as a result of adoption of the innovation. Estimating user-learning is critical because it holds experimentation responsible for trustworthiness, empowers organizations to make better decisions by providing a long-term view of expected impact, and prevents user dissatisfaction. In this article, we propose an observational approach, based on difference-in-differences technique to estimate user-learning at scale. We use this approach to test and estimate user-learning in many experiments at Microsoft. We compare our approach with the existing experimental method to show its benefits in terms of ease of use and higher statistical power, and to discuss its limitation in presence of other forms of treatment interaction with time.

## Sensitivity Prewarping for Local Surrogate Modeling

P. 535-547

Nathan Wycoff, Mickaël Binois & Robert B. Gramacy

**Abstract**

In the continual effort to improve product quality and decrease operations costs, computational modeling is increasingly being deployed to determine feasibility of product designs or configurations. Surrogate modeling of these computer experiments via local models, which induce sparsity by only considering short range interactions, can tackle huge analyses of complicated input–output relationships. However, narrowing focus to local scale means that global trends must be relearned over and over again. In this article, we propose a framework for incorporating information from a global sensitivity analysis into the surrogate model as an input rotation and rescaling preprocessing step. We discuss the relationship between several sensitivity analysis methods based on kernel regression before describing how they give rise to a transformation of the input variables. Specifically, we perform an input warping such that the "warped simulator" is equally sensitive to all input directions, freeing local models to focus on local dynamics. Numerical experiments on observational data and benchmark test functions, including a high-dimensional computer simulator from the automotive industry, provide empirical validation.

**Multi-Output Calibration of a Honeycomb Seal via On-site Surrogates** P. 548-563

Jiangeng Huang & Robert B. Gramacy

**Abstract**

We consider large-scale industrial computer model calibration, combining multi-output simulation with limited physical observation, involved in the development of a honeycomb seal. Toward that end, we adopt a localized sampling and emulation strategy called "on-site surrogates (OSSs)," designed to cope with the amalgamated challenges of high-dimensional inputs, large-scale simulation campaigns, and nonstationary response surfaces. In previous applications, OSSs were one-at-a-time affairs for multiple outputs leading to dissonance in calibration efforts for a common parameter set across outputs for the honeycomb. We demonstrate that a principal-components representation, adapted from ordinary Gaussian process surrogate modeling to the OSS setting, can resolve this tension. With a two-pronged—optimization and fully Bayesian—approach, we show how pooled information across outputs can reduce uncertainty and enhance efficiency in calibrated parameters and prediction for the honeycomb relative to the previous, "data-poor" univariate analog.

**Using BART to Perform Pareto Optimization and Quantify its Uncertainties** P. 564-574

Akira Horiguchi, Thomas J. Santner, Ying Sun & Matthew T. Pratola

**Abstract**

Techniques to reduce the energy burden of an industrial ecosystem often require solving a multiobjective optimization problem. However, collecting experimental data can often be either expensive or time-consuming. In such cases, statistical methods can be helpful. This article proposes Pareto Front (PF) and Pareto Set (PS) estimation methods using Bayesian Additive Regression Trees (BART), which is a nonparametric model whose assumptions are typically less restrictive than popular alternatives, such as Gaussian Processes (GPs). These less restrictive assumptions allow BART to handle scenarios (e.g., high-dimensional input spaces, nonsmooth responses, large datasets) that GPs find difficult. The performance of our BART-based method is compared to a GP-based method using analytic test functions, demonstrating convincing advantages. Finally, our BART-based methodology is applied to a motivating engineering problem. Supplementary materials, which include a theorem proof, algorithms, and R code, for this article are available online.