

ENBIS-21 PROGRAMME AND ABSTRACTS

Programme and abstracts of the ENBIS-21 Online Conference.

13-15 September 2021

©ENBIS (www.enbis.org)

All rights reserved

Edited by: Biagio Palumbo, Antonio Pievatolo, Jean-Michel Poggi, Johannes Rückert, Murat Testik

First edition

ISBN 979-12-20093-98-9

ENBIS-21 Online Conference

Monday, 13 September 2021 - Wednesday, 15 September 2021

Online Programme

Contents

A Digital Twin Approach for Statistical Process Monitoring of a High-Dimensional Micro-electronic Assembly Process	1
A Multivariate Non Parametric Monitoring Procedure Based on Convex Hulls	1
A Predictive Maintenance Model Proposal for a Manufacturing Company	2
A comparison of a new, open-source graphical user interface to R	2
A fixed-sequence approach for selecting best performing classifiers	3
A hybrid method for degradation assessment and fault detection in rolling element bearings	3
A novel fault detection and diagnosis approach based on orthogonal autoencoders	4
A novel online PCA algorithm for large variable space dimensions	4
A robust method for detecting sparse changes in high-dimensional (heteroskedastic) data	5
A tailored analysis of data from OMARS designs	5
Accreditation of statisticians	6
Active coffee break: A short poll on some fun and some constructive topics	7
AdaPipe: A Recommender System for Adaptive Computation Pipelines in Cyber-Manufacturing Computation Services	7
Adaptive Design and Inference for a Step-Stress Accelerated Life Test	8
Addressing statistics and data science educational challenges with simulation platforms	8
Adhesive bonding process optimization via Gaussian Process models	9
An Ode to Tolerance: beyond the significance test and p-values	10
An algorithm for robust designs against data loss	10
Analysis of resistance of spot welding process data in the automotive industry via functional clustering techniques	11
Analyzing categorical time series in the presence of missing observations	11

Application of domain-specific language models for quality and technical support in the Food and Beverage Industry	12
Application of machine learning models to discriminate tourist landscapes using eye-tracking data	12
Application of the Bayesian conformity assessment framework from JCGM 106 to lot inspection on the basis of single items	13
Attribute-Variable Alternating Inspection (AVAI): The use of $np_x - S^2$ mixed control chart in monitoring the process variance	13
Autocorrelated processes in metrology with examples from ISO and JCGM documents	14
Bayesian Designs for Progressively Type-I Censored Simple Step-Stress Accelerated Life Tests Under Cost Constraint and Order-Restriction	15
Bayesian I-optimal designs for choice experiments with mixtures	15
Bayesian Transfer Learning for modelling the hydrocracking process using kriging	16
CUSUM control charts for monitoring BINARCH(1) processes	16
Calibrating Prediction Intervals for Gaussian Processes using Cross-Validation method	17
Causal Rules Extraction in Time Series Data	17
Classification of On-Road Routes for the Reliability Assessment of Drive-Assist Systems in Heavy-Duty Trucks based on Electronic Map Data	18
Cleanliness an underestimated area when solving problems on Safety Critical Aerospace parts	18
Commented Summary of a Year of Work in Covid-19 Statistical Modeling	19
Constructing nonparametric control charts for correlated and independent data using resampling techniques	20
Copula-based robust optimal block designs	20
Customer prioritization for marketing actions	21
Data Mining for Discovering Defect Associations and Patterns to Improve Product Quality: A Case for Printed Circuit Board Assembly	21
Deciphering Random Forest models through conditional variable importance	22
Deep Multistage Multi-Task Learning for Quality Prediction and Diagnostics of Multistage Manufacturing Systems	22
Design Optimization for the Step-Stress Accelerated Degradation Test under Tweedie Exponential Dispersion Process	23
Design-Expert and Stat-Ease360: Easy and Efficient as Illustrated by Examples	23
Detecting changes in Multistream Sequences	24

Dubious new control chart designs – a disturbing trend	24
Entropy-based Discovery of Summary Causal Graphs in Time Series	25
Enumeration of large mixed four-and-two-level regular designs	25
Estimating the Time to Reach the Curing Temperature in Autoclave Curing Processes	26
Evaluating and Monitoring the Quality of Online Products and Services via User-Generated Reviews	27
Explainable AI and Predictive Maintenance	27
Explainable AI in preprocessing	27
Fault detection in continuous chemical processes using a PCA-based local approach	28
Forecasting count time series in retail	28
Generalized additive models for ensemble electricity demand forecasting	29
Greenfield Challenge 2021	30
Hands-on Projects for Teaching DoE	30
Harnessing the recondite role of randomization in today’s scientific, engineering, and in- dustrial world	31
Heteroscedastic Gaussian Process regression for assessing interpolation uncertainty of es- sential climate variables	31
Hypothesis-based acceptance sampling for modules F and F1 of the European Measuring Instruments Directive	32
Importance of Spatial Dependence in the Clustering of NDVI Functional Data Across the Ecuadorian Andes	33
Image-Based Feedback Control Using Tensor Analysis	33
Imbalanced multi-class classification in process industries. Case study: Emission levels of SO ₂ from an industrial boiler	34
In-Profile Monitoring for Multivariate Process Data in Advanced Manufacturing	34
Inference for the Progressively Type-I Censored K -Level Step-Stress Accelerated Life Tests Under Interval Monitoring with the Lifetimes from a Log-Location-Scale Family	35
Influence of process parameters on part dimensional tolerances: An Industrial Case Study	36
Infusing Statistical Engineering at NASA	36
Interactive tool for clustering and forecasting patterns of Taiwan COVID-19 spread	37
Lessons Learned from a Career of Design of Experiments Collaborations	37
Long short-term memory neural network for statistical process control of autocorrelated multiple stream process with an application to HVAC systems in passenger rail vehicles	37

MODELLING WIND TURBINE POWER PRODUCTION WITH FUZZY LINEAR REGRESSION METHODS	38
Machine Learning Approach to Predict Land Prices using Spatial Dependency Factors . . .	39
Modelling electric vehicle charging load with point processes and multivariate mixtures	39
Modern Methods of Quantifying Parameter Uncertainties via Bayesian Inference	40
Non-parametric multivariate control charts based on data depth notion	40
Online Hierarchical Forecasting for Power Consumption Data	41
Outlier detection in sensor networks	41
Outliers and the instrumental variables estimator in the linear regression model with endogeneity	42
PHEBUS, a Python package for the probabilistic seismic Hazard Estimation through Bayesian Update of Source models	43
PREDICTION OF PRECIPITATION THROUGH WEATHER VARIABLES BY FUNCTIONAL REGRESSION MODELS	43
Parameter Calibration in wake effect simulation model with Stochastic Gradient Descent and stratified sampling	44
Predicting migration patterns in Sweden using a gravity model and neural networks . . .	44
Prediction intervals for real estate price prediction	45
Predictive Maintenance in plasma etching processes: a statistical approach	45
Priors Comparison in Bayesian mediation framework with binary outcome	46
Railway track degradation prediction using Wiener process modelling	46
Randomizing versus not randomizing split-plot experiments	47
Real-time monitoring of functional data	48
Robust bootstrapped h and k Mandel's statistics for outlier detection in Interlaboratory Studies	48
ShapKit: a Python module dedicated to local explanation of machine learning models . . .	49
Six-Sigma and Obesity – Part 1	49
Six-Sigma and Obesity – Part 2	50
Sparse abnormality detection based on variable selection for spatially correlated multivariate process	50
Sparse and smooth cluster analysis of functional data	51
Spatial correction of low-cost sensors observations for fusion of air quality measurements	51
Spectral-CUSUM for Online Community Change Detection	52

Statistical analysis of simulation experiments: Challenges for industrial applications	52
Statistical models for measurement uncertainty evaluation in coordinate metrology	53
Strategies for Supersaturated Screening: Group Orthogonal and Constrained Var(s) Designs	54
Study of the effectiveness of Bayesian kriging for the decommissioning and dismantling of nuclear sites.	54
Tensor based Modelling of Human Motion	55
The Parameter Diagram as a DoE Planning Tool	55
The Shiryaev-Roberts Control Chart for Markovian Count Time Series	56
The numerical statistical fan and model selection	56
The study of variability in engineering design—An appreciation and a retrospective	57
Two questions of "class": Kind of quantity and Classification	57
Understanding and Addressing Complexity in Problem Solving	57
Univariate Self-Starting Shiryaev (U3S): A Bayesian Online Change Point Model for Short Runs	58
Variable importance analysis of railway vehicle responses	59
Vibration signal analysis to classify spur gearbox failure.	59
What's New In JMP 16	60
When, Why and How Shewhart Control Chart Constants need to be changed?	60

migraine patients, where the monitored qualitative time series on features such as pain peak severity or perceived stress are often incomplete.

The talk relies on the open-access publication

Weiß (2021) Analyzing categorical time series in the presence of missing observations.

Statistics in Medicine, in press.

<https://doi.org/10.1002/sim.9089>

Keywords:

incomplete data; nominal time series; ordinal time series

Quality 3 / 96

Application of domain-specific language models for quality and technical support in the Food and Beverage Industry

Authors: Peng Liu; Chiara Mondino; Noah Schellenberg; Alberto Barroso; Astrid Kyhl

Corresponding Authors: peng.liu@tetrapak.com, chiara.mondino@tetrapak.com

Issue Resolution is a critical process in the manufacturing sector to sustain productivity and quality, especially in the Food and Beverage Industry, where aseptic performance is critical. As a leader in this industry, Tetra Pak has built a database regarding quality events reported by Tetra Pak technicians, each containing domain knowledge from experts. In this paper, we present a model framework we have internally developed, which is using a domain-specific language model to address two primary natural language challenges impacting the resolution time:

1. Automatically classify a new reported event to the proper existing class
2. Suggest existing solutions when a new event is being reported, ranked by relevance of the descriptions of the issues (free text documented by the technician)

Our study shows that the language model could benefit from training on domain-specific data compared with those trained on open-domain data. For task 1, the language model is trained on the domain-specific data with an accuracy of over 85%. F1 score average is over 80%. For task 2, the domain-specific deep learning model is combined with a bag-of-words retrieval function-based algorithm to build an advanced search engine with an average precision of 53%.

Keywords:

Domain-Specific NLP, Text Classification, Prescriptive Analytics

Advances in Statistical Modeling and Applications (ISBIS) / 66

Application of machine learning models to discriminate tourist landscapes using eye-tracking data

Authors: Gianpaolo Zammarchi¹; Giulia Contu¹; Luca Frigau¹

¹ *University of Cagliari*

Corresponding Author: gpz.dottorato@gmail.com

Nowadays tourist websites make extensive use of images to promote their structure and the its location. Many images, such as landscapes, are used extensively on destination tourism websites to

draw tourists' interest and influence their choices. The use of eye-tracking technology has improved the level of knowledge of how different types of pictures are observed. An eye-tracker enables to accurately define the eye location and therefore to carry out precise measurement of the eye movements during the visualization of different stimuli (e.g. pictures, documents).

Eye-tracking data can be analyzed to convert the viewing behavior in terms of quantitative measurements and they might be collected for a variety of purposes in a variety of fields, such as grouping clients, improving the usability of a website, and in neuroscience studies. Our work aims to use eye-tracking data from a publicly available repository to get insight of user behavior regarding two main categories of images: natural landscapes and city landscapes. We choose to analyze these data using supervised and unsupervised methods. Finally, we evaluate the results in terms of which choice should be made between possible options to shed light on how decision-makers should take this information into account.

Keywords:

tourism, images, eye-tracking, machine learning

Statistical Standardization / 113

Application of the Bayesian conformity assessment framework from JCGM 106 to lot inspection on the basis of single items

Authors: Steffen Uhlig¹; Bertrand Colson¹

¹ *QuoData*

Corresponding Authors: steffen.uhlig@quodata.de, colson@quodata.de

The ISO 2859 and ISO 3951 series provide acceptance sampling procedures for lot inspection, allowing both sample size and acceptance rule to be determined, starting from a specific value either for the consumer or producer risk. However, insufficient resources often prohibit the implementation of "ISO sampling plans." In cases where the sample size is already known, determined as it is by external constraints, the focus shifts from determining sample size to determining consumer and producer risks. Moreover, if the sample size is very low (e.g. one single item), prior information should be included in the statistical analysis. For this reason, it makes sense to work within a Bayesian theoretical framework, such as that described in JCGM 106. Accordingly, the approach from JCGM 106 is adopted and broadened so as to allow application to lot inspection. The discussion is based on a "real-life" example of lot inspection on the basis of a single item. Starting from simple assumptions, expressions for both the prior and posterior distributions are worked out, and it is shown how the concepts from JCGM 106 can be reinterpreted in the context of lot inspection. Finally, specific and global consumer and producer risks are calculated, and differences regarding the interpretation of these concepts in JCGM 106 and in the ISO acceptance sampling standards are elucidated.

Keywords:

ISO 2859, ISO 3951, prior information

Quality 1 / 21

Attribute-Variable Alternating Inspection (AVAI): The use of $np_x - S^2$ mixed control chart in monitoring the process variance

Authors: Leandro Alves da Silva¹; Linda Lee Ho¹; Roberto Costa Quinino²

¹ *Universidade de São Paulo, São Paulo SP, Brazil.*