



## July 29-31, 2020 SCHEDULE

### Wednesday, July 29<sup>th</sup>

*Virtual - U.S. New York/Eastern Daylight Time*

8:00-8:50 Join the "Click Here to Test Audio/Video Connections" session by navigating to the "Agenda" tab in Whova. (*Note: we will not be able to assist with audio/visual issues once the meeting has begun*)

9:00-9:10 Opening – **David Banks**, Duke University and SAMSI

9:10-10:00 Chair: **David Banks**, Duke University and SAMSI  
**Cynthia Rudin**, Duke University  
*Seeing into Data and Models*

10:00-10:10 Break

10:10-11:25 Parallel Sessions

Statistical Learning - Org.: **Patrick Groenen**, Erasmus University

**Chun-houh Chen**, Academia Sinica

*Covariate-adjusted Heatmaps for Visualizing Biological Data via Correlation Decomposition*

**Patrick Groenen**, Erasmus University; **Michael Greenacre**, Universitat Pompeu Fabra  
*Interpretable Kernels for Explainable AI*

**Mikhail Zehlonkin**, Erasmus University

*Probabilistic Forecasting of Binary Outcomes in the Presence of Outliers*

Statistical Learning – Org.: **Jason Xu**, Duke University

**Jason Xu**, Duke University

*A Proximal Distance Algorithm for Likelihood-Based Sparse Covariance Estimation*

**Tianxi Li**, University of Virginia

*Linear Regression and its Inference on Noisy Network-linked Data*

**Aaron J. Molstad**, University of Florida

*Insights and Algorithms for the Multivariate Square-root Lasso*

Reproducible Computing and Reporting – Org.: **Jim Harner**, West Virginia University

**Dirk Eddelbuettel**, U of Illinois at Urbana-Champaign

*Reliable Reproducible Research via Containers from the Rocker Project*

**Brian Lee Yung Rowe**, Pez.AI

*Achieving Practical Reproducibility with Transparency and Accessibility*

**Jim Harner**, West Virginia University; **Chris Grant**, Rc2ai; **Mark Lilback**, Rc2ai

*Reproducible Computing and Reporting in a Complex Software Environment*



**DSSV 2020**  
DATA SCIENCE, STATISTICS & VISUALISATION

11:25-11:35 Break

11:35-12:50 Parallel Sessions

Visualisation - Org.: **Adalbert Wilhelm**, Jacobs University

**Adalbert Wilhelm**, Jacobs University

*Visual Story Telling of Covid-19: A Case Study*

**Xiaoyue “Zoe” Cheng**, University of Nebraska

*Visually Exploring Age-based Population Data over Time*

**Heike Hofmann**, Iowa State University

*Visualizing Elections in the U.S.*

**Susan VanderPlas**, University of Nebraska-Lincoln

*Perception and Visual Communication in a Global Pandemic*

Statistical Learning - Org.: **Peter Filzmoser**, TU Wien

**Sugnet Lubbe**, University of Stellenbosch

*Comparison of Zero Replacement Strategies for Compositional Data with Large Numbers of Zeros*

**Dorit Hammerling**, Colorado School of Mines

*Contained Chaos: Ensemble Consistency Testing for the Community Earth System Model*

**Matey Neykov**, Carnegie Mellon University

*High Temperature Structure Detection in Ferromagnets*

Data Science – Org.: **Ruda Zhang**, SAMSI

**Ruda Zhang**, SAMSI

*Normal-bundle Bootstrap*

**Deborshee Sen**, SAMSI

*Bayesian Neural Networks and Dimensionality Reduction*

**Jason Poulos**, SAMSI

*Retrospective Causal Prediction via Elapsed-Time and Propensity-Weighted Matrix Completion, with an Evaluation of the Effect of European Integration on Labour Market Outcomes*

12:50 Adjourn



## Thursday July 30<sup>th</sup>

Virtual - U.S. New York/Eastern Daylight Time

8:00-8:50 Join the "Click Here to Test Audio/Video Connections" session by navigating to the "Agenda" tab in Whova. (*Note: we will not be able to assist with audio/visual issues once the meeting has begun*)

9:00-10:15 Parallel Sessions

Statistical Learning - Org.: **Kohei Adachi**, Osaka University

**Kohei Adachi**, Osaka University, Japan

*Principal Component versus Factor Analyses with their Intermediate Procedure in Matrix Decomposition Formulation*

**Inge Koch**, University of Western Australia

*Principal Components for High-Dimensional and Directional Data*

**Giuseppe Vinci**, University of Notre Dame

*Graph Quilting: Graphical Model Selection from Partially Observed Covariances*

Data Science – Org.: **John Nardini**, SAMSI

**John Nardini**, SAMSI

*Learning Differential Equation Models for Noisy Biological Data*

**Glen Wright Colopy**, Cenduit

*Personalized Inference Protects Patients and Science*

**Xinyi Li**, SAMSI

*Sparse Learning and Structure Identification for Ultra-High-Dimensional Image-on-Scalar Regression*

10:15-10:25 Break

10:25-11:15 Chair: **Patrick Groenen**, Erasmus University

**David Dunson**, Duke University

*Generalized Bayes for Probabilistic Uncertainty Quantification in Unsupervised Learning*

11:15-11:25 Break

11:25-12:40 Parallel Sessions

Statistical Learning - Org.: **Richard Samworth**, University of Cambridge

**Hao Chen**, University of California, Davis

*Change-point Analysis for Modern Data*

**Yining Chen**, London School of Economics

*Jump or Kink: Super-efficiency in Segmented Linear Regression Break-point Estimation*

**Tengyao Wang**, University College London

*High-Dimensional, Multiscale Online Change-point Detection*



**DSSV 2020**  
DATA SCIENCE, STATISTICS & VISUALISATION

Data Science Technology - Org.: **James Harner**, West Virginia University

**Javier Luraschi**, RStudio

*Training ImageNet using TensorFlow and R*

**Soren Harner**, LayerJot & **Jim Harner**, West Virginia University

*Harnessing Big Data and Machine Learning with Arrow Data Frames in R and Python*

**Shih-Hsiung Chou**, Atrium Health & **Phil Turk**, Atrium Health

*CURVE: a Web Application for In-Hospital Resource Forecasting During the COVID-19 Outbreak*

New Ideas for Old Problems – Org.: **Deborshee Sen**, SAMSI

**Pulong Ma**, SAMSI

*Multifidelity Computer Model Emulation with High-Dimensional Output: An Application to Storm Surge*

**Kate Moore**, Wake Forest University

*Communities in Data*

**Wenjia Wang**, SAMSI

*Uncertainty Quantification for Bayesian Optimization*

12:40 Poster Session

1:10 Adjourn



## Friday July 31<sup>st</sup>

*Virtual - U.S. New York/Eastern Daylight Time*

- 8:00-8:50 Join the "Click Here to Test Audio/Video Connections" session by navigating to the "Agenda" tab in Whova. (*Note: we will not be able to assist with audio/visual issues once the meeting has begun*)
- 9:00-9:50 Chair: **Peter Filzmoser**, TU Wien  
**Robert Gramacy**, Virginia Polytechnic  
*Replication or Exploration? Sequential Design for Stochastic Simulation Experiments*
- 9:50-10:00 Break
- 10:00-11:15 Parallel Sessions
- JDSSV – Orgs.: **Patrick Groenen**, Erasmus University & **Stefan Van Aelst**, KU Leuven  
**Andreas Alfons**, Erasmus University  
*Cellwise and Rowwise Robust Regression with Compositional Covariates*
- Eun-Kyung Lee**, Ewha Womans University  
*Tree-structured Models using Projection Pursuit Method and their Explanation*
- Mu Zhu**, University of Waterloo  
*Some Statistical Applications of Generative Neural Networks*
- SAS – Orgs.: **Brett Wujek**, SAS Institute  
**Xan Gregg**, SAS Institute  
*Understanding Smoothers through Interactive Examples*
- Kelci Miclaus**, JMP Lifesciences  
*The Role of Visualization in Translational and Clinical Research*
- Guohui Wu**, SAS Institute  
*Location matters: Estimating Spatial Regression Models with Large Spatial Weights Matrices using SAS Econometrics*
- 11:15-11:25 Break
- 11:25-12:15 Chair: **David Banks**, Duke University and SAMSI  
**Ming Yuan**, Columbia University  
*Information Based Complexity of High Dimensional Sparse Functions*
- 12:15-12:25 Closing