Dr. David Rügamer

Research Statement

Combining Statistics and Machine Learning for Probabilistic, Scalable and Explainable Models

Modeling time-varying or complex systems in applied research fields such as biology, medicine, or psychology is one of the main challenges in the 21st century and requires joint efforts from experts of various fields and new innovative approaches. Some applications require fitting models for high-dimensional yet small sample size data, others the fusion of different sources of information. Methods need to account for spatio-temporal or multi-dimensional dependencies, need to work at large scale and still must be interpretable for researchers of the respective field. For safety-critical application, it is further required to measure uncertainties, both in the model and the data generating process.

Previous Research Experience

In the past, I have worked with researchers from various fields to analyze data and develop methods that bring new insights to the respective research domain. In the field of neuroscience I have worked with Prof. Scherer from the university of Geneva on time-varying brain signals to explain the emergence of human emotions [37]. The project not only led to various advancements in functional data boosting [31] but also inspired new methods to account for overoptimism in inference post model selection [15, 36, 38]. Through collaborations in radiology and medicine, I was inspired to develop methods that make generative models explainable to clinicians [29, 30] and allow researchers to combine methods from statistics and deep learning [8, 10, 13, 16, 22]. In the field of biomechanics, I closely work together with Prof. Fallah and Dr. Liew from Essex to better understand human motions and motion-induced pain [14, 20, 23, 24, 25, 32, 33, 34, 35] by also accounting for the dynamical nature of these signals. Last year, I started working with the department of geography at the LMU Munich to identify the atmospheric drivers of drought and heat in Western Europe [26]. Lastly, motivated by a collaboration with Prof. Stachl from St.Gallen, we are working on a scalable time-varying regression model to allow for the analysis of large-scale user behaviour data. In general, my methodological research is centered around models for time-dependent data, including time series (e.g., [7, 18]), functional data (e.g., [31, 37]) and time-to-event prediction (e.g., [13, 17, 22]).

Current and Future Research Directions

Researchers that bridge the gap between probabilistic statistical methods, machine learning (ML) and deep learning (DL) can make use of the best of both worlds while also profit from recent trends and the increased public interest in such methods. The ulterior goal of my research is an amalgamation of statistics, ML and DL that allows to

- make regression techniques more flexible,
- include unstructured data sources such as images into statistical models;
- make models more scalable;
- extend and better explain machine learning and deep neural networks;
- derive statistical uncertainty quantification methods for machine learning models.

A combination of these fields will provide flexible methods that are suitable for the steadily increasing amount of complex and dynamic data and offer larger applicability than existing methods. While this will allow to appropriately model complex data generating processes, there will also be the need for fundamental research to better understand the optimization, tuning, model selection, interpretation and uncertainty of these new approaches.

References

Preprints / Under Review

[1] F. Ott, D. Rügamer, L. Heublein, B. Bischl, and C. Mutschler. *Cross-Modal Common Representation Learning with Triplet Loss Functions*. Under Review. 2022. eprint: 2202.07901.

[2] F. Ott, D. Rügamer, L. Heublein, B. Bischl, and C. Mutschler. *Domain Adaptation for Time-Series Classification to Mitigate Covariate Shift*. Under Review. 2022. eprint: 2204.03342.

[3] F. Ott, D. Rügamer, L. Heublein, T. Hamann, J. Barth, B. Bischl, and C. Mutschler. *Benchmarking Online Sequence-to-Sequence and Character-based Handwriting Recognition from IMU-Enhanced Pens*. Under Review. 2022. eprint: 2202.07036.

[4] L. Bothmann, S. Strickroth, G. Casalicchio, D. Rügamer, M. Lindauer, F. Scheipl, and B. Bischl. *Developing Open Source Educational Resources for Machine Learning and Data Science*. Under Review. 2021. arXiv: 2104.02705.

[5] C. Fritz, G. D. Nicola, F. Günther, D. Rügamer, M. Rave, M. Schneble, A. Bender, M. Weigert, R. Brinks, A. Hoyer, U. Berger, H. Küchenhoff, and G. Kauermann. *Challenges in Interpreting Epidemiological Surveillance Data - Experiences from Germany*. Under Review. 2021.

[6] M. Rezaei, E. Dorigatti, D. Rügamer, and B. Bischl. *Learning Statistical Representation with Joint Deep Embedded Clustering*. Under Review. 2021. arXiv: 2109.05232.

[7] D. Rügamer, P. Baumann, T. Kneib, and T. Hothorn. *Probabilistic Time Series Forecasts with Autoregressive Transformation Models*. Under Review. 2021.

[8] D. Rügamer, C. Kolb, C. Fritz, F. Pfisterer, P. Kopper, B. Bischl, R. Shen, C. Bukas, L. B. de Andrade e Sousa, D. Thalmeier, P. Baumann, L. Kook, N. Klein, and C. L. Müller. *deepregression: a Flexible Neural Network Framework for Semi-Structured Deep Distributional Regression*. Under Review. 2021. arXiv: 2104.02705.

[9] D. Schalk, B. Bischl, and D. Rügamer. *Accelerated Componentwise Gradient Boosting using Efficient Data Representation and Momentum-based Optimization*. Under Review. 2021. arXiv: 2110.03513.

[10] D. Rügamer, C. Kolb, and N. Klein. *Semi-Structured Deep Distributional Regression: Combining Structured Additive Models and Deep Learning*. Under Review. 2020. arXiv: 2002.05777.

[11] D. Rügamer, F. Pfisterer, and B. Bischl. *Neural Mixture Distributional Regression*. 2020. arXiv: 2010.06889.

Published Articles

[12] C. Fritz, E. Dorigatti, and D. Rügamer. "Combining Graph Neural Networks and Spatio-temporal Disease Models to Predict COVID-19 Cases in Germany". In: *Scientific Reports* 12 (2022), pp. 2045–2322.

[13] P. Kopper, S. Wiegrebe, B. Bischl, A. Bender, and D. Rügamer. "DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Complex Hazard Structures in Survival Analysis". In: *Advances in Knowledge Discovery and Data Mining (PAKDD)*. Springer International Publishing, 2022.

[14] B. X. Liew, F. M. Kovacs, D. Rügamer, and A. Royuela. "Machine learning versus logistic regression for prognostic modelling in individuals with non-specific neck pain". In: *European Spine Journal* (2022), pp. 1–10.

[15] D. Rügamer, P. Baumann, and S. Greven. "Selective Inference for Additive and Mixed Models". In: *Computational Statistics and Data Analysis* 167 (2022), p. 107350.

[16] P. F. M. Baumann, T. Hothorn, and D. Rügamer. "Deep Conditional Transformation Models". In: *Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*. Springer International Publishing, 2021, pp. 3–18.

[17] A. Bender, D. Rügamer, F. Scheipl, and B. Bischl. "A General Machine Learning Framework for Survival Analysis". In: *Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*. Springer International Publishing, 2021, pp. 158–173.

[18] C. Berninger, A. Stöcker, and D. Rügamer. "A Bayesian Time-Varying Autoregressive Model for Improved Short- and Long-Term Prediction". In: *Journal of Forecasting* 41.1 (2021), pp. 181–200.

[19] S. Coors, D. Schalk, B. Bischl, and D. Rügamer. "AutoCompboost: Automatic Interpretable Boosting". In: *ECML-PKDD Workshop on Automating Data Science (ADS2021)*. 2021. arXiv: 2109.05583.

[20] D. Falla, V. Devecchi, D. Jimenez-Grande, D. Rügamer, and B. X. Liew. "Modern Machine Learning Approaches Applied in Spinal Pain Research". In: *Journal of Electromyography and Kinesiology* (2021).

[21] J. Goschenhofer, R. Hvingelby, D. Rügamer, J. Thomas, M. Wagner, and B. Bischl. "Deep Semi-Supervised Learning for Time Series Classification". In: *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2021.

[22] P. Kopper, S. Pölsterl, C. Wachinger, B. Bischl, A. Bender, and D. Rügamer. "Semi-Structured Deep Piecewise Exponential Models". In: *Proceedings of AAAI Spring Symposium on Survival Prediction – Algorithms, Challenges, and Applications, PMLR*. 2021, pp. 40–53. arXiv: 2011.05824.

[23] B. X. Liew, H. Y. Lee, D. Rügamer, A. M. D. Nunzio, N. R. Heneghan, D. Falla, and D. W. Evans. "A novel metric of reliability in pressure pain threshold measurement". In: *Scientific Reports (Nature)* 10.6944 (2021).

[24] B. X. Liew, D. Rügamer, K. Duffy, M. Taylor, and J. Jackson. "The mechanical energetics of walking across the adult lifespan". In: *PLOS ONE* (2021).

[25] B. X. Liew, D. Rügamer, X. J. Zhai, S. Morris, and K. Netto. "Comparing machine, deep, and transfer learning in predicting joint moments in running". In: *Journal of Biomechanics* 129 (2021), p. 110820.

[26] M. Mittermeier, M. Weigert, and D. Rügamer. "Identifying the atmospheric drivers of drought and heat using a smoothed deep learning approach". In: *NeurIPS 2021, Tackling Climate Change with Machine Learning*. 2021.

[27] F. Ott, D. Rügamer, L. Heublein, B. Bischl, and C. Mutschler. "Joint Classification and Trajectory Regression of Online Handwriting using a Multi-Task Learning Approach". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2021.

[28] B. Säfken, D. Rügamer, T. Kneib, and S. Greven. "Conditional model selection in mixed-effects models with cAIC4". In: *Journal of Statistical Software, Articles* 99.8 (2021), pp. 1–30.

[29] T. Weber, M. Ingrisch, B. Bischl, and D. Rügamer. "Towards modelling hazard factors in unstructured data spaces using gradient-based latent interpolation". In: *NeurIPS 2021, Deep Generative Models and Downstream Applications*. 2021. arXiv: 2110.11312.

[30] T. Weber, M. Ingrisch, M. Fabritius, B. Bischl, and D. Rügamer. "Survival-oriented embeddings for improving accessibility to complex data structures". In: *NeurIPS 2021, Bridging the Gap: From Machine Learning Research to Clinical Practice*. 2021. arXiv: 2110.11303.

[31] S. Brockhaus, D. Rügamer, and S. Greven. "Boosting Functional Regression Models with FDboost". In: *Journal of Statistical Software* 94.10 (2020), pp. 1–50.

[32] B. X. Liew, A. Peolsson, D. Rügamer, J. Wibault, H. Löfgren, A. Dedering, P. Zsigmond, and D. Falla. "Clinical predictive modelling of post-surgical recovery in individuals with cervical radiculopathy – a machine learning approach". In: *Scientific Reports (Nature)* 10.16782 (2020).

[33] B. X. Liew, D. Rügamer, D. Abichandani, and A. M. De Nunzio. "Classifying individuals with and without patellofemoral pain syndrome using ground force profiles – Development of a method using functional data boosting". In: *Gait & Posture* 80 (2020), pp. 90–95.

[34] B. X. Liew, D. Rügamer, A. D. Nunzio, and D. Falla. "Interpretable machine learning models for classifying low back pain status using functional physiological variables". In: *European Spine Journal* 29 (2020), pp. 1845–1859.

[35] B. X. Liew, D. Rügamer, A. Stocker, and A. M. De Nunzio. "Classifying neck pain status using scalar and functional biomechanical variables – development of a method using functional data boosting." In: *Gait & Posture* 75 (2020), pp. 146–150.

[36] D. Rügamer and S. Greven. "Inference for L2-Boosting". In: *Statistics and Computing* 30 (2020), pp. 279–289.

[37] D. Rügamer, S. Brockhaus, K. Gentsch, K. Scherer, and S. Greven. "Boosting factor-specific functional historical models for the detection of synchronization in bioelectrical signals". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 67.3 (2018), pp. 621–642.

[38] D. Rügamer and S. Greven. "Selective inference after likelihood-or test-based model selection in linear models". In: *Statistics & Probability Letters* 140 (2018), pp. 7–12.

[39] L. Klüser, P. Holler, J. Simak, G. Tater, P. Smets, D. Rügamer, H. Küchenhoff, and G. Wess. "Predictors of sudden cardiac death in Doberman Pinschers with dilated cardiomyopathy". In: *Journal of veterinary internal medicine* 30.3 (2016), pp. 722–732.

[40] J. Gillhuber, D. Rügamer, K. Pfister, and M. C. Scheuerle. "Giardiosis and other enteropathogenic infections: a study on diarrhoeic calves in Southern Germany". In: *BMC research notes* 7.1 (2014), p. 112.