Moritz N. Lang

Data Scientist

I currently work as a post-doc in data science at F. Hoffmann-La Roche Ltd, where I develop statistical forecasting tools in the form of web applications for drug formulation development. Previously, I worked at the Department of Statistics at the University of Innsbruck, from which I graduated in 2020. My PhD thesis, under the supervision of Georg J. Mayr and Achim Zeileis, was on advanced statistical methods for probabilistic forecasting within the domain of natural science. The statistical models employed range from parametric to non-parametric machine learning approaches, whereas the applications include one-dimensional, multivariate and circular responses.

My research stands at the intersection between computational statistics and natural science with a focus on probabilistic forecasting. In this framework, I am a (co-)developer of several R-packages for estimating distributional random forests and graphically evaluating probabilistic models. I enjoy working on descriptive and predictive problems while being comfortable with the entire data science pipeline, from restructuring different types of input data, to data analysis and statistical modeling, to building web applications to visualize the results.

Professional Experience

2022 -

Post Doctorate Scientist
Pharmaceutical R&D, F. Hoffmann-La Roche Ltd, Basel, Switzerland
- Development of statistical prediction tools to help design drug product formulations.
- Design and development of R Shiny web applications for data visualizations and statistical forecasting (incl. GitLab CI/CD).
- Data cleaning and wrangling with R and Python, as well as data management with Google Cloud databases.

2020

Post-Doctoral University Assistant
Department of Statistics, Faculty of Economics and Statistics, University of Innsbruck, Austria
- R-package development for estimating distributional random forests and graphically evaluating probabilistic models.
- Teaching and supervision of master students within the programs Economics, Atmospheric Sciences, and Data Science.
- Consulting of industry partners within Tiroler Data Science Bootcamp.

2017

Pre-Doctoral Research Assistant
Department of Statistics, Faculty of Economics and Statistics, University of Innsbruck, Austria
- Development of advanced statistical methods for probabilistic wind forecasting as part of an industry collaboration with the Austrian air navigation service provider.
- FFG Project: 'Profcast – Probabilistic nowcasting of wind profiles'

2015

Research Assistant
Section Model Applications - Division Data, Methods and Models, Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria
- Statistical post-processing of spatial (weather) forecasts.
- Contribution to the development of an operational probabilistic prediction system.

2014

Student Assistant
Atmospheric Sciences, Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Austria
- Data analysis and visualization with MATLAB and Python.
- FWF-Project: 'Quantifying exchange processes over mountainous terrain'

Contact Information

moritz.n.lang@gmail.com
https://moritzlang.org
moritz-n-lang
MoritzNLang

Research Interests

- Machine Learning
- Computational Statistics
- Data Visualization
- Probabilistic Forecasting
- Multivariate Statistics

Technical Skills

- Data Analysis and Modeling with R
- Programming in R and Python
- Data Management
- Web (App) Development
- R Package Development
- High-Performance Computing
- Source Control

Educational Background

2020 **Doctor of Philosophy**
Atmospheric Sciences, Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Austria

- Various applied statistics courses at PhD and Master's level such as microeconometrics, time series analysis, and advanced regression methods.
- Thesis: 'Probabilistic Wind Forecasting in the Framework of Distributional Modeling'
- Supervisor: Ao. Univ.-Prof. Dr. Georg J. Mayr, Univ.-Prof. Dr. Achim Zeileis
- Reviewer: Prof. Dr. Christophe Ley, Dr. Michael Scheuerer

2015 **Master of Science**
Atmospheric Sciences, Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Austria

- Thesis: 'The impact of embedded valleys on daytime pollution transport over a mountain range - Idealised large-eddy simulations'

2011 **Bachelor of Science**
Atmospheric Sciences, Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Austria

- Various courses in mathematics and physics at the Institute for Theoretical Physics.
- Thesis: 'Feinskalige Struktur von Kaltfronten im Inn- und Wipptal während MAP'

2006 **Allgemeine Hochschulreife**
Ignaz-Günther-Gymnasium, Rosenheim, Germany

2003 **Academic Year Abroad**
Hugh-Christie-Technology College, Tunbridge Wells, England

Teaching Experience

2021 | 2022 **Tiroler Data Science Bootcamp**
Consulting of industry partners funded by 'Leuchtturmprojekte im Bereich Digitalisierung', University of Innsbruck, Austria

2020 | 2022 **Statistical Data Analysis**
Courses for Bachelor's Programme Management and Economic, University of Innsbruck, Austria

2020 | 2022 **Data Science in Practice**
Supervision of master's theses within continuing education program Data Science, University of Innsbruck, Austria

Awards

2022 **Best PhD Thesis**
Award for the best PhD thesis in Applied Statistics 2021 by the Austrian Statistical Society (ÖSG)

2022 **Best PhD Thesis**
Award for the best PhD thesis of the Faculty of Geo- and Atmospheric Sciences in the academic year 2021/22

2012 **Scholarship for Excellence**
Master of Science, Atmospheric Sciences, Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Austria

R Package Development

topmodels: Infrastructure for Inference and Forecasting in Probabilistic Models
Senior Developer
- Unified infrastructure for probabilistic models and distributional regressions: Computation of probabilities, densities, scores, and Hessians.
- Diagnostic graphics such as rootograms, PIT histograms, reliagrams, quantile residual Q-Q plots, and worm plots.

disttree: Trees and Forests for Distributional Regression
Senior Developer
- Infrastructure for fitting distributional regression trees and random forests based on maximum-likelihood estimation.

circtree: Regression Trees and Forests for Circular Responses
Senior Developer
- Infrastructure for fitting distributional trees and random forests based on maximum-likelihood estimation of parameters for a circular response.

Selected Publications and Conferences

Full publication list available at:
https://moritzlang.org/publication

2021 **Probability distribution forecasts: Learning with random forests and graphical assessment**
Moritz N. Lang, Reto Stauffer, Lisa Schlosser, Achim Zeileis

Talk at useR! 2021, Zürich, Switzerland (virtual)

2020 **Circular regression trees and forests with an application to probabilistic wind direction forecasting**
Moritz N. Lang, Lisa Schlosser, Torsten Hothorn, Georg J. Mayr, Reto Stauffer, Achim Zeileis

Journal of the Royal Statistical Society: Series C (Applied Statistics), 69, 1357–1374, https://doi.org/10.1111/rssc.12437

2020 **Remember the past: A comparison of time-adaptive training schemes for non-homogeneous regression**
Moritz N. Lang, Sebastian Lerch, Georg J. Mayr, Thorsten Simon, Reto Stauffer, Achim Zeileis

Nonlinear Processes in Geophysics, 27, 23–34, https://doi.org/10.5194/npg-27-23-2020

2019 **Bivariate Gaussian models for wind vectors in a distributional regression framework**
Moritz N. Lang, Georg J. Mayr, Reto Stauffer, Achim Zeileis

Advances in Statistical Climatology, Meteorology and Oceanography, 5, 115–132, https://doi.org/10.5194/ascmo-5-115-2019, 2019.