

EGU22-6543

<https://doi.org/10.5194/egusphere-egu22-6543>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



High-resolution hybrid spatiotemporal modeling of daily relative humidity across Germany for epidemiological research: a Random Forest approach

Nikolaos Nikolaou¹, Laurens Bouwer², Mahyar Valizadeh¹, Marco Dallavalle¹, Kathrin Wolf¹, Massimo Stafoggia^{3,4}, Annette Peters¹, and Alexandra Schneider¹

¹Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany (nikolaos.nikolaou@helmholtz-muenchen.de)

²Climate Service Center Germany (GERICS), Helmholtz-Zentrum Hereon, Hamburg, Germany

³Department of Epidemiology, Lazio Regional Health Service, Rome, Italy

⁴Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

Introduction: Relative humidity (RH) is a meteorological variable of great importance as it affects other climatic variables and plays a role in plant and animal life as well as in human comfort and well-being. However, the commonly used weather station observations are inefficient to represent the great spatiotemporal RH variability, leading to exposure misclassification and difficulties to assess local RH health effects. There is also a lack of high-resolution RH spatial datasets and no readily available methods for modeling humidity across space and time. To tackle these issues, we aimed to improve the spatiotemporal coverage of RH data in Germany, using remote sensing and machine learning (ML) modeling.

Methods: In this study, we estimated German-wide daily mean RH at 1km² resolution over the period 2000-2020. We used several predictors from multiple sources, including DWD RH observations, Ta predictions as well as satellite-derived DEM, NDVI and the True Color band composition (bands 1, 4 and 3: red, green and blue). Our main predictor for estimating the daily mean RH was the daily mean Ta. We had already mapped daily mean Ta in 1km² across Germany through a regression-based hybrid approach of two linear mixed models using land surface temperature. Additionally, a very important predictor was the date, capturing the day-to-day variation of the response-explanatory variables relationship. All these variables were included in a Random Forest (RF) model, applied for each year separately. We assessed the model's accuracy via 10-fold cross-validation (CV). First internally, using station observations that were not used for the model training, and then externally in the Augsburg metropolitan area using the REKLIM monitoring network over the period 2015-2019.

Results: Regarding the internal validation, the 21-year overall mean CV-R² was 0.76 and the CV-RMSE was 6.084%. For the model's external performance, at the same day, we found CV-R²=0.75 and CV-RMSE=7.051% and for the 7-day average, CV-R²=0.81 and CV-RMSE=5.420%. Germany is characterized by high relative humidity values, having a 20-year average RH of 78.4%. Even if the

annual country-wide averages were quite stable, ranging from 81.2% for 2001 to 75.3% for 2020, the spatial variability exceeded 15% annually on average. Generally, winter was the most humid period and especially December was the most humid month. Extended urban cores (e.g., from Stuttgart to Frankfurt) or individual cities as Munich were less humid than the surrounding rural areas. There are also specific spatial patterns for RH distribution, including mountains, rivers and coastlines. For instance, the Alps and the North Sea coast are areas with elevated RH.

Conclusion: Our results indicate that the applied hybrid RF model is suitable for estimating nationwide RH at high spatiotemporal resolution, achieving a strong performance with low errors. Our method contributes to an improved spatial estimation of RH and the output product will help us understand better the spatiotemporal patterns of RH in Germany. We also plan to apply other ML techniques and compare the findings. Finally, our dataset will be used for epidemiological analyses, but could also be used for other research questions.