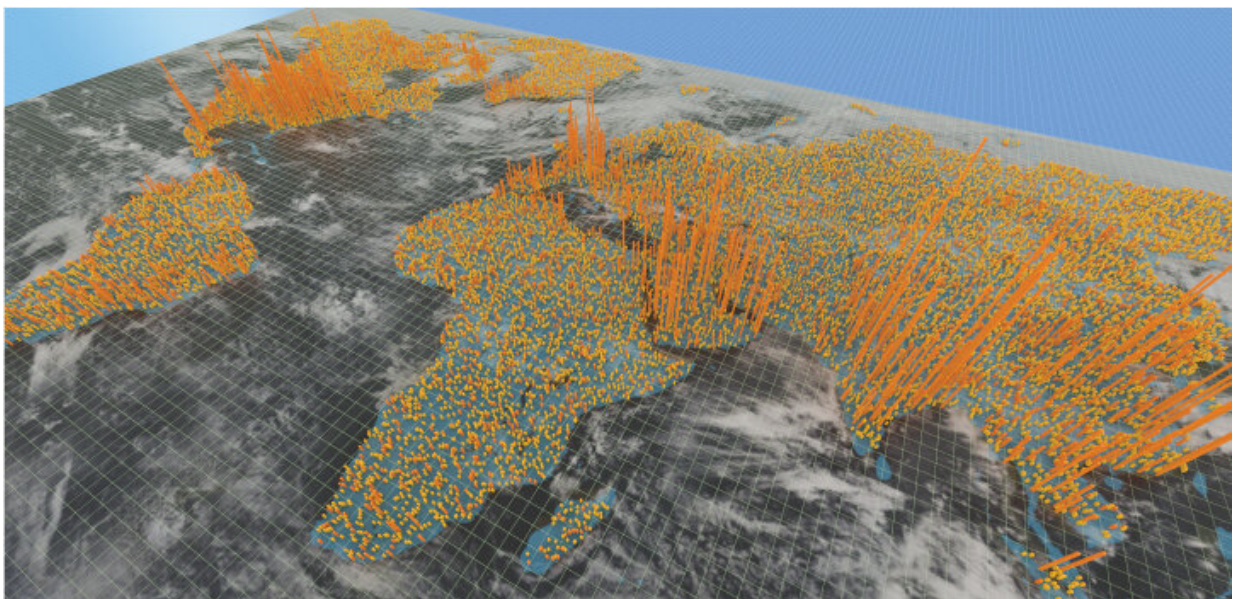# A model for millions of locations enables better prediction of climate and environmental conditions

August 8 2022

A framework developed by KAUST statisticians enables modeling of a range of meteorological and environmental datasets from up to 2 million locations globally. Credit: KAUST; Heno Hwang

A faster and more statistically accurate modeling scheme enables better prediction of climate and environmental conditions at very large scales.

Combining nuanced statistical methods with a robust parallel

computational platform has enabled a modeling scheme that better predicts [environmental conditions](link) while being efficient enough to cover millions of monitoring locations.

The new modeling approach developed by KAUST tackles a longstanding obstacle to improved weather and climate prediction: how to implement non-Gaussian statistics for very large geospatial datasets.

"In spatial statistics, the main objective is to use data observed at monitoring stations to predict the conditions at unobserved locations," explains Sagnik Mondal, a Ph.D. student from Marc Genton's statistics research group. "These types of predictions are necessary for many kinds of weather and climate applications. Nowadays, however, the number of observation locations can reach millions, which is beyond the capability of traditional computational approaches, and the traditional Gaussian models fail to statistically capture extreme values."

A Gaussian model is a straightforward statistical description of a dataset based on an average "mean" value and symmetric distributions to higher and lower values—the iconic "bell curve." However, many environmental variables and their derivates—like rainfall intensity, [wind speed](link), days without rain or days above a certain temperature—are not symmetric in their distribution. Rather, they have peak probabilities hovering close to zero but can, on rare occasions, reach significantly high extremes. This long "tail" to extreme values with very low probability cannot be captured by Gaussian models but is becoming increasingly important under climate change.

"In this work, we applied the Tukey g-and-h model, which is a non-Gaussian spatial model with two additional parameters to accommodate asymmetric distributions and better capture extreme values," says Mondal.

While the Tukey model is clearly beneficial for weather data, it is not efficient enough to apply in practice for large geospatial data sets as a traditional sequential computation. However, it can be significantly improved by parallelizing the computations.

"Gaussian models have already been parallelized, and so we set out to implement the Tukey model for the first time using a state-of-the-art parallel architecture," says Mondal.

Running the new modeling scheme on KAUST's Shaheen-II supercomputer, the research team demonstrated the model's performance using real precipitation data from more than 300,000 locations across Germany and using a synthetic dataset of more than 800,000 stations.

"Our framework enables us to fit the exact model to datasets as large as 1 million locations and, with additional approximations, up to 2 million locations," Mondal says. "By using parallel computations, we are providing an avenue for modeling large-scale geospatial data."

The study was published as part of the *2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*.

Provided by King Abdullah University of Science and Technology

Citation: A model for millions of locations enables better prediction of climate and

environmental conditions (2022, August 8) retrieved 9 December 2023 from
https://techxplore.com/news/2022-08-millions-enables-climate-environmental-conditions.html