

Competition on spatial statistics showcases the global state of the art in analyzing vast spatial datasets

March 6 2023



Credit: Pixabay/CC0 Public Domain

In 2021, KAUST ran the first international competition "Spatial Statistics for Large Datasets" using ExaGeoStat's reference datasets. Following that successful event, the Exascale Geostatistics Project (ExaGeoStat) led by Marc Genton and Ying Sun ran a second

competition in March-May 2022 attracting researchers from ten groups around the world.

The analysis and interpretation of large spatial datasets consisting of millions of monitoring locations and many parameters and observations over time is a major new frontier in [data science](#) and statistics. Such datasets require not only highly specialized computing systems to store and process the data, but new statistical methods that can reduce the amount of computation while retaining interpretability and accuracy. However, as the various statistics groups around the world hone methods using their own inhouse datasets, there has been no way to objectively compare the accuracy and performance of different statistical approaches.

Organizers of the event anticipate that the [competition](#) can help to address this need. "This competition was motivated by the absence of a general benchmarking suite for existing spatial statistics methods to assess their accuracy with different data types," says Sameh Abdulah, a research scientist and organizer of the competition.

"Using our ExaGeoStat software, we were able to generate different data types and sizes that can be used to evaluate existing methods in both modeling and prediction capabilities," Abdulah explains.

"Most existing tools can deal with [large datasets](#) using different approximation methods, but it is difficult to evaluate their efficiency since there is generally no exact solution for the datasets they are analyzing. ExaGeoStat can generate very large geospatial datasets and model them to calculate exact solutions that can be used to evaluate the accuracy and efficiency of a statistical method."

Setting six different challenges across three types of datasets, competitors applied their respective statistical methods to derive their

best model approximations. The winners across the six challenges were groups from KAUST as well as from France, Taiwan, China and the U.S.

"This competition helps demonstrate how our ExaGeoStat software can handle large geospatial data on leading-edge hardware architecture, as well as showcasing the best tools and methods for accurate modeling and prediction," says Abdulah. "We hope this will help improve the efficiency of modeling and prediction for applications such as climate and weather forecasting and increase investment in spatial statistics."

Abdulah's team is already organizing the third competition.

"We have had very good responses over the three years of this competition. We have received many submissions from different research groups worldwide, which shows the importance of this annual competition to the spatial statistics community," Abdulah says.

The paper is published in the *Journal of Data Science*.

More information: Sameh Abdulah et al, The Second Competition on Spatial Statistics for Large Datasets, *Journal of Data Science* (2022).

[DOI: 10.6339/22-JDS1076](https://doi.org/10.6339/22-JDS1076)

Provided by King Abdullah University of Science and Technology

Citation: Competition on spatial statistics showcases the global state of the art in analyzing vast spatial datasets (2023, March 6) retrieved 22 June 2024 from

<https://phys.org/news/2023-03-competition-spatial-statistics-showcases-global.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private

study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.