

Competition sheds light on approximation methods for large spatial datasets

January 19 2022



KAUST scientists organized a global competition with 21 competing teams to compare different approximation methods for analyzing large spatial datasets. Credit: KAUST; Anastasia Serin

Organizing a global competition between approximation methods used for analyzing and modeling large spatial datasets enabled KAUST researchers to compare the performance of these different methods.

Spatial datasets can contain many different types of data, from topographical, geometric or [geographic information](#), such as

environmental or [financial data](#), comprising measurements taken across many locations. The development of advanced observation techniques has led to increasingly [larger datasets](#) with high dimensionality, making statistical inference in spatial statistics computationally challenging and very costly.

Various [approximation](#) methods can be used to model and analyze these large real-world spatial datasets, where exact computation is no longer feasible and inference is typically validated empirically or via prediction accuracy with the fitted model. However, there have been few studies that compare the statistical efficiency of these approximation methods, and these have been limited to small- and medium-sized datasets for only a few methods.

This motivated Marc Genton, Huang Huang and colleagues from KAUST to organize a [competition](#) between different approximation methods to assess their model inference performance.

The competition "was designed to achieve a comprehensive comparison between as many different methods as possible and also involved more recently developed methodologies," said Huang. "It was also designed to overcome weaknesses in previous studies by incorporating several key features."

These features included synthetic spatial datasets generated by the ExaGeoStat software, which comprised datasets ranging from 100,000 to one million data points. "With these much larger synthetic datasets, where we know the true processes at scale, we could better compare the statistical efficiency of different approximation methods," Genton explains.

In addition, the data-generating models represented a wide range of statistical properties for both Gaussian and nonGaussian cases and

included both estimation and prediction that were assessed by multiple criteria.

Launched in November 2020, the competition motivated 29 research teams from the global spatial statistics community to register their interest, with 21 teams submitting their results by the close of the competition in February 2021. "By reviewing entries to the [competition](#), we were able to better understand when each approximation method became inadequate," said Huang, which provided "a unified framework for understanding the performance of existing [approximation](#) methods."

"We now plan to extend the comparison to more complex datasets from multivariate or spatio-temporal random processes." he adds.

More information: Huang Huang et al, Competition on Spatial Statistics for Large Datasets, *Journal of Agricultural, Biological and Environmental Statistics* (2021). [DOI: 10.1007/s13253-021-00457-z](https://doi.org/10.1007/s13253-021-00457-z)

Provided by King Abdullah University of Science and Technology

Citation: Competition sheds light on approximation methods for large spatial datasets (2022, January 19) retrieved 21 June 2024 from <https://techxplore.com/news/2022-01-competition-approximation-methods-large-spatial.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.