Exascale Geostatistics for Environmental Data Science

A seminar by Marc G. Genton

King Abdullah University of Science and Technology (KAUST)
Saudi Arabia

Monday 14 Oct 2024 | 2.30 p.m.
Room Benvenuti
Department of Statistical Sciences

Environmental data science relies on some fundamental problems such as: 1) Spatial Gaussian likelihood inference; 2) Spatial kriging; 3) Gaussian random field simulations; 4) Multivariate Gaussian probabilities; and 5) Robust inference for spatial data. These problems develop into very challenging tasks when the number of spatial locations grows large. Moreover, they are the cornerstone of more sophisticated procedures involving non-Gaussian distributions, multivariate random fields, or space-time processes. Parallel computing becomes necessary for avoiding computational and memory restrictions associated with large-scale environmental data science applications. In this talk, I will explain how high-performance computing can provide solutions to the aforementioned problems using tile-based linear algebra, tile low-rank approximations, as well as multi- and mixed-precision computational statistics. I will introduce ExaGeoStat, and its R version ExaGeoStatR, a powerful software that can perform exascale (10^{18} flops/s) geostatistics by exploiting the power of existing parallel computing hardware systems, such as shared-memory, possibly equipped with GPUs, and distributed-memory systems, i.e., supercomputers. I will then describe how ExaGeoStat can be used to design competitions on spatial statistics for large datasets and to benchmark new methods developed by statisticians and data scientists for large-scale environmental data science.

Università degli Studi di Padova

Dipartimento di Scienze Statistiche