

Exploiting mixed-precision to speed-up applications

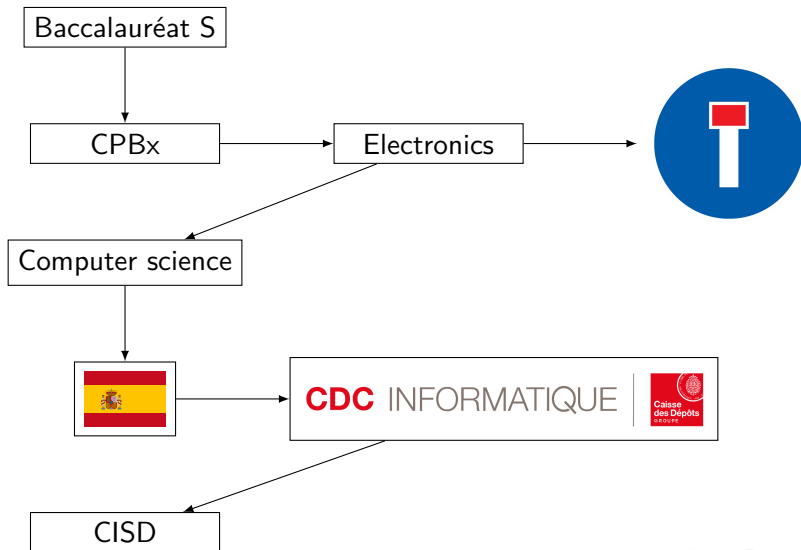
Ana Hourcau

Working group internship presentation

21 March 2024



Curriculum vitae



Exageostat

- Exascale GeoStatistics
- Predict missing observations in climate/weather
- KAUST Spatial Statistics group and Extreme Computing Research Center (ECRC)

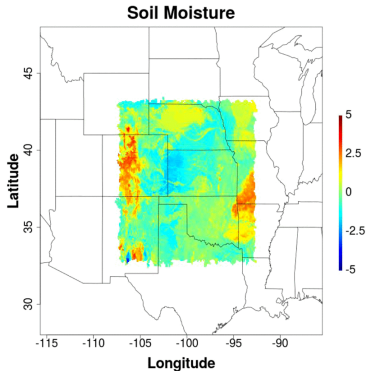
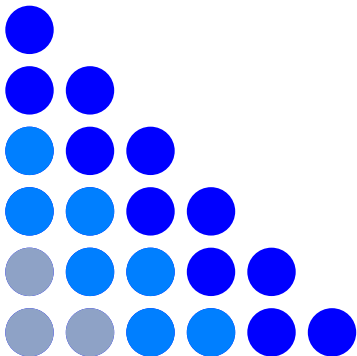


Figure 1: Soil moisture residuals at the topsoil of the Mississippi River basin.

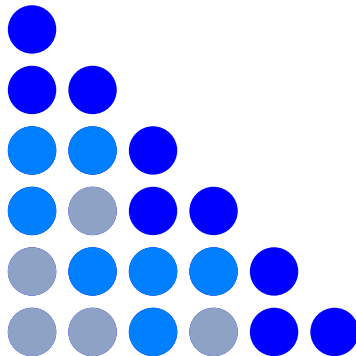
Reference : H. Ltaief, M. G. Genton, D. Gratadour, D. E. Keyes and M. Ravasi, "Responsibly Reckless Matrix Algorithms for HPC Scientific Applications," in *Computing in Science & Engineering*, vol. 24, no. 4, pp. 12-22, 1 July-Aug. 2022, doi: 10.1109/MCSE.2022.3215477.

Different ways of using mixed-precision

- Double precision
- Single precision
- Half precision



Based on the gradient pattern of the correlations, band regions are defined along the diagonal for applying corresponding precision



Based on the values, precision is determined on a tile-by-tile basis, using the Frobenius Norm

Using mixed-precision in StarPU/Chameleon

- Approach already tested with Exageostat + PaRSEC/DPLASMA
 - Limited performances
- Incorporate mixed-precision into StarPU/Chameleon with two leads to investigate:
 - Handling several versions of the data
 - Memory cost: when to discard or generate
 - Using mixed-precision within calculation kernels
 - Half precision only available only on GPUs and only GEMM, {HE,SY}RK

Conclusion

To be done:

- Understand how Exageostat is using mixed-precision with two versions (C and C++)
- Explore how it can be integrated with the operations performed by StarPU/Chameleon in Exageostat

Future work:

- Enable the use of mixed-precision in HPL-MxP
- Use low/mixed-precision in tensor operations