

State-of-the-art in Parallel Computing with R

Markus Schmidberger^{1,*}

1. IBE, Ludwig-Maximilians-Universität Munich, Germany

* Contact author: schmidb@ibe.med.uni-muenchen.de

Keywords: Parallel Computing, Computer Cluster, Multi-core Systems, Grid Computing, Benchmark

R is a mature open-source programming language for statistical computing and graphics. Many areas of statistical research are experiencing rapid growth in the size of data sets. Methodological advances drive increased use of simulations. A common approach is to use parallel computing.

This presentation presents an overview of techniques for parallel computing with R on computer clusters, on multi-core systems, and in grid computing. It reviews sixteen different packages, comparing them on their state of development, the parallel technology used, as well as on usability, acceptance, and performance.

Two packages (`snow`, `Rmpi`) stand out as particularly useful for general use on computer clusters. Packages for grid computing are still in development, with only one package currently available to the end user. For multi-core systems four different packages exist, but a number of issues pose challenges to early adopters. The presentation concludes with ideas for further developments in high performance computing with R.

References

- M. Schmidberger, M. Morgan, D. Eddelbuettel, H. Yu, L. Tierney, U. Mansmann (2009). State-of-the-art in Parallel Computing with R; *Journal of Statistical Software*; submitted. Preprint: <http://epub.ub.uni-muenchen.de/8991/>