

## R at NCAR

Matthew Pocernich, Eric Gilleland and Doug Nychka

The need to understand the complex physical, biological and societal relationships of the Earth system will pose a rich set of statistical problems. We are using our perspective at the National Center of Atmospheric Research to adapt the R environment for the geophysical community. Problems include dealing with very large datasets and working with models that use both observational data and data derived from numerical models. NCAR's mission is in part to "... foster the transfer of knowledge and technology for the betterment of life on Earth." R plays an important and growing role in fulfilling this mission. This is due to R's philosophical commitment to sharing, its acceptance within the statistical community and the increasing sophistication, stability and breadth of applications created by the community.

This presentation will consider the current use of R within the scientific community at NCAR and possible future efforts. In particular, examples of packages created at NCAR will be presented. Limitations (real and perceived) for using R at NCAR will also be discussed. An overview of the several packages that have been developed at NCAR and are ( or will be soon) available on CRAN will be presented. These packages include the following.

**extRemes** is a package developed to assist in teaching extreme value statistics. Developed by Eric Gilleland and others, this packages provides a graphical user interface to the **ismev** package developed by Stuart Coles and implemented in R by Alec Stephenson.

**fields** was developed by the Geophysical Statistics Project (GSP) at NCAR. This package focuses on curve and function fitting, with a special emphasis on spatial problems. Major topics addressed by functions in the package include thin plate spline regression, kriging, space-filling designs and other functions related to working with spatial data.

**RadioSonde** presents and summarizes data from radiosondes and drop-sondes. These devices are attached to balloons or dropped from planes in order to gather atmospheric information.

**verify** was developed internally to standardize some routine model and forecast verification diagnostics. While created with weather forecasts in mind, it is written to be readily applied to other types of forecasts and models. By making algorithms more readily available, we hope to extend discussions of their usefulness from the theoretical to the applied.

While R is universally used by the statisticians at NCAR, far fewer atmospheric scientists and software engineers use it. The following actions and activities could expand the number of R users at NCAR and will be briefly discussed.

- The NCAR Command Language (NCL) is a programming language designed for the analysis and visualization of data. NCL is particularly useful in handling and processing very large datasets stored in netcdf, grib and binary files. Presently, NCL has somewhat limited statistical capabilities. Graphs are produced using a low level graphics language. Allowing users of NCL to use R's statistical and graphics function could greatly help this community.
- Prototyping by software engineers is commonly done in Matlab and so many Matlab libraries exist. The ability to use these Matlab libraries would be key in getting more software engineers to proto-type in R.
- R's GPL license could be a benefit in attracting some software engineers to R. There have been instances when developers wanted to offer their code to an external audience but were prohibited because of Matlab's license.
- Some scientists use Excel to do graphs and simple analyses. When the analysis become too complicated or the datasets too large, software engineers often provide solutions either in Matlab or in a compiled code. Instruction and support for scientists who want to use R will help R gain acceptance amongst these users. NCAR currently has a users group, but it is not as active as it might be. A more active user community will help extend the uses of R at NCAR.