AquaFortR: Streamlining Atmospheric Science, Oceanography, Climate, and Water Research with Fortran-accelerated R

Ahmed Homoudi[1] and Klemens Barfus[1]

[1]Institute of Hydrology and Meteorology, Technische Universität Dresden, Pienner Street 23, Tharandt, D-01737, Germany

07.07.2023

I. Discipline(s)

Atmospheric Science, Oceanography, Climate and Water Research.

II. Format(s)

We expect the end products to include: a) An R-Package, including the reproducible workflow of integrating Fortran in R with examples from the disciplines mentioned above; b) an interactive swirl course will be provided; c) and lastly, for readers, an online book will be published. These resources can be deployed online (e.g., GitHub or OneStop4All). In addition, we can offer a half-day workshop within an NFDI$_4$Earth event or as a separate occasion.

III. Target group(s)

Bachelor's, Master's, PhD Students and Researchers.

IV. Project summary

Generally, simulation and modelling of the environmental processes are accomplished on the grid level in which the investigation region is discretised to numerous grid points in the three dimensions of space plus time. Consequently, these simulations produce enormous data sets and processing these data extends beyond the current average Personal Computer (PC) capacity. However, only some people have access to high-performance computing centres (i.e., generally scientists in the Global North). Additionally, the possibility of speeding up calculations and modelling exists in each PC through compiled programming languages such as Fortran. This solution speeds up computations and can reduce the CO_2 footprint drastically.

R is one of the languages widely used in data analysis, visualisation, and presentation, and it has a wide supporting community and thousands of packages. Nevertheless, Fortran is one of the fastest-performing languages -if not the fastest in number crunching- and one of the oldest. Due to the latter, the interest in Fortran is constantly low (Ref). Considering all the above, the need for educational material that links R and Fortran is essential.

Open Educational Resources (OERs) regarding the integration of Fortran and R exist on the internet; however, these resources are individual efforts, goal-unoriented and generally outdated. In this project, we aim to provide one OER platform that will be a one-stop for all R

users looking for speed in general and users from Environmental Science disciplines in particular.

Many developers have achieved efforts to speed up R using C++ (e.g., Rcpp package in R); however, integration of Fortran and R in such a package has yet to exist, to the best of the authors' knowledge. Filling this gap is important because Fortran is well-suited for numerical and scientific computations due to its array processing capabilities, performance, and efficiency. Generally, Computationally demanding models (e.g., climate models) are written in Fortran; thus, integrating Fortran and R will allow environmental modellers and researchers to minimise changes between different programming languages. Besides computational efficiency, code longevity is a feature of Fortran which supports the sustainability and interoperability goals of NFDI$_4$Earth.

Our goal is to present three challenging computational examples when considering large datasets: 1) two-dimensional cross-correlation (Xcorr2D), 2) two-dimensional convolution (conv2D), and 3) Convective Available Potential Energy (CAPE). Xcorr2D is a similarity measure widely used in image processing and Particle Image Velocimetry (e.g., cloud tracking). We have already tested the performance of multiple calculation approaches, e.g., using FFT (Ref). Secondly, conv2D is similar to Xcorr2D but is used to blur images and remove noises, and it can be accelerated using FFT. For example, it can be used in delineating clouds and precipitation objects and as a Gaussian filter for Radar data (For DE: 5min, 1km). Lastly, CAPE is important for evaluating atmospheric instability because higher values indicate potential hazards such as thunderstorms and hail. Various factors, including the vertical resolution of the atmospheric profile, influence the accuracy of calculating CAPE. The computations above may pose no challenge for single task, but it becomes a complex process when dealing with vast atmospheric datasets such as climate model output. The work will include the integration of Fortran in R scripts and R Packages (using C). Finally, these efforts will be documented concurrently with other tasks in an R Swirl course, an R package, and a book.

Inclusivity is an important and basic human right; therefore, we will aim to provide the OERs materials as accessible as possible to include people with disabilities. The material will be in English to reach a wider public. Individuals aiming to use the materials from this project are expected to have access to an average PC and the internet. Basic R or any programming language knowledge is beneficial but optional. The expected table of contents is (but is not limited to):

- Brief Introduction to R, Rstudio, and Fortran
- Integration of Fortran in R scripts, and R packages (using C)
- The three examples of Fortran integration in R including performance tests.

V. Duration & requested funding

- Project budget requested from NFDI$_4$Earth with breakdown by category (e.g., funds requested for a research assistant, media production, etc.):
  - A Senior Student Assistant (WHK) with 10 Hour/Week (17.28 EUR/Hour = 5750.78 EUR). The fund is calculated as follows: Average monthly hours (10 Hours x 4/12 Weeks x 13) x 6 Months x 17.28 EUR x 28% Arbeitgeberanteil→ 5750.78 EUR
- Proposed duration of grant in months: Six months.