EuroHPC:
Requirements from Weather and Climate

Bryan Lawrence (bryan.lawrence@ncas.ac.uk)
V1.0 12/06/2018. On behalf of the ESIWACE Centre of Excellence and the wider weather and climate community.

Strategic context: the 2017 update of the 2012-2022 European Network for Earth System modelling (ENES) infrastructure strategy made seven substantive recommendations, all of which impact on EuroHPC. They include the need to *accelerate the preparation for exascale computing, exploit a blend of national and European HPC, maximise bandwidth between facilities, grow the size of, and skills within, the community, and strengthen the collaboration with European actors providing services to ENES.*

Scientific Context: Weather and Climate science depends on simulation, but simulations do not produce knowledge, they produce data. Sustained post-processing is necessary to turn data into knowledge, and this process can involve large downstream communities and take (in the case of climate) years. Modelling campaigns (whether for pure science, or to develop new weather prediction systems) take years from the planning phase through to the first use of the data in post-processing. Hence the full set of requirements cover what is necessary for planning, preparing, running models, and then managing data through to post-processing, potentially with data live for years.

Planning and Logistics: require allocation mechanisms that recognise wider context, and need to synchronise HPC resource with grant mechanisms and other ways of ensuring the necessary human resources are in place throughout the lifecycle from conception to data analysis. It is important to realise that it is the science data analysis (or weather products) that are the purpose of the simulation, and so the simulation is only necessary, and far from sufficient!

Runtime Environments: must support multiple executables, and deliver high performance well maintained Fortran, C, C++, MPI, and hybrid MPI/OpenMP compilers and libraries. These environments need to be stable, and predictable, but support rapid bug-fixing and the necessary modules and branches. Systems must be able to support large volumes of both input and output data. Queuing and user environments must reflect workflow requirements.

Performance Evaluation: needs to be cast in terms of scientific units, such as *Simulated Years Per Day* (SYPD) for well defined codes using normal full input and output – benchmarking complete weather and climate codes without I/O is very misleading.

Pre-Exascale: In the near term, for the pre-exascale, existing codes will need to be modified, and so we expect to see benchmarking using evolution of existing codes, and machines selected in terms of their scientific performance using the sum of weighted performance of selected benchmark production science codes with respect to a reference machine.

Exascale (1): In the medium term, the weather and climate community is active in many projects developing the tools necessary to co-design exascale machines using “mini-apps” or “dwarfs” which expose particular numerical and data handling behaviours. These benchmarks will be integral in the design of Exascale platforms, however their success will require not only the procurement of the exascale machines themselves but also new codes which exploit the co-design process..

Exascale (2): The complete scientific exploitation of next generation machines will require massive investment in revolutionary new codes (based on the co-design), which will require a workforce which both has a different skillset and is larger than the existing workforce. This will require significant community collaboration – which is reflected in both the ENES recommendations, and the development of a strong community bid for a FET flagship - ExtremeEarth.

Data Infrastructure is crucial to the success of weather and climate simulation. Not only do HPC systems themselves have to include high performance (and large capacity) storage systems, but the longer-term analysis needs even higher capacity, and larger sustained I/O performance. These requirements lead to the need for dedicated customised high performance data storage and analysis platforms.