



Data preservation and standardization in computational fluid dynamics

Our project is meant to preserve and standardize a first set of state-of-the-art numerical datasets in computational fluid dynamics, concerning: (i) fully homogeneous and isotropic turbulence evolved on a fractal Fourier set, (ii) a world record simulation of a turbulent flow with rotation at 40963 collocation points (iii) multi-component microfluidics in complex geometries. Data-sets include both Eulerian and Lagrangian data, i.e. snapshot of the velocity field and trajectories of particles affected by the flow. All data are of potential interest for a vast community of researchers, mostly in Europe and in the USA, in the fields of theoretical physics, geophysics, meteorology, chemical and bio-engineering.

The Scientific Challenge

The Computational Fluid Dynamics (CFD) community is increasingly facing the problem of data preservation, data standardization and data analysis (by both the data owners and by third parties). It is therefore mandatory to develop user friendly supports and optimal interfaces to make the data available and useful for a long period of time. Besides the obvious scientific interests of the owner groups, the availability of these large data sets is potentially crucial for a much wider audience of theoretical and applied scientists working in different cross-disciplinary domains, who do not intend --or cannot do-- numerical simulations on their own. Moreover, it is important to mention that these accurate and high-resolution (both in space and time) datasets cannot be obtained by any commercial CFD software because of the very strict requirements about the precision, error control, statistical accuracy etc. Systematic analysis and classification of huge datasets is a challenge for both the needed man power and for the storage requirements.

Our research group is involved in many collaborations all over Europe and worldwide, including the International Collaboration for Turbulence Research (ICTR) and the European High-Performance Infrastructures in Turbulence (EuHIT) project, two initiatives that count more than 100 scientists in the domain of numerical, theoretical and experimental fluid mechanics.

Who benefits and how?

The most important goal of this pilot is to provide fluid dynamics scientists with a platform to test, assess or dismiss ideas and models in the field of turbulence, microfluidics and related applications. The direct target is a broad community, encompassing people working on complex flows, active matter, bio- and chemical-engineering, material processing, cloud physics, air quality and ocean sciences. The benefit is clear and of fundamental importance: performing state-of-the-art numerical simulations is a difficult task, requiring a team of people working full time on the problem. On the other hand, in view of the expensive and complex nature of experiments, numerical simulations data are of extreme utility: sometimes they access quantities that actual sensors or probes cannot deal with. So the availability of these large data sets is crucial for theoretical, but mostly for applied scientists working in different cross-disciplinary domains, who do not intend or cannot do numerical simulations on their own.

The expected impact of the pilot project is very large: in particular the data standardization, in addition to a continuous in time and fully reliable accessibility can make a set of data a renewing source of research and applied activities.

Technical Implementation

Since our project is meant to preserve and standardize a first set of state-of-the-art numerical datasets in computational fluid dynamics, it's main nature is of data storage. Our simulation data are already present in two

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.



PRACE facilities, so we used mainly B2STAGE by grid-ftp to move data to the repository we are building. No particular problems were found during this phase: dealing with big files (ranging from 25GB to 1.5TB) carried some delay/trouble when checksumming for the integrity check, but this was fixed in the starting phase of the pilot.

Preliminary Results

As of Nov 2016, in Turbase-DNS there are already 70TB of datasets representative of several numerical experiments with different forcing, and filtering/decimation in Fourier space. This is only the first step, since the repository is expected to grow till 200TB. Work is ongoing to create a good catalog for reaching a good searchability of data, further investigating the possibility to interoperate with the EUHIT repository for experimental data. Typical files in Turbase-DNS are encoded in the HDF5 file format, which permits to link some extra information in the file itself, avoiding the possibility of mismatch between metadata files and computational results. At the time of the simulations some information were already added in the files, mainly to remember detailed physical parameters used. Meanwhile the analysis work on the datasets is progressing, and we are adding new attributes to these files, aiming at producing well formatted self-consistent datasets, which are of primary importance.

Contacts

- Luca Biferale, University of Rome, Tor Vergata, [luca.biferale\(a\)roma2.infn.it](mailto:luca.biferale(a)roma2.infn.it)
- Fabio Bonaccorso, INFN, [fabio.bonaccorso\(a\)roma2.infn.it](mailto:fabio.bonaccorso(a)roma2.infn.it)

Further Information

[Scientific Challenges behind the Pilot](#)

[Learn more about the EUDAT/TURBASE-DNS Collaboration here](#)

[Read more](#)