

Too big or not too big... Big-data challenges in hydraulic structures applications

José Barateiro^{1,2}

¹*National Laboratory for Civil Engineering, Av Brasil, 101, Lisboa, Portugal*

²*NOVA Information Management School, Universidade Nova de Lisboa, Campus de Campolide, Lisboa, Portugal*

E-mail: jbarateiro@lnec.pt

ABSTRACT: As originally explained by Jim Gray, the world of science is changing to an era where data can drive research in multiple fields. In fact, the beginning of science was experimental, and then highly theoretical with Newton's, Maxwell's, etc. laws and equations. Then, many problems grew very fast, making them so complex to be analytically solved that researchers started to simulate solutions. These simulations drove the research in multiple fields during the last decades. Nowadays, gathering data that describe scientific events is much easier, allowing scientists to run their processes on top of data-intensive systems, facing new challenges related to data volume, velocity and variety (the 3 *Vs* of big data). In fact, data-intensive processing is used to verify/validate existing knowledge, but also to produce new knowledge, represented by complex relations and data dependencies, that would be impossible to produce with traditional research paradigms. The civil engineering domain in general, and hydraulics in particular, can also take advantage of big datasets produced by monitoring sensors and simulations, being capable to generate high-quality results and running data-intensive algorithms that would fail in a traditional (non-big-data) environment. This talk will follow a top-down approach, presenting the main technical challenges and solutions in big data applications for any domain. It will clarify and distinguish the use of high-performance computing (HPC), high-throughput computing (HTC) and big data. The fundamental technical concepts are then used to motivate real examples in scenarios and applications where the Portuguese National Laboratory for Civil Engineering is an important stakeholder, from both the engineering and technological perspective.

Keywords: Big Data, Science Paradigms, High Performance Computing