

## New decimation method for fusion research data

*Tuesday, 14 May 2019 10:05 (5 minutes)*

New fusion research experiments will generate massive experimental data. For example, ITER will have above one million of variables from control signals and diagnostic systems. Some of these variables will produce data during long pulse (about 30 minutes) experiments and other will generate data continuously. Just to have a clearer idea of the problem, ITER estimates more than 10 GBytes/second of data flow during experiment pulses. In this context, fusion research appears in the scope of the big data where both search and access functions require new approaches and optimizations.

One common data access functionality is decimation. It allows to retrieve a limited number of points of the total. One classical mechanism of decimation is downsampling by an integer factor called step. It consists on selecting a value every 'step' number of values. The main characteristic of classical decimation is that selected values are uniformly distributed along the total. However, in case of time evolution experiment signals, the relevancy of data is not uniform. There are some intervals where the provided information is more complex and richer, and usually more interesting from user point of view.

The contribution presents a new data decimation technology for unidimensional time evolution signals where the limited number of accessed points are distributed based on data interest level. The new method implements, on one hand, a heuristic function which is able to determine the level of interest of an interval based on its data characteristics, and on the other hand, a selection algorithm where points are distributed based on weighted intervals. The work includes detailed explanation of this new decimation method, results of its application to real experiment signals with different sampling rate and frequency, and comparatives with classic decimation method.

**Primary authors:** CASTRO ROJO, Rodrigo (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas); Dr VEGA, Jesús (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas)

**Presenter:** CASTRO ROJO, Rodrigo (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas)

**Session Classification:** Minioral

**Track Classification:** Database Techniques for Information Storage and Retrieval