

Abstract Submitted
for the DFD17 Meeting of
The American Physical Society

Algorithm for computing descriptive statistics for very large data sets and the exa-scale era IZAAK BEEKMAN, ParaTools Inc. — An algorithm for Single-point, Parallel, Online, Converging Statistics (SPOCS) is presented. It is suited for *in situ* analysis that traditionally would be relegated to post-processing, and can be used to monitor the statistical convergence and estimate the error/residual in the quantity—useful for uncertainty quantification too. Today, data may be generated at an overwhelming rate by numerical simulations and proliferating sensing apparatuses in experiments and engineering applications. Monitoring descriptive statistics in real time lets costly computations and experiments be gracefully aborted if an error has occurred, and monitoring the level of statistical convergence allows them to be run for the shortest amount of time required to obtain good results. This algorithm extends work by Pébay (Sandia Report SAND2008-6212). Pébay’s algorithms are recast into a converging delta formulation, with provably favorable properties. The mean, variance, covariances and arbitrary higher order statistical moments are computed in one pass. The algorithm is tested using Sillero, Jiménez, & Moser’s (2013, 2014) publicly available UPM high Reynolds number turbulent boundary layer data set, demonstrating numerical robustness, efficiency and other favorable properties.

Izaak Beekman
ParaTools Inc.

Date submitted: 01 Aug 2017

Electronic form version 1.4