



A proposal for a compressed format for discharge flows time series

Lorenzo Campo

CIMA Research Foundation, Savona, Italy (lorenzo.campo@cimafoundation.org)

The large amount of data in hydrological analysis and observations needs efficient ways to store time series. In literature there are several works in which the Information Theory is employed in order to build efficient compression algorithms based on the concept of entropy. In general these algorithms exploit peculiarities of the hydro-meteorological time series characteristics, like the frequent presence of long periods of null values (flows time series in small rivers/catchments), probability distribution of the values and of the intervals (rainfall time series), correlation structures in multivariate datasets (discharge flows in near streams of the same catchment). Generally, these methods assume to employ a lossless algorithm (i.e. that allows to reconstruct the whole information in the original data to less than some sort of quantization).

In this work, on the opposite, a lossy compression algorithm for discharge flow time series is proposed that aims to store only the characteristics of a set of “main features” in which the original signal can be decomposed, while the remaining signal is neglected as background noise. The main concept is that the most part of the “interesting” information in terms of hydrological analysis (especially for what concern extreme events) is retained in this compressed format, and the compressed data are informative per se.

The algorithm is tested on a series of observed time series of discharge flows, and results are discussed in terms of conversion factor and fraction of preserved information.