Compressing the Sky - an attempt to compress the CFHTLS time series

Hanno Spreeuw

February 27, 2017

Abstract

"Compressing the Sky" is a NLeSC pathfinder project, engineered by Vincent van Hees. The general idea is that, since $> 99\%$ of the sky consists of sources with constant flux (over timescales of decades, at least), it should be possible to compress databases with astronomical time series data, by applying approriate statistical tests. These tests could determine, at some appropriate confidence level, if a source is constant (null hypothesis) or not. If this is the case, all, say, 365 magnitudes with error bars from a year of nightly exposures, could be replaced by its weighted mean and a much reduced error margin. We performed this research with time series from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS). These are optical data in six different colours. From these data, we select only time series with at least 100 measurements, in all six colours.

The code is hosted on github, on a separate branch:

https://github.com/NLeSC/compressing-the-sky/tree/Attempt-to-compress-CFHTLS-timeseries

0.1 Reduced χ^2 test

Naturally, arrays of artifical random noise pass this test. However, when e.g. a single 5σ spike is added to artificial noise, the time series still pass the test when the series is long enough. That is a problem. If we require that the variability should be detected in all six colours with at least 100 time stamps, 44% of the time series are marked as constant, despite the added 10σ peak.

If we don't add any spike, but replace the artificial Gaussian noise by the CFHTLS timeseries, only 44% are marked as constant. This is a very disappointing result. Again, this is with 95% confidence for constancy in all six colours with at least 100 elements per time series.

0.2 Augmented Dickey - Fuller test

Only 5% of the CFHTLS time series with at least 100 elements in all six colours pass the test for constancy at 95% confidence.

0.3 Ljung-Box test

96% of the CFHTLS time series with at least 100 elements in all six colours pass the test for constancy at 95% confidence.

However, when a single 10σ spike is added to all time series, the same fraction passes the test. When this spike is added to five consecutive time bins, still 78% of the time series pass the test for constancy.

0.4 Jarque-Bera test

Only 8% of the CFHTLS time series with at least 100 elements in all six colours pass the test for constancy at 95% confidence.

0.5 Omnibus test for normality

Only 9% of the CFHTLS time series with at least 100 elements in all six colours pass the test for constancy at 95% confidence.

0.6 Conclusions

The CFHTLS time series are unsuitable for compression, because too few pass the test for constancy. This is probably due to incorrect error bars.