A SIGKDD 2009 paper, "Efficient Anomaly Monitoring Over Moving Object Trajectory Streams", compares to the discord definition and claims: "*similar results have been observed on "spaceshuttle" dataset, which confirm that our definition is superior*."

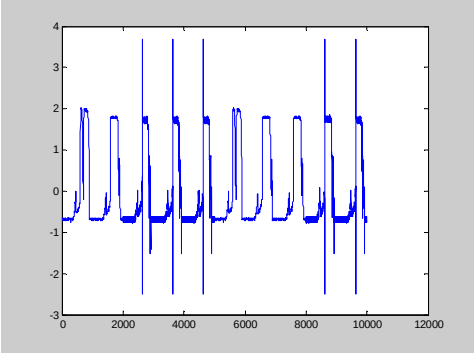
Is this really true? After all, a recent paper that tests on an order of magnitude more datasets says… "*..on 19 different publicly available data sets, comparing 9 different techniques, time series discords is the best overall technique among all techniques.*" Chandola, Cheboli, and Vipin Kumar 2009

In fact, the apparently good results are due only to the way the data was contrived. The spaceshuttle dataset is presented as being a real dataset of length 10,002. This is *not* the case. It was built from a shorter dataset, and the authors repeated a section "cut-and-pasted" fashion, without explaining why (or noting they did so). Below is a visual demonstration.

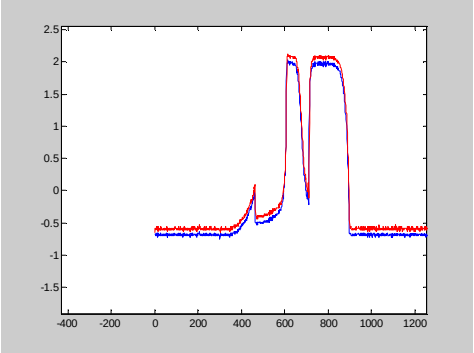
Note the following:

1. On the **real** dataset, the discord definition has PERFECT accuracy.
2. Even on this contrived version, the discord definition has PERFECT accuracy, if you simply consider the k^{th} discord version (the 'twin freak' problem [a]).

Eamonn



The *spaceshuttle dataset*



Here the section beginning at 1 (blue) is shown with section beginning at 6,600 (red, shifted in the y-axis for visual clarity)

[a] www.cs.ucr.edu/~eamonn/Shape_Discords.pdf