



Visitor rhythms

Description

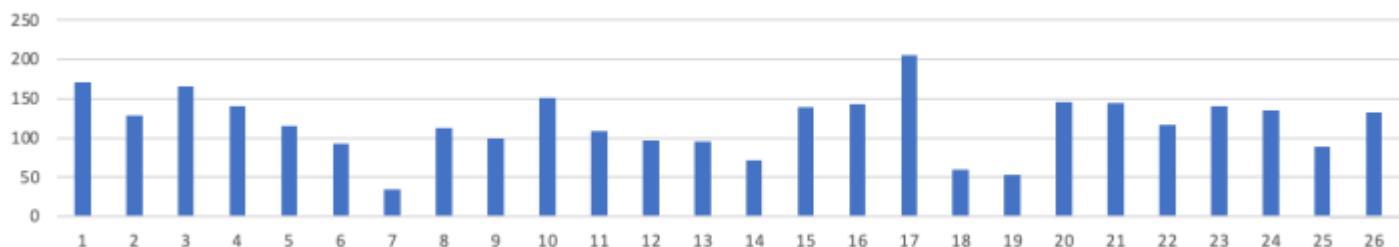
August 25, 2020 marked the tenth anniversary of this blog site. To mark the occasion I thought I would experiment with visitor statistics and their rhythms.

WordPress provides helpful visitor statistics that indicate topics that have most attracted readers or that people are most likely to stumble across. My most viewed post by far is called [Structuralism in architecture: not a style but a tool for critique](#), followed by [What's wrong with parametricism](#), and [Why cartoons have animals](#), and [The opposite of architecture](#).

To further test the [DCT \(discrete cosine transform\) method](#) as a kind of rhythm analysis I thought I would run some of the WordPress stats through a DCT process, to see if visits to particular posts follow significant cycles. I analysed the monthly visitor frequency figures for [Structuralism in architecture: not a style but a tool for critique](#) for the period Jan 2018 to May 2020, i.e. 29 months.

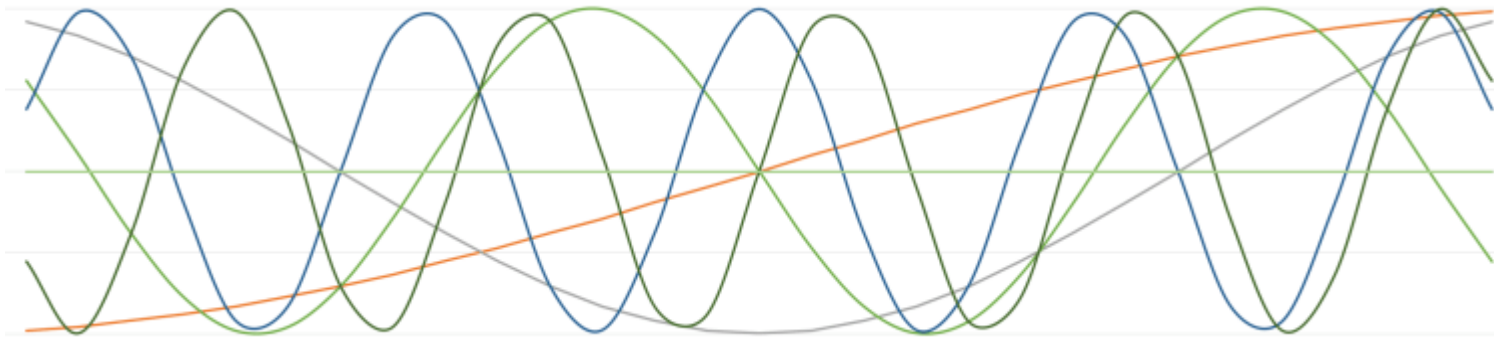
Monitoring the beat

Looking at the visitor frequency chart provided by WordPress, the data does look cyclical, perhaps under the influence of university and college assignment submission dates in different institutions in different parts of the world. I'm not going to prove that is the case just that cycles are present, or at least that the data can be filtered through a cyclical model.



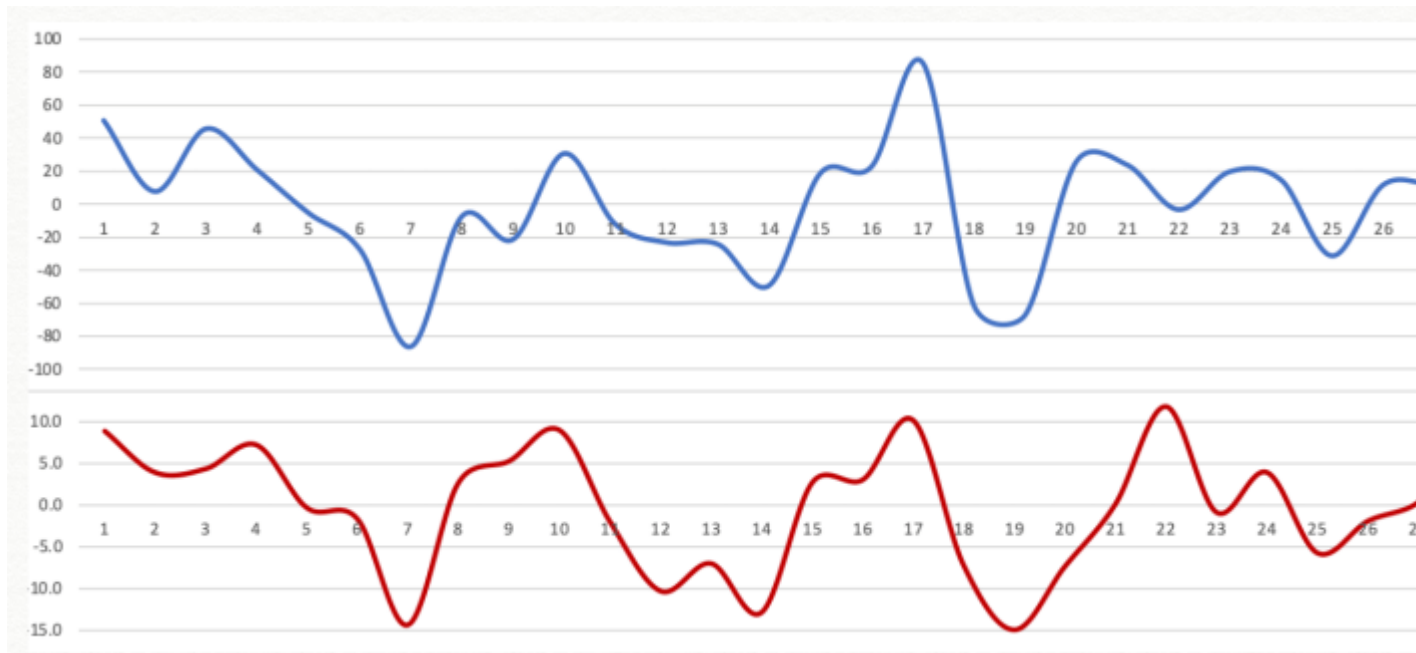
I also wanted to see how sensitive the DCT method is to shifts in the start of the temporal data sequence, whatever the source of data. In the DCT method, the data sequence is matched against a standard set of cosine curves of different frequencies (the 'basis function' curves) to produce

similarity coefficients. As shown in a previous post ([DCT](#)), these standard constituent curves are generated to be symmetrical about the midpoint of the data sequence.



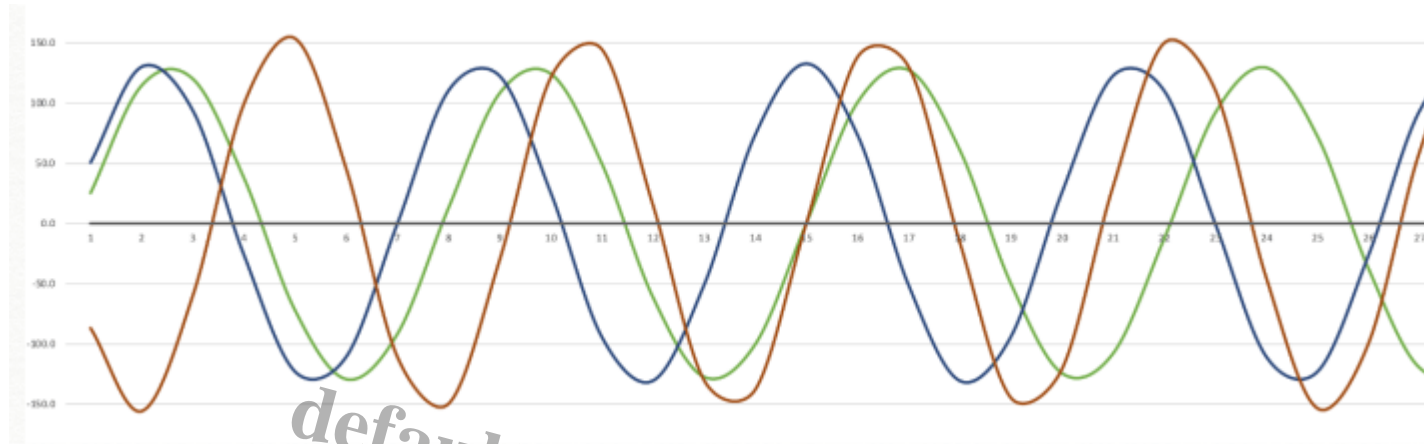
Does this bias towards the centre give different results if the data sequence is shifted right or left by one or more data points?

Here's the original blog visitor data as a smoothed curve (in blue) and moved down to the 0 axis (which is the axis along which cosine curves move). I ran the data through the DCT method as described in [COVID-19 Rhythmanalysis](#) to produce a series of coefficients. The coefficients were then run through the reverse DCT process to reconstruct the original data curve. That's the red curve below, which looks to me like a reasonable approximation of the blue curve. That verifies to me that the DCT method works. I could translate a sequence of data points to a series of coefficients indexed to some standard cosine curves and back again to recreate the original data.

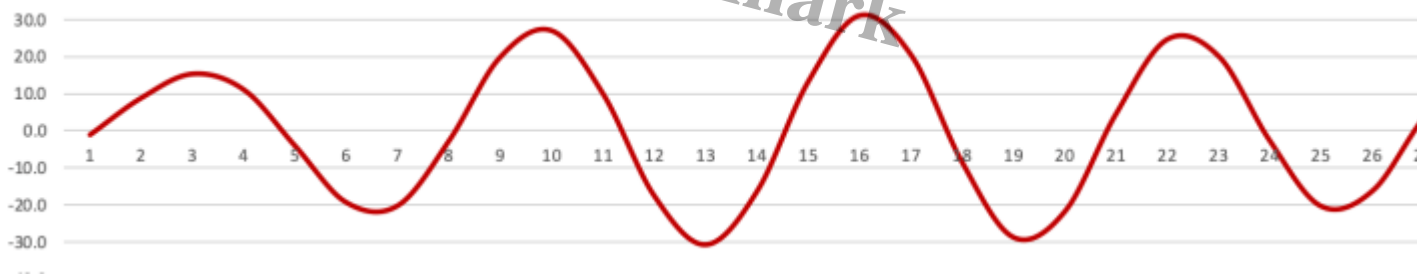


Of greater interest however is the set of constituent cosine curves and their amplitudes as identified by the procedure and that capture the cyclical nature of the data.

In this exercise the procedure revealed some high frequency cosine data that was hard to graph due to the limits of the Excel curve smoothing algorithm. I treated that as random noise and filtered it out. I also filtered out low frequency curves. They may hint of longer term cycles beyond the 29 months, but I wanted to keep the graphs simple. Here are the cosine curves that most strongly match the data when combined.

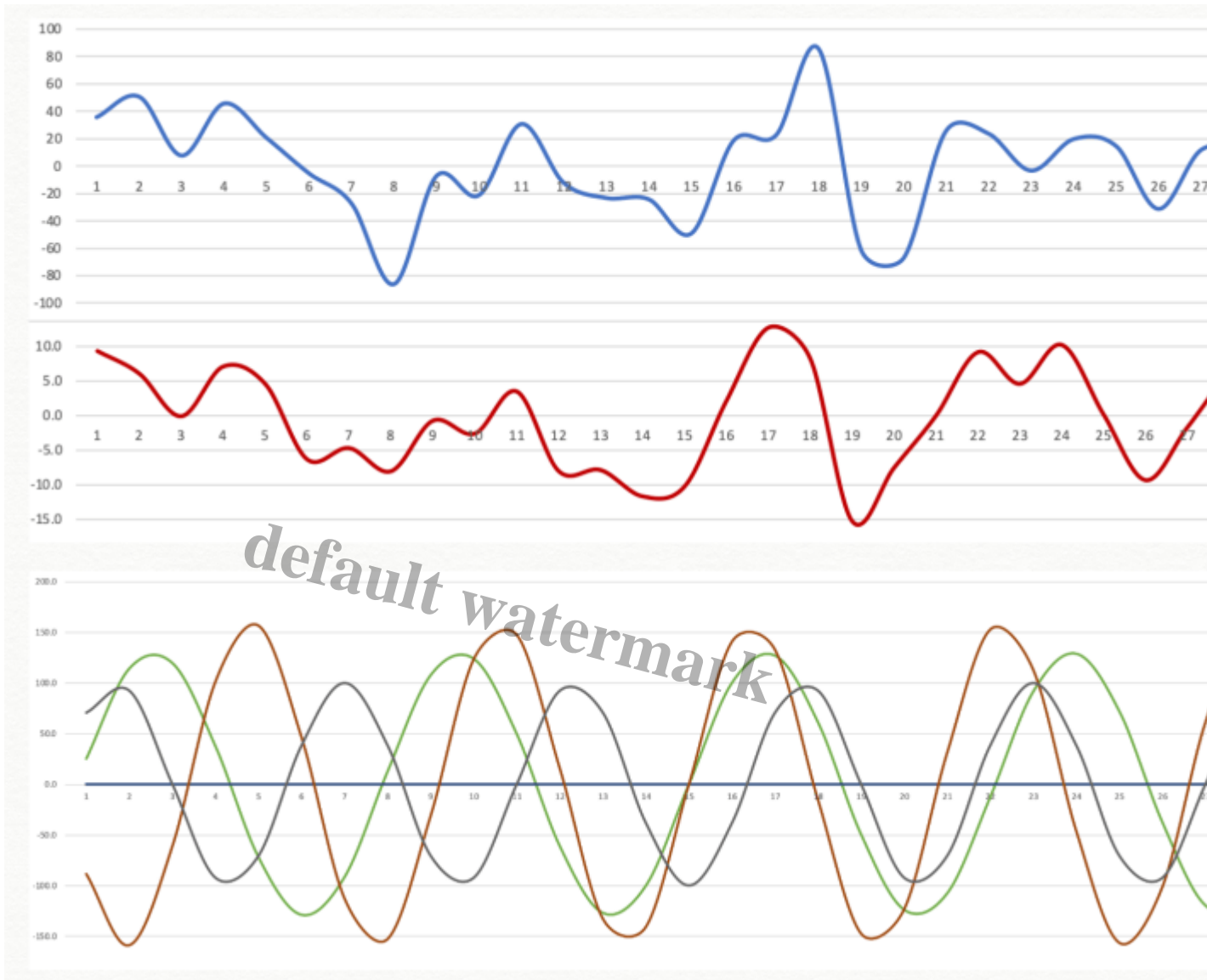


They average out to the following gross approximation of the original data curve:

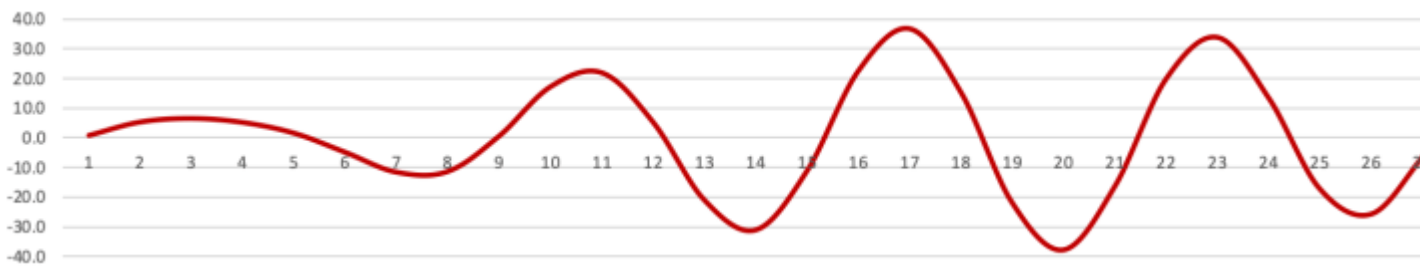


Does it matter where you start?

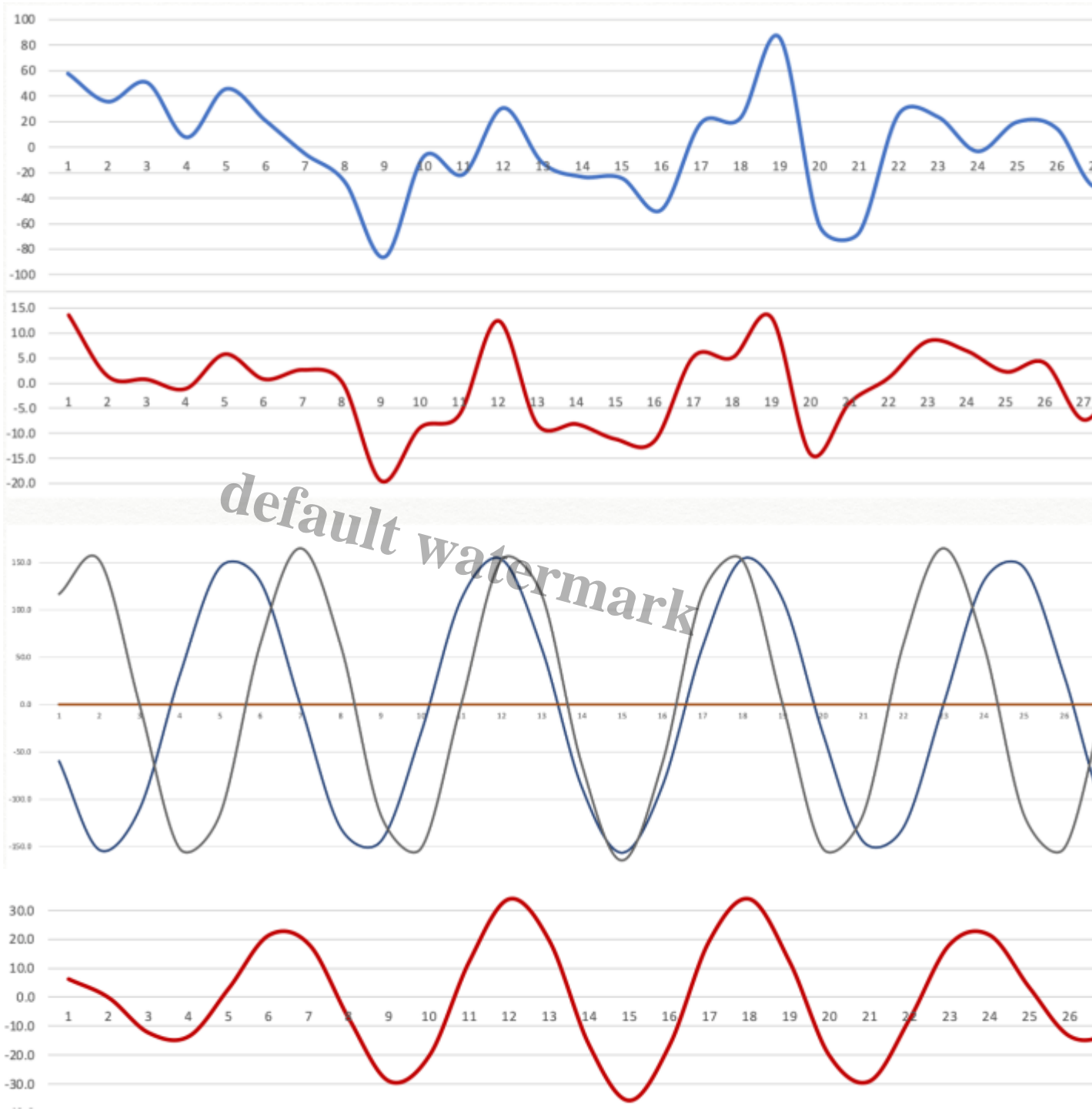
Do these curves look substantially different if I start the data sequence a month earlier? Here are the same graphs for data across 29 months, but including data from one month earlier. That's a phase shift of one month, i.e. a single data point.



These average out to something a bit different than the previous gross average graph.



Here's the same data sequence starting two months earlier instead of one month.



It seems to me that the DTC method, at least as I have implemented it, is not particularly sensitive to slight differences in where you start measuring the data in a time sequence. That's the case at least without filtering.

The differences between the gross average graphs (the smooth curves in red) could be due to (i) the high and low points in the original data being out of step with the basis functions to differing degrees, and (ii) the bias (tendency) in the method towards reflective symmetry in a data sequence.

I won't take this any further here. There's a lot of human intervention in selecting and filtering curve data. That could presumably be automated with some if-then rules. There are other methods, including the more general Discrete Fourier Transform (DFT) approach, and more statistical approaches for detecting cycles, periods, and rhythms in data. E.g. Satinder Gill alerted me recently to an interesting online book called [City Rhythms](#) that I'll look into.

Reference

- Nevejan, Caroline, Pinar Sefkatli, and Scott Cunningham. 2018. *City Rhythm*. Delft: TU <http://www.nevejan.org/caroline-nevejan/2018/4/17/city-rhythm-logbook-of-an-exploration>



Category

1. Media

Date Created

September 5, 2020

Author

rcoyne99