Innate Cycle Correlation

R G M Crockett

Data Analysis Group, School of Science and Technology, University of Northampton, NN2 6JD, UK.

Introduction

Innate Cycle Correlation (ICC) is a generalisation of the Base Number Correlation technique previously developed for integer data containing integer-periodic features, such as count / enumeration data (Crockett et al., 2004).

It is a novel technique, under development, for investigating and quantifying periodic features in time-series and other sequential data-sets.

At an intuitive and straightforward level, ICC is capable of generating a spectrum of periodic components directly comparable to the familiar spectrum of frequency components generated by the Discrete/Fast Fourier Transform. At more advanced levels, it can be used to enhance information in DFT/ FFT spectra and investigate phase-domain clustering of cyclic and sub-cyclic features.

Hypothesis

If a data-set contains an innate cyclic behaviour / feature then mapping its index from arbitrary units to cyclic-period units and aggregating according to phase will result in an aggregate cycle showing reinforcement of the cyclic features.

Conversely, if there is no cyclic feature at a given period, then mapping its index and aggregating according to phase around that period will not result in reinforcement and the aggregate cycle will be uniform and featureless.

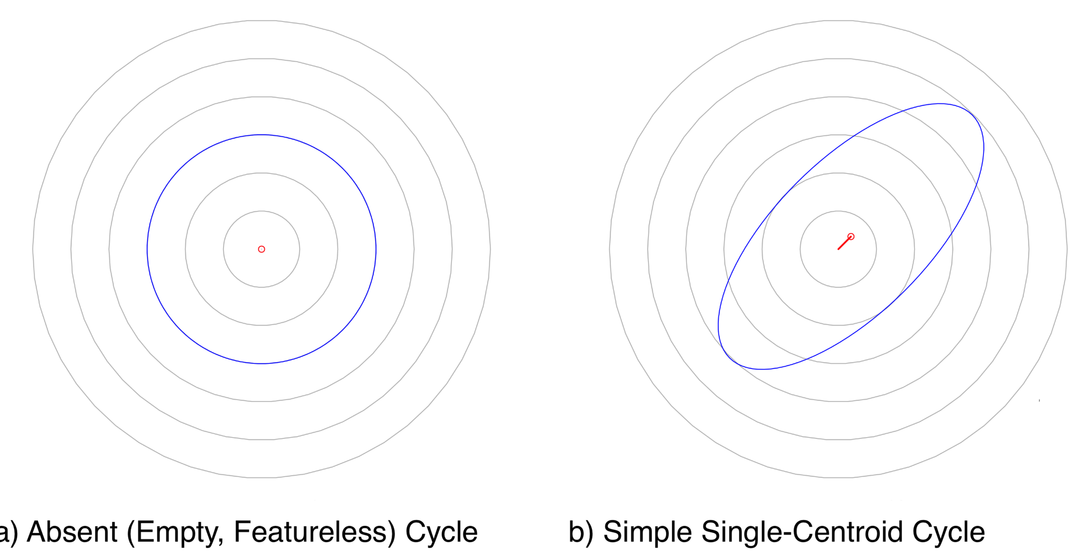

Fig 1. Simple Illustrative Clusters (blue) and Centroids (red).

Outline Method

1. The time-index is mapped onto a phase-index for the period under investigation, i.e. scaling from arbitrary time units to period units.

2. The resulting phase-magnitude pairs can be polar-plotted (modulo-2π), explicitly or implicitly.

3. The aggregate cycle can be either simple, i.e. one cyclic feature (e.g. single maximum, minimum), or complex, i.e. many cyclic features (e.g. multiple maxima, minima).

4. Cluster centroids can be calculated according to the individual cyclic features and the overall centroid can be calculated. The angular position of a centroid indicates the phase with respect to the start of the time-series.

Simple and Complex Cycles

In the simplest case, simple or no cycle, the phase-magnitude pairs will form a single cluster with:

i) the centroid at (an insignificant distance from) the origin, in the absence of an innate cycle in the data;

ii) the centroid at a significant distance from the origin, in the presence of an innate cycle in the data.

In complex cycles, it is possible for there to be both:

i) a distinct sub-cyclic structure with more than one cluster-centroid;

ii) the centroid of all the phase-magnitude pairs to be an insignificant distance from the origin.

In his 1897 paper which set-out the "Schuster Test", Schuster did not consider complex cycles: he only considered isolated simple cycles using an equivalent vector formalisation (using a random-walk Rayleigh probability model to calculate the significance of the resultant vector). Thus, the "Schuster Test" can readily yield false negatives for non-simple behaviour.

Amplitude Spectrum

Assuming that there are only simple cycles in the data, ICC can be used to generate an amplitude spectrum (of centroid positions / distances) comparable to a DFT/FFT spectrum, see Figure 2 (data from Crockett et al., 2010). The significance of any cycles revealed in the data can be assessed with respect to the baseline noise in the amplitude spectrum.

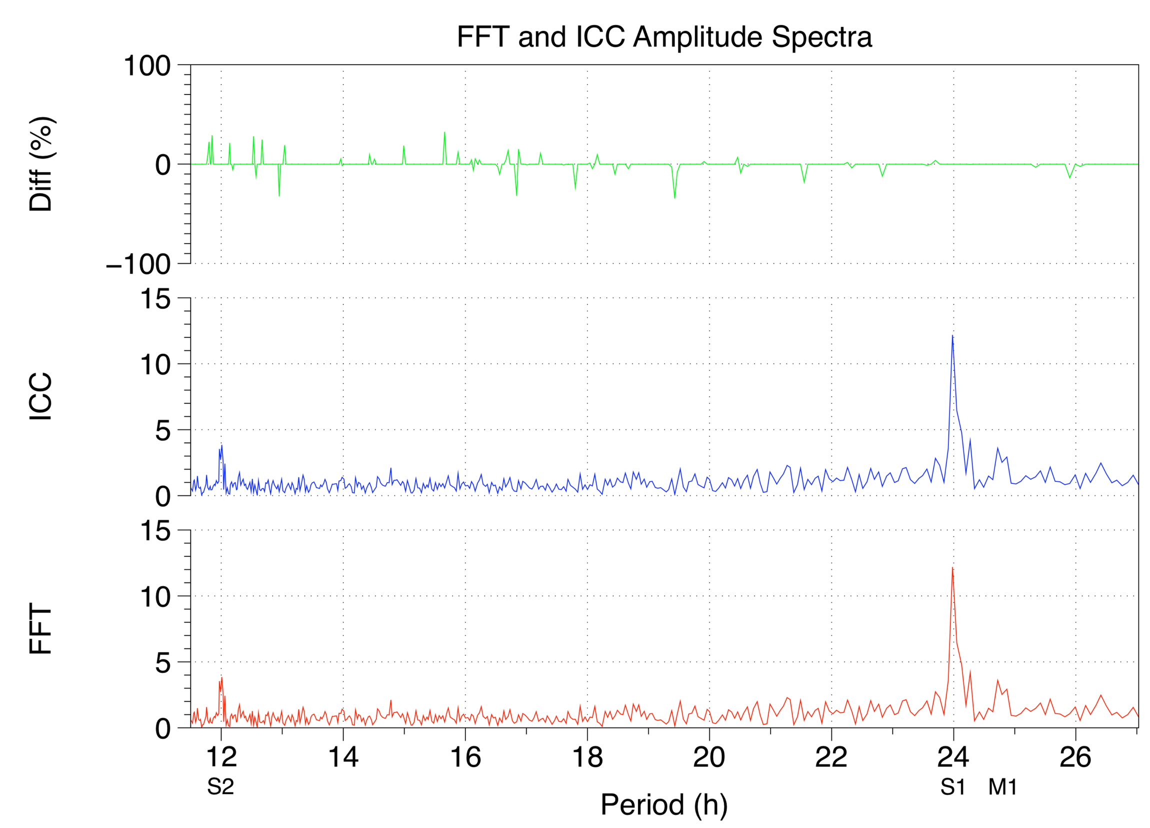

Fig 2. FFT and ICC Spectra Compared (Radon Data).

Note that the spectra are normalised to unit mean across the period-range to indicate similar relative sensitivities.

In the data investigated at this preliminary stage, the magnitudes of the amplitude spectrum are closely Gumbel-distributed, allowing an assessment of the significance of components using a Standard-Gumbel *z*-statistic (analogous to the Standard-Normal *z*-statistic for normally-distributed data).

This is shown in Figure 3, for the same data as Figure 2, using non-FFT sample intervals as described below.

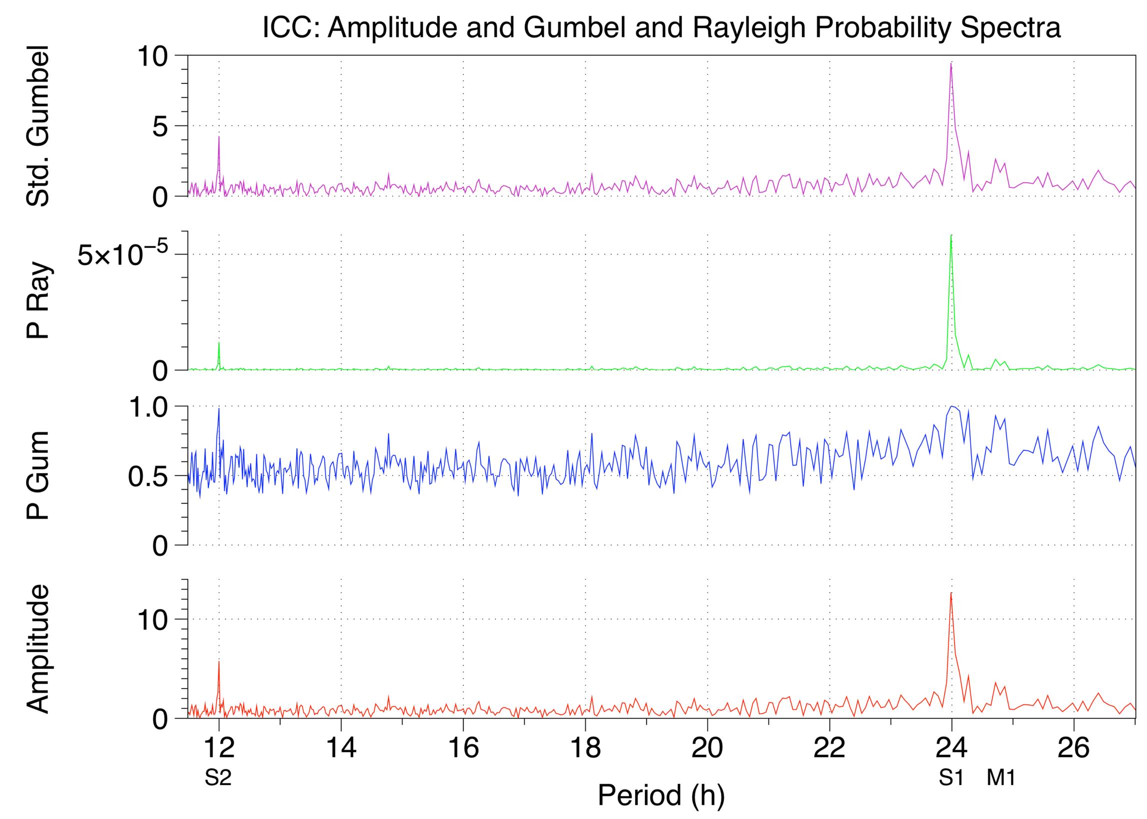

Fig 3. ICC Amplitude and Probability Spectra Compared.

Spectra: Period (Frequency) Intervals

Given an innate cycle, period T_i, in a time-series duration T, such that $n_i T_i = T$

we can sample at T_s and still resolve the T_i-cycle provided

$n_i - 1 < n_s < n_i + 1\,; \quad n_s T_s = T$

Sampling outside these bounds means that T_i will not be resolved. Thus, for an initial sample period, T_1, the maximum interval, ΔT, to the next sample period, T_2, so as to be able to resolve T_i for $T_1 < T_i < T_2$

is given by:

$$\Delta T = T_2 - T_1\,; \quad n_1 T_1 = n_2 T_2 = T\,; \quad n_2 = n_1 - 1$$

$$\Delta T = \frac{T_1^2}{T - T_1} = \frac{T}{n_1(n_1 - 1)} = \frac{T}{n_2(n_2 + 1)}$$

Note that:

i) shorter intervals can be used

ii) specific periods can be quantified

iii) an ICC spectrum can be calculated for periods (frequencies) as defined by DFT/FFT algorithms (as is the case in Figure 2).

Variable Sample Intervals

ICC can be used to produce period / frequency spectra from variable-interval time-series, such as earthquake catalogues.

This is shown in Figure 4, for Sumatran earthquake data (Crockett et al., 2006). The 'blur' in the features is due, in part, 'jitter' of major earthquakes around bi-weekly tidal maxima and the 'aftershock lag', and also to the variable data-intervals.

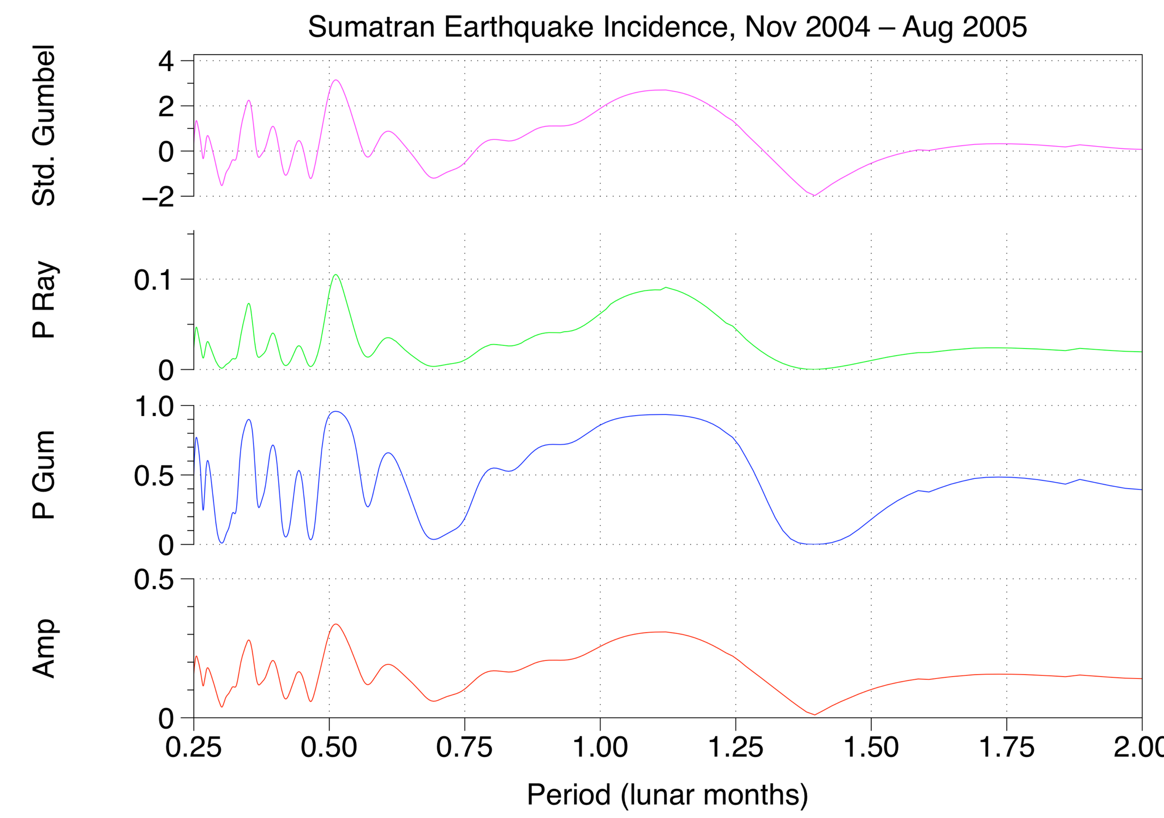

Fig 4. ICC Spectra of Variable-Interval Earthquake Data.

Conclusions

Used in simple-cycle spectrum mode, subject to statistical considerations, ICC has advantages over DFT/FFT:

i) The time (sequential) index does not have to be uniform throughout the data, i.e. the sample interval does not have to be constant, and the data can include interruptions which do not have to be interpolated.

ii) It is possible to investigate user-specified periods, whereas frequencies in a DFT/FFT spectrum are dictated by the number of data-points and duration of the (constant) sample interval.

iii) It is possible to 'seed' the spectrum with random periods as well as user-specified periods, thereby enabling statistical evaluation of significance of specified periods – a bootstrap-type approach assuming all periods are present in the data.

However, ICC is susceptible to harmonics.

Used in multi-centroid mode, with implied cluster analysis, ICC is able to resolve complex periodic behaviour.

To the memory of A C Crockett, 1968-2006.

References

1. Crockett R G M, Crockett A C, Turner S J; 'Base-Number Correlation': a new technique for investigating digit preference and data heaping. History and Computing, 13, 2, (2001) 161-179. 2004.
2. Schuster A; On lunar and solar periodicities of earthquakes, Proc. R. Soc. London, 61, 455– 465. 1897.
3. Crockett R G M, Perrier F, Richon P; Spectral-decomposition techniques for the identification of periodic and anomalous phenomena in radon time-series. NHESS, 10, 559–564, 2010.
4. Crockett R G M, Gillmore G K, Phillips P S, Gilbertson D D. Tidal Synchronicity of the 26 December 2004 Sumatran Earthquake and its Aftershocks. Geophys. Res. Lett., 33, 2006.