ChaoticModelingandSimulation(CMSIM) 1:11-19, 2014

Towards a Novel Method to Distinguish Random from Under-Sampled Light Curves

N. Jevtic[a,b], S. Grib[a], M. Ashton[a], J.S. Schweitzer[b], P. Stine[a]

[a] Bloomsburg University, 400 Second St, Bloomsburg, PA 17815, USA
(Email: njevtic@bloomu.edu)
[b] Physics Department, University of Connecticut, Storrs, CT 06269, USA

Abstract: A novel application of nonlinear time series analysis was developed thanks to the high-quality almost-continuous *Kepler Space Telescope* data. It makes it possible to distinguish truly random from under-sampled signals and to ascertain whether the selected sampling time is short enough to access all the complexity of the light curve using nonlinear time series analysis. This is true irrespective of whether the data are acquired in space or from the ground and the methodology is independent of the details of the data acquisition.
Keywords: Nonlinear time series analysis, Phase-space portrait, Average mutual information, *Kepler Space Telescope*, Variable stars

1 Introduction

Gravity reigns in our Universe. And even in its Newtonian form, gravity is a nonlinear force. So one would expect astronomy to be the playground of the nonlinear time series analysis practitioner. (For a more detailed introduction to the methodology see Jevtic et al. (2005).) However, for nonlinear time series analysis, the requirements on the data are very stringent [7]:

a) The observable has to couple all the active degrees of freedom. Thus energy, power and hence, in astronomy, brightness and luminosity are "good" variables.
b) Data must be sampled uniformly.
c) Data must be continuous.
d) Data sets have to be as long as possible since the longer the data set the more efficient the methodology.
e) The data should be finely digitized.
f) Data should yield access to a large dynamic range.
g) There should be as little additive noise as possible.
h) The nature of the process should not change during observation.

As luck would have it brightness, the observable of choice in astronomy, is a "good" observable (a). Requirement (b) (with the exception of valiant efforts such as the Whole Earth Telescope) and requirements (c) and (d) are almost

Received: 17 September 2013 / Accepted: 10 January 2014

ISSN 2241-0503

never satisfied for astronomical data collected from the ground. It has been possible to satisfy requirements (e) and (f) since the advent of CCDs which do single-photon counting and can accommodate large dynamic ranges with no difficulty. The issue of noise (g) is a complex one, the distinction between dynamic and additive noise being a whole subfield. Dynamic noise can span the whole range of frequencies present and its identification is crucial to understanding such processes as granulation. The last requirement, (h), is completely out of our control. At best, all we can hope for is that the nature of the process does not change over the observation time.

2 The *Kepler Space Telescope* as a source of variable-star data

Launched in 2009, The *Kepler Space Telescope* was designed to detect extrasolar Earth-like planets using the transit method. The telescope has a modest 1.4 meter diameter mirror and 21 CCD modules with 2200x1024 pixel resolution. It is in a heliocentric Earth-trailing orbit with a period of 372.5 days that is slightly more eccentric (ε=0.03188) than that of the Earth. With *Kepler's* 105 square-degree field-of-view (FOV), it is able to continuously monitor the brightness of nearly 200,000 stars [3,12] in our spiral arm of the Milky Way Galaxy in the direction of the constellation Cygnus with a search area extending about 3,000 light years and covering ~0.28 % of the sky. Though designed primarily for observing 9^{th} to 16^{th} magnitude stars, it can also collect data on stars outside this range with a bandpass of 430-890 nm. The photometric precision for the telescope is ~50 parts per million for a Kp = 12 magnitude G2V star when integrating for 30 minutes.

Kepler's observations are subdivided into quarters. Two binning modes are available: "long cadence" (LC) with 28.5 min ("30 min") bins and "short cadence" (SC) with 58.85 s ("one-minute") bins [13].

Because the *Kepler Space Telescope* gives long, continuous data sets of large S/N for a wide range of variable stars processed in the same manner it is easy to make comparisons. *Kepler Space Telescope* data have been instrumental in our work on nonlinear time series analysis and were used to analyze the NGC 6826 light curve in Jevtic at al. (2012) and (2011).

3 Time-Delay Portrait and Reconstruction Parameters

A standard method to explore nonlinear systems is a time-delay phase-space reconstruction. According to the embedding theorems [16, 17, 18] the geometry of this time-delay phase-space portrait is diffeomorphic to the geometry of the phase-space representation that would be obtained if the equations governing the system were known. This object can serve as a surrogate for the system. To obtain a time-delay phase-space portrait of a uniformly-sampled time series, the phase-space dimension and the time delay are first determined from the data.

3.1 Dimension of the Phase Space - False Nearest Neighbours

The dimension of the reconstructed phase space is obtained using the False Nearest Neighbor (FNN) method [11]. This dimension is important because it is related to the number of degrees of freedom needed to model the system.

3. 2 Time Delay - Average Mutual Information

The optimal time delay ensures the greatest possible independence of coordinates in phase-space. It is chosen at a minimum of average mutual information [5]. Average mutual information (AMI) is the information-theory analog to the autocorrelation function that is more general in the statistical sense and represents the expectation of the average degree of interdependence incorporating all higher orders. It is the amount of information (in bits) shared by the signal and its time shifted value averaged over the orbit.

For a uniformly sampled time series such as a stellar light curve AMI is defined as:

$$\mathrm{AMI}(\tau) = \sum_{s(t),s(t+\tau)} P(s(t), s(t+\tau)) \log_2 \left[\frac{P(s(t), s(t+\tau))}{P(s(t))P(s(t+\tau))} \right] \tag{1}$$

Here “s” is the sampled scalar time series and τ is the time delay. The range of the time series is divided into m sub-intervals and a histogram is obtained that yields the probability p_i for a point to be in the interval i, the probability p_j for a point to be in the interval j, and the joint probability $p_{i,j}$ that if s_k is in interval i then $s_{k+\tau}$ is in interval j. Whether the first, local, minimum or the global minimum best determine the optimal delay is still open to discussion.

The optimal time delay for time-delay embeddings is at the minimum of AMI and ensures the greatest possible independence of axes in phase space. This ensures the greatest possible ”unfolding” i.e. the greatest amount of information about the system. However, AMI’s probabilistic nature gives it even wider applicability [4]. One such application is the use of AMI to preview the power spectrum under the noise as a guide for nonlinear noise reduction, This was discussed in Jevtic et al (2011). We shall focus on a novel application unique to astronomy: estimating whether the sampling time is short enough to distinguish random from under-sampled signals.

4 Distinguishing Truly Random from Under-Sampled Signals

Space telescopes such as *Kepler* and Corot and the future TESS provide us with well-sampled almost noise-free light curves. However, for the foreseeable future, we sill continue to combine data from space and ground-based telescopes. It is not uncommon that the question arises whether the sampling rate with which we are observing variable targets is short enough to sample all the complexity of a process. Though an inspection of the light curve can often answer this question for stars whose light curves do not suffer from significant noise, a more general answer based on nonlinear dynamics may be obtained using phase-space reconstruction.

Traditionally, time is not an explicit variable in phase-space. However, time on the order of the sampling time can be estimated [10] using nonlinear time series analysis. Initially nonlinear time series analysis developed in the search for chaotic systems where the focus was on finding dynamically related phenomena and avoiding time-correlations. This resulted in the prescription to select the time delay at the minimum of AMI and the injunction to avoid delay times that are too short for which time correlations dominate. The objective is to obtain an unfolded phase-space portrait so as to maximize the amount of information obtained. In this context "unfolded" refers to a portrait that fills the largest fraction of phase space possible. If instead we focus on the shortest delays on the order of the sampling time, we obtain a tool to distinguish truly random from under-sampled signals. If a delay of the shortest available sampling time results in a phase-space portrait that is "unfolded" [7, 15] i.e. fills a significant fraction of the phase space available, the sampling time is too long. If on the other hand, the points in phase space fall along one of the diagonals, the sampling time is short enough to properly sample the process.

The two available sampling rates, long and short cadence available for the *Kepler Space Telescope* data have allowed us to explore and develop a tool that may be used to determine if the sampling time is short enough to sample the dynamics properly.

5 Under-sampled vs. Optimally Sampled Light Curves

5.1 The Central Star of Planetary Nebula NGC 6826

The central star of planetary nebula NGC 6826 [10, 6] (KIC 12071221) was observed by the *Kepler Space Telescope* in both long and short sequence. The phase-space portrait for one month of long-cadence data with nonlinear noise reduction for a delay of one is shown in Fig. 1.a. The 3D phase-space portrait is unfolded indicating that the sampling time of ~30 min is too long. The time-delay phase space portrait of a month of short-cadence data with a delay of one sampling time is shown in Fig. 1.b. The points lie along a diagonal, indicating that the short-cadence sampling rate is short enough to capture all the dynamics.

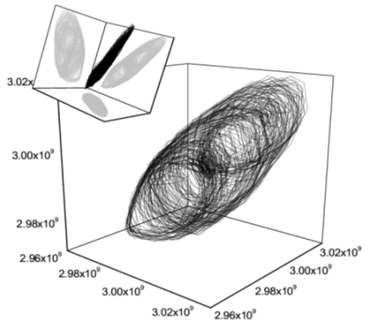

Fig. 1.a Unfolded phase-space portrait of a month of nonlinear noise reduced long-cadence data with a delay of one sampling time

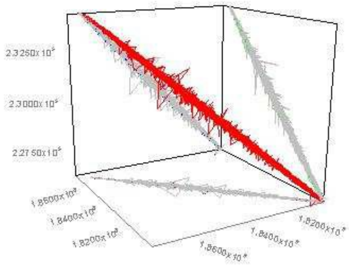

Fig. 1.b Time-delay phase space portrait of a month of short-cadence data with a delay of one sampling time

The difference between the long and short cadence light curves, due to contributions at the higher frequencies, is observable due to the higher resolution of the latter. For the range of frequencies accessible for the long-cadence data, the power spectra are comparable for the two sampling times.

5. 2 A δ-Scuti Star in a Triple-Star System

A starker example is that of KIC 4840675, a triple system with a rapidly-rotating A-type δ-Scuti variable and two solar-type fainter companions [2].

A section of the light curve for the long-cadence data is shown in Fig. 2.a. The fact that the sampling time is too long can easily be observed. The three-dimensional portrait of this data is shown in Fig. 2.b. where the phase-space portrait looks unfolded.

A section of the light curve for the short-cadence data during the same month is shown in Fig. 3.a. The three-dimensional portrait of this data is shown in Fig. 3.b. where all the points are clustered around the diagonal. When Fig.2 and 3 are

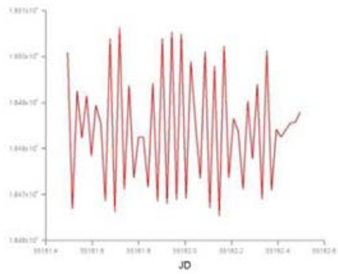

Fig. 2.a Section of long-cadence light curve

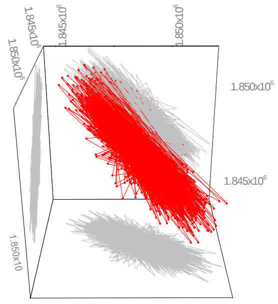

Fig. 2.b Time-delay phase space portrait of a month of long-cadence data with a delay of one sampling time (~30 min).

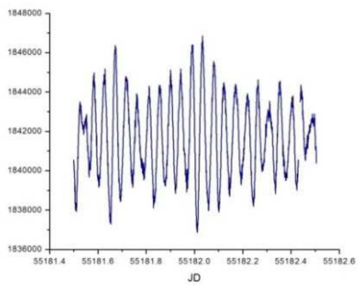

Fig. 3.a Section of short-cadence light curve

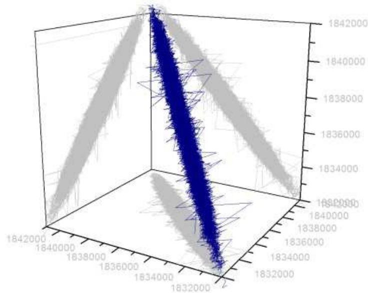

Fig. 3.b Time-delay phase space portrait of a month of short-cadence data with a delay of one sampling time

compared, it is easy to observe that in the short-cadence data the sampling time is sufficiently short to access the variations on all the time scales.

5. 3 The Only *Kepler* White Dwarf – KIC 8626021

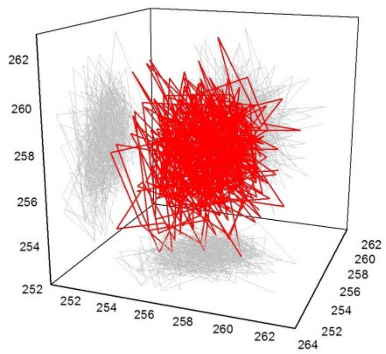

Fig.4.a Time-delay phase space portrait of a month of long-cadence data with a delay of one sampling time (~30 min).

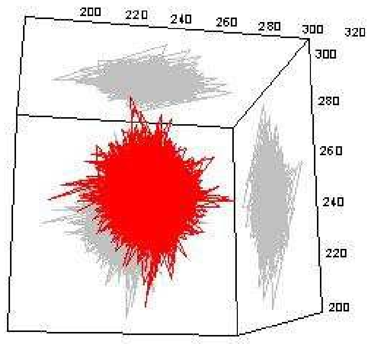

Fig. 4.b Time-delay phase space portrait of a month of noise reduced short-cadence data with a delay of one sampling time (~ 1min)

For comparison, we also look at the magnitude Kp=18 KIC 8626021, the only identified pulsating white dwarf in the *Kepler Space Telescope* [14] field. Pulsating, white dwarfs can pulsate with frequencies on the order of 1 min. This is the case for KIC 8626021 that has a dominant period of 3.2 minutes. The three-dimensional portraits of the long and short cadence data are shown in Fig. 4.a and Fig. 4.b, respectively. However, here for the short cadence data the points do not line up along the diagonal. No structure can be observed either. One possible interpretation is that even the short cadence time is still too long to properly access information. However, this target has a magnitude of 18 so a significant noise contribution can't be ruled out.

6 From the Ground: Observations of PG1351+489

As a test case we look at Whole Earth Telescope XCov12 observations of PG1351+489 [1] with two binning times of 30s (WET) and 10s (A. Kaanan).

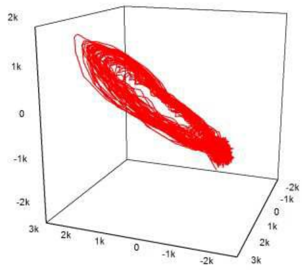

Fig. 5.a Phase-space portrait for a delay of one of the 10s binning with noise reduction (pca eigenvalues 1:27:370)

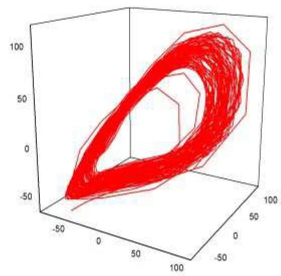

Fig. 5.b Phase-space portrait for a delay of one of the 30s binning with noise reduction (pca eigenvalues 1:4:106).

The phase-space portraits for a delay of one in three dimensions for 10s and 30s binning with noise reduction are shown in Fig. 5.a and 5.b, respectively. These light curves were obtained at different observatories and normalized differently requiring a principle component decomposition for the comparison. For the 10s binned data the eigenvalues in respect to the largest component are 1:27:370. For the 30s binned data the eigenvalues in respect to the largest component are 1:4:106 indicating that the 10s binning may still be too long.

7 Conclusions

When data are collected from a controlled experiment in the laboratory, it is easy to choose the optimal sampling time. When observing stellar objects, we *a priori* do not know down to what time scale is needed to access all the information about the source. The *Kepler Space Telescope* light curves have allowed us to explore this problem. On an example of ground-based data obtained at different observatories with different processing we show that time-delay reconstruction can be used to reliably estimate when sampling time is short enough to capture all the complexity of a signal. Since it is based on the dynamics at the source, this approach will particularly be useful for longer-term periodic targets observed from the ground.

References

1. V. M. Alves, et al. 2003, Baltic Astron., 12, 33
2. L. Balona et al., 2012, MNRAS, 424, 1187B
3. W.J. Borucki, et al,2010, Sci, 327, 977B
4. J. Chen, et al., Geophys Res, VOL. 113, A05217
5. A.M. Fraser, H.L. Swinney, Phys.Rev. A, 33, 2, 1986, pp.1134-1140
6. G. Handler et al. 2013, MNRAS,.430, 2923H

7. R. Regger, H. Kantz and T, Schreiber, CHAOS, 9, 413, (1999)
8. N. Jevtic et al., ApJ, 635:527-539, 2005
9. N. Jevtic et al. CMSIM, Oct 2011, pp 187-197
10. N. Jevtic et al, 2012, ApJ, 756:9, 2012
11. M. B. Kennel et al, Phys.Rev. A 45, 3403 (1992)
12. D.G. Koch et al., ApJL 713, L79, 2010
13. S.J. Murphy et al., 2013, MNRAS, 430, 2986M
14. R. Ostensen et al., 2011, ApJL, 736, L39
15. M. Small Applied Nonlinear Time Series Analysis, World Scientific, 2005
16. J. Stark, JNonlinSci, 9, 255-332, 1999
17. J. Stark et al. JNonlinSci , 13, 519, 2003.
18 Takens-Mane, Detecting Strange Attractors in Turbulence, Dynamical Systems and Turbulence, 898, Lecture Notes in Mathematics (Springer-Verlag, 1981)

Acknowledgements

We thank the entire Kepler team for the development and operations of this outstanding mission and all involved in KASC for the handling of the data. Their efforts have allowed us to conduct the theoretical work on nonlinear time series methodology presented herein. Funding for the Kepler Mission is provided by NASA's Science Mission Directorate.

We also wish to thank:

Steve Kawaler and Antonio Kanaan of the WET for the data on PG 1351.

The Bloomsburg University Faculty Travel Fund and the Bloomsburg University Foundation.

The Max Planck Institute for Complex Systems for the TISEAN software:
R. Hegger, H. Kantz, and T. Schreiber, Practical implementation of nonlinear time series methods: The TISEAN package, CHAOS, 9, 413 (1999)