Clustering of snapping shrimp snaps on long time scales: a simulation study

M.W. Legg and M.A. Chitre

Acoustic Research Laboratory, Tropical Marine Science Institute

National University of Singapore, Singapore

ABSTRACT

Fields of snapping shrimp produce a great number of impulsive snapping sounds throughout the world's tropical and sub-tropical shallow water regions. Snaps from snapping shrimp appear as short spikes in timeseries obtained from hydrophone measurements. Clustering of the spikes in the timeseries can occur on at least three different time scales corresponding to three spatio-temporal mechanisms. The mechanism for clustering on short time scales is multipath propagation, because a single snap from a shrimp evolves into a cluster of snap signals at the hydrophone. Mechanisms for clustering on medium and long time scales remain intriguing and inconclusive. This paper investigates diurnal snap-rate variations and their effect on clustering over long time scales. Simulation of spike timeseries made use of the Cox-Ingersoll-Ross driven doubly stochastic Poisson process with diurnal variation introduced by varying one model parameter. Fano-factor analysis of the simulation results showed that diurnal variations have a profound effect on clustering of snapping shrimp snaps on long time scales.

INTRODUCTION

Clustering of the snaps from snapping shrimp remains an intriguing and unresolved problem in underwater acoustics. Snapping shrimp noise can have a significant impact on sonar and underwater acoustic applications, so understanding when and why these snaps are likely to occur is important. Sonar and underwater telemetry systems can make use of unique perspectives and greater understanding of shrimp noise to improve signal processing (Chitre et al. 2006, Mallawaarachchi et al. 2008) while other applications make use of the shrimp noise for ambient noise imaging (Buckingham et al. 1992, Kuselan et al. 2011) and acoustic inversion (Chitre 2010).

Clustering in snapping shrimp noise has long been observed (Johnson et al. 1947, Cato & Bell 1992) and possible mechanisms that explain the cause of clustering have been proposed including increased activity due to foraging and feeding (Johnson et al. 1947), response to tactile stimulus (Schmitz & Herberholtz 1998) and territorial defence (Toth & Duffy 2005). There has been speculation that the shrimp may be communicating with each other, although it has been difficult to prove. Cato & Bell (1992) report that higher level pulses seem to occur in bursts and suggest that shrimp may snap in response to other nearby snaps to enhance survival in the presence of an aggressor. Potter & Koay (2000) investigated spatial clustering and temporal chorus and found spatial anisotropy of ±3 dB but no evidence for temporal chorus. Legg (2010) analysed the snaps received at a single hydrophone as a counting process and found temporal clustering on at least three different time scales. Short time clustering (less than a second) was caused by multipath propagation. A Cox-Ingersoll-Ross driven doubly stochastic Poisson process (CIR driven DSPP) was used to model medium time clustering (one second to two hundred seconds) but was not linked to any causal mechanism. Long-time clustering (greater than two hundred seconds) was discovered while conducting an experiment to investigate an asymptote predicted by the CIR driven DSPP model.

One very important question regarding the clustering of snaps on long time scales is if the temporal point process has a fractal nature. Fractal characteristics can be identified through a power law increase in the Fano-factor time curve for long counting times (Teich & Lowen, 1994).

This study investigates diurnal variation as a candidate mechanism for clustering on long time scales. There are three questions that require answers:

- Does diurnal variation in snapping have any effect on the Fano-factor time curve?
- What characteristics might we expect from a diurnal varying process (does it have a fractal nature)?
- How long does real shrimp noise need to be observed to allow these features to be recognised?

These three questions are particularly important for the motivation and design of future experiments, observation studies and potential signal processing methods. Motivations for measurements over long periods and analysis of very large data sets need to be strong since these activities can be logistically difficult and computationally expensive. Distinguishing any fractal features in a diurnal varying process is of fundamental importance for signal processing applications.

The paper is organised in the following way: a theory section provides a brief review of diurnal variation of snapping shrimp noise then introduces a point process model and the Fano-factor time curve analysis technique. Theory is followed by a simulation section that details the methods, codes and model parameters used to provide a set of simulated data. The simulated data is subjected to analysis and the results presented, followed by a short discussion. Conclusions are drawn from the simulation study to answer the three questions that have been raised.

THEORY

Diurnal variations

Diurnal variations in the noise levels produced by fields of snapping shrimp are well known. Johnson et al. (1947) report a small diurnal variation with night time levels 2 dB to 5 dB higher than during the day and a slight peak shortly after sunset and before sunrise. They speculate that the increased noise levels are due to increased activity of shrimp at these times. In a related publication Everest et al. (1948) present diurnal change in wide band shrimp noise from four different locations: San Diego yacht harbour, Scripps Institution of Oceanography, Kaneohe Bay and Midway Island. They also report higher noise levels at night (by as much as 6 dB) and peaks after sunset and before sunrise. Radford et al. (2008) report significant diurnal variation in shrimp noise from observations at a shallow water reef in north-eastern New Zealand.

Cato & Bell (1992) report a contrasting result from the Timor Sea where overall variations over very long periods were about 10 dB but no consistent diurnal variation in noise levels were observed. Lammers et al. (2008) also report slightly different results from observations in Kaneohe Bay, with diurnal peaks centred on sunset and sunrise and the daytime noise levels consistently *higher* than night time levels. Subsequent observations from a Waikiki Marine Life Conservation District report (the more frequently observed) lower daytime levels than night levels and sharp increases at sunrise and sunset.

In Legg (2010) the snapping rate (rather than noise level) was observed over a 24 hour period from a wharf near Rockingham in Western Australia. The snapping rate followed a similar pattern to noise levels with higher snap rate at night and peaks just after sunset and just before sunrise. The template diurnal function used in this paper was based on a piecewise approximation of the snap rate trend observed at that site.

The CIR driven DSPP model

If shrimp snaps occurred purely at random (in time) without any influence, or interactions such as response to other shrimp, then the snap event times would be consistent with a temporal homogeneous Poisson process with fixed rate λ independent of time (Cox & Lewis 1966). However, the snaps from snapping shrimp are not consistent with a homogeneous Poisson process; there are influences that affect when the shrimp snap (Cato & Bell 1992, Tóth & Duffy 2005).

A Poisson process that is influenced by a time dependent mechanism will have a rate parameter λ that varies with time; λ becomes λ(t). Allowing λ to vary with time opens up the possibility for many Poisson process variants. One such variant occurs when λ(t) is itself a random process and this leads to what is called a doubly-stochastic Poisson or Cox process (Cox & Lewis, 1966).

The doubly stochastic Poisson process was originally used to model stops of a loom (in textile assembly) because of breakages in the weft threads (Cox 1955) and has subsequently been used to model many processes in physics, biology, neuroscience and finance (see for example Snyder & Miller, 1991). The random process that controls the rate λ(t) is referred to as the *driving* process or the *information* process because it is the quantity that conveys useful information (Snyder & Miller, 1991). For the Cox-Ingersoll-Ross driven doubly-stochastic Poisson process (CIR driven DSPP) the driving process is a Cox-Ingersoll-Ross process.

The Cox-Ingersoll-Ross (CIR) process was developed to study the term structure of interest rates (Cox et al. 1985). It is a generalisation of the squared Ornstein-Uhlenbeck (OU) process (LePage et al. 2006) characterised by the stochastic differential equation of Equation (1), where a, b and σ are parameters, X_t is a random variable, dX_t is an incremental change in the random variable with incremental change in time dt, and dW_t is a differential Wiener process.

$$dX_t = a(b - X_t)dt + \sigma\sqrt{X_t}dW_t \quad (1)$$

A solution to Equation (1) is the non-central χ^2 distribution. Equation (2) shows the non-central χ^2 probability density function $P(x)$ adapted from Weisstein 2012, with degree of freedom φ and non-centrality parameter ζ. Equations (3) and (4) show, respectively, expressions for the degree of freedom and non-centrality parameter for a CIR driven DSPP with parameters a, b and σ. Note that for this process the non-centrality parameter $\zeta(t)$ is a function of time but the degree of freedom φ is not. $I_n(x)$ is a modified Bessel function of the first kind. The value x_o that appears in Equation (4) is an initial value of x (chosen to be CIR parameter b).

$$P(x) = \frac{e^{-(x+\zeta)/2}x^{(\varphi-1)/2}\sqrt{\zeta}}{2(\zeta x)^{\varphi/4}} I_{\varphi/2-1}\left(\sqrt{\zeta x}\right) \quad (2)$$

$$\varphi = \frac{4ab}{\sigma^2} \quad (3)$$

$$\zeta(t) = \frac{4ax_o e^{-at}}{\sigma^2(1-e^{-at})} \quad (4)$$

The time-dependent covariance of the stationary process $\rho(t)$ is given in Equation (5). This covariance function can be used to give a theoretical expression for the Fano-factor time curve (Equation (6)). Each of the CIR parameters can be loosely described in terms of the rate. Parameter b is in some sense an average or expected rate, parameter a controls the dependence between samples (evidenced by an exp(-at) term in the covariance function) and σ controls the level of variability of the rate.

$$\rho(t) = \frac{b\sigma^2}{2a}e^{-at} \quad (5)$$

The CIR process has been shown to provide a good model for clustering of snapping shrimp noise over medium counting times (between 1 and 200 seconds) albeit without a physical or biological mechanism (Legg 2010). It is likely that the model has sufficient degree of freedom and generality to model many processes that display clustering but have numerous external influences. Furthermore, it is one of only a few driving models for the doubly-stochastic Poisson process that has an analytic solution for the Fano-factor time curve. This fact, combined with the property that negative rates cannot occur, make the CIR process an attractive model.

The CIR driven DSPP model will be used as the foundation for all of the processes presented in this paper, with diurnal influences introduced by varying the CIR process model parameter b. For a detailed description of the CIR driven DSPP model see LePage et al. (2006).

Fano-factor analysis

Fano-factor analysis is one technique in a spectrum of analysis techniques available for investigating point processes. Other techniques include inter-event interval histogram analysis (Weiss 1966), higher order interval analysis (Perkel et al. 1967), serial correlation coefficients (Cox & Lewis 1966), rescaled range (Lowen et al. 2001) and spectral techniques (Bartlett 1963). Fano-factor analysis is appropriate for temporal processes that have importance in the order of the events, or when only a count of the number of events is available, and if important information exists for time scales much greater than the average inter-event time (Turcott et al. 1994). Fano-factor analysis of the counting process is particularly useful because the results maintain the same time reference as the point process, whereas techniques based on intervals have a deformed time reference (Lowen et al. 2001).

The theoretical Fano-factor time curve for a CIR driven DSPP model is given in Equation (6), where w is counting time and σ and a are parameters of the CIR process. As $w \to \infty$ the Fano-factor time curve for this process has a finite asymptote at $1+ (\sigma/a)^2$. This finite asymptote is an important feature that can be used to help distinguish the CIR driven DSPP process from other processes.

$$FF(w) = 1 + \frac{\sigma^2}{a^3 w}(e^{-aw} + aw - 1) \qquad (6)$$

Fano-factor values less than one reflect a process that is sub-Poisson (having increased regularity) and values greater than one indicate a super-Poisson process (having increased clustering). The CIR driven DSPP is a super-Poisson model and will generate Fano-factor values greater than one. Fractal processes display a power law relationship between the Fano-factor and counting time that can be approximated by Equation (7) with counting time w, a scaling factor w_o and fractal exponent α_F (Lowen & Teich 1996). This representation is valid for $0< \alpha_F <1$ (Lowen et al. 2001).

$$FF(w) \approx 1 + \left(\frac{w}{w_o}\right)^{\alpha_F} \qquad (7)$$

A measure of significance is required for the Fano-factor time curve because normal variations of the sampling distributions can be quite large and are dependent on counting time. χ^2 approximations useful for deviations from a homogeneous Poisson process (Hoel 1943, Selby 1965, Eden & Kramer 2010) do not apply to a test for deviation from a CIR driven DSPP model. Alternative methods such as shuffling (random permuting) (Lowen & Teich 1992) can produce guide levels to pseudo-test for deviations from a homogeneous Poisson process but again these techniques cannot be used to test for deviations from a CIR driven DSPP. The only remaining option for setting guide levels for the expected variability of the Fano-factor for a CIR driven DSPP process is to use a Monte Carlo approach. In this paper Monte Carlo guide levels are produced using ten independent realisations of the CIR driven DSPP process used as a control and set guide levels at two standard deviations beyond the mean (assuming a normal approximation). Deviations observed for a diurnal varying process (shown in later figures) are large enough to justify the use of Monte Carlo guide levels.

SIMULATION

Simulations were conducted using MATLAB with the signal processing and statistics toolboxes. Random numbers were drawn from a CIR random number generator and a Poisson random number generator.

CIR random numbers were generated using the `cirpath.m` matlab code (Shvorob 2007). The cirpath code is based on the method described in Glasserman (2004) but uses `ncx2rnd.m` (from the matlab statistics-toolbox) to generate random numbers from a non-central χ^2 distribution. The cirpath code requires an input time vector t and input CIR parameters a,b,s and r_0, and returns a single sample path from a Cox-Ingersoll-Ross process. Parameters a,b and s correspond with parameters a,b and σ from the theory, and r_0 is the initial value of the process at time t_0 corresponding with x_0 in Equation (4).

Poisson random numbers were generated using the `poissrnd.m` matlab function. The random number generator requires only one input parameter λ (the rate) and outputs one Poisson random number for each input λ. This method of generating Poisson random numbers assumes a small incremental time homogeneity. For each time increment, the process is assumed locally homogeneous with constant rate equal to the value of λ from the driving process for that particular instant in time. The assumption is valid provided the time increment is small compared with important clustering times. A series of these Poisson random numbers provide the counts that represent an observation of the process and these are the primary simulation output.

To achieve the objectives the following activities and simulations were performed:

- Simulate persistent snapping, with no diurnal variations, to use as a control
- Define and compute diurnal variation functions
- Simulate diurnal variations by changing parameters of the persistent snapping process

Each of these simulations is described in the following sections.

Persistent snapping (control) process

Persistent snapping was modelled as a CIR driven DSPP with clustering times expected between 1 and 60 seconds and average snap rate of 20 snaps per second. To simulate using this model the CIR parameter values chosen were a=0.6, b=20 and σ=3. These parameter choices reflect an environment of vigorous snapping, such as near a wharf or pier, where on average 20 snaps are close enough to a hydrophone to exceed a fixed pressure amplitude threshold.

Ten independent realisations of the persistent process were simulated. Each simulation was of 31 day duration and used a small time homogeneity interval of 0.1 seconds, giving a total of 26784000 Poisson random numbers. These random numbers were then counted into one second bins giving a total of 2678400 count values as the basic counting process. These ten realisations were used to compute Monte Carlo guide levels; estimates of normal levels of variability of the process.

A Fano-factor time curve was produced using the control data sets and is shown in Figure 1. The curves show the mean Fano-factor (solid black line), upper and lower guide levels set at two standard deviations beyond the mean Fano-factor (grey lines) and a theoretical curve evaluating Equation (6) with the parameters used for the simulation (dashed red line in agreement with the mean level).

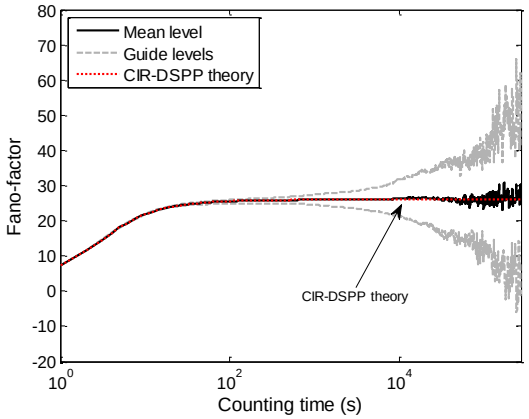

Figure 1: Fano-factor time curve (solid black line) derived from ten independent realisations of the persistent snapping process (CIR driven DSPP). This data set provides the control for no diurnal influence on the process. Guide levels using the Monte Carlo method (grey lines) were set at plus and minus 2 standard deviations beyond the mean. A theoretical curve for the CIR driven DSPP (red dashed line) is in good agreement with the mean Fano-factor.

Diurnal variation functions

Diurnal variations were introduced by a time dependent scaling function applied to the average value of the driving function. For a CIR driving function the average value, in the limit of infinite counting time, is equal to the parameter b. Given a scaling function $m(t)$ the modified value of b at time t is $\hat{b}(t) = bm(t)$. The effect of using this multiplier on the CIR parameter b is to change the degree-of-freedom of the non-central χ^2distribution of rate as a function of time; it does not change the non-centrality parameter (which is also a function of time).

A realistic diurnal function was constructed using a piecewise function followed by interpolation resampling and various (possibly random) adjustments. 26 initial values were chosen to roughly follow the average snap rate observed over a full day of measured data from Legg (2010). These 26 values give the basic structure of diurnal variations that were then resampled at the rate required for the simulations (ten times per second).

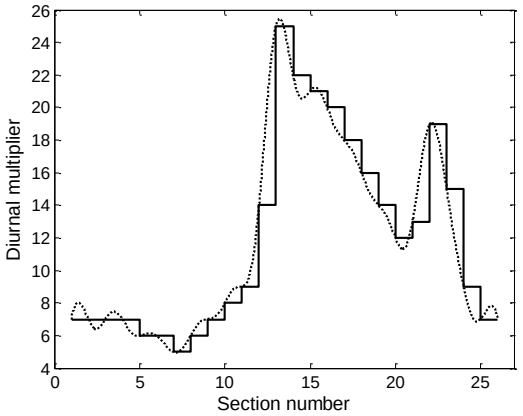

Figure 2: Piecewise (solid line) and resampled (dotted line) diurnal functions, or multipliers, for one day. These template forms are modified in various ways to allow day-to-day variations over the course of a month.

Shown in Figure 2 is a plot of the 26 point piecewise function (solid line) used for the simulations and includes a curve of resampled data (dotted line) that provides a template of variations over a single day. This plot shows only the diurnal function without any variations.

The form of the template diurnal function was adjusted in several ways to allow changes in the diurnal multiplier from day to day over the course of one month. The following adjustments were made:

- Random change of start time
- Addition of (independent Gaussian) random noise
- Random permutation of the order of sections in the piecewise function (destroying most, but not all, of the diurnal pattern).

Using these adjustments and combinations of adjustments gave the following diurnal function set:

- FR - Constant multiplier with no variation from day to day (same as the control data)
- DD - Diurnal multiplier with no variation from day to day (an exact replica of the diurnal function is used for each day)
- DDDO -Diurnal multiplier with random start time offsets (the diurnal function is randomly time shifted each day)
- SSDO - Shuffled diurnal multiplier with random start time offsets (each day uses a new randomly permuted diurnal multiplier and a new time shift)
- RRDO – Diurnal multiplier with independent Gaussian noise added and random start time offsets (each day has a different set of random noise added and a new time shift)
- SRDO – Shuffled diurnal multiplier with random noise and random offset (a combination of all variations applied to each new day).

A consequence of the variation structure is that many different scenarios can be simulated and their effect on the Fano-factor time curve investigated with reference to long time clustering. This freedom allows for testing of the known structure of the shrimp noise against any arbitrary changes that may occur over the period of one day and with variations from day to day. The characteristic features of a daytime, a night-time and two peak time regions are retained by some but not all of the variations listed above.

Diurnal varying process

The diurnal process was simulated as a medium time process with time varying parameter b, controlled by the diurnal multiplier. Diurnal multiplier values were chosen to simulate low snap rates during the "daytime" and higher snapping rates during the "night-time". The terms "daytime" and "night-time" are used loosely here to represent different periods during a single (24-hour) day. A choice of higher snapping rate during the night-time compared with the daytime reflects a typical pattern for many locations (Radford et al. 2008) although the reverse has been observed (Lammers et al. 2008).

Given the CIR driving process modified by the diurnal multiplier function, the series of counts were generated assuming incremental time homogeneity. For each time increment, the process was assumed locally homogeneous with constant rate equal to the value of λ from the CIR process for that particular instant in time. The assumption is valid provided the time increment is small compared with important clustering times.

This assumption can be tested through comparison of simulated data with the expected theoretical curve for a CIR driven DSPP process. For long time clustering (in this case for counting times greater than 200 seconds), the effect of the incremental time homogeneity is considered negligible provided the time increments are a fraction of a second.

Simulation data was produced by generating one Poisson random number for each time interval using the λ generated by the (modified) CIR process for that particular time. The diurnal process was generated using parameter values a=0.6, $b=\hat{b}(t)$, σ =3 and $r_0=\hat{b}(0)$, where $\hat{b}(0)$ is the starting value for b at time t=0. These parameter values are identical to those used for the control with the exception of b.

RESULTS

Description of the analysis and plots

Results from each of the simulations were subjected to a Fano-factor analysis and plotting routine. The plotting routine produced figures with two sub-plots, one on the left and one on the right. The left sub-plot was used to display an overlay of the diurnal function as a function of time (in seconds) on linear-linear scales for the duration of one day. Each subsequent day is overlayed. The right sub-plot shows the Fano-factor as a function of counting time (in seconds) on log-log scales. Normally the Fano-factor would be plotted on a linear scale but the effect of diurnal variation in subsequent figures is so large that the guide levels could not be resolved without using a logarithmic scale. A dashed black line was used on each of the Fano-factor time curve plots to show the 0.95 quantile level. This quantile level was chosen as the nominal peak level of the Fano-factor time curve and the point at which each of the curves started to return back toward the CIR driven DSPP asymptote. An estimate of the peak counting time (the counting time at which the Fano-factor reaches the peak value) was computed as the midpoint between the first increasing transition through the 0.95 quantile level and the last decreasing transition through the same quantile level.

Simulation results

Figure 3 shows the results from simulation FR, having a fixed rate for the entire month of simulated data. The diurnal function is just a single multiplier value of 12× with no variations. The right sub-plot shows a Fano-factor time curve for this data (black solid line) along with guide levels from the control (grey lines). The shape of the curve follows the guide levels indicating that data are consistent with a CIR driven DSPP (the expected result). Although the 0.95 quantile line has been plotted on this figure it does not relate to a peak.

The DD simulation data was the first to use a diurnal variation. In this data the diurnal function was repeated for every day in the month as seen in the left sub-plot in Figure 4 (the overlays lie exactly over each other). The diurnal function is a replica of the resampled piecewise diurnal function shown in Figure 2 displaying the characteristic double peak with a peak level near 25× and normal daytime level around 7×. The effect of this diurnal variation on the Fano-factor time curve is dramatic and extremely significant with respect to the guide levels. The curve rises with power law characteristic from 65 s (1.08 minutes) to 6524 s (1.81 hours). This power law section of the curve describes a straight line on the log-log plot corresponding with a fractal exponent of 0.97. Following the power law increase the Fano-factor value peaks at 43308 when the counting time is 30155 s (8.38 hours) then turns back toward the CIR driven DSPP asymptote. A ringing effect can be seen following the peak; with pseudo-nulls occurring at 87160 s (24.21 hours), 174469 s (48.46 hours) and 262428 s (72.90 hours). A distinct transient anomaly in the Fano-factor curve can be seen at counting time 43101 s (11.97 hours) corresponding with one half of a day.

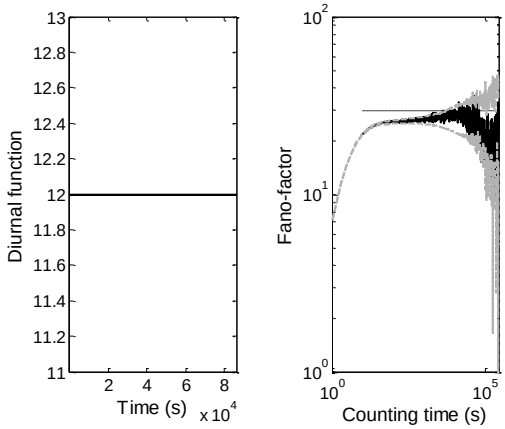

Figure 3: FR (fixed rate) simulation results showing the diurnal function (left sub-plot) and Fano-factor time curve (right sub-plot). This simulation used a fixed value for the diurnal function (no variations) and as expected the Fano-factor time curve shown in the right sub-plot (black solid line) is bounded by the guide levels (grey lines).

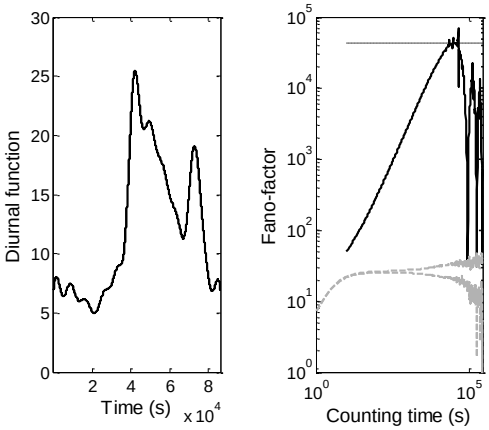

Figure 4: DD (diurnal varying) simulation results showing the diurnal function (left sub-plot) and Fano-factor time curve (right sub-plot). The diurnal function varied through the day but was not changed from day to day. The effect of diurnal variation is reflected in a large deviation of the Fano-factor time curve (black solid line) from the control (grey lines).

The DDDO simulation introduced a random offset to the start time of the diurnal function from day to day. The random offset can be seen as a spread of the diurnal functions in the overlay (left sub-plot) in Figure 5. Inspection of the Fano-factor time curve shows results similar to the DD simulation. A Fano-factor peak of 46970 occurs at a counting time of 28948 s (8.04 hours). The curve follows a power-law between 78 s (1.3 minutes) and 6074 s (1.69 hours) with fractal exponent 0.97. Pseudo-nulls are also present at 86275 s (23.97 hours), 172697 s (47.97 hours) and 259763 s (72.16 hours) although they are not as pronounced as the nulls in the DD data.

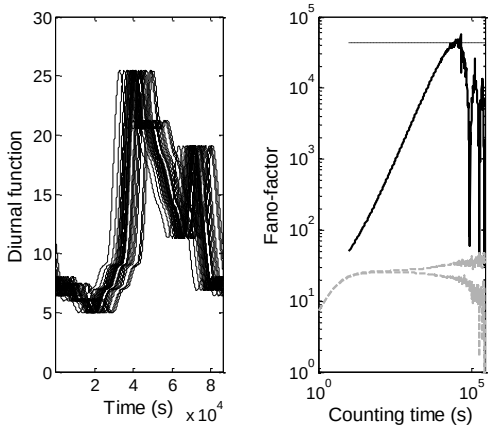

Figure 5: DDDO (random time offset) simulation results showing the diurnal function overlay (left sub-plot) and Fano-factor time curve (right sub-plot). Fano-factors (solid black line) are shown with guide levels (grey lines) and peak level (black dashed line).

The SSDO simulations introduced shuffling of the diurnal function for each day prior to the resampling. This level of change to the diurnal function results in loss of the characteristic dual peaks. Day to day changes display much more variation in the overlay (left sub-plot) on Figure 6. The Fano-factor results shown in the right sub-plot of Figure 6 retain the characteristic power-law curve that now occurs at lower counting times between 38 s (0.63 minutes) and 2449 s (0.68 hours) and has a slightly lower fractal exponent of 0.94. The Fano-factor peak is reduced to 9127 and occurs at a counting time of 14911 s (4.14 hours). Pseudo-nulls exist at counting times 86275 s (23.97 hours), 172697 s (47.97 hours) and 257126 s (71.42 hours) although they are not obvious.

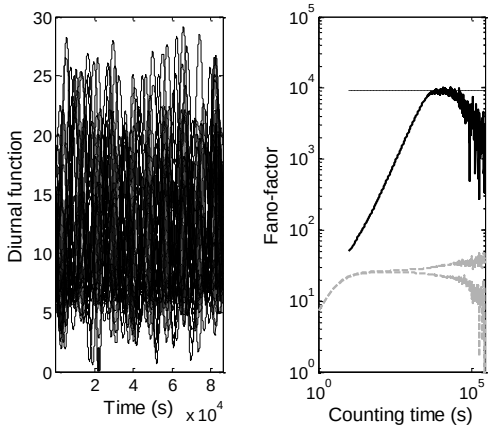

Figure 6: SSDO (shuffled) simulation results showing the diurnal function overlay (left sub-plot) and Fano-factor time curve (right sub-plot). The Fano-factor sub-plot contains guide levels (grey lines) and a peak level (black dashed line).

The RRDO simulation used random noise to change the diurnal function both before and after the resampling operation. Results are shown in Figure 7. The result is a spread of the diurnal function from day to day and distinct changes in the form on small scales. The form of the function over each day retains the characteristic double peak expected of snapping shrimp noise. The Fano-factor results for this simulation are consistent with the DD and DDDO simulations. A Fano-factor peak of 40531 occurs at a counting time of 30464 s (8.46 hours) following a power-law rise with fractal exponent 0.97. The power-law curve occurs between counting times 83 s (1.38 minutes) and 5951 s (1.65 hours). This simulation also has prominent pseudo-nulls at counting times 86275 s (23.97 hours), 172697 s (47.97 hours) and 254515 s (70.70 hours)

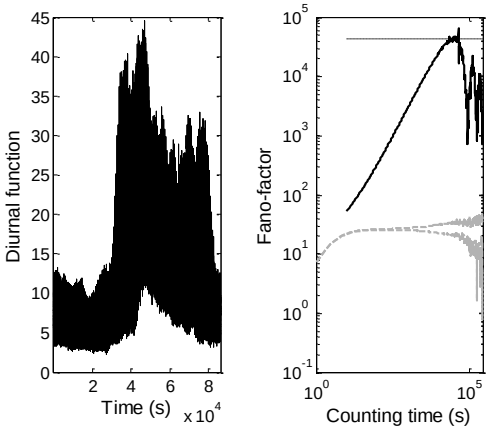

Figure 7: RRDO (random noise) simulation results showing the diurnal function overlay (left sub-plot) and Fano-factor time curve (right sub-plot). The Fano-factor sub-plot contains guide levels (grey lines) and a peak level (black dashed line).

The last simulation in the series (SRDO) used a combination of all variations that had been used previously. The piecewise diurnal function was shuffled and random noise added prior to resampling. After resampling a random start time offset was applied and more random noise added. The resulting overlay for a month of day to day changes is shown in the left sub-plot of Figure 8. The Fano-factor time curve displays a strong power-law region with fractal exponent 0.95 and a Fano-factor peak of 8203 at a counting time of 8771 s (2.44 hours). There was less structure for counting times greater than the peak counting time and no prominent pseudo-nulls. Close inspection revealed a pseudo-null near the first expected cycle time (24 hours) however the pseudo-null was not more prominent than the surrounding noise and could easily have occurred by chance. A null-like feature occurs at a counting time of 201266 s (55.91 hours). This null-like feature does not relate to any obvious cycle time in the process, reinforcing that these features can occur by chance.

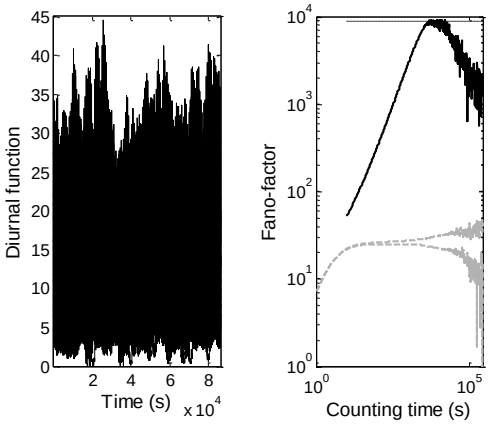

Figure 8: SRDO (combination) simulation results showing the diurnal function (left sub-plot) and Fano-factor time curve (right sub-plot). The Fano-factor time curve (black solid line) deviates from the control (grey lines) up to the Fano-factor peak (black dashed line). The diurnal function was changed using a combination of all previous randomizing schemes.

Fano-factor time curve results for selected simulation results are shown in Figure 9 on a semi-logarithmic plot. This plot shows the location of the 24-hour and 48-hour pseudo-null features. Also visible in the figure is an anomaly at approximately 12 hours. These features appear to a greater or lesser extent depending on the type of randomisation applied to the template diurnal function. The exact repeating function (DD) showed quite smooth results with deep pseudo-nulls. Adding a random time offset (DDDO) introduced some additional noise and reduced the depth of the pseudo-nulls. Shuffling the piecewise diurnal function (SSDO and SRDO) destroyed most of the structure; however pseudo-nulls still appear.

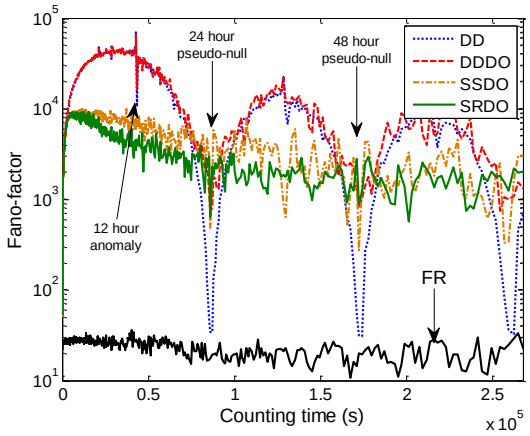

Figure 9: Fano-factor time curves for each of the simulation plotted on a semi-logarithmic scale. The pseudo-null features can be clearly seen (but the preceding power-law curve cannot) due to the linear time scale. The 24-hour and 48 hour pseudo-nulls have been annotated along with an anomaly that appears on the DD and DDDO curves around 12 hours.

DISCUSSION

A series of simulations were conducted to investigate the effect of diurnal variation on clustering of snapping shrimp snaps, particularly for long time scales. Fano-factor analysis was used as a measure of clustering and the Fano-factor time curve was computed and interpreted for a range of counting times and for each set of simulated data. Simulated data was produced to represent one month of continuous snapping observation.

A CIR driven DSPP model was used as a control and to provide a model for the normal snapping without any diurnal variations. This model precluded the use of guide levels from χ^2 approximation and shuffle techniques, requiring the use of less satisfactory Monte Carlo guide levels. Monte Carlo guide levels are extremely approximate and cannot be used if the deviations are not persuasively large. Fortunately the deviations caused by the introduction of diurnal variations were so large that their significance was not in doubt. Diurnal variations were introduced using a diurnal multiplier function applied to the CIR parameter *b*. A series of simulations were conducted to investigate how various randomising schemes affected the Fano-factor time curve for large counting times. Randomising schemes were used to offset the daily start time, shuffle the order of level changes in the daily cycle, add random noise or combinations of these.

The overall result was very persistent; all forms of diurnal variation produced significant deviation of the Fano-factor time curve from the control in the positive (clustering) direction. Furthermore, each Fano-factor curve had a power-law section for a specific counting time interval prior to a peak. The peak was followed by a region that contained additional structure depending on the cyclic information in the process and the choice of diurnal randomisation. Non shuffled randomisation schemes displayed a series of pseudo-nulls that coincided (within a fraction of an hour) with integer multiples of 24 hours, which was the characteristic cycle time. Shuffled randomisation schemes tended to reduce or eliminate this cyclic information. The location of pseudo-nulls remained relatively constant despite some large changes in location and strength of the power-law curve and Fano-factor peak.

The simulation results have provided answers to the important questions that were posed in the introduction. These questions are reiterated and accompanied by an answer in the following paragraphs.

Does diurnal variation in snapping have any effect on the Fano-factor time curve? Diurnal variations do affect the Fano-factor time curve in the super-Poisson (clustering) sense for long counting times, and the effect can be very strong.

What characteristics might we expect from a diurnal varying process (does it have a fractal nature)? Several characteristics may be observed including a region of power-law behaviour indicating some fractal nature. However, the power-law curve does not continue for all counting times (as would be expected for a truly fractal process). For large counting times the power-law curve transitions to a peak and then progresses through a series of pseudo-nulls that are related to important cycle times in the process.

How long does real shrimp noise need to be observed to allow these features to be recognised? The month (31 days) of simulated data allowed the power-law region to be observed along with three pseudo-nulls related to a 24 hour cycle time. If three pseudo-nulls are sufficient to establish the existence of the cycle time then (at least) one month of data is needed. If only one pseudo-null is desired then the duration could be reduced to ten days, however the result would not be conclusive.

CONCLUSIONS

Diurnal variations have a profound effect on the Fano-factor time curve at long counting times. For all of the scenarios investigated, the Fano-factor time curve had a strong power law rise and peak followed by additional structure, including pseudo-nulls, related to important cycle times in the process. It is anticipated that Fano-factor analysis of real shrimp noise will show similar structure, and the results presented will enhance our ability to interpret long time clustering characteristics of real snapping shrimp noise.

One month of data was sufficient to observe a region of power-law behaviour in the Fano-factor time curve followed by a peak and several features related to important cyclic time scales in the process. However, one month of data was not sufficient to observe asymptotic behaviour and therefore other questions, such as a return to the CIR driven DSPP asymptote, remain open. Further understanding of the cyclic and non-cyclic features for counting times greater than the peak counting time is required. The next step will be to simulate over many months or possibly years in an attempt to observe the asymptotic behaviour of the simulated process, however the measurement and analysis of real shrimp noise over such time scales will be very difficult.

REFERENCES

Bartlett, MS 1963, 'The spectral analysis of point processes', *J. R. Statist. Soc. B*, vol. 25, no. 2, pp. 264-296.

Buckingham, MJ, Berkhout, BV & Glegg, SA 1992, 'Imaging the ocean with ambient noise', *Nature*, vol. 356, no. 6367, pp. 327-329.

Cato, DH & Bell MJ 1992, 'Ultrasonic ambient noise in Australian shallow waters at frequencies up to 200 kHz', MRL Technical Report MRL-TR-91-23, Materials Research Laboratory, DSTO, Australia.

Chitre, M 2010, 'Acoustic sensing in snapping shrimp dominated environments', in *Proceedings of 20th International Congress on Acoustics, ICA 2010*, Sydney, Australia.

Chitre, MA, Potter, JR & Ong, SH 2006, 'Optimal and near-optimal signal detection in snapping shrimp noise', *IEEE J. Ocean. Eng.*, vol. 31, no. 2, pp. 497-503.

Cox, DR 1955, 'Some statistical methods connected with series of events', *J. R. Statist. Soc. B*, vol. 17, no. 2, pp. 129-164.

Cox, DR & Lewis, PAW 1966, *The statistical analysis of series of events*, Methuen's Monographs on Applied Probability and Statistics, Methuen & Co Ltd, London.

Cox, JC, Ingersoll, JE & Ross, SA 1985, 'A Theory of the Term Structure of Interest Rates', *Econometrica*, vol. 53, no. 2, pp. 385-407.

Eden, UT & Kramer, MA 2010, 'Drawing inferences from Fano factor calculations', *Journal of Neuroscience Methods*, vol. 190, pp. 149–152.

Everest, FA, Young, RW & Johnson, MW 1948, 'Acoustical characteristics of noise produced by snapping shrimp', *J. Acoust. Soc. Am.*, vol. 20, no. 2, pp. 137-142.

Glasserman, P 2004, *Monte Carlo methods in financial engineering*, Applications of mathematics, Springer Science + Business Media Inc., New York, NY.

Hoel, PG 1943, 'On indices of dispersion', *Ann. Math. Statist.*, vol. 14, no. 2, pp. 155-162.

Johnson, MW, Everest, FA & Young, RW 1947, 'The role of snapping shrimp (Crangon and Synalpheus) in the production of underwater noise in the sea', *Biol. Bull.*, vol. 93, no. 2, pp. 122-138.

Kuselan, S, Chitre, M & Pallayil, V 2011, 'Ambient noise imaging through joint source localization', *in IEEE/MTS Oceans'11 Conference*, Hawaii, USA, pp. 1-6.

Lammers, MO, Brainard, RE, Au, WWL, Mooney, TA & Wong, KB 2008, 'An ecological acoustic recorder (EAR) for long-term monitoring of biological and anthropogenic sounds on coral reefs and other marine habitats', *J. Acoust. Soc. Am.*, vol. 123, no. 3, pp. 1720-1728.

Legg, MW 2010, 'Non-Gaussian and non-homogeneous Poisson models of snapping shrimp noise', PhD thesis, Curtin University, Western Australia.

Lepage, T, Lawi, S, Tupper, P & Bryant, D 2006, 'Continuous and tractable models for the variation of evolutionary rates', *Mathematical Biosciences*, vol. 199, no. 2, pp. 216-233.

Lowen, SB & Teich, MC 1992, 'Auditory-nerve action potentials form a nonrenewal point process over short as well as long time scales', *J. Acoust. Soc. Am.*, vol. 92, no. 2, pp. 803-806.

Lowen, SB & Teich, MC 1996, 'The periodogram and Allan variance reveal fractal exponents greater than unity in auditory-nerve spike trains', *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3585-3591.

Lowen, SB, Tsuyoshi, 0, Kaplan, E, Saleh, BEA & Teich, MC 2001, 'Fractal features of dark, maintained, and driven neural discharges in the cat visual system', *Methods*, vol. 24, pp. 377-394.

Mallawaarachchi, A, Ong, SH, Chitre, M & Taylor, E 2008, 'Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles', *J. Acoust. Soc. Am.*, vol. 124, no. 2, pp. 1159-1170.

Perkel, DH, Gerstein, GL & Moore, GP 1967, 'Neuronal spike trains and stochastic point processes 1: The single spike train', *Biophysical Journal*, vol. 7, pp. 391-418.

Potter, JR & Koay, TB 2000, 'Do snapping shrimp chorus in time or cluster in space? Temporal-spatial studies of high-frequency ambient noise in Singapore waters', in *European Conference on Underwater Acoustics 2000*, Lyons, France.

Radford, CA, Jeffs, AG, Tindle, CT & Montgomery, JC 2008, 'Temporal patterns in ambient noise of biological origin from a shallow water temperate reef', *Oecologia*, vol. 16, pp. 921-929.

Schmitz, B & Herberholtz, J 1998, 'Snapping behaviour in intraspecific agnostic encounters in the snapping shrimp (*Alpheus heterochaelis*)', *Journal of Biosciences*, vol. 23, no. 5, pp. 623-632.

Selby, B 1965, 'The index of dispersion as a test statistic', *Biometrika*, vol. 52, no. 3/4, pp. 627-629.

Svorob, D 2007, 'Simulate a Cox-Ingersoll-Ross process', online resource, viewed 14 March 2012, <http://www.mathworks.com/matlabcentral/fileexchange/16670-simulate-a-cox-ingersoll-ross-process/content/cirpathdemo/cirpath.m>

Snyder, DL & Miller, MI 1991, *Random point processes in time and space*, 2nd edn., Springer Texts in Electrical Engineering, Springer, New York.

Teich, MC & Lowen, SB 1994, 'Fractal patterns in auditory nerve-spike trains', *IEEE Eng. Med. Biol. Mag*, vol. 13, pp. 197-202.

Tóth E & Duffy JE 2005, 'Coordinated group response to nest intruders in social shrimp', *Biol. Lett.*, vol. 1, pp. 49-52.

Turcott, RG, Lowen, SB, Li, E, Johnson, DH, Tsuchitani, C & Teich, MC 1994, 'A stationary Poisson point process describes the sequence of action potentials over long time scales in lateral-superior-olive auditory neurons', *Biol. Cybern.*, vol. 70, pp. 209-217.

Weiss, TF 1966, 'A model of the peripheral auditory system', *Biological Cybernetics*, vol. 3, no. 4, pp. 153-175.

Weisstein, EW 2012, *Noncentral Chi-Squared Distribution* From MathWorld--A Wolfram Web Resource, viewed 21 May 2012, <http://mathworld.wolfram.com/NoncentralChi-SquaredDistribution.html>.