INTERNATIONAL JOURNAL OF
INFORMATION AND SYSTEMS SCIENCES
Volume 6, Number 1, Pages 49–60

A COMPARATIVE STUDY OF COHERENCE, MUTUAL INFORMATION AND CROSS-INTENSITY MODELS

TING WANG, MARK S. BEBBINGTON AND DAVID S. HARTE

Abstract. Coherence is a measure of the time invariant linear dependence of two processes at certain frequencies, and provides a measure of the degree of linear predictability of one process from another process. The coherence is inadequate as a measure of general association for it may be identically 0 when two series are in fact related. However, such behavior does not occur for the coefficient of mutual information, which is a measure of the amount of information that one random variable contains about another random variable. The Lin-Lin model, which describes the influence of an input on a point process output, can identify linear causal relationships between one sequence of events and another. This paper presents a comparative study of the three approaches using a case study of the relationship between groundwater level data from Tangshan Well and global earthquakes with minimum magnitude 5.8.

Key Words. Coherence, mutual information, Lin-Lin model, Groundwater level, Earthquakes.

1. Introduction

Correlation between two or more series of events is often an issue in statistics or other scientific subjects. This paper will review three very important and popular statistical methods for ascertaining the association between two point processes: coherence, mutual information and the Lin-Lin model. Previous papers in the literature have compared the coherence with mutual information (e.g. Brillinger, 2002; Brillinger, 2003). However, to date, none have discussed the three approaches together.

Coherence is a stationary process analog of the traditional correlation coefficient, taking on values between 0 and 1 at any given frequency. It is a measure of the time invariant linear dependence of the two processes at frequency λ, (Brillinger, 1975; Brillinger, 1994; Brillinger and Villa, 1997), and provides a measure of the degree of linear predictability of one process from another process. When the coherence function is identically zero, one process is of no use in linearly predicting the other. When it is identically one, one process gives a perfect linear prediction of the other (Brillinger, 1994; Iyengar, 2001). If it is significantly greater than zero over a limited frequency range, it implies association between the two processes over this frequency range.

Mutual information was originally introduced by Shannon (1948). It is a measure of the amount of information that one random variable contains about another random variable (Cover and Thomas, 1991). The coherence is inadequate as a

Received by the editors September 12, 2008 and, in revised form, October 31, 2008.
This research was supported by the Marsden Fund, administered by the Royal Society of New Zealand.

measure of general association for it may be identically 0 when two series are in fact related (Brillinger, 2003; Brillinger and Guha, 2007). Brillinger (2003) proved that such behavior does not occur for the coefficient of mutual information for random variables. The mutual information coefficient has the property of taking on the value 0 if and only if the variables are statistically independent. In this work, we transform the possible pairs of random variables, in the point process context, to pairs of intervals by considering the point process as a 0-1 time series, and then calculate the mutual information as a function of the time lag.

A significant cross correlation between two series of events, N and M, doesn't mean that we can determine whether

(i) N causes M;
(ii) M causes N;
(iii) N and M cause each other; or
(iv) some other process causes both N and M

(Ogata, 1999). To discriminate among the first three cases, as well as to test the significance, Ogata et al. (1982) suggested a parametric model based on the self-exciting and mutually exciting point processes introduced by Hawkes (1971). The model describes the influence of an input on a point process output, or in other words, it can identify linear causal relationships between one sequence of events and another (Ogata, et al., 1982; Ma and Vere-Jones, 1997). Hence the model is named the Lin-Lin model. Unlike the previous two cases which can be defined very generally, this method applies specifically to point process models.

In this paper, we first review the formulation of coherence, mutual information and the Lin-Lin model, and how to ascertain the relationship between two processes using the three methods. Then we apply them to a real case study, to determine whether the water level fluctuations in a well are correlated with, and furthermore caused by, earthquakes.

2. Three statistics characterizing the association between two or more series of events

The data under consideration are functions, particularly realizations of stationary processes. In particular, two types of processes, 0-1 time series and point processes, are studied. In this section we briefly review three statistical methodologies, coherence and mutual information analyses, and the Lin-Lin model, which can be used to characterize the association between two or more series of events.

2.1. Coherence. Let (M, N) represent a bivariate point process. Let σ_m, $m = 0, \pm 1, \cdots$ and τ_n, $n = 0, \pm 1, \cdots$ denote the times of occurrence of the M and N events, respectively. For real-valued λ, set

$$d_M^T(\lambda) = \int_0^T \exp\{-i\lambda t\} \mathrm{d}M(t) = \sum_m e^{-i\lambda\sigma_m} \tag{1}$$

$$d_N^T(\lambda) = \int_0^T \exp\{-i\lambda t\} \mathrm{d}N(t) = \sum_n e^{-i\lambda\tau_n} \tag{2}$$

where T denotes the length of the time period of observation. Then one frequency domain measure of association which can be used to assess the linear dependency between processes M and N is defined by

$$\lim_{T\to\infty} |\mathrm{Corr}\{d_M^T(\lambda), d_N^T(\lambda)\}|^2 = \lim_{T\to\infty} \frac{|\mathrm{Cov}\{d_M^T(\lambda), d_N^T(\lambda)\}|^2}{\mathrm{Var}\{d_M^T(\lambda)\}\mathrm{Var}\{d_N^T(\lambda)\}} = |R_{MN}(\lambda)|^2, \tag{3}$$

which is called the coherence of the two processes at frequency λ. This can be interpreted as the magnitude squared of the correlation between the finite Fourier transforms of processes M and N. The definition of the correlation in variance and covariance terms, $\text{corr}\{A, B\} = \text{cov}\{A, B\}/\sqrt{\text{var}\{A\}\text{var}\{B\}}$, leads to an alternative definition of the coherence as

$$|R_{xN}(\lambda)|^2 = \frac{|f_{xN}(\lambda)|^2}{f_{xx}(\lambda)f_{NN}(\lambda)}, \tag{4}$$

where $f_{MN}(\lambda)$ is a hybrid cross-spectrum between point process M and N, which is defined as

$$f_{MN}(\lambda) = \lim_{T\to\infty} \frac{1}{2\pi T}\mathbf{E}\{d_M^T(\lambda), \overline{d_N^T(\lambda)}\}. \tag{5}$$

The auto-spectra $f_{NN}(\lambda)$ and $f_{MM}(\lambda)$ can be similarly defined.

Using the method of disjoint sections (Halliday, et al., 1995), the complete record, denoted by T, is divided into L non-overlapping disjoint sections each of length K, where $T = LK$. The finite Fourier transform of the lth segment ($l = 1, \cdots, L$) from process N at frequency λ is defined as

$$d_N^K(\lambda, l) = \int_{(l-1)K}^{lK} e^{-i\lambda t}\mathrm{d}N(t) \approx \sum_{(l-1)K\leq\tau_n\leq lK} e^{-i\lambda\tau_n}, \tag{6}$$

A similar definition holds for $d_M^K(\lambda, l)$. A consistent estimate of $f_{MN}(\lambda)$ can be given by

$$\hat{f}_{MN}(\lambda) = \frac{1}{2\pi LK}\sum_{l=1}^{L} d_M^K(\lambda, l)\overline{d_N^K(\lambda, l)}, \tag{7}$$

with similar expressions for the estimation of the auto-spectra $\hat{f}_{NN}(\lambda)$ and $\hat{f}_{MM}(\lambda)$. For large K and $\lambda \neq 0$, the estimated cross-spectrum $\hat{f}_{MN}(\lambda)$ can be interpreted as the covariance between the components, at frequency λ, of processes M and N. The estimated auto-spectrum, $\hat{f}_{NN}(\lambda)$, can be interpreted as the variance at frequency λ of the process N.

The coherence function can be estimated by direct substitution of the estimates of the appropriate spectra as

$$|\hat{R}_{MN}(\lambda)|^2 = \frac{|\hat{f}_{MN}(\lambda)|^2}{\hat{f}_{MM}(\lambda)\hat{f}_{NN}(\lambda)}. \tag{8}$$

Coherence functions provide a bounded and normative measure of association. In the case of independence, $|R_{xN}(\lambda)|^2 = 0$, the distribution of $|R_{xN}(\lambda)|^2$ can be evaluated in terms of the incomplete Beta function with parameters 1 and $(L-1)$ (Brillinger, 2001). The pointwise $100\alpha\%$ confidence limit is given by $1-(1-\alpha)^{1/(L-1)}$. Thus we will include the level

$$1 - 0.05^{1/(L-1)} \tag{9}$$

in the following coherence plots as a benchmark of the upper 95% confidence limit under the hypothesis of independence. Estimated values of coherence lying below this line can be taken as evidence for the lack of a linear association between the two processes, i.e. that zero coherence is plausible at that frequency (Brillinger, et al., 1976; Halliday, et al., 1995).

The second moment

$$d_N^T(\lambda) = \int_0^T \exp\{-i\lambda t\} \mathrm{d}N(t) \tag{10}$$

needs to be corrected by subtracting the mean $\bar{n}\mathrm{d}t$, (see, for example, Vere-Jones and Ozaki, 1982) where $\bar{n}$ is the average rate. Thus we can use

$$\begin{aligned} d_N^*(\lambda) &= \int_0^T \exp\{-i\lambda t\}(\mathrm{d}N(t) - \bar{n}\mathrm{d}t) \\ &= d_N^T(\lambda) - \int_0^T \exp\{-i\lambda t\}\bar{n}\mathrm{d}t \\ &= d_N^T(\lambda) - \frac{\bar{n}i}{\lambda}(\exp\{-i\lambda T\} - 1) \end{aligned} \tag{11}$$

The estimates are

$$\hat{d}_N^T(\lambda) = \sum_n e^{-i\lambda\tau_n} \quad \text{and} \quad \hat{\bar{n}} = \frac{N(T)}{T}, \tag{12}$$

where $N(T) = \#(0, T)$. Therefore we obtain

$$\hat{d}_N^*(\lambda) = \sum_n e^{-i\lambda\tau_n} - \frac{iN(T)}{\lambda T}(\exp\{-i\lambda T\} - 1). \tag{13}$$

For the method of disjoint sections, we use

(14)

$$\hat{d}_N^*(\lambda, l) = \sum_{(l-1)K<\tau_n\le lK} e^{-i\lambda\tau_n} - \frac{iN(K)}{\lambda K}(\exp\{-i\lambda lK\} - \exp\{-i\lambda(l-1)K\}),$$

where $N(K) = \#((l-1)K, lK]$.

2.2. Mutual information. The mutual information of a bivariate random variable (U, V) is defined as

$$I_{UV} = \mathbf{E}\left\{\log_2\left(\frac{p_{UV}(u,v)}{p_U(u)p_V(v)}\right)\right\},$$

where $p_{UV}(u, v)$ is the joint probability mass function, and $p_U(u)$ and $p_V(v)$ are the marginal probability mass functions. The mutual information is non-negative and measures the strength of dependence in that $I_{UV} = 0$ if and only if U and V are independent, $I_{UV_1} \le I_{UV_2}$ if U is independent of V_1 given V_2, and for the continuous case, $I_{UV} = \infty$ if $V = g(U)$.

When a bivariate variable (U, V) has a continuous distribution, the mutual information is

$$I_{UV} = \int\int \log_2\left(\frac{p_{UV}(u,v)}{p_U(u)p_V(v)}\right) p_{UV}(u,v)\mathrm{d}u\mathrm{d}v.$$

One popular mutual information estimator is obtained by substituting suitable density estimators $\hat{p}_{UV}(u, v)$, $\hat{p}_U(u)$ and $\hat{p}_V(v)$ into the above formula (Strong et al., 1998; Antos and Kontoyiannis, 2001), which takes the form

$$\hat{I}_{UV} = \int\int \log_2\left(\frac{\hat{p}_{UV}(u,v)}{\hat{p}_U(u)\hat{p}_V(v)}\right) \hat{p}_{UV}(u,v)\mathrm{d}u\mathrm{d}v.$$

We can use either parametric density estimators (Brillinger, 2004) or nonparametric ones (either histogram-based, Moddemeijer, 1989; or kernel-based, Mars and van Aragon, 1982; Joe, 1989; Granger and Lin, 1994; Moon et al., 1995).

For a bivariate discrete variable (U, V) with U taking on the values $1, \cdots, K$ and V taking on $1, \cdots, J$ and

$$P\{U = k, V = j\} = p_{kj}.$$

The marginals are then

$$p_{k+} = P\{U = k\} = \sum_{j=1}^{J} p_{kj}, \quad p_{+j} = P\{V = j\} = \sum_{k=1}^{K} p_{kj},$$

and the mutual information becomes

$$I_{UV} = \sum_{k,j} p_{kj} \log_2 \frac{p_{kj}}{p_{k+}p_{+j}},$$

assuming that $p_{jk} \neq 0$. Let $w = \{w_{kj}\}$, with

$$w_{kj} = \begin{cases} 1, & U = k, V = j, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that there are n independent realizations, $\{w_{kjl}, l = 1, \cdots, n\}$, of w. The maximum likelihood estimates of p_{kj} are $\hat{p}_{kj} = \sum_l w_{kjl}/n$ and the plug-in estimate of the mutual information (Brillinger, 2004; Brillinger and Guha, 2007) is

$$\hat{I}_{UV} = \sum_{k,j} \hat{p}_{kj} \log_2 \frac{\hat{p}_{kj}}{\hat{p}_{k+}\hat{p}_{+j}}, \tag{15}$$

where $\hat{p}_{k+} = \sum_j \hat{p}_{kj}$ and $\hat{p}_{+j} = \sum_k \hat{p}_{kj}$. Note that a point process can be transformed into a 0-1 time series, with 1 at the occurrence times (of the process) and 0 otherwise. The likelihood ratio test statistic of the null hypothesis that the two variables U and V are independent is

$$G^2 = 2n \sum_{k,j} \hat{p}_{kj} \log_2 \frac{\hat{p}_{kj}}{\hat{p}_{k+}\hat{p}_{+j}}$$

(Christensen, 1997; Brillinger, 2004; Brillinger and Guha, 2007). The asymptotic null distribution of G^2 under the hypothesis of the independence of U and V is $\mathcal{X}^2_{(J-1)(K-1)}$. Noticing the proportional relationship between the estimate $\hat{I}_{UV}$ in (15) and G^2, we have that under the null hypothesis that the two variables U and V are independent, and the large sample distribution of the estimate (15) is $\mathcal{X}^2_{(J-1)(K-1)}/2n$.

2.3. Ogata's Lin-Lin Model. Consider a point process $\{N_t\}$ with intensity function

$$\lambda(t) = \mu + \int_0^t g(t - s)\mathrm{d}N_s + \int_0^t h(t - s)\mathrm{d}X_t, \tag{16}$$

where $\{X_t\}$, the input process, may be either a point process or a cumulative process

$$X_t = \int_0^t x(s)\mathrm{d}s$$

of a stochastic process $x(t)$. The self-exciting term $g(t)$ describes the nature of the point process, while the transfer term $h(t)$ indicates the strength of the causal relation between the input process X_t and the output N_t. When the function $h(t) \equiv 0$, it means there is no causal relationship between the input and output processes, whereas when the function $g(t) \equiv 0$ and $h(t) \neq 0$, it means that the output process is a doubly stochastic Poisson process whose intensity is modulated only by the input process (Ogata, et al., 1982). The parametrization of the two response functions we will use follows Ogata, et al. (1982), where

$$g(t) = \sum_{k=1}^{K} a_k t^{k-1} e^{-ct}, \qquad h(t) = \sum_{k=1}^{L} b_k t^{k-1} e^{-ct}. \tag{17}$$

Given the occurrence times of two sequence of events $\{t_i : i = 1, \cdots, I\}$ (output) and $\{\tau_m : m = 1, \cdots, M\}$ (input) over the time interval $[0, T]$, the parameters are estimated by maximizing the partial log likelihood

$$\log L_T(\theta) = \sum_{i:0 \le t_i < T} \log \lambda_\theta(t_i) - \int_0^T \lambda_\theta(t) \mathrm{d}t,$$

where θ denotes $(\mu, c, a_1, \cdots, a_K, b_1, \cdots, b_L)$. To determine the orders K and L of the response functions, we use Akaike's information criteria (AIC) (Akaike, 1974)

$$\mathrm{AIC}(K, L) = -2 \max_\theta (\log L_T(\theta)) + 2(K + L + 2),$$

and choose K and L which minimize the AIC.

Ogata (1983) suggested the following model to examine case (iv),

$$\lambda(t) = a_0 + P_J(t) + C_K(t) + \sum_{t_i < t} g_N(t - t_i) + \sum_{\tau_m < t} h_X(t - \tau_m).$$

The second term on the right-hand side

$$P_J(t) = \sum_{j=1}^{J} a_j \phi_j(t/T), \quad 0 < t < T,$$

represents the evolutionary trend. The total length of the observed interval is T, and $\phi_j(\cdot)$ is a polynomial of order j. The third term

$$C_K(t) = \sum_{k=1}^{K} \{b_{2k-1} \cos(2k\pi t/T_0) + b_{2k} \sin(2k\pi t/T_0)\}$$

is the Fourier expansion for cyclic effects with a given fixed cycle length T_0. The fourth term is the clustering effect, i.e., a response function of the output process. The last term describes the causal relation from the input process to the output process. The last two terms are parameterized as in (17). If there is no causal relation from $\{\tau_m\}$ to the conditional intensity function $\lambda(t)$, or to the occurrence of $\{t_i\}$, then $h_X(x) = 0$ is expected. Maximum likelihood estimation can be used to estimate the parameters and the AIC can be used to determine the order(s).

3. Groundwater level data and earthquake catalogue

The data under consideration here are four years of groundwater level observations at Tangshan Well, a well located in DaZhao Park in Tangshan City, 100km southeast of Beijing, China, with longitude 118.18E, latitude 39.62N and altitude 23.43m, and global (i.e. located anywhere in the world) earthquakes with minimum magnitude 5.8 within the same time period. The groundwater level data have precision 0.0005m with a sampling rate of one observation per minute. According to the digital records of the groundwater level and the earthquake catalogue, a great number of fluctuations of the groundwater level at this well may have been induced by many global teleseismics, a possible example of which is shown in Figure 1.

We are going to use the three methods to examine whether the earthquakes are associated with, and further cause, the water level fluctuations at this well, using the data from the four years from January 2002 to September 2005. Groundwater

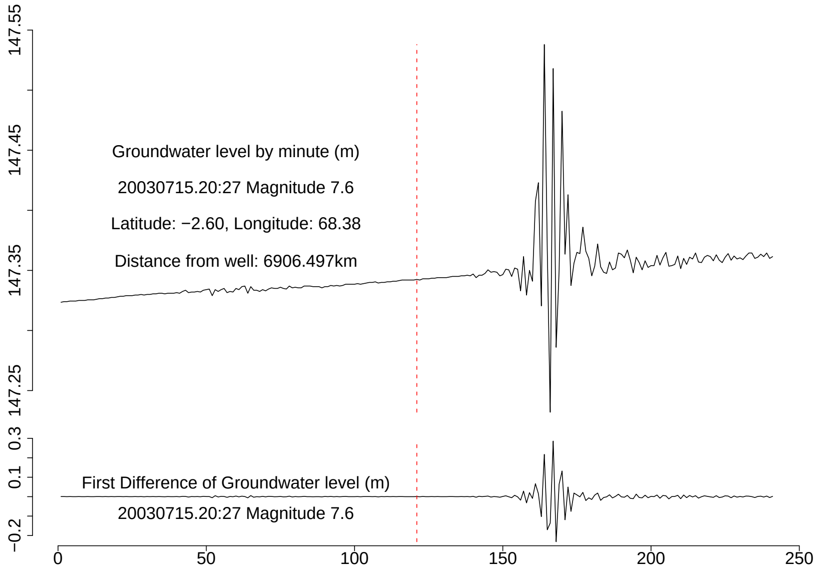

FIGURE 1. *An example of the fluctuations of the groundwater level at Tangshan Well. The dashed line indicates the origin time of the magnitude 7.6 earthquake on the Carlsberg Ridge, on July 15th, 2003.*

level changes are caused not only by tectonic factors, such as earth crust deformation, related to earthquakes; but also by non-tectonic factors, such as rainfall, air pressure changes, earth tides, exploitation of mines and underground water. To eliminate 'slow' (or non-tectonic) factors such as air pressure, rainfall, earth tides and pumping, we examine the first differences of the groundwater level data.

After cleaning the abnormal spikes of the data which were caused by electrical failure, sensor adjustment and malfunction, 1,971,360 data from January 2002 to September 2005 remain to be investigated, including 88335 missing data, in the data set of the first differences of groundwater level. Wang et al. (2008) extracted 754 *well signals*, representing the times when well fluctuations are initiated, from this data set.

We extracted the earthquakes with minimum magnitude 5.8 from the USGS-NEIC catalogue (http://wwwneic.cr.usgs.gov/neis/epic/epic_global.html) from January 2002 to September 2005. The number of such earthquakes is 926, the occurrence times of which will be used to analyze the correlation between the groundwater level fluctuations and the earthquakes. The cumulative number of events versus time plots for the earthquakes and well signals are shown in Figure 2.

4. Data analysis

Figure 3 shows the spectra of the well signal sequence and the earthquake occurrence sequence, and the coherence between them, using disjoint intervals of length 32 hours. We corrected the second moment by subtracting the mean. The graph shows strong association between the first difference signals and the earthquake occurrences at low frequencies, apparent association up to about 6hr^{-1} or more than 10 minutes. The figure also shows strong association near 20 and 40hr^{-1} (corresponding to 3 or 1.5 minutes) as well.

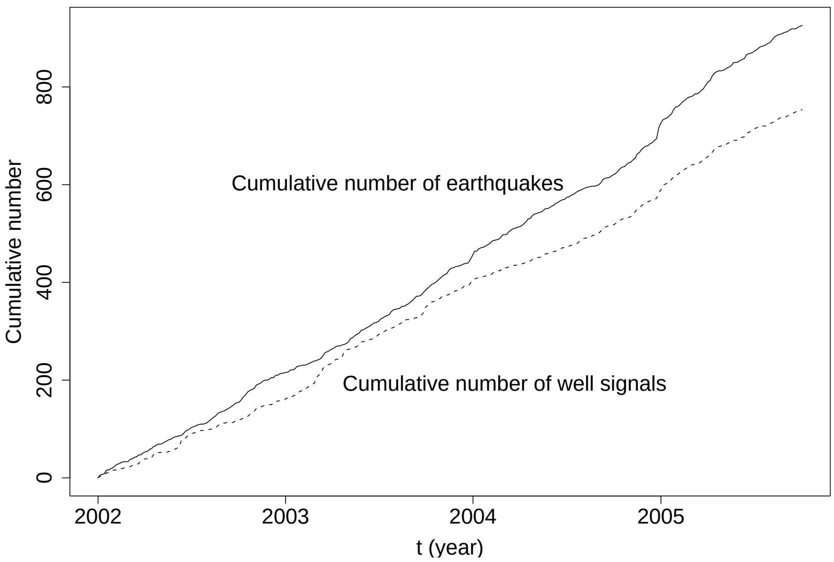

FIGURE 2. *The cumulative number of events plots for earthquakes (solid) and well signals (dashed).*

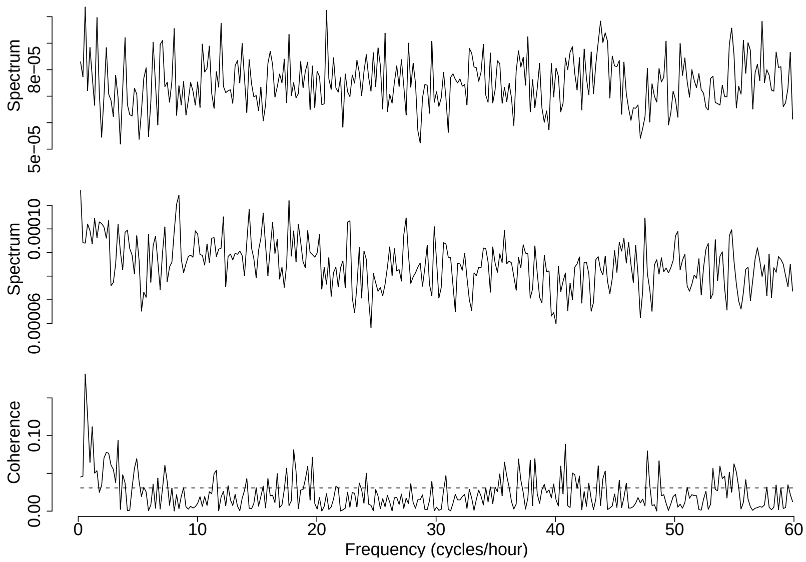

FIGURE 3. *Spectra and coherence plots for well signals and earthquake occurrence times with disjoint intervals each having a length of 32 hours. The dashed line shows the upper level of the approximate* 95% *confidence interval under the hypothesis that the two processes are independent.*

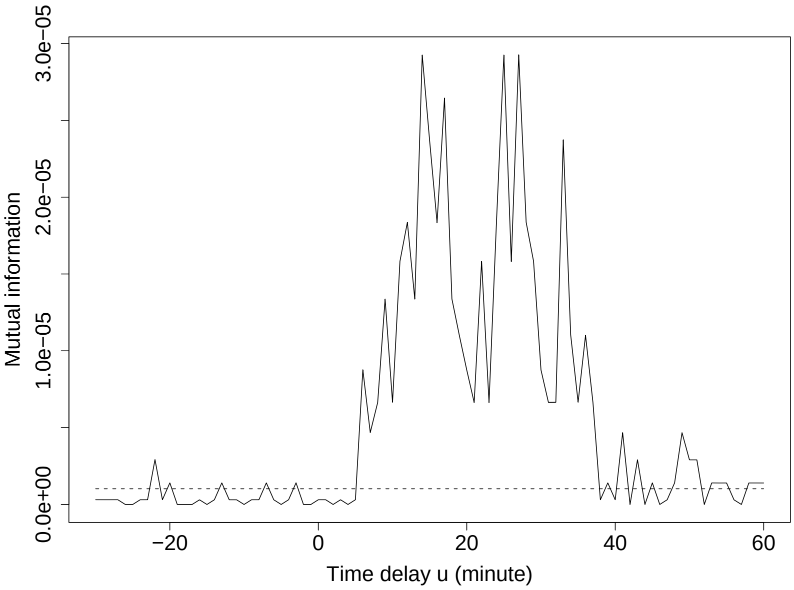

FIGURE 4. *Mutual information for well signals and earthquake occurrence times. The dashed line shows the upper level of the approximate* 95% *confidence interval under the hypothesis that the two processes are independent.*

In order to consider the mutual information between well signals, denoted by U, and the earthquake occurrence times, denoted by V, we first transform the well signal series and the earthquake series into 0-1 sequences. A 1 indicates the beginning of the well signal, 0 otherwise. The earthquake series has a 1 in the minute of the earthquake occurrence, 0 otherwise. We examine the mutual information between two series $\{U_{t+u}\}$ and $\{V_t\}$, where u is the time lag. The mutual information results are shown in Figure 4. The figure indicates that the earthquake occurrences appear to induce the occurrences of the water level fluctuations with a time delay, which is due to the travel times of the earthquake waves from the hypocenter to the well.

Since we are interested in whether or not the earthquakes have induced the water level fluctuations, we treat the well signal sequence as the output process and the earthquake occurrence times as the input process. The causal relation between the well signals and the earthquake occurrences can then be considered using the Lin-Lin model (16). The orders of the response functions were determined to be $K = L = 4$. The estimated response functions for the self-exciting part and for the earthquake occurrence input are shown in Figure 5. The self-exciting part is small compared to the mutual-exciting part, which clearly shows that the earthquake occurrences do play a part in inducing the water level fluctuations at Tangshan Well.

5. Discussion

We note that the three methods discussed in this work examine different types of association. When the problem of whether or not we can predict an output series from some input series through a linear relation is considered, coherence may be a very useful approach to use. It plays a diagnostic role in detecting association

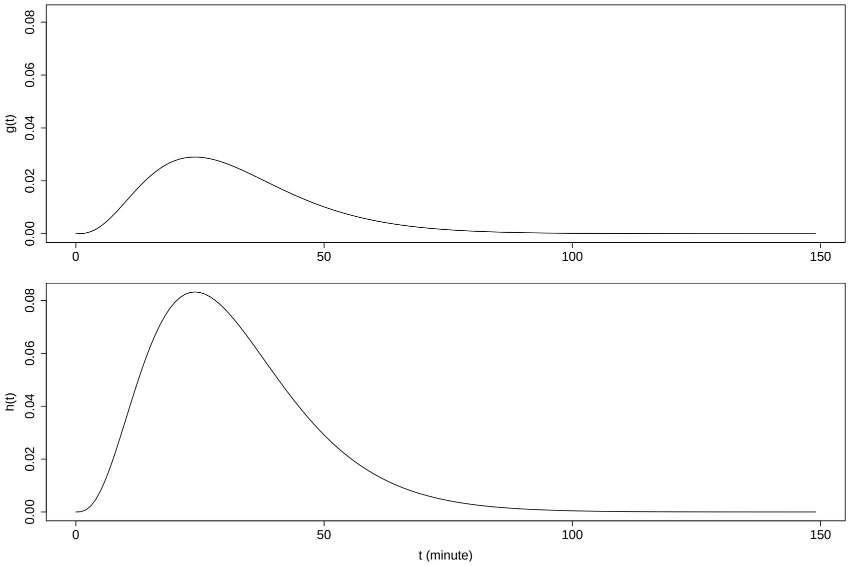

FIGURE 5. *Estimated response functions (top: self-exciting part; bottom: earthquake occurrence input) of well signals.*

between two processes at certain frequency range. The estimation of coherence using the method of disjoint sections can work around missing data, which appears to be an advantage of using coherence.

However, the coherence may be identically 0 when two processes are actually correlated. The coefficient of mutual information overcomes this drawback and takes on the value 0 if and only if the two processes under consideration are statistically independent and thus forms a test of dependence. As can be seen from Figure 3 and 4, it also presents a far more detailed conclusion. Another important advantage of mutual information is that it does not really pay any attention to the values of the processes. Mutual information can tell us more than that the hypothesis of independence is rejected, it measures the strength of the correlation between two series of events. The drawback of the mutual information may be that it does not set up naturally for point processes, as we can see from the formulation as well as the application. But we can transform a point process into a time series which is suitable or convenient for mutual information analysis.

When we are interested in more than whether the two processes are correlated or not, or in other words, when we want to make clear which process is the driven process, the coherence and mutual information analyses become inadequate. The Lin-Lin model can identify cases in which either one of the two processes causes the other, or they cause each other, or some other process causes both of them. Unlike the generally defined coherence and mutual information, this model was designed specifically for point processes. The mutual-exciting term being identically 0 implies that there is no causal relationship between the two processes; whereas the self-exciting term being identically 0 while the mutual-exciting term being none zero means that the intensity of the output process is modulated only by the input process.

The application of the three methods to the groundwater level data from Tangshan Well and the global earthquakes shows that the well signals and earthquake occurrences have significant linear correlation, especially at lower frequencies. These lower frequencies may be related to the travel times from earthquakes to the well, and perhaps to the time necessary to set up fluctuations in the groundwater level. The mutual information analysis reveals that the fluctuations of the groundwater level are strongly associated with earthquake occurrences, which appear to trigger water level fluctuations. The Lin-Lin model then confirms that earthquake occurrences are likely to be inducing water level fluctuations at Tangshan Well.

Acknowledgments

We sincerely thank Professor David Vere-Jones for valuable discussions on the interpretation of the results, and for reading and commenting on an earlier version of this paper, with detailed criticisms which have helped improve the paper greatly. We are grateful to Tangshan Earthquake Administration for providing the well data. Helpful suggestions from Baojun Yin and Jianping Huang on the well data are greatly appreciated. This work was supported by the Marsden Fund, administered by the Royal Society of New Zealand.

References

[1] Akaike, H., (1974) 'A new look at the statistical model identification', *IEEE Trans. Autom. Control*, Vol. 19, pp. 716-723.

[2] Antos, A. & Kontoyiannis, Y., (2001) 'Convergence properties of functional estimates for discrete distributions', *Random Structures and Algorithms*, Vol. 19, pp. 163-193.

[3] Brillinger, D. R., (1975) 'The Identification of Point Process Systems', *The Annals of Probability*, Vol. 3, No. 6, pp. 909-929.

[4] Brillinger, D. R., (1994) 'Examples of scientific problems and data analyses in demography, neurophysiology and seismology', *J. Computational and Graphical Statistics*, Vol. 3, pp. 239-249.

[5] Brillinger, D. R., (2001) *Time Series: Data Analysis and Theory,* SIAM, Philadelphia, USA.

[6] Brillinger, D. R., (2002) 'Second-order Moments and Mutual Information in the Analysis of Time Series', in *Recent Advances in Statistical Methods.* Imperial College Press, London, pp. 64-76.

[7] Brillinger, D. R., (2003) 'Modelling and analysis of some random process data from neurophysiology', *Investigacion Operacional*, Vol. 24, pp. 206-218.

[8] Brillinger, D. R., (2004) 'Some data analyses using mutual information', *Brazilian J. Probab. Statist.*, Vol. 18, pp. 163-183.

[9] Brillinger, D. R., Bryant, H. L. & Segundo, J. P., (1976) 'Identification of synaptic interactions', *Biological Cybernetics*, Vol. 22, pp. 213-228.

[10] Brillinger D. R. & Guha A., (2007) 'Mutual information in the frequency domain', *Journal of statistical planning and inference*, Vol. 137, pp. 1076-1084.

[11] Brillinger, D. R. & Villa, A. E. P., (1997) 'Assessing Connections in Networks of Biological Neurons', in *The Practice of Data Analysis: Essays in Honor of John W. Tukey*, pp. 77-92, ed. D. R. Brillinger, L. T. Fernholz and S. Morgenthaler, Princeton Univ. Press, Princeton.

[12] Christensen, R., (1997) *Log-linear Models for Logistic Regression,* Springer, New York.

[13] Cover, T. M. & Thomas, J. A., (1991) *Elements of Information Theory,* New York: Wiley.

[14] Granger, C. W. J., Lin, J.-L., (1994) 'Using the mutual information coefficient to identify lags in nonlinear models', *J. Time Series Anal*, Vol. 15, pp. 371-384.

[15] Halliday, D. M., Rosenberg, J. R., Amjad, A. M., Breeze, P., Conway, B. A. & Farmer, S.F., (1995) 'A framework for the analysis of mixed time series/point process data–Theory and application to the study of physiological tremor, single motor unit discharges and electromyograms', *Progress in Biophysics and molecular Biology*, Vol. 64, No. 2/3, pp. 237-278.

[16] Hawkes, A. G., (1971) 'Spectra of some self-exciting and mutually exciting point processes', *Biometrika*, Vol. 58, pp. 83-90.

[17] Iyengar, S., (2001) 'The analysis of multiple neural spike trains', In *Advances in Methodological and Applied Aspects of Probability and Statistics*, pp. 507-524, ed. N. Balakrishnan, Gordon and Breach.

[18] Joe, H., (1989) 'Estimation of entropy and other functionals of a multivariate density', *Ann. Inst. Statist. Math.*, Vol. 41, pp. 683-697.

[19] Ma, L. & Vere-Jones, D., (1997) 'Application of M8 and Lin-Lin algorithms to New Zealand earthquake data', *New Zealand Journal of Geology and Geophysics*, Vol. 40, pp. 77-89.

[20] Mars, N. J. I., van Aragon, G. W., (1982) 'Time delay estimation in non-linear systems using average amount of mutual information analysis', *Signal Process.*, Vol. 4, pp. 139-153.

[21] Moddemeijer, R., (1989) 'On estimation of entropy and mutual information of continuous distributions', *Signal Processing*, Vol. 16, pp. 233-248.

[22] Moon, Y. I., Rajagopalan, B., Lall, U., (1995) 'Estimation of mutual information using kernel density estimators', *Phys. Rev. E*, Vol. 52, pp. 2318-2321.

[23] Ogata, Y., (1983) 'Likelihood analysis of point processes and its applications to seismological data', *Bull. Int. Statist. Inst.*, Vol. 50, Book 2, pp. 943-961.

[24] Ogata, Y., (1999) 'Seismicity analysis through point-process modeling: A review', *Pue Appl. Geophys.*, Vol. 155, pp. 471-507.

[25] Ogata, Y., Akaike, H. & Katsura, K., (1982) 'The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process', *Ann. Inst. Statist. Math.*, Vol. 34, Part B, pp. 373-387.

[26] Shannon, C. E., (1948) 'A mathematical theory of communication', *Bell System Tech. J.*, Vol. 27, pp. 379-423, 623-656.

[27] Strong, S. P., Koberle, R., Steveninck, R. R. de Ruyter van & Bialek, W., (1998) 'Entropy and information in neural spike trains', *Phys. Rev. Lett.*, Vol. 80, pp. 197-200.

[28] Vere-Jones, D. & Ozaki, T., (1982) 'Some examples of statistical estimation applied to earthquake data', *Ann. Inst. Statist. Math.*, Vol. 34, Part B, pp. 189-207.

[29] Wang, T., Bebbington, M., Vere-Jones, D., Harte, D. & Ma, L., (2008) 'Coseismic Responses for Global Teleseismics at Tangshan Well', Submitted.

IFS/Statistics, Massey University, Private Bag 11222, Palmerston North, New Zealand.
E-mail: t.wang@massey.ac.nz and m.bebbington@massey.ac.nz

Statistics Research Associates Ltd, Wellington, New Zealand
E-mail: david@statsresearch.co.nz