Res. Lett. Inf. Math. Sci., (2000) **1**, 119-127
Available online at http://www.massey.ac.nz/~wwiims/rlims/

A renewal cluster model for the inter-arrival times of rainfall events*

Paul S.P. Cowpertwait
I.I.M.S., Massey University Albany Campus, Auckland, N.Z.
P.S.Cowpertwait@massey.ac.nz

A statistical model, based on a renewal cluster point process, is proposed and used to infer the distributional properties of dry periods in a continuous-time record. The model incorporates a mixed probability distribution in which inter-arrival times are classified into two distinct types, representing cyclonic and anticyclonic weather. This results in rainfall events being clustered in time, and enables objective probabilistic statements to be made about storm properties, e.g. the expected number of events in a storm cluster. The model is fitted to data taken from a gauge near Wellington, New Zealand, by maximising the likelihood function with respect to the parameters. The Akaike Information Criteria is used to select the best fitting distributions from a range of candidates. The log-Normal distribution is found to provide the best fit to the times between successive storm clusters, whilst the Weibull distribution is found to provide the best fit to the times between successive events in the same storm cluster. Harmonic curves are used to provide a parsimonious parameterisation, allowing for the seasonal variation in precipitation. Under the fitted model, the interval series is transformed into a residual series, which is assessed to determine overall goodness-of-fit.

1. Introduction

Various types of automatic rain gauges are available for recording data in continuous-time. For example, a digitised tipping-bucket gauge will automatically record the tipping times of a 0.2mm bucket. Some of these gauges record changes in rainfall intensity using a pluviograph trace on a rotating drum, where zero rain is recorded with horizontal lines and high intensity rain as steep gradients (Samson, 1992). Data from automatic gauges are usually digitised into hourly or daily series, which can be fitted using discrete-time stochastic models, or using derived moments of continuous-time stochastic models (Cowpertwait, 1994, 1998). In this paper, we analyse a digitised pluviograph record, which contains the starting and finishing times of rainfall events over a 41-year continuous-time record.

Previous studies have provided empirical evidence that rainfall events cluster in time (e.g. Cowpertwait 1994, 1998). However, most of these studies use stochastic models that are fitted to discrete-time data; models which are usually unsuitable for modelling the continuous-time process. Thus, it seems appropriate to postulate a model for the analysis of continuous-time data, which also incorporates clustering. This is achieved here by using a mixed probability density function in which inter-arrival times are classified into two distinct types, representing cyclonic and anti-cyclonic weather. The methodology has the advantage in that it enables formal statistical inferential methods to be used in model fitting and selection.

The model represents a point process of inter-arrival times, i.e. event depths and durations are not explicitly modelled in this paper. There are many examples of depth-duration analyses of rainfall data that could be combined with the fitted model for use in hydrologic simulation studies. For example, see Samson and Thomson 1992 for a continuous-time analysis of pluviograph data. For a general review of stochastic rainfall modelling see Foufoula-Georgiou and Krajewski (1995), or, for applications in hydrology, refer to O'Connell and Todini (1996).

The paper is organised as follows. In Section 2, the renewal cluster model is formulated and mathematical properties are given. The fitting procedure and inferential methodology are discussed in Section 3. In Section 4, the model is fitted to data from Wellington, New Zealand. The adequacy of fit is discussed in Section 5, using residual errors for the fitted model. Finally, some overall conclusions are given in Section 6.

2. Model Formulation

---

* Due to appear in the International Journal of Climatology, a journal of the Royal Meteorological Society

Let $Y(u) \geq 0$ be a random variable representing rainfall intensity at time u ($-\infty < u < \infty$) so that the rainfall depth over an arbitrary time interval $[a, b]$ is given by: $\int_a^b Y(v)dv$. A rainfall *event* consists of non-zero values of $Y(u)$ immediately preceded and followed by a zero intensity, so that any event has a starting time t, a lifetime l, and a mean intensity z, where $z = l^{-1}\int_t^{t+l} Y(v)dv$. Thus, if an event starts at time t and finishes at time $t+l$, then $Y(u) > 0$ for all u in $[t, t+l]$, and $Y(t-\varepsilon) = Y(t+l+\varepsilon) = 0$ for some arbitrarily small $\varepsilon > 0$.

Consider a stochastic point process $\{t_i\}$ representing the starting times of rainfall events in a time interval [0, T], where each event has a lifetime l_i, $i = 1, \ldots, N$. Let $X_i = t_i - t_{i-1} - l_{i-1}$ be a random variable representing the ith dry period or inter-arrival time between two successive events (taking $t_0 = l_0 = 0$; $i = 1, \ldots, N$), and suppose each X_i is independently marked as '*type* 1' or '*type* 2', where type 1 inter-arrivals represent atmospheric conditions suitable for precipitation (cyclone or frontal weather), whilst type 2 inter-arrivals represent conditions unsuitable for rain (high pressure or anticyclone). Let $\langle X_i \rangle$ denote the mark associated with the ith interval X_i, and let p be the probability that a dry interval chosen at random is of type 2, i.e. $p = P(\langle X_i \rangle = 2) = 1 - P(\langle X_i \rangle = 1)$, $i = 1, \ldots, N$.

The marks $\langle X_i \rangle$ form a stochastic process, for example {2112122111} is a possible realisation when N = 10. Using the associated marks, an inter-arrival process $\{X_i\}$ can be broken down into sequences of clusters, where a cluster of size C is defined to be a sequence of inter-arrival times beginning with a type 2 inter-arrival time followed by C-1 type 1 inter-arrivals. The random variable C follows a Geometric distribution with mean $\mu_C = p^{-1}$, and probability function: $P(C = j) = p(1-p)^{j-1}$, for $j = 1, 2, 3, \ldots$. Therefore, provided $\mu_C > 1$, the inter-arrival process $\{X_i\}$ forms a cluster point process. In the example above, the realisation {2112122111} contains four clusters represented by: {211}, {21}, {2} and {2111}, with C taking the values 3, 2, 1 and 4 respectively. In this example, the clusters of inter-arrival times are: $\{X_1, X_2, X_3\}$, $\{X_4, X_5\}$, $\{X_6\}$, and $\{X_7, X_8, X_9, X_{10}\}$.

We may also define a 'storm' to be a cluster of C rainfall events, where each event in the storm has a starting time determined by a cluster of inter-arrival times $\{X_i: i = k, \ldots, k-1+C\}$ and the duration process $\{l_i: i = k, \ldots, k-1+C\}$, so that starting times in the storm are given by: $t_i = X_i + t_{i-1} + l_{i-1}$, $i = k, \ldots, k-1+C$, where t_{k-1} and l_{k-1} are the starting time and lifetime of the last event in the preceding storm. In the previous example, the cluster of inter-arrival times $\{X_7, X_8, X_9, X_{10}\}$, where $k = 7$ and $C = 4$, gives a storm of four rainfall events with arrival times: $t_7 = X_7 + t_6 + l_6$, $t_8 = X_8 + t_7 + l_7$, $t_9 = X_9 + t_8 + l_8$, $t_{10} = X_{10} + t_9 + l_9$.

Let f_1 be the probability density function (PDF) for a type 1 inter-arrival time and f_2 be the PDF for a type 2 inter-arrival time. Then, the probability density function (PDF) for X_i ($i = 1, \ldots, N$) is given by:

$$g(x) = (1 - \mu_c^{-1})\, f_1(x) + \mu_c^{-1}\, f_2(x) \qquad (1)$$

Thus, the $\{X_i\}$ form a series of independent identically distributed random variables with PDF (1). Hence, the series $\{X_i\}$ is essentially a renewal process using the mixed density (1) to give clusters of rainfall events, i.e. $\{X_i\}$ is a 'Renewal Cluster Process'.

3. Fitting Procedure and Inference

Let $\{x_i; i = 1, \ldots, N\}$ be a series of observed inter-arrival times. From (1), the log-likelihood function is given by:

$$LL = \sum_{i=1}^{N} \ln g(x_i) = \sum_{i=1}^{N} \ln\left\{(1 - \mu_c^{-1})\, f_1(x_i) + \mu_c^{-1}\, f_2(x_i)\right\} \qquad (2)$$

To fit the model, some distributions need to be postulated for f_1 and f_2. The following were considered as they represent a wide range of positive-valued random variables:

A. Exponential: $f(x) = e^{-x/\alpha}/\alpha$

B. Gamma: $f(x) = x^{\beta-1}e^{-x/\alpha} / \left\{\Gamma(\beta)\alpha^{\beta}\right\}$

C. Weibull: $f(x) = \beta x^{\beta-1}e^{-x^{\beta}/\alpha^{\beta}} / \alpha^{\beta}$

D. Log-Normal: $f(x) = \exp\left\{-\frac{1}{2\beta^2}(\ln x - \alpha)^2\right\} / \left(x\beta\sqrt{2\pi}\right)$

The model parameters to be estimated include the mean cluster size (μ_c), and the scale and shape parameters (α_j, β_j; j = 1, 2) for each type of inter-arrival time. For each combination of f_1 and f_2 (A-D above), the PDF (1) can be fitted by maximising the log-likelihood (2) with respect to the parameters. The Akaike Information Criteria (AIC = -2×LL + 2×number of parameters, Akaike 1974) can be used to choose the best distributions for f_1 and f_2 from A-D above, i.e. a distribution for (1) that gives the best fit to the data.

Using the mixed distribution (1) enables objective inferences to be made about the statistical properties of events within and between storms, for example the mean cluster size μ_c can be estimated without having to subjectively separate storms in the data. The fitted model can also be used to estimate a conditional probability that two successive events are within the same storm given the observed time between the events. This is the probability that the inter-arrival time is of type 1 given an observed inter-arrival time x, and is given by:

$$(1 - \mu_c^{-1}) f_1(x)/g(x) \tag{3}$$

4. Analysis of Data

A continuous-time record of rainfall data (Kelburn, near Wellington; 41.283°S, 174.767°E; 1955-95) was provided by the New Zealand National Institute of Water and Atmospheric Research (NIWA) for use in this study. This record was the longest complete data set available containing starting and finishing times of events in continuous-time. To describe the various fitting procedures, and data analysis, it is helpful to adopt the following notation.

Let N_i be the number of event starting times that occur in the ith year and N_{ij} be the number of event starting times that occur in the jth month of the ith year respectively, so $\sum_j N_{ij} = N_i$ (i = 1, …, 41; j = 1, …, 12). Furthermore, let the starting time and lifetime (in hours) of the kth event starting in the jth month of the ith year be t_{ijk} and l_{ijk} respectively (measured relative to the starting time of the record) and let $x_{ijk} = t_{ijk} - t_{ij,k-1} - l_{ij,k-1}$ (i = 1, 2, …, 41; j = 1, 2, …, 12; k = 1, …, N_{ij}). Note that whilst t_{ijk} must be in the jth month, it is possible for $t_{ijk} + l_{ijk}$ to be in the (j+1)th month for events that overlap two adjacent months. The total number of events N in the 41-year record was 24560, i.e. $N = \sum_i \sum_j N_{ij} = \sum_i N_i = 24560$.

To ensure that long dry intervals would be included when fitting the model, some care was needed when choosing values for $t_{i,j,0}$ and $l_{i,j,0}$. When $i = j = 1$ (January of the first year), these were taken to be zero, i.e. $t_{1,1,0} = l_{1,1,0} = 0$. Otherwise they were taken to be $t_{i,j,0} = t_{i,j-1,N_{i,j-1}}$ ($j > 1$) or $t_{i,1,0} = t_{i-1,12,N_{i-1,12}}$ ($j = 1$), corresponding to the starting time of the last event in the previous month, and $l_{i,j,0} = l_{i,j-1,N_{i,j-1}}$ ($j > 1$) or $l_{i,1,0} = l_{i-1,12,N_{i-1,12}}$ ($i > 1$, $j = 1$), the lifetime of the last event with starting time in the previous month. For months in which the last event did not overlap the next month (but excluding December of the last year), a total of $N_{ij} + 1$ inter-arrival times were considered, where the last interval was taken to be $x_{i,j,N_{ij}+1} = t_{i,j+1,1} - t_{i,j,N_{ij}} - l_{i,j,N_{ij}}$ (i = 1, 2, …, 41; j = 1, 2, …, 11, excluding i = 41 and j = 12, with the obvious adaptation for j = 12 and i < 41) . Thus, inter-arrival times spanning two adjacent months are included in the fitting for both months.

The parameter estimates for each calendar month j were thus obtained by maximising the following:

$$LL_j = \sum_{i=1}^{41} \sum_{k=1}^{N_{ij}+1} \ln g(x_{i,j,k}) \tag{4}$$

where each combinations of A-D for f_1 and f_2 was used in (4), j = 1, …, 12. To ensure a unique solution exists for each combination of f_1 and f_2, the minimisation is subject to: $\mu_2 > \mu_1$, where μ_1 and μ_2 are the mean type 1 and type 2 inter-arrival times respectively. The maximisation of (4) was carried out numerically using the Simplex method (Nelder and Mead 1965), implemented on a micro-computer

using the algorithm by O'Neil (1985). This produced 12 estimates of each parameter for each fitted distribution (1).

Table 1 summarises the log-likelihood and AIC values for each fitted distribution of f_1 and f_2, where the log-likelihoods LL are summed over the months, i.e. $LL = \sum_{j=1}^{12} LL_j$. The overall best fitting distributions are the Log-Normal distribution (D) for f_1 and the Weibull distribution (C) for f_2 (Table1).

TABLE 1: Log-likelihood and AIC values for f_1 and f_2 (given to 4 significant figures)

f_1	F_2	-LL	AIC
A	A	65090	130300
A	B	64870	129800
A	C	64780	129600
A	D	64730	129600
B	A	65080	130200
B	B	64790	129700
B	C	64630	129400
B	D	64310	128700
C	A	65060	130200
C	B	64860	129800
C	C	64740	129600
C	D	64450	129000
D	A	63790	127700
D	B	63720	127600
D	C	63710	127500
D	D	63850	127800

The monthly parameter estimates for the best fitting models are plotted in Figures 1-5, where it can be seen that the estimates reflect some well-known observed seasonal changes in precipitation. For example, over (Southern Hemisphere) winter months there are more storms on average, corresponding to a decrease in the scale parameter in Figure 4. In addition, the mean cluster size increases during Winter months which represents an increase in frontal weather systems (Figure 1). Also, note that the mean cluster size is always greater than one, which provides statistical evidence that rainfall events are clustered in time.

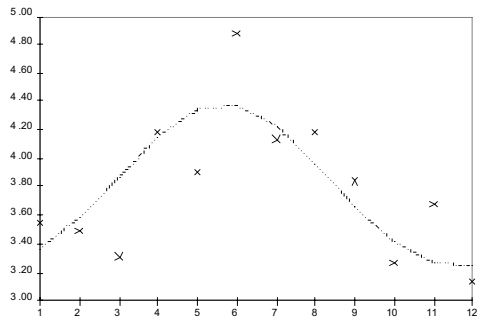

Figure 1: Estimates of μ_c; fitted values (×) for each month and the fitted harmonic equation 5 (dotted curve).

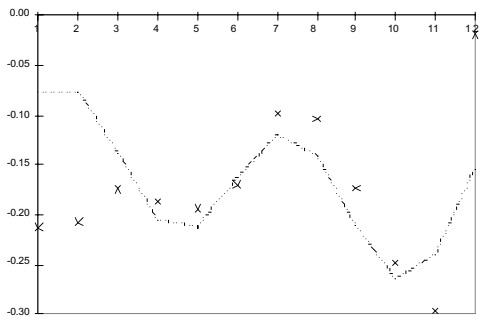

Figure 2: Estimates of α_1; fitted values (×) for each month and the fitted harmonic equation 6 (dotted curve).

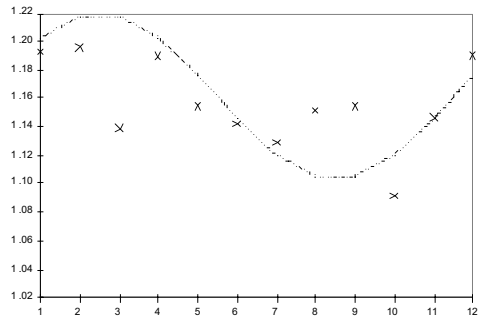

Figure 3: Estimates of β_1; fitted values (×) for each month and the fitted harmonic equation 7 (dotted curve).

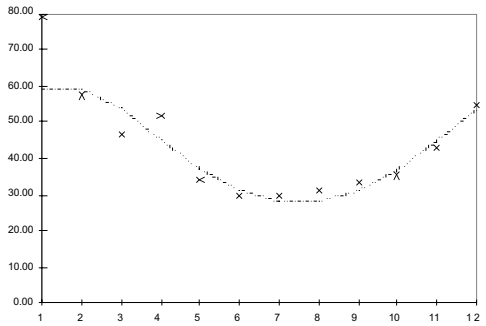

Figure 4: Estimates of α_2; fitted values (×) for each month and the fitted harmonic equation 8 (dotted curve).

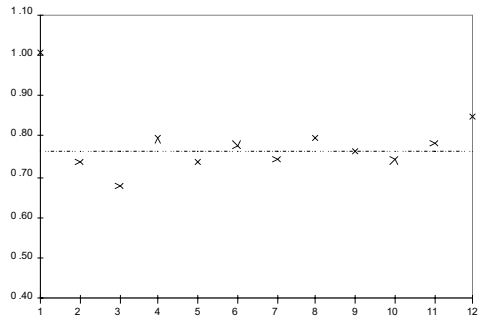

Figure 5: Estimates of β_2; fitted values (×) for each month and the fitted harmonic equation 9 (dotted line).

The seasonal variation in the parameter estimates suggested it might be reasonable to use harmonic curves for the best fitting distributions. This reduces the number of estimates and provides a smooth transition over a year, avoiding discontinuities between adjacent months. The adequacy of this approach can be tested using AIC.

A single harmonic wave seemed appropriate for the monthly estimates in Figures 1, 3, and 4, whilst the more complex seasonal pattern in Figure 2 suggested a second-order harmonic was needed. In Figure 5, the monthly estimates appear to follow no seasonal pattern, which suggested no harmonic wave was needed and the parameter estimates can be treated as constant throughout the year. We thus considered the following equations for the model parameters:

$$\mu_c(t) = \exp\left\{m_c + A_c \sin\left(2\pi t / T + 2\pi/(1+e^{\theta_c})\right)\right\} \quad (5)$$

$$\alpha_1(t) = m_{\alpha_1} + A_{\alpha_1} \sin\left(2\pi t / T + 2\pi/(1+e^{\theta_{\alpha_1}})\right) + B_{\alpha_1} \sin\left(2\pi t / T + 2\pi/(1+e^{\phi_{\alpha_1}})\right) \quad (6)$$

$$\beta_1(t) = \exp\left\{m_{\beta_1} + A_{\beta_1} \sin\left(2\pi t / T + 2\pi/(1+e^{\theta_{\beta_1}})\right)\right\} \quad (7)$$

$$\alpha_2(t) = \exp\left\{m_{\alpha_2} + A_{\alpha_2} \sin\left(2\pi t / T + 2\pi/(1+e^{\theta_{\alpha_2}})\right)\right\} \quad (8)$$

$$\beta_2(t) = \exp\left\{m_{\beta_2}\right\} \quad (9)$$

where t is time of year (in hours), T is the total number of hours in the year (T = 8784 for leap years; otherwise T = 8760). The exponential functions exp{.} are used to ensure the estimates take positive values, which is essential for the distribution parameters in (5), (7)-(9). The logistic functions, which

take the form $2\pi/(1+e^{\theta})$, are used to ensure all angles lie between 0 and 2π radians. Under this re-parameterisation, the density function of the (i+1)th interval between the ith and (i+1)th event depends on the finishing time $t_i + l_i$ of the ith event, so that the log-likelihood function takes the form:

$$LL = \sum_{i=1}^{N} \ln\left\{\left(1-\mu_c^{-1}(t_{i-1}+l_{i-1})\right) f_1(t_{i-1}+l_{i-1}, x_i) + \mu_c^{-1}(t_{i-1}+l_{i-1})\, f_2(t_{i-1}+l_{i-1}, x_i)\right\} \qquad (10)$$

where $t_0 = l_0 = 0$ and N is the number of events in the 41-year record (analogous to the monthly intervals, the first interval in all years after 1955 was obtained by measuring from the last event in the preceding year).

The mean type 1 and type 2 inter-arrival times (μ_1 and μ_2) come directly from the expressions for the Log-Normal and Weibull random variables, and are given by:

$$\mu_1(t) = \exp\left\{\alpha_1(t) + \tfrac{1}{2}\beta_1^2(t)\right\} \qquad (11)$$

$$\mu_2(t) = \alpha_2(t)\Gamma\left(1+\beta_2^{-1}(t)\right) \qquad (12)$$

where Γ denotes the Gamma function.

The harmonic parameters on the right hand side of equations 5-9 were estimated by maximising the likelihood function and are given in Table 2 and shown as dotted lines in Figures 1 to 5. For the harmonic parameterisation, the AIC was 127100, which is less than the AIC for all the models in Table 1. Thus, the reduction in the number of parameters is well justified, and the harmonic estimates can be used in preference to the monthly values.

TABLE 2: Harmonic Parameter Estimates and their Standard Errors (in parenthesis)

Parameter	Estimate
m_c	1.33 (0.061)
A_c	-0.151 (0.036)
θ_c	0.744 (0.24)
m_{α_1}	-0.167 (0.062)
A_{α_1}	0.0354 (0.0077)
θ_{α_1}	2.79 (0.36)
B_{α_1}	0.0739 (0.016)
ϕ_{α_1}	2.14 (0.85)
m_{β_1}	0.149 (0.016)
A_{β_1}	0.0498 (0.0051)
θ_{β_1}	2.39 (0.40)
m_{α_2}	3.70 (0.041)
A_{α_2}	0.387 (0.047)
θ_{α_2}	1.61 (0.39)
m_{β_2}	-0.271 (0.018)

The estimated harmonic coefficients were used in equations 5-12 to determine how the mean cluster size μ_C, and the mean inter-arrival times (μ_1 and μ_2) varied over the year (Figures 6, 7, and 8). Again, the figures reflect some well-known seasonal properties of rainfall. For example, on average, storms are more frequent and contain larger clusters during the winter months, which is characteristic of frontal weather (Figures 6 and 8). Due to equation 6, a more complex seasonal pattern is evident in Figure 7 for the expected time between successive events in the same storm system, with a tendency for winter events to be clustered closer together.

In Figure 9, the approximate probability that two successive events come from the same storm system, as a function of their temporal separation, is plotted, using expression (3). For any temporal separation, summer events have a slightly higher probability of coming from the same system compared

with winter events, which is due to summer storms being less frequent (Figure 9). In the absence of other meteorological information, the model predicts that two consecutive events, which are separated by less than about 8 hours, are more likely to belong to the same storm cluster (probability > 0.5). Conversely, events separated by more than about 8 hours are more likely to belong to a different storm cluster (Figure 9).

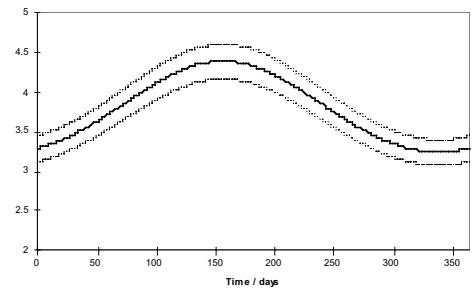

Figure 6: The mean cluster size μ_C (± 1 SE) against time of year.

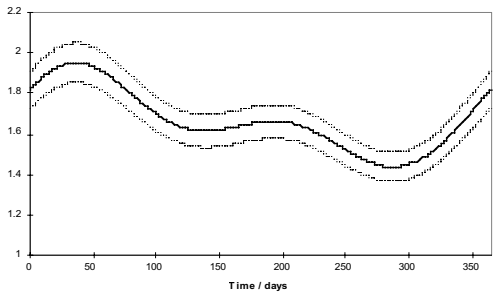

Figure 7: The mean inter-arrival time μ_1 (± 1 SE; in hours) between successive events in the same storm cluster plotted against time.

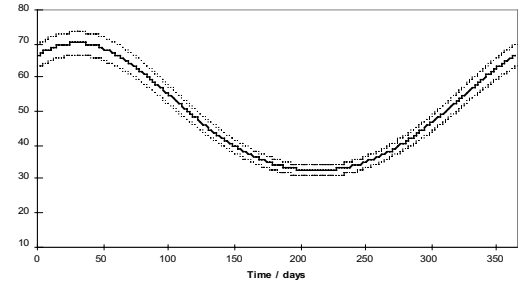

Figure 8: The mean inter-arrival time μ_2 (± 1 SE; in hours) between successive storm clusters plotted against time.

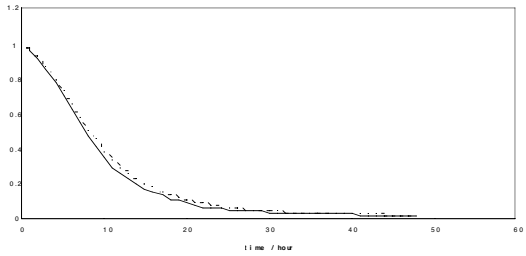

Figure 9: The approximate probability (*y*-axis) that two successive events come from the same storm cluster plotted against their temporal separation (*x*-axis). The dotted line is for winter events (taken as the midpoint in the year, i.e. with $t = 4380 = \frac{1}{2} \times 365 \times 24$ in equations 5-9), whilst the solid line is for summer events (taken as $t = 0$ in equations 5-9).

5. Residual Analysis

Having obtained the best fitting model out of those considered, we now move on to the problem of assessing goodness-of-fit and determining whether a better fitting model is likely to exist. This is achieved via a general analysis of residuals (e.g. see Cox and Snell, 1968), where the 'residuals' in the present context are defined as follows.

As before, let $x_i = t_i - t_{i-1} - l_{i-1}$ ($i = 1, \ldots, N$) be the observed times between successive events, and let $\hat{G}(x) = P(X \le x)$ be the fitted distribution function of the times X between successive events. We define the ith residual r_i to be:

$$r_i = -\ln\left\{ 1 - \hat{G}(x_i) \right\} \qquad (13)$$

Under the above transformation, the residuals will be a series of independent standard exponential random variables, provided the model adequately fits the data. It is sometimes more convenient to work with $u_i = \hat{G}(x_i)$ which forms a series of independent uniform random variables over the interval (0, 1), again assuming the model adequately fits the data. Departures from these distributions indicate lack-of-fit and may suggest another model is more appropriate. Appropriate tests include assessing goodness-of-fit to the exponential and uniform distributions and looking for lack of independence (e.g. serial correlation) in the residuals. Ogata (1988) gave a similar analysis in the context of earthquake modelling.

Plots for the residuals and transformed residuals are given in Figures 10-13. The cumulative distribution plot (Figure 10) indicates that the residuals are very close to exponential, because they lie approximately on a straight line of unit slope, and that an overall good fit has been obtained. However, discrepancies in the upper tail are not readily seen in this plot, so quantiles were also plotted and are shown in Figure 11. Some discrepancies are evident in the upper 1% tail, which implies the fitted model will under-predict extreme dry periods (Figure 11). For applications in which a good fit to the extremes is important, the model may therefore need to be modified, but this was beyond the scope of the research presented here.

Residual serial correlations (i.e. the correlation between r_i and r_{i+k}) are plotted against lag (k) in Figure 12. A small persistent correlation is present and indicates dependence in the residual series (Figure 12). This was investigated further by plotting the (i+1)th uniform residual against the corresponding ith residual (for all i). The result is shown in Figure 13, from which it is clear that the dependence in the residuals is very weak (slightly higher densities of points are evident in the top right hand and bottom left hand corners of the figure). Some further research would be needed to find the cause of this correlation. It may be due to small underlying trends in the data, caused by climate change or the El Nino effect, which could be modelled by including lower frequency harmonics in the equations for the model parameters.

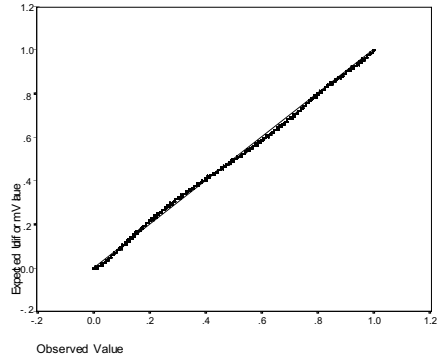

Figure 10: The cumulative distribution function evaluated for the residuals (r). The points, which appear as a slight curve, are the expected values for a standard exponential distribution plotted against the empirical cumulative distribution function. Departures from the line indicate lack-of-fit of the residuals to the exponential distribution.

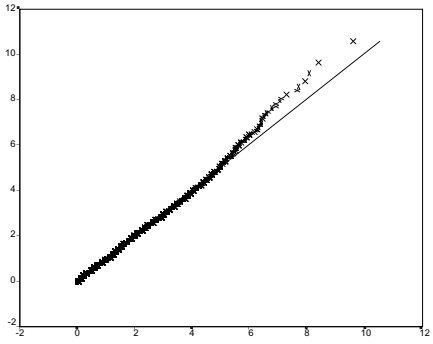

Figure 11: A quantile-plot for the residuals, where a quantile is a percentile expressed as a decimal. The points (×) are the expected quantiles under a standard exponential distribution plotted against the empirical quantiles. Departures from the line indicate lack-of-fit of the residuals to the exponential distribution.

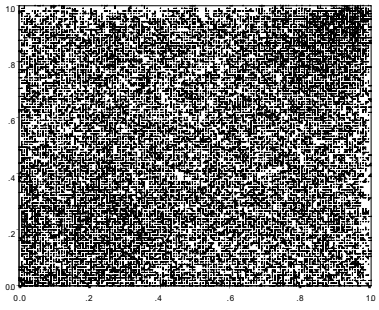

Figure 12: The residual autocorrelations plotted against lag k, i.e. the correlation between r_i and r_{i+k} (k = 1, …, 50).

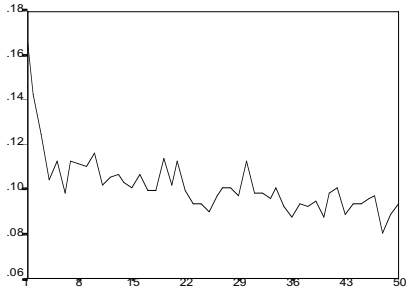

Figure 13: Plot of the (i+1)th residual (uniform transformation, u_{i+1}) against the ith residual (u_i).

6. Conclusions

A renewal cluster model was proposed for the analysis of inter-arrival times of rainfall events in continuous-time. A fitting procedure was given, which used maximum likelihood for parameter estimation and the AIC to choose the best fitting distributions. For the Wellington data, the log-Normal distribution was found to provide the best fit to the times between successive storm clusters, whilst the Weilbull distribution was found to provide the best fit to the times between successive events in the same storm cluster. It was found that the mean cluster size and parameters for the distributions of inter-arrival times could be represented as harmonic curves, without a significant reduction in the likelihood. The mean number of events in a storm cluster was always greater than one, providing statistical evidence that the recorded events were clustered in time.

The plots of the residual series showed that overall the model fitted the data, although a slight under-prediction of extreme values was evident. The residual series were slightly dependent, which may be due to small underlying trends caused by El Nino or climate change. Some further research would be needed to address these problems should they be deemed of practical significance.

In conclusion, the renewal cluster model is recommended for the statistical analysis of rainfall data in continuous-time, as it provides an objective basis on which to infer probability distributions for the inter-arrival process and the expected number of events in a storm cluster.

Acknowledgements

The New Zealand National Institute of Water and Atmospheric Research (NIWA) are gratefully acknowledged for supplying the data.

References

Akaike H. (1974), A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, AC-19, 716-723.

Cowpertwait P.S.P. (1994), A generalized point process model for rainfall, Proc. R. Soc. (A), 447, 23-37.

Cowpertwait P.S.P. (1998), A Poisson-cluster model of rainfall: high-order moments and extreme values, Proc. R. Soc. (A), 454, 885-898.

Cox D.R. and Snell E.J. (1968), A General Definition of Residuals, *J. R. Statist Soc.*, B, 30, 248-275.

Foufoula-Georgiou E. and Krajewski W. (1995), Recent Advances in Rainfall Modeling, Estimation, and Forecasting, *Rev. Geophys.,* 1125-1137.

Nelder J.A. and Mead R. (1965), A Simplex Method for Function Minimization, *Computer Journal,* 7, 308-313.

O'Connell P.E. and Todini E. (1996), Modelling of Rainfall, Flow and Mass Transport in Hydrological Systems: An Overview, *J. Hydrol.,* 175, 3-16.

Ogata Y. (1988), Statistical Models for Earthquake Occurrence and Residual Analysis for Point Processes, Journal of the American Statistical Association, 83, 9-27.

O'Neil R. (1985), Function minimization using a Simplex procedure. Algorithm AS 47 in *Applied Statistics Algorithms*, edited by Griffiths P. and I.D. Hill. Published by Ellis Horwood Ltd. for the Royal Statistical Society, London.

Sansom J. and Thomson P.J. (1992), Rainfall Classification using Breakpoint Pluviograph Data, *J. Climate.,* 5, 7, 755-764.

Sansom J. (1992), Breakpoint Representation of Rainfall, *J. App. Meteor.,* 31, 12, 1514-1519.