



Intermittency of point processes in warfare

Prepared by:

Corey M. Yanofsky, Ph.D.
David R. Bickel, Ph.D.

University of Ottawa
Faculty of Medicine
Roger-Guindon Hall
451 Smyth, Ottawa, ON, K1H 8M5

DRDC CORA SCIENTIFIC AUTHORITY
Dr. Peter Dobias
DRDC CORA
101 Colonel By Dr.
Ottawa, ON, K1A 0K2

PWGSC Contract No.: L9-42310

The scientific or technical validity of this Contract Report is entirely the responsibility of the Contractor and the contents do not necessarily have the approval or endorsement of Defence R&D Canada.

- © Her Majesty the Queen as represented by the Minister of National Defence, 2014
- © Sa Majesté la Reine, représentée par le ministre de la Défense nationale, 2014

This page intentionally left blank.

Abstract

Since instances of criticality are ubiquitous in nature, the intensity of conflicts has been interpreted in terms of self-organized criticality. In this report, interest focused on the statistical properties of series of events in warfare; by investigating the fractal nature of these time series, it is possible in principle to characterize the underlying event-generating process.

We applied three methods to explore the scale-invariant behaviour of the time series. With the first two, we examined the tail of each count distribution for evidence of fractal scaling. First, we estimated the density of the logarithm of the counts, and estimated the slope of the graph of the log-counts versus the log-density; second, we fit a truncated power law distribution to the upper-tail cumulative distribution function of the count data using weighted non-linear least squares. Using the third method, we estimated the intermittency (a measure of the propensity of the time series to suddenly increase above typical values) of each time series. In all cases, we applied a bootstrap approach to correct bias and provide levels of confidence. We found that estimates of the scaling exponents of the count distributions were all close to minus one (-1), suggesting that these distributions have wide tails (possibly up to some upper cut-off). For 13 out of 15 data sets, the probability that the scaling exponent is compatible with that of a discrete stable process is 75% or more. In terms of intermittency, the results suggest that the time series are fairly intermittent (values between 0.06 and 0.16 on a scale from zero (0) to one (1); attempted bias correction widened the range to between zero (0.00) and 0.27.). However, reliable confidence intervals of intermittency could not be obtained using currently available statistical methods.

Executive summary

Corey M. Yanofsky and David R. Bickel, Intermittency of point processes in warfare, DRDC-RDDC-2014-CXX

Introduction

Since instances of criticality are ubiquitous in nature, the intensity of conflicts has been interpreted in terms of self-organized criticality. In this report, interest focused on the statistical properties of series of events in warfare; by investigating the fractal nature of these time series, it is possible in principle to characterize the underlying event-generating process.

Results

We found that estimates of the scaling exponents of the count distributions were close to -1 for all event data sets provided by the DRDC. For 13 out of 15 data sets, the probability that the scaling exponent is compatible with that of a discrete stable process is 75% or more. In terms of intermittency, the results suggest that the time series are fairly intermittent, taking values between 0.06 and 0.16 on a scale from 0 to 1; bias correction widened the range to between 0.00 and 0.27.

Conclusions

The time series show significant evidence of scale-invariant behaviour. The estimates of the scaling exponent of the count distribution and the intermittency indicate that these time series may be expected to show sudden, sharp increases above typical values. Thus, statistical methods that account for intermittency are expected to perform better in terms of predictive power than conventional methods of data analysis.

This page intentionally left blank.

Table of contents

Abstract.....	ii
Executive summary	iii
Table of contents	v
List of figures	vi
List of tables	vii
1. Introduction	1
1.1. Background	1
1.2. Aim.....	1
1.3. Objectives.....	1
2. Methodology.....	3
2.1. Time series	3
2.2. Count density fitting.....	8
2.3. Upper-tail cumulative distribution fitting.....	13
2.4. Intermittency estimation.....	17
3. Summary.....	23
4. References	25
List of symbols/abbreviations/acronyms	27

List of figures

Figure 1. Time series for data sets with prefix DATA.	5
Figure 2. Time series for data sets with prefix HELDATA.....	6
Figure 3. Time series for data sets with prefix KANDATA.....	7
Figure 4. Logarithmic plots of kernel density estimates of the log-counts for data sets with prefix DATA. A least-squares linear fit to the region within the gray vertical lines are plotted in black. The maximum-likelihood Poisson distribution is plotted in red.	9
Figure 5. Logarithmic plots of kernel density estimates of the log-counts for data sets with prefix HELDATA. A least-squares linear fit to the region within the gray vertical lines are plotted in black. The maximum-likelihood Poisson distribution is plotted in red.....	10
Figure 6. Logarithmic plots of kernel density estimates of the log-counts for data sets with prefix KANDATA. A least-squares linear fit to the region within the gray vertical lines are plotted in black. The maximum-likelihood Poisson distribution is plotted in red.....	11
Figure 7. Upper-tail CDFs of counts for data sets with prefix DATA. A least-squares weighted nonlinear fit to the region within the gray vertical lines are plotted in black. The maximum-likelihood Poisson distribution is plotted in red.	14
Figure 8. Upper-tail CDFs of counts for data sets with prefix HELDATA. A least-squares weighted nonlinear fit to the region within the gray vertical lines are plotted in black. The maximum-likelihood Poisson distribution is plotted in red.	15
Figure 9. Upper-tail CDFs of counts for data sets with prefix KANDATA. A least-squares weighted nonlinear fit to the region within the gray vertical lines are plotted in black. The maximum-likelihood Poisson distribution is plotted in red.	16
Figure 10. Log-log plots of estimated counting process second moment versus for data sets with prefix DATA.	20
Figure 11. Log-log plots of estimated counting process second moment versus for data sets with prefix HELDATA.	21
Figure 12. Log-log plots of estimated counting process second moment versus for data sets with prefix KANDATA.	22

List of tables

Table 1. Maximum-likelihood parameter estimates for the Poisson distribution.....	4
Table 2. Estimated slopes for linear fits of log-counts versus log-density. Densities were estimated using kernel density estimation. Due to the higher variance of nonparametric estimation compared to parametric estimation, these values are less reliable than those of Table 3.....	12
Table 3. Estimated scaling exponents for truncated power-law fits to the upper-tail CDFs of the counts.	17
Table 4. Estimated bias of intermittency estimates as a function of the ratio of the highest possible time between events to the lowest possible time between events.	19
Table 5. Estimated intermittencies and bias-corrected estimates.	19

This page intentionally left blank.

1. Introduction

1.1. Background

Self-organized criticality is thought to be a mechanism by which natural systems may develop scale-invariant behaviour. The phrase *self-organized* refers to the idea that the emergence of scale-invariance in such systems is robust to a wide range of conditions. In particular, the intensity of conflicts has been interpreted in terms of self-organized criticality (Roberts and Turcotte 1998).

In this report, interest focused on the statistical properties of series of events in warfare. The occurrence of events can be modeled as a stochastic point process. Such processes may display fractal (i.e., scale-invariant) behaviour in a variety of statistical quantities over a range of time scales. The series of events in warfare are modeled as realizations of fractal-based stochastic point processes (Lowen and Teich 2005). By investigating the fractal nature of the process, it is possible to characterize the underlying event-generating process under the model assumptions.

1.2. Aim

The stochastic nature of the data is such that a deeper analysis of the deviation from expected values is required in order to determine the validity of the conclusions made from the data. In addition, the deviation from normalcy (intermittency) of the data needs to be determined due to its relationship with the nature of the underlying assessed system, the counterinsurgency campaign in Afghanistan.

1.3. Objectives

The objectives of the study were as follows:

- Assess the nature of the data, in particular determine if the data exhibits fractal behaviour or is random (i.e. corresponding to a Poisson process).
- Calculate intermittency of the data and provide implications of the findings for the statistical nature of the provided data.

Provide the results of the analysis in the form of a technical report including raw data and details of the methodology.

This page intentionally left blank.

2. Methodology

For each locale, the start of the time series was treated as the day before the earliest event in the all of the time series at that locale. These start dates are 2002-04-06, 2003-08-11, and 2003-03-22, for the DATA, HELDATA, and KANDATA times series, respectively. To achieve a timescale appropriate for estimating the scaling exponent of the distribution tails, the counting time was increased from 1 day to 10 days, that is, the original daily count data were aggregated by taking blocks of 10 days and summing all counts within each block. Plots of the data may be found in Section 2.1.

Three approaches were used to examine the fractal nature of the data. The first two methods estimate scaling in the tails of the count distributions, whereas the third method estimates the degree to which the underlying process is intermittent. First, an approach that combined log-log plotting of the count probability density with a non-parametric kernel density estimator was used to investigate scale-invariant behaviour in the count probability densities (Section 2.2). Second, the upper-tail cumulative probability distributions of the counts were fit to a truncated power law (Section 2.3). Third, the process intermittency was estimated (Section 2.4). A bootstrap approach was used to correct estimator bias and quantify confidence in the inferences.

2.1. Time series

Here we present plots of the time series (Figures 1 to 3). Visual inspection of the data served to ensure that no errors have occurred while importing the data into the R environment. The visual inspection also revealed that the time series all show a substantial trend with time, whereas the homogeneous Poisson would make such trends extremely unlikely.

In subsequent sections containing the analyses of count probability densities and distributions, we include the maximum-likelihood Poisson fits for comparison; in all cases the Poisson fits are inadequate. Table 1 gives the estimated parameter values, which correspond to the expected number of counts in 10 days under the Poisson model.

Table 1. Maximum-likelihood parameter estimates for the Poisson distribution

Data set name	Estimated mean number of events per 10 days
DATA_allevents	286.4
DATA_attevents	119.2
DATA_devents	92.6
DATA_idevents	21.9
DATA_isevents	19.3
HELDATA_allevents	63.7
HELDATA_attevents	36.2
HELDATA_devents	32.4
HELDATA_idevents	5.0
HELDATA_isevents	3.7
KANDATA_allevents	42.5
KANDATA_attevents	17.6
KANDATA_devents	13.3
KANDATA_idevents	4.8
KANDATA_isevents	4.3

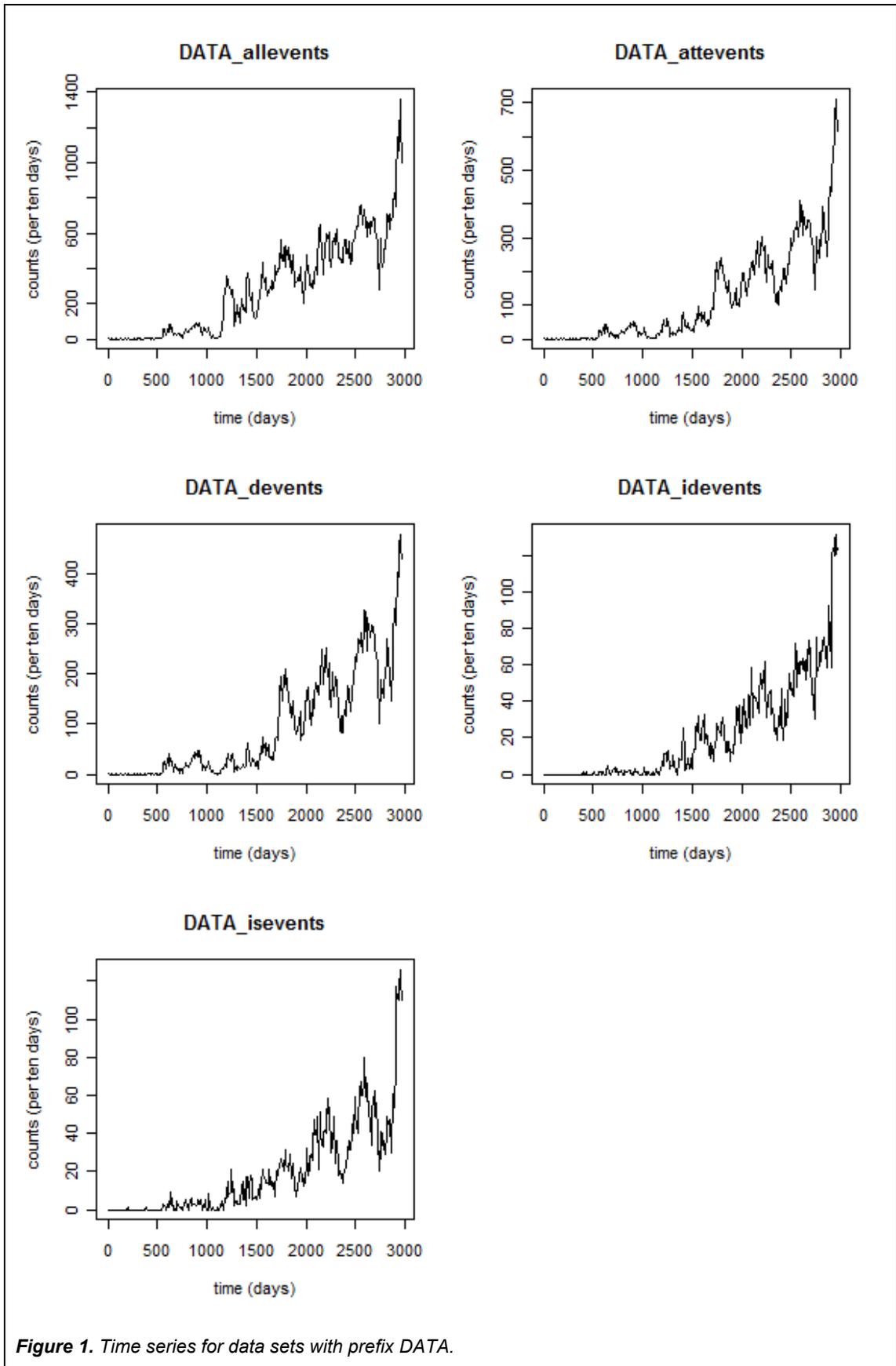


Figure 1. Time series for data sets with prefix DATA.

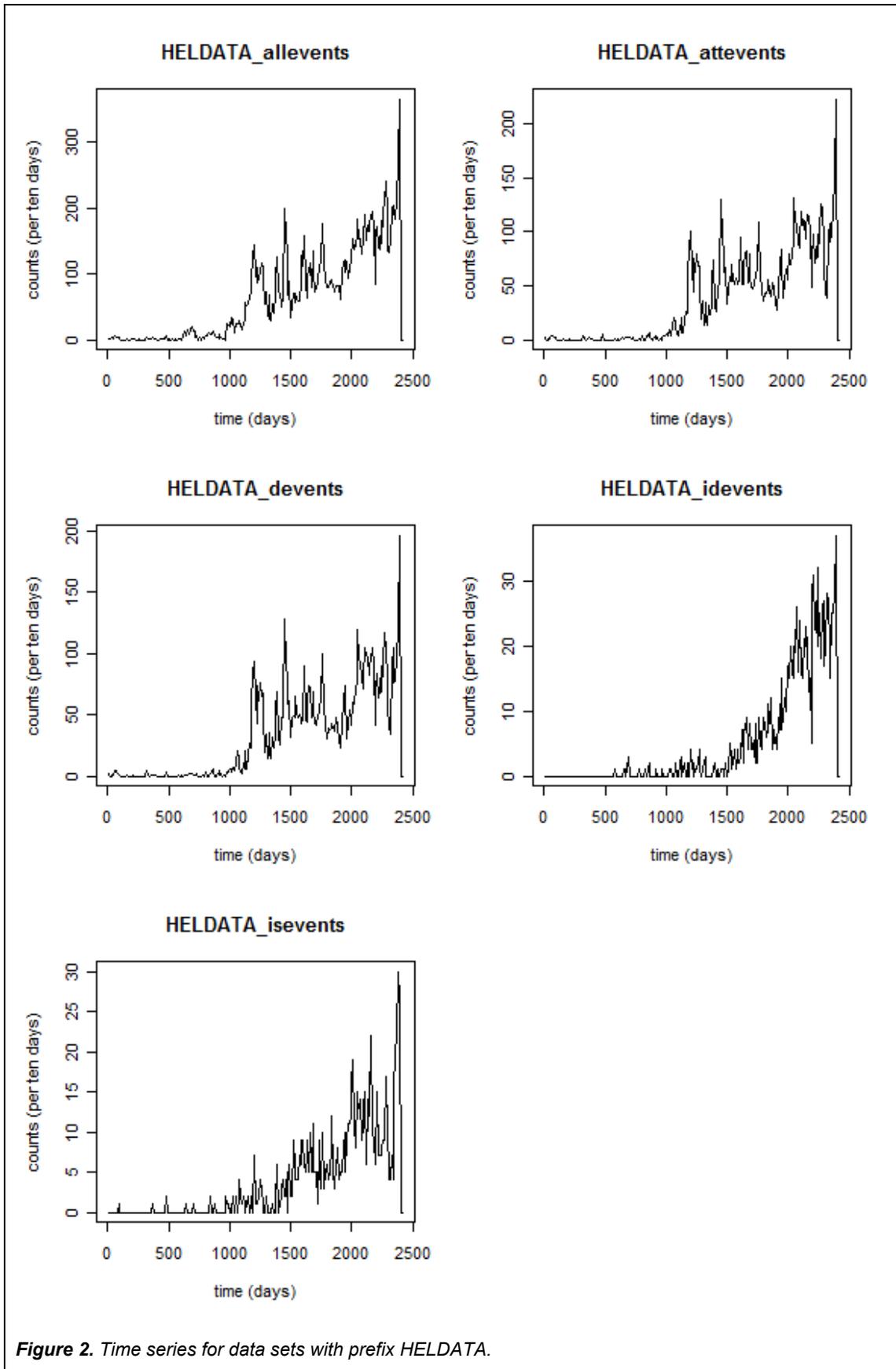


Figure 2. Time series for data sets with prefix HELDATA.

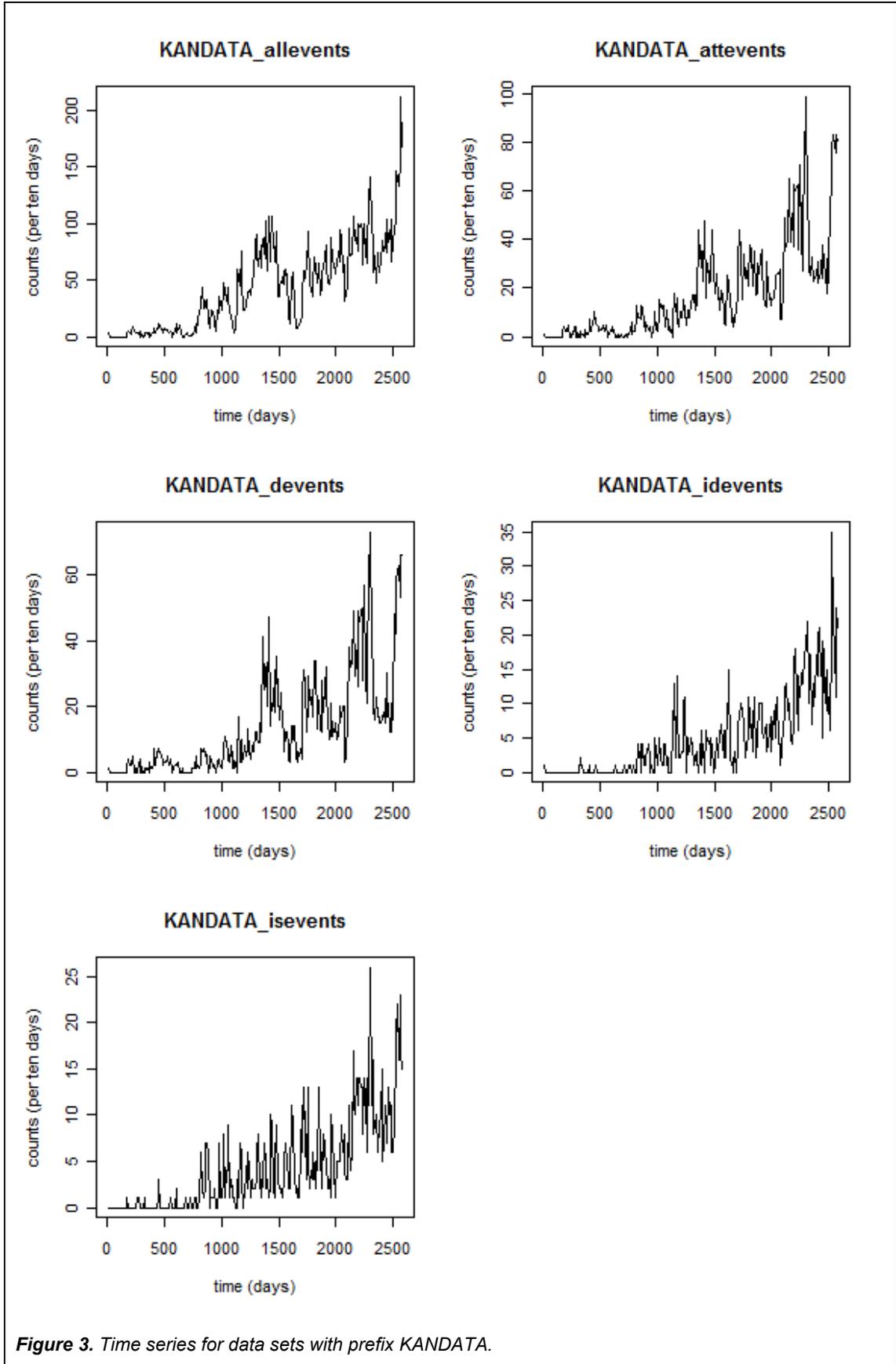


Figure 3. Time series for data sets with prefix KANDATA.

2.2. Count density fitting

When a probability density has power-law behaviour over some range, a log-log plot of the density yields a straight line over that range. One method of detecting such behaviour in real data involves estimating the density in a way that does not constrain it to a particular parametric form. The histogram is a common non-parametric density estimator, but when the data set is small, the histogram is subject to high variance. In order to make more efficient use of the data, we used a kernel density estimator to estimate the count density.

A kernel density estimator is defined by

$$\hat{f}(x; h) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right),$$

in which x_i are the N observations, $K(\cdot)$ is a kernel function (a symmetric, real-valued function of total integral one), and h is a bandwidth parameter that controls the smoothness of the resulting estimate.

In the present application, we applied a kernel density estimate to the logarithm (base 10) of the count data. This transformation improved the performance of the kernel estimator, but note that the logarithm transformation changes the slope of the resulting log-log plot relative to the untransformed data: let x denote the number of counts and suppose that

$$f(x) \propto x^\alpha$$

where α is the scaling exponent. Let $g(\log_{10}x)$ denote the probability density of $\log_{10}x$. Then

$$f(x) = g(\log_{10}x) \left| \frac{d\log_{10}x}{dx} \right|,$$

$$x^\alpha \propto g(\log_{10}x) \frac{1}{x \ln 10},$$

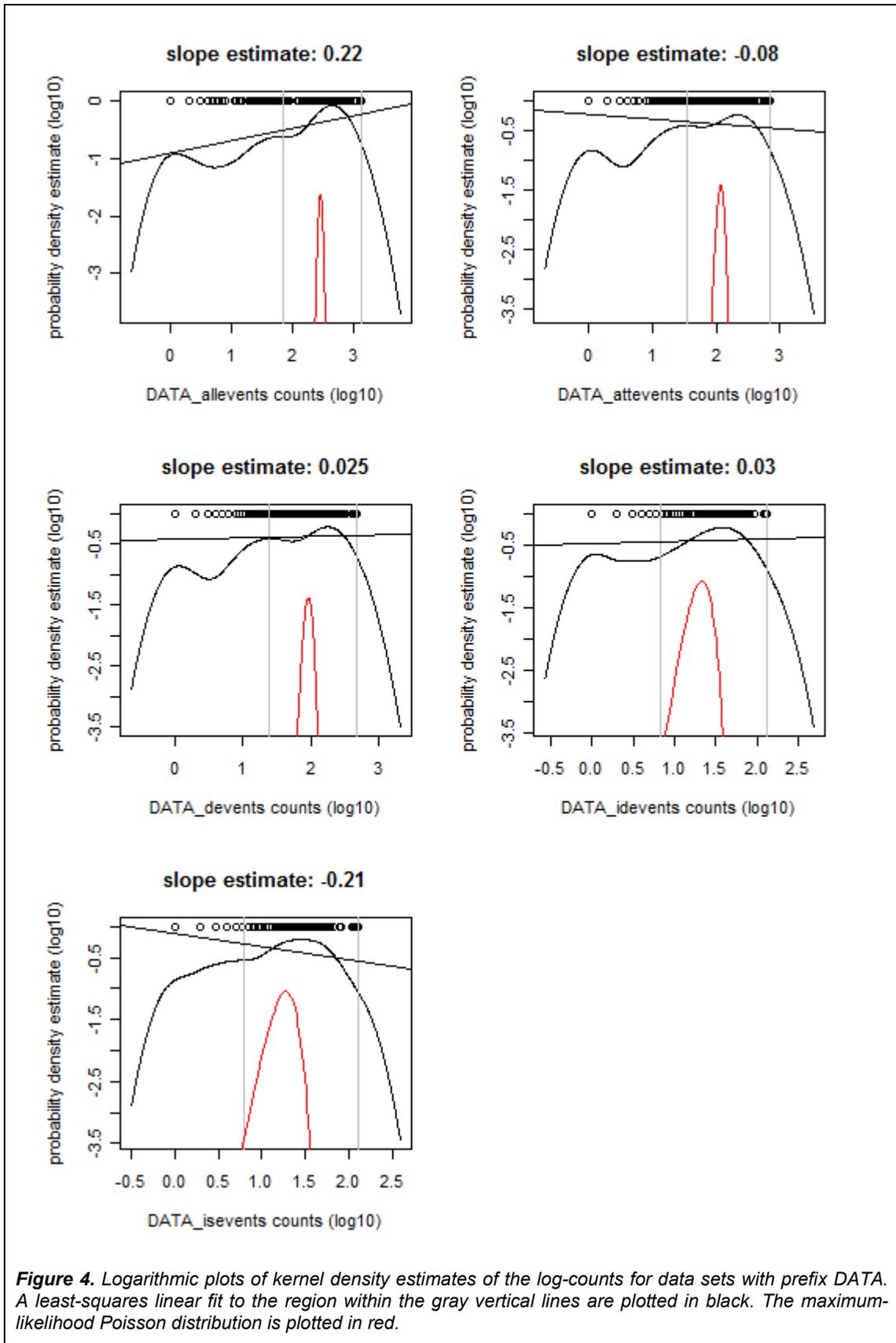
$$g(\log_{10}x) \propto [x^{(1+\alpha)}] \ln 10,$$

$$\log_{10}[g(\log_{10}x)] = k + (1 + \alpha)\log_{10}x.$$

Therefore, as with the usual log-log plots, a plot of the log-density estimated in terms of the log-counts will be linear in regions where scale-invariant behaviour occurs, but the slope is equal to the scaling exponent plus one.

Figure 4 to Figure 6 show logarithmic plots of kernel density estimates of the density of the log-counts. The kernel density estimates were calculated using R's default kernel (Gaussian) and bandwidth parameter ($h = 0.282$). The data themselves are plotted as points at the top of each plot. The least-squares regression line is plotted in black; only the interval of the density estimate from one-twentieth of the maximum datum to the maximum datum (shown by the gray vertical lines) was used for the fit. As discussed in Section 2.1, the maximum-likelihood Poisson distributions are also plotted, in red.

In all plots, the estimated densities plots do not show downward trends with increasing log-counts, even in cases where the linear fit has a negative slope. The count densities are apparently roughly inversely proportional to counts over a large range of the data. This effect dominates the estimation of the scaling exponent. We could not improve the fit even when decreasing the bandwidth of the kernel estimator to as small as 0.02.



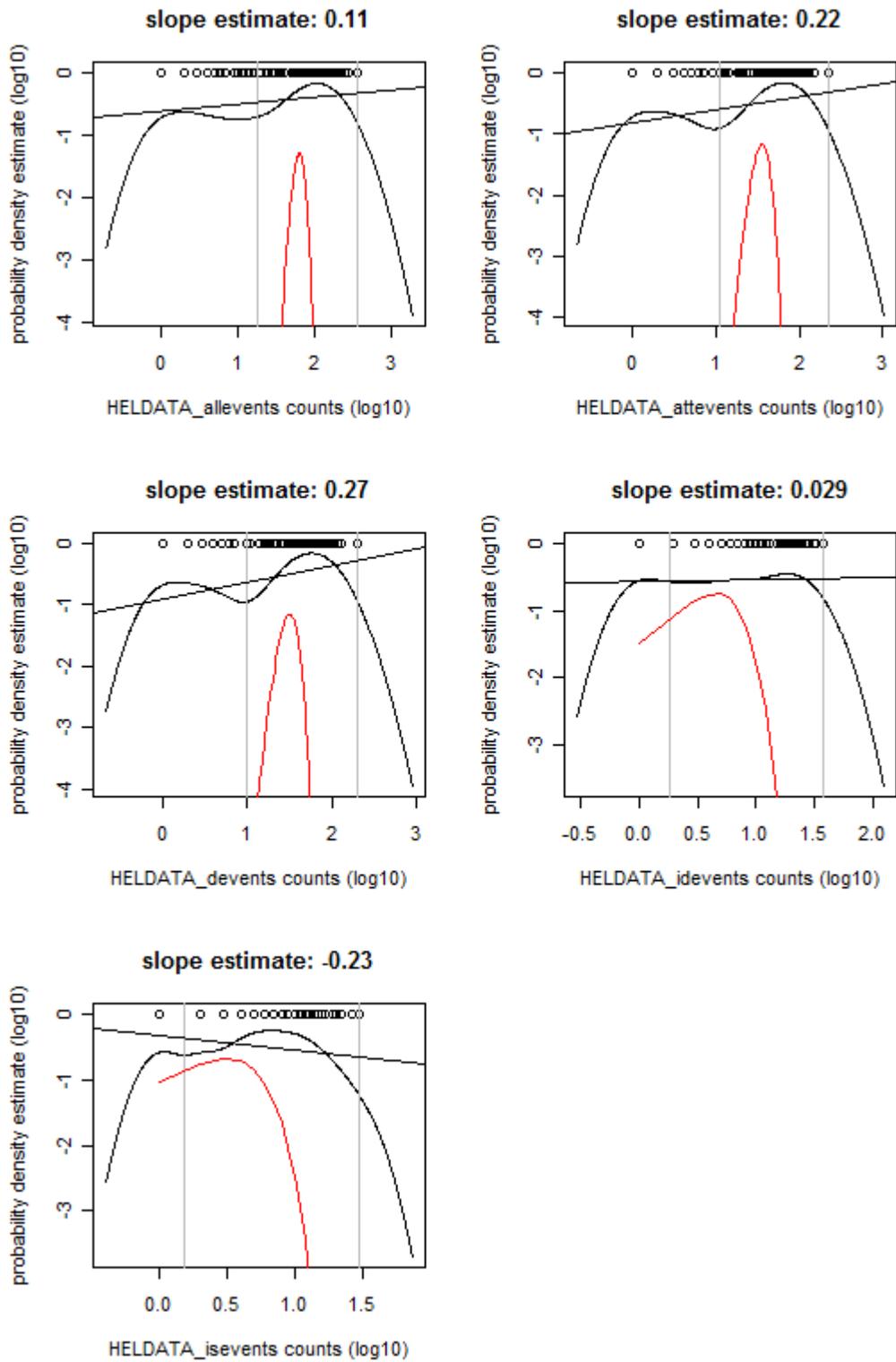


Figure 5. Logarithmic plots of kernel density estimates of the log-counts for data sets with prefix HELDATA. A least-squares linear fit to the region within the gray vertical lines are plotted in black. The maximum-likelihood Poisson distribution is plotted in red.

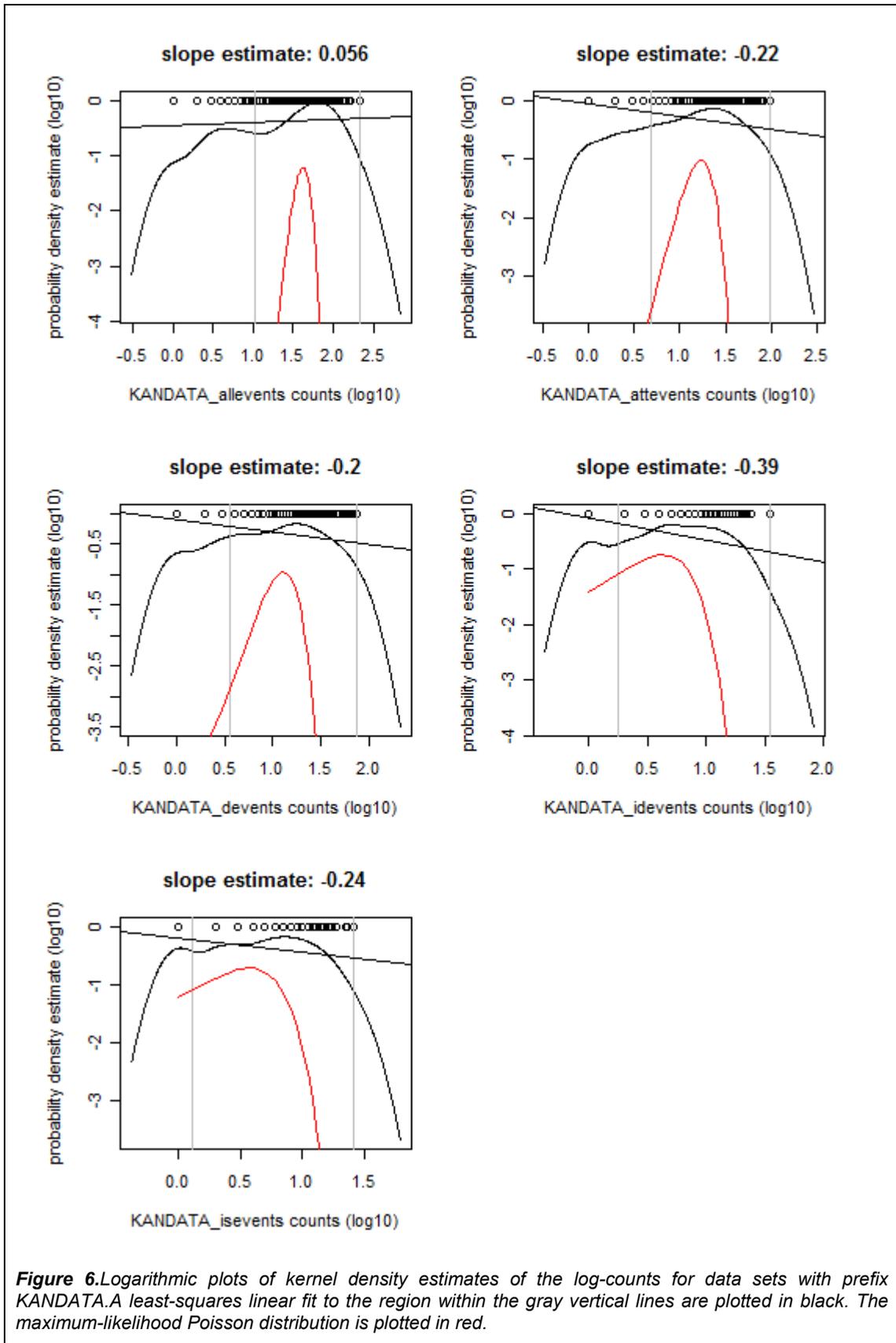


Figure 6. Logarithmic plots of kernel density estimates of the log-counts for data sets with prefix KANDATA. A least-squares linear fit to the region within the gray vertical lines is plotted in black. The maximum-likelihood Poisson distribution is plotted in red.

We used the non-parametric bootstrap (Shao and Tu 1995) to provide bias correction and confidence distributions (Singh, Xie and Strawderman 2007) for the parameter. By encoding confidence intervals at all levels of confidence, confidence distributions enable assigning a level of confidence to a given hypothesis, a property that we exploit to compute the probability that the scaling exponent is compatible with a certain stochastic process found in critical systems. In this approach, replicate data sets are generated by independently resampling the original data with replacement. The estimation procedure is applied to the replicate data, providing information about the estimators' sampling distributions which can then be used to correct bias and generate confidence distributions. We applied the bias-corrected bootstrap percentile method to the intermittency estimates without transformation, but further research would be required to verify that the resulting confidence distributions are valid in the sense that they generate confidence intervals that cover the true parameter value at a rate equal to the stated probability.

In Table 2 we report uncorrected and bias-corrected point estimates of the scaling exponent (equal to the slope in Figure 4 to Figure 6 minus 1). In the last column we report a confidence-distribution-based probability that the scaling exponent is in the interval (-2, -1), this interval corresponding to the discrete stable distribution. The stable processes underlying the discrete stable processes (Lee, Hopcraft and Jakeman 2008) are stationary solutions to a renormalization group equation (West, Bologna and Grigolini 2003) and thus can prove useful as descriptions of systems in a critical state. Probabilities reported in the last column are based on the non-Bayesian posterior distribution (Bickel 2009) as approximated by the bootstrapping method mentioned above.

Table 2. Estimated slopes for linear fits of log-counts versus log-density. Densities were estimated using kernel density estimation. Due to the higher variance of nonparametric estimation compared to parametric estimation, these values are less reliable than those of Table 3.

Data set name	Estimated scaling exponent	Bias-corrected Estimate	Probability that parameter is in discrete stable region
DATA_allevents	-0.78	-0.74	4%
DATA_attevents	-1.08	-1.07	92%
DATA_devents	-0.97	-0.96	43%
DATA_idevents	-0.97	-0.96	40%
DATA_isevents	-1.21	-1.19	99%
HELDATA_allevents	-0.89	-0.77	41%
HELDATA_attevents	-0.78	-0.62	35%
HELDATA_devents	-0.73	-0.53	41%
HELDATA_idevents	-0.97	-0.95	47%
HELDATA_isevents	-1.23	-1.15	98%
KANDATA_allevents	-0.94	-0.84	50%
KANDATA_attevents	-1.22	-1.15	99%
KANDATA_devents	-1.20	-1.18	99%
KANDATA_idevents	-1.39	-1.19	94%
KANDATA_isevents	-1.24	-1.18	100%

2.3. Upper-tail cumulative distribution fitting

Another way to inspect the data for evidence of scale-invariant behaviour is through its upper-tail cumulative distribution function (CDF), defined as

$$P(X > x) = F(x) = \int_x^{\infty} p_X(x') dx',$$

in which X is the random variable, x is the realized value, and x' is a dummy variable. An unbiased estimator of its value at the observed data is

$$\hat{F}(x_i) = \frac{1}{N+1} \sum_{j=1}^N I(x_j \geq x_i),$$

in which $I(\cdot)$ is the indicator function that yields 1 if its argument is true and 0 otherwise.

Relative to the density estimation method, this approach has the advantage of requiring fewer data for stable estimation. However, log-log plots of the CDF are not sufficient to detect scale invariance because in real data such invariance generally has an upper cut-off, reflecting the fact that the time between events has some physical upper bound. As a result, the expression for $F(x)$ does not give a line after taking logarithms. For example, if

$$p_X(x) = \begin{cases} g(x), & x < x_1 \\ Cx^\alpha, & x_1 \leq x < x_2 \\ 0, & x \geq x_2 \end{cases}$$

Then, for $x \geq x_1$,

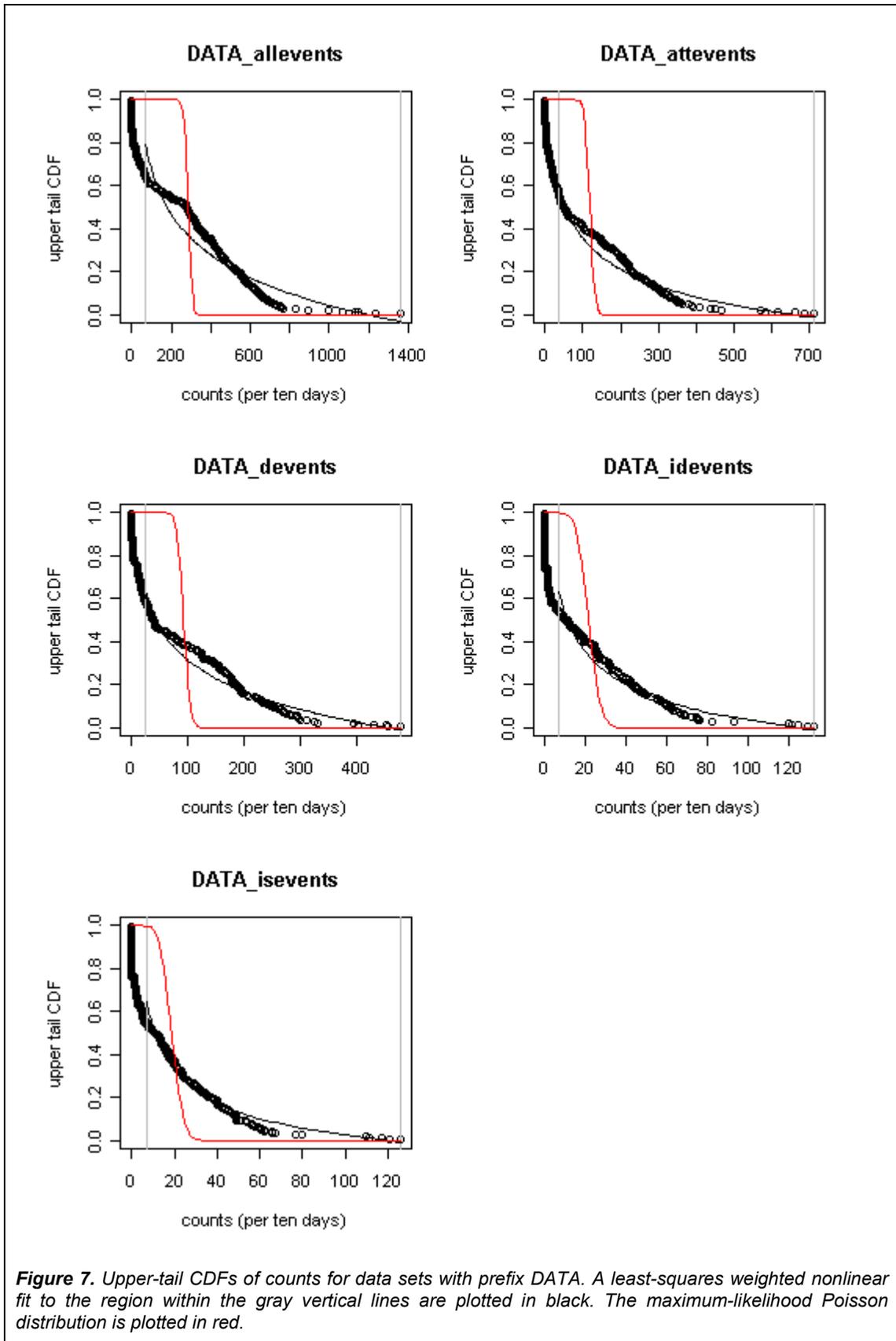
$$F(x) = \int_x^{\infty} p_X(x') dx',$$

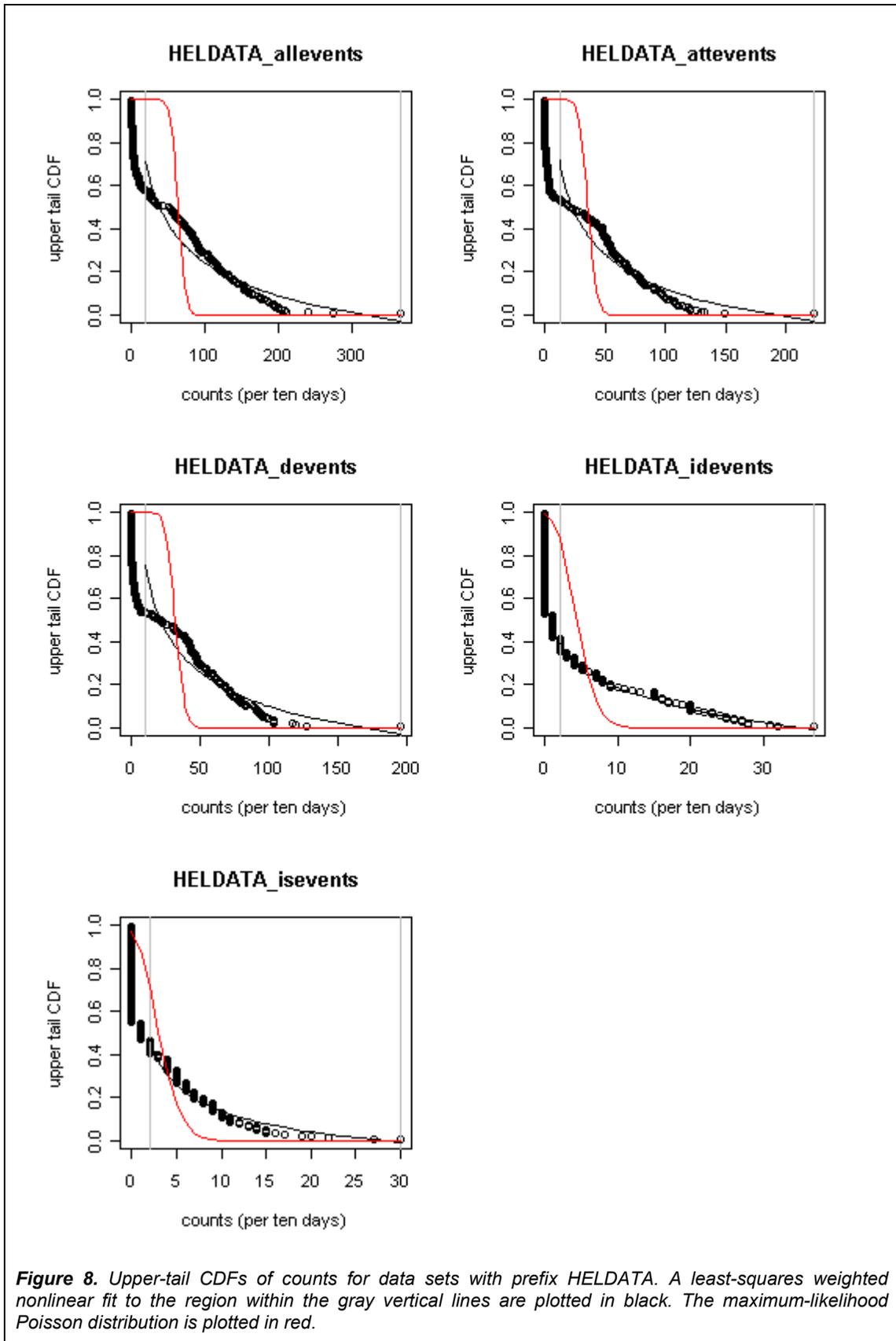
$$F(x) = \frac{C}{1+\alpha} (x_2^{1+\alpha} - x^{1+\alpha}).$$

Therefore, we selected a three-parameter method that proved useful in application to a scale-free network (Bickel 2005).

We used R's non-linear least-squares fitting function, `nls()`, to fit the truncated power-law distribution (given above) to the estimated upper-tail CDF of the count data; results are shown in Figure 7 to Figure 9. As before, only data between one-twentieth of the maximum number of counts and the maximum number of counts were used in the fits; the gray vertical lines show these limits on the plots. Weights were applied to the data to emphasize the tails: the range of the fitted data was divided into five bins whose lengths were a geometric sequence, each bin (except the left-most) being half the size of the bin to its left. Data points in the left-most bin were given weight one, and data in all other bins were given weight such that each bin contained as much total weight as the left-most bin.

In almost all cases, the plots of the upper-tail CDFs show that the largest counts are well separated from the bulk of counts, as would be expected for a distribution with wide tails. In many cases, the model overestimates probability mass in the tails of the CDFs; this lack of fit will tend to cause the scaling exponent to be overestimated. Table 1 shows the parameter estimates; as in the previous section, we report uncorrected and bias-corrected point estimates for the scaling parameter and the confidence-based probability that the true value lies between -2 and -1.





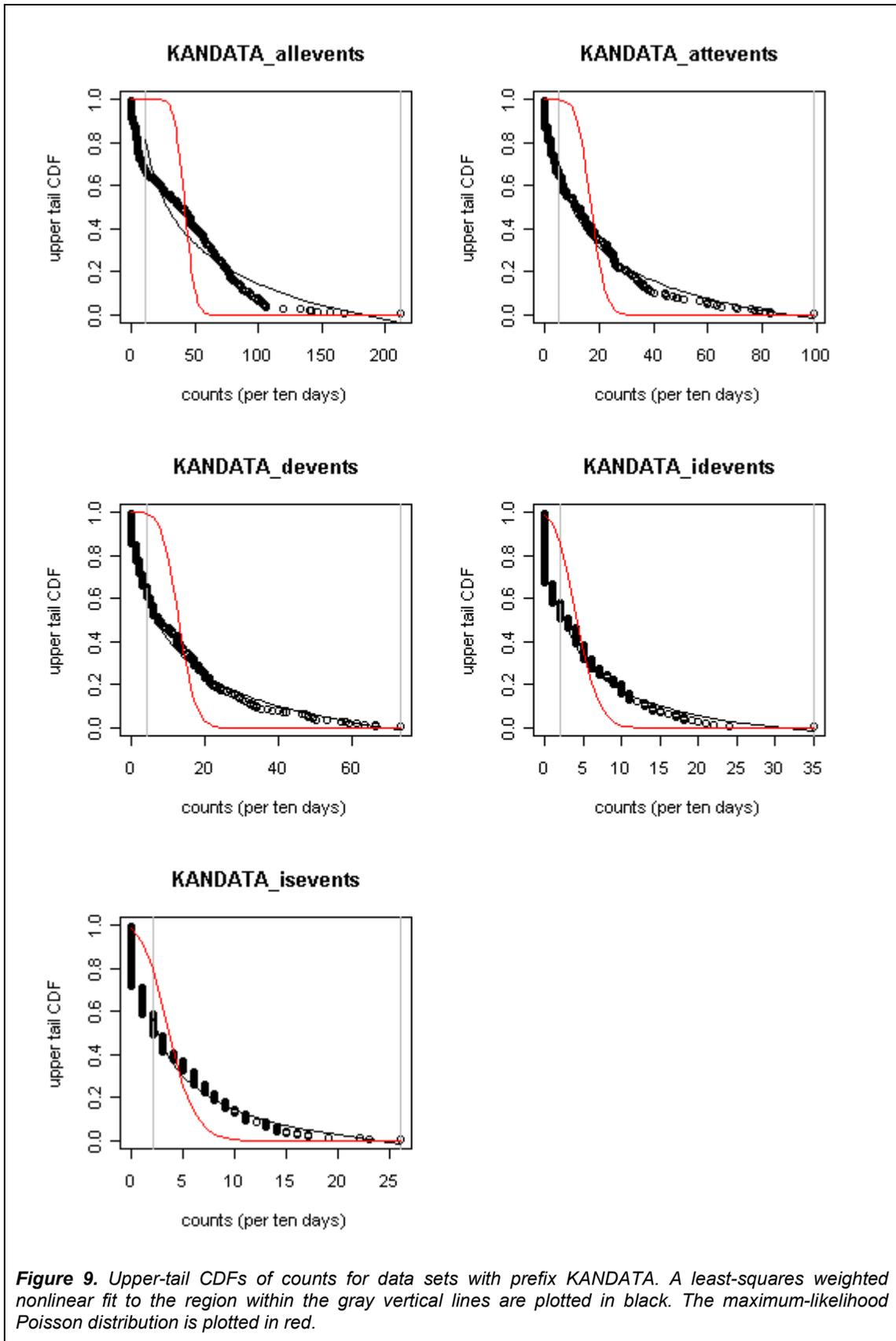


Table 3. Estimated scaling exponents for truncated power-law fits to the upper-tail CDFs of the counts.

Data set name	Estimated scaling parameter	Bias-corrected estimate	Probability that parameter is in discrete stable region
DATA_allevents	-1.10	-1.05	97%
DATA_attevents	-1.18	-1.16	100%
DATA_devents	-1.02	-0.99	70%
DATA_idevents	-1.24	-1.24	100%
DATA_isevents	-1.38	-1.36	100%
HELDATA_allevents	-1.21	-1.04	96%
HELDATA_attevents	-1.28	-1.05	88%
HELDATA_devents	-1.26	-0.99	85%
HELDATA_idevents	-0.89	-0.83	38%
HELDATA_isevents	-1.31	-1.27	100%
KANDATA_allevents	-1.16	-1.08	100%
KANDATA_attevents	-1.22	-1.20	100%
KANDATA_devents	-1.19	-1.18	99%
KANDATA_idevents	-1.38	-1.28	100%
KANDATA_isevents	-1.35	-1.30	100%

2.4. Intermittency estimation

The count distribution is expected to provide evidence of scale-invariant intermittency if it is present, but fitting the count density makes no use of temporal structure of the data. Therefore, we estimated the intermittency of each times series, using the definition and methods described in (Bickel 1999). Roughly speaking, the intermittency parameter measures the propensity of the counting process to suddenly deviate from typical values.

Let $N(T)$ be the *counting process* associated with a stationary point process; a realization of $N(T)$ is the number of events that occur within a time period T . It has a distribution that depends on the underlying point process and on the length of the counting time. In particular, we can examine how the second moment of the distribution varies with T . If the underlying point process is scale-invariant, then it has been shown that

$$\langle N^2(T) \rangle \sim T^{2-C_2},$$

in which C_2 is known technically as the correlation codimension, which measures the intermittency of the point process on a scale from 0 (no intermittency) to 1 (completely intermittent) (Bickel 1999).

The second moment of the counting process can be estimated from the series of the observed counts as follows. Let $\{Z_1, Z_2, \dots, Z_n\}$ be the count time series, with counting time T_0 . Construct the counting process as the cumulative sum of the count process:

$$N_k = \sum_{i=1}^k Z_i, k = 1, 2, \dots, n$$

Then, for $k = 1, 2, 4, 8, \dots, k_{\max} \leq n$, set

$$S_k = \frac{1}{n - k + 1} \sum_{j=1}^{n-k+1} (N_{j+k} - N_j)^2$$

S_k is an estimate of $\langle N^2(kT_0) \rangle$; for point processes with scale-invariant behaviour, a log-log plot of kT_0 versus S_k will have a linear region with slope $2 - C_2$.

To generate confidence distributions, we used a parametric bootstrap approach. In that approach, the estimated parameters are plugged into a parametric model from which replicates of the data are sampled. The nonparametric, IID bootstrap used in the above two methods of estimating the exponent in the distribution tails cannot be applied to the estimation of intermittency because it would destroy the temporal structure of the time series.

We instead modeled the data as a finite-time realization of a fractal renewal process. In the fractal renewal process (FRP), inter-event intervals are independent and identically distributed with a truncated power law distribution. The FRP is thus defined by the three parameters of the truncated power law distribution: the scaling exponent, the lower cut-off, and the upper cut-off. The scaling exponent is estimated by the slope of the linear region of a log-log plot of kT_0 versus S_k , but no estimates for the cut-offs are available.

We therefore performed a sensitivity analysis in which the bias of the estimator was estimated using simulations from the FRP. We fixed the lower cut-off to 1 hour, used the maximum and minimum intermittencies estimated in the actual data (0.06 and 0.16, respectively, as reported in Table 5) as the true intermittency of the simulation, and estimated the bias as a function of the ratio of the upper cut-off to the lower cut-off, R , for each of the six values in $\{10^1, 10^2, 10^3, 10^4, 10^5, 10^6\}$. For each condition, 100 simulated FRP samples were generated.

Table 4 presents the results of the simulation. The second column provides a human-readable interpretation of the R value by stating the upper cut-off (i.e., the maximum possible time period between events, within the context of the FRP model) in decipherable units. The third and fourth columns report the bias. When the upper cut-off is small, the bias is essentially equal to the negative of the true intermittency. Under these conditions, no intermittency can be detected because the intermittency is occurring on time scales much smaller than the time window used to summarize of the underlying point process into a time series of counts. When the upper cut-off is large, the bias is strongly positive. Under these conditions, the FRP will tend to generate inter-event intervals on the same order of magnitude as the length of the time series, creating large gaps in the simulated count time series that essentially “run out the clock”. This gives the appearance of greater intermittency than is actually present; a longer time series would ameliorate this effect. These results highlight the need for either an accurate estimator of R or an estimator of intermittency which is insensitive to R .

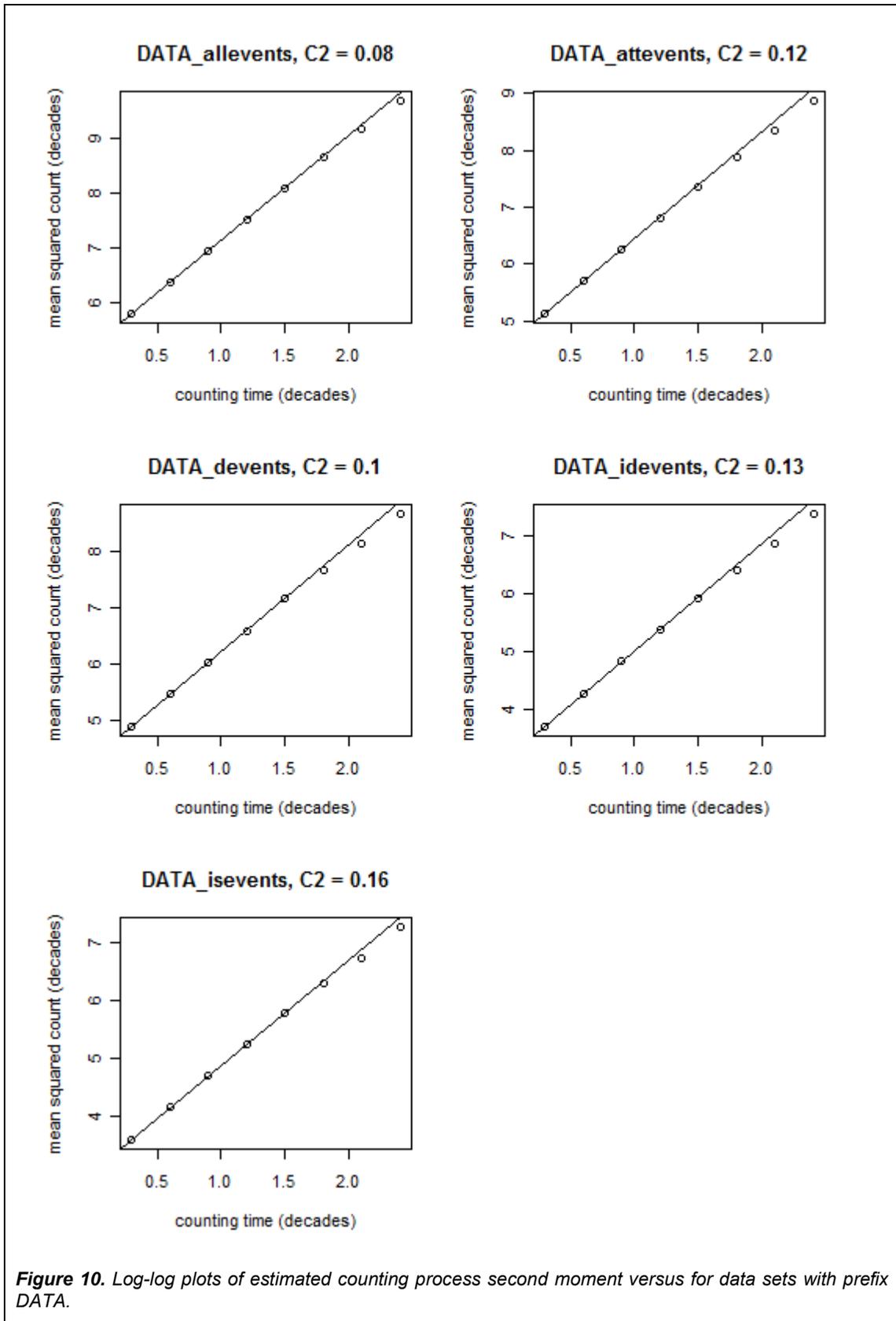
The bias-corrected estimates of intermittency are presented in Table 5 for values of R in $\{10^3, 10^4, 10^5, 10^6\}$. In Table 4, the $R = 10^3$ row showed extreme negative bias and the $R = 10^6$ row showed strong positive bias. Thus, at least one of the corresponding columns in Table 5 contains over-corrected estimates.

Table 4. Estimated bias of intermittency estimates as a function of the ratio of the highest possible time between events to the lowest possible time between events.

<i>R</i>	upper cut-off	bias	
		true intermittency = 0.06	true intermittency = 0.16
10 ¹	10 hours	-0.06	-0.16
10 ²	4 days, 4 hours	-0.06	-0.16
10 ³	42 days	-0.06	-0.15
10 ⁴	1 year, 52 days	-0.04	-0.13
10 ⁵	11 years, 5 months	0.07	0.01
10 ⁶	114 years	0.09	0.08

Table 5. Estimated intermittencies and bias-corrected estimates.

Data set name	Estimated intermittency	Bias-corrected intermittency estimate			
		<i>R</i> = 10 ³	<i>R</i> = 10 ⁴	<i>R</i> = 10 ⁵	<i>R</i> = 10 ⁶
DATA_allevents	0.08	0.15	0.12	0.02	-0.01
DATA_attevents	0.12	0.23	0.20	0.08	0.04
DATA_devents	0.10	0.18	0.16	0.04	0.01
DATA_idevents	0.13	0.26	0.23	0.10	0.06
DATA_isevents	0.16	0.32	0.28	0.13	0.09
HELDATA_allevents	0.06	0.12	0.10	0.00	-0.04
HELDATA_attevents	0.07	0.12	0.10	0.00	-0.03
HELDATA_devents	0.06	0.12	0.10	0.00	-0.04
HELDATA_idevents	0.07	0.14	0.12	0.01	-0.03
HELDATA_isevents	0.13	0.25	0.22	0.08	0.03
KANDATA_allevents	0.09	0.18	0.13	0.04	0.00
KANDATA_attevents	0.13	0.25	0.21	0.09	0.05
KANDATA_devents	0.14	0.27	0.23	0.10	0.06
KANDATA_idevents	0.13	0.25	0.20	0.09	0.05
KANDATA_isevents	0.14	0.27	0.23	0.10	0.06



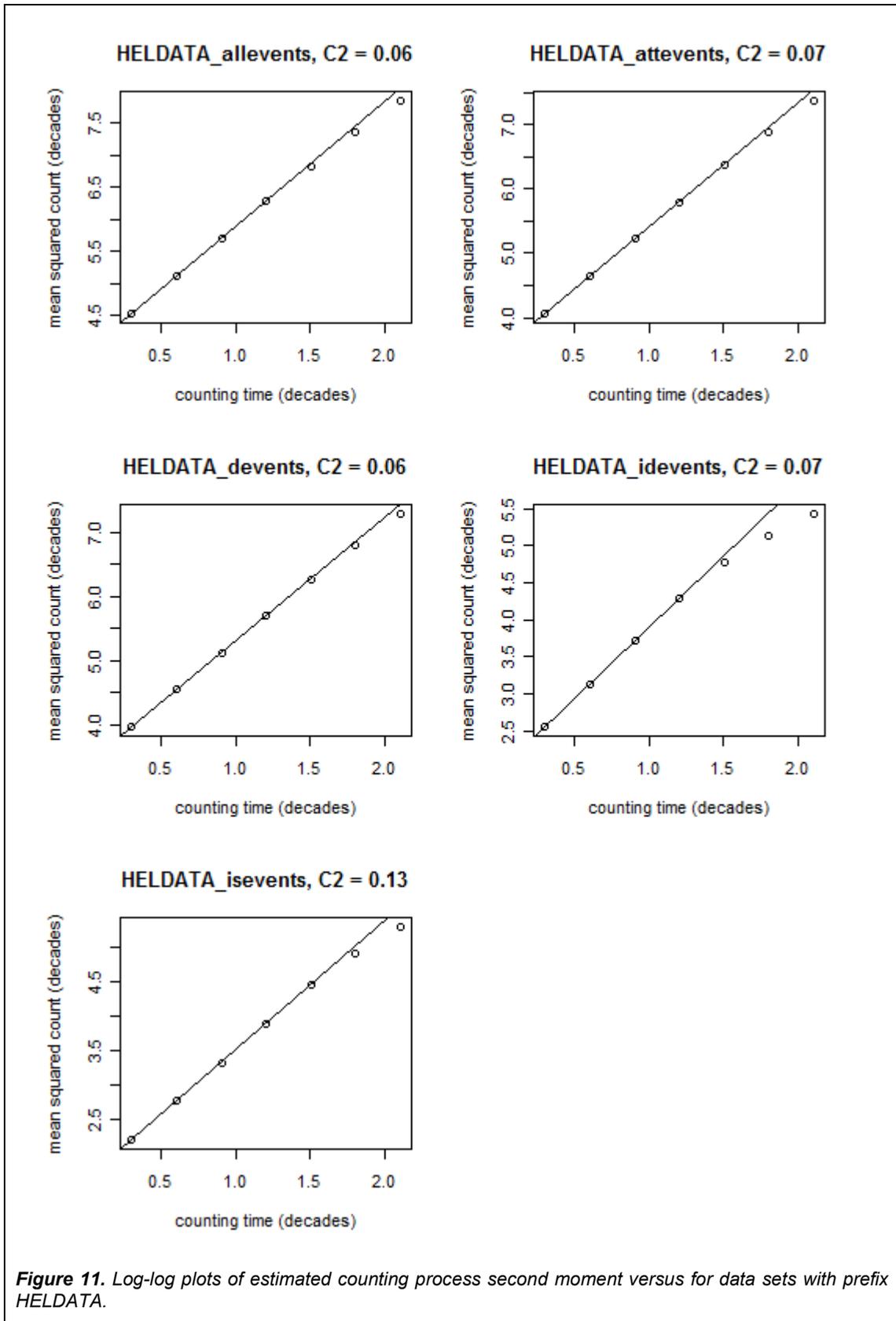


Figure 11. Log-log plots of estimated counting process second moment versus for data sets with prefix HELDATA.

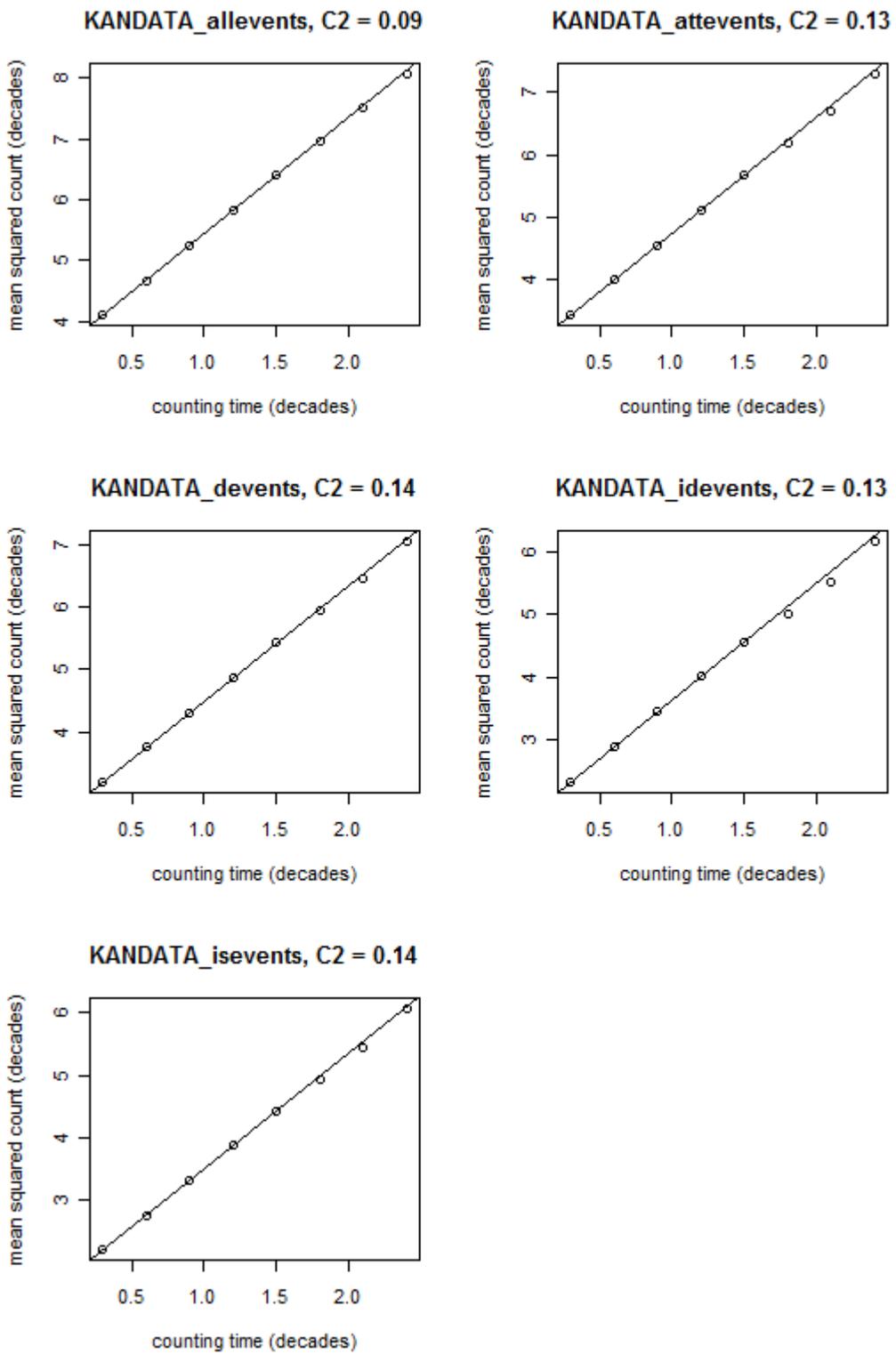


Figure 12. Log-log plots of estimated counting process second moment versus for data sets with prefix KANDATA.

3. Summary

Series of events in warfare were modeled as discrete count processes derived from underlying stochastic point processes. The scaling exponent of the tail of each count distribution was estimated in two ways: first, the scale invariance of the count density was estimated using a linear fit to kernel density estimate of the log-density of the log-counts; second, a parametric model was fit by weighted non-linear least squares to the upper-tail CDF of the count data. In both cases, plots of the fits included the maximum-likelihood Poisson distribution for comparison. Finally, the temporal structure of the times series was investigated by the estimation of the process intermittency. In all three analyses, bootstrapping was used to correct estimator bias and to obtain confidence levels.

In the first analysis (Section 2.2), the log-densities were roughly flat, leading to slope that were close to zero and scaling exponents around -1. This suggests that in the bulk of the data, the count densities are nearly inversely proportional to counts. In the second analysis (Section 2.3), the fit imposed an upper cut-off to reflect the fact that the time between events has some physical upper bound and was weighted to emphasize data at the tails. The fitted curves tended to overestimate the probability mass in the tails, likely causing overestimation of the scaling exponents (negative estimates may be closer to 0 than they should be). Again, the estimated scaling exponents were very close to -1, and in some cases even exceeded -1. For 13 out of 15 data sets, the probability that the scaling exponent is compatible with that of a discrete stable process is 75% or more.

In the third analysis (Section 2.4), the intermittencies of the series of events in warfare were estimated. For this analysis, the bootstrap approach required an unknown parameter (the upper cut-off of the assumed inter-event interval distribution), so the bootstrap bias correction used a set of six possible values. Intermittencies were estimated to be between 0.06 and 0.16; bias correction widened the range to between 0.00 and 0.27.

This page intentionally left blank.

4. References

1. Bickel, David R. "Probabilities of spurious connections in gene networks: application to expression time series." *Bioinformatics* 21, no. 7 (2005): 1121-1128.
2. Bickel, David R. "Coherent frequentism." 2009, <http://arxiv.org/abs/0907.0139v3>.
3. Bickel, David R. "Estimating the intermittency of point processes with applications to human activity and viral DNA." *Physica A* 265 (1999): 634-648.
4. Lee, W.H., K.I. Hopcraft, and E. Jakeman. "Continuous and discrete stable processes." *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 77, no. 1 (2008): 011109.
5. Lowen, S.B., and M. C. Teich. *Fractal-Based Point Processes*. Hoboken, New Jersey: Wiley-Interscience, 2005.
6. Roberts, D. C., and D. L. Turcotte. "Fractality and Self-Organized Criticality of Wars." *Fractals* 6, no. 4 (1998): 351-357.
7. Shao, Jun, and Dongsheng Tu. *The Jackknife and the Bootstrap*. New York: Springer-Verlag New York Inc., 1995.
8. Singh, Kesar, Minge Xie, and William E Strawderman. "Confidence distribution (CD) – distribution estimator of a parameter." *IMS Lecture Notes-Monograph Series – Complex Datasets and Inverse Problems: Tomography, Networks and Beyond* 54 (2007): 132-150.
9. West, B.J., M. Bologna, and P. Grigolini. *Physics of Fractal Operators*. New York: Springer, 2003.

This page intentionally left blank.

List of symbols/abbreviations/acronyms

CDF	Cumulative distribution function
CORA	Centre for Operational Research and Analysis
DND	Department of National Defence
DRDC	Defence Research and Development Canada
FRP	Fractal renewal process