Biostatistics, Bioinformatics and Biomathematics
Volume …, Issue …, 2010, Pages …

SHOT NOISE PROCESSES FOR CLUMPED PARASITE INFECTIONS WITH TIME-DEPENDENT DECAY DYNAMICS

DOMINIK HEINZMANN[1,2], A. D. BARBOUR[1] and PAUL R. TORGERSON[2,3]

[1]Institute of Mathematics
University of Zurich, Switzerland
E-mail: dominik.heinzmann@math.uzh.ch

[2]Institute of Parasitology
University of Zurich, Switzerland

[3]School of Veterinary Medicine
Ross University, West Indies

Abstract

Shot noise processes are introduced to model aggregated parasitic count data arising from clumped superinfections coupled with different decay mechanisms of the ingested parasite clumps. The corresponding likelihood functions are derived by using Laplace transforms. The models are fitted to samples with *Echinococcus granulosus* parasites in dogs from Kazakhstan, Tunisia and China. It is shown that parameter estimates take plausible values and that the decay dynamics is comparable in the three samples. The results indicate that dogs cease to be infectious after about 8 months, and that infections of dogs occur at a low rate, but the ingested parasite load per clump is in the thousands.

1. Introduction

In macroparasitic diseases, count data on parasites in hosts is often the result of multiple clumped superinfections and a decay pattern of the established clumps [4]. The number and time points of the clumped

2010 Mathematics Subject Classification: Please provide.

Keywords: shot noise process, inverse Laplace transform, clumped infection, infection duration, Echinococcus.

Revised November 2, 2010

superinfections are in general not observable and thus the decay pattern is difficult to determine. The resulting parasite counts are most likely strongly aggregated, with many zero and skewed positive counts [15, 35]. Analysis of such data is most commonly done by using descriptive tools, as for example negative binomial distributions in [5, 9], negative binomial regression in [8] and [19] to account for the host's age, or zero-inflated models in [24] and two-part conditional models in [11] to account for the excess of zeros. However, the parameters of these models are in general not directly linked to the underlying transmission dynamics of the parasite. Thus important biological features, such as superinfection and parasite life history, do not appear explicitly and thus are difficult to address.

[18] used a mechanistic model based on clumped infections to describe such data. They assumed that parasites ingested by clumps evolve independently of all others, going through different parasite stages. Their model can be used to test different scenarios such as influence of life time of the parasite or parasite-induced host mortality. However, parameter estimation is difficult since common data sets of macroparasitic diseases only contain ages and parasite counts of the hosts. [16] and [25] also used mechanistic models based on clumped infections to describe macroparasitic data sets. Both models allow for superinfection, and incorporate nonlinear effects such as parasite-induced immunity. As above, the resulting models are difficult to fit to commonly available data sets of macroparasitic infections. In a previous paper [17], we have used mechanistic models based on compound processes to describe the clumped infection mechanism and the age-dependent distribution of Echinococcus granulosus hydatid cysts in sheep. Based on the biology of the parasite in the sheep, their models do not incorporate a decay pattern of the ingested parasite clumps.

In this paper, we present shot noise processes [10, p. 135] as models of aggregated parasitic count data arising from clumped infections coupled with a death process of the parasites. We focus on the infection process of the parasite *Echinococcus granulosus* in the definitive host, a dog (see Section 2). A shot noise process is a superposition of independent and identically distributed shots which occur at different time points and whose effects decay over time [22] and is thus a natural extension of compound processes, to allow for some dynamics such as a decrease of the accumulated shots [10, p. 135]. Biologically, the shots represents clumped infections so that a host

ingests multiple parasites per infection. The shot noise process allow to model multiple clumped infections of a host, so that the parasite burden can decrease between consecutive infections. The numbers and time points of the infections are described by a counting process. Hence shot noise processes can used to describe the prevalence (presence/absence of disease) and the number of parasites in each individual host as function of time.

The models are fitted by the method of maximum likelihood to three samples of *Echinococcus granulosus* in dogs from different countries. It is shown that the estimates are plausible with regard to other experimental data and that all three data sets are adequately described using the models. In addition, the mean duration of a single infection is derived for the models. The models provide a reasonable mechanistic description of the transmission dynamics of *Echinococcus granulosus* in dogs, and describe well the observed zero-inflated and aggregated parasite distributions.

2. Biology and Data Sets

The life-cycle of Echinococcus granulosus takes place primarily between dogs as definitive hosts and sheep as intermediate hosts [12]. Dogs harbor the adult parasite in the small intestine; it releases eggs that are passed in the feces. Sheep ingest the eggs on pasture, and some of them develop into hydatid cysts. Cysts become fertile if they produce protoscoleces. Humans are ecologically aberrant intermediate hosts. Dogs acquire the infection by ingesting organs from the sheep that contain fertile cysts. The protoscoleces then hatch in the small intestine and develop into adult worms. In general, the adult parasite does not proliferate within the dog, and the tapeworm infection does not cause significant harm. The parasite is endemic in many parts of the world [13, 32] and continues to exert a burden on human health, livestock production and wildlife ecology [12].

The data samples used in this paper contain the age and the corresponding *Echinococcus granulosus* parasite count of dogs from South Kazakhstan [31] with a sample size of 606 dogs, from the Testour and Bouzid regions of Tunisia [21] with 140 dogs and from Sichuan Province, People's Republic of China [9] with 371 dogs. Parasite burdens of *Echinococcus granulosus* in dogs in these samples are obtained by purging and then collecting the intestinal contents. The dog ages are derived from an interview with the owner and a

personal assessment of the animals by the interviewer. In the Kazakhstan sample, most of the dogs are free of parasites (76.9%), and 86.4% of the infected dogs harbor 1000 or fewer parasites. Similar patterns are observed in the Tunisia and China samples, where 72.9% and 91.6% respectively of the dogs are parasite free, and 86.8% and 90.3% respectively of the infected animals respectively harbor 1000 or fewer parasites. The maximal parasite loads in the Kazakhstan, Tunisia and China samples are 150000, 67000 and 20000 adult parasites. The mean ages of dogs are 3.133 years for the Kazakhstan, 4.630 years for the Tunisia and 4.237 years for the China sample.

3. Models

The models used to describe the parasite distribution in the individual dogs as function of time are introduced as follows. Starting with introducing the general concept of shot noise processes, we then specify decay patterns and the shot distributions and derive the likelihood functions. Finally, the mean survival times of single clumped infections are evaluated. Related computational aspects to model building and fitting are provided in Appendix A.

3.1. Shot noise models

A dog normally ingests only a few fertile cysts per infection, but the number of protoscoleces in such cysts is rather high. Hence a clumped infection mechanism is realistic. If the infection rate of dogs is low, as suggested in [14, 26, 31 and 32], acquired immunity of dogs can be neglected as shown by experimental studies with the parasite in [15]. Furthermore, the low infection rate suggests that the infection process can be modelled by a Poisson process. A reasonable assumption for *Echinococcus granulosus* in our study regions is that the transmission system is in a steady state [12, 26, 32], so that since infective sheep typically harbor only one fertile cyst and since the number of protoscoleces per fertile cyst is well described by a log-normal distribution (reanalysis of data from [33]), the clump sizes can be supposed to be identically distributed. We further assume that the clump sizes of infections of a single dog are independent since they come from different sheep, and since acquired immunity in dogs can be taken to be negligible. The number of surviving parasites from a clump decreases over time [3, 20, 30], so that shot noise process is reasonable as a model for Echinococcus granulosus in dogs.

A shot noise process $(X_t)_{t\geq 0}$ is a continuous-time piecewise deterministic stochastic process. Events of $(X_t)_{t\geq 0}$ occur at times $0 < \tau_1 < \tau_2 \ldots$, given by the realization of a point process N_t on the nonnegative integers. Henceforth, we shall always take N_t to be a Poisson process of rate β. At each τ_i, there is a realization of a non-negative random variable U_i known as shots, and $X_{\tau_i} - X_{\tau_{i-}} = U_i \geq 0$. Between the τ_i's, $(X_t)_{t\geq 0}$ undergoes a death process determined by a non-increasing function $h(t)$. Assume that $X_0 = 0$ almost surely. Then

$$X_t = \sum_{k=1}^{N_t} U_k h(t - \tau_k), \quad t \geq 0, \tag{1}$$

where U_k $(k = 1, 2, \ldots)$ are i.i.d. random variables with density function f_U independent of N_t, $h(t) = 0$ if $t < 0$, and $X_t = 0$ if $N_t = 0$. In this paper, the shots U_k correspond to the number of successfully established *Echinococcus granulosus* parasites per infection of a dog.

Since $X_t \geq 0$ almost surely, the one-sided Laplace transform

$$L_{X_t}(s) := \mathbb{E}(e^{-sX_t})$$

exists throughout $\mathrm{Re}(s) > \gamma$ for some $\gamma \leq 0$. The random variable X_t has a point mass $F_{X_t}(0)$ at zero, and a density f_{X_t} over the interval $(0, \infty)$. The point mass $F_{X_t}(0)$ is equal to $\exp(-\beta t)$ if $h(t) > 0$ for all positive t, and it is the probability that no shots arrive during the time interval $[0, t]$. The continuous portion f_{X_t} is due to the arrival of one or more shots during $[0, t]$. The function $L_{X_t}(s)$ is completely monotonic for $\mathrm{Re}(s) > \gamma$ [34, p. 161].

Assuming that all moments of X_t exist, one has for $\mathrm{Re}(s) > \gamma$

$$\mathbb{E}(X_t^j) = (-1)^j \frac{d^j L_{X_t}(s)}{ds^j}\bigg|_{s=0}. \tag{2}$$

Since $(N_t)_{t\geq 0}$ is a Poisson process, it follows that $\mathbb{E}(X_t^j) < \infty$ if $\mathbb{E}(U_k^j) < \infty$ for any $j \geq 1$, as is seen by using $\sum_{k=1}^{N_t} U_k$ as upper bound of X_t $(t \geq 0)$.

3.2. Decay pattern

For modeling of the decay of the parasite loads in dogs, we introduce different forms of $h(t)$ in (1) leading to three different models which we then compare based on simulation studies in Section 4.

Assume that $h(t) = \exp(-\lambda t)$, indicating that we have an exponential decay dynamics of the parasite loads in dogs. Since $X_0 = 0$ almost surely, [10, p. 136] have shown that

$$L_{X_t}(s) = \exp\left\{-\beta \int_{0+}^{t} [1 - L_U(se^{-\lambda z})]dz\right\}, \tag{3}$$

where L_U is the common Laplace transform of the U_k ($k = 1, 2, \ldots$). Using (3) in (2),

$$\mathbb{E}(X_t) = \frac{\beta \mathbb{E}(U)}{\lambda}[1 - e^{-\lambda t}], \quad \mathrm{Var}(X_t) = \frac{\beta \mathbb{E}(U^2)}{2\lambda}[1 - e^{-2\lambda t}], \tag{4}$$

with $\mathbb{E}(U^j)$ the j-th moment of the U_k's.

There is experimental evidence that dogs eventually lose *Echinococcus granulosus* infections [3, 12, 15]. We incorporate this with two possibilities. First, define X_t^0 to be a random variable such that $X_t^0 = 0$ if $X_t \in [0, 1)$ and $X_t^0 = X_t$ else. Thus $X_t^0 \in 0 \cup [1, \infty)$ so that

$$\begin{aligned} F_{X_t^0}(0) &= F_{X_t}(0) + \int_0^1 f_{X_t}(u)du \quad && \text{if } z = 0, \\ F_{X_t^0}(z) &= f_{X_t}(z) && \text{if } z \geq 1. \end{aligned} \tag{5}$$

This model will be referred to as MA (mass accumulation) model.

The second possibility is to use the Poisson transform. Let Y_t have the mixed Poisson distribution $\mathrm{Po}(X_t)$; that is, for a given $t \geq 0$,

$$\mathbb{P}(Y_t = y) = \int_0^\infty \frac{e^{-x}x^y}{y!} f_{X_t}(x)dx, \quad y \in \mathbb{N}, \tag{6}$$

and

$$\mathbb{P}(Y_t = 0) = e^{-\beta t} + \int_0^\infty e^{-x} f_{X_t}(x)dx = \mathbb{E}(e^{-X_t}). \tag{7}$$

Now, for $y \in \mathbb{N}$,

$$(-1)^y \frac{d^y L_{X_t}(s)}{ds^y} = \mathbb{E}(X_t^y e^{-sX_t}) = \int_0^\infty x^y e^{-xs} f_{X_t}(x) dx,$$

so that from (6)

$$\mathbb{P}(Y_t = y) = \frac{(-1)^j}{y!} \frac{dL_{X_t}^y(s)}{ds^y}\bigg|_{s=1}, \quad y \in \mathbb{N}, \tag{8}$$

and also $\mathbb{P}(Y_t = 0) = L_{X_t}(1)$. This model will be referred to as PT (Poisson transform) model.

Alternatively, we define the decay pattern as $h(t) = I(t \le t_d)$, where I is the indicator function and t_d a fixed (survival) time, so that that all shots disappear after t_d. Biologically this means that at each time, a dog catches an infection with multiple parasites, the parasites stay at life for a fixed time td and die then all on the same time. In this setting, we have

$$X_t = \sum_{k=1}^{N_t} U_k I(t - \tau_k \le t_d) =_d \sum_{k=1}^{N_{t \wedge t_d}} U_k,$$

where $=_d$ indicates equal in distribution, and so

$$L_{X_t}(s) = e^{-\beta(t \wedge t_d)(1 - L_U(s))}, \tag{9}$$

and

$$\mathbb{E}(X_t) = \beta(t \wedge t_d)\mathbb{E}(U), \quad \mathrm{Var}(X_t) = \beta(t \wedge t_d)\mathbb{E}(U^2). \tag{10}$$

This model will be referred to as CS (constant survival) model.

3.3. Shot distribution

Since observed parasite counts in dogs are heavily skewed [12, 31], shots are assumed to be lognormally distributed. Let U_k $(k = 1, 2, \ldots)$ be independent and lognormally distributed random variables, so that $\log(U_k) \sim N(\mu, \sigma^2)$. Then (4) transforms into

$$\mathbb{E}(X_t) = \frac{\beta e^{\mu + \sigma^2/2}}{\lambda}[1 - e^{-\lambda t}], \quad \mathrm{Var}(X_t) = \frac{\beta e^{2\sigma^2 + 2\mu}}{2\lambda}[1 - e^{-2\lambda t}], \tag{11}$$

and (10) into

$$\mathbb{E}(X_t) = \beta(t \wedge t_d)^{e^{\mu+\sigma^2/2}}, \quad \mathrm{Var}\,(X_t) = \beta(t \wedge t_d)^{e^{2\sigma^2+2\mu}}. \tag{12}$$

The Laplace transform of the U_k's, $L_U(s)$, is needed to evaluate (3) and thus the PT model in (8). We will see later that $L_U(s)$ is also needed to compute the MA and the CS models. There is no general closed-form expression for $L_U(s)$. However, it can be represented through a series expansion based on Gauss-Hermite integration for s real [23], so that

$$\begin{aligned} L_U(s) = \mathbb{E}(e^{-sU}) &= \int_0^\infty \exp(-su)\frac{1}{u\sigma\sqrt{2\pi}}\exp\left[-\frac{(\log(u)-\mu)}{2\sigma^2}\right]du \\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{\pi}}\exp\left[-s\exp(\sqrt{2}\,\sigma z+\mu)\right]\exp(-z^2)dz \\ &= \sum_{i=1}^{N}\frac{\omega_i}{\sqrt{\pi}}\exp\left[-s\exp(\sqrt{2}\,\sigma a_i+\mu)\right] + R_N, \end{aligned}$$

where N, ω_i and a_i $(1 \le i \le N)$ in the final expression are the order, weights and nodes (abscissas) of the Gauss-Hermite integration. For small N, ω_i and a_i can be found in [2, p. 924]. For larger N, the integration nodes a_i are found as a root of the Hermite polynomial H_N of order N [2, p. 509], and the corresponding weights ω_i are calculated by using

$$\omega_i = \frac{2^{N-1}N!\sqrt{\pi}}{N^2[H_{N-1}(a_i)]^2}.$$

The error term R_N decreases with N and several upper bounds can be found in [23]. However, they are difficult to compute for N large and not sharp enough for many applications [29, p. 171]. Thus numerical methods are necessary to choose an appropriate N (see Appendix A.1).

Let $A_i := \omega_i/\sqrt{\pi}$ and $B_i := \exp(\sqrt{2}\sigma a_i + \mu)$, so that $\sum_{i=1}^{N} A_i = 1$ [23]. An approximation to the Laplace transform of the U_k's is then given by

$$\hat{L}_U(s) = \sum_{i=1}^{N} A_i \exp[-sB_i]. \tag{13}$$

Since $\sum_{i=1}^{N} A_i = 1$, we can let W denote a discrete random variable taking values B_i with probabilities A_i $(1 \le i \le N)$, and (13) implies that $\hat{L}_U$ is the Laplace transform of W and that $\hat{L}_U$ is completely monotone. Hence

$$\hat{L}_{X_t}(s) = \exp\left\{-\beta \int_0^t [1 - \hat{L}_U(se^{-\lambda z})]dz\right\}$$

$$= \exp\left\{-\beta \int_0^t \left[1 - \sum_{i=1}^{N} A_i \exp\left(-se^{-\lambda z} B_i\right)\right] dz\right\}, \tag{14}$$

where $h(t) = \exp(-\lambda t)$ as before. A similar argument as above for $\hat{L}_U$ shows that $\hat{L}_{X_t}(s)$ is a Laplace transform and that it is completely monotonic.

Equation (14) indicates that

$$\lim_{s \to 0} \hat{L}_{X_t}(s) = 1, \quad \lim_{s \to \infty} \hat{L}_{X_t}(s) = e^{\beta t}.$$

Substituting $c = sB_i \exp(-\lambda z)$ in (14),

$$\hat{L}_{X_t}(s) = \exp\left\{-\beta t + \frac{\beta}{\lambda} \sum_{i=1}^{N} A_i \int_{sB_i e^{-\lambda t}}^{sB_i} \left[\frac{e^{-c}}{c}\right] dc\right\},$$

where the integral is well-defined, since $sB_i \exp(-\lambda t) > 0$ if $s > 0$, and is bounded above by λt. Using $E_1(a) = \int_a^{\infty} \frac{e^{-b}}{b}\, db$, $a > 0$,

$$\hat{L}_{X_t}(s) = \exp\left\{-\beta t + \frac{\beta}{\lambda} \sum_{i=1}^{N} A_i [E_1(sB_i e^{-\lambda t}) - E_1(sB_i)]\right\}. \tag{15}$$

If $h(t) = I$ $(t \le t_d)$, with t_d the constant duration of a single infection, an approximation to $L_{X_t}(s)$ is given by

$$\hat{L}_{X_t}(s) = e^{-\beta(t \wedge t_d)(1 - \hat{L}_U(s))}.$$

For the PT model, the probabilities (8) can now be approximated by using (15). Then $dE_1(a)/da = -\exp(-a)/a$ $(a > 0)$ implies that for $s > 0$

$$\frac{d\hat{L}_{X_t}(s)}{ds} = \frac{\beta}{\lambda} \left\{\sum_{i=1}^{N} A_i \left[\frac{e^{-sB_i}}{s} - \frac{e^{-sB_i e^{-\lambda t}}}{s}\right]\right\} \hat{L}_{X_t}(s)$$

and $d^2\hat{L}_{X_t}(s)/ds^2$ is

$$\frac{\beta}{\lambda}\left\{\sum_{i=1}^{N} A_i\left[\frac{B_i e^{-\lambda t}e^{-sB_i e^{-\lambda t}} - B_i e^{-sB_i}}{s} - \frac{e^{-sB_i} - e^{-sB_i e^{-\lambda t}}}{s^2}\right]\right\}\hat{L}_{X_t}(s)$$

$$+\left\{\frac{\beta}{\lambda}\sum_{i=1}^{N} A_i\left[\frac{e^{-sB_i}}{s} - \frac{e^{-sB_i e^{-\lambda t}}}{s}\right]\right\}^2 \hat{L}_{X_t}(s),$$

and hence (8) yields the probabilities. Higher order analytical derivatives can be derived using the above formulas.

3.4. Likelihood inference

We have seen that X_t in (1) with $h(t) = \exp(-\lambda t)$ has a point mass of size $\mathbb{P}(X_t = 0) = \exp(-\beta t)$ at zero and a continuous density f_{X_t} on $(0, \infty)$. Define

$$\tilde{M}_{X_t}(s) := \int_0^\infty e^{-sx} f_{X_t}(x)dx,$$

so that $\tilde{M}_{X_t}(s) = L_{X_t}(s) - \mathbb{P}(X_t = 0)$. Then the density of X_t on $x > 0$ can be written as

$$f_{X_t}(x) := \frac{1}{2\pi i}\int_{\gamma - i\infty}^{\gamma + i\infty} e^{sx}\tilde{M}_{X_t}(s)ds, \tag{16}$$

where γ is chosen such that the line $s = \gamma$ lies in the complex plane to the right of all singularities of $\tilde{M}_{X_t}(s)$. Here, we can take $\gamma = 0$.

The contour integral in (16) cannot be evaluated analytically and thus needs to be numerically approximated. The Stehfest algorithm for numerical inversion [28] can be used. Its scheme is given for $x > 0$ by

$$f_{X_t, M}(x) = \frac{\log(2)}{x}\sum_{k=1}^{2M}\zeta_k \tilde{M}_{X_t}\left(\frac{k\log(2)}{x}\right) = \frac{\log(2)}{x}\sum_{k=1}^{2M}\zeta_k L_{X_t}\left(\frac{k\log(2)}{x}\right), \tag{17}$$

since $\sum_{k=1}^{2M}\zeta_k = 0$ so that $\sum_{k=1}^{2M}\zeta_k \mathbb{P}(X_t = 0) = 0$. The coefficients ζ_k are defined in [28] and M is such that increasing values of M imply a more

accurate inversion. Using the approximation $\hat{L}_{X_t}$ (15) in (17) implies an approximation for the density of X_t on $x > 0$, which we denote as $\hat{f}_{X_t,M}(x)$. In this case, the Stehfest algorithm is a suitable choice since it is stable for completely monotonic functions [1].

Given observations of loads x_l and ages t_l $(1 \le l \le n)$, the log-likelihood function of the MA model can be approximated as

$$\sum_{k=1}^{M}\left\{I_{\{x_l<1\}}\log\left[e^{-\beta t}+\int_{0+}^{1}\hat{f}_{X_{T_l},M}(x)dx\right]+I_{\{x_l<1\}}\log\left[\hat{f}_{X_{T_l},M}(x_l)\right]\right\}. \qquad (18)$$

[36] presented an estimation of the Poisson shot noise process (1) based on the method of moments. However, they estimated only the product $\beta\mathbb{E}(U)$ rather than the Poisson rate β and the full distribution of the shots simultaneously. Hence the above likelihood approach is favored.

The probabilities $\mathbb{P}(X_t = y)$ for the Poisson transform Y_t in (6) can be computed by using (8) which requires (analytical) evaluation of higher order derivatives of $\hat{L}_{X_t}$. From (6) we can see that for sufficient large y, $\mathbb{P}(X_t = y)$ is close to $f_{X_t}(y)$ and thus the latter can be used to simplify computations. Denote such an appropriate y as y_0. The determination of y_0 is discussed in Appendix A.2. Given y_0, the log-likelihood for the PT model becomes

$$\sum_{l=1}^{n}\{I_{\{x_l<y_0\}}\log\left[\mathbb{P}(Y_t = x_l)\right]+I_{\{x_l<y_0\}}\log\left[\hat{f}_{X_t l,M}(x_l)\right]\}. \qquad (19)$$

For the CS model, $X_t =_d \sum_{k=1}^{N_t \wedge t_d} U_k$ and thus the likelihood can be computed based on sums of lognormals for the compound Poisson process. Even if there is no general explicit formula for the distribution of sums of lognormals, several approximations exist in the literature [6, 7, 23, 27]. A numerically stable, very accurate and flexible approach is to approximate the sum of lognormals by a single lognormal random variable [23]. Given K independent and lognormally $\mathrm{LN}(\mu, \sigma)$ distributed random variables U_1, …, U_K the Laplace transform of $\sum_{k=1}^{K} U_k$ is matched with the Laplace

transform of $U(K)$, the approximating lognormal random variable with parameters $\mu(K)$ and $\sigma(K)$, at two different, real and positive values s_1 and s_2 [23].

Thus $\mu(K)$ and $\sigma(K)$ are computed by solving the system of nonlinear equations

$$\sum_{i=1}^{N} \frac{\omega_i}{\sqrt{\pi}} \exp\left[-s_l \exp\left(\sqrt{2\sigma}(K) a_i + \mu(K)\right)\right] = \left[\hat{L}_U(s_l;\, \mu,\, \sigma,\, u)\right]^K, \tag{20}$$

for $l = 1, 2$. The choice of s_1 and s_2 is discussed in Appendix A.1. Hence the log-likelihood for the CS model is

$$\sum_{k=l}^{M} \left\{ I_{\{x_l=0\}}(-\beta t) + I_{\{x_l=0\}} \log\left[\sum_{m=1}^{\infty} \frac{e^{-\beta(t \wedge t_d)} (\beta(t \wedge t_d))^m}{m!} f_{U(m)}(x_l) \right] \right\}, \tag{21}$$

where $f_{U(m)}$ is the pdf of $U(m)$ with parameters $\mu(m)$ and $\sigma(m)$ computed based on (20).

3.5. Mean survival times for shots

Given the model specifications above, the mean times for the survival if clumped infections can be derived. Let T be a random variable for the duration of infection for a single shot. For the MA model, $\mathbb{P}(T \geq t) = \mathbb{P}(U \geq \exp(\lambda t)) = 1 - \Phi((\lambda t - \mu)/\sigma)$ for any t fixed, so that

$$\mathbb{E}(T) = \frac{\lambda}{\sigma} \int_{-\mu/\sigma}^{\infty} [1 - \phi(y)] dy.$$

For the PT model, given U, we have $Y_t \sim \mathrm{Po}(U \exp(-\lambda t))$. Hence $\mathbb{P}(T \leq t) = \mathbb{E}[\mathbb{P}(Y_t = 0 \,|\, U)] = \mathbb{E}(\exp(-U \exp(-\lambda t)))$, so that given $U = u$, $T \sim V/\lambda + \log(u)/\lambda$, where V is a Gumbel random variable. It follows that

$$\mathbb{E}(T) = \frac{\mathbb{E}(V) + E(\log(U))}{\lambda} = \frac{\gamma^* + \mu}{\lambda},$$

where γ^* is the Euler-Mascheroni constant.

Finally for the CS model, the survival time is constant and thus $\mathbb{E}(T) = t_d$.

Table 1. Maximum likelihood estimates of the mass accumulation model (MA) (equation (6)), the Poisson transform model (PT) (8) and from the constant survival model (CS) (9) for Kazakhstan, Tunisia and China, together with 95% confidence intervals computed by the bootstrap percentile method. The mean survival times $\mathbb{E}(T)$ of a single infection is computed as described in Subsection 3.5. Note that "–" indicates that the corresponding parameter is not specified in the model.

	Sample	MA	PT	CS
$\hat{\beta}$	Kaza.	0.501 (0.359, 1.112)	0.445 (0.317, 0.918)	0.340 (0.213, 0.881)
	Tuni.	0.689 (0.417, 1.243)	0.662 (0.401, 1.192)	0.487 (0.312, 0.986)
	China	0.330 (0.226, 0.469)	0.308 (0.197, 0.423)	0.127 (0.014, 0.410)
$\hat{\mu}$	Kaza.	5.474 (4.505, 6.754)	6.001 (4.305, 7.054)	4.302 (3.723, 4.928)
	Tuni.	4.340 (3.096, 5.321)	4.046 (3.462, 5.109)	3.560 (3.101, 4.137)
	China	3.263 (2.535, 4.672)	3.398 (2.403, 4.766)	3.261 (2.718, 3.764)
$\hat{\sigma}$	Kaza.	2.804 (2.417, 3.321)	2.955 (2.437, 3.306)	2.616 (2.182, 2.882)
	Tuni.	2.879 (2.509, 3.471)	3.079 (2.393, 3.399)	2.693 (2.184, 3.157)
	China	2.598 (2.396, 2.988)	2.635 (2.285, 3.108)	2.535 (2.207, 2.890)
$\hat{\lambda}$	Kaza.	9.620 (7.238, 15.617)	8.833 (6.319, 13.176)	–
	Tuni.	9.293 (6.783, 18.431)	8.728 (6.210, 16.527)	–
	China	8.413 (7.084, 14.981)	7.916 (6.811, 13.672)	–
$\hat{t}_d$	Kaza.	–	–	0.744 (0.580, 1.108)
	Tuni.	–	–	0.640 (0.397, 1.064)
	China	–	–	0.713 (0.474, 1.216)
$\mathbb{E}(T)$	Kaza.	0.572	0.745	0.744
	Tuni.	0.476	0.530	0.640
	China	0.403	0.502	0.713

4. Application

4.1. Comparison

The MA, PT and CS models are fitted by the maximum likelihood method to the three data sets using the log-likelihoods (18), (19) and (21) respectively. Table 1 shows the resulting estimates. The mean survival times $\mathbb{E}(T)$ in Table 1 are computed based on Subsection 3.5. All three models attest a significantly higher infection pressure β in Kazakhstan and Tunisia than in China, which is reasonable given that a lower prevalence of infection is observed in China (see Figure 1). All models suggest that the mean μ of the log-transformed

shots is significantly higher in the Kazakhstan sample than in the others but that the corresponding standard deviation σ is similar throughout the samples. The exponential decay rate λ for the MA and PT models is not significantly different in any of the three samples. The mean survival times $\mathbb{E}(T)$, computed based on Subsection 3.5, are lowest for the MA model, but largest for the CS model. They are discussed in more details in the following subsection.

Table 2 shows the observed prevalences of infection q and the means m of the log-transformed observed positive loads of the three samples together with the corresponding model values $\hat{q}$ and $\hat{m}$ computed by simulation as follows. Let n be the sample size and let $t_1, \ldots, t_n$ be the observed ages of dogs in the sample. For $1 \le k \le m$, generate a realization for the k-th dog with age t_k from the MA, PT or CS model respectively with the parameters set as their estimated values in Table 1, to attribute a simulated load to him. This yields a new sample from which the prevalence and the mean of the log-transformed positive loads can be computed. Repeating the procedure 2000 times, $\hat{q}$ and $\hat{m}$ are computed as the averages of the resulting corresponding 2000 values, and the corresponding 2.5% and 97.5% quantiles can easily be determined. Overall, the observed values q and m of the models agree with the simulated values of the models. However, for the MA model, the observed values q for the China sample and m for the Kazakhstan sample lie outside the 95% interval of the corresponding simulated quantities.

Table 2. Observed prevalences of infection q and means of the log-transformed loads m in dogs, together with the corresponding model mean values $\hat{q}$, $\hat{m}$, and 2.5% and 97.5% quantiles, computed by simulation from the MA, PT and CS model as described in the text.

Model	Sample	q	$\hat{q}$	m	$\hat{m}$
MA	Kaza.	0.230	0.243 (0.208, 0.278)	4.503	3.682 (3.331, 4.061)
	Tuni.	0.271	0.277 (0.193, 0.357)	3.826	3.283 (2.611, 4.005)
	China	0.084	0.125 (0.096, 0.156)	3.322	2.594 (1.891, 3.409)
PT	Kaza.	"	0.237 (0.204, 0.273)	"	4.265 (3.843, 4.699)
	Tuni.	"	0.261 (0.191, 0.331)	"	3.548 (2.827, 4.291)
	China	"	0.116 (0.081, 0.148)	"	2.939 (2.206, 3.738)
CS	Kaza.	"	0.211 (0.173, 0.259)	"	4.685 (4.301, 5.077)
	Tuni.	"	0.245 (0.174, 0.309)	"	4.259 (3.505, 5.082)
	China	"	0.078 (0.051, 0.105)	"	3.802 (3.022, 4.691)

Figure 1 displays the prevalences of infection and means of the log-transformed positive parasite loads computed by simulation from the MA, PT and CS models for different age classes in all three samples, together with the observed quantities (grey points). The simulation is done as described above. Let $b_{0.025}$ and $b_{0.975}$ denote the 2.5% and 97.5% quantiles from the corresponding simulated values. Given the (simulated) mean prevalence $\hat{q}(t)$ for an age class, the intervals $[b_{0.025}, \hat{q}(t)]$ and $(\hat{q}(t), b_{0.975}]$ were each similar for the three models and in all age classes. Thus instead of plotting all 95% simulation intervals for the three models, we plot the averaged interval lengths over the three models at the observed sample values (grey points). Averaging is done over the sub-interval lengths $\hat{q}(t) - b_{0.025}$ and $b_{0.975} - \hat{q}(t)$, and the resulting (averaged) values are then plotted around the observed quantity (grey bars). This gives a reasonable indication of the variation of the models. Analogously, we proceed for the case of the mean load $\hat{m}(t)$ of the log-transformed positive loads in all age classes. The results in Figure 1 indicate that all models suggest an asymptotic prevalence of infection much lower than 1 in the three samples, and a decreasing mean load of the log-transformed positive loads in young dogs which stabilizes after about 1 year. The PT and CS models are well in line with the observed quantities. In plot (b1), the mean prevalence of the MA model for the age class $6.5+$ is slightly larger than the upper bound of the (averaged) simulation interval. In plot (a2), the mean load of the log-transformed positive loads computed by the MA model is close to the lower bounds of the simulation intervals in age classes (0.5. 1.5], (1.5, 2.5] and (3.5, 4.5], and even slightly below in age class (2.5, 3.5]. The MA model produces the highest prevalences and the lowest means of the log-transformed positive loads in all samples.

4.2. Bounding models

The results in Table 2 and Figure 1 suggest that the PT and CS perform better than the MA model. Figure 2 shows a plot of the estimated pdfs of the PT and CS models for the log-transformed positive loads in the Kazakhstan sample for different age classes, together with a histogram of the corresponding observed quantities. The Kazakhstan sample is chosen since it has the largest number of observations of positive loads. For simplicity, the densities are computed by simulation. Given the middle point of an age class t, we generate

100000 realizations of the PT respectively CS model for t fixed, and take the logarithms of the positive loads. Then a kernel-estimator is applied to the log-transformed loads to obtain an approximation to the true pdf, resulting in the solid and dashed lines in Figure 2.

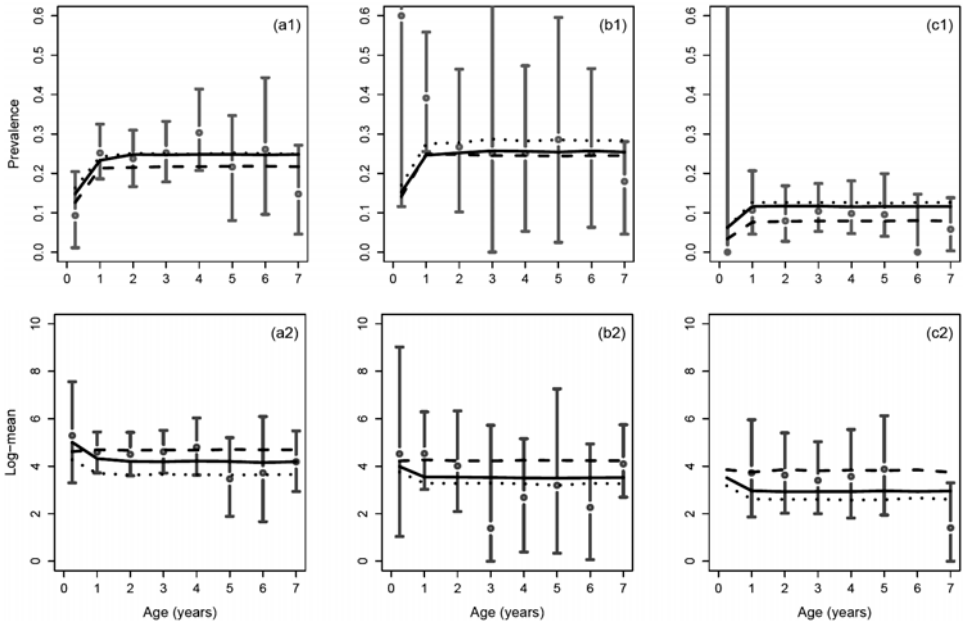

Figure 1. Prevalences of infection (a1-c1) and means of the log-transformed positive loads (a2-c2) in dogs from the MA model (dotted line), PT model (solid line) and CS model (dashed line) (computed by simulation) versus the observed quantities (grey points) for the age classes $(0, 0.5]$, $(0.5, 1.5]$, $(1.5, 2.5]$, $(2.5, 3.5]$, $(3.5, 4.5]$, $(4.5, 5.5]$, $(5.5, 6.5]$ and $6.5+$ of the samples (a) Kazakhstan, (b)Tunisia and (c) China. The grey bars indicate the averaged 95% simulation intervals for the three models, computed as described in the text. Note that there are no observed positive loads in age classes $(0, 0.5]$ and $(5.5, 6.5]$ of the China sample and hence the simulation intervals are not plotted there.

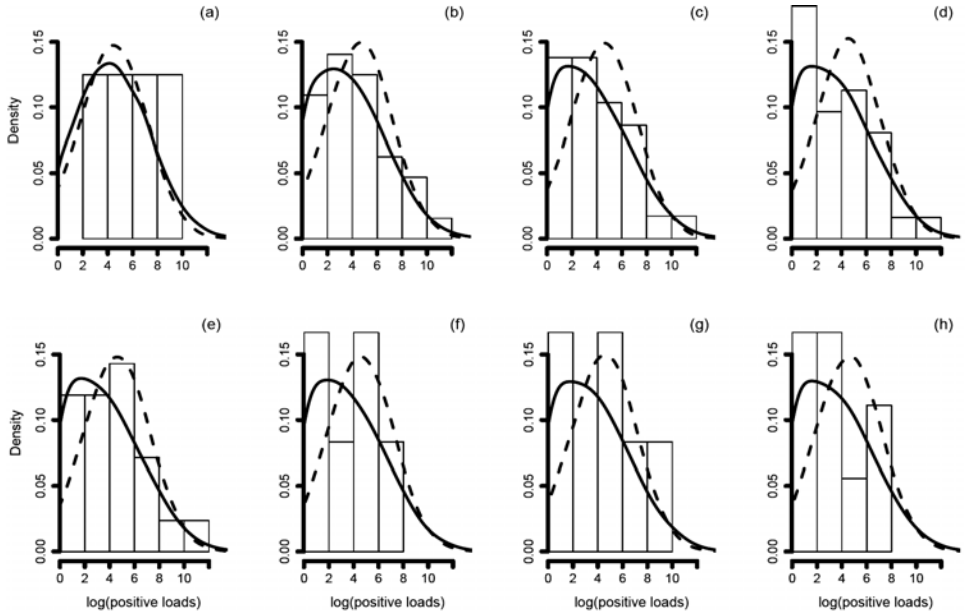

Figure 2. Plot of the estimated density functions for the log-transformed positive loads in the Kazakhstan sample by the PT model (solid line) and the CS model (dashed line) for the age classes (a) $(0, 0.5]$, (b) $(0.5, 1.5]$, (c) $(1.5, 2.5]$, (d) $(2.5, 3.5]$, (e) $(3.5, 4.5]$, (f) $(4.5, 5.5]$, (g) $(5.5, 6.5]$ and (h) $6.5+$, together with a histogram of the observed log-transformed positive loads for the above age classes. Note that there are 4, 32, 29, 31, 21, 8, 6 and 9 observed positive loads in the above age classes.

In the PT model, each of the parasites ingested by a dog has an independent, exponentially distributed lifetime, and in the CS model, the ingested parasites survive for a constant time. The two models are mathematically tractable, but may not exactly picture the true decay dynamics for ingested parasite clumps in dogs which are likely intermediate between the PT and CS mechanisms. We have thus averaged the values of the quantities such as infection pressure, mean load and mean survival time of a clumped infection obtained from these models. Alternative approaches not considered here would be to define the models in a Bayesian framework and use Bayesian model selection or to use intermediate decay functions as mixtures of exponentials (phase-type distributions).

Using Table 1, the averaged infection rates β are $(0.445 + 0.340)/2 = 0.393$, 0.575 and 0.218 infections per dog per year in Kazakhstan, Tunisia and China respectively. This could indicate that the prevalences of infection of sheep in Kazakhstan and China are lower than in Tunisia, or that dogs in Kazakhstan and China consume less viscera of sheep. The averaged mean μ of the log-transformed positive loads is with 5.152 in Kazakhstan larger than in Tunisia with 3.803 and China with 3.330. The averaged σ of 2.806, 2.886 and 2.585 for the Kazakhstan, Tunisia and China samples respectively are comparable. The corresponding mean clump sizes of a single infection are then 9000, 3000 and 1000 parasites for the Kazakhstan, Tunisia and China sample. The higher average clump size in Kazakhstan could be due to a higher number of fertile cysts in infected sheep, or that the age of sheep at slaughter are higher, or that there is a higher infection pressure in sheep from Kazakhstan, implying that cysts are acquired by sheep at younger ages and thus have more time to develop and become fertile.

The averaged mean durations of a single infection are with 0.745, 0.585 and 0.608 years for the Kazakhstan, Tunisia and China samples respectively similar. Hence we have an overall mean survival time of about 0.65 years ≈ 8 months, which is in line with the results in [14], who suggested that the mean duration of infection is slightly lower than 1 year. Our estimate is also in line with the about 10 months suggested in [3]. They experimentally infected 48 3-4 month old dogs with 25000 protoscoleces, and killed groups of 4 dogs at times regularly distributed over the year.

5. Discussion

In this paper, different shot noise processes are used to describe the ingestions of clumps containing *Echinococcus granulosus* parasites in dogs. The processes model clumped superinfections coupled with a time-dependent decay mechanism of the ingested parasite burden, namely an exponential decay with absorption around zero (MA model), a Poisson transform of the exponential shot noise process (PT model) and a constant duration of infection (CS model). Based on the skewness in the data, a lognormal distribution is chosen as distribution for the number of parasites per clump.

Maximum likelihood estimation is used to fit the models to samples from Kaza-khstan, Tunisia and China. The PT and CS model are shown to perform best. The results suggest that the infection rate is about 0.4, 0.6 and 0.2 infections per dog per year in Kazakhstan, Tunisia and China respectively. The lower values in Kaza-khstan and China could be the result of a lower consumption of infected viscera of sheep, due to a lower prevalence of infection in sheep or a different feeding behavior of dogs as compared to Tunisia. The mean number of parasites in a clumped infection is about 9000 in Kazakhstan, 3000 in Tunisia and 1000 China. The higher average clump size in Kazakhstan could be due to a higher number of fertile cysts in infected sheep, or that the age of sheep at slaughter are higher, or that there is a higher infection pressure in sheep from Kazakhstan, implying that cysts are acquired by sheep at younger ages and thus have more time to develop and become fertile. Hence the infection of dogs with *Echinococcus granulosus* occur at a low rate, but the ingested parasite load per clump is in the thousands. The mean duration of a single clumped infection is about 8 months, comparable in all three samples. The value is in line with other studies, [14] suggesting a mean time of slightly lower than 1 year, and [3] suggesting a mean time of about 10 months.

Acknowledgements

This work was supported by the Schweizerischer Nationalfonds (SNF), project no. 107726.

References

[1] J, Abate and P. P. Valko, Multi-precision Laplace inversion, Int. J. Numer. Meth. Eng. 60 (2004), 979-993.

[2] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, Edition 9, Dover, 1972.

[3] M. Aminzhanov, Duration of the life of Echinococcus granulosus in the organism of dogs, Veterinariia 12 (1975), 70-72.

[4] R. M. Anderson and R. M. May, Regulation and stability of host parasite population interactions, I, Regulatory processes, J. Anim. Ecol. 47 (1978), 219-249.

[5] T. E. Balling and W. Pfeiffer, Frequency distributions of fish parasites in the perch Perca fluviatilis *L*. from Lake Constancee, Parasitol Res. 83 (1997), 370-373.

[6] R. Barakat, Sums of independent lognormally distributed random variables, J. Opt. Soc. Am. 66 (1976), 211-216.

[7] N. C. Beaulieu and Q. Xie, An optimal lognormal approximation to lognormal sum distributions, IRE Trans. Comm. Sys. 53 (2004), 479-489.

[8] C. Braga, R. Ximenes, J. Miranda and N. Alexander, Bancroftian filariasis in an endemic area of Brazil: Differences between genders during puberty, Rev. Soc. Bras. Med. Trop. 38 (2005), 224-228.

[9] C. M. Budke, J. Qiu and P. S. Craig, Modeling the transmission of Echinococcus granulosus and Echinococcus multilocularis in dogs for a high endemic region of the Tibetan plateau, Int. J. Parasitol. 35 (2005), 163-170.

[10] D. R. Cox and V. Isham, Point Processes, Edition 2, Chapman and Hall, New York, 1980.

[11] P. K. Das, S. Subramanian, A. Manoharan, K. D. Ramaiah, P. Vanamail, B. T. Grenfell, D. A. P. Bundy and E. Michael, Frequency distribution of Wuchereria bancrofti infection in the vector host in relation to human host: Evidence for density dependence, Acta. Tropica. 60 (1995), 159-165.

[12] J. Eckert and P. Deplazes, Biological, epidemiological and clinical aspects of echinococcosis, a zoonosis of increasing concern, Clin. Microbiol. Rev. 17 (2004), 107-135.

[13] P. Economides and G. Cristofi, Cestode zoonoses: Echinococcosis and Cysticercosis, An Emergent and Global Problem, Edition 3, NATO Science Series, IOS Press Amsterdam, 2002.

[14] M. A. Gemmell, Hydatid diseases in Australia, IV, Observations on the incidence of Echinococcus granulosus on stations and farms in endemic regions of New South Wales, Aust. Vet. J. 35 (1959), 396-402.

[15] M. A. Gemmell, J. R. Lawson and M. G. Roberts, Population dynamics in echinococcosis and cysticercosis: Biological parameters of Echinococcus granulosus in dogs and sheep, Parasitology 92 (1986), 599-620.

[16] B. T. Grenfell, K. Wilson, V. S. Isham, H. E. Boyd and K. Dietz, Biological, Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study, Parasitology 111 (1995), 135-151.

[17] D. Heinzmann, A. D. Barbour and P. R. Torgerson, Compound processes as models for clumped parasite data, Math. Bios. 222 (2009), 27-35.

[18] J. Herbert and V. Isham, Stochastic host-parasite interaction models, J. Math. Biol. 40 (2000), 343-371.

[19] R. J. Irvine, A. Stien, J. F. Dallas, O. Halvorsen, R. Langvatn and S. D. Albon, Life-history strategies and population dynamics of abomasal nematodes in Svalbard reindeer (Rangifer tarandus platyrhynchus), Parasitology 120 (2000), 297-311.

[20] C. M. Kapel, P. R. Torgerson, R. C. A. Thompson and D. Deplazes, Reproductive potential of Echinococcus multilocularis in experimentally infected foxes, dogs, raccoon dogs and cats, Int. J. Parasitol. 36 (2006), 79-86.

[21] S. Lahmar, M. Kilani and P. R. Torgerson, Frequency distributions of echinococcus granulosus and other helminths in stray dogs in Tunisia, Ann. Trop. Med. Parasitol. 95 (2001), 69-76.

[22] R. B. Lund, R. W. Butler and R. L. Paige, Prediction of shot noise, J. Appl. Prob. 36 (1999), 374-388.

[23] N. B. Mehta, J. Wu, A. F. Molisch and J. Zhang, Approximating a sum of random variables with a lognormal, IEEE Trans. Wireless Comm. 6 (2007), 2690-2699.

[24] A. Nodtvedt, I. Dohoo, J. Sanchez, G. Conboy, L. DesCoteaux, G. Keefe, K. Leslie and J. Campell, The use of negative binomial modelling in a longitudinal study of gastrointestinal parasite burdens in Canadian dairy cows, Can. J. Vet. Res. 66 (2002), 249-257.

[25] A. Pugliese, R. Rosa and M. L. Damaggio, Analysis of a model for macroparasitic infection with variable aggregation and clumped infections, J. Math. Biol. 36 (1998), 419-447.

[26] M. G. Roberts, J. R. Lawson and M. A. Gemmell, Population dynamics in echinococcosis and cysticercosis: Mathematical model of the life-cycle of echninococcus granulosus, Parasitology 92 (1986), 621-641.

[27] S. Schwartz and Y. Yeh, On the distribution function and moments of power sums with lognormal components, IRE Trans. Comm. Sys. 8 (1982), 1441-1462.

[28] H. Stehfest, Algorithm 368: Numerical inversion of Laplace transforms, Comm. ACM. 13 (1970), 47-49.

[29] J. Stoer and R. Bulirsch, Introduction to Numerical Analysis, Edition 2, Springer, Berlin, 1983.

[30] R. C. A. Thompson and A. J. Lymbery, The Biology of Echinococcus and Hydatid Disease, George Allen and Unwin, London, 1986.

[31] P. R. Torgerson, K. K. Burtisurnov, B. S. Shaikenov, A. T. Rysmukhambetova, A. M. Abdybekova and A. E. Ussenbayev, Modelling the transmission dynamics of echinococcus granulosus in dogs in rural Kazakhstan, Parasitology 126 (2003), 417-424.

[32] P. R. Torgerson, B. Oguljahan, M. E. Muminov, R. R. Karaeva, O. T. Kuttubaev, M. Aminjanov and B. Shaikenov, Present situation of cystic echinococcosis in Central Asia, Int. J. Parasitol. 55 (2006), 207-212.

[33] P. R. Torgerson, I. Ziadinov, D. Aknazarov, R. Nurgaziev, P. Deplazes, Modelling the age variation of larval protoscoleces of echinococcus granulosus in sheep, Int. J. Parasitol. 39 (2009), 1031-1035.

[34] D. V. Widder, The Laplace Transform, Princton University Press, Princton, 1946.

[35] M. E. J. Woolhouse, C. Dye, J. F. Etard, T. Smith, J. D. Charlwood, G. P. Garnett, P. Hagan, J. L. K. Hii, P. D. Ndhlovu, R. J. Quinnell, C. H. Watts, S. K. Chandiwana and R. M. Anderson, Heterogeneities in the transmission of infectious agents: Implications for the design of control programs, PNAS 94 (1997), 338-342.

[36] Y. Xiao and R. Lund, Inference for shot noise, Stat. Inf. Stoch. Proc. 9 (2006), 77-96.

A. Computational Aspects

A.1. Laplace transform and sum of lognormals

All computations are carried out using Matlab® (Version 7.4.0). Theoretically, the approximation (13) becomes more accurate the greater N. In practice, there is an optimal value, beyond which numerical error starts to increase the total error (as a function of N) even if the theoretical error of the approximation continues to decrease. Set $\hat{L}_U(s) = \hat{L}_U^N(s)$ to make explicit the dependence of the approximation (13) on N. An appropriate N can for example be chosen such that for different parameter settings and different s, the relative error of $\hat{L}_U^N(s)$ in approximating $L_U(s)$ is smaller than 10^{-3} and that the difference $|\hat{L}_U^N(s) - \hat{L}_U^{N+1}(s)|$ is smaller than 10^{-6}. The value of $L_U(s) = \mathbb{E}(\exp(-sU))$ is approximated by simulation. For our implementation, we obtained $N = 20$.

To compute $\hat{L}_{X_t}$ for the shot noise process with $h(t) = \exp(-\lambda t)$, one can now either solve (14) by numerical integration or (15) by using some approximation for $E_1(a)$. Both approaches were carried out by standard routines implemented in Matlab for different parameter settings, and the latter approach was found to be significantly faster with almost identical accuracy to the numerical integration approach.

There is a trade-off in the choice for s_l $(l = 1, 2)$ when approximating the sum of lognormals (20), in that increasing s_l yields better estimation of the density function for small arguments, whereas reducing s_l yields a better estimation in the tails. As suggested in [23], we set $s_1 = 1$ and $s_2 = 2$ in (20). This setting provides a reasonable fit for the parameter configurations and values of K that we tested.

A.2. Likelihood inference

As for N in $\hat{L}_U(s) = \hat{L}_U^N(s)$, we need to find an optimal value for M in $\hat{f}_{X_t, M}$ defined in (17). An appropriate M is chosen by applying the algorithm

to test transforms and their (analytical known) inverses given in Table 1 of [1], where the test functions in the table have all singularities on the real axis to the left of $s = a$, $a \geq 0$, and are infinitely differentiable. The Laplace transform of one of the test functions is $\exp(-2\sqrt{s})$ which has a comparable curve to our function $\hat{L}_{X_t}(s)$. We obtain $M = 9$ for the present application. Integrating now $\hat{f}_{X_t, M}(x)$ over $(0, \infty)$ and adding the point mass $F_{X_t}(0)$ for different parameter settings yields values which are approximately 1. The results indicate that the approximation implemented works well for our case.

In addition, the performance of the maximum likelihood method for the shot noise process, first with exponential decay (without absorption at zero) and then with constant survival time, is evaluated by estimation from simulated populations with known parameter values of the corresponding model. The size for the simulated dog populations is set to 400 and the ages are drawn from an exponential distribution with mean 3, so that we produce a population somewhat similar to the samples from Kazakhstan, Tunisia and China. Then, based on the specified shot noise process, realizations of the process for the given ages of the dogs are taken as loads for the dogs in the sample. The shot noise process is then fitted by using maximum likelihood to the simulated data. The procedure is repeated 1000 times.

Applying this approach with different parameter settings reveals that the methods work well. Representatively, Figure 3 shows the application of the approach to the exponential decay model with fixed parameters $(\beta, \lambda, \mu, \sigma) = (2, 4, 6, 2)$ and for the constant survival model with fixed parameters $(\beta, t_d, \mu, \sigma) = (2, 1, 6, 2)$.

As indicated in the likelihood function (19) of the PT model, computations of higher order derivatives of the Poisson transform $\mathbb{P}(Y_t = y)$ can be approximated by $f_{X_t}(y)$ for $y \geq y_0$. To determine the parameter y_0, one can successively implement higher derivatives of $\hat{L}_{X_t}$ and test if $|\mathbb{P}(Y_t = y) - \hat{f}_{X_t, M}(y)|$ is smaller than some predefined threshold, say 10^{-4}. Our implementation yields $y_0 = 5$ and thus analytical derivatives of the Poisson transform are implemented only up to order 4.

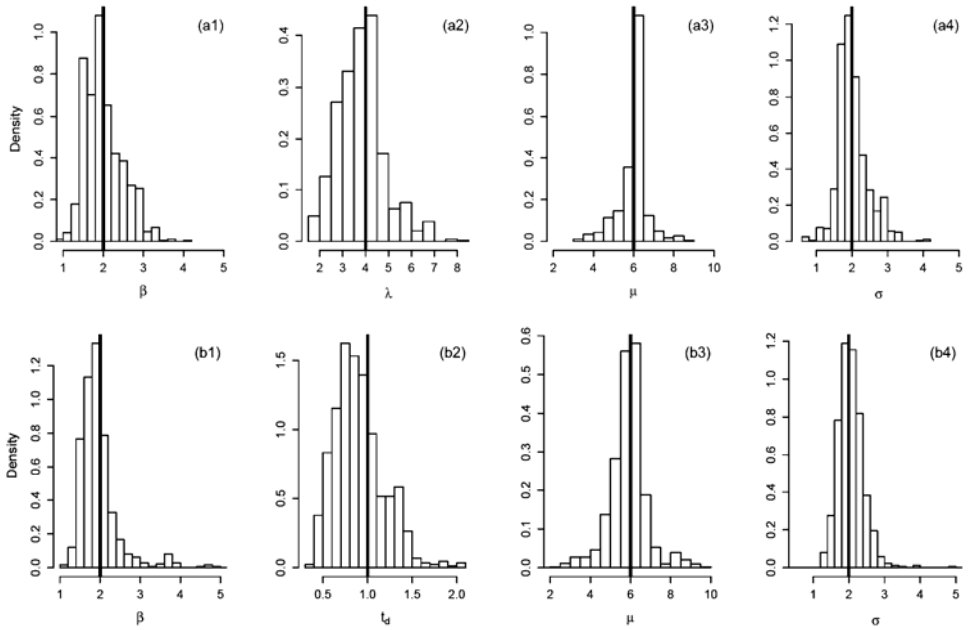

Figure 3. Histograms of the estimated model parameters from simulated data sets, where (a1-a4) the shot noise process with exponential decay with parameters $(\beta, \lambda, \mu, \sigma) = (2, 4, 6, 2)$ and (b1-b4) the constant survival model with parameters $(\beta, t_d, \mu, \sigma) = (2, 1, 6, 2)$ are used to generate the loads for given ages of the dogs for the generation of the samples. The true values are indicated by bold vertical lines.