Spatial Modelling of Emergency Service Response Times

Benjamin M. Taylor

January 28, 2016

Abstract

This article concerns the statistical modelling of emergency service response times. We apply advanced methods from spatial survival analysis to deliver inference for data collected by the London Fire Brigade on response times to reported dwelling fires. Existing approaches to the analysis of these data have been mainly descriptive; we describe and demonstrate the advantages of a more sophisticated approach. Our final parametric proportional hazards model includes harmonic regression terms to describe how response time varies with time-of-day and shared spatially correlated frailties on an auxiliary grid for computational efficiency.

We investigate the short-term impact of fire station closures in 2014. Whilst the London Fire Brigade are working hard to keep response times down, our findings suggest there is a limit to what can be achieved logistically: the present article identifies areas around the now closed Belsize, Downham, Kingsland, Knightsbridge, Silvertown, Southwark, Wesminster and Woolwich fire stations in which there should perhaps be some concern as to the provision of fire services.

1 Introduction

The thought of a fire in the home terrifies most people. In 2013-2014 UK fire and rescue services attended over half a million calls; there were a total of 322 fire-related deaths and 9748 non-fatal casualties due to fires of which 80% occurred in dwellings. There were 39600 dwelling fires in the UK in 2013-2014, with most occurring between the hours of 8 and 9pm at night and with misuse of equipment/appliances being the leading cause of around 1/3

these incidents (Department for Communities and Local Government, 2014). In London, 70% of fire-related deaths have been attributed to reporting delays, but crucial in saving lives is the efficient response of emergency services to calls to 999/112 (London Fire Brigade, 2013b).

The choice of where to locate emergency service stations (police, fire, ambulance) in cities has a direct impact on possible response times. From an academic perspective, this can be treated as an optimisation problem: how do we balance the need to respond quickly to emergency situations given a finite resource allocation (Toregas et al., 1971; Kolesar and Blum, 1973)? Whilst such approaches are very useful in helping to decide where might be best to build emergency service stations in the first place, the urban environment is constantly changing and there is therefore a need to continually monitor and improve information on response times in order to ensure safety standards are maintained. One aspect of this is the profiling and mapping of high risk groups, which has been undertaken in a limited way in the UK, one example being a pair of studies in Merseyside (Higgins et al., 2013, 2014). Another aspect is the study of response times to emergency calls, the subject of the present article.

The analysis of emergency response times has received a modest amount of attention in the literature. Scott et al. (1978) is one exception to this, these authors sought to form a mathematical model for ambulance response times in Houston. Other statistical approaches have focussed on predicting demand for emergency services such as Matteson et al. (2011), Vile et al. (2012) and Zhou et al. (2014). Recent concern over the UK Government's cuts to public services and their potential impact on the ability of fire services to maintain safety standards has resulted in a resurgence of interest, albeit primarily from the media and opposition parties (London Fire Brigade, 2013a,d; Open Data Institute, 2013; Bannister, 2014; Caven, 2014; Foley, 2014; Johnson, 2014; Westminster's Labour Councillors, 2014; Read, 2014; McCartney, 2015). The year 2014 saw 10 fire stations close: Belsize, Bow, Clerkenwell, Downham, Kingsland, Knightbridge, Silvertown, Southwark, Westminster and Woolwich.

In the present article our goal is to form a model for response times with the aim of providing emergency services with probabilistic information on where *in space* response times could be improved; clearly, faster response to fires in the home means more saved lives. Whilst in this article we focus on urban fires, it is worth noting the marked difference in response times to fires in urban compared with rural areas (Claridge and Spearpoint, 2013; Torney, 2013). In rural areas response times are usually longer and the results presented in the present article corroborate these findings: response times in the outskirts of the city, where there are fewer

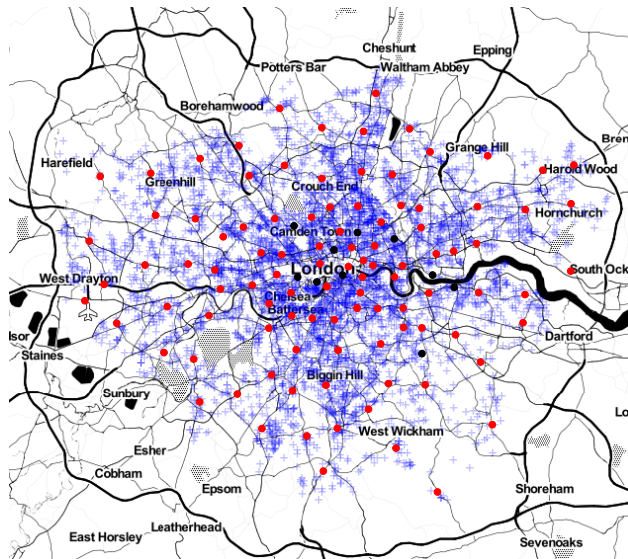fire stations and there is generally more open space, are typically longer than near the centre of the city.

Figure 1: Locations of the 5769 dwelling fires in 2014 (blue crosses) and the LFB stations (red dots are open stations, black dots are closed stations).

The London Fire Brigade (LFB) is one of the largest fire and rescue services in the world; they collect and analyse substantial amounts of data on incident response times (London Fire Brigade, 2014a). In addition to presenting tables of average response time by ward, there has also been some investment in the mapping of response times at this level of aggregation (London Fire Brigade, 2013c, 2014b) and in preparation for the proposed closures of 2014, an assessment of the potential impact of the closures was also carried out (Open Data Institute, 2013). Whilst these analytical efforts are to be highly commended, they could be further improved by the use of formal statistical models.

One of the main purposes of this article is to demonstrate ways in which the modelling-oriented approach could help to improve the description and presentation of emergency response time data and consequently inform city planners in their decision-making processes. To achieve this aim, we apply recently-developed techniques in survival analysis specifically designed for the modelling of large spatially referenced time-to-event data like the LFB response times (Taylor, 2015). Specific goals of the analysis are to (i) compare how well targets were met in 2014 to how well they were met in 2013 and 2012 and (ii) to identify regions around the closed fire stations that are not performing well with regards to the 6 minute target.

3

The remainder of this article is organised as follows. In Section 2 we introduce some basic concepts in spatial survival analysis and give details of the modelling approach proposed; in Section 3 we present results from the analysis of the LFB data; and in Section 4, the article concludes with a discussion.

2 Data and Model

The statistical analysis of time-to-event data is the realm of survival analysis (Cox and Oakes, 1984; Klein et al., 2013). Most often, survival methods are applied in clinical studies assessing the potential effect of treatments or exposures on the survival time of patients. Due to patients dropping out and the fact that studies are finite in duration, survival data are typically 'censored', which means the event of interest was not necessarily observed for all individuals. Survival methods handle the time to observed and censored events in a formal way.

Let T be a random variable, denoting the time after the call to 999/112 that the first fire engine arrives at the scene of a dwelling fire at location s. We will shortly introduce a model for the hazard function at location s defined formally by,

$$h(t,s) = \lim_{\Delta t \to 0} \left\{ \frac{\mathbb{P}(t \leq T \leq t + \Delta t | T \geq t, \text{the fire is at location } s)}{\Delta t} \right\}.$$

For any time t, the interpretation of $h(t,s)$ in this case is as the instantaneous arrival rate of the first engine at the scene of the fire at s conditional on the engine having not arrived before time t. The survival function, which is of particular interest in our current context can be derived from the hazard function,

$$S(t,s) = \mathbb{P}(T > t | \text{the fire is at location } s) = \exp \left\{ - \int_0^t h(x,s) \mathrm{d}x \right\},$$

and represents the probability that the engine will arrive after time t at location s.

Our ideal proportional hazards spatial survival model for the response times postulates the following form for the hazard function for the ith call:

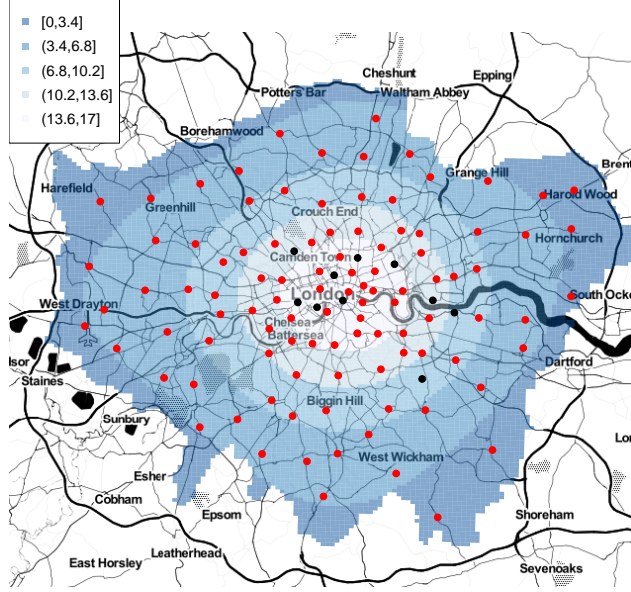$$h(t,s;\beta,Y(s)) = h_0(t) \exp\{X(s)\beta + Y(s)\}, \tag{1}$$

4

Figure 2: Smoothed intensity of fire stations. The red points are open fire stations, the black points are fire stations that closed in 2014.

where $X(s)$ are covariates associated with a call at location s, β are parameters, h_0 is the baseline hazard function (see below) and $Y(s)$ is the value of a spatially continuous Gaussian process, $Y(s)$, at s. We assume that the Gaussian process $Y(s)$ has associated covariance function,

$$\text{cov}[Y(s_1), Y(s_2)] = \sigma^2 \rho(\|s_1 - s_2\|; \phi)$$

where the parameter σ^2 is the unconditional variance of the process at any point and ρ is the correlation function with $\|s_1 - s_2\|$ being the Euclidean distance between s_1 and s_2.

In this model h_0 describes the part of the hazard function that is common to all individuals and the remaining term, $\exp\{X(s)\beta + Y(s)\}$, describes the relative risk for a call at location s. The relative risk splits into two parts: the first, $\exp\{X(s)\beta\}$, is the part of the risk that we can explain by the available covariates; and the second, $\exp\{Y(s)\}$ is the unexplained risk. With regards to the latter, we choose $\mathbb{E}[Y(s)] = -\sigma^2/2$, so that $\mathbb{E}[\exp\{Y(s)\}] = 1$ for all s in the observation window of interest, a subset of $\mathbb{R}^2$. When survival models are used to measure time to death, a high hazard is regarded as bad, since it means there is a high chance of an individual experiencing the event. In the present context however, we need to adopt the opposite meaning in which a low hazard is interpreted as bad: at time t if no engine has yet arrived at location s and $h(t, s)$ is low, the chance of one arriving in the immediate future is also low.
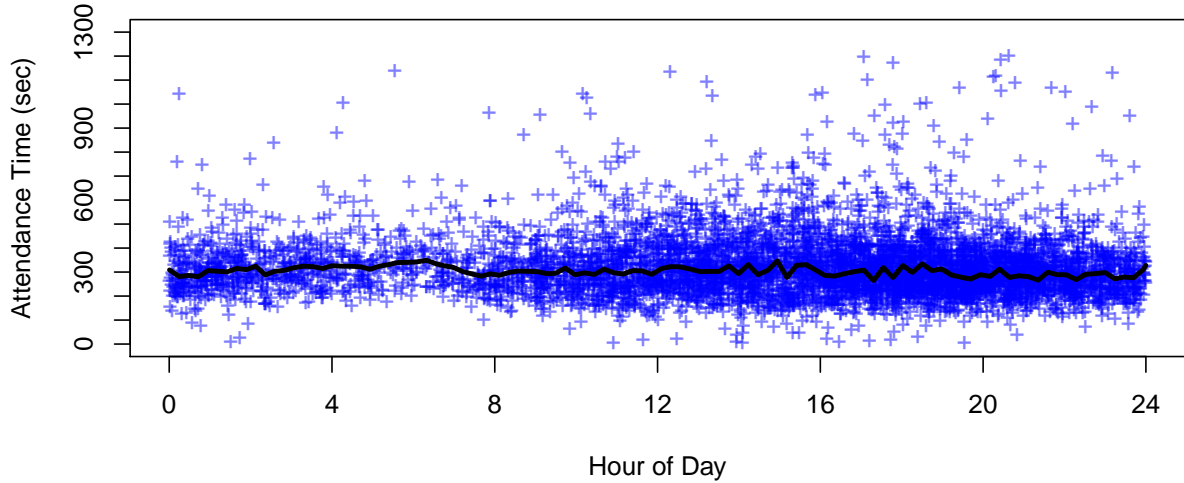
5

Figure 3: First pump attendance time by hour of day, the black line is a lowess smoother.

The two main options for modelling the baseline hazard are: (i) to assume a parametric form for h_0; or (ii) to leave h_0 unspecified, which results in a semiparametric model. Whilst the main advantage of the semiparametric approach is flexibility, in this article we opt for the former of these modelling paradigms because we are interested in probabilistic prediction. We considered two different parametric models for the hazard function for these data. The first is a simple Weibull model where,

$$h_0(t; \alpha, \lambda) = \alpha \lambda t^{\alpha - 1}, \qquad \alpha, \lambda > 0, \tag{2}$$

so that the baseline cumulative hazard takes the form:

$$H_0(t; \alpha, \lambda) = \lambda t^{\alpha}; .$$

In the second parametric model we mimic the flexibility of a semiparametric approach by modelling the baseline hazard using B-splines as in Rosenburg (1995), setting:

$$h_0(t; \omega) = \sum_{i=1}^{p} \exp\{\omega_i\} B_i^{(d)}(t), \tag{3}$$

where $\omega_1, \ldots, \omega_d$ are parameters to be estimated and $\{B_i^{(d)}(t) : i = 1, \ldots, p\}$ is the collection of B-spline basis functions of degree d (each of which is a piecewise positive polynomial), see

Younes and Lachin (1997) for details on how to construct these. Being piecewise polynomial, the baseline cumulative hazard function $H_0(t) = \int_0^t h_0(s)\mathrm{d}s$, required in likelihood computation, is trivial to compute provided we store the (piecewise) coefficients of the integrated basis functions:

$$H_0(t;\omega) = \sum_{i=1}^p \exp\{\omega_i\} \int_0^t B_i^{(d)}(x)\mathrm{d}x,$$

	False Alarm	Fire	Special Service
Aircraft	54	13	198
Boat	36	40	97
Dwelling	61680	18769	60367
Non Residential	62281	6460	8197
Other Residential	16699	1363	1937
Outdoor	6079	13845	6554
Outdoor Structure	2452	15824	1278
Rail Vehicle	68	47	156
Road Vehicle	3410	6089	14865

Table 1: Number of call-outs by incident group and property category between 2012 and 2014 inclusive.

Our final model is a slight modification of (1) introduced in Taylor (2015); this modification concerns the frailties $Y(s)$. Rather than assume these are spatially continuous we adopt a shared frailty approach, introducing an auxiliary grid of cells on which we wish to predict the response times; in other words, we approximate the spatially continuous process $Y(s)$ by a piecewise constant process on a fine grid. Our model for the hazard takes the form:

$$h(t, s_i; \omega, \beta, Y(s_i)) = h_0(t;\omega) \exp\{X_i(s)\beta + Y_{\mathcal{G}[s_i]}\}, \tag{4}$$

where s_i is the location of the ith fire and $Y_{\mathcal{G}[s_i]}$ denotes the value of the process $Y(s)$ in the cell containing s_i. Note that for each i, $\mathcal{G}[s_i] \in \{1,\ldots,M\}$, where M is the total number of grid cells in the grid. The reason for introducing this auxiliary grid is primarily computational efficiency: model (1) incurs $O(n^3)$ computational cost, where n is the number of observations, whereas model (4) is $O(n)$. With there being around 6000 dwelling fires each year, a Markov chain Monte Carlo (MCMC) algorithm used to deliver inference for model (1) would be impractical.

For a discussion of other potential models for these data, see Section 4.

3 An Analysis of the 2012–2014 Data

The data on LFB response times analysed in this article is available from the London Datastore (London Fire and Emergency Planning Authority, 2015). All analyses were carried out in the R statistical software (R Core Team, 2014). In this section, we present an analysis of the LFB data from 2012 to 2014 inclusive.

3.1 Preliminary Analyses and Model Choice

Table 1 shows the number of call-outs to fires, false alarms and special services by property type for the years 2012–2014 inclusive. We restrict our attention to analysing data from call-outs to fire events in dwellings because as mentioned above, dwellings are where most fire-related deaths occur. Figure 1 shows the locations of the 5769 dwelling fires in 2014 as well as the locations of all fire stations, including those that were closed in 2014. The pattern of the points roughly follows the distribution of the population in the city with higher concentration in the centre and with large open park areas being free of dwelling fires, for instance.
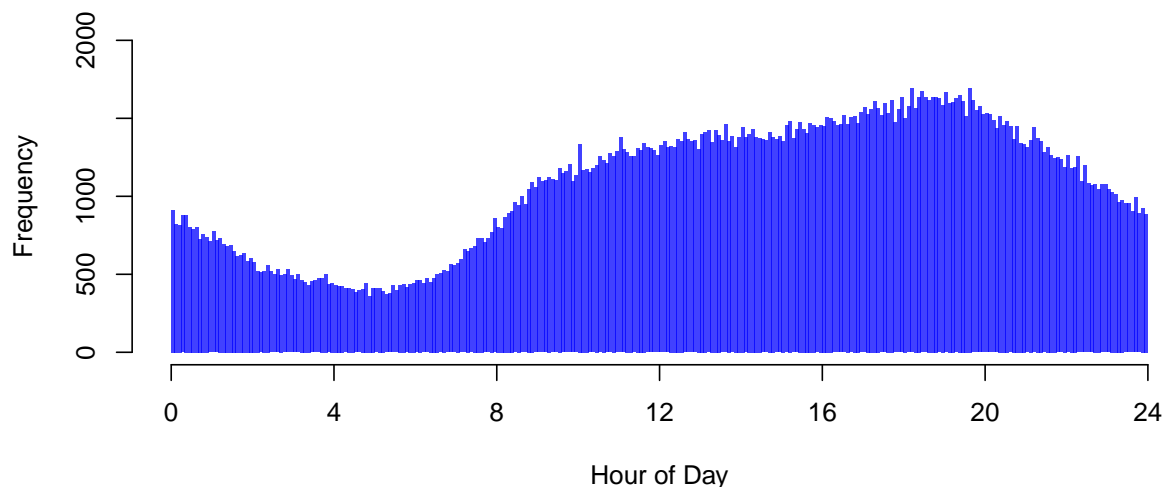

Figure 4:   Number of calls between 2012–2014 inclusive by hour of the day; each bar represents a 5 minute interval. The histogram includes call-outs to fires, special service and also false alarms.

It can also be seen from Figure 1 that the intensity of locations of fire stations is also more

8

concentrated towards the centre of the city, a bivariate isotropic Gaussian smoothing of these points is shown in Figure 2 (the intensity was scaled by a factor of 10^8). The bandwidth used to compute the smooth intensity was chosen using the rule of thumb method of Baddeley and Turner's `density.ppp` function from the `spatstat` package, resulting in a kernel standard deviation of 5737 (Taylor and Rowlingson, 2014). The result is quite a smooth approximation to the intensity of the fire stations which is desirable for the use we put it to. The relationship between response time to proximity of the nearest fire station is not clear cut: it is not necessarily the closest station that will respond to a call and for this reason, we use the smoothed intensity as a covariate in our model. We expect the coefficient of this covariate to be positive, since in places where there are a greater concentration of fire stations, fires are likely to be attended more swiftly.

Whilst the plot of response time by time of day, shown in Figure 3 suggests no obvious trends, it is reasonable to assume that time of day does in fact influence response time due to traffic congestion and local demand on the fire service. The demand by time of day is illustrated in Figure 4, this shows that it is lowest at around 5:30am, highest around 7pm and generally higher in the day-time than at night. There was no obvious difference in response times comparing weekdays with weekends: the median response time in 2014 for weekdays was 299 seconds and for weekends it was 296.5; again, though there is no obvious difference between these values, it is reasonable to assume that we might expect a difference in response times due in part to differing traffic patterns on weekdays compared with weekends.

Our overall model for the hazard function therefore includes the intensity of fire stations and a time trend through harmonic regression terms: $\sin(2\pi kt/24)$ and $\cos(2\pi kt/24)$ with $k = 1, 2, 3, 4$ i.e. with periodicity 1 day, 1/2 day, 1/3 day and 1/4 day. The number of harmonic terms was chosen by fitting a non-spatial version of the model to the 2010 data using forward selection, terminating the process when both the sin and cos contributions were not significant at the 5% level. We initially fitted the spatial Weibull model which included a weekday/weekend indicator variable as an additional covariate, but since this was not found to be significant, we fitted the B-spline model without this term. For the B-spline model, the baseline hazard function, h_0 was modelled using a piecewise cubic B-spline function of the form (3) with 5 internal knots at the minimum, 0.2, 0.4 and 0.6 quantiles and at the maximum response time; with a repeated knot at each end-point the spline has a total of seven parameters. We used an exponential covariance function for the spatial random effects, that is

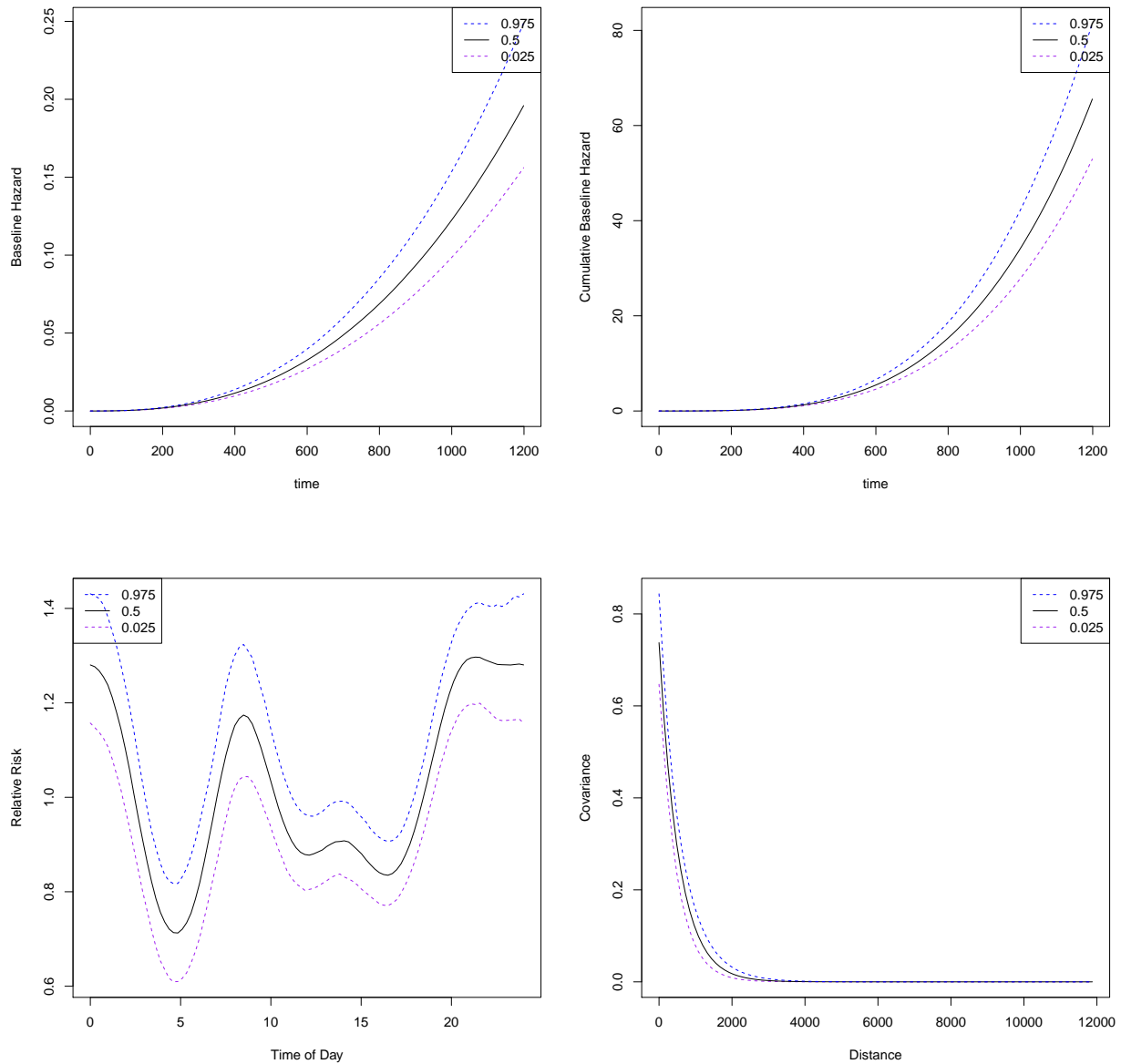$$\text{cov}[Y(s_1), Y(s_2)] = \sigma^2 \exp\{-\|s_1 - s_2\|/\phi\}.$$

Figure 5: Top left: baseline hazard. Top right: baseline cumulative hazard. Bottom left: relative risk by time of day (compare this with Figure 3, which shows no obvious trend). Bottom right: posterior covariance function. These plots are based on the results from 2014, similar plots from 2012 and 2013 are not materially different.

.

	Year	50%	2.5%	97.5%
intensity	2012	0.101	7.64×10^{-2}	0.129
weekend	2012	4.83×10^{-2}	-1.55×10^{-2}	0.113
α	2012	3.85	3.76	3.95
λ	2012	1.05×10^{-10}	6.05×10^{-11}	1.89×10^{-10}
σ	2012	0.879	0.823	0.94
ϕ	2012	772	652	929
intensity	2013	9.82×10^{-2}	7.85×10^{-2}	0.118
weekend	2013	-4.63×10^{-4}	-6.69×10^{-2}	6.45×10^{-2}
α	2013	3.73	3.64	3.82
λ	2013	2.18×10^{-10}	1.21×10^{-10}	3.76×10^{-10}
σ	2013	0.859	0.806	0.916
ϕ	2013	583	503	687
intensity	2014	6.4×10^{-2}	4.52×10^{-2}	8.2×10^{-2}
weekend	2014	1.64×10^{-2}	-5.79×10^{-2}	8.52×10^{-2}
α	2014	3.58	3.5	3.69
λ	2014	6.09×10^{-10}	3.41×10^{-10}	1.02×10^{-9}
σ	2014	0.861	0.809	0.926
ϕ	2014	537	458	634

Table 2: Table of parameter estimates from the three models fitted.

3.2 Description of Inferential Procedure

Model (4) was fit separately to the 2012, 2013 and 2014 datasets using the R package
spatsurv (Taylor and Rowlingson, 2014). This package implements a fully Bayesian adaptive MCMC algorithm, which delivers inference for the model parameters β and ω, the shared
frailties $Y_{\mathcal{G}[i]}$ and the parameters of the process $Y(s)$ (σ and ϕ). We used Gaussian priors for
all parameters on an appropriately transformed scale: $\pi(\beta) \sim \mathrm{N}(0, 100^2)$, $\log\omega \sim \mathrm{N}(0, 10^2)$,
$\log\alpha \sim \mathrm{N}(0, 10^2)$, $\log\lambda \sim \mathrm{N}(0, 10^2)$, $\log\sigma \sim \mathrm{N}(0, 0.5^2)$ and $\log\phi \sim \mathrm{N}(\log 1000, 0.5^2)$. As per
the method described in Taylor (2015), we used a $\mathrm{N}(0, 1)$ prior for a whitened version of the
spatial process, Γ where $Y(s) = -\sigma^2/2 + \Sigma_{\sigma,\phi}^{1/2}\Gamma$, where $\Sigma_{\sigma,\phi}$ is the covariance matrix of $Y(s)$
on the auxiliary grid; we retained the transformed samples $\{Y_j\}_{j=1}^{16384}$. We chose the size of
cells in the computational grid to be 500m×500m.

For all of the Weibull models and for the 2014 and 2013 B-spline models, we ran the samplers
for 500,000 iterations with a 10,000 iteration burn in and retaining every 490th sample for
inference and verified convergence by examining a plot of the log posterior over the retained

11

Year	B-spline	Weibull
2012	71937.75	67976.35
2013	70454.97	65221.69
2014	66222.56	62177.38

Table 3: Table comparing DIC values between the Weibull and B-spline models for each year.

iterations (Figure 10) which showed that the retained chain had left the transient phase and was at stationarity. The 2012 chain for the B-spline model required a longer burn in period and was run for 600,000 iterations with a 110,000 iteration burnin, again retaining every 490th sample. Plots of autocorrelation in the $Y(s)$ chains (for our final Weibull model) at lags 1, 5 and 10 are shown in Figure 11 confirm that the the chain was mixing well: the autocorrelation in all cases had dropped to a negligible amount on or before the 5th lag. The B-spline chain mixed more slowly but by lag 10 autocorrelation in the 2013 and 2014 chains was low; the 2012 chain was mixing a little more slowly. For model comparison purposes we considered these chains to be sufficiently well mixing to decide between the Weibull and B-spline models. Diagnostic plots for the other chains in our final Weibull model are available by following the links appearing in Section 5.

This *is* a challenging sampling problem: with around 6,000 observations and for technical reasons $M = 256 \times 256 = 65,536$ prediction points (reducing to an output grid of size 128×128) each chain takes around 7.5 hours to run; we estimate that the full spatially continuous model would take over 5 months to run on a dataset of this size.

We chose between the Weibull and B-spline models using the DIC, shown in Table 3. It is interesting to note that the DIC from the simpler Weibull model is lower than that for the more flexible B-spline model in each of the years considered. The parameters of the baseline hazard function are well identified by the data since every observation provides information about them; in contrast each observation provides limited information about the Gaussian process $Y(s)$ except in the locality of the observation. The baseline hazard function for the Weibull model does not capture the shape of the hazard function well compared with the B-spline model (compare the top left plot in Figure 5 to the left plot in Figure 12). The lower DIC values for the Weibull model therefore suggest that the Gaussian process term in that model explains the spatial variation in residual response times better when compared with the B-spline model. Note that although the two baseline hazard functions look dissimilar, for times less than about 370 seconds they are in fact quite similar; 74% of the response times in 2014 were under 370 seconds.

3.3   Summary of Results

The result of fitting this model is a sample from the joint posterior density of all model parameters, $\pi(\beta, \omega, Y_{1:65536}, \eta|\text{data})$, from which we can compute expectations of quantities of interest; Figure 5 shows several such quantities.

Beginning with the baseline hazard, we report here the shape of the baseline hazard and cumulative hazard. The plot of the baseline hazard gives the instantaneous arrival rate conditional on there not having been an arrival so far, with the covariates and frailties in the model set to zero. Whilst the hazard function itself is more easily interpretable, it is more difficult to visualise as it varies according to location because both the fire station intensity and the frailties vary over space. The baseline hazard function is nevertheless a useful plot because it provides a global representation of the hazard, which is then scaled by the intensity of fire stations and the frailty terms in different areas of space. Under the assumed conditions, this plot shows the remarkable speed with which engines arrive on the scene of a fire: the baseline hazard is quite flat for around the first 100 seconds and then starts to increase steadily. The baseline cumulative hazard (top right) can be interpreted as the expected number of arrivals of first fire engines up to a given time if those events were repeatable, the other covariates in the model being set to zero. It is difficult to see from this plot because of the scale, but under the assumed conditions, we would expect one fire engine to arrive at the scene of a fire by around 6 minutes, the London Fire Brigade target response time.

Whilst the baseline hazard and cumulative hazard describe global properties of the process generating these data, the arrival rate of fire engines does depend on time of day and space as will now be illustrated. The bottom left plot in Figure 5 shows the relative risk by time of day: we do not present the coefficients of the harmonic regression terms here, as a plot is much simpler to interpret. This plot shows that there are two main times of day when services take longer than usual to arrive on the scene of a fire: between 3am and 7am and between 11am and 6pm the relative risk is significantly below 1 and reaches its lowest value of around 0.7 at 5am. It is interesting to note that this pattern was not observed in Figure 3, in which arrival times appear to be independent of the time of day. Under our proposed model, the variation in response times has been broken down into variation we can explain through the use of covariates and variation that we cannot explain through the inclusion of spatially correlated random effects. In Figure 3, both of these sources of variation contribute to the observed variation in response times over the day; it would appear therefore that
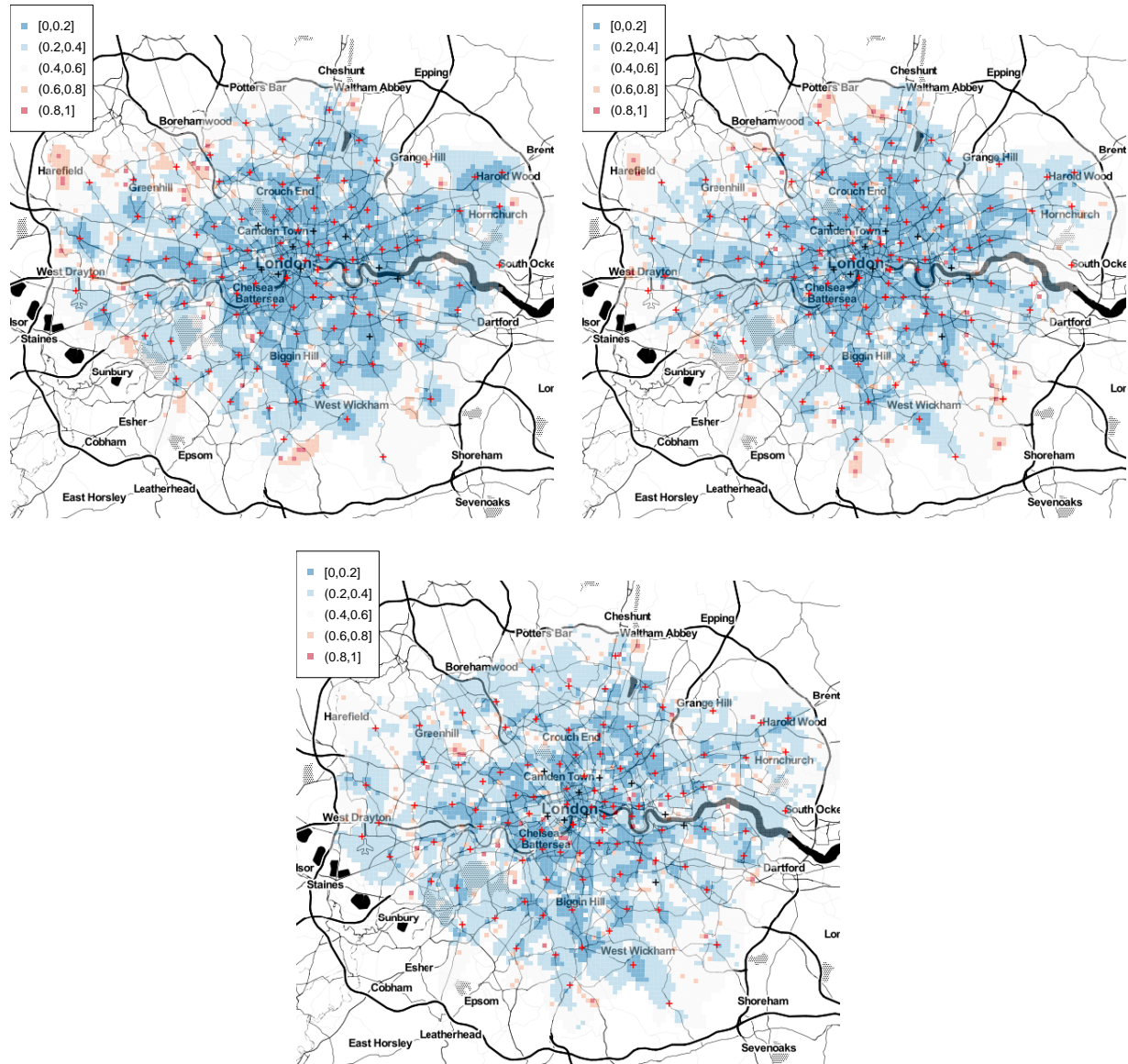
Figure 6: Plots showing $\mathbb{E}[S(360, s)]$, the expected probability that the response time will be greater than 360 seconds for each location s that is within the Greater London region; top left: for the year 2012, top right: for the year 2013 and in the bottom plot: for the year 2014. Red crosses are open stations, black crosses are stations that closed in 2014.

adjusting for the spatial variation in risk has bought out this trend.

The bottom right plot in Figure 5 shows the posterior covariance function with 95% confidence interval. This shows that spatial correlation is over quite a short range, around 0-1000 metres. Plots comparing the prior to the posterior for the parameters σ and ϕ showed that these were well identified by the data (the identifiability of ϕ is a common problem in spatial analyses). Table 2 gives the estimated coefficients from the Weibull model for the three years under consideration, it can be seen that the coefficients are quite similar for each year.

Using the spatial survival modelling framework, we can also illustrate answers to questions of substantive interest including (i) where in space is the London Fire Brigade's target response time of 6 minutes not being met; and (ii) what have been the effects on target response times of the 2014 fire station closures?

We answer the first question using the expected probability that the response time to a call will take longer than 6 minutes. For a response time in cell i of the computational grid and at location s, this is evaluated as,

$$\mathbb{E}[S(360, s)] = \frac{1}{1000} \sum_{j=1}^{1000} S(360, s; \beta^{(j)}, \omega^{(j)}, \eta^{(j)}, Y_{\mathcal{G}[s]}^{(j)}), \tag{5}$$

where $S(360, s; \beta^{(j)}, \omega^{(j)}, \eta^{(j)}, Y_{\mathcal{G}[s]}^{(j)})$ is the survival function at location s evaluated for the jth retained sample of each parameter in the model. In computing S in the above, we assumed that the fire occurred on a weekday at 8:30pm, since as stated above, most dwelling fires occur between 8 and 9pm. The resulting plot is shown in Figure 6, where we have masked the computational grid cells appearing outside of the London boroughs. This plot shows that whilst there has been an improvement over the last three years in responding to reported fires in the outskirts of the city, there are some areas in the inner part of the city in 2014 where the expected probability the response will take longer than 6 minutes is slightly elevated compared with the other years – the colour has shifted from dark blue to light blue/neutral and red in some small areas, the largest stretching from below Westminster to around Camden Town and also of note the area around Gallions Point Mariana. Some caution must be maintained in not over-interpreting these maps, as these are point estimates, they are subject to uncertainty.

To account for this uncertainty and to make a comparison between years and thus address the second substantive question, in Figure 7, we plotted $\mathbb{P}[S_{y_1}(360, s) > S_{y_2}(360, s) + c]$ for
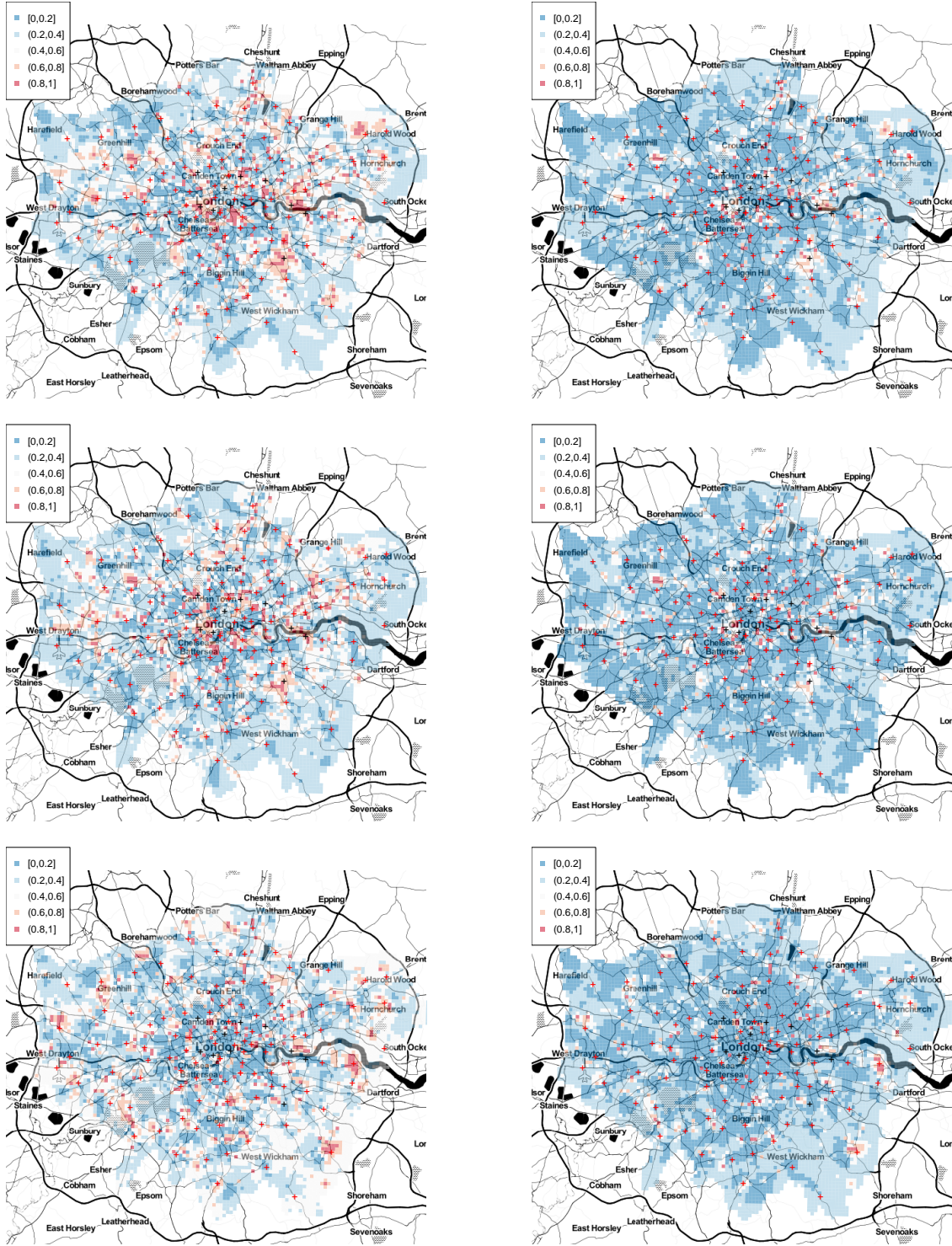
Figure 7: Top row: the left plot shows $\mathbb{P}[S_{2014}(360, s) > S_{2012}(360, s) + 0.1]$ and the right plot shows $\mathbb{P}[S_{2014}(360, s) > S_{2012}(360, s) + 0.25]$. The middle row of plots compares 2014 with 2013 and the bottom row of plots compares 2013 with 2012; see the main text for further details. Red crosses are open stations, black crosses are stations that closed in 2014.
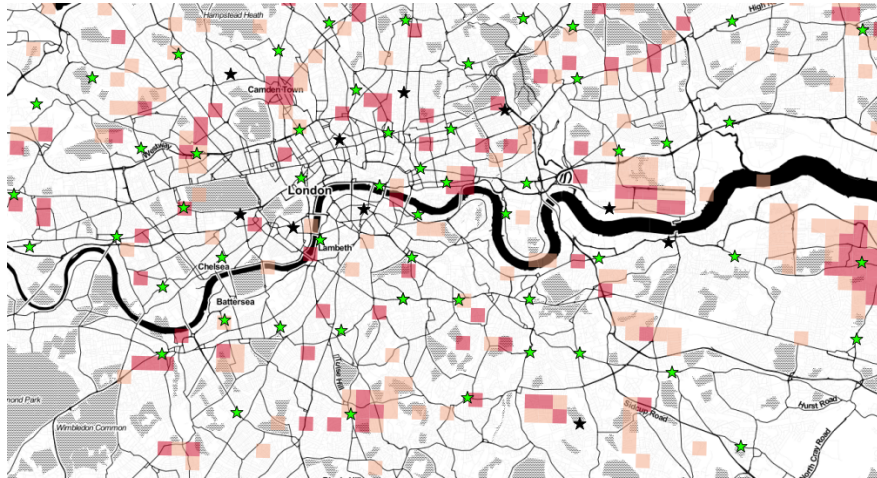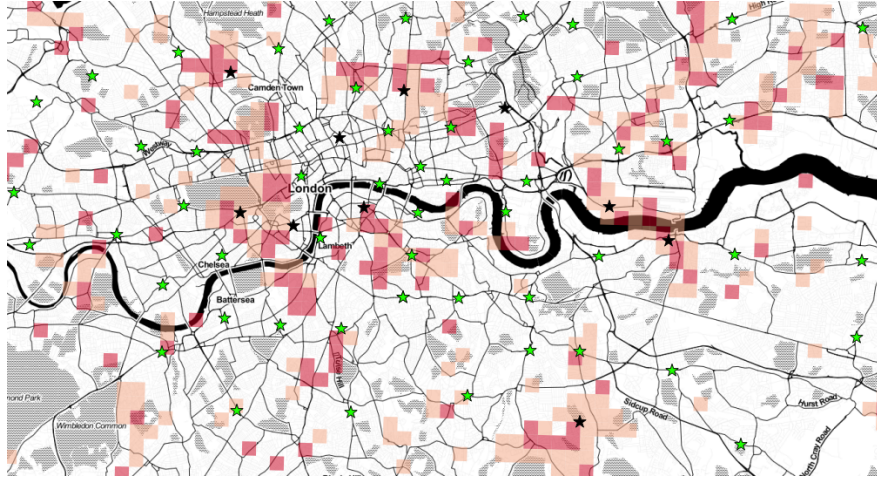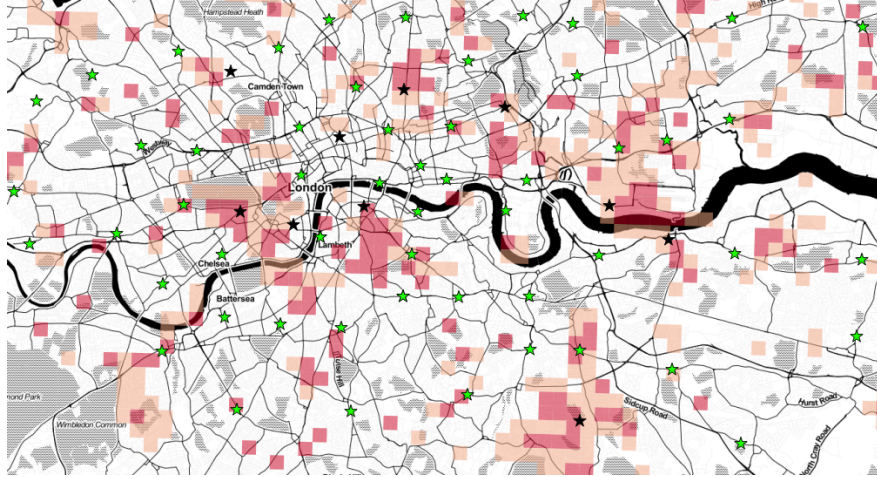
Figure 8: Top: $\mathbb{P}[S_{2014}(360, s) > S_{2012}(360, s) + 0.1]$. Middle Top: $\mathbb{P}[S_{2014}(360, s) > S_{2013}(360, s) + 0.1]$. Bottom Top: $\mathbb{P}[S_{2013}(360, s) > S_{2012}(360, s) + 0.1]$. Light red represents a probability between 0.6 and 0.8 and dark red represents a probability between 0.8 and 1. Green stars are open stations, black stars are stations that closed in 2014.

$y_1, y_2 \in \{2014, 2013, 2012\}$ with $y_1 > y_2$ and $c \in \{0.1, 0.25\}$ and s inside the Greater London region. In the case that $y_1 = 2014$, $y_2 = 2012$ and $c = 0.1$ these probabilities were computed as:

$$\frac{1}{1000} \sum_{j=1}^{1000} \mathbb{I}[S_{2014}(360, s; \beta^{(j)}, \omega^{(j)}, \eta^{(j)}, Y_{\mathcal{G}[s]}^{(j)}) > S_{2012}(360, s; \beta^{(p(j))}, \omega^{(p(j))}, \eta^{(p(j))}, Y_{\mathcal{G}[s]}^{(p(j))}) + 0.1],$$

(6)

for each location s and where $p(j)$ denotes a permuted index of the sample. In computing S, we again assumed that the fire occurred on a weekday at 8:30pm. The interpretation of $\mathbb{P}[S_{2014}(360) > S_{2012}(360, s) + 0.1]$, for example, is the proportion of times the probability of the response time in 2014 exceeding 360 seconds is at least 0.1 bigger than the probability of the response time in 2012 exceeding 360 seconds. Figure 7 shows these probabilities on a map of London; the top row compares 2014 with 2012, the middle row compares 2014 with 2013 and the bottom row compares 2013 with 2012. The main points of interest from these plots are the red areas, around the now closed Belsize, Downham, Kingsland, Knightsbridge, Silvertown, Southwark, Wesminster and Woolwich fire stations. Around or near to these stations there are regions where we are over 80% confident that the probability the response time is greater than 6 minutes in 2014 were at least 0.1 bigger than in 2013 or 2012. These plots give an idea about the size and location of regions potentially affected by the closures. The area around the now closed Bow station has not been so badly affected: whilst there was an increase compared with 2012, there was not compared with 2013. Areas around the Clerkenwell station do not currently appear to have been affected, at least with respect to responses to dwelling fires.

These areas are more easily seen in Figure 8, which shows the plots for $c = 0.1$ in an area around the closed stations. The important point to note here is that the spatial pattern of these probabilities comparing 2014 with 2012 and 2014 with 2013 are very similar, whereas the pattern of these probabilities comparing 2013 with 2012 is completely different. The red areas in the 2014/2012 and 2014/2013 plots are the areas that have been most affected by the closures.

We used these probabilities to identify small 500×500m squares near the closed fire stations that are of potential concern in terms of response time. We identified regions as those squares satisfying conditions (i) $\mathbb{P}[S_{2014}(360, s) > S_{2013}(360, s) + 0.1] > 0.7$ (ii) $\mathbb{P}[S_{2014}(360, s) > S_{2012}(360, s) + 0.1] > 0.7$ and (iii) $\mathbb{P}[S_{2013}(360, s) > S_{2012}(360, s) + 0.1] < 0.3$; i.e. areas in which response times seemed higher in 2014 compared with 2013 and 2012, but in which the probability of an increase between 2012 and 2013 was low. We identified an area of

interest around the closed fire stations by constructing the convex hull of the locations of the closed fire stations and extending it by a buffer zone of 4km. The left plot in Figure 9 shows in red the regions meeting criteria (i) to (iii) above. The middle plot in this figure is a box and whisker plot of response times in the red regions; there were a total of 705 calls in 2012–2014 in these small areas. The right hand plot is the same but illustrates times for all small regions inside the area of interest; there were are total of 8276 of these. Calls within the red regions of potential concern therefore accounted for 9% of all calls in the area of interest surrounding the closed stations. It can be seen from the two box and whisker plots that whilst in the region as a whole, the fire brigade appears to mainly be meeting their 6 minute target (right plot), in these small areas of potential concern near to the closed fire stations, there is a definite increase in response times in 2014; the median in these areas is above the 6 minute target.
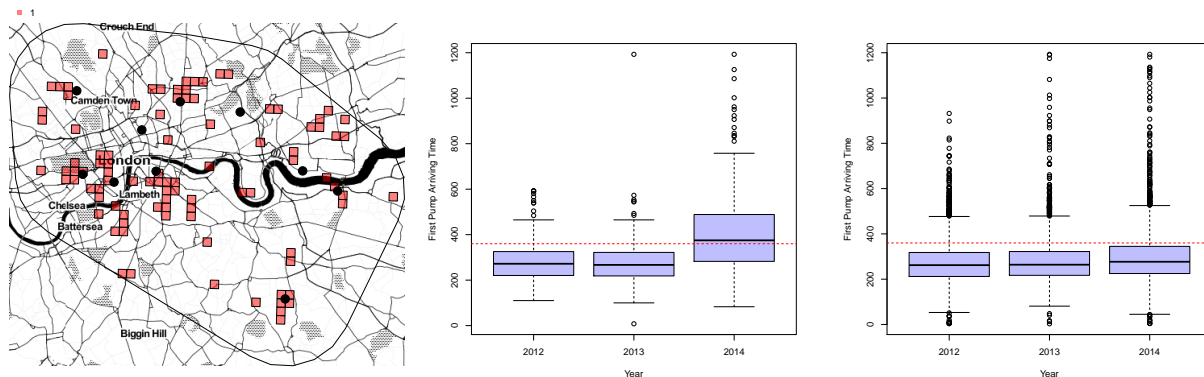

Figure 9: The left plot shows the central area of London under consideration: the convex hull of the locations of the closed fire stations with a 4km buffer; closed fire stations appear as black dots. The middle plot shows is a box and whisker plot of 705 response times in potentially problematic areas (see text for explanation), coloured in red in the left plot. The right hand plot is a box and whisker plot of all the 8276 response times in the buffered zone shown.

4   Discussion

In this article we have shown how advanced methods from spatial survival analysis can be used to model emergency service response times. We have applied these methods to the London Fire Brigade data and have illustrated the impact of the 2014 closures on response times. We have identified areas of potential concern surrounding the recently closed stations; in these areas the median response time exceeds the London Fire Brigade's target of six minutes.

While there may be simpler ways to model these data such as kriging the log-response times, the fitting of a spatial survival model is advantageous. The survival analysis framework is a 'natural' way to handle these time-to-event data and also the hazard, cumulative hazard and survival functions are useful for describing properties of the data that are of genuine interest in this context (particularly the survival function). One possible criticism of the choice of methodology in this article is that it is too complex because of the absence of censoring in our data, however survival analyses are frequently used in the absence of censoring.

Other possible models for the data discussed here could have accounted for spatiotemporal dependence between the years under consideration in our analysis. For example a hazard function of the form,

$$h(t, s; \beta, Y(s, t)) = h_0(t) \exp\{X(s)\beta + Y(s, t)\}, \tag{7}$$

where $Y(s, t)$ is a spatiotemporal Gaussian process would account for such temporal dependence in the frailty process. To the author's knowledge there is no available software to fit such a model, however such software is currently under development in the `spatsurv` R package (Taylor and Rowlingson, 2014). Such a model would account for the correlation between years that is currently ignored in Equation 6. One issue with such a spatiotemporal model for the hazard in the present context is the assumption that $Y(s, t)$ is a Gaussian process: it is not clear that such a modelling choice could adequately model the abrupt changes in response times due to the station closures in 2014.

Another alternative to the model in Equation (1) could include non-spatial frailties as well as spatial frailties:

$$h(t, s; \beta, Y(s)) = h_0(t) \exp\{X(s)\beta + Y(s) + U(s)\}, \tag{8}$$

where $U(s) \sim \mathrm{N}(-\sigma_U^2/2, \sigma_U^2)$ are non-spatial random effects. In practice, these non-spatial random effects could be fit at either the individual fire level, or at the level of the computational grid. We have not fitted such models in the present analysis, and these may lead to different areas of London being identified as 'potentially problematic'. However, the areas identified in Figure 9 together with the associated box plots in that figure remain an important issue to residents living in those areas, regardless of the manner in which these regions have been identified. Further analyses on these data could examine jointly the arrival times of first and second fire engines using a bivariate model; again we have not pursued this matter in the present article.

The proposed inferential framework is advantageous as it provides a sample from the joint posterior of all model parameters including the spatial process $Y(s)$ on all prediction cells. These samples can be used to deliver posterior expectations of functionals of interest, some of which have been illustrated in this article. The main drawback with the proposed method is computational cost. We estimate it would take over 5 months to run the full model in Equation (1), so the 7.5 hours our sampler takes represents a substantial reduction in cost. Other techniques to speed up the MCMC such as using a Gaussian Markov random field to represent the spatial process on the auxiliary grid are also not as fast as the proposed method: the Fourier methods applied here scale as $O(M \log M)$ compared with $O(M^{3/2})$ for sparse matrix methods, though the latter have the advantage of being able to work on non-regular and non-rectangular grids.

Ackowledgements

5   Appendices

- 2012 diagnostic plots: `http://www.lancaster.ac.uk/staff/taylorb1/londonfires/2012/traceplots_2012.html`

- 2013 diagnostic plots: `http://www.lancaster.ac.uk/staff/taylorb1/londonfires/2013/traceplots_2013.html`

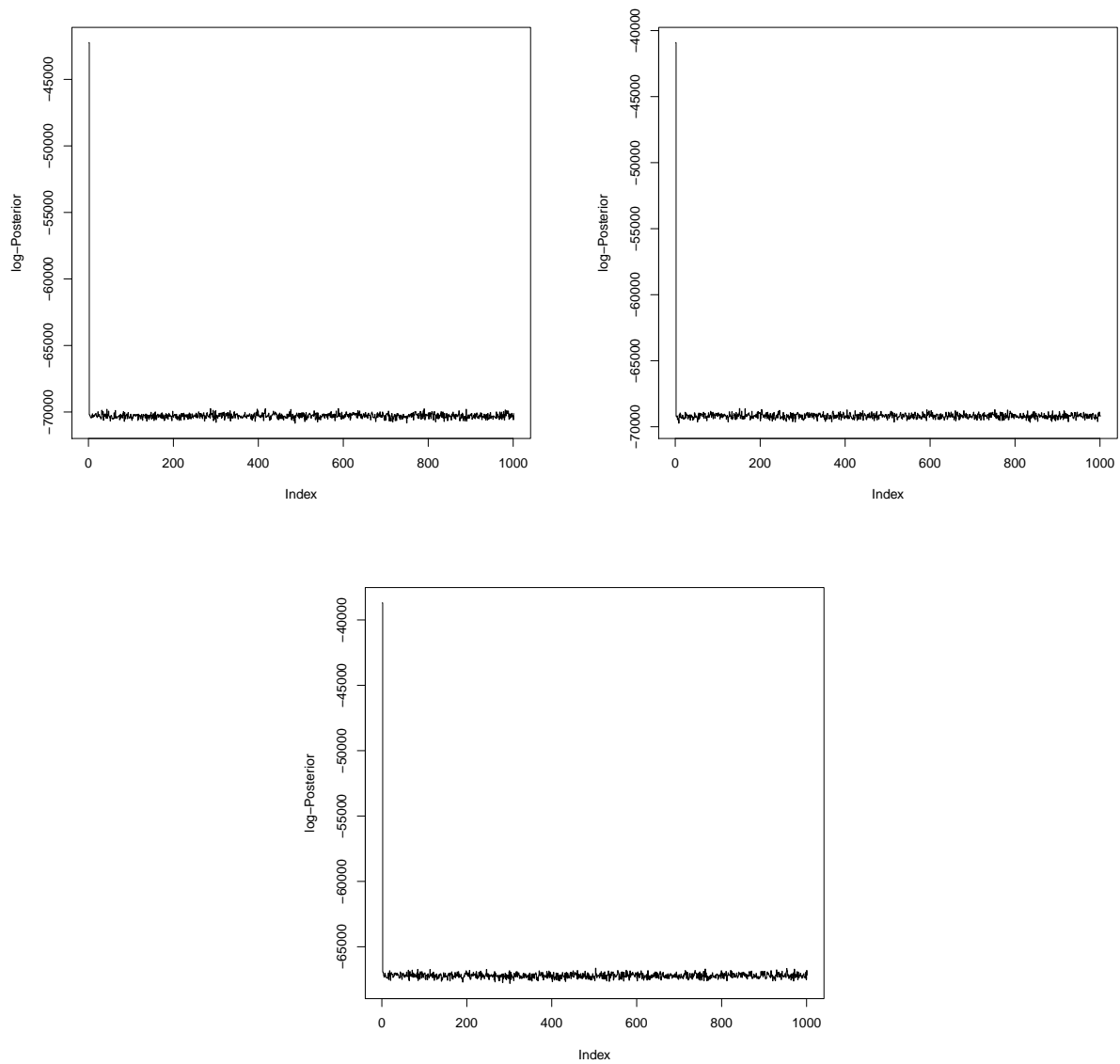- 2014 diagnostic plots: `http://www.lancaster.ac.uk/staff/taylorb1/londonfires/2014/traceplots_2014.html`

Figure 10: Plot of log posterior evaluated at the initial value and over all retained iterations of the chain. This shows that initially the chain started far away from a mode, but had found one before the burnin had finished.
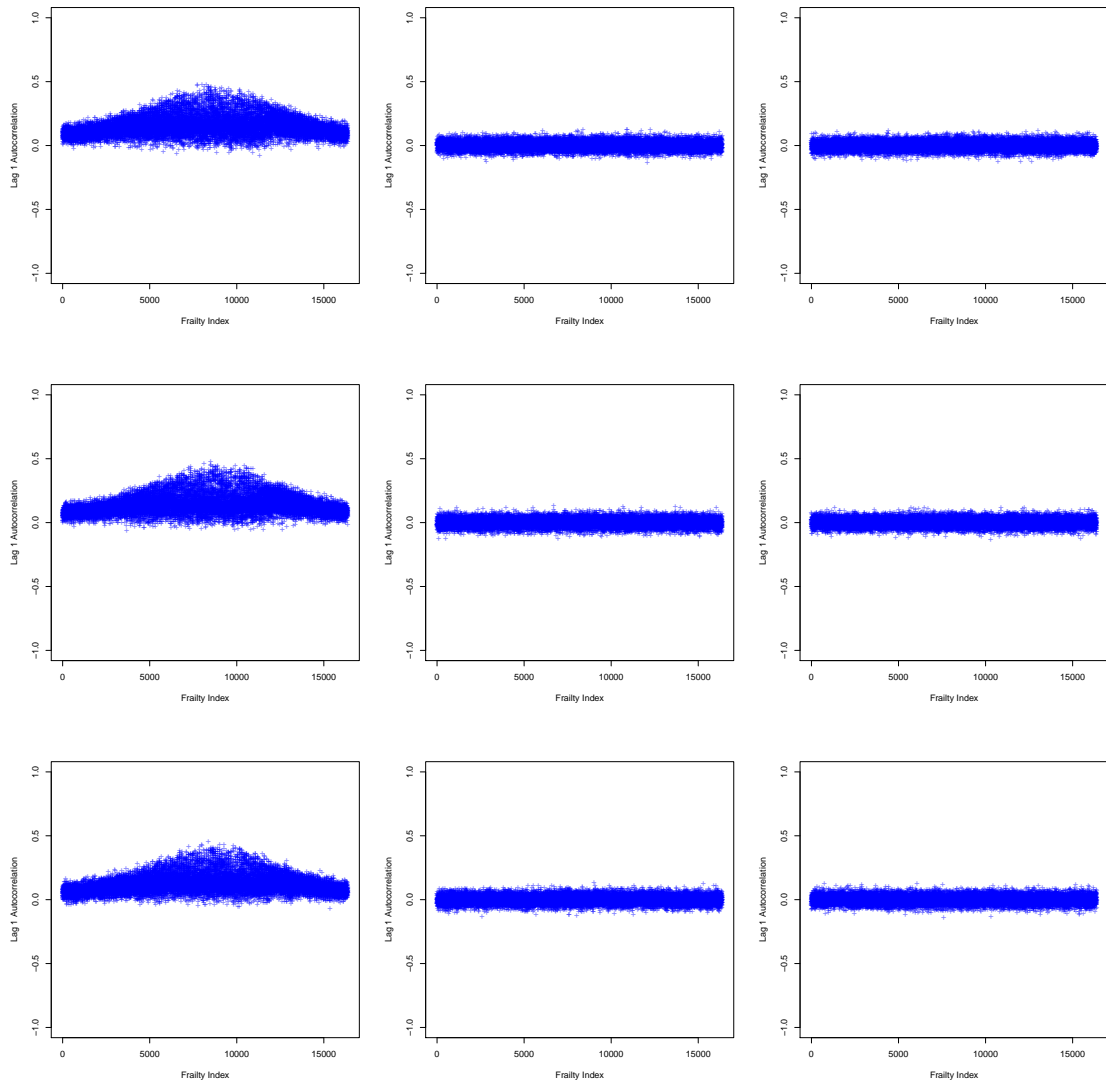
Figure 11: Autocorrelation in all sampled Y_is. Left column: lag1, middle column: lag 5, right column: lag 10. Top row: 2014, middle row: 2013, bottom row: 2012.
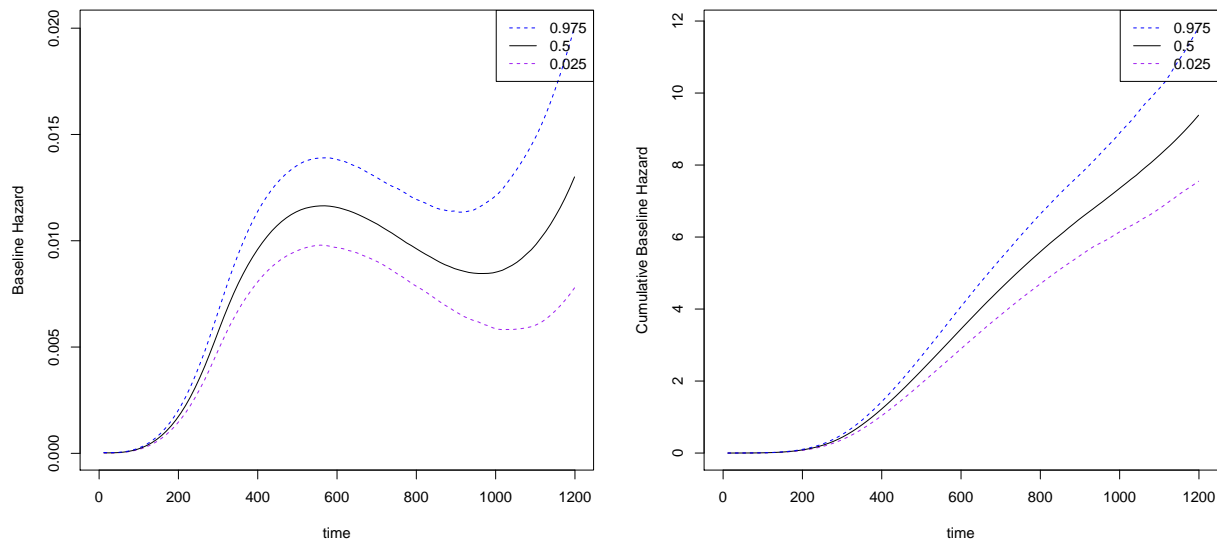
Figure 12: Plot showing the estimated B-spline baseline hazard function for the 2014 data (left), and baseline cumulative hazard (right).

References

Bannister, A. (2014). Fire service cuts felt as response times rise in capital. http://www.ifsecglobal.com/effects-fire-service-cuts-felt-average-response-time-rises/.

Caven, J. (2014). Fire brigade response times increase across barnet. http://www.times-series.co.uk/news/11601145.print/.

Claridge, E. and M. Spearpoint (2013). New zealand fire service response times to structure fires. *Procedia Engineering 62*(0), 1063 – 1072.

Cox, D. and D. Oakes (1984). *Analysis of Survival Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Department for Communities and Local Government (2014). Fire statistics: Great britain april 2013 to march 2014. https://www.gov.uk/government/statistics/fire-statistics-great-britain-2013-to-2014.

Foley, F. (2014). Increase in fire brigade response time. http://www.kcwtoday.co.uk/2014/11/19/increase-in-fire-brigade-response-time/.

Higgins, E., M. Taylor, H. Francis, M. Jones, and D. Appleton (2014). The evolution of geographical information systems for fire prevention support. *Fire Safety Journal 69*(0), 117 – 125.

Higgins, E., M. Taylor, M. Jones, and P. Lisboa (2013). Understanding community fire riska spatial model for targeting fire prevention activities. *Fire Safety Journal 62, Part A*(0), 20 – 29.

Johnson, A. (2014). Brigade response time increase 'a risk to lives' months after closure of clerkenwell fire station. `http://www.islingtontribune.com/news/2014/nov/brigade-response-time-increase-%E2%80%98-risk-lives%E2%80%99-%E2%80%93-months-after-closure-clerkenwell-fire-s`.

Klein, J., J. Ibrahim, T. Scheike, J. van Houwelingen, and H. Van Houwelingen (2013). *Handbook of Survival Analysis*. Chapman and Hall/CRC Handbooks of Modern Statistical Methods Series. Taylor & Francis Group.

Kolesar, P. and E. H. Blum (1973). Square root laws for fire engine response distances. *Management Science 19*(12), 1368–1378.

London Fire and Emergency Planning Authority (2015). London fire brigade incident records. `http://data.london.gov.uk/dataset/london-fire-brigade-incident-records`.

London Fire Brigade (2013a). Fifth london safety plan 2013–2016. `http://www.london-fire.gov.uk/lsp5.asp`.

London Fire Brigade (2013b). Incident response times. `http://www.london-fire.gov.uk/responsetimes.asp`.

London Fire Brigade (2013c). London fire brigade borough stats package. `http://www.london-fire.gov.uk/borough-stats-package/`.

London Fire Brigade (2013d). London fire brigade publishes savings proposals. `http://www.london-fire.gov.uk/news/C92C30A17C5547278C728EF4B9A8B530_LSPsavingsproposals-110113.asp#.VNiCwDX7uS8`.

London Fire Brigade (2014a). Fire facts: Incident response times 2005–2013.

London Fire Brigade (2014b). Lfb incident maps. `http://www.london-fire.gov.uk/LFBIncidentMaps.asp`.

Matteson, D. S., M. W. McLean, D. B. Woodard, and S. G. Henderson (2011, 06). Forecasting emergency medical service call arrival rates. *The Annals of Applied Statistics 5*(2B), 1379–1406.

McCartney, J. (2015). Fire brigade response times missing 6 minute target in enfield and haringey after boris johnsons fire brigade cuts. `http://joannemccartney.co.uk/2015/01/22/fire-brigade-response-times-missing-6-minute-target-in-enfield-and-haringey-after-bor`

Open Data Institute (2013). London fire stations. `http://london-fire.labs.theodi.org/`.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Read, C. (2014). Residents kept waiting longer than target response time for london fire brigade following woolwich and downham station closures. `http://www.thisislocallondon.co.uk/news/11624413.Greenwich_and_Lewisham_wards_see_London_Fire_Brigade_response_times_rise/?ref=mr`.

Rosenburg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics 51*(3), 874–887.

Scott, D. W., L. E. Factor, and G. A. Gorry (1978). Predicting the response time of an urban ambulance system. *Health Services Research 13*(4), 404–417.

Taylor, B. M. (2015). Auxiliary variable Markov chain Monte Carlo for spatial survival and geostatistical models. Under review, preprint available from `http://arxiv.org/abs/1501.01665`.

Taylor, B. M. and B. S. Rowlingson (2014). spatsurv: an R package for Bayesian inference with spatial survival models. To Appear in Journal of Statistical Software. Preprint available from `http://www.lancaster.ac.uk/staff/taylorb1/preprints/spatsurv.pdf`.

Toregas, C., R. Swain, C. ReVelle, and L. Bergman (1971). The location of emergency service facilities. *Operations Research 19*(6), 1363–1373.

Torney, K. (2013). How quickly did the fire service respond to emergencies where you live? `http://www.thedetail.tv/issues/187/fire-response-times/how-quickly-did-the-fire-service-respond-to-emergencies-where-you-live`.

Vile, J. L., J. W. Gillard, P. R. Harper, and V. A. Knight (2012). Predicting ambulance demand using singular spectrum analysis. *Journal of the Operations Research Society 63*, 1556–1565.

Westminster's Labour Councillors (2014). Westminster residents at risk as fire brigade response times go up by more than a minute. `https://labourwestminster.wordpress.com/2014/11/14/` `westminster-residents-at-risk-as-fire-brigade-response-times-go-up-by-more-than-a-min`

Younes, N. and J. Lachin (1997). Link-based models for survival data with interval and continuous time censoring. *Biometrics 53*(4), pp. 1199–1211.

Zhou, Z., D. S. Matteson, D. B. Woodard, S. G. Henderson, and A. C. Micheas (2014). A spatio-temporal point process model for ambulance demand. *Journal of the American Statistical Association 0*(ja), 00–00.