# Generalised additive point process models for natural hazard occurrence

## B. D. Youngman[a]* and T. Economou[a]

**Summary:** Point processes are a natural class of model for representing occurrences of various types of natural hazard event. Flexibly implementing such models is often hindered by intractable likelihood forms. Consequently, rates of point processes tend to be reduced to parametric forms, or the processes are discretised to give data of readily modelled 'count-per-unit' type. This work proposes generalised additive model forms for point process rates. The resulting low-rank spatio-temporal representations of rates, coupled with the Laplace approximation, makes the restricted likelihood relatively tractable, and hence inference for such models possible. The models can also be interpreted from a regression perspective. The proposed models are used to estimate different types of Cox process and then spatio-temporal variation in European windstorms. Through a combination of thin plate and cubic regression splines, and their tensor product, established relationships between where windstorms occur and the state of the North Atlantic Oscillation are confirmed, and then expanded to bring detailed understanding of within-year variation, which has otherwise not been possible with count-based models.

**Keywords:** Point process; Generalised additive model; Restricted maximum likelihood; Nonhomogeneous Poisson process; Natural hazard; European windstorms.

## 1. INTRODUCTION

Natural hazards have serious impacts on human society, such as causing disruption and loss of life. Understanding and quantifying the associated risk from natural hazards is therefore of

[a] *Department of Mathematics and Computer Science, University of Exeter, UK.*

* *Correspondence to: B.D. Youngman, Department of Mathematics and Computer Science, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QF, UK. E-mail: b.youngman@exeter.ac.uk*

particular importance. A main component of risk is of course the probability of occurrence (the other being exposure), which can be quantified using various data sources, such as simulations from physical models or noisy observations.

Various challenges can arise when analysing such data. For example, when modelling the rate at which hazard events occur over some domain, it is often necessary to allow for spatial and temporal variability. The events, however, are likely to be spatially and temporally irregular, as opposed, for example, to measured on a grid or at fixed locations. A common approach to remedy this (Møller and Waagepetersen, 2003) is to impose an arbitrary lattice structure and aggregate occurrence data. This achieves 'count-per-unit' data, which can be conventionally modelled using Poisson regression, for example. Beyond representing all parts of the study domain, the choice of the lattice structure is arbitrary, and results are conditional on its choice. Although the resolution of the lattice may be increased to reduce error, this increases computational demand, which can become prohibitive.

Point process models allow direct modelling of event occurrence data; see, for example, Cox and Isham (1980). Various parametric forms of rate variation exist for which point process models are relatively straightforward to implement; Ogata (1998) gives examples based on modelling earthquakes and aftershocks. However, such forms may lack flexibility. Implementing more flexible forms of variation is typically hindered by the corresponding likelihood of the point process model being analytically intractable. Similar motivation for Gaussian process models in geostatistics (Diggle and Ribeiro, 2007) makes log Gaussian Cox processes an attractive class of point process model; see Møller *et al.* (1998) for various properties. Simpson *et al.* (2016) have recently proposed an efficient approach to inference for log Gaussian Cox processes, which utilises the integrated nested Laplace approximation (Rue *et al.*, 2009) and stochastic partial differential equation models of Lindgren *et al.* (2011).

As opposed to an explicit stochastic process representation of rate variation, we adopt a generalised additive model approach (Hastie and Tibshirani, 1990). This allows variation in

the occurrence rate to be represented through covariates, each of which is represented by a finite set of basis functions and coefficients. Simpson *et al.* (2016) used a fixed (typically low) rank basis representation to represent log Gaussian Cox processes with Matérn covariance. We consider spline basis functions, which can also be used to provide low rank approximations to stochastic processes (Kammann and Wand, 2003), but also give various flexible choices for representing other forms of smooth and non-smooth variability. Such models can be estimated from count data once data are discretised using the method of Baddeley and Turner (2000) to produce data of standard generalised additive model form, for example, but this again relies on choosing an arbitrary lattice structure to achieve discretisation. We propose to base inference for the generalised additive point process model on the restricted likelihood (Wood, 2011; Wood *et al.*, 2016). This avoids the need to discretise occurrence data, and offers much greater tractability than cross-validation; see Wood (2011) for comparison.

This paper is structured as follows. The following section introduces the application of this paper and provides some further motivating examples. Section 3 presents the modelling framework. Section 4 presents a simulation study into estimation of Cox processes and then section 5 presents an analysis of European windstorm occurrence. Finally, a summary and discussion are given in section 6.

## 2. BACKGROUND

The present application involves understanding and quantifying European windstorm (extratropical cyclone) occurrence over the North Atlantic and Europe, although the proposed framework is applicable to other hazards, including tropical cyclones, hail, and tornadoes.

A particular question arising from the study of windstorm risk is whether temporal occurrence is completely random (in the Poisson sense) or whether storms tend to cluster, and if so why. Mailier *et al.* (2006) argued that windstorms do cluster over Europe, by showing

that windstorm counts are overdispersed with respect to a Poisson distribution. The spatial unit was defined by imposing an arbitrary grid, and the temporal unit by considering a season, so that aggregate counts in each spatio-temporal unit were used. Vitolo *et al.* (2009) extended that work by fitting Poisson regression models to illustrate that certain climate covariates potentially explain some of the temporal clustering. The inclusion of covariates reinforces the arbitrariness of choosing the spatio-temporal unit, as they then have to be aggregated as well.

Similar approaches were also used to model tropical cyclone (hurricane) occurrences. Elsner *et al.* (2012) discuss various ways of defining the spatial grid for the aggregation. Various studies of tornado occurrence both in the UK (Mulder and Schultz, 2015) and the USA (Elsner *et al.*, 2013) have also used aggregated counts over fairly arbitrary spatio-temporal units. Furthermore, Allen *et al.* (2015) impose a spatial grid to capture the relationship between hail occurrence and large-scale meteorological variables.

The imposition of arbitrary spatio-temporal grids on data available at much higher resolution seems to be common practice, presumably because modelling simplifies. However, aggregation can lead to loss of information especially if the spatial or temporal unit is not chosen to reflect the spatio-temporal variability of the data-generating process. Furthermore, comparison of results across studies becomes more difficult as the results are conditional on the choice of aggregation level. The unifying framework of point processes considered here, enables direct modelling of the rate of occurrence, making use of actual occurrence locations and times of events to construct the likelihood, thus avoiding the need to define aggregation units and then aggregate data.

## 3. METHOD

### 3.1. Non-homogeneous Poisson processes

We assume that occurrences of a natural hazard event can be modelled by a non-homogeneous Poisson process (NHPP; Cox and Isham, 1980). Whereas a Poisson process can be shown to represent complete spatio-temporal randomness by assuming the rate of occurrence is constant, a NHPP allows the rate to vary in space and time making it a very flexible model able to capture deviation from complete randomness, in terms of either clustering or regularity.

We consider a NHPP on a domain $\mathcal{V} = \mathcal{S} \times \mathcal{T}$, where $\mathcal{S} \subset \mathbb{R}^2$ and $\mathcal{T} \subset [0, \infty)$ represent spatial and temporal sub-domains, respectively. Let $\mathbf{v} = (\mathbf{s}, t)$ denote an arbitrary point in $\mathcal{V}$, where $\mathbf{s} = (x, y)$ denotes an arbitrary point in $\mathcal{S}$, defined in terms of its longitude and latitude, and $t$ represents an arbitrary point in $\mathcal{T}$. The rate of the process, quantifying the propensity of occurrence at any given point and time, will be denoted by $\lambda(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v}))$, where $\boldsymbol{\theta}$ represents unknown parameters and $\mathbf{z}(\mathbf{v}) = \big(z_1(\mathbf{v}), \dots, z_p(\mathbf{v})\big)$ represents $p$ covariates. Consider the sub-region $V = S \times T$, where $S \subseteq \mathcal{S}$ and $T \subseteq \mathcal{T}$. It follows that the number of hazard events in $V$ is Poisson distributed with mean

$$\Lambda(V; \boldsymbol{\theta}, \mathbf{z}(V)) = \int_V \lambda(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})) \mathrm{d}\mathbf{v}, \tag{1}$$

where $\mathbf{z}(V) = \cup_{\mathbf{v} \in V} \mathbf{z}(\mathbf{v})$.

Now consider a finite data set of hazard events, indexed by $i = 1, \dots, n$, occurring on $\mathcal{V}$. Denote data on events by $\{\mathbf{v}_i\}_{i=1,\dots,n}$, where for event $i$ $\mathbf{v}_i = (\mathbf{s}_i, t_i)$ collects its location, $\mathbf{s}_i = (x_i, y_i)$, defined by longitude $x_i$ and latitude $y_i$, and occurrence time $t_i$. Inference for the Poisson process model is then based on the likelihood

$$L(\boldsymbol{\theta}) = \exp\{|V| - \Lambda(V; \boldsymbol{\theta}, \mathbf{z}(V))\} \prod_{i=1}^n \lambda(\mathbf{v}_i; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v}_i)). \tag{2}$$

Parametric models for $\lambda(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v}))$ are conventionally chosen so that $\Lambda\big(V; \boldsymbol{\theta}, \mathbf{z}(V)\big)$ in equation (2) is tractable, which typically compromises flexibility. Here we propose a semi-parametric approach to modelling $\lambda\big(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})\big)$, and discuss solutions to computing equation (1) in §3.4.1.

### 3.2. Generalized additive point process models

This work focuses on generalized additive model (GAM) forms for $f\big(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})\big) = \log\big\{\lambda\big(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})\big)\big\}$. Such forms allow the rate (on the log scale) to be modelled as a potentially smooth and unknown function of space, time and/or any covariates. Subtracting larger multiples of the penalty

$$\int_V f''\big(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})\big)\mathrm{d}\mathbf{v} \tag{3}$$

from log likelihood (2) promotes smoother estimates of $f(\cdot)$. GAMs are particularly appealing since $f(\cdot)$ can be represented using regression splines, which essentially involve linear combinations of regression coefficients.

More specifically, we consider a basis representation for $f(\cdot)$ so that

$$f\big(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})\big) = \sum_{k=1}^{K}\sum_{d=1}^{D_k} \beta_{kd} b_{kd}\big(\mathbf{z}(\mathbf{v})\big), \tag{4}$$

for basis functions $b_{kd}(\cdot)$ (to be specified in the subsequent section). This formulation allows $f(\cdot)$ to be represented by a combination of different bases, $k = 1, \ldots, K$, each of which have a finite $D_k$-dimensional representation. The upshot of this representation is that we maximise a penalised log-likelihood of the form

$$\ell_p(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) - \frac{1}{2}\sum_{k=1}^{K} \rho_k \boldsymbol{\beta}_k^T \mathbf{S}_k \boldsymbol{\beta}_k, \tag{5}$$

where for $k = 1, \ldots, K$, $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kD_k})^T$ are basis coefficients, smoothing parameter $\rho_k > 0$ control the smoothness of $\sum_{d=1}^{D_k} \beta_{kd} b_{kd}(\mathbf{z}(\mathbf{v}))$, and $\mathbf{S}_k$ is a penalty matrix determined by the basis functions $b_{k1}(\cdot), \ldots, b_{kD_k}(\cdot)$.

### 3.3. Suggested basis representations

For most types of natural hazard, $\lambda(\cdot)$ should capture spatial, temporal and also spatio-temporal variation. As opposed to two one-dimensional bases with potential interactions, we suggest representing spatial variation using two-dimensional bases; for example, $b_{kd}(z_k(\mathbf{v})) = b_{kd}(x, y)$. The thin plate regression spline is a robust two-dimensional basis; see Duchon (1977), Wood (2003) and Wood (2006, Section 4.1.5) for further details. Thinking of spatial variability as a stochastic process with spatial correlation, this basis imposes the same amount of correlation in all directions. If such isotropy is inappropriate, then Duchon splines (Duchon, 1977) are a natural anisotropic extension, and are readily incorporated in the proposed framework.

For other covariates, such as time itself, one-dimensional bases are suggested. In general we choose cubic regression splines, or special cases thereof; for example, cyclic cubic regression splines. For two covariates, represented by bases $b_{k.}(\cdot)$ and $b_{k'.}(\cdot)$, the following additive form may be assumed:

$$f_{kk'}\big(z_k(\mathbf{v}), z_k(\mathbf{v}')\big) = \sum_{d=1}^{D_k} \beta_d b_{k,d}(z_k(\mathbf{v})) + \sum_{d'=1}^{D_{k'}} \beta_{d'} b_{k',d'}(z_{k'}(\mathbf{v})). \tag{6}$$

We should also consider that some covariates may interact. This can be captured through a tensor product representation where

$$f_{kk'}(z_k(\mathbf{v}), z_{k'}(\mathbf{v})) = \sum_{d=1}^{D_k} \sum_{d'=1}^{D_{kk'}} \beta_{kk'} b_{k,d}(z_k(\mathbf{v})) b_{k',d'}(z_{k'}(\mathbf{v})), \tag{7}$$

which can be extended to higher-order interactions.

We reiterate at this point that these bases are merely suggestions, albeit ones that are intended to be robust yet flexible, and thus to suit various types of hazard event. The model of §3.2 readily accepts other representations for $\lambda(\cdot)$.

## 3.4. Model fitting

Prior to fitting the model, i.e. estimating the basis coefficients $\boldsymbol{\beta}$, there are two difficulties that must be addressed. The first is evaluating the integral of equation (1), which occurs in likelihood (2), and the second is choosing optimal values for the smoothing parameters $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_K)'$.

### 3.4.1. Integral approximation

To evaluate $\Lambda\big(V; \boldsymbol{\theta}, \mathbf{z}(V)\big)$ in equation (1), with $V$ as defined in §3.1, we use quadrature, as in Simpson *et al.* (2016). Noting that the choice of quadrature scheme is effectively arbitrary, the simplest scheme is to partition $V$ as a lattice and represent each of its $N$ parallelotopes as nodes $\{\tilde{\mathbf{v}}_i\}_{i=1,\ldots,N}$ with corresponding quadrature weights $\{w_i\}_{i=1,\ldots,N}$. We then assume that

$$\Lambda\big(V; \boldsymbol{\theta}, \mathbf{z}(V)\big) \simeq \sum_{i=1}^{N} w_i \lambda\big(\tilde{\mathbf{v}}_i; \boldsymbol{\theta}, \mathbf{z}(\tilde{\mathbf{v}}_i)\big).$$

Larger $N$ increases the computational cost of this approximation. In general, though, a fairly large range for $N$ exists such that the approximation is sufficiently accurate. This range decreases, however, if smoothing parameters are small, i.e. surfaces are rough.

Note that, if $V$ comprises a spatial component, and the corresponding slice of its lattice representation is a regular grid, then the weights will be equal. However, as we propose allowing covariates in $\lambda(\cdot)$, although each dimension of the lattice may be represented by equal-width bins, weights may not be equal. An example is given in §5, where we allow $\lambda(\cdot)$

to vary with the North Atlantic Oscillation index, which is not uniformly distributed over the time period studied. The two simplest solutions to this are to define bins by equally-separated quantiles, which will give equal weights, or to approximate the distribution of the covariate (e.g. fixing bins and using a histogram), which may give non-equal weights.

### 3.4.2. Smoothing parameter estimation

We follow the notation and method of Wood (2011) and recognise that the penalty term on the right-hand side of equation (5), $\frac{1}{2} \sum_{k=1}^{K} \rho_k \boldsymbol{\beta}_k^T \mathbf{S}_k \boldsymbol{\beta}_k$, can be re-written as $\frac{1}{2} \boldsymbol{\beta}' \mathbf{S} \boldsymbol{\beta}$, where $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \dots, \boldsymbol{\beta}_K')$ and $\mathbf{S} = \sum_{k=1}^{K} \rho_k \mathbf{S}_k^*$, where $\mathbf{S}_k^*(D_{k-1}^\dagger + i, D_{k-1}^\dagger + j) = \mathbf{S}_k(i, j)$ for $i, j = 1, \dots, D_k$ and $D_{k-1}^\dagger = \sum_{k=1}^{k-1} D_k$, and $\mathbf{S}_k^*(i, j) = 0$ otherwise. This formulation has the interpretation that $\ell_p$ is proportional to the log posterior of $\boldsymbol{\beta}$ if $\boldsymbol{\beta}$ has a zero-mean multivariate Gaussian prior with covariance matrix $\mathbf{S}^-$, where $\mathbf{S}^-$ is the generalised inverse of $\mathbf{S}$. It follows that the approximate sampling distribution of $\hat{\boldsymbol{\beta}}$ is multivariate Gaussian with mean $\boldsymbol{\beta}$ and covariance matrix $\hat{\mathbf{V}}_{\boldsymbol{\beta}} = \mathbf{H}^{-1}$, where $\mathbf{H} = -\nabla^2 \log \ell_p(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$.

As proposed by Wood (2011), the Laplace approximation can be used to integrate out $\boldsymbol{\beta}$ from $\ell_p$, which results in the restricted log likelihood taking the form

$$\ell_R(\boldsymbol{\rho}) = \ell_p(\hat{\boldsymbol{\beta}}) + \frac{1}{2} \log |\mathbf{S}^-| - \frac{1}{2} \log |\mathbf{H}| + \text{constant}. \tag{8}$$

Maximising the restricted log likelihood of equation (8) w.r.t. $\boldsymbol{\rho}$ involves inner and outer iterations, which are described in detail in Wood *et al.* (2016, Section 3). Inner iterations give $\hat{\boldsymbol{\beta}}$ given $\boldsymbol{\rho}$ by maximising $\ell_p(\boldsymbol{\beta})$ w.r.t. $\boldsymbol{\beta}$ given $\boldsymbol{\rho}$. This will almost inevitably require numerical methods. Gradient-based methods, such as quasi-Newton algorithms, are found to perform robustly for our simulation study of §4 and application of §5. Significant speed increases are had by supplying analytical gradients instead of using numerical estimates based on finite differences. Outer iterations give $\hat{\boldsymbol{\rho}}$ by maximising $\ell_R(\boldsymbol{\rho})$ w.r.t. $\boldsymbol{\rho}$, each iteration of which requires obtaining $\hat{\boldsymbol{\beta}}$ given $\boldsymbol{\rho}$. Numerical methods are again likely to be required to maximise

$\ell_R(\boldsymbol{\rho})$. However, efficiency gains from supplying the analytical gradient tend to be much less when using gradient-based methods as $\boldsymbol{\rho}$ usually comprises many fewer elements than $\boldsymbol{\beta}$. Convergence is relatively simple to monitor when using gradient-based methods by requiring that elements of $\nabla \log \ell_p(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$ and $\nabla \log \ell_R(\boldsymbol{\rho})|_{\boldsymbol{\rho}=\hat{\boldsymbol{\rho}}}$ are close to zero.

### 3.5. Model checking

From a statistical perspective, there are various aspects of the fit of the model described in §3.1–3.4 that can be checked. When modelling natural hazard events, and in the light of the suggested basis representations of §3.3, we focus on assessing whether events' spatial variability, variation with covariates, and spatial variation with covariates are all well represented. Each of these can be checked in a general and yet intuitive way. As in §3.4, we suppose that the domain $\mathcal{V}$ can be partitioned into parallelotopes and, for simplicity, let $V^* \subset \mathcal{V}$ denote an arbitrary parallelotope, and define $\mu = \Lambda\big(V^*; \boldsymbol{\theta}, \mathbf{z}(V^*)\big)$ and $Y = \sum_{i=1}^n I(\mathbf{v}_i \in V^*)$. Deviance residuals are then given by

$$R = \text{sign}(Y - \mu)\sqrt{Y \log(Y/\mu) - (Y - \mu)}. \tag{9}$$

These can be used to identity areas of space, values or ranges of covariates, or combinations of the two, for which the model fit is poor. Multiple parallelotopes can also be simultaneously considered and compared to the Chi-squared distribution with suitable degrees of freedom. High- and low-resolution partitions of $\mathcal{V}$ may also be used to avoid sensitivity to paralleotope choice. These are chosen to be robust tests of model fit. If results are inconclusive, more thorough tests, such as those described in detail in and Baddeley *et al.* (2005), are recommended.

## 4. SIMULATION STUDY

This section presents a two-dimensional simulation study into the effect of basis dimension and numerical integration grid resolution on rate estimate accuracy, and compares the proposed point process model to kernel estimation. The study considers two domains: $\mathcal{S}_{\mathrm{sim}} = [-1, 11] \times [-1, 11]$ and $\mathcal{S}_{\mathrm{est}} = [0, 10] \times [0, 10]$. The former is used for point simulation and model estimation, whereas the latter is used for model comparison. A reduced domain is used for comparison to avoid adverse boundary effects, which may occur with the point process model or kernel estimation. Rates of the form $\lambda(\mathbf{s}) = g(Z(\mathbf{s}))$ are considered, where $g$ is a link function and $Z(\mathbf{s}) \sim GP(0, \sigma^2 c(\mathbf{s}, \cdot))$ is a Gaussian process with $c(\mathbf{s}, \mathbf{s}') = \exp\{-(||\mathbf{s} - \mathbf{s}'||^2)/\phi\}$, $\sigma^2 = 0.5$ and $\phi = 0.3$. This rate form and covariance function combination is intended to mimic a plausible natural hazard occurrence rate that is reasonably smooth. Three forms for $g$ are considered: $g_1(x) = \exp(x)$, $g_2(x) = \Phi(x)$, where $\Phi(x)$ is the standard Gaussian distribution function, and $g_3(x) = -\log(1 - \Phi(x))$. These give rise to Cox processes in which rates have log-Gaussian, uniform and exponential distributions, respectively. Cox processes are simulated for a regular $500 \times 500$ grid on $\mathcal{S}_{\mathrm{sim}}$. Realisations from each of these processes are shown in Figure 1, each represented by simulated points superimposed on the underlying rate that generated the points.

[Figure 1 about here.]

The point process model is specified as $\log \lambda(\mathbf{s}) = f_{\mathrm{TPRS},D}(\mathbf{s})$, where $f_{\mathrm{TPRS},D}$ denotes a thin-plate regression spline with basis rank $D$. Here $D = 20, 40$ and $80$ are used. Knots are automatically chosen according to a truncated eigendecomposition of points simulated on $\mathcal{S}_{\mathrm{sim}}$, as detailed in Wood (2003). Kernel estimation uses the `kde2d` function from the `R` package `MASS` (R Core Team, 2016; Venables and Ripley, 2002), which forms a two-dimensional kernel from the product of two one-dimensional Gaussian kernels. Here the bandwidth, $h$, is set at $h = 2, 4$ and $8$. Let $N$ denote both the number of integration nodes

for the point process model and knots for kernel estimation. Nodes and knots are chosen as a regular $\sqrt{N} \times \sqrt{N}$ grid on $\mathcal{S}_{\text{sim}}$, where $\sqrt{N}$ ranges from 15 to 45 in steps of 5. Each of these models is fitted to each of the three Cox processes described above, and 1,000 simulations performed for each. Examples of point process model and kernel estimates can be seen in Figure 1 for a single simulation. These estimates use $D = 80$, $h = 4$, and $\sqrt{N} = 30$. Models are assessed by comparing the mean squared error (MSE) between the generated Cox process rate, $\lambda(\mathbf{s})$, and its estimate, $\hat{\lambda}(\mathbf{s})$, say, which is evaluated at each point of $\mathcal{S}_{\text{est}}$ (which is a $416 \times 416$ grid comprising the points of $\mathcal{S}_{\text{sim}}$ lying within $\mathcal{S}_{\text{est}}$). MSE values for different point process model, kernel estimation and Cox process specifications are shown in Figure 2.

[Figure 2 about here.]

Based on Figure 2 the number of integration nodes used for the point process model has little effect on MSE values. The effect of basis rank, $D$, is that MSE is lowest for $D = 80$ and highest for $D = 20$, although this difference is relatively small for Cox processes with uniformly distributed rates. For kernel estimation, the relative performance of the bandwidth, $h$, in terms of reduced MSE, depends on both type of Cox process and number of knots. For rates with a log-Gaussian distribution, $h = 8$ gives highest MSE. Comparable MSE values are seen for $h = 2$ and 4 with between 15 and 25 knots; however $h = 4$ notably outperforms $h = 2$ with 30 or more knots. If rates are uniformly distributed $h = 2$ is outperformed by $h = 4$ and 8, which perform similarly; the number of knots has little effect. Results for exponentially distributed rates are similar to those for the log-Gaussian case, apart from $h = 4$ consistently outperforms $h = 8$ and is less sensitive to the number of knots.

In comparison to kernel estimation, the point process model is seen to be less sensitive to $N$, and to outperform kernel estimation for the simulated Cox processes with log-Gaussian rates and to a lesser extent Cox processes with exponential rates for the model specifications considered. Point process models and kernel estimates with higher basis ranks and bandwidths, respectively, perform similarly for Cox processes with uniform rates,

irrespective of number of integration nodes or knots; best performance is found for the kernel estimate with $h = 8$. Overall point process models with $D = 80$ can be seen to be at least as good as the best-performing kernel estimates, except for the case of uniform rates when estimates are slightly inferior. Rates for Cox processes range from heavy- to light-tailed. Altering kernels or link functions relating $f(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v}))$ and $\lambda(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v}))$ may help improve performance across a range of tail behaviour. These results also only apply to Cox processes based on a single Gaussian process specification.

## 5. APPLICATION TO EUROPEAN WINDSTORMS

Intense extra-tropical cyclones or windstorms are a major source of risk in Europe, and can cause severe damage, disruption and insurance/human losses due to extreme wind speeds and flooding. An interesting problem, both from a meteorological and societal point of view, is understanding where the storm peaks occur and how this might vary with time—both in terms of seasonal variation and background large-scale climatic conditions.

[Figure 3 about here.]

### 5.1. Data

Storm tracks are extracted from the ERA-Interim reanalysis (Dee *et al.*, 2011) using the tracking algorithm of Hodges (1995, 1999). These reanalysis data may be considered a historical realisation of weather produced by a physical coupled model of the atmosphere and ocean. Actual observations are assimilated in the process to provide a data constrained best estimate of historical weather. Hodges' (1995, 1999) tracking algorithm extracts storm tracks by locating maxima in the vorticity field from the ERA-Interim reanalysis. The data span 1979–2014 and the study domain is most of Europe and the North Atlantic. Windstorms typically form over the North Atlantic and track east. The spatial domain was chosen to

reflect the main area of windstorm activity. Figure 3 shows the spatial domain and windstorm tracks for the year 2014, with locations at which storm peaks occur indicated. Each track is defined as the location of the centre of the windstorm in three-hourly time steps.

[Figure 4 about here.]

[Figure 5 about here.]

## 5.2. Windstorm peaks

The location of the peak of a windstorm is defined here as the point where the lowest mean sea level pressure (MSLP) occurs along the track, which results in 3133 storm peaks. MSLP is a variable quantifying windstorm intensity, and has been shown to be highly correlated with wind speed (Economou *et al.*, 2014), both at the extremes and non-extremes. Figure 4 shows a gridded summary (counts, on annual scale) of where storm peaks occurred during 1979–2014. It is clear that storm peaks do not occur uniformly over the study domain: most tend to occur just south-east of Iceland, west of the United Kingdom. Storm peaks seldom occur over land, especially in the western part of the study domain. Figure 5 also shows a summary of storm counts per month, indicating that storm frequency is at its highest mid-winter and lowest mid-summer, ranging by a factor of approximately 15 times.

We also consider the North Atlantic Oscillation (NAO) as a possible covariate affecting the occurrence of storm peaks. The NAO is defined as the difference between sea-level pressure in Iceland and the Azores, and is standardised to have zero mean and unit variance. It is a large-scale climate index and has been shown to influence the development of extra-tropical cyclones (Pinto *et al.*, 2009). Observed values of the NAO index matching the reanalysis tracks were taken from the National Oceanic and Atmospheric Administration (NOAA) webpage (http://www.noaa.gov; accessed 2 November 2016). A summary of storm peak occurrence with the NAO is shown in Figure 6, which suggests a tendency for storm peaks to occur further south if the NAO is below -1. The distribution of NAO is not uniform over

its approximate range of [-3, 3]: instead it qualitatively resembles a Gaussian distribution. If only NAO values coinciding with the times at which storm peaks occur are considered, a slightly positively skewed distribution results in which approximately 50% of NAO values lie between -0.2 and 0.9.

[Figure 6 about here.]

## 5.3. Model specification

Let $s \in \mathcal{S}$ represent a storm peak location, $t$ represent its occurrence time in years and $nao(t)$ represent the value of NAO at time $t$, so that $\mathbf{z}(\mathbf{v}) = (\mathbf{s}, t, nao(t))$. We model the rate of storm occurrence, $\lambda(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v}))$, i.e. the propensity of storm peak occurrence at point $s$, time $t$ and NAO value $nao(t)$. We assume that

$$\log\left\{\lambda(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v}))\right\} = f_{\text{tensor}}\left(f_{\text{TPRS},49}(\mathbf{s}), f_{\text{CR},5}(nao(t))\right) + f_{\text{CC},10}(t - \lfloor t \rfloor), \tag{10}$$

where $f_{\text{tensor}}$ represents a tensor product, $f_{\cdot,D}$ is a spline of basis dimension $D$, and $f_{\text{TPRS}}$, $f_{\text{CR}}$ and $f_{\text{CC}}$ denote thin plate, cubic and cyclic cubic regression splines, respectively. In equation (10), $f_{\text{TPRS},49}(\mathbf{s})$ and $f_{\text{CR},7}(nao(t))$ allow smooth variation over space and with NAO, while their tensor product lets the spatial distribution of storm peaks smoothly vary with NAO. Function $f_{\text{CC},10}(t - \lfloor t \rfloor)$ captures within year variability; its cyclic form ensures continuity from year to year. Knots for the thin plate regression spline are defined by an eigendecomposition of the 3133 data points, as in §4 and Wood (2003). The number of knots for the cubic and cyclic cubic regression splines matches their corresponding basis dimensions and their positions are defined by equally-spaced quantiles on $[0, 1]$ of the variable that they represent. The model is formulated so that $\lambda(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v}))$ represents the *annual* occurrence rate for given $\mathbf{z}(\mathbf{v})$.

The model is fitted following the method described in §3.4.2. Smoothing parameters are estimated by maximising the restricted log likelihood of equation (8) using a quasi-Newton algorithm, each iteration of which requires that equation (5) is maximised to find $\hat{\boldsymbol{\beta}}$ given $\boldsymbol{\rho}$. Analytical gradients are supplied throughout. These estimates are based on the assumption that storm peak occurrences are independent from one storm to another, conditional on NAO, which is consistent with physical understanding of windstorm occurrence (Vitolo *et al.*, 2009). The total number of integration nodes is $N = 15^4$: each variable (longitude, latitude, NAO and within-year time) has 15 equally-spaced values, which form a $15 \times 15 \times 15 \times 15$ regular lattice of integration nodes. Uncertainty estimates for $\lambda\big(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})\big)$ are obtained by drawing samples of $\boldsymbol{\beta}$ from a multivariate Gaussian distribution with mean $\hat{\boldsymbol{\beta}}$ and covariance matrix $\mathbf{V}_{\boldsymbol{\beta}}$ (see §3.4.2), and using these to calculate $\lambda\big(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})\big)$ (for example, by plugging samples in to equation (4)). Confidence intervals are based on quantiles of the sampled values of $\lambda\big(\mathbf{v}; \boldsymbol{\theta}, \mathbf{z}(\mathbf{v})\big)$.

## 5.4. Results

[Figure 7 about here.]

Figure 7 shows estimated storm rates aggregated over space. The spatial estimate consistently matches that empirically estimated (shown in Figure 4). The model-based estimate, however, makes clearer that although most storms reach their peak south-west of Iceland, a secondary group reach theirs in a channel south-east of Iceland and north-west of the United Kingdom. The accompanying uncertainty estimate of Figure 7 broadly shows that rate uncertainty is greatest where model-based estimates are highest. Model-based estimates of storm rates aggregated over NAO and (within-year) time, together with uncertainty estimates, are shown in Figure 7. These consistently match the empirical estimates of Figure 5, once the non-uniform distribution of NAO has been taken into account.

[Figure 8 about here.]

We use the aggregated model-based and empirical counts to assess the model's fit, using the method outlined in §3.5. Specifically, we compare counts aggregated over a spatial grid and bins for NAO and time within year. Resulting deviance residuals are shown in Figure 8. In general the deviance residuals, defined in equation (9), are evenly distributed about zero, and not overly large in absolute value. For the spatial part, there are no striking signs of systematic areas where the model is over- or under-estimating rates, which indicates that, although smooth, the thin plate spline is sufficiently flexible at capturing storm rate variability over the domain. Similar support for the model is found by analysing the model-based counts aggregated over NAO bins and months. We also analyse residuals collectively through $\chi^2$ tests based on the sums of squared deviance residuals. Studying spatial, NAO and within-year parts as a whole yields $\chi^2$ test statistics of 99.21, 4.54 and 19.99, none of which clearly indicate model inadequacy when compared to the $\chi^2_{100}$, $\chi^2_{20}$ and $\chi^2_{12}$ distributions, respectively.

[Figure 9 about here.]

Finally, we show how the locations of storm peaks vary with NAO. Figure 9 shows model-based estimates of the rate for different NAO values between -3 and 3. The plots show that for each value of NAO, storm peaks tend to occur out in the Atlantic. However, if we consider the NAO ranging from -3 to 3, storm peaks can be seen to occur south-west of the United Kingdom and move north until over Iceland. This is consistent with the findings of Mailier *et al.* (2006) and Vitolo *et al.* (2009), although they studied locations of *tracks*, whereas our analysis is of *peaks*. Note though that relatively few storm peaks occur when the NAO index is outside $[-1, 1]$, giving relatively large uncertainty estimates.

## 6. SUMMARY

We have presented and illustrated a point process framework for modelling spatio-temporal occurrence of natural hazards. The framework relies on generalised additive model forms for flexibly modelling the intensity or rate of hazard event occurrence. This formulation allows rates to be represented as potentially smooth—yet unknown—functions of space and time, and other covariates, thus avoiding making assumptions about regularity in the occurrence of events in any dimension; for example, assuming linearity. We capture spatio-temporal variability by combining thin plate and cubic regression splines though a tensor product form. Irrespective of spline choice, model fitting based on restricted maximum likelihood is efficient, even for large data set, such as the one used in §5.

A simulation study of different types of spatial Cox process was given in §4. This compared various specifications of point process model to kernel estimation with different bandwidths and numbers of knots. The study showed that, irrespective of the number of integration nodes, and provided sufficiently high basis dimensions are used, the point process model offers similar or better accuracy than the best performing kernel model. The latter, however, required careful choice of bandwidth and, for some Cox processes, number of knots. We therefore think that the speed, off-the-shelf nature, ability to quantify uncertainty and reliability make generalised additive point process models fitted by restricted maximum likelihood useful for understanding and quantifying natural hazard occurrence.

The model was then applied to windstorm peak occurrence over Europe and the North Atlantic. Results indicated almost all peaks occur over the Atlantic ocean, although this varies significantly with the NAO: during negative phases peaks tend to have similar latitudes to central Europe and the UK, whereas during positive phases most peaks occur around Iceland.

From an impact perspective, an analysis of whole storm tracks rather than peaks might have been more relevant, as damage does not necessarily coincide with a peak's location;

instead it may be anticipated anywhere along a storm's track. The proposed model can be fitted to track data using exactly the same method as used in §5.4, apart from uncertainty estimation, as it would be unreasonable to assume that locations of points from the same storm occur independently over the domain, given the time of year or NAO state. This is because, given a point on a storm track, the next point to occur in time is likely to be relatively close to the previous point. This can be resolved by applying the sandwich correction to the Hessian matrix (White, 1980), to allow for this dependence.

## SUPPLEMENTARY MATERIAL

The file `ppgam.tar.gz` contains `R` functions to fit point process models of the form described in §3.2. The function `ppgam` is inspired by—and uses—the `gam` function from the `mgcv` package, and allows generalised additive point process models to be fitted using a single function; e.g. we use

```
> ppgam(te(lon, lat, nao, d=c(2, 1), k=c(49, 7), bs=c('tp', 'cr'))
        + s(cyc, bs='cc'), data, quadlist, ...)
```

where `data` comprises a data frame of points and corresponding covariates and `quadlist` can be used to provide representative values of non-uniformly distributed covariates, such as NAO, used in §5.3. The file `ppgam.tar.gz` also gives the data used here and `R` code to recreate the model fitted and plots presented in the present paper.

## ACKNOWLEDGEMENTS

## REFERENCES

Allen JT, Tippett MK, Sobel AH, 2015. An empirical model relating U.S. monthly hail occurrence to large-scale meteorological environment. *Journal of Advances in Modeling Earth Systems* **7**(1): 226–243.

Baddeley A, Turner R, 2000. Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics* **42**(3): 283–322.

Baddeley A, Turner R, Møller J, Hazelton M, 2005. Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(5): 617–666.

Cox DR, Isham V, 1980. *Point processes.* Chapman and Hall.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F, 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**(656): 553–597.

Diggle P, Ribeiro P, 2007. *Model-based Geostatistics.* Springer Series in Statistics, Springer New York.

Duchon J, 1977. *Splines minimizing rotation-invariant semi-norms in Sobolev spaces.* Springer Berlin Heidelberg, Berlin, Heidelberg, 85–100.

Economou T, Stephenson DB, Ferro CAT, 2014. Spatio-temporal modelling of extreme storms. *Annals of Applied Statistics* **8**: 2223–2246.

Elsner JB, Hodges RE, Jagger TH, 2012. Spatial grids for hurricane climate research. *Climate Dynamics* **39**(1): 21–36.

Elsner JB, Michaels LE, Scheitlin KN, Elsner IJ, 2013. The Decreasing Population Bias in Tornado Reports across the Central Plains. *Weather, Climate, and Society* **5**(3): 221–232.

Hastie TJ, Tibshirani RJ, 1990. *Generalized additive models*, volume 43. CRC Press.

Hodges K, 1995. Feature tracking on the unit-sphere. *Monthly Weather Review* **123**(12): 3458–3465.

Hodges K, 1999. Adaptive constraints for feature tracking. *Monthly Weather Review* **127**(6): 1362–1373.

Kammann EE, Wand MP, 2003. Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **52**(1): 1–18.

Lindgren F, Rue H, Lindström J, 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series*

*B (Statistical Methodology)* **73**(4): 423–498.

Mailier PJ, Stephenson DB, Ferro CA, Hodges KI, 2006. Serial clustering of extratropical cyclones. *Monthly weather review* **134**(8): 2224–2240.

Møller J, Syversveen AR, Waagepetersen RP, 1998. Log gaussian cox processes. *Scandinavian Journal of Statistics* **25**(3): 451–482.

Møller J, Waagepetersen RP, 2003. *Statistical inference and simulation for spatial point processes*. CRC Press.

Mulder KJ, Schultz DM, 2015. Climatology, Storm Morphologies, and Environments of Tornadoes in the British Isles: 1980–2012. *Monthly Weather Review* **143**(6): 2224–2240.

Ogata Y, 1998. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics* **50**(2): 379–402.

Pinto JG, Zacharias S, Fink AH, Leckebusch GC, Ulbrich U, 2009. Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO. *Climate Dynamics* **32**(5): 711–737.

R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rue H, Martino S, Chopin N, 2009. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)* **71**(2): 319–392.

Simpson D, Illian JB, Lindgren F, Srbye SH, Rue H, 2016. Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Biometrika* .

Venables WN, Ripley BD, 2002. *Modern Applied Statistics with S*. Springer, New York, fourth edition, iSBN 0-387-95457-0.

Vitolo R, Stephenson DB, Cook IM, Mitchell-Wallace K, 2009. Serial clustering of intense European storms. *Meteorologische Zeitschrift* **18**(4): 411–424.

White H, 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**(4): 817–838.

Wood SN, 2003. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(1): 95–114.

Wood SN, 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Wood SN, 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(1): 3–36.

Wood SN, Pya N, Säfken B, 2016. Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* **111**(516): 1548–1563.
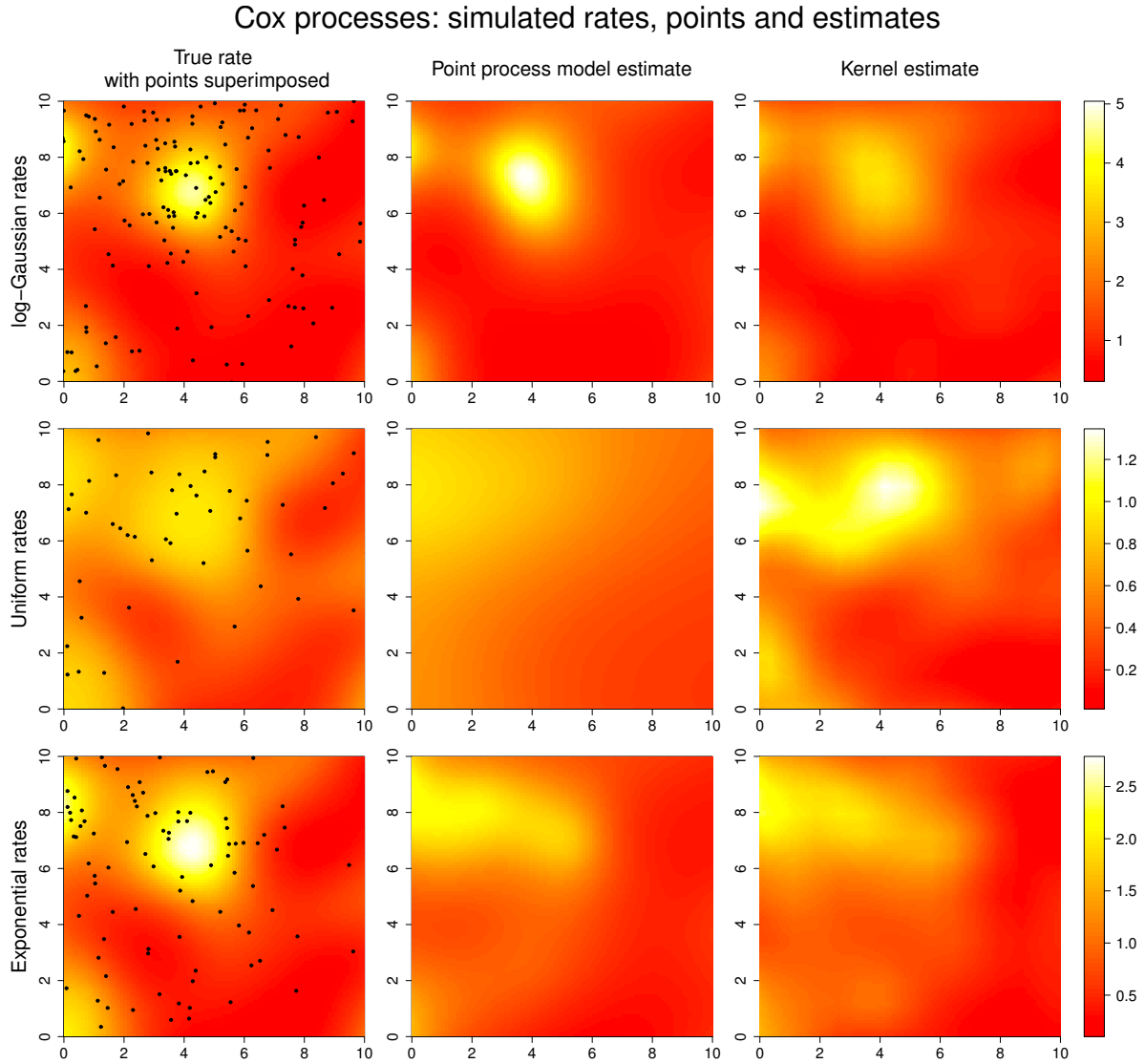
**FIGURES**

**Figure 1.** Cox processes with log-Gaussian (row 1), uniform (row 2) and exponential (row 3) rates. Each process is represented by a realisation of points superimposed on the underlying process rate (column 1), and corresponding point process model (column 2) and kernel (column 3) estimates. Note that each Cox process is based on the same Gaussian process realisation. Point process models used thin plate spines basis rank 64 and a $30 \times 30$ grid of integration nodes. Kernel estimates use a bandwidth of 4 and a $30 \times 30$ grid of knots.

**Point process rate mean squared error**
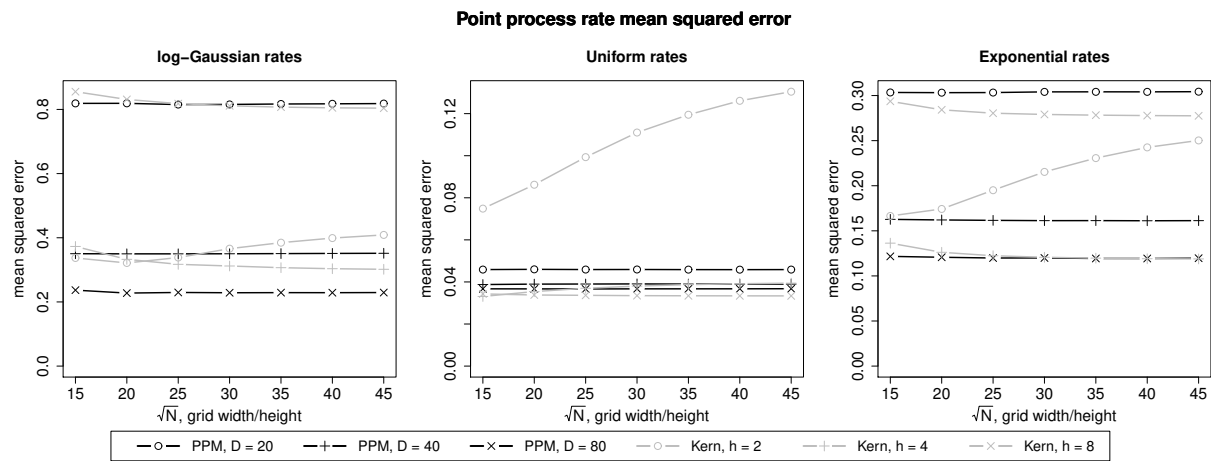


**Figure 2.** Mean squared error estimates comparing point process model (PPM) and kernel (Kern) estimates.
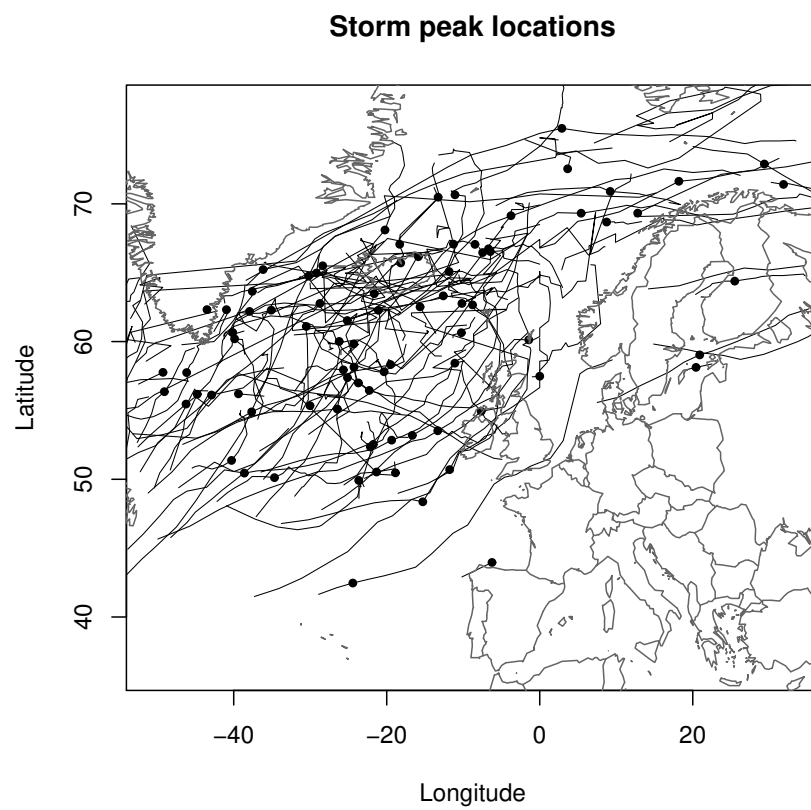
**Figure 3.** Storm tracks for the year 2014 extracted from the ERA-Interim reanalysis with locations of storm peaks indicated.
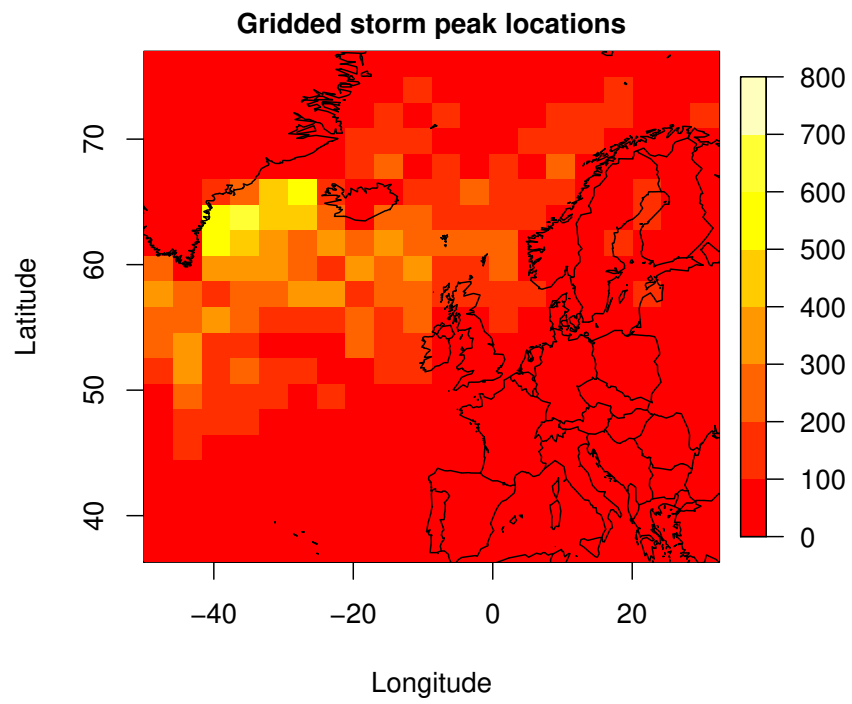
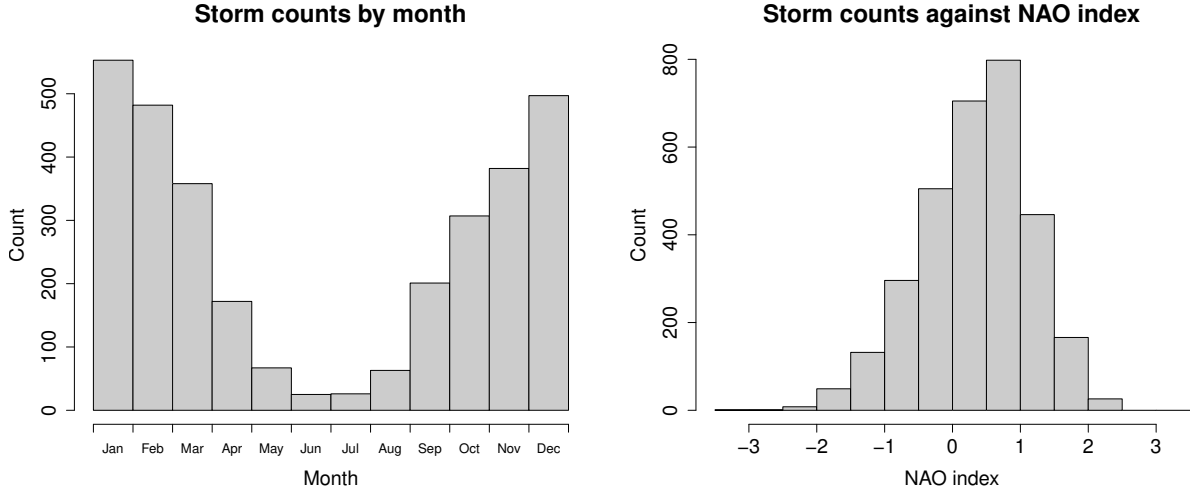**Figure 4.** Gridded empirical storm peak counts on annual scale.

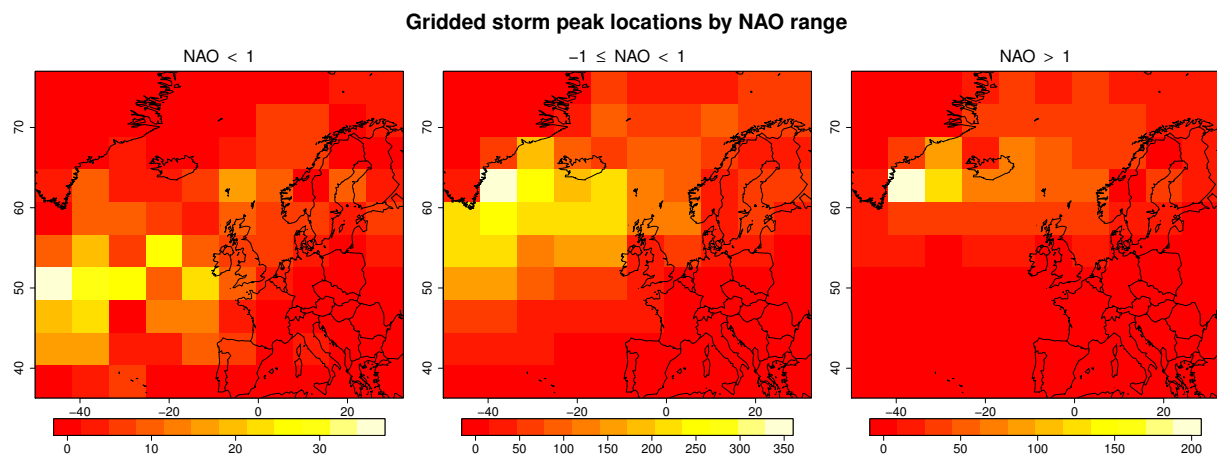**Figure 5.** Storm peak occurrence by month (left) and North Atlantic Oscillation (NAO) value (right) on annual scale.

**Figure 6.** Plots summarising the spatial distribution of storm peaks for different NAO ranges: $(-\infty, -1]$ (left), $(-1, 1]$ (center) and $(1, \infty)$ (right).
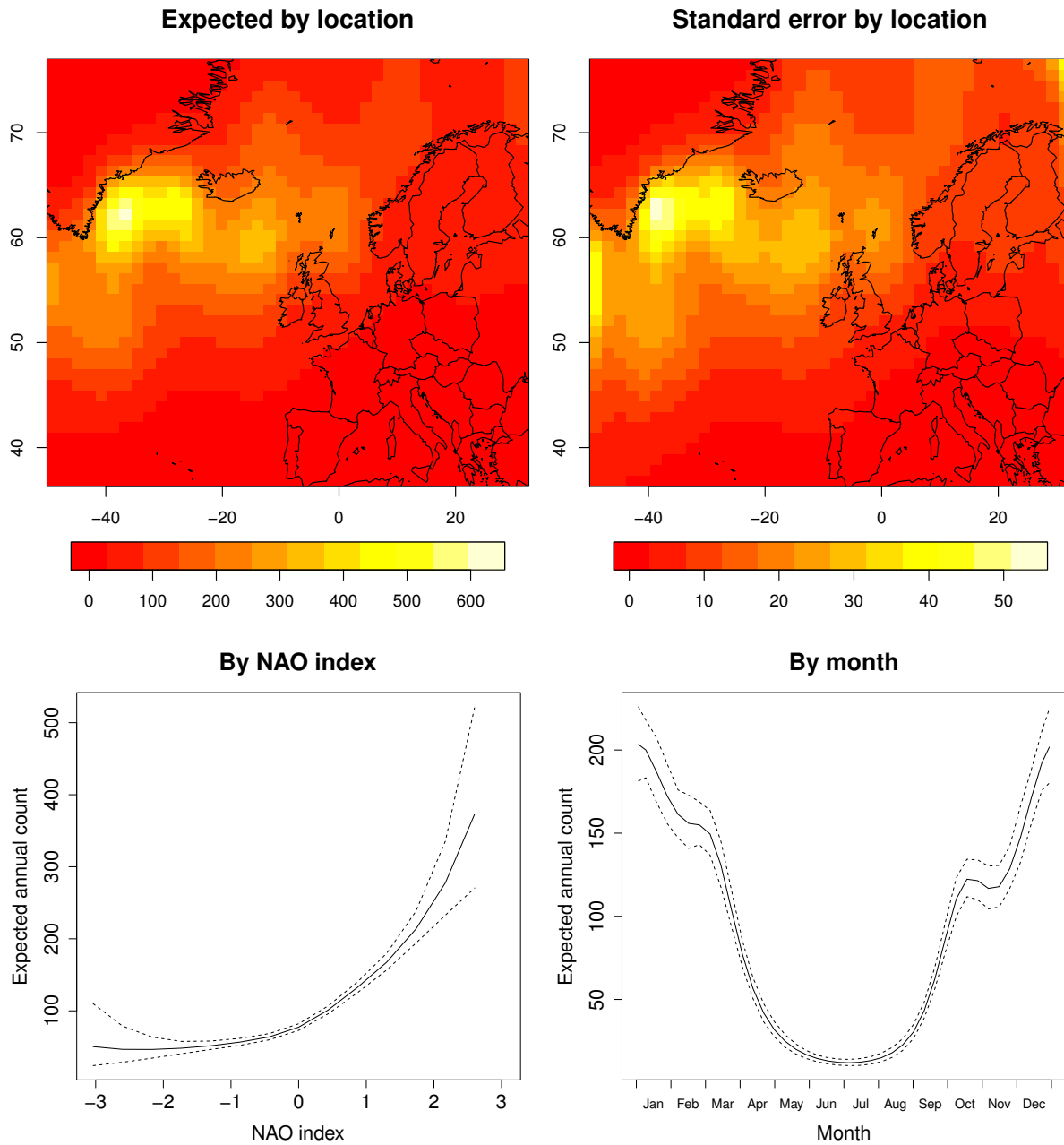
# Model−based annual storm rates

### Expected by location

### Standard error by location

### By NAO index

### By month

**Figure 7.** Model-based estimates of storm peak occurrence represented on annual scale. Plotted are spatial mean estimates (upper left) and standard errors (upper right), and estimates by NAO (lower left) and month (lower right), which are represented as mean estimates (—-) with 95% uncertainty bounds (- - -).
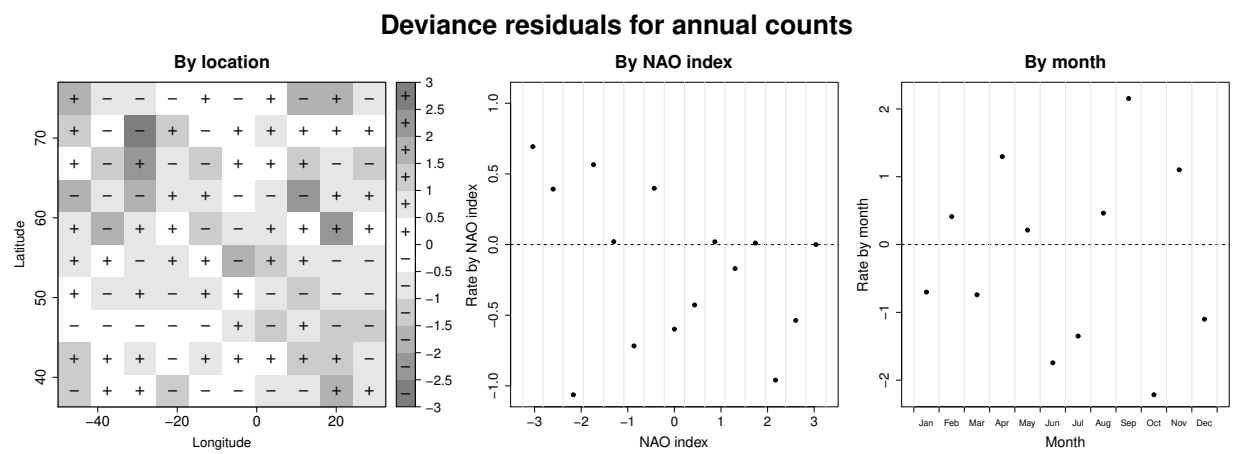
**Figure 8.** Deviance residuals comparing model-based and empirical counts aggregated over space (left), NAO (center) and months (right). Deviance residuals are calculated according to §3.5.
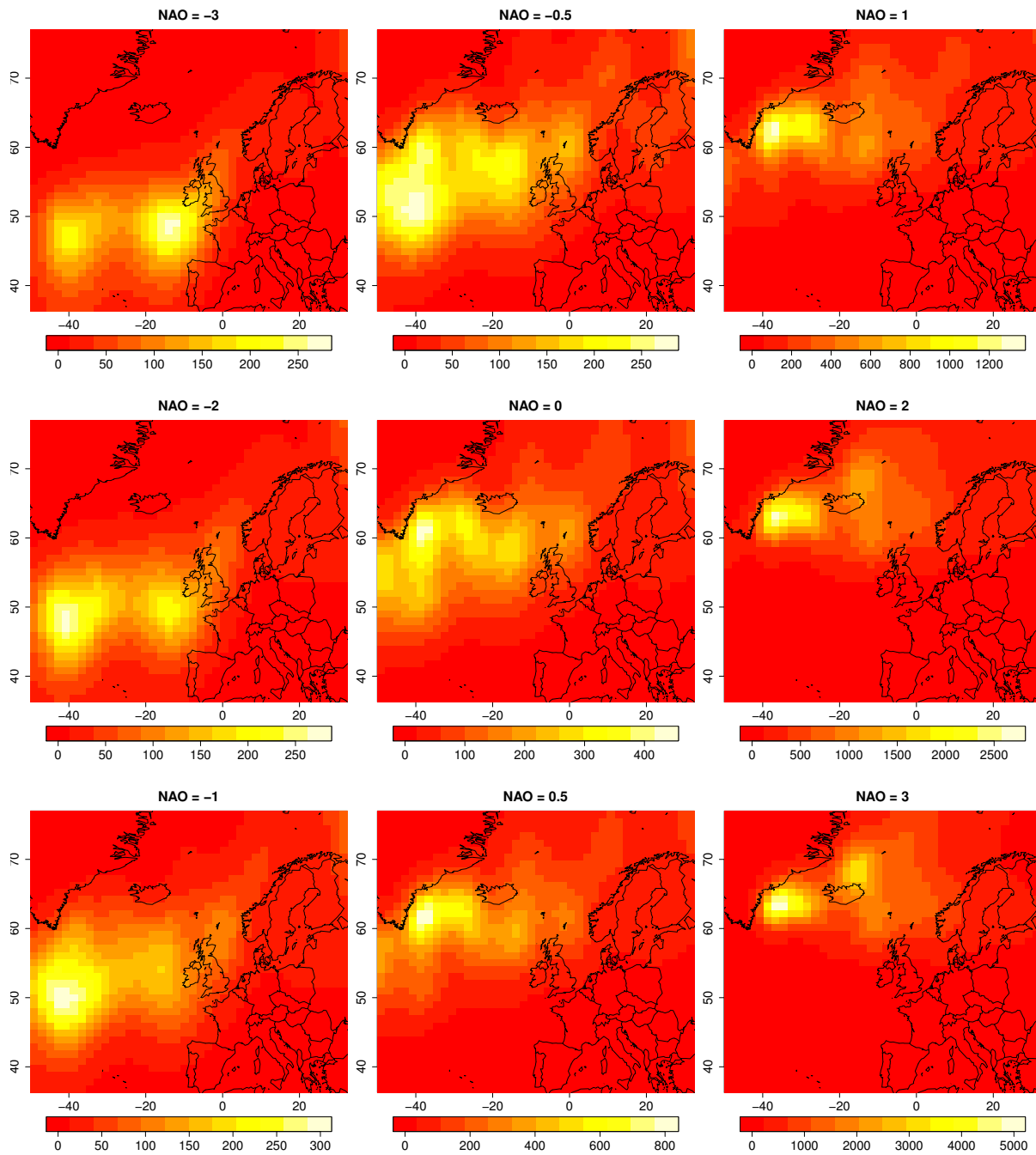
**Figure 9.** Model-based estimates of annual storm peak occurrence locations for different values of the NAO index.