Modelling the spatial extent and severity of extreme European windstorms

Paul Sharkey\textsuperscript{1}, Jonathan A. Tawn\textsuperscript{1} and Simon J. Brown\textsuperscript{2}

Email: p.sharkey1@lancs.ac.uk, j.tawn@lancs.ac.uk, simon.brown@metoffice.gov.uk

\textsuperscript{1}STOR-i Centre for Doctoral Training, Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, U.K.

\textsuperscript{2}Met Office Hadley Centre, Exeter, EX1 3PB, U.K.

Abstract

Windstorms are a primary natural hazard affecting Europe that are commonly linked to substantial property and infrastructural damage. Extreme winds are typically generated by extratropical cyclone systems originating in the North Atlantic, which are often characterised by a track of local vorticity maxima. While there have been numerous statistical studies on modelling extreme winds, little has been done to model the influence of the extratropical cyclone on the wind speeds that they generate. By modelling the development of windstorms in a Lagrangian frame of reference, we can assess the marginal and joint risk of severe events occurring at certain sites. We present a novel approach to modelling windstorms that preserves the physical characteristics linking the windstorm and the cyclone track by exploring the dependence structure of these characteristics in a Lagrangian frame of reference. Our model allows simulation of synthetic windstorm events, which one can use to quantify the risk associated with previously unobserved events at different sites, thus representing a useful tool for practitioners with regard to risk assessment. In particular, we show the spatial extent of windstorms become more localised as its magnitude increases, while our model captures the varying degrees of spatial dependence at different sites.

Keywords: Climate extremes, extratropical cyclones, extreme value analysis, Lagrangian model, spatial dependence.

1 Introduction

While the winter climate of the United Kingdom and northern Europe is typically associated with mild, wet weather that poses little infrastructural or societal risk, there has been an increased focus
in recent years on the impact of windstorms in this part of the world. These events are often the consequence of extratropical cyclones, and are directly linked to the occurrence of flooding, transport chaos and considerable damage to infrastructure. Roberts et al. (2014) describe a comprehensive catalogue of European windstorms in the period 1979-2012 that contains extensive information related to the meteorology and monetary impact of each storm. Storm Daria, which occurred in January 1990, is believed to be the most destructive windstorm in this period, with an estimated insured loss of $8.2bn.

Windstorms are often a consequence of the passage of extratropical cyclones. Extratropical cyclones are synoptic-scale weather systems associated with low pressure that generally originate in the North Atlantic and progress northeasterly towards northern Europe. These systems can be characterised by paths of local vorticity maxima, which we refer to as tracks. Cyclones are typically formed as a result of horizontal temperature gradients and evolve according to a particular lifecycle with associated frontal systems (Shapiro and Keyser, 1990). Windstorms tend to occur along the boundary where cold and warm air masses converge, commonly referred to as a weather front (Hewson and Neu, 2015). A large body of research exists on cyclone identification, storm tracking and feature extraction in reanalysis datasets (Murray and Simmonds, 1991; Hodges, 1995), which produce good approximations of how a track develops in space and time. However, methods don’t currently exist to track the evolution of windstorms relative to the cyclone centre, and how to quantify these in a robust way.

The data record is relatively short with regard to storm tracks, and even more so with regard to windstorms, which motivates the need for a statistical model to provide extra information about the possible extreme, long-term characteristics of windstorms that are generated by the extratropical cyclone. In particular, we would like to assess the likelihood of observing more severe storms than those observed, where these might occur, and how large the spatial extent of the event might be. We would also like to assess the joint risk of multiple locations experiencing the same windstorm event. This is particularly difficult to model as the sites experiencing the event are largely determined by the position of the track relative to these sites. We therefore require a method that accounts for the spatial variability of the storm track in quantifying the risk associated with windstorms at multiple sites. Our paper describes an approach to simulate synthetic windstorms that statistically represents the winds within the cyclone, which can be used in practice to assess the marginal and joint risk of these weather systems over Europe while accounting for the varying probability of storm tracks over the region.
A common approach to statistical modelling of extreme wind speeds is to use techniques from extreme value analysis, which use models built on asymptotic arguments to estimate probabilities of events beyond the range of the data. In meteorological applications, such probabilities are commonly used by practitioners to design infrastructure appropriately to defend against the natural hazard being studied. The most widely-used approach in extreme value analysis is to consider excesses above a suitably high threshold. Consider a sequence of independent and identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$. Under weak conditions on the $X_i$, the unique, non-degenerate distribution that the scaled excesses of a threshold by $X_i$ converges to, as the threshold tends to the upper limit $x_F$ of $X_i$, is the generalised Pareto distribution (GPD) (Pickands, 1975; Davison and Smith, 1990). We make the assumption that this limiting result holds for a large enough threshold $u$. The GPD takes the form

$$\Pr(X_i - u > x|X_i > u) = \left(1 + \frac{\xi x}{\sigma_u}\right)^{-1/\xi}, \quad x > 0$$

(1)

where $c_+ = \max(c, 0)$ and where $\sigma_u > 0$ and $\xi \in \mathbb{R}$ denote the scale and shape parameters respectively. The shape parameter is invariant to the choice of threshold but the scale parameter is threshold-dependent. The threshold $u$ is typically determined using some standard selection diagnostics (Coles, 2001) such as ensuring that the parameters are stable with respect to the threshold for all threshold choices larger than $u$.

There have been numerous studies using extreme value models to estimate extreme wind speeds (Coles and Walshaw, 1994; Fawcett and Walshaw, 2006; Ribatet, 2013). However, these models have no consideration of the physical processes generating the extremes. Some recent studies have, however, modelled extreme winds in the context of an extratropical cyclone. Della-Marta and Pinto (2009) use a GPD model to assess changes in extreme wind intensity under climate change scenarios, which led to results showing that the frequency of intense wind events in Europe is predicted to increase. Sienz et al. (2010) extended this approach to model the effect of the North Atlantic Oscillation (NAO) index. Bonazzi et al. (2012) modelled the tail dependence of wind speeds between locations over Europe using a bivariate extreme value copula and found dependence to be greater in the west-east direction, which is consistent with the passage of extratropical cyclone tracks over Europe. More recently, Youngman and Stephenson (2016) use extreme value analysis coupled with a geostatistical model to capture the spatio-temporal development of windstorms over Europe, but again the direct influence of the storm track is not accounted for.
The approaches described above share a common philosophy in that windstorms are modelled in an Eulerian frame of reference. In fluid mechanics, this refers to the scenario whereby an observer measures observations of a process at a fixed location while the process, e.g., a windstorm, passes over. This is the most common approach to statistical modelling, with the advantage that one can build large spatial data sets with time series at each location being observed. However, if one’s concern is focused more on modelling the evolution and influence of the process itself, a Lagrangian frame of reference is required. Historically, Eulerian approaches have been used to model observational data from sources such as weather gauges. However, recent advances in climate modelling have resulted in the increased availability of high-resolution datasets that are spatially and temporally complete, in which large-scale processes can be modelled in a Lagrangian framework. Previous climate studies on extratropical cyclones in a Lagrangian framework include Catto et al. (2010), Rudeva and Gulev (2011) and Dacre et al. (2012). This approach to modelling requires following the process and collecting observations as it moves through space and time. This is a natural framework on which to build a model for windstorms, as it allows us to explore and model the behaviour of extreme winds relative to the storm track as it moves across the North Atlantic. This is especially useful as many sites have not experienced events of a large spatial extent or magnitude during the observational record, possibly as a consequence of the path storms have taken.

In Sharkey et al. (2017), a model is described such that synthetic storm tracks are generated that replicate the climatology of extratropical cyclones in the North Atlantic. This model also allows the generation of larger extreme storms, with regard to vorticity, than previously observed. In this paper, we present an extension to this work that allows for the simulation of synthetic wind events relative to the storm tracks generated by the model of Sharkey et al. (2017). In particular, we are interested in modelling in a Lagrangian framework the most damaging wind events, which we refer to as windstorms, relative to the centre of an extratropical cyclone. We first describe a model for the area affected by strong winds in the vicinity of the storm centre, which we refer to as a footprint. We represent the footprint as an ellipse and model the evolution of its characteristics through time relative to the storm centre. Next, we describe an approach for modelling the magnitude and spatial distribution of the extreme winds within the footprint. These approaches allow us to generate a series of footprints for multiple windstorms associated with the synthetic storm tracks of Sharkey et al. (2017), providing a method for estimating the risk associated with extreme windstorms over the North Atlantic and Europe.

The paper is structured as follows. In Section 2, we introduce the data and our methods for
extracting the features of the windstorm from the data. We also detail an exploratory analysis of these features in this section. We discuss our modelling strategy in Section 3, introducing our approaches to modelling the evolution of the windstorm footprints and the extreme winds within the footprints. We examine some key results in Section 4, before concluding in Section 5 with some discussion.

2 Windstorm definition and exploratory analysis

2.1 Data description

As in Sharkey et al. (2017), our work uses storm track data covering the North Atlantic and European domain. In particular, our dataset consists of storm track locations at 3-hourly intervals with an associated vorticity measure representing the strength of the storm. Storms are identified and tracked over the period 1979-2014 from the ERA-Interim reanalysis dataset (Dee et al., 2011) using a feature extraction approach outlined in Hoskins and Hodges (2002) based on the tracking algorithm introduced in Hodges (1995). We restrict our attention to the set of storm tracks produced during an extended winter period (October-March), when storms are widely regarded to be most intense. We exclude Mediterranean storms as these often arise as a result of convective behaviour in the atmosphere and are not captured well by reanalysis data (Akhtar et al., 2014). We denote the longitude and latitude coordinates of the storm track at time $t$ by $\text{Lon}_t$ and $\text{Lat}_t$ respectively. The vorticity associated with the track at $(\text{Lon}_t, \text{Lat}_t)$ is denoted by $\Omega_t$.

Our model is based on wind speed data from the EURO4 numerical weather prediction model (Standen et al., 2017), which is a downscaled version of the ERA-Interim reanalysis dataset. Data are available on a 4 km spatial resolution over Europe and part of the North Atlantic, amounting to $1,100,000$ cells (see Figure 1). Values are obtained at hourly intervals over the period 1979-2014. We linearly interpolate the storm track locations and vorticity within each 3-hourly interval to match the hourly temporal resolution of the wind speed data. We select only the wind speed fields at times corresponding to the set of storm tracks. In particular, as we are looking to model the effect of the storm track on the spatio-temporal evolution of wind speeds in the vicinity of the track, we isolate the field of interest as a square-shaped region centred at the storm centre with sides of approximately 1,600 km in length (see Figure 2, which in the left panel shows such a region at a time step when storm Daria was located over the UK). We believe that this field is large enough so that the extreme winds generated by a windstorm are sufficiently captured.
Figure 1: The region corresponding to the availability of data from the EURO4 numerical weather prediction model.

2.2 Marginal model

Initial investigation of the data confirms, as expected, that winds over the sea are markedly stronger than those over land (see Figure 2, left panel). This is largely due to open water exerting significantly less drag on the atmosphere in contrast with the land surface, orography and man-made structures that impede strong winds. The contrast in scale over land and sea, and to a lesser extent, over low-lying and high-lying land, motivates a standardisation of wind speeds in each cell to have a common marginal distribution. Let $X(s,t)$ be a random variable denoting the wind speed in cell $s$ at time $t$, for $s = 1, \ldots, 1,100,000$. We propose a marginal model for $X(s,t)$ of the form

$$F_s(x) = \begin{cases} \hat{F}_s(x) & x \leq u_s \\ 1 - \lambda_{u_s} \left( 1 + \xi_{u_s} \frac{x - u_s}{\sigma_{u_s}} \right)^{-1/\xi_{u_s}} & x > u_s, \end{cases}$$

where $\hat{F}_s$ denotes the empirical distribution function for realisations of $X(s,t)$. For realisations above $u_s$, the GPD in (1) is used as a conditional model for excesses above $u_s$, with cell-specific parameters $(\sigma_{u_s}, \xi_{u_s})$. To undo this conditioning, a third parameter $\lambda_{u_s}$, denoting the probability of an exceedance of $u_s$, must be specified. Parameter stability plots were checked at a number of cells, which indicated that a threshold corresponding to the 98% quantile would be a good choice for all cells. We therefore choose this quantile to be the cell-specific threshold in each cell. Parameter estimates are obtained using maximum likelihood techniques. We note that we do not attempt to impose spatial smoothness on the form of $F_s$. 
Figure 2: Wind speeds, in m/s, at 3pm on January 25th, 1990 in the vicinity of Storm Daria (left) and standardised onto Exp(1) margins (right). The storm centre is represented by the cross. The white box contains an example of a localised convective event. Land/sea borders are not explicitly shown on in the left panel, but can be seen due to the contrast in magnitude between winds over land and sea.

Figure 3 shows the parameter estimates of the GPD corresponding to each cell in the region over Europe as shown in Figure 1 along with the threshold, which corresponds to the 98% quantile in each cell. This shows explicitly the contrast in wind speed magnitudes between locations on land and sea. This contrast is also reflected in the estimation of the scale parameter, but the shape parameter exhibits no such contrast between land and sea, with most estimates occurring in the region $(-0.2, 0)$, indicating that the distribution of wind speeds has a finite endpoint in general. The numerical maximisation algorithm used to obtain the parameter estimates mostly converges, however there are certain regions that exhibit unusual behaviour. For example, the 98% quantile corresponding to the threshold is a lot higher in areas over Iceland than other land locations, while the Italian Alps see unusually high estimates of the shape parameter. As weather variables are not represented well by reanalysis data in regions of high orography, as evidenced by these two particular locations, we exclude regions with orography $> 500$m in our analysis. Marginal bias of this type has been studied, with recalibration methods proposed in Howard and Clark (2007),
which offers a possible extension of our method.

We propose a marginal standardisation to unit exponential Exp(1) margins using a probability integral transform using the marginal model (2). We define $X^E(s, t)$ to be the relative wind speed on Exp(1) margins in cell $s$, such that for all cells and all times

$$X^E(s, t) = -\log\{1 - F_s(X(s, t))\},$$

where $F_s$ is defined as in (2). We use the term relative wind speeds to describe $X^E(s, t)$ as this quantity defines wind speeds relative to the marginal characteristics of cell $s$. Figure 2 (right panel) shows the effect of the standardisation on one time step of Daria. In particular, we see spatially correlated values of high relative wind speeds over both land and sea as a result of the transform and the land/sea contrast is almost entirely removed. This approach provides some intuition regarding the shape of the windstorm event without the influence of the marginal characteristics at each location, which should allow for a simpler approach to modelling the spatial extent of the event. In this sense, our approach is based on a copula modelling strategy over space (Joe, 1997).

### 2.3 Feature extraction

Figure 2 shows that, as well as the large band of strong relative winds clustered near the storm centre, small localised fragments of high relative wind speeds are visible on the western edge of the footprint. Such events are due to localised convection and not due to the larger scale dynamics of the storm. Since we believe them not to be directly linked with the extratropical storm system, we are not concerned with modelling these features. Our work is focused on modelling the features of
an extratropical cyclone that occur on larger spatial scales and have the potential to produce much larger impacts than localised convective events. This motivates developing some methodology to extract the main features of interest from the standardised fields, in particular, the larger cluster of high relative wind speeds. We apply a spatio-temporal Gaussian filter (Nixon and Aguado, 2012) to each field, which removes the effect of the small-scale convective events. From the locations for which the filtered data exceed some arbitrary threshold \( v \), we extract the relative wind speeds. Relative winds corresponding to the other locations are set to zero, thus masking convective wind events and other non-extreme winds.

In many situations, this filtering step masks the entire field, which we interpret as there being no significant windstorm activity, that is, the windstorm is in an inactive phase. When this isn’t the case, the filtered footprint consists of at least one cluster of cells corresponding to non-zero relative wind speeds. Since our objective is to model the primary effects of the storm on the surrounding winds, we extract the largest cluster of cells in terms of size and use this to define the features of the windstorm event. We do this using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm (Ester et al., 1996), which recursively groups cells into distinct clusters based on adjacency to neighbouring cells. Unlike other standard clustering approaches, this can be achieved without having to specify the number of clusters in advance. We then extract the largest cluster of cells in size as defined by the algorithm, an example of which is shown for storm Daria in Figure 4 (left panel).

Since we are interested in modelling how the storm influences the magnitude and spatial extent of wind speeds, we construct a set of variables to adequately represent and summarise these characteristics. We therefore impose the shape of the region obtained as a result of the filtering and clustering steps to take the form of an ellipse. We believe an ellipse is a suitably flexible shape to choose as it can describe well the position (by the centre of the ellipse), size and shape (by the area defined by the semi-major and semi-minor axes) of an event. To define this ellipse, we use Khachiyan’s optimisation algorithm for fitting the minimum-area ellipse enclosing a set of cells (Moshtagh, 2005; Todd and Yıldırım, 2007). Specifically, consider a set \( S \) containing \( m \) locations in 2-dimensional space such that \( S = \{s_1, \ldots, s_m\} \in \mathbb{R}^2 \). We can define an ellipse by the set \( \mathcal{E} \) such that

\[
\mathcal{E} = \{ s \in \mathbb{R}^2 : (s - c)^T E (s - c) \leq 1 \},
\]

where \( c \) is the centre of the ellipse and \( E \) is a positive-definite matrix. The area of \( \mathcal{E} \) is \( \{ \det E^{-1} \}^{1/2} \). Therefore, to determine the ellipse of minimum area containing the cells in \( S \), we must find a vec-
tor $c$ and positive definite matrix $E$ which minimises $\det E^{-1}$ subject to $(s_i - c)^T E (s_i - c) \leq 1$ for $i = 1, \ldots, m$. This optimisation problem is solved using conditional gradient ascent methods, the details of which can be found in Moshtagh (2005). We remove the masking of wind speeds below $v$ for the empty regions of the ellipse so that a full ellipse of relative winds is defined. Figure 4 (centre panel) shows the ellipse fitted to the largest cluster of cells for the same time step of Daria as in Figure 2. For the remainder of the paper, we refer to the ellipse of high relative wind speeds associated with an extratropical cyclone as a footprint. In the scenario whereby a series of footprints are associated with a single cyclone, we refer to this as a windstorm. If this series has gaps without footprints between periods with footprints, we refer to these as inactive and active phases of the windstorm.

Figure 5 shows the set of random variables we extract from the set of footprints that summarise the shape and magnitude of windstorms at a given time step relative to the storm centre $(\text{Lon}_t, \text{Lat}_t)$ with associated vorticity $\Omega_t$. We use the notation $\mathcal{E}_t$ to denote the subset of cells in the spatial domain in Figure 1 contained within the footprint. We define $A_t$ and $B_t$ to be the semi-major and semi-minor axes of the ellipse respectively at time $t$. The quantity $R_t^E$ represents the grid cell distance between the storm centre and the centre of the ellipse at time $t$, while $\Theta_t^E \in [-\pi, \pi]$ denotes the bearing of the centre of the ellipse relative to due south at time $t$. We define $\Gamma_t \in [-\pi/2, \pi/2]$
to be the orientation of the ellipse relative to due north, while \( W_t \) denotes the maximum relative wind speed observed in the footprint at time \( t \). We define \( R_t^W \) as the grid cell distance between the centre of the ellipse and the cell corresponding of maximum relative wind speed at time \( t \), while \( \Theta_t^W \in [-\pi, \pi] \) denotes the bearing between the two cells relative to due south.

2.4 Exploratory analysis

We undertake an extensive exploratory analysis on a number of aspects driving and influencing the evolution of windstorm activity, of which some findings are reported here. First, we investigate the dependence structure of characteristics of the ellipse representing the windstorm footprint, as shown in Figure 5. We assess the factors influencing the activation and termination of windstorm events. We also look at some quantities representing the spatial distribution of wind speeds relative to the storm centre, and how these compare with previous studies. Finally, we explore how the distribution of relative wind speeds within the footprint varies with respect to characteristics of the storm track and footprint.

Figure 6 shows boxplots illustrating some key dependencies between variables of a footprint. The area of the ellipse, which is proportional to \( \Delta_t = A_t \times B_t \), tends to increase as vorticity and maximum relative wind speed increases, indicating that the strongest events tend to occur on a larger spatial scale. The radius \( R_t^E \) tends to decrease as \( \Omega_t \) increases, although the effect is small, suggesting that footprints tend to occur closer to the storm centre when a large vorticity is observed. Maximum relative wind speed \( W_t \) tends to increase as \( \Omega_t \) increases, though this dependence is weak.
Figure 6: Boxplots showing the dependence between $\Omega_t$ and $R_t^E$, $\Omega_t$ and $\Delta_t^{1/2}$, $\Delta_t^{1/2}$ and $W_t$, and $\Omega_t$ and $W_t$. 
We also examine partial autocorrelation plots (not shown) to determine how individual components of the ellipse depend on their lags. This shows evidence of a second-order temporal dependence structure in most components reflecting the smooth evolution of footprints through the windstorm.

Figure 7: The spatial density of track locations that are associated with observed (top) and simulated (bottom) windstorms.

A consequence of our extraction of footprints is the loss of a large number of wind fields that have been removed for being non-extreme. Thus, if we are to construct our model that simulates footprints in time, we need a model that explains the factors influencing whether a windstorm is active or not. With regards to windstorm activation, we investigate the components of the track
that may trigger an event. We find that the probability of windstorm activations tends to increase as vorticity increases, signalling a direct link between the intensity of the storm track and the occurrence of windstorm events. With regard to an inactive phase caused by termination of an active phase, we use the variables of the windstorm footprint in addition to information from the track. We find that a storm is more likely to terminate if it is associated with small values of relative maximum wind speed, vorticity and area. This indicates a termination is more likely if the windstorm weakens both in terms of its magnitude and its spatial scale.

This exploratory analysis also shows some spatial variation in the occurrence of footprints. Figure 7 (top panel) shows the density of storm track locations when windstorms are in an active phase. This shows that windstorms tend to occur over the North Atlantic and western Europe, with the density decaying as one moves to the edges of the domain. The reductions in density on the western boundary are not considered to be physical but the result of both the filtering of footprints to avoid their truncation by the EURO4 boundary and the effect of the boundary within the EURO4 simulation of weather. Most extratropical cyclones enter the EURO4 domain through this western boundary but their intensity is diminished due to the coarser resolution of the driving GCM. Their subsequent intensification takes a number of time steps which results in smaller footprints and lower winds towards this boundary (see Figure 3).

We collect footprints over all fields centred at a point on the storm track, like in Figure 4, and assess the spatial distribution of these fields relative to the storm centre. We find that some events are picked up spuriously by the feature extraction algorithm that may not be generated as a result of the extratropical cyclone. With this in mind, we exclude ellipses whose centre occurs on the outer edges of the footprint as well as events that are sufficiently small. Figure 8 shows the mean, 95% quantile and 99% quantile of wind speeds over the remaining fields after this filtering is implemented. This shows that relative wind speeds tend to be larger in regions southwest of the storm centre, which arise with the passage of cold fronts. This figure also shows the density of events over all footprints, illustrating that windstorms are most likely to occur southwest of the storm centre, with very few events occurring in the northern half of the field in comparison. These are consistent with the observations of Catto et al. (2010) and Rudeva and Gulev (2011).

Unusual events with large magnitudes are detected on the northwest and southeast edges of the domain. Despite our filtering, these events may not be generated by the extratropical cyclone; however, they are very rare and their impact minimal. The mean behaviour shows a local minimum
Figure 8: The mean, 95% quantile, 99% quantile and density of wind speeds relative to the storm centre (represented here as a cross) over all observed footprints.
occurring close to the storm centre, which likely arises as a result of low wind speeds occurring at pressure minima (Catto et al., 2010), which is known to be spatially adjacent to where the maximum vorticity occurs (Hoskins and Hodges, 2002).

We investigate the factors influencing the distribution of relative wind speeds within each footprint through analysis of their mean and standard deviation. This reveals, intuitively, that the mean and standard deviation of winds tend to increase as the maximum relative wind increases. In contrast, an increase in vorticity tends to slightly reduce the mean relative wind, while increasing the standard deviation. Additionally, the standard deviation of wind increases as the area of the ellipse increases, which one would expect given the increased chances of observing smaller and larger values of wind speed. By construction the strongest relative winds tend to occur near the location of the maximum, determined by $R_{Wt}^t$ and $\Omega_{Wt}^t$, while weaker relative winds are more likely closer to the perimeter of the ellipse. Additionally, the relative winds tend to exhibit anisotropic properties, which tends to manifest in a ‘stretching’ of the band of strongest winds oriented in the direction perpendicular to $\Theta_t$. As the windstorm evolves, this gives the effect of the winds ‘bending’ around the storm centre.

3 Windstorm modelling

3.1 Introduction

We propose an approach for simulating windstorms relative to an extratropical cyclone track. In particular, we describe methods, motivated by our findings in Section 2.4, for simulating a footprint at each time step of a windstorm along with the spatial distribution of wind speeds within a footprint. The times at which a windstorm is in an active phase with respect to a cyclone track are denoted by $\{t_S, \ldots, t_A, \ldots, t_T\}$, where $t_S$ and $t_T$ denote the times corresponding to the beginning and end of an active phase respectively and $t_A$ denotes the time at which the windstorm is judged to be active. If a windstorm becomes active at time $t_A$, forwards and/or backwards propagation routines, outlined in Section 3.2, are used to model the evolution of the footprint through the windstorm until times $t_T$ and/or $t_S$ are reached, which we refer to as the termination of the active phase. In practice, it can be more subtle as a windstorm can have repeated phases of activity and inactivity along a cyclone track. Details of this feature are discussed in Section 3.3. Stochastic models for determining $t_A$, $t_S$ and $t_T$ are also described in Section 3.3, along with a more detailed description of the simulation procedure. Our approach for modelling the wind speed fields within the footprint is outlined in Section 3.4.
3.2 Footprint modelling

Our approach aims to model the spatio-temporal evolution of footprints using the ellipse structure of Section 2.3. As this structure describes the spatial extent of a windstorm at a given time step, we can model the temporal evolution of the footprint by exploiting the time series structure of the footprint variables shown in Figure 5, which determine the position, size and magnitude of a footprint relative to the storm centre. Supported by the exploratory data analysis in Section 2.4, we assume that the multivariate time series \( Z_t = \{ A_t, B_t, W_t, R_t^E, \Theta_t^E, R_t^W, \Theta_t^W, \Gamma_t \} \) jointly follows a stationary \( k \)th order Markov process during active phases along the cyclone track and are conditionally independent given the track information over tracks and on different active phases along a track. By the Markov property, the distribution of the current value of a process is affected only by the previous \( k \) time steps of the process. We define an arbitrary \( d \)-dimensional \( k \)th order stationary multivariate Markov process \( Z_{1:n} = \{ Z_{tj} : t = 1, \ldots, n; j = 1, \ldots, d \} \), where \( Z_{tj} \) denotes the \( t \)-th time step of the \( j \)-th component and \( n \) is the length of the time series. Consequently, it is only necessary to model the joint distribution of \( Z_{t:t+k} \), from which the conditional density function of \( Z_{k+1} \mid Z_{1:k} \) can be derived. This joint distribution is determined by its marginal distributions and its copula. This structure also allows us to capture dependence between the different components of the time series, e.g., vorticity and distance from the storm centre.

Under the assumption of stationarity of \( Z_{t:t+k} \), observations of this \( d \times (k+1) \) vector are identically distributed over \( t \) and with joint density function \( f \). A simple choice is to model \( f \) nonparametrically using a multivariate kernel smoothed density function \( \hat{f} \), such that

\[
\hat{f}(z) = \frac{1}{n} \sum_{i=1}^{n} K_H(z - z_i),
\]

where \( K_H \) denotes the kernel function, defined with respect to a \((k+1) \times (k+1)\) bandwidth matrix \( H \). The function \( K_H \) is chosen to be the multivariate Gaussian density function with variance \( H \). Our exploratory analysis in Section 2.4 helps to identify which components of \( Z_{t:t+k} \) are independent or conditionally independent, which simplifies the form of \( H \) and helps to identify \( k \). For example, we found \( k = 2 \) and a weakly dependent relationship between \( \Omega_t \) and \( W_t \) with \( \Omega_t \) and \( W_t \) conditionally independent given \( A_t \) and \( B_t \). This allows us to simplify our model for \( \Pr(Z_{t+k} \leq z \mid Z_{t:t+k-1} = z_{t:t+k-1}) \), for which we use the kernel estimate of the conditional distribution function, the formulation of which can be found in Appendix A.

Our approach is designed to construct the joint distribution for \( \{ Z_{t_S}, \ldots, Z_{t_A}, \ldots, Z_{t_T} \} \), where \( \{ t_S, \ldots, t_A, \ldots, t_T \} \) are the consecutive time steps at which the windstorm is active, as discussed
in Section 3.1. The characteristics of the footprint are influenced by the storm track components (Lon_t, Lat_t, Ω_t) so our footprint propagation routine needs to reflect this. To account for the spatial variation of the footprint characteristics, we simulate realisations of \( Z_t \) only using conditional kernel density estimates based on footprints associated with values of (Lon_t, Lat_t) in \( \nabla_t \), where \( \nabla_t \) is a region of size 20° × 14° centred at (Lon_t, Lat_t). First, consider forward propagation from a given value of \( Z_{t_A} \) whilst the windstorm remains in an active phase. For all times \( t_A + 1 \leq j \leq t_A + 2 \).

We simulate forwards realisations of \( Z_j \) from a conditional kernel density derived from \( \hat{f}(z) \), as described in Appendix A, such that

\[
\begin{align*}
r_j^E &\sim R_j^E | R_{t_A;j-1}^E = r_{t_A;j-1}^E, \Omega_j = \omega_j, (Lon_j, Lat_j) \in \nabla_j \\
\theta_j^E &\sim \Theta_j^E | \Theta_{t_A;j-1}^E = \theta_{t_A;j-1}^E, R_j^E = r_j^E, (Lon_j, Lat_j) \in \nabla_j \\
a_j &\sim A_j | A_{t_A;j-1} = a_{t_A;j-1}, \Omega_j = \omega_j, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, (Lon_j, Lat_j) \in \nabla_j \\
b_j &\sim B_j | B_{t_A;j-1} = b_{t_A;j-1}, \Omega_j = \omega_j, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, (Lon_j, Lat_j) \in \nabla_j \\
w_j &\sim W_j | W_{t_A;j-1} = w_{t_A;j-1}, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, A_j = a_j, B_j = b_j, (Lon_j, Lat_j) \in \nabla_j \\
\gamma_j &\sim \Gamma_j | \Gamma_{t_A;j-1} = \gamma_{t_A;j-1}, \Theta_j^E = \theta_j^E, (Lon_j, Lat_j) \in \nabla_j \\
r_j^W &\sim R_j^W | R_{t_A;j-1}^W = r_{t_A;j-1}^W, A_j = a_j, (Lon_j, Lat_j) \in \nabla_j \\
\theta_j^W &\sim \Theta_j^W | \Theta_{t_A;j-1}^W = \theta_{t_A;j-1}^W, (Lon_j, Lat_j) \in \nabla_j.
\end{align*}
\]

When \( j > t_A + 2 \), we simulate from the conditional kernel model

\[
\begin{align*}
r_j^E &\sim R_j^E | R_{j-k;j-1}^E = r_{j-k;j-1}^E, \Omega_j = \omega_j, (Lon_j, Lat_j) \in \nabla_j \\
\theta_j^E &\sim \Theta_j^E | \Theta_{j-k;j-1}^E = \theta_{j-k;j-1}^E, R_j^E = r_j^E, (Lon_j, Lat_j) \in \nabla_j \\
a_j &\sim A_j | A_{j-k;j-1} = a_{j-k;j-1}, \Omega_j = \omega_j, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, (Lon_j, Lat_j) \in \nabla_j \\
b_j &\sim B_j | B_{j-k;j-1} = b_{j-k;j-1}, \Omega_j = \omega_j, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, (Lon_j, Lat_j) \in \nabla_j \\
w_j &\sim W_j | W_{j-k;j-1} = w_{j-k;j-1}, R_j^E = r_j^E, \Theta_j^E = \theta_j^E, A_j = a_j, B_j = b_j, (Lon_j, Lat_j) \in \nabla_j \\
\gamma_j &\sim \Gamma_j | \Gamma_{j-k;j-1} = \gamma_{j-k;j-1}, \Theta_j^E = \theta_j^E, A_j = a_j, (Lon_j, Lat_j) \in \nabla_j \\
r_j^W &\sim R_j^W | R_{j-k;j-1}^W = r_{j-k;j-1}^W, A_j = a_j, (Lon_j, Lat_j) \in \nabla_j \\
\theta_j^W &\sim \Theta_j^W | \Theta_{j-k;j-1}^W = \theta_{j-k;j-1}^W, (Lon_j, Lat_j) \in \nabla_j.
\end{align*}
\]

Backwards simulation is implemented similarly, but with substituting \( Z_j \) for \( Z_{t_A-j} \). The details of whether a forward or backward routine should be used are found in Section 3.3. Realisations of \( R_j^W \) and \( \Theta_j^W \) are rejected if the simulated position of the maximum occurs outside of the simulated ellipse. For marginal components of \( Z_t \) where the extremal behaviour is of interest, e.g., \( W_t \), one
could follow the approach of Sharkey et al. (2017) by specifying a model for the extremal marginal and temporal dependence structure; this is discussed in more detail in Section 5.

### 3.3 Activation and termination

As discussed in Section 3.1, we require an approach for determining the times at which a windstorm is in an active phase. Whether the windstorm is active or not is modelled by a Bernoulli logistic generalised additive model (Wood, 2006) based on covariate information from the physical structure of an extratropical cyclone. Section 2.4 demonstrated that the windstorm is typically active at the time when the maximum vorticity is observed, which we denote by \( t_\Omega \), as this time is associated with the strongest winds over the cyclone track. However, we observe some instances where the windstorm is inactive at \( t_\Omega \), but is active at other times on the cyclone track. Similarly, the termination of a windstorm is highly uncertain, but it is shown in Section 2.4 to be linked to weakening events in terms of spatial scale and magnitude.

Specifically, first consider the activation of a windstorm phase. Let \( T_t \) be a Bernoulli random variable such that:

\[
T_t = \begin{cases} 
1 & \text{the storm is active at time } t \\
0 & \text{otherwise.} 
\end{cases}
\]

So \( T_t \sim \text{Bernoulli}(p_t) \), where

\[
p_t = \frac{\exp \left\{ \sum_{i=1}^{q} \beta_i(\nu_{i,t}) \right\}}{1 + \exp \left\{ \sum_{i=1}^{q} \beta_i(\nu_{i,t}) \right\}},
\]

where \( \beta_i \) is a smooth non-linear function of covariate \( \nu_i \) with \( i \in (1, \ldots, q) \), where \( q \) is the number of covariates and \( \nu_{i,t} \) denotes the realisation of the \( i \)th covariate at time \( t \). The smooth functions are represented by penalised regression splines, where the smoothing parameter is determined using generalised cross validation (GCV) and the model is fitted using penalised maximum likelihood. For more details on additive models, see Wood (2006). An identical formulation is used for the termination of a windstorm.

For a storm track, let \( t \) denote the time from its start \( t = 1 \) until its end \( t = l \). We first attempt to activate the windstorm simulation process at time \( 1 \leq t_\Omega \leq l \). If a windstorm is determined to be active at \( t_\Omega \), then \( t_A = t_\Omega \) and we propagate forwards and backwards in time using the methods described in Section 3.2 until the active phase terminates. If a windstorm is inactive at \( t_\Omega \), we proceed successively forwards in time along the track and check whether a windstorm becomes active at times \( t_\Omega < t \leq l \). If a windstorm is inactive at all these times, we proceed successively backwards along the track and check whether an activation occurs at times \( 1 \leq t < t_\Omega \).
If a windstorm is first found to be active during the forwards procession such that \( t_A > t_\Omega \), we set \( t_S = t_A \) and propagate forwards using the approach in Section 3.2 until termination at time \( t_T \leq l \). Likewise, if a windstorm is first found to be active during the backwards procession such that \( t_A < t_\Omega \), we set \( t_T = t_A \) and propagate backwards until termination at time \( t_S \geq 1 \). We allow for the possibility that multiple phases of consecutive footprints can occur on the same track. If \( t_S > 1 \), we proceed backwards along the track to check if the windstorm reactivates, in which the same procedure applies until \( t = 1 \). Similarly, if \( t_T < l \), we proceed forwards along the track to check if the windstorm reactivates, in which the same procedure applies until \( t = l \).

Given that a windstorm is active, some initial footprint characteristics are required for propagating the storm in time. In doing so, we would like to model the joint distribution of \( Z_{t_A} = \{A_{t_A}, B_{t_A}, W_{t_A}, R_E^{t_A}, \Theta_E^{t_A}, P_W^{t_A}, \Theta_W^{t_A}, \Gamma_{t_A}\} \), where \( Z_{t_A} \) represents the characteristics of the footprint at time \( t_A \). Our model should account for the dependence structure between \( Z_{t_A} \) and the components of the track, specifically \( (\text{Lon}_{t_A}, \text{Lat}_{t_A}, \Omega_{t_A}) \). We therefore simulate a realisation of \( Z_{t_A} \) jointly from \( Z_{t_A} | (\text{Lon}_{t_A}, \text{Lat}_{t_A}, \Omega_{t_A}) \), which we estimate using a conditional kernel density (see Appendix A). If \( t_A = t_\Omega \), we use all observed footprint data \( z_{t_\Omega} \) in constructing our density function. Otherwise, if \( t_A \neq t_\Omega \), we use \( z_\tau \) where \( \tau = \{t_A : t_A \neq t_\Omega\} \), that is, footprint data corresponding to when the windstorm is activated during the forwards and backwards procession along the track.

3.4 Modelling wind speeds within a footprint

We also require an approach for modelling the spatial distribution of relative wind speeds within a footprint at each time step of a windstorm. A natural class of models to consider is Gaussian processes, which are widely used in geostatistics to model spatial data. A Gaussian process can be used to describe the joint distribution of random variables over a continuous domain such as space inside a footprint, while for any finite collection of locations in the space the variables follow a multivariate Gaussian distribution. For a comprehensive overview of Gaussian process modelling for spatial data, see Cressie (1993), Stein (1999) and Diggle and Ribeiro (2007).

Let \( \{X_E(s,t) : s \in E_t\} \) denote the field of relative wind speeds in the ellipse \( E_t \) at time \( t \), where \( X_E(s,t) \) is marginally \( \text{Exp}(1) \) distributed over \( t \) for each \( s \). Let \( D_t \) be the distribution function of \( X_E(s,t) \) for all \( s \in E_t \) at time \( t \). We use a probability integral transform to convert to a Gaussian field, which we denote by \( X^G(s,t) \), with

\[
X^G(s,t) = \Phi^{-1} \left[ D_t(X_E(s,t)) \right],
\]

(4)
for all \( s \in E_t \) and all \( t \), where \( \Phi \) denotes the standard Gaussian distribution function. We make the assumption that \( \{X^G(s, t) : s \in E_t \} \) follows a Gaussian process with zero mean and unit variance for each \( t \) such that

\[
X^G(s, t) \sim \text{GP}(0, 1, \rho((s_1, s_2) J_t \Psi_t)),
\]

where \( \rho(\cdot) \) denotes an isotropic correlation function, \( s_1, s_2 \in E_t \) and

\[
\Psi_t = \begin{bmatrix}
\cos \psi_t & -\sin \psi_t \\
\sin \psi_t & \cos \psi_t
\end{bmatrix},
\]

\[
J_t = \begin{bmatrix}
1 & 0 \\
0 & \zeta_t
\end{bmatrix},
\]

where \( \psi_t \) is the time-varying anisotropy angle representing the counter-clockwise rotation of the space and \( \zeta_t > 1 \) is the time-varying anisotropy ratio, which controls the degree of stretching along the angle where correlation decays most slowly with increasing distance. Supported by the exploratory analysis in Section 2.4, we fix \( \psi_t \) to be the angle perpendicular to \( \Theta_t \), while empirical evidence suggests that a good choice of \( \zeta_t \) would be the ratio \( A_t/B_t \).

The correlation function is typically chosen so that the correlation between \( X^G(s_1, t) \) and \( X^G(s_2, t) \) decreases as the distance \( |s_2 - s_1| \) between the sites increases. A common choice of correlation function is the Matérn family, which has the form

\[
\rho(u) = \left\{2^{\kappa-1} \Gamma(\kappa)\right\}^{-1}(u/\alpha_t)^{\kappa}K_\kappa(u/\alpha_t),
\]

(5)

where \( \kappa > 0 \) is a shape parameter that determines the smoothness of the underlying process, \( K_\kappa \) denotes a modified Bessel function of the second kind of order \( \kappa \), and \( \alpha_t > 0 \) is a time-varying scale parameter with dimensions of distance. The Matérn family is a generalisation of other common choices of correlation functions. For example, choosing \( \kappa = 0.5 \) gives the exponential correlation function, while as \( \kappa \to \infty \), the form reduces to a Gaussian correlation function \( \rho(u) = e^{-\alpha_t u^2} \). The parameters \( \kappa \) and \( \alpha_t \) are highly non-orthogonal, but larger values of \( \kappa \) are typically associated with smoother fields, while larger values of \( \alpha_t \) gives fields that are correlated at larger distances. We fix \( \kappa = 0.6 \), which in practice generates fields of similar smoothness to the observed fields. We estimate the parameter \( \alpha_t \) corresponding to each footprint using variogram methods. We avoid the use of likelihood methods due to the computational difficulties that arise with large spatial datasets like ours. For more details on inference for Gaussian processes, see Diggle and Ribeiro (2007). Exploratory analysis suggests that \( \alpha_t \) is constrained by the value of \( \Delta_t \), which is proportional to the area of the ellipse. We therefore model \( \alpha_t \mid \Delta_t \) using a conditional kernel density, the formulation of which is outlined in Appendix A. At time \( t \), we generate a realisation of \( \alpha_t \) conditional on the simulated realisation of \( \Delta_t \) determined by the model in Section 3.2.
We require an approach for determining the marginal distribution $D_t$, used in (4), of relative wind speeds within the simulated footprint. As discussed in Section 2.4, the relative wind speeds within the ellipse are linked to the vorticity of the track, the size of the event and the maximum relative wind speed. We construct a weighted nonparametric estimate $\tilde{D}_t$ of the distribution function $D_t$. Using a kernel, we weight observed relative wind speeds in $E_t$ conditional on the corresponding values of $W_t$, $\Delta_t$ and $\Omega_t$, from which $\tilde{D}_t$ can be estimated using Monte Carlo methods.

We incorporate information about the physical structure of the footprint in determining the structure of the Gaussian field $\{\tilde{X}_G(s,t) : s \in E_t\}$ using conditional simulation (Diggle and Ribeiro, 2007). We impose three conditions on the simulated fields: that the maximum relative wind speed is simulated at the position determined by $R_t^W$ and $\Theta_t^W$; that the lower limit of $\tilde{D}_t$ occurs on the outer perimeter of the ellipse; and that the lower limit of $\tilde{D}_t$ occurs everywhere in the region corresponding to the local minimum of wind speeds (see Section 2.4) near the storm centre when a footprint is simulated in the vicinity of this region. The first and second conditions create a two-dimensional pseudo-Brownian bridge between the position of maximum and the perimeter of the footprint. To impose the third condition, we specify a second ellipse centred at the local minimum with random dimensions for size; specifically we fix the maximum length of the semi-major and semi-minor axes of this ellipse to be 40 and 35 units of grid-cell distance respectively, with random perturbations modelled as Exp(0.05) random variables. These values were found to replicate well the average behaviour of wind speeds in the region at which the local minimum occurs.

After simulation of $\{\tilde{X}_G(s,t) : s \in E_t\}$, we transform this field to obtain a field of relative wind speeds conditional on the characteristics of the footprint such that

$$\tilde{X}^E(s,t) = \tilde{D}_t^{-1} \left( \Phi[\tilde{X}_G(s,t)] \right).$$

An example of this is shown in Figure 4, in which we have simulated a field of relative wind speeds conditional on the footprint of storm Daria at this particular time step. Our model captures the correlation structure of the field quite well, with the weakest winds occurring on the outer perimeter of the ellipse and the large winds occurring in similar locations to the observed footprint. For this simulated field, the fine structure in terms of the decay from the maximum relative wind speed being more isotropic than the observed field. When viewed in an Eulerian framework, this level of spatial difference is not too important as these footprints move over space with the cyclone track, so blur out this distinction. Having obtained the simulated relative wind fields, we can then transform
these onto the observed margins, such that for each $s \in \mathcal{E}_t$

$$\tilde{X}(s, t) = F_s^{-1}(\tilde{X}^E(s, t)),$$

where $F_s$ denotes the marginal model for cell $s$, as defined in (2). With this formulation, we are assuming the relative wind fields at consecutive time steps are conditionally independent given temporally correlated realisations of $W_t$ and $\alpha_t$. While this assumption gives good results in practice, further investigation may be necessary to assess whether performance could be improved by specifying a spatio-temporal structure in the correlation function $\rho(\cdot)$.

4 Results

We examine the performance of the windstorm model in terms of simulated footprint characteristics and then the wind speeds within the footprint. The joint risk from extreme windstorms at locations in northern England and eastern Germany is then explored by combining the windstorm model presented here with the track model of Sharkey et al. (2017) to produce estimates of joint event probabilities through simulation.

4.1 Validation of footprint model

We explore first whether the characteristics of windstorm footprints are being captured through an assessment of the marginal distributions of the individual components and their dependence structure. QQ plots based on the simulation of footprints using the model described in Section 3.2, applied to 2,944 synthetic storm tracks, the same number as in the observed record, were assessed both for the observed tracks and tracks generated by the model of Sharkey et al. (2017). Both showed similar positive results, so we illustrate only the latter (see Figure 9). They show that the marginal distributions of radius, bearing, proportional area and relative maximum wind speed are being captured well by the model. Figure 10 shows that, when compared with Figure 6, the dependence structure of these components is also consistent with the observations. We can thus conclude that the physical structure of the observed windstorm footprints is replicated sufficiently by the model.

We examined the components of a windstorm influencing the activation and termination models that were outlined in Section 3.3. In both cases, we investigate multiple combinations of covariates and compare model fit using AIC. The best fitting activation model includes functions of vorticity, longitude and latitude. Figure 11 (left panel) shows the estimated smooth function $\beta_i$ associated
Figure 9: QQ plots, with 95% tolerance intervals, comparing the observed and simulated marginal distributions of $R_t^E$, $\Theta_t$, $\Delta_t^{1/2}$ and $W_t$. Simulated values are based on footprints relative to 2,944 synthetic tracks from the model of Sharkey et al. (2017).
with vorticity, which shows that \( \beta_i \) tends to increase approximately linearly as vorticity increases. This has the effect the probability of activation tends to increase as vorticity increases, which reflects our findings from Section 2.4. The best fitting termination model includes functions of \( \Delta_i^{1/2} \) and \( W_i \). Figure 11 (centre and right panels) shows that \( \beta_i \) tends to decrease non-linearly as both \( \Delta_i^{1/2} \) and \( W_i \) increase, which means that the probability of termination also tends to decrease. The effect of \( \Delta_i^{1/2} \) tends to level off at high values, though wide confidence intervals suggest that this effect is highly uncertain. This analysis confirms our belief that weakening events in terms of spatial extent and maximum relative wind speed are consistent with the termination of an active phase of a windstorm. Figure 7 (bottom panel) shows the spatial density of windstorm occurrence in the simulations. We see that the large-scale spatial variation of windstorms from the model reflects that of the observations in Figure 7 (top panel).
Figure 11: The smooth functions $\beta_i$ (with 95% confidence intervals) showing the effect of $\Omega_t$ on the probability of windstorm activation (left) and of $\Delta_1^{1/2}$ (centre) and $W_t$ (right) on the probability of termination.

4.2 Validation of model for wind speeds within a footprint

We check that the model replicates well the physical structure and location of winds relative to the storm centre. For this task we simulate the wind speeds within the footprints generated in Section 4.1 using the wind model of Section 3.4. Figure 12 shows the mean, 95% quantile, 99% quantile and spatial density of simulated winds from all simulated footprints relative to the centre of the storm, quantities that were previously studied for the observed data in Figure 8. The comparison of observed and simulated winds is very good; in particular it replicates the feature that the largest magnitudes tend to be found in the region southwest of the storm centre. The local minimum that occurs near the storm centre as a result of small pressure gradients is also captured. The upper quantiles of the spatial distribution are slightly more dispersed compared to the observed characteristics; however, we are satisfied that the large-scale features have been captured by the model. Figure 12 also shows that high magnitude events can be generated by the model on the outer edges of the domain. These are rare occurrences, but we believe these to be attributed to the detection of spurious events in the feature extraction algorithm. Section 5 discusses possible improvements to the algorithm so that the detection of spurious events is minimised.

By simulating windstorms relative to synthetic tracks, our model also allows us to perform an Eulerian analysis at different locations over the North Atlantic and Europe. We assess how our
Figure 12: The mean, 95% quantile, 99% quantile wind speed and density of within-footprint winds relative to the storm centre (represented by the cross) over a set of simulated windstorms.

model captures the distribution of extreme winds at a number of locations to examine whether our approach succeeds in generating physically realistic synthetic values at these different sites. We compare our simulated values with winds that are contained within the observed footprints at each site. QQ plots for six locations, three on land and three at sea, are shown in Figure 13. The 100-year return level (not shown), estimated from our model for the spatial regions discussed in Section 4.3, compares favourably with estimates from the marginal model in (2), with an average percentage error of 2.4% over all cells. This demonstrates that our model can be used to assess marginal risk at different locations for events beyond the range of the observational record.
Figure 13: QQ plots, with 95% tolerance intervals, at six locations comparing the distribution of wind speeds from the observed and simulated windstorms.

4.3 Joint risk from windstorms

We use our approach to estimate quantities related to joint risk, that is, the probability that multiple locations are affected by extreme wind speeds simultaneously. As our model captures the spatial extent of these meteorological events, we should therefore capture the risk of multiple locations experiencing extreme wind speeds from the same storm. The results are based on 50,000 extratropical cyclone tracks generated using the track model of Sharkey et al. (2017). For each track, a windstorm is simulated using the windstorm model of Section 3. This dataset represents approximately 600 years of data, under the assumption that there are on average 811 windstorms per year as found in the observed record.

Figure 14 (left panel) compares the joint behaviour on common Exp(1) margins between wind speeds in Lancaster and Manchester, two cities in northern England 73km apart. The joint correlation structure is largely captured by the model, which has the added benefit of being able to simulate joint events of magnitudes beyond the range of the observation record. One way of summarising the joint extremal behaviour of a process at arbitrary locations $s_1$ and $s_2$ is to estimate the quantity...
\(\chi(q; s_1, s_2)\) (Coles et al., 1999), defined as

\[
\chi(q; s_1, s_2) = \Pr(X^E(s_2, t) > x^q \mid X^E(s_1, t) > x^q),
\]

where \(x^q\) is the 100\(q\)% quantile of the common \(\text{Exp}(1)\) distribution. Estimates of the quantity \(\chi(q; s_1, s_2)\) obtained from the observed and simulated data are shown for a range of \(x^q\) in Figure 14 (right panel), where \(s_1\) and \(s_2\) are chosen to be sites in Lancaster and Manchester respectively. Estimates from the data and the large simulated sample from the model are obtained as conditional proportions. Figure 14 shows that extremal dependence tends to decrease as the magnitude of the event increases. Here, 95% binomial confidence intervals are used to assess the uncertainty for the observed and the Monte Carlo uncertainty for the simulated data. To obtain these intervals, we use an effective sample size \(n\theta(x^q)\) defined in terms of the sample size \(n\) and a threshold-based extremal index \(\theta(x^q)\) (Ferro and Segers, 2003; Eastoe and Tawn, 2012) to account for temporal dependence. For the model-based estimates, the confidence intervals do not represent the uncertainty due to the model parameter estimation. A fuller assessment of model uncertainty can be obtained using a parametric bootstrap, which would have the effect of widening the model-based confidence intervals. Despite not representing the full uncertainty in the model-based estimates, it is clear that there is no statistically significant difference between data and model-based estimates of \(\chi(q; s_1, s_2)\) here over \(q\) within the range of the observed data. Critically though, this figure also illustrates how our model allows estimation of \(\chi(q; s_1, s_2)\) beyond the range of the observational record, indicating that \(\chi(q; s_1, s_2)\) continues to decay to 0 beyond the observed data.

Estimating \(\chi(q; s_1, s_2)\) at a fixed critical level \(x^q\) at a set of sites \(s_2\) can allow us to explore the spatial extent of extreme events. Figure 15 (top panels) shows \(\chi(q; s_1, s_2)\) calculated across a number of locations in northern England, with \(s_1\) being Lancaster in this instance. We explore two cases where \(x^q\) is chosen to the 90\% quantile and the 10-year return level at each site. In particular, we see that the probability surface decays more steeply as \(|s_2 - s_1|\) increases for the more extreme events. We also see that Liverpool, Manchester and Leeds are more likely than not to experience an event on the 90\% quantile simultaneously to Lancaster; however, this scenario is less likely for an event corresponding to the 10-year return level. Similarly, in Germany, as shown in the bottom panels in Figure 15, the probability of experiencing an extreme windstorm event simultaneously with Berlin decreases as the event becomes more extreme. The spatial extremal dependence is slightly stronger for Berlin than Lancaster, as might be expected given Berlin is more inland on a large land mass. In both regions, there is little evidence of anisotropy in the extremal dependence structure, with perhaps some indication of stronger dependence in the northwest-southwest direction centred at Berlin. The results in Figures 14 and 15 illustrate how the spatial extent of an extreme windstorm
Figure 14: Scatter plot (left) showing observed (red) and simulated (black) wind speeds on an Exp(1) scale in Manchester and Lancaster. Estimates of $\chi(q; s_1, s_2)$ measuring extremal dependence between these locations as a function of $x^q$ (right) using the observed (red) and simulated (black) data, with 95% binomial confidence intervals using an effective sample size. The vertical line denotes the maximum observed wind speed on the Exp(1) scale.

event becomes more localised as the magnitude increases. This implies that in the limit, extreme values at each location tend not to occur simultaneously, which corresponds to the property of asymptotic independence. Models for asymptotic dependence, that is, when $\chi(q; s_1, s_2) \rightarrow c > 0$ as $q \rightarrow 1$ for $s_1 \neq s_2$, lead to extreme values tending to occur simultaneously, are well-established but tend to over-estimate the probability of extreme joint events given that the underlying process is asymptotically independent, i.e., when $c = 0$. Models that capture asymptotic independence are less well-established; see Ledford and Tawn (1996), Heffernan and Tawn (2004), Wadsworth and Tawn (2012) and Winter et al. (2016) for some examples. We have shown that our model captures the property of asymptotic independence over space, while accounting for the complex non-stationarity of the extratropical cyclone system.

5 Discussion

We have presented a novel approach to modelling windstorms in a Lagrangian framework. We described two models; first, for the evolution and development of the footprint relative to the storm track, and second, for the spatial distribution of extreme winds within the footprint. The Lagrangian framework allows us to pool information regarding events over the spatial domain being studied, which allows extrapolation of the characteristics of windstorms over space. The model provides a mechanism for generating synthetic windstorm events, the analysis of which allows im-
Figure 15: Estimates of $\chi(q; s_1, s_2)$ for northern England (top) and eastern Germany (bottom) conditioning on a critical value $x^q$, where $s_1$ is the cell where Lancaster (top) and Berlin (bottom) are located. In the left panels, $x^q$ is taken to be the 90% quantile, while $x^q$ is taken to be the 10-year return level in the right panels. Both regions are of equal size.

proved estimation of joint risk associated with extreme windstorms over Europe.

There are, however, opportunities to improve the performance of the model. While the feature extraction approach introduced in Section 2.3 appears to extract the main features of a windstorm, steps could be taken to improve the robustness of this procedure. Firstly, one could conduct a sensitivity study on the choice of threshold $v$, which controls the level under which the wind speed fields are masked. Ideally, one should choose a high enough value of $v$ such that convective events and non-extreme winds are masked, but small enough so that the localised features of the windstorm aren’t excluded. “Sting jets”, meteorological phenomena associated with rapidly developing storms, can produce damaging winds on very small spatial scales (Baker, 2009; Hewson and Neu, 2015) and therefore it is important that the extraction algorithm should not exclude such features.
The feature extraction algorithm could also be improved to minimise the detection of spurious footprints that may not be generated by the extratropical cyclone. We see examples of this through high magnitude events occurring on the outer edges of the domains in Figure 8. After the spatio-temporal filtering step, our algorithm selects the largest cluster in size to define the footprint. One could alternatively define a score function that incorporates multiple criteria in a detection strategy, which could be motivated by physical intuition regarding the structure of a windstorm. For example, the score function could give a higher weight to clusters with a bearing relative to the storm centre closest to southwest, where most footprints seem to occur. To induce some smoothness between consecutive time steps, one could assign a higher weight to a cluster such that the Euclidean distance between the position of this cluster and the selected cluster at the previous time step is minimised. Exploration of different score functions could add valuable improvements to our feature extraction algorithm, and ultimately, the model.

Our conditional kernel strategy for modelling \( \{Z_t\} \) appears to perform well. However, we could alternatively model the extremal behaviour of \( W_t \) using the approaches described in Winter and Tawn (2017) and Sharkey et al. (2017), whereby a GPD model is defined above a high threshold and the extremal temporal dependence structure is modelled using a kth order extremal Markov process. This approach stems from the conditional multivariate extreme value methodology of Heffernan and Tawn (2004). We did not implement this approach as part of this study due to the additional complexity involved and that extrapolation occurred naturally through the features of our model. In addition, the observed values of \( W_t \) correspond to upper tail values of \( X_i \), the distribution of wind speeds in cell \( i \). Because we have a large dataset of observations in the upper tail, we felt that standard statistical modelling approaches were sufficient in this case. However, investigation of the benefits of imposing an extremal temporal dependence structure on the upper tail of \( W_t \) represents an interesting avenue of future research.

The potential future risk associated with extreme windstorms is of pressing concern in addition to how their characteristics might be affected by a changing climate. For the North Atlantic, previous studies have indicated the winter storm track will potentially increase over the UK and northern Europe but eliciting this signal is difficult (Zappa et al., 2013). This uncertainty, in combination with the low probability of a cyclone producing extreme winds makes estimating future windstorm risk very challenging. The windstorm model presented here and the track model of Sharkey et al. (2017) together provide a new tool that can be applied to future climate simulations to potentially
provide improved estimates of such risk.

Acknowledgements

The authors gratefully acknowledge the support of the EPSRC funded EP/H023151/1 STOR-i Centre for Doctoral Training, the Met Office and EDF Energy. We extend our thanks to Hugo Winter for helpful discussions and support. We thank Kevin Hodges for the storm track data. Simon Brown was supported by and Paul Sharkey partially supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101).

Appendix

A Conditional kernel density estimation

Consider an arbitrary $d$-dimensional random vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_d)$, which is observed $n$ times $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(n)}$. As a way of estimating $f(\mathbf{z})$, the joint probability density of $\mathbf{Z}$, we define the multivariate kernel density estimator as

$$
\hat{f}(\mathbf{z}) = \frac{1}{n} \sum_{i=1}^{n} K_H \left( \mathbf{z} - \mathbf{z}^{(i)} \right),
$$

where $K_H$ is the kernel function and $H$ denotes the bandwidth matrix which is symmetric and positive-definite. For our purposes, we choose $K_H$ to be the multivariate Gaussian density function

$$
K_H(\mathbf{z}) = (2\pi)^{-d/2}|H|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{z}^T H^{-1} \mathbf{z} \right\}
$$

and the bandwidth matrix $H$ chosen to be proportional to the rule-of-thumb selection of Scott (1992). The bandwidth matrix $H$ can be chosen to be diagonal or oriented. To simulate from the kernel density, we first sample uniformly a tuple $\mathbf{z}^{(i)}$, where $i \in \{1, \ldots, n\}$. We then simulate a vector $\tilde{\mathbf{z}}$, say, such that $\tilde{\mathbf{z}} \sim \text{MVN}(\mathbf{z}^{(i)}, H)$.

Let $\mathbf{Z}$ be decomposed such that $\mathbf{Z} = (\mathbf{Z}_m, \mathbf{Z}_m)$. Consider the case when values $\mathbf{Z}_{-m} = \mathbf{z}_{-m}$ have been observed and we wish to estimate the conditional density of $\mathbf{Z}_m$ given these values. We can then define the conditional kernel density estimator as

$$
\hat{f}(\mathbf{z}_m|\mathbf{z}_{-m}) = \sum_{i=1}^{n} w_i(\mathbf{z}_{-m}) K_H \left( \mathbf{z}_m - \mathbf{z}_m^{(i)} \bigg| \mathbf{z}_{-m} - \mathbf{z}_{-m}^{(i)} \right),
$$
where

\[ w_i(z_{-m}) = \frac{K_H(z_{-m} - z^{(i)}_m)}{\sum_{j=1}^{n} K_H(z_{-m} - z^{(j)}_m)}, \]

where \( K_H(\cdot) \) is the multivariate Gaussian kernel function and \( K_H(\cdot | \cdot) \) is the conditional Gaussian kernel function with bandwidth matrix \( H \) as defined in equation (7). Let \( H \) be partitioned such that

\[
H = \begin{bmatrix}
H_{m,m} & H_{m,-m} \\
H_{-m,m} & H_{-m,-m}
\end{bmatrix}.
\]

Conditional on having observed \( z_{-m} \), we choose a tuple \( z^{(i)} \) with probability \( w_i(z_{-m}) \). Then we simulate

\[ Z_m|(Z_{-m} = z_{-m}) \sim \mathcal{N}(\bar{\mu}, \bar{\Sigma}), \tag{9} \]

where \( \bar{\mu} = z^{(i)}_m + H_{m,-m}H^{-1}_{-m,-m}(z_{-m} - z^{(i)}_{-m}) \) and \( \bar{\Sigma} = H_{m,m} - H_{m,-m}H^{-1}_{-m,-m}H_{-m,m} \).

References


36


