

## Stochastic Simulation Model for Tropical Cyclone Tracks, with Special Emphasis on Landfall Behavior

Björn Kriesche · Helga Weindl ·  
Anselm Smolka · Volker Schmidt

Received: date / Accepted: date

**Abstract** We consider a spatial stochastic model for the simulation of tropical cyclone tracks, which has recently been introduced. Cyclone tracks are represented as labeled polygonal lines, which are described by the movement directions, translational speeds, and wind speeds of the cyclones in regular six-hour intervals. In the present paper, we compare return levels for wind speeds of historically observed cyclone tracks with those generated by the simulator, where a mismatch is shown for most of the considered coastal regions. To adjust this discrepancy, we develop a stochastic algorithm for acceptance and rejection of simulated cyclone tracks with landfall. It is based on the fact that the locations, translational speeds, and wind speeds of cyclones at landfall constitute three-dimensional Poisson point processes, which are a basic model type in stochastic geometry. Due to that, a well-known thinning property of Poisson processes can be applied. This means that to each simulated cyclone an acceptance probability is assigned, which is higher for cyclones with suitable landfall characteristics and lower for implausible ones. More intuitively, the algorithm comprises the simulation of a more comprehensive cyclone event set than needed and the random selection of those tracks that best match historical observations at landfall. A particular advantage of our algorithm is its applicability to multiple landfalls, i.e. to cyclones that successively make landfall at two geographically distinct coastlines, which is the most relevant case in applications. It turns out that the extended simulator provides a much better accordance between landfall characteristics of historical and simulated cyclone tracks.

**Keywords** Tropical cyclone · Landfall characteristics · Stochastic model · Acceptance-rejection · Poisson point process · Hazard assessment

---

Björn Kriesche · Volker Schmidt  
Ulm University, Institute of Stochastics, 89069 Ulm, Germany  
E-mail: bjoern.kriesche@uni-ulm.de

Helga Weindl · Anselm Smolka  
Munich Reinsurance Company, 80791 Munich, Germany

## 1 Introduction

### 1.1 Motivation

Natural disasters caused by tropical cyclones (TCs) pose a huge threat to both human life and property. As population and property development in areas susceptible to tropical cyclones grows, so does the risk for insurance companies. Hence, it is of increasing importance for insurers to assess hazards constituted by cyclones as precisely as possible. Since reliable cyclone data only exist for 50-100 years, stochastic modeling and simulation turned out to be a useful approach, in particular for estimating the impact of TCs having a very low occurrence probability (e.g. once in 10,000 years).

### 1.2 Model types and state of research

In the context of simulation-based hazard assessment, we typically distinguish between local models and basin-wide cyclone track models. Local methods estimate future TC characteristics using the corresponding historical events only, whereas track models involve the simulation of entire cyclone tracks from genesis to lysis. A more detailed description of both model types as well as a comparison of two example models for the prediction of landfall rates is given in Hall and Jewson (2008). One result of that analysis is that local models are considered to be more accurate and that they produce better landfall estimates than track models in regions with high TC activity. For a broad overview of local models for various aspects of TC hazard we refer to Elsner and Jagger (2013). However, insurers are sometimes interested in questions that can not be answered sufficiently well by local methods. Typical questions of this kind include: How much loss do we have to expect from an extreme TC in the entire ocean basin? How far does a TC move inland after landfall? What is the probability of a loss caused by a TC with landfall at two geographically distinct coastal regions? The consideration of complete TC tracks is necessary to deal with such problems. Furthermore, the comparison presented in Hall and Jewson (2008) shows that landfall estimates based on track simulation are more precise and produce better results in historically inactive regions than local methods. This advantage of basin-wide methods is particularly important, since many insurance-relevant areas are located in regions with only few historical data (e.g. the metropolitan areas of New York and Boston).

Therefore, in this paper we concentrate on basin-wide cyclone track models. Various simulation models for TCs have been introduced for both the North Atlantic (NA), see Vickery et al (2000); Emanuel et al (2006); Hall and Jewson (2007); Hallegatte (2007); Rumpf et al (2009), and the western North Pacific (WNP), see e.g. Rumpf et al (2007); Yonekura and Hall (2011). In the literature, however, only minor attention is paid to landfall behavior in the context of basin-wide models. In Vickery et al (2000), means of five landfall characteristics of historical and simulated cyclone tracks at a sequence of mileposts are compared for validation. In other papers, only

landfall rates at some regions of interest (e.g. Rumpf et al, 2007, 2009) or along an entire coastline (Hall and Jewson, 2007) are compared. In addition, exceedance probabilities for cyclone wind speeds at landfall have been considered (Emanuel et al, 2006; Hallegatte, 2007). To our knowledge, however, no thorough analysis of the (joint) distribution of landfall characteristics (landfall locations, wind speeds, translational speeds, number of landfalls) has been conducted yet. This and the high influence of landfall characteristics on hazard analysis motivate an increased focus on landfall behavior of simulated TC tracks.

### 1.3 Overview

The present paper deals with the landfall behavior of cyclone tracks that were generated by the stochastic simulator as described in Rumpf et al (2007, 2009). At first, the available data and the simulation model are recalled in Sections 1.4 and 2.1-2.3. By comparing estimated return levels from historical and simulated cyclone tracks, we motivate a more detailed consideration of landfall characteristics (see Section 2.4). Thus, in Section 3, a stochastic acceptance-rejection algorithm for simulated cyclone tracks is proposed, which produces tracks in such a way that the number of landfalls and further landfall characteristics are statistically comparable to historical observations. This involves an approximation of coastlines to identify landfalls, the modeling of landfall characteristics as spatial Poisson processes, and the description of acceptance-rejection itself based on a famous invariance property of Poisson processes. Single and multiple landfalls are handled separately in this context. Section 4 concerns the validation of the extended simulation model proposed in this paper, where several types of illustrations are depicted to verify the accordance of distributions for historical and simulated landfall characteristics. A summary in Section 5 concludes the paper.

### 1.4 Data

We focus on stochastic cyclone simulation in the NA and WNP, as in both ocean basins, numerous TCs with landfall occur every year. Furthermore, these basins contain some of the most endangered coastal areas worldwide and provide comprehensive historical cyclone data. In the NA, we use cyclone records from the North Atlantic hurricane database (HURDAT), which is compiled by the National Oceanic and Atmospheric Administration (NOAA) of the United States. HURDAT contains historical cyclone data from 1851 to 2010, where measurements before 1900, however, are considered to be of doubtful reliability and are therefore excluded in this paper. Cyclone records for the WNP are obtained from the International Best Track Archive for Climate Stewardship (IBTrACS), which is compiled by the NOAA, too. IBTrACS records provide reliable historical cyclone data for the 64 years between 1945 and 2008. For both ocean basins, the time span for which cyclone records are available is denoted by  $T_{hist}$ .

The considered data for each cyclone track includes the time of observation, the geographical coordinates, and the maximum wind speed in regular six-hour intervals. For simplicity, we consider the tracks as polygonal lines by connecting the measured geographical locations. In addition, translational speeds and movement directions between successive cyclone measurement points are easily computable. Each cyclone track can thus be represented by its initial location, movement direction, translational speed, and wind speed and by the consecutive changes of those parameters in regular six-hour intervals. To provide more homogeneity, tracks are divided into six cyclone classes for each ocean basin. Cyclone simulation and acceptance-rejection will be performed for each class separately.

## 2 Basic components of the stochastic simulation model

We consider the stochastic simulation model for TC tracks introduced in Rumpf et al (2007, 2009). The aim of the present paper is the advancement of the simulation output concerning landfall behavior. However, since the model has a relatively high level of complexity, it might be convenient for the reader to get a brief overview of the simulation procedure first. Therefore, this section recalls the particular components of the simulation model and makes a comparison between return levels estimated from historical and simulated TC tracks.

### 2.1 Points of genesis

The stochastic simulation model is built incrementally. The first step, naturally, is the modeling of points of genesis. Historical points of genesis form irregular point patterns, which suggests the use of inhomogeneous Poisson point processes, a basic model type from stochastic geometry. The Poisson process is considered to be a model for complete spatial randomness, i.e. the (random) points are located independently of each other and the number of points in any arbitrary observation window is Poisson distributed, see Illian et al (2008), p. 118. This model choice is justified by the following observations:

1. Meteorology does not provide evidence for interaction between TC genesis points. Hence, points of genesis can be considered to be independent of each other.
2. A Pearson-Fisher-Goodness-of-Fit test does not reject the hypothesis that the annual number of historical cyclone geneses is Poisson distributed, see Rumpf et al (2007).

The distribution of an inhomogeneous Poisson process is completely determined by its nonnegative intensity function  $\lambda : \mathbb{R}^2 \rightarrow [0, \infty)$ . When modeling a set of genesis points by a Poisson process, the corresponding intensity function is estimated from historical points of genesis. For that purpose, a generalized nearest-neighbor estimator, see Silverman (1986), p. 97, is applied. A synthetic set of TC genesis points can be simulated by generating a realization of the fitted Poisson point process model. Note that for the simulation of TCs representing a time span  $T \neq T_{hist}$ , the estimated intensity function has to be scaled properly by  $T/T_{hist}$ .

## 2.2 Track propagation and wind speeds

As described in Section 1.4, a TC track is modeled as a polygonal line, with each line segment representing the movement of the cyclone over a six-hour time span. Thus, it suffices to consider a cyclone's movement directions and translational speeds. Assuming these values to be constant over six hours, all consecutive cyclone locations can be calculated. In addition, the maximum wind speeds attained at these locations are taken into consideration to allow for a meaningful hazard assessment. Following this approach, the direction of movement  $X_i$ , the translational speed  $Y_i$ , and the wind speed  $Z_i$  after the  $i$ -th cyclone segment are given by

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + \sum_{j=1}^i \begin{pmatrix} \Delta X_j \\ \Delta Y_j \\ \Delta Z_j \end{pmatrix}, \quad (1)$$

with  $X_0$ ,  $Y_0$ , and  $Z_0$  denoting the TC's initial direction, translational speed, and wind speed, and  $\Delta X_j$ ,  $\Delta Y_j$ , and  $\Delta Z_j$  describing the changes of these values in regular six-hour intervals. Connecting each two successive locations gives the polygonal line that models the TC track, and at each location the maximum wind speed is provided.

Since the model is of a stochastic nature, all TC characteristics are considered to be random variables. The distributions of  $X_0$ ,  $Y_0$ ,  $Z_0$ ,  $\Delta X_j$ ,  $\Delta Y_j$ , and  $\Delta Z_j$  are supposed to depend on the cyclone's current position. The change in wind speed  $\Delta Z_j$  is additionally assumed to depend on the wind speed  $Z_{j-1}$  at the previous location. Realizations of the random variables are obtained through sampling from nearby historical cyclone observations. Essentially, the model follows the same basic assumptions as for example the approaches introduced in Emanuel et al (2006) and Hall and Jewson (2007), namely that TCs occurring in the same regions behave similarly. For further, more detailed information on track propagation, particularly on the simulation of  $X_0$ ,  $Y_0$ ,  $Z_0$ ,  $\Delta X_j$ ,  $\Delta Y_j$ , and  $\Delta Z_j$ , we refer to Rumpf et al (2007).

## 2.3 Track termination

For the termination of simulated cyclone tracks, Rumpf et al (2007) proposed a random mechanism. After the generation of each cyclone segment, a Bernoulli experiment is performed to decide whether the track is terminated or not. The success probability (success means 'the track terminates') is calculated as the maximum of a location-dependent termination probability  $p_x$  and a wind-speed-dependent probability  $p_z$ . On the one hand, this is based on the observation that cyclones located close to each other behave similarly. Thus, simulated tracks should be likely to terminate in regions where historical ones do. On the other hand, it is a meteorological fact that TCs with lower wind speeds are more likely to terminate than those with higher ones, which motivates the consideration of  $p_z$ .

Both termination probabilities are estimated from historical observations. The probability  $p_x$  is determined as the relative frequency of termination points among historical cyclone points in a neighborhood of the current location, whereas  $p_z$  is obtained

by fitting an exponential function to historical termination probabilities.

Combining the components from Sections 2.1 - 2.3 gives a complete simulation model for TC tracks. It has been implemented using classes and methods from the GeoStoch library, which is a Java-based software developed at Ulm University, see Mayer et al (2004).

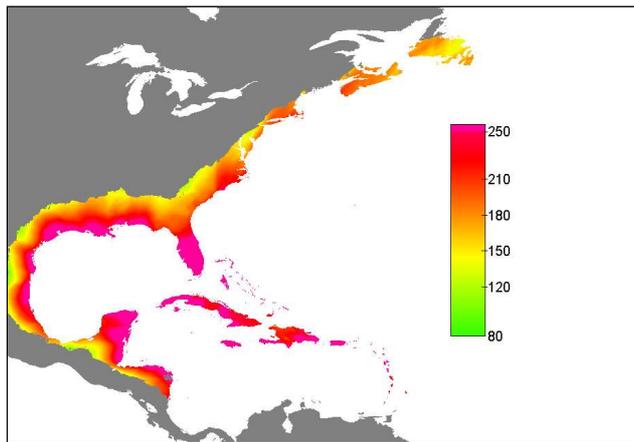
## 2.4 Comparison of return levels

To evaluate the quality of the simulation output, we estimate return levels for historical and simulated cyclone event sets representing the same time span  $T_{hist}$ . A comparison gives information about how well historical and synthetic cyclone characteristics coincide. First, for each TC and each location of interest inside the observation window, the maximum wind impact of the cyclone at the location is computed. The underlying cyclone shape model and the resulting computation algorithm are of minor importance here and can be looked up in Rumpf et al (2009).

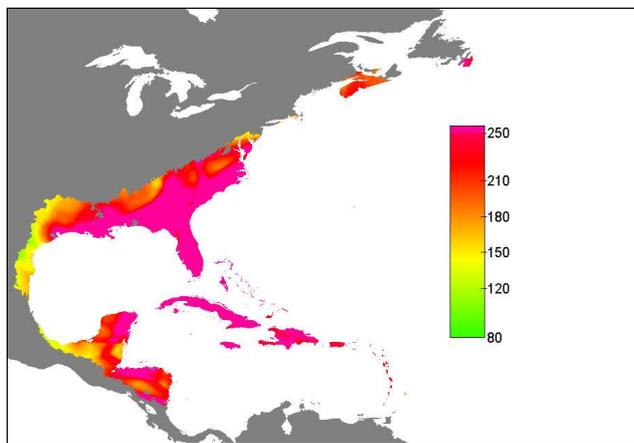
For a fixed return period, a hazard map defines a function, which assigns a return level to each location of interest inside the observation window. Intuitively, a return level can be interpreted as the maximum wind impact (in km/h) that is expected to be exceeded once during the corresponding return period. Return levels can be determined by applying a peak-over-threshold (POT) method from extreme value theory to the wind impacts of a (historical or simulated) cyclone event set. A particular advantage of POT methods is the possibility to calculate plausible return levels for return periods that exceed the time span of the underlying event set. The POT method used in this section is similar to that introduced in Jagger and Elsner (2006) and is therefore not described here.

In Fig. 1 and 2, hazard maps for a return period of 100 years are compared. In both the NA and WNP, a mismatch between the overland behavior of simulated and historically observed cyclone tracks is shown. In Fig. 1, return levels of synthetic tracks are far too high in the southeast of the US and slightly too high at the islands of the Caribbean Sea. This indicates that there are too many TCs making landfall in these regions and that their wind and translational speeds are inordinately high. At the western Gulf Coast, however, low return levels indicate a lack of intense simulated cyclone tracks. In the WNP (see Fig. 2), discrepancies are even more obvious. Return levels of simulated tracks are excessively high and reach too far inland at the northern Chinese coast, Korea, and Japan. At the southern Chinese and Vietnamese coast, the opposite is observed. Here, the large differences are an indication for an insufficient simulation of TC overland characteristics, too. Note that in the previous version of the simulator the Philippines were not considered as a location of interest, and therefore no return levels are given for this region in Fig. 2(b).

To figure out reasons for this mismatches, we investigated characteristics of simulated cyclone tracks. Our results correspond well to those of Hall and Jewson (2008).



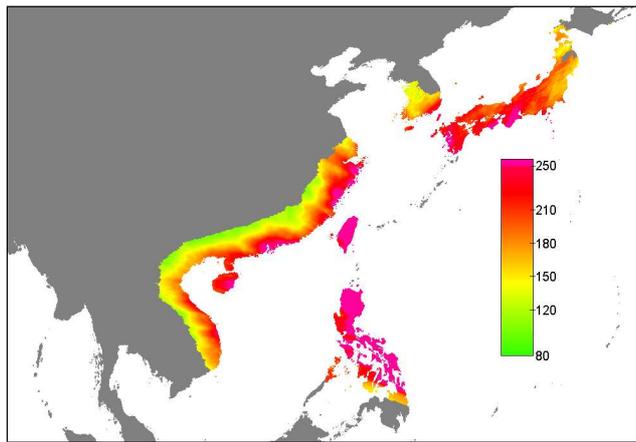
(a) historical



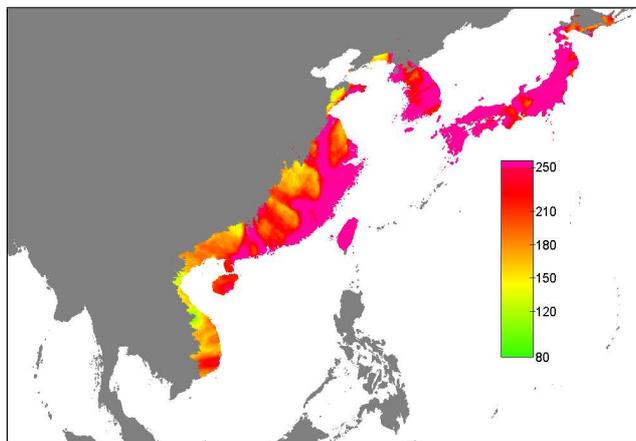
(b) simulated

Fig. 1: Hazard maps for return period 100 years in the NA. Return levels are indicated by colors in km/h.

TC characteristics derived from the stochastic simulator as described in Rumpf et al (2007, 2009) have certain biases, i.e. systematic errors occur. In particular, some translational speeds of cyclones are unrealistically high, whereas maximum wind speeds seem to be slightly too low. Additionally, differences in landfall locations of historical and simulated tracks are observed at most coastlines. These errors could be caused by the incremental nature of the simulation approach proposed in Rumpf et al (2007, 2009) or by other modeling components. On the other hand, they could be due to data inconsistencies. Anyhow, the systematic errors have to be corrected



(a) historical



(b) simulated

Fig. 2: Hazard maps for return period 100 years in the WNP. Return levels are indicated by colors in km/h.

to allow for a meaningful hazard assessment. Therefore, we propose to add a local acceptance-rejection method to the stochastic simulator in order to adjust the landfall locations, translational speeds, and wind speeds of simulated cyclone tracks to historical observations. This will remove the systematic error in simulated landfall characteristics and there is reason to believe that more accurate landfall characteristics also lead to an improved overland behavior, particularly in coastal regions.

### 3 Improvements of the landfall behavior by stochastic acceptance-rejection

In this section, we introduce a stochastic method that allows to generate synthetic TC tracks, whose landfall characteristics match historical observations in a statistical sense. However, this method is not a direct modification of the existing simulation model we described in Section 2. It rather extends the model by adding an acceptance-rejection component, which is performed at the end of the simulation procedure. The basic idea of this approach is to simulate a cyclone event set representing a much larger time span than needed and to select those tracks at random that best match historical cyclones concerning landfall behavior. Nevertheless, the resulting synthetic tracks do not simply copy historical landfall characteristics, but still feature some random variation, which is a desired property.

#### 3.1 Approximation of coastlines

To adjust the landfall characteristics of TCs, we first have to specify when a cyclone is considered to make landfall. Naturally, this is the case if its track crosses a coastline and moves over land. Since most coastlines look rather rough, however, checking this condition is computationally intensive. A simplification is proposed in Hall and Jewson (2007), where the North American and Central American Atlantic coastline is approximated by 39 connected line segments. Note that this approximation is not part of the simulation model but serves for the validation of landfall rates. We use this approach here, with some minor modifications, as a basic component of our acceptance-rejection algorithm. In each considered ocean basin, we approximate four important coastal areas by polygonal lines. Approximated coastlines are adapted to the historical landfall regions of different cyclone classes. Therefore, some coastlines may overlap at several line segments. All approximated coastlines are displayed in Fig. 3, the covered regions are described in Table 1.

The approximation of coastal areas by polygonal lines simplifies the identification of landfalls. We consider a TC to make landfall at any coastline if its track intersects one of the approximating line segments. Each cyclone track can, depending on its corresponding class, make landfall at zero, one, or two coastlines. The second case is denoted as single landfall, the latter one as multiple landfall.

#### 3.2 Landfall vectors

We consider three important characteristics of TC tracks at landfall, which are regarded as the most essential variables for a meaningful hazard assessment: the location, the translational speed, and the maximum wind speed. The approximation of coastal areas by polygonal lines as described above allows for a quick computation of these characteristics. Given a TC makes landfall, its landfall location  $X$  is determined as the cumulative distance along the coastline between the coastline's starting point and the point of intersection with the cyclone track. A particular advantage of this approach is that the (two-dimensional) landfall location of a TC can be described by one

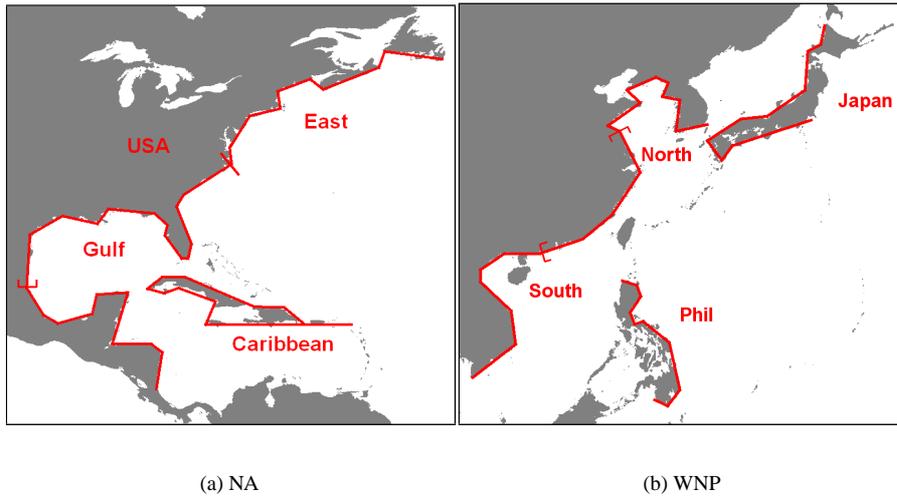


Fig. 3: Coastlines approximated by polygonal lines

Table 1: Regions covered by approximated coastlines

coastline	covered region
'USA'	Canadian and US-American Atlantic coast, small part of northern Mexico
'East'	Canadian and northern US-American Atlantic coast
'Gulf'	Mexican and US-American Gulf coast, southern US-American Atlantic coast
'Caribbean'	Greater Antilles
'South'	Vietnamese and Chinese coast up to Lianyungang
'North'	Chinese coast from Hong Kong northwards, parts of Korean coast
'Japan'	large parts of the Japanese coastline
'Phil'	Philippine east coast

single value. Due to the fact that the movement of a cyclone is assumed to be constant in regular six-hour intervals, the translational speed  $Y$  is set to the translational speed that is attained during the time interval in which landfall occurs. The wind speed  $Z$  at landfall is linearly interpolated between the wind speeds of the last cyclone measurement point over sea and the first point over land. Since all cyclone attributes are assumed to be of a random nature, see Section 2.2, the landfall characteristics  $X$ ,  $Y$ , and  $Z$  form random variables, too. They are summarized to a (nonnegative) random vector  $S = (X, Y, Z)^\top$ , which is called the TC's (random) landfall vector. There are of course further landfall characteristics having influence on the hazard constituted by TC tracks (e.g. track bearing), which are, however, not considered here to avoid the model from getting too complex and too difficult to handle.

According to results from Section 2.4, landfall vectors of simulated TC tracks are assumed to differ significantly from historical ones. In order to create a possibility of comparison, we show that it is plausible to assume that landfall vectors of cyclone event sets constitute (three-dimensional) Poisson processes. This property is traced

back to the modeling of points of genesis (see Section 2.1), where Poisson processes are applied, too. In particular, we use the fact that by our model assumptions the (random) number  $N$  of simulated tracks is Poisson distributed. We first consider cyclone classes (see Section 1.4), whose tracks do not make multiple landfalls; only landfall at one coast line, denoted by  $C$  here, is possible. Consequently,  $N$  can be written as the sum of the number  $N_C$  of cyclones with landfall and the number  $N_{ntf}$  of cyclones without landfall. A fundamental invariance property of the Poisson distribution gives that both  $N_C$  and  $N_{ntf}$  again have a Poisson distribution. Furthermore, according to Section 2, TC tracks are simulated independently of each other. Thus, we get that the number of landfall vectors with values in a given subset of the three-dimensional Euclidean space  $\mathbb{R}^3$  is Poisson distributed, and the numbers of landfall vectors with values in disjoint subsets of  $\mathbb{R}^3$  are independent. This implies that by the set of landfall vectors a Poisson process is given, see Kingman (1993), Chapter 2.5.

The statement described above can easily be generalized to cyclone classes where multiple landfalls at two coastlines  $C_1$  and  $C_2$  occur. Note that classes and coastlines are designed in such a way that  $C_1$  and  $C_2$  do not overlap and that, if a multiple landfall occurs,  $C_1$  is always hit first. Furthermore, one has to distinguish clearly between landfall vectors computed at  $C_1$  and landfall vectors computed at  $C_2$ . To emphasize this, they are denoted as  $C_1$ -landfall vectors or  $C_2$ -landfall vectors here. In the case of multiple landfalls, tracks from a simulated cyclone event set can be divided into four groups: those that make landfall at both  $C_1$  and  $C_2$ , those that make landfall only at  $C_1$ , those that make landfall only at  $C_2$ , and cyclones without landfall at  $C_1$  and  $C_2$ . Analogously to the case of single landfalls, we can show that the following classes of landfall vectors can be assumed to form (three-dimensional) Poisson processes:

- 1)  $C_1$ -landfall vectors of all TCs that make landfall only at  $C_1$ ,
- 2)  $C_2$ -landfall vectors of all TCs that make landfall only at  $C_2$ ,
- 3)  $C_2$ -landfall vectors of all TCs that make landfall at both  $C_1$  and  $C_2$ ,
- 4)  $C_1$ -landfall vectors of all TCs that make landfall at  $C_1$  (regardless of whether a second landfall occurs or not).

Note that all statements made in this section remain valid if (one-dimensional) landfall locations are considered instead of (three-dimensional) landfall vectors. In that case, sets of landfall locations at coastlines as described above form (one-dimensional) Poisson processes.

### 3.3 Landfall intensity functions

The distribution of the points of a (three-dimensional) Poisson process is completely determined by its corresponding intensity function  $\lambda : \mathbb{R}^3 \rightarrow [0, \infty)$ . Thus, the landfall behavior of simulated TCs (i.e. the distribution of landfall vectors) is characterized by the intensity functions of the Poisson processes that model landfall vectors (see Section 3.2). To emphasize this, these functions are referred to as landfall intensity functions (LFIFs) here. However, LFIFs are not directly given by the simulation model. They have to be estimated from simulated landfall vectors. In default

of a suitable parametrized family of three-dimensional density functions, the use of a non-parametric estimation technique is preferred. To be more precise, we apply a multivariate kernel density estimator proposed in Scott (1992), p. 153. For a given random sample of landfall vectors  $S_1, \dots, S_N$  that is considered to be a Poisson process, an estimator  $\hat{\lambda}(\cdot)$  for the LFIF  $\lambda(\cdot)$  is given by

$$\hat{\lambda}(x) = \frac{1}{|H|} \sum_{i=1}^N K(H^{-1}(x - S_i)) \quad \text{for all } x \in [0, \infty)^3 \quad (2)$$

with a kernel function  $K(\cdot)$  and a bandwidth matrix  $H$ , where  $|H|$  and  $H^{-1}$  denote the determinant and the inverse matrix of  $H$ , respectively. As kernel function we choose the probability density of the three-dimensional standard normal distribution (standard normal kernel), but other kernels would be possible, too. As a matter of fact, changing the type of the kernel function has only minor influence on the shape of  $\hat{\lambda}(\cdot)$ . A more difficult issue is the choice of an appropriate bandwidth matrix  $H$ , since varying  $H$  has an significant effect on the smoothness and shape of the estimated LFIF. To simplify the computations, we assume that  $H = \text{diag}(h_1, h_2, h_3)$ , i.e.  $H$  is a diagonal matrix. But common approaches to choose a suitable diagonal bandwidth matrix still turned out to be computationally too intensive. Therefore, a reference rule proposed in Scott (1992), p. 152 is applied, which is computationally efficient and provides reasonable bandwidth values  $h_1$ ,  $h_2$ , and  $h_3$ .

Note that the kernel density estimator given in (2) is applicable to compute LFIFs from both historically observed and simulated landfall vectors. Since historical and simulated cyclone tracks should feature the same landfall behavior, we thereby get reference functions for the final simulation output. Altogether, LFIFs allow to describe the landfall behavior (i.e. the distribution of landfall vectors) of all TCs from an entire event set by one single function and, more importantly, provide a tool to make a comparison between historical and simulated landfall vectors. Since LFIFs depend on three variables, a visual comparison, however, is not possible.

### 3.4 Acceptance-rejection for single landfalls

Based on LFIFs introduced above, we propose an acceptance-rejection method to obtain TC tracks with adjusted landfall characteristics. The method allows for the generation of cyclone event sets representing any arbitrary time span  $T$ . First, we consider a coastline  $C$  where only single landfalls occur. The LFIF computed from historical cyclone tracks is denoted by  $\lambda_C^{\text{hist}}(\cdot)$ , which is, however, based on observations made over  $T_{\text{hist}}$  years. If  $T \neq T_{\text{hist}}$ , it thus is not a suitable reference function for landfall behavior of simulated tracks. Instead, the Poisson process that models simulated landfall vectors at  $C$  should have a LFIF given by  $\frac{T}{T_{\text{hist}}} \lambda_C^{\text{hist}}(\cdot)$ . If we simply generated TC tracks representing  $T$  years with the simulator as described in Section 2, their estimated LFIF would be very likely to differ considerably from  $\frac{T}{T_{\text{hist}}} \lambda_C^{\text{hist}}(\cdot)$ . To solve this problem, we first generate a synthetic cyclone event set representing  $\alpha T$  years. The scaling parameter  $\alpha > 1$  has to be chosen large enough, such that for the

LFIF  $\lambda_C^{sim}(\cdot)$  of simulated TC tracks it holds that

$$\lambda_C^{sim}(x) \geq \frac{T}{T_{hist}} \lambda_C^{hist}(x) \quad \text{for all } x \in [0, \infty)^3. \quad (3)$$

Since LFIFs can not be visualized and a pointwise comparison would be too time-consuming, condition (3) is rather difficult to check. To be sure,  $\alpha$  should be chosen very large; values between 50 and 200, depending on the cyclone class, turned out to be sufficient. Consider the random landfall vectors  $S_1, \dots, S_{N_C}$  of those  $N_C$  (simulated) cyclone tracks that make landfall at  $C$  (i.e. the Poisson process with LFIF  $\lambda_C^{sim}(\cdot)$ ). Let furthermore  $U_1, U_2, \dots$  be a sequence of independent random variables that are uniformly distributed on  $[0, 1]$  and independent of  $S_1, \dots, S_{N_C}$ . Then, the independent thinning property of Poisson processes, see e.g. Møller and Waagepetersen (2004), Chapter 3.2.2, yields that by the subset of landfall vectors

$$\left\{ S_i : 1 \leq i \leq N_C, U_i < \frac{\frac{T}{T_{hist}} \lambda_C^{hist}(S_i)}{\lambda_C^{sim}(S_i)} \right\} \quad (4)$$

a (three-dimensional) Poisson process with the desired LFIF  $\frac{T}{T_{hist}} \lambda_C^{hist}(\cdot)$  is given.

A more intuitive interpretation of the procedure described above is given as follows. First, a cyclone event set is generated, which is much more comprehensive than the intended result. In particular, by (3) we require that statistically speaking, any arbitrary landfall vector is observed more often in the simulated event set than among historical tracks. Then, to each simulated TC that makes landfall at  $C$  with landfall vector  $s_i$ , an acceptance probability  $p(s_i)$  is assigned. To be more precise, we put  $p(s_i) = \frac{T}{T_{hist}} \lambda_C^{hist}(s_i) / \lambda_C^{sim}(s_i)$ , i.e. the acceptance probability  $p(s_i)$  is lower if we have considerably more simulated tracks with landfall vectors similar to  $s_i$  than historical ones, and higher if the (scaled) numbers of simulated and historical tracks with landfall vectors in a vicinity of  $s_i$  differ only slightly. Condition (3) ensures that by  $p(s_i)$  indeed a probability is given. After that, a Bernoulli experiment with success probability  $p(s_i)$  is performed, to decide whether the synthetic cyclone is accepted or rejected. Result of this procedure is a set of TC tracks that make landfall at  $C$  and whose landfall behavior is statistically equal to that of historical data.

To obtain a complete synthetic cyclone event set representing a time span  $T$ , the tracks gained through acceptance-rejection procedure described above have to be merged with tracks that do not make landfall at  $C$ . Since  $\alpha$  is chosen large enough, a sufficiently large number of such cyclone tracks is contained in the simulated event set for the time span  $\alpha T$ , and we propose the following selection procedure. An integer  $n_{nlf}$  is sampled from a Poisson distributed random variable  $N_{nlf}$  with expectation  $\frac{T}{T_{hist}} n_{nlf}^{hist}$ , where  $n_{nlf}^{hist}$  denotes the number of historical cyclones without landfall at  $C$ . Then, finally,  $n_{nlf}$  of such tracks are drawn at random from the previously simulated event set.

### 3.5 Acceptance-rejection for multiple landfalls

A particular advantage of the presented acceptance-rejection method is its applicability to multiple landfalls at two different coastlines  $C_1$  and  $C_2$ . The necessity of treating this case arises from the fact that there is a considerable number of TCs that affect several distinct coastal areas. However, it seems to be impossible to adjust  $C_1$ -landfall vectors and  $C_2$ -landfall vectors simultaneously. Instead, we apply our acceptance-rejection procedure several times consecutively to different sets of landfall vectors and with different historical reference LFIFs. The method is subdivided into four sequential steps, where only the third one is described in detail. The remaining steps do not involve any new ideas and were already explained in Section 3.4.

We use the decomposition of the family of all TCs in the presence of multiple landfalls mentioned in Section 3.2, namely that TCs can be divided into four groups: those with multiple landfall, those with landfall only at  $C_1$ , those with landfall only at  $C_2$ , and those cyclones that do not make landfall at  $C_1$  or  $C_2$ . Furthermore, we consider a simulated cyclone event set representing a time span of  $\alpha T$  years, which constitutes the basis for the further steps of the procedure. At first, acceptance-rejection is performed for  $C_2$ -landfall vectors of cyclone tracks that make landfall only at  $C_2$ . Here, we proceed analogously to the case of single landfalls. We require the validity of (3) and consider a decision rule for acceptance-rejection according to (4). The only difference is that LFIFs of historical and simulated cyclone tracks have to be replaced by respective LFIFs of  $C_2$ -landfall vectors. In doing so, we obtain a set of synthetic tracks that make landfall only at  $C_2$  with adjusted landfall characteristics. In the second step of the method, the procedure is repeated for  $C_2$ -landfall vectors of cyclone tracks that make landfall at both  $C_1$  and  $C_2$ . Again, a decision rule for acceptance-rejection is obtained by inserting the corresponding LFIFs into (3) and (4).

Note that up to this point,  $C_2$ -landfall vectors of all simulated TC tracks with landfall at  $C_2$  have been adjusted to the scaled LFIFs of historical cyclone observations. In the second step, however, we gained synthetic cyclone tracks that make landfall at  $C_1$  as well. Since these tracks are already accepted, it is not possible to adjust their  $C_1$ -landfall vectors. Thus, in the third step of the method, we proceed as follows. In order to adjust  $C_1$ -landfall vectors of simulated tracks that only make landfall at  $C_1$ , acceptance probabilities are now constructed in such a way that all accepted tracks with landfall at  $C_1$  have  $C_1$ -landfall vectors that were observed among all historical tracks with landfall at  $C_1$  (regardless of whether a second landfall occurs or not). In other words, a suitable reference LFIF is given by the scaled LFIF of all historical  $C_1$ -landfall vectors subtracted by the LFIF of the simulated  $C_1$ -landfall vectors which have already been accepted in the second step. More precisely, we introduce the LFIF  $\lambda_{C_1}^{acc}(\cdot)$  of simulated  $C_1$ -landfall vectors that were accepted in step 2 and assume that  $\alpha$  is chosen large enough, such that

$$\lambda_{C_1}^{sim}(x) \geq \frac{T}{T_{hist}} \lambda_{C_1}^{hist}(x) - \lambda_{C_1}^{acc}(x) \quad \text{for all } x \in [0, \infty)^3. \quad (5)$$

Here,  $\lambda_{C_1}^{sim}(\cdot)$  denotes the LFIF of  $C_1$ -landfall vectors of simulated cyclone tracks with landfall only at  $C_1$ , whereas  $\lambda_{C_1all}^{hist}(\cdot)$  denotes the LFIF of all historical  $C_1$ -landfall vectors (regardless of whether tracks make single or multiple landfall). Furthermore, the right-hand side of (5) is assumed to be always nonnegative. Based on these conditions, to each simulated cyclone track that makes landfall only at  $C_1$  with  $C_1$ -landfall vector  $s_i$  the acceptance probability

$$p(s_i) = \frac{\frac{T}{T_{hist}} \lambda_{C_1all}^{hist}(s_i) - \lambda_{C_1}^{acc}(s_i)}{\lambda_{C_1}^{sim}(s_i)} \quad (6)$$

is assigned. Then, by the independent thinning property of Poisson processes we get that after performing acceptance-rejection with the probabilities given in (6), accepted  $C_1$ -landfall vectors form a Poisson process with LFIF  $\frac{T}{T_{hist}} \lambda_{C_1all}^{hist}(\cdot) - \lambda_{C_1}^{acc}(\cdot)$ . Merging the set of these landfall vectors with those already obtained in step 2 (i.e. with LFIF  $\lambda_{C_1}^{acc}(\cdot)$ ) gives a Poisson process with the desired LFIF  $\frac{T}{T_{hist}} \lambda_{C_1all}^{hist}(\cdot)$ .

In this way, the LFIF of all simulated  $C_1$ -landfall vectors and the scaled LFIF of all historical  $C_1$ -landfall vectors coincide. However, a certain disadvantage of the presented approach is that  $C_1$ -landfall vectors of simulated tracks with multiple landfall are not adjusted separately. Thus, the statistical properties of these vectors can differ from those observed in historical data. The same applies for  $C_1$ -landfall vectors of tracks with landfall only at  $C_1$ . As described in Section 3.4, accepted tracks are merged with those cyclone tracks that do not make landfall at  $C_1$  or  $C_2$  to complete TC generation.

#### 4 Implementation, results, and validation

The presented acceptance-rejection method for simulated TC tracks has been incorporated into the stochastic simulation model using classes from the Java-based GeoStoch library, see Mayer et al (2004). For both the NA and WNP, synthetic cyclone event sets representing a period of  $T_{hist}$  years have been generated and evaluated. We provide three different types of illustrations that allow for a comparison of historical and simulated landfall characteristics.

Since a visualization of (three-dimensional) LFIFs as introduced in Section 3.3 is not possible, we consider one-dimensional LFIFs of landfall locations instead. These functions indicate how many cyclones with landfall occur and how landfalls are distributed along a coastline but do not distinguish between cyclones with different wind and translational speeds. They are estimated for each considered coastline using a one-dimensional version of the kernel density estimator given in (2). In Fig. 4 and 5, LFIFs of historical and simulated landfall locations are shown for some selected coastlines of the NA and WNP. On average, LFIFs show a good matching. Single simulation runs feature some more variation, but as long as no systematic bias occurs, this effect is absolutely desired.

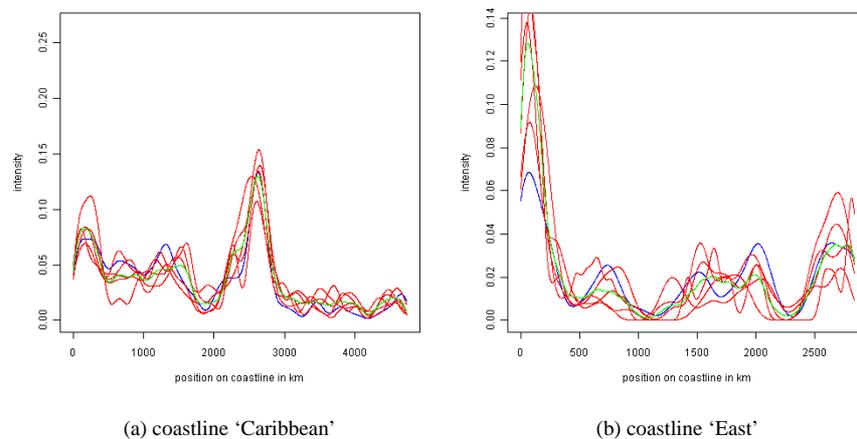


Fig. 4: Estimated LFIFs for landfall locations of historical cyclones (blue) and five simulated cyclone event sets (red) for selected coastlines of the NA. The green curve denotes the average of the five red curves.

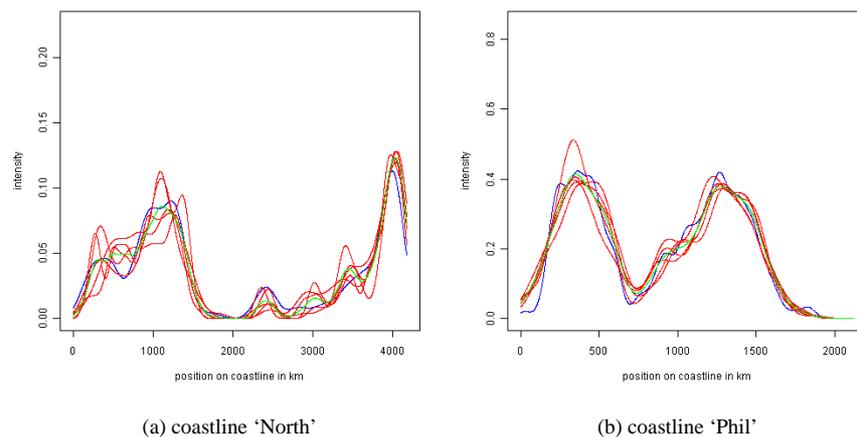


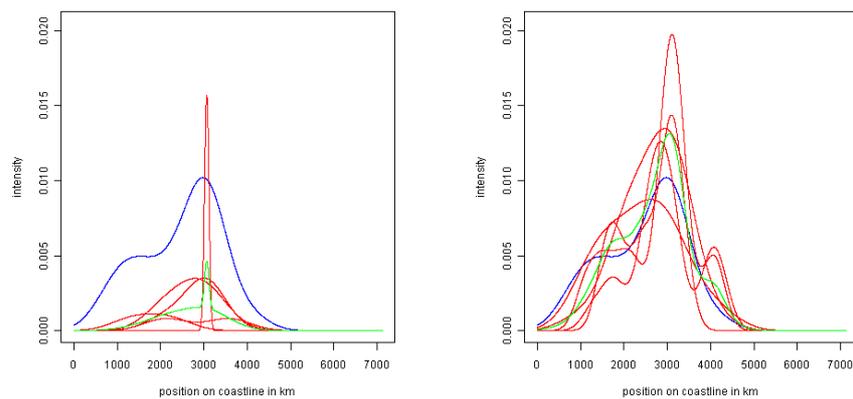
Fig. 5: Estimated LFIFs for landfall locations of historical cyclones (blue) and five simulated cyclone event sets (red) for selected coastlines of the WNP. The green curve denotes the average of the five red curves.

Unfortunately, one-dimensional LFIFs do not provide information about further landfall characteristics. On the one hand, similar (one-dimensional) comparisons could be made for wind and translational speeds at landfall of historical and simulated cyclone tracks that make landfall anywhere at a considered coastline. However, we are rather interested in joint distributions of landfall vectors, whose adjustment was the main aim of acceptance-rejection procedure. An example would be the question whether historical and simulated cyclones with certain wind and translational speeds make landfall in the same regions. Since a direct comparison of three-dimensional distri-

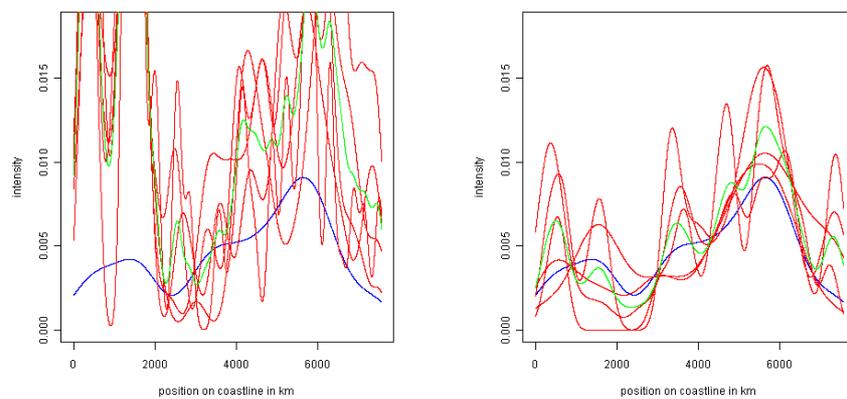
butions is difficult, we choose the following simplified approach. For each coastline, wind and translational speeds are subdivided into three groups ('low', 'medium', 'high'), respectively. Thus, all TC tracks with landfall at a considered coastline can be allocated to exactly one of nine groups according to their wind and translational speeds at landfall. For each of the nine groups, again LFIFs of landfall locations are estimated. Some examples are illustrated in Fig. 6 and 7. In addition to LFIFs of TCs obtained by the acceptance-rejection procedure, we estimated LFIFs of cyclone tracks generated by the former version of the simulator as described in Section 2, i.e. without correction of landfall behavior. Here, clear differences are visible, see the left-hand sides of Fig. 6 and 7. At coastlines 'USA' and 'Japan', for example, we previously had far too few synthetic cyclone tracks with low translational speeds and high wind speeds at landfall, a discrepancy that occurred systematically. After using the acceptance-rejection method, the total numbers and locations of landfalls of such tracks are, on average, much closer to those of historical observations. At coastlines 'Gulf' and 'South', on the other hand, we previously had far too many synthetic tracks with high translational speeds and low or medium wind speeds at landfall. Here again, the acceptance-rejection method provides a set of cyclone tracks with more realistic landfall behavior. This trend is also observed at the remaining coastlines. In general, the integration of the acceptance-rejection method into the stochastic track simulation model has led to a much better coincidence of LFIFs computed from historical and simulated data. Note, however, that when dividing TCs into nine groups, LFIFs of different simulation runs fluctuate to a greater extent than observed in Fig. 4 and 5. This effect is quite natural since for each single group fewer landfall vectors are available, which leads to more variability.

Additionally, we compared once again estimated return levels as introduced in Section 2.4. In Fig. 8 and 9, hazard maps are displayed for synthetic cyclone event sets that are obtained through the acceptance-rejection method. A comparison with Fig. 1 and 2 reveals significant improvements. On the one hand, a consequence of adjusting landfall locations and wind speeds is that in coastal areas, return levels of historical and simulated cyclone event sets coincide almost completely. In particular, return levels increased at the western Gulf coast in the NA and the southern Chinese and Vietnamese coast in the WNP and decreased in regions where previously an overestimation of return levels was observed, e.g. the southeast of the US, the Greater Antilles, northern China, South Korea, and Japan. On the other hand, the adjustment of translational speeds prevents the occurrence of high return levels unrealistically deep inland as previously observed north of Florida or in northern China. However, we recognize that in areas farther away from sea, slight differences between return levels of historical and simulated event sets still occur.

Altogether, the results obtained in the present paper indicate that there is indeed an essential adjustment of the (three-dimensional) landfall vectors of generated cyclone tracks to those of historical observations, which allows for a more precise hazard assessment.



(a) LFIFs estimated from cyclones that make landfall at coastline 'USA' with low translational speeds and high wind speeds; before (left) and after (right) using the acceptance-rejection procedure

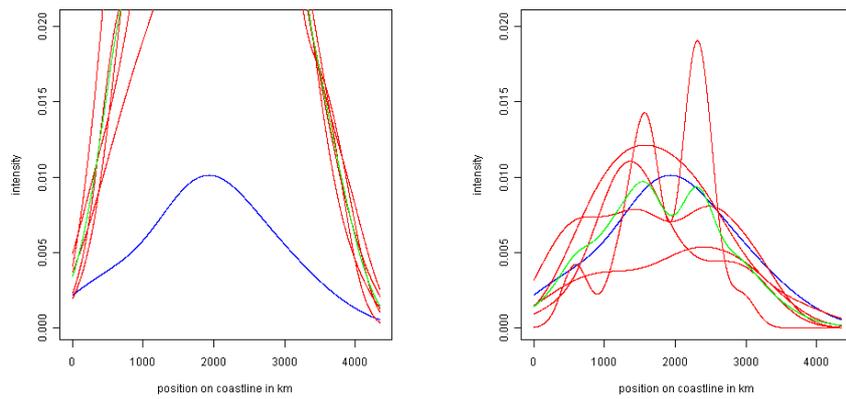


(b) LFIFs estimated from cyclones that make landfall at coastline 'Gulf' with high translational speeds and low wind speeds; before (left) and after (right) using the acceptance-rejection procedure

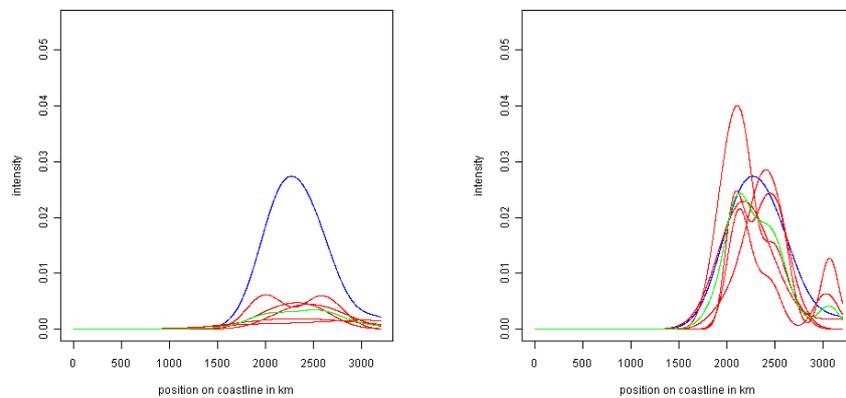
Fig. 6: Estimated LFIFs for landfall locations of historical cyclones (blue) and five simulated cyclone event sets (red). The green curve denotes the average of the five red curves.

## 5 Summary

In the present paper, the model introduced in Rumpf et al (2007, 2009) for the stochastic simulation of TC tracks is discussed and extended. Our aim is to adjust the landfall behavior of simulated cyclone tracks to that of historical observations to increase the low accuracy of the track simulation model. Important coastlines in the NA and WNP are approximated by polygonal lines in order to simplify the identification of landfalls and the computation of three different landfall characteristics: location,



(a) LFIFs estimated from cyclones that make landfall at coastline 'South' with high translational speeds and medium wind speeds; before (left) and after (right) using the acceptance-rejection procedure



(b) LFIFs estimated from cyclones that make landfall at coastline 'Japan' with low translational speeds and high wind speeds; before (left) and after (right) using the acceptance-rejection procedure

Fig. 7: Estimated LFIFs for landfall locations of historical cyclones (blue) and five simulated cyclone event sets (red). The green curve denotes the average of the five red curves.

translational speed, and wind speed. These landfall characteristics are combined to (three-dimensional) landfall vectors and it is shown that sets of landfall vectors can be assumed to constitute spatial Poisson point processes. Due to that, a thinning property of Poisson processes is applied to derive an acceptance-rejection procedure for simulated cyclone tracks. In this context, the cases of single and multiple landfalls are treated separately. Different types of visual validation indicate that after using the acceptance-rejection method, the joint distributions of simulated landfall characteristics are much closer to historical ones than before.

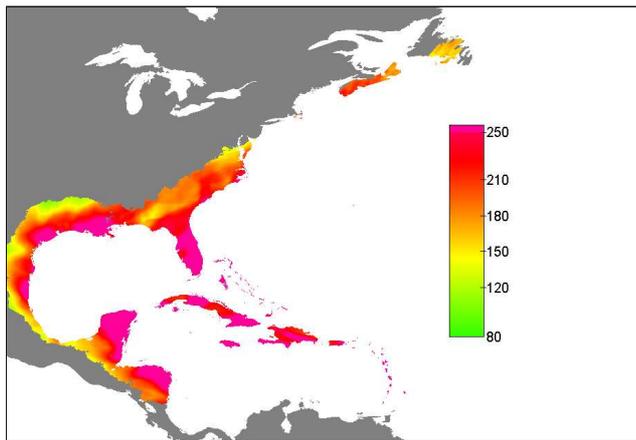


Fig. 8: Hazard maps for return period 100 years in the NA. Return levels are estimated from a simulated cyclone event set where the acceptance-rejection method is applied.

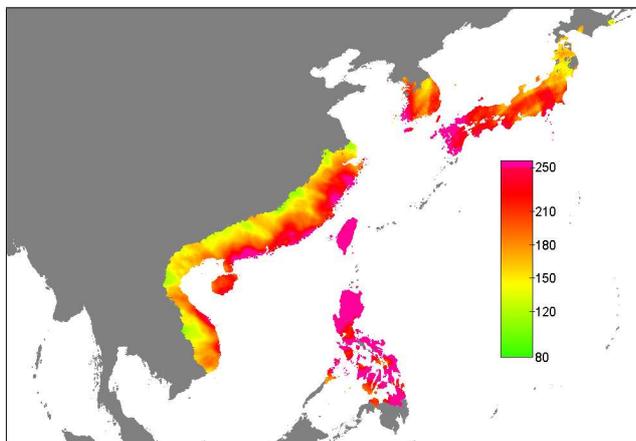


Fig. 9: Hazard maps for return period 100 years in the WNP. Return levels are estimated from a simulated cyclone event set where the acceptance-rejection method is applied.

The simultaneous adjustment of different landfall characteristics (location, translational speed, wind speed, as well as total number of landfalls) has not been discussed so far in the literature on stochastic cyclone track simulation. It allows for the generation of synthetic cyclone tracks with plausible landfall behavior. However, a certain limitation of the proposed model in comparison to other statistical simulation models is the strong simplification of complex meteorological aspects. This involves that neither information on oceanic and atmospheric temperature and pressure (ENSO, AMO, NAO), as e.g. in Yonekura and Hall (2011), nor effects of global warming, as e.g. in Hallegatte (2007), are taken into consideration. Another limitation is the use of a quite simple historical data base, which only contains cyclone measurements

in six-hour intervals. A promising alternative, at least for the NA, is the use of the hourly interpolated best track data introduced in Elsner and Jagger (2013). This and a stronger integration of climatic phenomena into our track simulation model could be subject of future research. Moreover, a transfer of the proposed simulation model including acceptance-rejection to other ocean basins, e.g. the Indian Ocean and the South Pacific, is possible.

**Acknowledgements** The authors would like to thank Scot Johnson for helpful comments on an earlier version of the manuscript.

## References

- Elsner JB, Jagger TH (2013) Hurricane climatology: a modern statistical guide using R. Oxford University Press, Oxford
- Emanuel K, Ravela S, Vivant E, Risi C (2006) A statistical deterministic approach to hurricane risk assessment. *Bull Am Meteorol Soc* 87(3):299–314
- Hall TM, Jewson S (2007) Statistical modelling of North Atlantic tropical cyclone tracks. *Tellus* 59A(4):486–498
- Hall TM, Jewson S (2008) Comparison of local and basinwide methods for risk assessment of tropical cyclone landfall. *J Appl Meteorol Climatol* 47(2):361–367
- Hallegatte S (2007) The use of synthetic hurricane tracks in risk analysis and climate change damage assessment. *J Appl Meteorol Climatol* 46(11):1956–1966
- Illian J, Penttinen A, Stoyan H, Stoyan D (2008) Statistical analysis and modelling of spatial point patterns. J. Wiley and Sons, Chichester
- Jagger TH, Elsner JB (2006) Climatology models for extreme hurricane winds near the United States. *J Climate* 19(13):3220–3236
- Kingman JFC (1993) Poisson processes. Oxford University Press, Oxford
- Mayer J, Schmidt V, Schweiggert F (2004) A unified simulation framework for spatial stochastic models. *Simul Model Pract Theory* 12(5):307–326
- Møller J, Waagepetersen RP (2004) Statistical inference and simulation for spatial point processes. Chapman & Hall/CRC, Boca Raton
- Rumpf J, Weindl H, Höppe P, Rauch E, Schmidt V (2007) Stochastic modelling of tropical cyclone tracks. *Math Meth Oper Res* 66(3):475–490
- Rumpf J, Weindl H, Höppe P, Rauch E, Schmidt V (2009) Tropical cyclone hazard assessment using model-based track simulation. *Nat Hazards* 48(3):383–398
- Scott DW (1992) Multivariate density estimation: theory, practice and visualization. J. Wiley and Sons, Chichester
- Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall, London
- Vickery PJ, Skerlj PF, Twisdale LA (2000) Simulation of hurricane risk in the U.S. using empirical track model. *J Struct Eng* 126(10):1222–1237
- Yonekura E, Hall TM (2011) A statistical model of tropical cyclone tracks in the western North Pacific with ENSO-dependent cyclogenesis. *J Appl Meteorol Climatol* 50(8):1725–1739