Math Meth Oper Res (2007) 66:475–490
DOI 10.1007/s00186-007-0168-7

ORIGINAL ARTICLE

Stochastic modelling of tropical cyclone tracks

Jonas Rumpf · Helga Weindl · Peter Höppe · Ernst Rauch · Volker Schmidt

Received: 27 October 2006 / Revised: 2 March 2007 / Published online: 26 June 2007

Abstract A stochastic model for the tracks of tropical cyclones that allows for the computerised generation of a large number of synthetic cyclone tracks is introduced. This will provide a larger dataset than previously available for the assessment of risks in areas affected by tropical cyclones. To improve homogeneity, the historical tracks are first split into six classes. The points of cyclone genesis are modelled as a spatial Poisson point process, the intensity of which is estimated using a generalised version of a kernel estimator. For these points, initial values of direction, translation speed, and wind speed are drawn from histograms of the historical values of these variables observed in the neighbourhood of the respective points, thereby generating a first 6-h segment of a track. The subsequent segments are then generated by drawing changes in theses variables from histograms of the historical data available near the cyclone's current location. A termination probability for the track is determined after each segment as a function of wind speed and location. In the present paper, the model is applied to historical cyclone data from the western North Pacific, but it is general enough to be transferred to other ocean basins with only minor adjustments. A version for the North Atlantic is currently under preparation.

Keywords Stochastic model · Monte-Carlo simulation · Inhomogeneous Poisson point process · Generalised spatial random walk · Tropical cyclones · Risk assessment

J. Rumpf (✉) · V. Schmidt
Ulm University, Institute of Stochastics, 89069 Ulm, Germany
e-mail: jonas.rumpf@uni-ulm.de

H. Weindl · P. Höppe · E. Rauch
Munich Reinsurance Company, 80791 Munich, Germany

1 Introduction

1.1 Motivation

Catastrophies caused by tropical cyclones are not only a threat to human lives but also a massive financial risk to insurance and reinsurance companies. These companies need to assess as precisely as possible the risk and extent of losses in areas affected by tropical cyclones. Since reliable data on cyclone tracks is only available for a relatively short period of time, it is not sufficient to make a risk assessment based solely on historical storm tracks. Therefore, as one possible approach to this problem, a basin-wide stochastic simulation model for the western North Pacific as outlined in Rumpf et al. (2006) is introduced. To the authors' knowledge, this is a novel concept for this ocean basin. Approaches with similar ideas for the North Atlantic and the South Pacific have been discussed in James and Mason (2005), Vickery et al. (2000), and most recently in Emanuel et al. (2006) and Hall and Jewson (2007).

1.2 Overview

After a short description of the available data, an explanation is given in Sect. 2 as to why and how this data is homogenised by splitting it into six different classes. Section 3 is concerned with modelling the starting points of the cyclone tracks as an inhomogeneous Poisson point process and the justification for this choice of model. The actual track model, in which resampling methods play an important role, is explained in Sect. 4. An algorithm for the simulation of cyclone tracks is specified in Sect. 5, along with some sample results of the algorithm and a brief discussion of the assessment of these results. After providing a brief insight into the methods of risk assessment applied to the simulated tracks in Sect. 6, the paper concludes with an outlook on possibilities for further model development.

1.3 Data

The available data consists of the tracks of all tropical cyclones recorded during the period 1945–2004 in the western North Pacific. Since there is no "definitive" dataset for this ocean basin in the same way as the *HURDAT best track* is for the North Atlantic (see Jarvinen et al. 1984), we use data Munich Re has compiled from different sources, mainly the Japanese Meteorological Agency (JMA), the Joint Typhoon Warning Center (JTWC) and Unisys Weather. Figure 1 shows the tracks of all 1,519 storms considered. Each track is given as a polygonal trajectory connecting between 2 and 100 points of measurement. Besides the date and time of measurement, the storm's current position (latitude and longitude) and its current maximum windspeeds are given for each point. The measurements within each individual storm are taken at regular intervals of 6 h, so the storm's translational speed can be easily calculated. All observations fall into an observation window that is delimited by the equator in the south, 60°N in the north, 100°E in the west, and the date line in the east.

2 Classification

As can be easily seen from Fig. 1, there are strong inhomogeneities in the shapes of the cyclone tracks. To improve the quality of the simulation, the storm tracks are therefore split into 6 more homogeneous classes. Since the shape of a storm track is a spatial characteristic, we split the observation window into 4 different zones, to correspond roughly to the map's major areas of land and sea, respectively; see Fig. 2. Tracks are then classified on the basis of the zones they touch. The classification procedure consists of two parts: First, storm tracks are split into classes 0, 1, 2, and 3 according to the criteria listed in Table 1. Then, class 4 is created with those storms from class 2 that have their starting point in zone 1, and class 5 is created with storms of class 1 which have their starting point east of 122°E, i.e. in the South China Sea. This

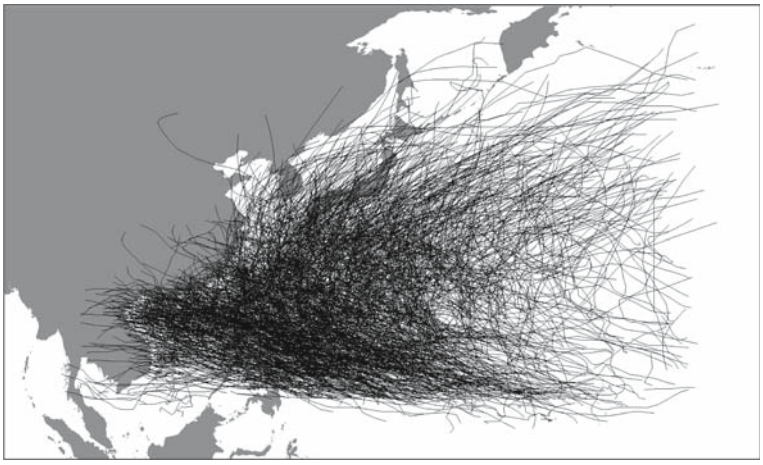

Fig. 1 Tracks of all storms in the dataset

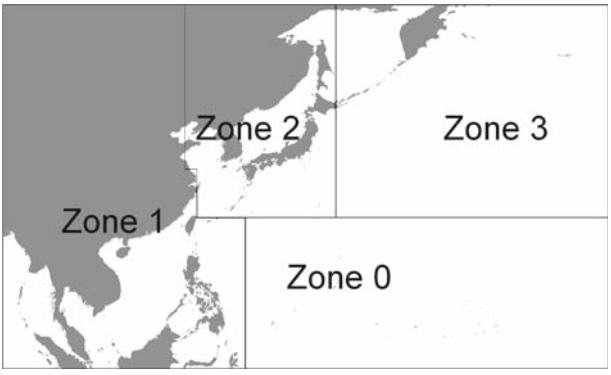

Fig. 2 Observation window split into 4 zones

Table 1 Criteria for the classification of the cyclone tracks

Start in zone	Touched zones	End in zone	Class
0	0	0	0
1	1	1	1
2	2	2	2
3	3	3	3
0 or 1	0 and 1	0 or 1	1
0 or 2	0 and 2	0 or 2	2
0 or 3	0 and 3	0 or 3	3
1 or 2	1 and 2	1	1
1 or 2	1 and 2	2	2
2 or 3	2 and 3	2 or 3	2
0, 1 or 2	0, 1 and 2	0	0
0, 1 or 2	0, 1 and 2	1	1
0, 1 or 2	0, 1 and 2	2	2
0, 1 or 3	0, 1 and 3	0	0
0, 1 or 3	0, 1 and 3	1	1
0, 1 or 3	0, 1 and 3	2	2
0, 2 or 3	0, 2 and 3	0, 2 or 3	2
1, 2 or 3	1, 2 and 3	1	1
1, 2 or 3	1, 2 and 3	2 or 3	2
0, 1, 2 or 3	0, 1, 2 and 3	0	0
0, 1, 2 or 3	0, 1, 2 and 3	1	1
0, 1, 2 or 3	0, 1, 2 and 3	2 or 3	2

particular choice of classes has the desirable properties of creating a decent amount of homogeneity among the track shapes with not too many classes. The authors do recognize, however, that the vague notion of homogeneity among track shapes introduces a certain amount of arbitrariness, though this is not considered a detriment to the model, since no information is lost and a helpful tool for the simulation is gained.

Two of the resulting classes of storm tracks can be seen in Figs. 3 and 4, respectively. The class sizes are given in Table 2. A more intuitive description of the classification criteria can be given as follows:

- Class 0 contains all the storms whose tracks are situated completely in the open sea in the southeastern part of the observation window.
- Class 1 contains the storms, whose tracks start in the southeastern part of the observation window and then head mostly northwest in an almost straight line towards the Philippines, the South China Sea, and the Asian continent; see Fig. 3.
- Class 2 contains the storms whose tracks start in the eastern part of the observation window. After initially heading northwest, they head to the right towards the northeast and in one way or the other affect Japan and/or the Korean Peninsula; see Fig. 4.

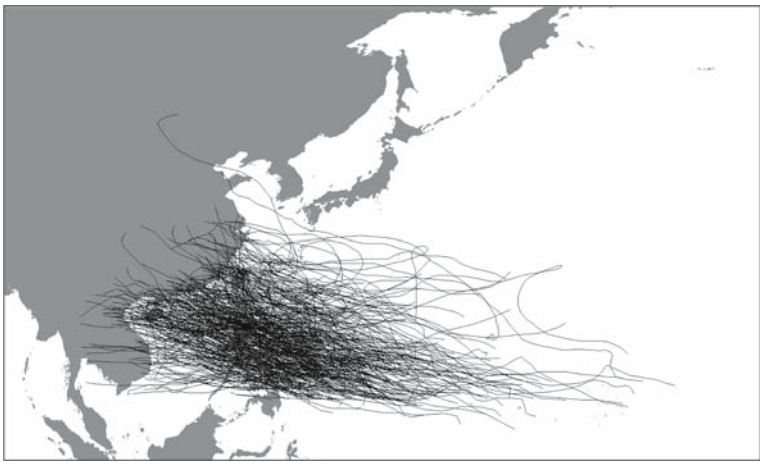

Fig. 3 Tracks of all storms in class 1

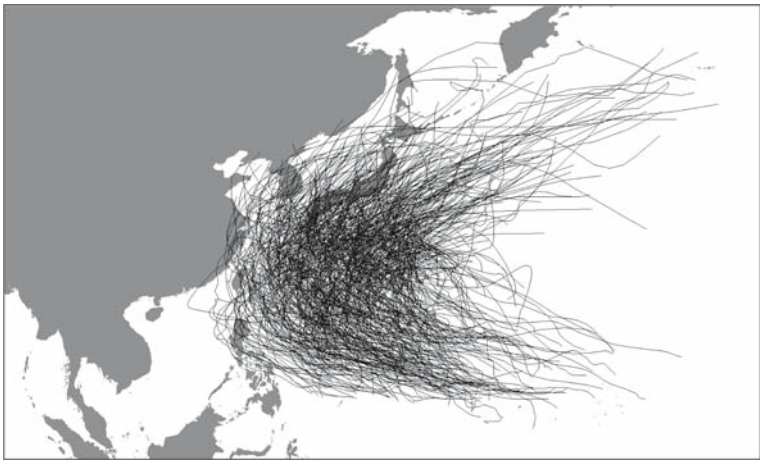

Fig. 4 Tracks of all storms in class 2

- Class 3 contains storms whose tracks are situated completely in the open sea in the eastern part of the observation window and which move further up north than those of class 0.
- Class 4 contains those storms that start off from the South China Sea and then head northeast towards Japan in a nearly straight line.
- Class 5 contains storms whose tracks are almost completely limited to the vicinity of the South China Sea. Most of them move in a straight line towards the Asian continent.

After creating the six different storm classes, all subsequent steps of the modelling process are done separately for each class.

Table 2 Class sizes of the storm classes

Class	Number of tracks	Number of data points
0	115	1,939
1	470	11,958
2	470	14,695
3	178	4,086
4	84	2,032
5	202	2,667
Total	1,519	37,377

3 Points of genesis

3.1 Basic model

For a stochastic model of the tracks of tropical cyclones, first a model for the points of cyclone genesis, i. e. the first points of the tracks, is needed. Figure 5 shows, as an example, the points of genesis of storms in class 2. The points are clearly distributed inhomogeneously within the observation window. Therefore, an inhomogeneous Poisson process is chosen as a model, which is further justified by the central properties of this point process model (see, for example, Stoyan and Stoyan 1994, p. 228):

- **No interaction:** The Poisson process is considered to be a model for complete spatial randomness, in the sense that its points are placed in the observation window completely independent of each other. They exhibit no interaction; neither attraction (or clustering) nor repulsion is incorporated into the model. This property reflects the fact that meteorology provides no evidence of interactions between storms' starting points, especially not over the number of years contained in the data. However, the

Fig. 5 Points of genesis of storms in class 2

authors do recognise that in any given year a clustering of storms in a certain region could occur.

- **Poisson distribution:** In a Poisson process, the number of points in a given area is Poisson-distributed. This corresponds to the result of a Pearson-Fisher-Goodness-of-Fit test, which does not reject the hypothesis that the number of storms per year in the given data is Poisson-distributed. This result is obtained not only for the numbers of storms in the different classes, but also for the total number of storms within the observation window (see Sect. 3.2).

3.2 Tests for Poisson distribution

To test the hypothesis that the number of storms in the data is Poisson-distributed, i.e.

$$H_0 : \ P \in \{Poi(\lambda), \lambda > 0\} \qquad \text{vs.} \qquad H_1 : \ P \notin \{Poi(\lambda), \lambda > 0\}, \tag{1}$$

where P is the distribution of the number of cyclones per year within the observation window, an asymptotic Pearson-Fisher-Goodness-of-Fit test is performed (see Cramér 1971, Chap. 30). For this test, consider the numbers of storms in the 59 years contained in the data as realisations $u_1, \ldots, u_n$ of independent and identically distributed random variables $U_1, \ldots, U_n$, where $n = 59$. These realisations are then grouped into r disjoint subsets $A_0 = \{0, \ldots, a_0\}$, $A_1 = \{a_0 + 1, \ldots, a_1\}, \ldots, A_{r-2} = \{a_{r-3} + 1, \ldots, a_{r-2}\}$, $A_{r-1} = \{a_{r-2} + 1, a_{r-2} + 2, \ldots\}$. The values of $a_0, \ldots, a_{r-2}$ are chosen in the following way: starting with a_0, the a_j are determined iteratively as the minimal values that ensure that the condition

$$p_j(\widehat{\lambda}) \cdot n > 5 \ \forall j = 0, \ldots, r-1, \tag{2}$$

holds, where

$$p_0(\widehat{\lambda}) = \sum_{i=0}^{a_0} \frac{\widehat{\lambda}^i}{i!} \mathrm{e}^{-\widehat{\lambda}}, \tag{3}$$

$$p_j(\widehat{\lambda}) = \sum_{i=a_{j-1}+1}^{a_j} \frac{\widehat{\lambda}^i}{i!} \mathrm{e}^{-\widehat{\lambda}} \ \forall j = 1, \ldots, r-2, \tag{4}$$

$$p_{r-1}(\widehat{\lambda}) = \sum_{i=a_{r-2}}^{\infty} \frac{\widehat{\lambda}^i}{i!} \mathrm{e}^{-\widehat{\lambda}} \tag{5}$$

denote the Poisson probabilities of the r subsets. Condition (2) requires the subsets A_i to each contain more than a certain minimum number of observations. While there seems to be a consensus in the literature that this is an important condition for the validity of the Pearson-Fisher test (see e.g. Cramér 1971, p. 420 and Gibbons 1985, p.72f), the actual values stated vary. In choosing the minimum number in (2) equal to 5, we follow Gibbons (1985).

Table 3 Tests for Poisson distribution of the number of storms per year

Class	$\bar{u}_n$	r	$T_n(u_1, \ldots, u_n)$
0	1.95	5	7.85
1	7.97	8	2.49
2	7.97	8	9.85
3	3.02	6	5.02
4	1.42	4	0.05
5	3.42	6	4.32
All	25.75	8	1.97

Table 4 Tests for Poisson distribution of the number of storms per year

Class	$\chi^2_{r-2,0.99}$	H_0 rej.?	$\chi^2_{r-2,0.95}$	H_0 rej.?	$\chi^2_{r-2,0.90}$	H_0 rej.?
0	11.34	No	7.81	Yes	6.25	Yes
1	16.81	No	12.59	No	10.64	No
2	16.81	No	12.59	No	10.64	No
3	13.28	No	9.49	No	7.78	No
4	7.38	No	5.99	No	4.61	No
5	13.28	No	9.49	No	7.78	No
All	16.81	No	12.59	No	10.64	No

Note that condition (2) is explicitly required to hold for $j = r - 1$, thereby determining uniquely not only the A_j, but also r. The maximum likelihood estimator $\widehat{\lambda}$ for the parameter λ of the hypothetical Poisson distribution from the grouped data can be approximated by the well-known maximum likelihood estimator for λ for ungrouped data, the sample mean $\overline{U}_n$. Then the test statistic

$$T_n(U_1, \ldots, U_n) = \sum_{j=0}^{r-1} \frac{\left(Z_j(U_1, \ldots, U_n) - np_j(\overline{U}_n)\right)^2}{np_j(\overline{U}_n)} \tag{6}$$

is asymptotically χ^2_{r-2}-distributed, where

$$Z_j(U_1, \ldots, U_n) = \#\{i : 1 \leq i \leq n, U_i \in A_j\}. \tag{7}$$

Therefore, the hypothesis H_0 is rejected at a given level of significance α if

$$T_n(u_1, \ldots, u_n) > \chi^2_{r-2,1-\alpha}, \tag{8}$$

recalling that $(u_1, \ldots, u_n)$ denotes a realization of $(U_1, \ldots, U_n)$. As can be seen from Tables 3 and 4, for the tropical cyclone data, the hypothesis is not rejected except for storms of class 0 for the higher α-levels of 0.05 or 0.1. This result holds for the different storm classes as well as for the total number of storms.

3.3 Intensity estimation

The distribution of points of an inhomogeneous Poisson point process within the observation window W is determined by its intensity function $\lambda(\mathbf{t})$. This function can be interpreted in a way that $\lambda(\mathbf{t})d\mathbf{t}$ describes the infinitesimal probability of a point of the Poisson process being located in the infinitesimally small disc with area $d\mathbf{t}$ centred at $\mathbf{t}$ (see, for example, Stoyan et al. 1995, p. 42). Since there is no obvious parametric trend visible in the data (see Fig. 5), a non-parametric estimation technique was chosen. The generalised nearest neighbour estimator (see Silverman 1986, p. 97) is given by

$$\widehat{\lambda}(\mathbf{t}) = r_k(\mathbf{t})^{-2} \sum_{i=1}^{m} K_e \left\{ r_k(\mathbf{t})^{-1} (\mathbf{t} - \mathbf{T}_i) \right\} , \tag{9}$$

where $r_k(\mathbf{t})$ is the distance to the k-th nearest point of genesis seen from the location $\mathbf{t}$, $\mathbf{T}_i$ the location of the i-th historical point of cyclone genesis, and K_e the Epanechnikov kernel:

$$K_e(\mathbf{t}) = \begin{cases} \frac{2}{\pi}(1 - \mathbf{t}^\top \mathbf{t}) & \text{if } \mathbf{t}^\top \mathbf{t} < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The parameter k is chosen such that $k = \lfloor \sqrt{m} \rfloor$, where m is the number of historical points of genesis.

A simplified interpretation of this estimator is given in the following: while the kernel K_e determines the size and the shape of the "probability mass" which is assigned to a measurement point, the bandwidth $r_k(\mathbf{t})$ is the radius over which this mass is spread. Note that the estimator $\widehat{\lambda}(\mathbf{t})$ is nowhere zero: at all points within the observation window, there is a non-zero probability mass from exactly k historical points of genesis. This probability mass decreases with increasing distance to the k-th nearest historical point of genesis, but in theory never reaches zero. This effect is intended, because it allows, if only rarely, for the genesis of tropical cyclones within the model at locations that are far away from most historical initial points of cyclones, where there are no physical reasons against cyclone genesis. In some areas, of course, the intensity is set to zero because cyclone genesis is meteorologically impossible there. These areas include all locations closer to the equator than 3° of latitude, motivated by the negligible Coriolis force in these regions, as well as all locations not over sea, because of their lack of necessary heat sources for a cyclone.

4 Cyclone tracks and wind speeds

4.1 Direction, translational speed and wind speed

Once a model for the points of cyclone genesis is available, the propagation of the tracks is the next step in the modelling process. Here our model relies on the same basic assumption as the models introduced in, for example, Emanuel et al. (2006),

Hall and Jewson (2007), and Vickery et al. (2000) that cyclones located in similar areas of the observation window behave similarly. An appropriate model of the tracks following the points of genesis needs to include the direction of movement (denoted by X in the following) and the translational speed (Y), i.e. the velocity at which the cyclone is moving in the given direction. By assuming these characteristics to be constant for intervals of 6 h (see Sect. 1.3) and updating them instantaneously after each interval, the cyclone's location can be calculated in 6-h steps, thereby generating a complete trajectory. For risk assessment (see Sect. 6), additional information is needed, namely the maximum wind speeds attained (Z) at each of the cyclone's positions. To combine these characteristics, consider a 3-dimensional state vector S_i that contains their values after the i-th track segment. These values are considered to be the sum of an initial value and the changes in these values after each step:

$$S_i = S_0 + \sum_{j=1}^{i} \Delta S_j = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + \sum_{j=1}^{i} \begin{pmatrix} \Delta X_j \\ \Delta Y_j \\ \Delta Z_j \end{pmatrix} \tag{11}$$

Since a stochastic model is being developed, all of the characteristics X_0, Y_0 and Z_0 as well as ΔX_j, ΔY_j, ΔZ_j are considered to be random variables. The distributions of these random variables depend on the storm's current location $\mathbf{t}$ within the observation window W. To generate a realisation of S_0 at a certain location, data is resampled from the distributions of the historical measurements of X_0, Y_0 and Z_0 near that location, i.e. the probability distribution function of X_0 at location $\mathbf{t}$ is estimated by

$$F_{X_0}(x, \mathbf{t}) = \frac{\#\{l : 1 \le l \le k_{X_0},\ x_l^{(0)}(\mathbf{t}) \le x\}}{k_{X_0}}, \tag{12}$$

where $x_l^{(0)}(\mathbf{t})$, $l = 1, \ldots, k_{X_0}$, denote the k_{X_0} historical realisations of X_0 closest to the location $\mathbf{t}$. In short, the distribution of the initial direction of a track in the model is determined by all historical initial directions of storm tracks. Similar formulae are used in order to estimate the location-dependent distributions of Y_0 and Z_0, respectively. In analogy to this, the probability distribution functions of a change in direction ΔX_{j_0} is given by

$$F_{\Delta X}(x, \mathbf{t}) = \frac{\#\{l : 1 \le l \le k_{\Delta X},\ \Delta x_l(\mathbf{t}) \le x\}}{k_{\Delta X}}, \tag{13}$$

where $\Delta x_l(\mathbf{t})$, $l = 1, \ldots, k_{\Delta X}$, now denote the $k_{\Delta X}$ historical realisations of $\Delta X_j\ \forall j$ closest to the location $\mathbf{t}$. This means that the distribution of any change in direction ΔX_{j_0} depends on the historical realizations of *all* changes in direction ΔX_j of tropical cyclones, no matter after which step of a storm they occurred. A similar formula is used for ΔY_j.

To conform with reality, certain boundary conditions are imposed on the components of S_i. For example, the translational speeds Y_i must be nonnegative at all

times. Also note that all calculations involving the direction of a cyclone are made "*mod* 360°" to keep the value of the X_j within the interval $[0°, 360°)$, where 0° is considered North.

For the changes in wind speed ΔZ_j, a modified version of the probability distribution function given in (12) is considered. To reflect the fact that the wind speeds of stronger storms tend to decrease, while weaker storms tend to intensify in their early stages, the distribution of the changes in wind speed was made dependent on the previous wind speed z:

$$F_{\Delta Z}(x, \mathbf{t}, z) = \frac{\#\{l : 1 \le l \le k_{\Delta Z},\ \Delta z_l(\mathbf{t}, z) \le x\}}{k_{\Delta Z}}, \tag{14}$$

where the $\Delta z_l(\mathbf{t}, z),\ l = 1, \ldots, k_{\Delta Z}$, denote $k_{\Delta Z}$ historical realisations of those ΔZ_j closest to the location $\mathbf{t}$ that had previous wind speeds in the same range ("low", "medium-low", "medium-high", "high") as z.

4.2 Termination probabilities

Since the proposed model creates synthetic cyclone tracks in 6-h steps, a mechanism is needed to determine whether the track should be terminated after the current step or if it should be continued. This is done stochastically via a Bernoulli experiment with a success probability $p(\mathbf{t}, Z)$ depending on the storm's current location and wind speed. Here, the event "success" is considered equivalent to "the cyclone terminates". The termination probability is determined as the maximum of two probabilities p_Z and $p_{\mathbf{t}}$, since this approach to combining the two probabilities delivered the best results in eliminating spurious synthetic storm tracks penetrating unrealistically deep into the Asian continent.

Although in theory tropical cyclones should only be considered as such as long as their wind speeds exceed 62 km/h (34 knots), the wind speeds at the last points of measurement of the cyclones vary greatly in the available data. Therefore, to match the data, a curve of the form

$$p_Z = c \cdot \mathrm{e}^{-\lambda Z^{\alpha}} \tag{15}$$

depending on a cyclone's current wind speed Z was fitted to historical termination probabilities, which have been determined as follows: all points of measurement and their respective wind speeds are grouped into 10 km/h-wide bins. The termination probability of every bin is then given as the fraction of cyclone termination points among all points in this bin. To avoid artefacts resulting from imperfect data, points with speeds less than 30 km/h are omitted. The curve given in (15) is then fitted to the resulting 28 points using a least-squares method. It is well known that tropical cyclones behave differently over land than they do over sea. For example, cyclones are subject to higher friction and a lower energy supply over land than they are over sea. Therefore, this procedure is applied separately to the points of measurement over land and to those over sea. As an example, the data points and the fitted curve for the

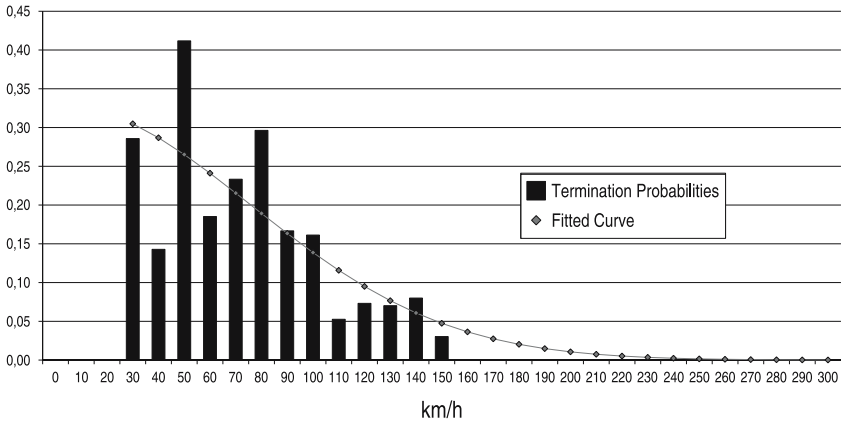

Fig. 6 Curve fitted to historical termination probabilities of storms of class 2 over land

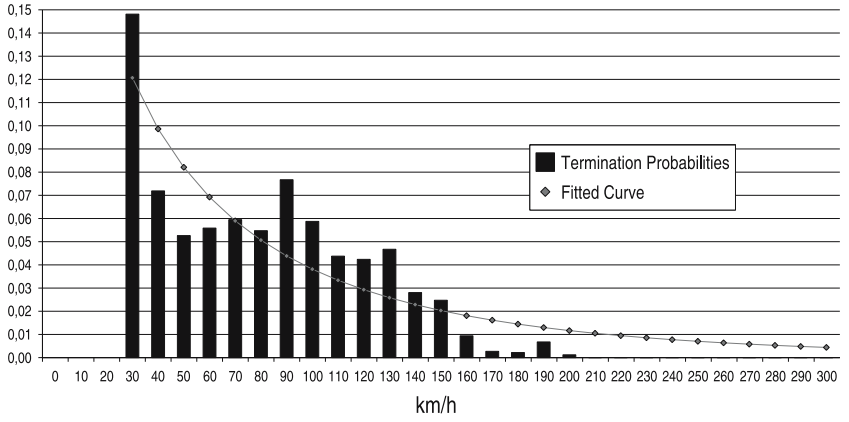

Fig. 7 Curve fitted to historical termination probabilities of storms of class 2 over sea

storms of class 2 over land and over sea are shown in Figs. 6 and 7, respectively. A possible lack in fit of the curve to the data is considered acceptable because of the fact that only the general form of the curve is needed for the model. In particular, an exact replication of the data is not intended, considering the goal of creating a larger dataset for risk assessment and bearing in mind the existence of imperfections in the data. Additionally, fitting errors are also compensated in part by the second termination probability, whose calculation is described in the following.

To account for the fact that the weakening of tropical cyclones and their wind speeds is significantly influenced by the geographical conditions at the storm's current location, a second probability $p_{\mathbf{t}}$ is calculated for all locations $\mathbf{t}$ in the observation window. Similarly to the probabilities given in (12), $p_{\mathbf{t}}$ is calculated as the fraction of termination points among the n nearest points of measurement of the location $\mathbf{t}$.

The termination probability used in the Bernoulli experiment is then taken to be

$$p(\mathbf{t}, Z) = \max\{p_Z, p_{\mathbf{t}}\}. \tag{16}$$

This allows for the quicker termination of storms that have reached specific locations while exhibiting uncommonly high wind speeds, such as storms that penetrate far inland with high wind speeds. On the other hand, it also accounts for the termination of storms that do not reach particularly high wind speeds in areas where most storms in the data have continued to exist.

5 Simulation and results

In this section, an algorithm for generating synthetic cyclone tracks is described, summarising the different parts of the model and illustrating their interaction. To create a complete set of synthetic storm tracks from the model described above, the procedure is as follows for each of the 6 storm classes (see Sect. 2):

0. **Initialisation:** Find all needed estimators and probabilities as they were defined in (9), (12), (13), (14), (15), and (16), respectively, and go to step 1.
1. **Points of genesis:** Generate a realisation of the inhomogeneous Poisson point process with the intensity function defined in (9) and go to step 2.
2. **Choose a point:** From the point process realisation generated in step 1, pick one point that does not yet have a corresponding cyclone track and go to step 3. If there are no such points left, terminate the algorithm.
3. **Initial segment:** Generate a realisation of S_0 from the distribution functions defined in (12) according to the location $\mathbf{t}$ of the cyclone's starting point from step 2. With this, find the storm's new location after its first segment and go to step 4.
4. **Termination probability:** Perform a Bernoulli experiment with the success probability given by (16) according to the storm's current location and wind speed. If the result is "success", terminate the storm track and go to step 6. Otherwise go to step 5.
5. **Additional segment:** Generate a realisation of ΔS_j from the distribution functions defined in (13) and (14) according to the storm's current location and wind speed. Add ΔS_j to S_{j-1} to get S_j and from this a new location and wind speed for the storm. Then go to step 4.
6. **Class verification:** Determine the classification of the generated storm track as described in Sect. 2. If it matches the class of storm tracks for which this algorithm is being performed, accept the storm track for the given point of genesis and go to step 2. Otherwise, reject the storm track and go to step 3 with the same point of genesis.

Note that the possibility of a storm being rejected mentioned in step 6 of the algorithm is not just theoretical, but that in fact quite a few rejections do happen during simulation. For example, when a storm of class 1 (see Fig. 3) is being simulated, the random combination of a starting point far to the east, an initial direction towards the west and several changes of the direction of movement to the right early on is

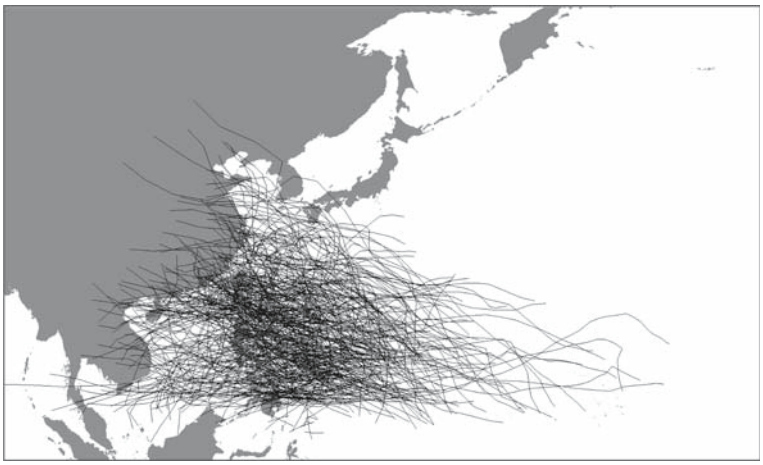

Fig. 8 Synthetic tracks of storms in class 1

improbable, but possible. This track will then be classified into class 3 instead of class 1 (see Table 1) and therefore will be rejected in step 6 of the algorithm.

This algorithm has been implemented using Java, which creates the possibility for the generation of a large number of synthetic cyclone tracks. A sample of synthetic storm tracks in class 1 with the same expected number of storms as in the original data is plotted in Fig. 8.

To evaluate the results of the model, 150 samples of synthetic storm tracks, where every sample is considered to consist of 59 years of data, are generated by simulating a random number of storms for each year which is Poisson-distributed with a parameter given by the mean number of storms per year in the historical data. Then the number of storms affecting Japan (which is the area of highest interest within the observation window), denoted by V_i, is counted for each year. From every sample with sample size $n = 59$, the expected number of storms affecting Japan per year and the variance of this number are estimated by the sample mean $\overline{V}_n$ and the sample variance S_n^2. The simulated data is then compared to the historical data by 150 realisations of an asymptotic one-sample test. The test statistic $T(V_1, \ldots, V_n)$ given by

$$T(V_1, \ldots, V_n) = \sqrt{n}\frac{\overline{V}_n - \mu_0}{\sqrt{S_n^2}} \tag{17}$$

is approximately $N(0, 1)$-distributed for sufficiently large n (see, for example, Lehmann and Romano 2005, p. 444). The hypothesis that the expected number of storms affecting Japan in the simulated data correctly reflects the corresponding number from the historical data (denoted by μ_0), is therefore rejected at a given level of significance α if $|T(V_1, \ldots, V_n)| > z_{1-\alpha/2}$ where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution. Table 5 shows that the percentage of samples where the hypothesis is rejected is approximately α, which suggests that the model correctly represents the number of storms affecting Japan per year.

Table 5 Test results for the comparison between simulated and historical data

α	1%	5%	10%
Rejections	0.7%	5.3%	10.0%

6 Risk assessment

From a cyclone track consisting of the storm's locations, translational speeds and maximum attained wind speeds, a two-dimensional wind field can be calculated. This windfield is obtained using an empirical relation between the maximum wind speed v_{max}, which is considered to be attained at the so-called "radius of maximum wind speeds" r_{max} of the storm and wind speeds $v(r)$ at a radial distance $r \geq r_{max}$ from the centre. The general form of this relation is given by

$$v(r) = v_{\max} \cdot \left(\frac{r}{r_{\max}} \right)^{-\gamma}, \tag{18}$$

where the exponent $\gamma \in (0, 1)$ and $r_{\max}$ are determined empirically.

In the lower troposphere, the wind in a cyclone is dominated by the tangential wind speed. Therefore, the cyclone wind profile (18) is originally derived using the conservation of relative angular momentum of tangential winds ($v(r) \cdot r = \text{const}$) outside of $r_{\max}$. However in reality, the wind in the boundary layer is spiraling inwards, and, as a consequence, it loses relative angular momentum due to frictional dissipation at the surface (see Depperman 1947, Holland 1980). Due to this fact, the exponent $\gamma < 1$ is introduced into (18), which then reflects a typical typhoon wind profile adequately for the purposes of this investigation.

With this method, wind speeds caused by a cyclone at locations of interest are calculated. These wind speeds can then be used to calculate an estimate for the damage the cyclone causes at these locations. This creates a possibility for a long-term risk assessment since, with the described model and its implementation in the programming language Java, it is possible to generate a large number of realistic storm tracks. For example, one could simulate tracks for a time horizon of 10,000 years and then calculate the "10,000-year damage" or damages with return periods of less than 10,000 years.

7 Summary and outlook

A stochastic model for the simulation of tropical cyclone tracks in the western North Pacific was developed. The model relies mostly on the historical track data available. Complex meteorological aspects of tropical cyclone movement are greatly simplified, thereby creating the possibility for the simulation of large numbers of synthetic cyclone tracks. With these synthetic tracks, the assessment of damage risks at locations of interest in areas affected by the cyclones can be improved.

The transfer of our model to the North Atlantic basin is currently under development. In addition, the model is constantly being enhanced. For example, it is intended

to model dependencies of S_j and S_{j-1} by Markov chains, related to what was suggested in, for example, Emanuel et al. (2006). Another possibility for refining the model could be to follow e.g. Hall and Jewson (2007) by weighting the historical data used in formulae (12) through (14) according to their distance from the storm's current location $\mathbf{t}$. Furthermore, additional methods for the comparison of simulated and historical cyclone tracks are being developed. With these methods, it will also be possible to assess the relevance of the previously mentioned enhancements to the performance of the model.

Acknowledgments The authors would like to thank an anonymous referee for a report that was very helpful in the improvement of the manuscript.

References

Cramér H (1971) Mathematical methods of statistics. Princeton University Press, Princeton

Depperman RCE (1947) Notes on the origin and structures of Philippine typhoons. Bull Am Met Soc 29:399–404

Emanuel K, Ravela S, Vivant E, Risi C (2006) A statistical deterministic approach to hurricane risk assessment. Bull Am Met Soc 87:299–314

Gibbons JD (1985) Nonparametric statistical inference. Marcel Dekker, New York

Hall T, Jewson S (2007) Statistical modeling of North Atlantic tropical cyclone tracks. Tellus B (in press)

Holland GJ (1980) An analytic model of the wind and pressure profiles in hurricanes. Mon Wea Rev 108:1212–1218

James MK, Mason LB (2005) Synthetic tropical cyclone database. J Waterw Port Coast Ocean Eng 131:181–192

Jarvinen BR, Neumann CJ, Davis MAS (1984) A tropical cyclone data tape for the North Atlantic basin, 1886-1983, contents, limitations and uses. NOAA Tech Memo, NWS NHC 22, Miami, Florida

Lehmann EL, Romano JP (2005) Testing statistical hypotheses, 3rd edn. Springer, New York

Rumpf J, Rauch E, Schmidt V, Weindl H (2006) Stochastic modeling of tropical cyclone track data. In: Proceedings of the 27th Conference on hurricanes and tropical meteorology, April 24–28, 2006, Monterey, CA

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall, New York

Stoyan D, Kendall WS, Mecke J (1995) Stochastic geometry and its applications. 2nd edn. Wiley, Chichester

Stoyan D, Stoyan H (1994) Fractals, random shapes and point fields. Methods of geometrical statistics. Wiley, Chichester

Vickery P, Skerlj PF, Twisdale LA (2000) Simulation of hurricane risk in the US using empirical track model. J Struct Eng 126:1222–1237