Clim Dyn
DOI 10.1007/s00382-016-3461-4

A climatological model of North Indian Ocean tropical cyclone genesis, tracks and landfall

Mohammad Wahiduzzaman[1,2] · **Eric C. J. Oliver**[1,3] · **Simon J. Wotherspoon**[1,4] · **Neil J. Holbrook**[1,3]

Received: 10 June 2016 / Accepted: 16 November 2016

Abstract Extensive damage and loss of life can be caused by tropical cyclones (TCs) that make landfall. Modelling of TC landfall probability is beneficial to insurance/re-insurance companies, decision makers, government policy and planning, and residents in coastal areas. In this study, we develop a climatological model of tropical cyclone genesis, tracks and landfall for North Indian Ocean (NIO) rim countries based on kernel density estimation, a generalised additive model (GAM) including an Euler integration step, and landfall detection using a country mask approach. Using a 35-year record (1979–2013) of tropical cyclone track observations from the Joint Typhoon Warning Centre (part of the International Best Track Archive Climate Stewardship Version 6), the GAM is fitted to the observed cyclone track velocities as a smooth function of location in each season. The distribution of cyclone genesis points is approximated by kernel density estimation. The model simulated TCs are randomly selected from the fitted kernel (TC genesis), and the cyclone paths (TC tracks), represented by the GAM together with the application of stochastic innovations at each step, are simulated to generate a suite of NIO rim landfall statistics. Three hindcast validation methods are applied to evaluate the integrity of the model. First, leave-one-out cross validation is applied whereby the country of landfall is determined by the majority vote (considering the location by only highest percentage of landfall) from the simulated tracks. Second, the probability distribution of simulated landfall is evaluated against the observed landfall. Third, the distances between the point of observed landfall and simulated landfall are compared and quantified. Overall, the model shows very good cross-validated hindcast skill of modelled landfalling cyclones against observations in each of the NIO tropical cyclone seasons and for most NIO rim countries, with only a relatively small difference in the percentage of predicted landfall locations compared with observations.

Keywords Tropical cyclone · Genesis · Tracks · Landfall · North Indian Ocean · Generalised additive model · Kernel density estimation

1 Introduction

Tropical cyclones (TCs) are considered to be the most devastating weather phenomena that affect North Indian Ocean (NIO) rim countries, producing major impacts over significantly large areas (Tyagi et al. 2010; Girishkumar and Ravichandran 2012; Nath et al. 2015). TCs have affected NIO rim communities since the earliest days of settlement (O'Hare 2001) and can have substantial impacts on the coastal countries of the Bay of Bengal and the Arabian Sea (Belanger et al. 2012).

The NIO is a breeding ground for tropical cyclones (Mohapatra et al. 2014; Shaji et al. 2014) but accounts for only 7% (Mohapatra et al. 2014; Sahoo and Bhaskaran 2016) of the world's tropical cyclones [about 5/year (Mohapatra et al. 2014)]—four in the Bay of Bengal

✉ Neil J. Holbrook
neil.holbrook@utas.edu.au

1 Institute for Marine and Antarctic Studies (IMAS), University of Tasmania, Hobart, TAS 7001, Australia

2 Department of Geography and Environment, Jahangirnagar University, Dhaka, Bangladesh

3 Australian Research Council Centre of Excellence for Climate System Science, University of Tasmania, Hobart, TAS, Australia

4 Australian Antarctic Division, 203 Channel Highway, Kingston, TAS 7050, Australia

Published online: 05 December 2016

(Alam et al. 2003; Vissa et al. 2013; Balaguru et al. 2014; Mohapatra et al. 2014; Rajasekhar et al. 2014) and one in the Arabian Sea (Rajeevan et al. 2013; Mohapatra et al. 2014; Rajasekhar et al. 2014)—but the number of deaths from TCs that make landfall in the region can be staggering (Webster 2008; Islam and Peterson 2009; Lin et al. 2009; McPhaden et al. 2009; Ng and Chan 2012; Pattanaik and Mohapatra 2016). Examples of the devastating effects of TCs in this region include the impacts of the Category-5 TC (*TC 02B*) that hit the Chittagong district of Bangladesh in 1991, and *TC Nargis* that hit Myanmar in 2008, with each resulting in around 140,000 lives lost and US$2.4 billion and $10 billion in damages respectively (Mydans and Cowell 2008; Alam and Collins 2010; Nath et al. 2015). Further, between 2007 and 2010, there were significant losses associated with at least one event in each year (Haggag and Badry 2012). In short, the human and financial costs annually due to TCs making landfall across NIO rim coastlines can be enormous.

In the present study, we develop a climatological TC model that takes account of the characteristics of past tropical cyclones and that can be used as a fundamental baseline for understanding the seasonality of TC genesis, tracks and landfall. Modelling of seasonal TC landfall probabilities has the potential to assist decision-makers, and residents in vulnerable coastal areas, to consider and plan. At the government planning and policy levels, seasonal modelling can be usefully adopted to inform and assist decision-making. Ultimately skilful prediction and careful assessment of the possible risk and extent of losses in areas affected by tropical cyclones can be beneficial to insurance companies (Rumpf et al. 2007).

Seasonal forecast modelling of tropical cyclone activity represents a major challenge for dynamical models (Camargo 2013; Shaevitz et al. 2014; Camp et al. 2015). Dynamical seasonal forecast model skill depends on the model used, the model resolution, and the intrinsic predictability of the large-scale circulation regimes (Camargo et al. 2005; Bengtsson et al. 2007; Camargo and Barnston 2009). Camargo et al. (2007) suggest that at least part of the reason for poor performance of dynamical models (European Centre for Medium-range Weather Forecasts (ECMWF) and the UK Met Office predict tropical storm frequency based on dynamical models) in predicting tropical storm frequency is due to model error and lack of predictability.

Camp et al. (2015) describe the seasonal forecast skill of the UK Met Office Hadley Centre GloSea5 model, which is a high-resolution dynamical seasonal forecast system with an atmospheric horizontal grid of 0.83° longitude $\times$0.55° latitude ($\sim$53 km at 55°N) and 0.25° in the global ocean. Model skill of tropical storm predictions was evaluated across various basins for the period 1992–2013. While GloSea5 and ERA-Interim have been shown to be remarkably skilful in most basins across the globe, these model systems show no seasonal forecast skill over the NIO region (Camp et al. 2015)—a region notably characterised by a double-peak in tropical cyclone occurrences throughout the annual cycle [maxima during the pre-monsoon (April–May) and post-monsoon (October–November) periods]. Yahyai (2014) describes the ability of the dynamical High Resolution Model (HRM) and Consortium for Small-scale Modelling (COSMO) models in the NIO to predict the probabilities of tropical cyclone landfall, but their performance was insufficient for predicting landfall points compared to those observed, due to limited spatial resolution.

In a separate study, Ray et al. (2012) used an artificial neural network system for the North Indian Ocean region to point out the prediction error between the landfall points of the simulated eastward tracks and the observed tracks. Zhao et al. (2009) explored global atmospheric dynamical models (with a horizontal grid scale of 50 km) to show that tropical cyclone activity (the frequency and interannual variability of tropical cyclones with intensities above a threshold of 33 m/s) are simulated quite realistically in the Atlantic and Northwest Pacific basin. However, the simulation for the NIO is not as close to the observations, with too many storms simulated in the Arabian Sea and too few in the Bay of Bengal. Knutson et al. (2014) concluded that while reasonable results have been found for the North Atlantic and northwest Pacific basins, a substantially larger intensity bias was found for simulations in the NIO region using the Geophysical Fluid Dynamics Laboratory (GFDL) hurricane model (with a horizontal grid scale of 9 km) in a seasonal forecast mode.

Interestingly, statistical models have shown some promise for TC count prediction in the NIO region (Nath et al. 2015). Skilful TC track studies have been previously also demonstrated using statistical forecast models in the western North Pacific (Rumpf et al. 2010), North Atlantic (Emanuel et al. 2006) and South Pacific (James and Mason 2005) regions. There is no guarantee that statistical TC models, developed for the Atlantic and Pacific, will be similarly beneficial in the NIO in a seasonal context, since the seasonal cycle is characterised by the double peak that exists in the pre-monsoon and post-monsoon periods. Compared to other ocean basins across the globe, the Indian Ocean has also received much less attention for statistical seasonal prediction of cyclones (Nath et al. 2015). Therefore, in the first instance, an attempt has been made here to develop a climatological model of the four TC seasons in the NIO region, i.e. that best characterises the winter (December–February), pre-monsoon (March–May), monsoon (June–September) and post-monsoon (October–November) seasons.

This paper discusses the development of a statistical seasonal model of tropical cyclone genesis, tracks and landfall

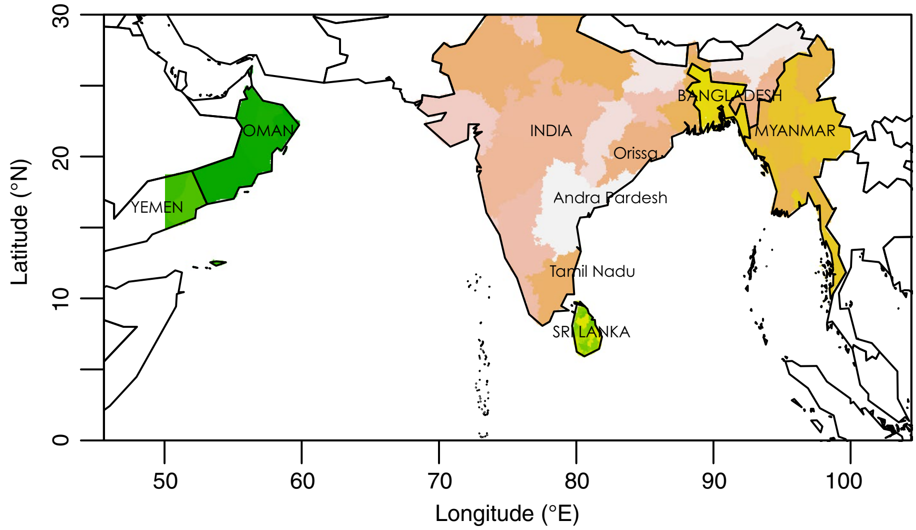

Fig. 1 Geographic map showing the countries (and states) along the North Indian Ocean rim relevant to this study. The six countries affected by tropical cyclone landfall include India, Bangladesh, Myanmar, Sri Lanka, Oman and Yemen

for the North Indian Ocean—that represents a climatological model of tropical cyclones for the region. The genesis model is based on kernel density estimation, and the tracks are estimated using a generalised additive model (GAM). Three hindcast methods are used to cross-validate the model. The structure of the paper is as follows: The data on which the model is based and how the model works are described in Sect. 2. The kernel density estimation for tropical cyclone genesis distribution, the GAM used for cyclone tracks, and country and state-based hazard zone mapping for landfall probabilities, are also described in Sect. 2. Section 3 presents the simulation and cross-validation results from the model, and evaluates these against observed data. Results are discussed in Sect. 4, and some concluding remarks are provided in Sect. 5.

2 Data and methods

2.1 Data from IBTrACS

Tropical cyclone track data for the North Indian Ocean region (Fig. 1) were obtained from the International Best Track Archive for Climate Stewardship (IBTrACS), National Climatic Data Center (www.ncdc.noaa.gov/oa/ibtracs/).

IBTrACS provides global tropical cyclone best track data, in a centralised location, to aid understanding of the distribution, frequency, and intensity of TCs worldwide. In combining track and intensity estimates from many sources, this consolidated collection of TC data provides an extensive global climatology and insight into data uncertainty, which is a critical consideration for climate trending (Knapp et al. 2010; Kossin et al. 2013). Specifically, we have used data from the Joint Typhoon Warning Centre (JTWC) for the 35-year period from 1979 to 2013. The JTWC data are a contributed subset within IBTrACS version 6, which are the data most commonly used by researchers worldwide (Knapp et al. 2010). The parameters used in our analysis of 'storms' include date, position (degrees of latitude and longitude) and wind speed. Note that we also take account of tropical depressions that have not reached tropical cyclone intensity, as well as tropical cyclone intensity storms (i.e. those storms that occurred where wind speeds have reached at least 34 knots, or 17.5 m s^{-1}).

For the genesis model, we determined the time and position along each track that the tropical storm winds first exceeded the critical speed of 34 kt to determine that the storm had reached tropical cyclone intensity. For tropical cyclone pathways in the North Indian Ocean (NIO) region, we note that both tropical depressions and tropical cyclones follow tracks that are not really any different (Warrick et al. 2000; Mohapatra et al. 2012). By including the tracks from both tropical cyclones and storms, we increase the overall sample size [since the historical database of tropical cyclones for the NIO region is not large (Paliwal et al. 2011)] and hence improve the overall statistical representation, based on this assumption that the tracks are not significantly different between tropical storms or tropical cyclones. Notably, we have ignored TCs prior to 1979 to improve reliability of the observational database (Evan and Camargo 2010; Weinkle et al. 2012).

2.2 Methods

Skilful basin-wide statistical models have been developed by several researchers (Gray 1984; Nicholls 1985; Wilks 1995; Nicholls et al. 1998; Elsner and Jagger 2004, 2006, 2010; Camargo et al. 2005; James and Mason 2005; Saunders and

Lea 2005; Emanuel et al. 2006; Hall and Jewson 2007; Holland and Webster 2007; Landsea 2007; Rumpf et al. 2007; Leroy and Wheeler 2008; Jagger and Elsner 2010; Hall and Yonekura 2013; Kang et al. 2016). A crucial component of basin-wide TC modelling is the TC track—i.e. the trajectory from genesis (Hall and Jewson 2007). First, TC genesis modelling can be performed in several ways, including sampling from historical genesis sites (Vickery et al. 2000), interpolating historical genesis locations (James and Mason 2005), binning historical genesis events through a probability density function (PDF) (Emanuel et al. 2006), or the application of a near-neighbour approach to develop and sample a genesis kernel PDF (Rumpf et al. 2007). Second, TC track modelling can also be approached in different ways. For example, Vickery et al. (2000) and James and Mason (2005) used autoregressive models to increment TC track direction. Casson and Coles (2000) draw from the historical tracks, translating TC tracks through small random displacements. Emanuel et al. (2006) propagate tracks by sampling a transition matrix for TC track direction, and Rumpf et al. (2007) described a stochastic model for the separation of tropical cyclone tracks based on the geographic characteristics and applied kernel probability density function and direction increment to simulate the tracks.

Our approach has been to develop a statistical climatological model of TC genesis, tracks and landfall for the NIO, and based on some similar methods employed by Hall and Jewson (2007) for the North Atlantic Ocean and Yonekura and Hall (2011) for the Western North Pacific region. The basic modelling approach comprises of kernel density estimation for TC genesis, application of a generalised additive model for the TC tracks (fitting and baseline for simulation) and Euler integration, and takes account of a country mask for determining points of landfall.

The genesis model developed here closely follows the previous work of Hall and Jewson (2007) for the North Atlantic and Yonekura and Hall (2011) for the western North Pacific. In their research, genesis is sampled from a Poisson distribution and an empirical kernel-density function and genesis has no climate state sensitivity. Our approach fits and utilises a kernel density function for random sampling (using a standard plug in estimator), with the kernel density estimates masked over shallow water and land for the genesis distribution.

TC tracks are modelled as successive 4 step (i.e. 6-h) track displacements each day. The choice of a 6-h increment is consistent with that applied by Hall and Jewson (2007) and Yonekura and Hall (2011). However, in those previous studies, auto regression was applied for the tracks. In a separate study, Mestre and Hallegatte (2009) used a GAM approach to predict the number and intensity of TCs in the North Atlantic. Here, we used a GAM approach instead to model the storm track velocity field in the NIO.

The landfall detection closely follows the work of Weinkle et al. (2012) for the North Atlantic, Eastern and Western Pacific, North Indian Ocean, and the Southern Hemisphere. They used an operational sea surface temperature product as a land mask with 1/20° global grid spacing to detect the first landfall. In the present study, we used a 1/12° mask approach with Euler integration to detect landfall across country and state boundaries.

In the following subsections, we present the approaches that we have applied to model tropical cyclone genesis, tracks and landfall, and also to cross-validate the models.

2.2.1 *Tropical cyclone genesis: kernel density estimation*

In the present study, we have applied a Kernel density approach to model the spatial distribution of observed tropical cyclone genesis locations. This approach has been previously used successfully by, for example, Vickery et al. (2000); James and Mason (2005); Emanuel et al. (2006) and Rumpf et al. (2007). Kernel density estimation is a method to estimate the probability density function (PDF) of a random variable in a non-parametric way. This distribution is defined by a smoothing function and a bandwidth value (length scale) that controls the smoothness of the resulting density curve. Bandwidth selection can be undertaken in various ways, including by "rule of thumb", by cross-validation, by "plug-in" methods, or by other means (Turlach 1993; Bashtannyk and Hyndman 2001).

More formally, Kernel estimators smooth out the contribution of each observed data point over the local neighbourhood of that data point, with the neighbourhood defined by the bandwidth. The contribution of data point $x(i)$ to the estimate at some point x depends on how far apart $x(i)$ and x are situated. If we denote the kernel function as K and its bandwidth by h, the estimated density at any point x is

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - x(i)}{h}\right)$$

In essence, the convolution smoothes out the contribution of each observed data point over a local neighbourhood of that data point. The extent of this contribution is dependent upon the shape of the kernel function, K, and the chosen bandwidth, h. A larger bandwidth increases the region influenced by each x_i, resulting in a smoother estimate.

The kernel K must satisfy the constraint $\int K(x)dx = 1$, but is otherwise arbitrary. The Gaussian kernel,

$$K(x) = (2\pi)^{-1} \exp\left(-\frac{x^2}{2}\right)$$

is a common choice and the one we use here.

In our analysis, we calculate the kernel bandwidth for the kernel density function using a standard plug in estimator. This approach has been shown to be a simple but effective method for estimating relevant bandwidths (Rigollet and Vert 2009) and makes for pragmatic calculations. Hence, kernel density estimation is also used in the present study to model and quantify the probability distribution (function) of observed tropical cyclone genesis locations and occurrences from which we can randomly sample.

The choice of bandwidth is a crucial issue in kernel density estimation. The plug in method is based on a simple idea that gives an estimate ($\hat{f}$) of the bias of the unknown function f. A pilot estimate of f is then "plugged in" to derive an estimate of the bias and hence an estimate of the mean integrated squared error. The optimal bandwidth (h) minimises the error/estimated measure of fit (Loader 1999). Essentially, the kernel density estimator places a Gaussian kernel of specified bandwidth over each observation and sums. Samples are drawn from the estimated density by treating the estimate as a Gaussian mixture—first a component of the mixture is chosen and then a sample deviate is drawn from that component of the mixture.

In the absence of any modification, simple application of the kernel estimator enables TC genesis over land, which is most apparent for longer smoothing (bandwidth) scales. While there are a few isolated cases when TCs have been observed to form over land, such occurrences are largely atypical. To address this issue, we restrict genesis to ocean regions by masking the kernel density function over land areas and also choose to reject genesis in regions where water depths are shallower than some critical depth—we assume this to be for water depths <200 m, i.e. typically across the continental shelf. This process is iterated to generate a sample of the correct size.

For genesis, we determined the time and position along each track that the tropical storm winds exceeded the critical speed of 34 knots, defined as becoming tropical cyclone strength (WMO 1997). It was found that <10% (9.4%) of the estimated genesis locations were either very close to the coast or over land, which were removed by our mask. As a simulation set, we extracted 50 genesis points from the kernel to set up the climatology.

2.2.2 Tropical cyclone tracks: generalised additive model

Cyclone trajectories were modelled by fitting a generalised additive model (GAM) to the incremental track velocities. The GAM is an extension of the generalised linear model (GLM), which in turn is an extension of the standard linear model (LM) (Guisan et al. 2002). GAMs are a generalisation of linear regression in which the linear terms are replaced by smooth transformations of the predictors.

One of the most popular and useful tools in data analysis is the linear regression model or simply the "linear model". Standard regression models assume the response, y, is normally distributed about its mean, μ, with variance σ^2

$$y \sim N\left(\mu, \sigma^2\right).$$

In the regression case, the mean μ, can be modelled as a linear combination of predictor variables, $X_1, X_2, \ldots, X_m$

$$\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m$$

where β_0, β_1, …, β_m are the regression coefficients to be estimated.

The GLM extends the linear model in two key ways. First, the GLM assumes the response may be distributed about its expected value according to any distribution F from any exponential family of distributions (including the Poisson, Binomial and Normal families)

$$y \sim F(\mu)$$

And secondly, that the predictors enter the model through the linear predictor η which is related to the expected response μ by a monotonic function ℓ, called the link function

$$\ell(\mu) = \eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

These extensions relax the restrictions imposed by the linear model on both the distribution of the data and the functional relation between the response and predictors. Note that if F is the Normal family of distributions and the link function ℓ is the identity, then the generalised linear model reduces to the linear model.

The GAM further relaxes the functional relation between the response and the predictors by assuming the linear predictor is related to the predictors through a number of smooth transformations $f_1, f_2, \ldots, f_p$

$$\ell(\mu) = \eta = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_p\left(x_p\right)$$

where a LM or GLM seek to estimate the regression coefficients $\beta_0, \beta_1, \ldots, \beta_p$, a GAM seeks to estimate the smooth functions $f_1, f_2, \ldots, f_p$ of the predictors.

The advantage of the GAM scheme is that it is a data driven approach and can automatically discover relationships in the data. The smooth functions $f_1, f_2, \ldots, f_p$ are not prescribed by any rigid parametric representation, but instead are typically estimated by smoothing.

The disadvantage of the GAM is that it cannot easily represent interactions amongst terms. Rather, it assumes each term contributes additively. To a limited degree, interactions can be modelled by including multivariate smooth terms, that is terms of the form $f_{ij}\left(x_i, x_j\right)$. However, multivariate smooth terms require much greater volume of data to be reliably estimated (Hastie et al. 2009).

In this paper, GAMs are used to model the two components of the velocity field that define the trajectory of the cyclones. First, track velocities are generated by calculating the increment in identifiable tropical storm positions at each time step to estimate the storm track velocities along their tracks. Then, the GAM is fitted to these storm track velocities. A separate GAM is fitted to each velocity component in the x (east) and y (north) directions, to allow the vector velocity field to be predicted. In this simple model we fit each velocity as a smooth function of location in each season.

Cyclone trajectories are simulated using the fitted GAM to predict mean increment velocities at each step along the simulated track, together with random (stochastic) innovations also at each step—starting from randomly selected genesis locations drawn from the kernel for each NIO region TC season. Based on the mean vector field estimated by the GAM fit, an Euler step is used to project trajectories forward in time from this genesis point. A stochastic innovation is added at each time step to account for the variable nature of the vector field that is not captured by the GAM, with a variance determined by the residual error of the GAM fit. For each simulated genesis point, 50 trajectories (and each trajectory has an independent set of innovations applied) were simulated 7 days forward in time with a 6 hourly time step.

We have endeavoured to be both pragmatic with our lifetime choice and consistent with typical TC lifetime scales for the region in our modelling. This choice of a 7-day lifetime for TCs in the model is based on both reported research and our own analysis of historical NIO region TC observations from genesis to landfall. Cyclone lifetimes over the NIO range from 1 to 12 days (1–9 days has been reported by Evan and Camargo (2010) for the Arabian Sea in the period 1979–2008), with an average TC lifetime of 4.8 days, with most TCs making landfall by the 7th day. Almost 80% (79.3%) of NIO region TC lifetimes are within 7 days of genesis, and we therefore consider this to be a reasonable and appropriate choice for our climatological model. Our analysis shows that the model-simulated landfall is 80% for a choice of 7-days for the cyclone lifetime, which is consistent with observations (75%). If we run the model with a lifetime of 12 days the simulated landfall becomes too large (94%) and if we run the model with a lifetime of 2 days then the simulated landfall becomes too small (23%). Overall, by considering the maximum TC lifetime over the NIO and our sensitivity analysis the 7-day choice of lifetime is the most representative.

2.2.3 *Landfall locations*

To determine the points of landfall, each trajectory is interpolated in time to finer increments, with successive points along the trajectory compared to the land mask and the first land mask crossing is taken to be the point of landfall. The estimated precise point of landfall accuracy depends on the resolution to which the trajectory is interpolated. One limitation of this approach is that only the first point of landfall is identified. For a cyclone that passes over the Andaman or Nicobar Islands, this will be the recorded point of landfall, even if the cyclone carries on to possibly strike elsewhere. To functionally solve this problem, we removed the islands from the mask in our analysis (although they appear on the geographical map), so the model simply allows the cyclone track to continue.

From the literature and our own analysis, we note that multiple landfalls are rare in the NIO region; it is almost never seen that a cyclone strikes a country in the NIO rim and then comes back again to strike land a second time (Evan and Camargo 2010; Weinkle et al. 2012; Alam and Dominey-Howes 2015). In the study period, 95.2% of NIO region TCs made landfall (95.7% is reported by Alam and Dominey-Howes (2015) based on their long-term catalogue (from 1000 AD to 2009) of landfall occurrences in surrounding countries of the Bay of Bengal) and none make secondary landfalls. Based on this observation, we have only considered the first point of landfall for this region, while being consistent with other previous research across other basins by Weinkle et al. (2012) where they assume that even in TC multiple landfall cases, only the first point of landfall is counted.

2.2.4 *Model validation*

To assess the integrity and potential utility of the model, we have applied three separate model validation methods. These are introduced as follows.

2.2.4.1 Cross-validation Cross-validation is a resampling procedure where the available data are repeatedly divided into development and verification (prediction) subsets (Wilks 1995). It tries to simulate actual forecasts and provide an accurate estimate of the predictive skill of the model or algorithms (Hess and Elsner 1994; McDonnell and Holbrook 2004). Cross validation endeavours to assess how well the developed model will perform in forecasting the unknown future (Elsner and Schmertmann 1993).

We develop instead a climatological model of tropical cyclones, where the model represents our best fit to the observed seasonal cycle. The model is developed through standard cross-validation procedures, whereby we leave out individual tracks and fit on the remaining data, which is in itself a test of model capability against data not included in the fit. Given that our climatological represents our best estimate of the seasonal cycle, holding out the last 10 years of observations would not be particularly instructive since

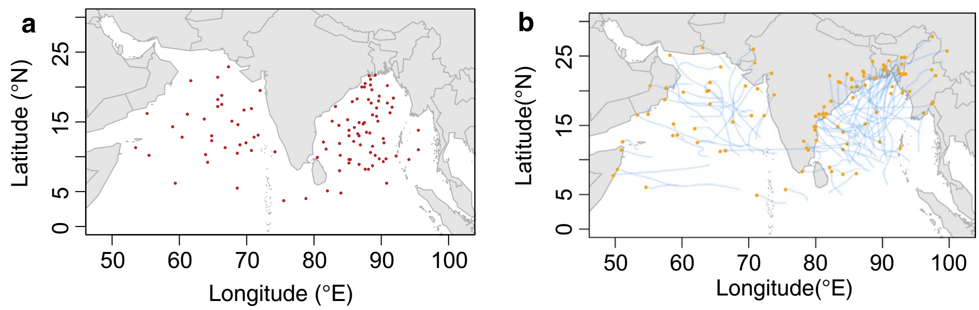

Fig. 2 Observed **a** tropical cyclone genesis locations, and **b** tracks of the corresponding storms that reach tropical cyclone intensity at the genesis points. *Red dots* in **a** identify the tropical cyclone genesis points in the North Indian Ocean region (0°–30° N and 50°–100°E) for the JTWC storm record in the 35-year period from 1979 to 2013. The *blue lines* indicate the tropical cyclone tracks with the last recorded location of the tropical storm intensity tracks indicated as a *yellow dot*

it would compare climatological seasonal forecasts against interannually changing conditions. We point out though that in ongoing separate work, we build forecast model versions using indices of El Niño—Southern Oscillation and the stratospheric Quasi-Biennial Oscillation as predictor variables for forecasting. In those model cases, we indeed do assess forecast skill against independent data and our climatological model forecasts.

Here, we apply leave-one-out cross-validation where we:

1. Remove each storm track in turn, and fit the model to the remaining data;
2. Simulate each of the removed storm trajectories based on the model fit to the reduced data set; and
3. For each simulated track determine the country of landfall, and from all landfalls determine the most likely country of landfall by majority vote (we only identified one state/country where most of the tracks make landfall).

2.2.4.2 Probability distribution Because India has the longest coastal boundary of any country along the NIO rim, India is exposed to higher rates of landfall than other neighbouring countries, such as Bangladesh. As such, a 'majority vote' approach means that India results in greater occurrences of landfall. To address this, we considered an alternative approach whereby we examined the probability distribution of simulated landfall with relation to the observed landfall. Here, we ran the GAM simulation 40 times and averaged the results in turn for each TC season. From the results, we also calculated the standard deviation and compared the simulation with the observed data.

2.2.4.3 Distance between simulated and observed landfall Finally, we evaluated the model by calculating the distance between the observed and simulated landfall points, for tracks that correspondingly originate from the same observed genesis points. Each landfall location, from a simulated track, was measured against the corresponding observed landfall location, and a probability distribution was developed based on the differences.

3 Results

3.1 Model fits of TC genesis and tracks

Figure 2 shows the observed genesis locations (panel a) and tracks (panel b) for the 35-year record of storms contained in the JTWC database for the period of 1979–2013. During this period, a total of 105 storms reached tropical cyclone strength. Most of the TCs in the NIO region originate between 5°N and 15°N and east of 85°E. A greater proportion of TCs formed in the Bay of Bengal (72.4%) compared to the Arabian Sea (27.6%). In the Bay of Bengal, most TCs initially move towards the northwest or north; later on in their lifetime a few have been seen to recurve towards the northeast. In the Arabian Sea, these storms generally move westward. Not all TCs made landfall: 25% decayed to below TC strength while still located over the ocean. In the Bay of Bengal, cyclones typically made landfall along the southeast coast of India, Bangladesh and Myanmar. For the Arabian Sea, TC landfall mainly occurred along the Oman and Yemen coasts, and very few crossed the western coast of India—with most cyclones dying out in the Arabian Sea.

Annually, tropical cyclones form most frequently (44% of annual totals) during the post-monsoon (October–November) season (Fig. 3), whereas winter (December–February) represents the quietest season (15% of annual totals) for tropical cyclones, with only about one third of the post-monsoon season occurrences. The monsoon (June–September) season (18% of annual totals) displays similar numbers to winter while the pre-monsoon season (March–May) contributes slightly more than half of the

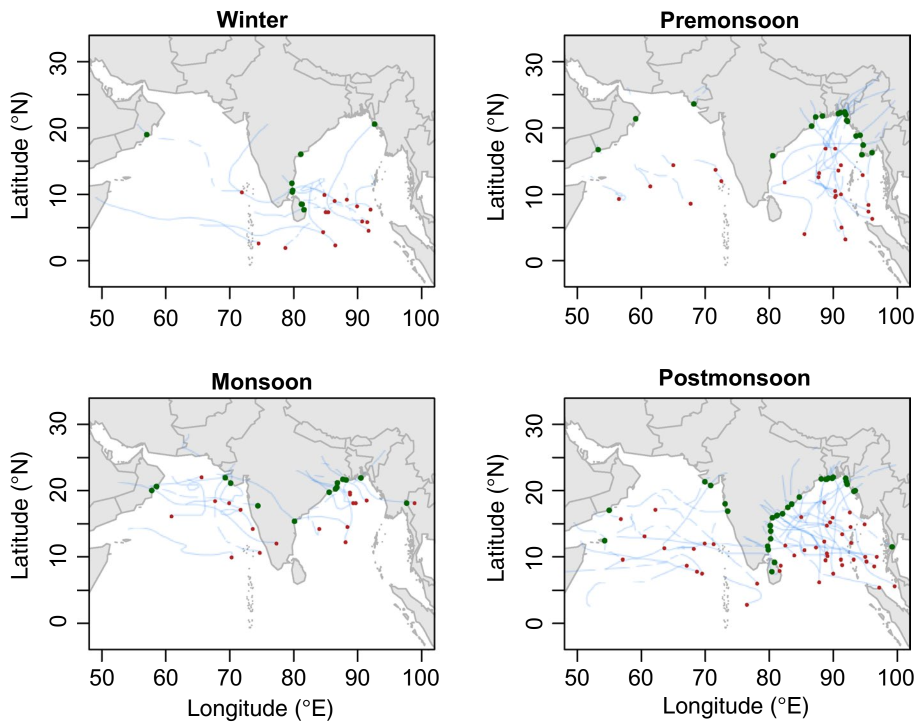

Fig. 3 Seasonal distribution of North Indian Ocean region (0°–30°N and 50°–100°E) observed tropical cyclone genesis locations (identified by *red dots*), tracks (identified by *blue lines*) and landfall locations (identified by *green dots*) over the 35-year period of 1979 to 2013

total number of TCs formed during the post-monsoon season (23% of annual totals). There is a clear latitudinal shift of genesis locations with season. In winter, genesis tends to occur south of 10°N, but occurs north of 10°N in the other seasons. In winter and during the monsoon season, cyclones move in a westward or north-westward direction. In the Bay of Bengal, these storms tend to track westwards and north-westwards into the east coast of India during the post-monsoon season, and north-eastwards into Bangladesh and Myanmar during the pre-monsoon. The distribution of observed storm tracks corresponding to each TC season (Fig. 3) illustrates that most TCs during the post-monsoon season make landfall along the western and north-western coastal fringe of the Bay of Bengal—in particular, the Indian state of Andra Pradesh (see Fig. 1) and the south-west Bangladesh state of Khulna. During the pre-monsoon season, TCs tend to make landfall in the north-eastern part of the Bay of Bengal, including the south-eastern part of Bangladesh and Myanmar. In winter, TCs tend to track towards the Tamil Nadu coast of India (i.e. the south-western fringe of the Bay of Bengal). During the monsoon season, storms tend to track towards the Orissa coast of India (i.e. north-western fringe of the Bay of Bengal).

The modelled distribution of genesis points, approximated by the kernel density estimation, is shown in Fig. 4. Highest density estimates are found in the Bay of Bengal (consistent with the observed genesis locations seen in Fig. 2). By season, highest densities were found in the post-monsoon season followed by the pre-monsoon. In the post-monsoon season the highest density is nearest to the Indian state of Andra Pradesh, while in the monsoon and winter seasons the highest genesis densities are closer to the Bangladesh and Sri Lanka coasts, respectively.

For each country around the NIO rim, the total numbers (and percentage within season) of TCs that make landfall as a function of season are provided in Table 1. Looking along rows (and thus across columns) we can see how landfall rates vary by season for each country ($\%^*$). Alternatively, looking down columns (and thus across rows) we can see how landfall rates vary by country within each season ($\%^\#$). We see that 71% of NIO region TCs make landfall in two of the four TC seasons, annually—these are the post-monsoon (44% of the time) and pre-monsoon (23% of the time) seasons. Further, the pre-monsoon period represents the season with the highest annual number of TC landfall occurrences in Myanmar (50%) and Bangladesh (47%). The most landfall occurrences for India have been observed during the post-monsoon season, with 52% of India's annual tropical cyclone landfalls. For Yemen, there was only one recorded landfall during the entire record, and this occurred during the post-monsoon. For Oman, the seasons with the most landfalls were the monsoon and pre-monsoon seasons (each with 33%). For Sri Lanka, the most landfall occurrences (60%) have been in winter (a total number of 15 (14% of annual total) TCs make landfall).

Among the NIO rim countries, the total numbers, and percentage, of observed tropical cyclones that make landfall as a function of season are also provided in Table 1.

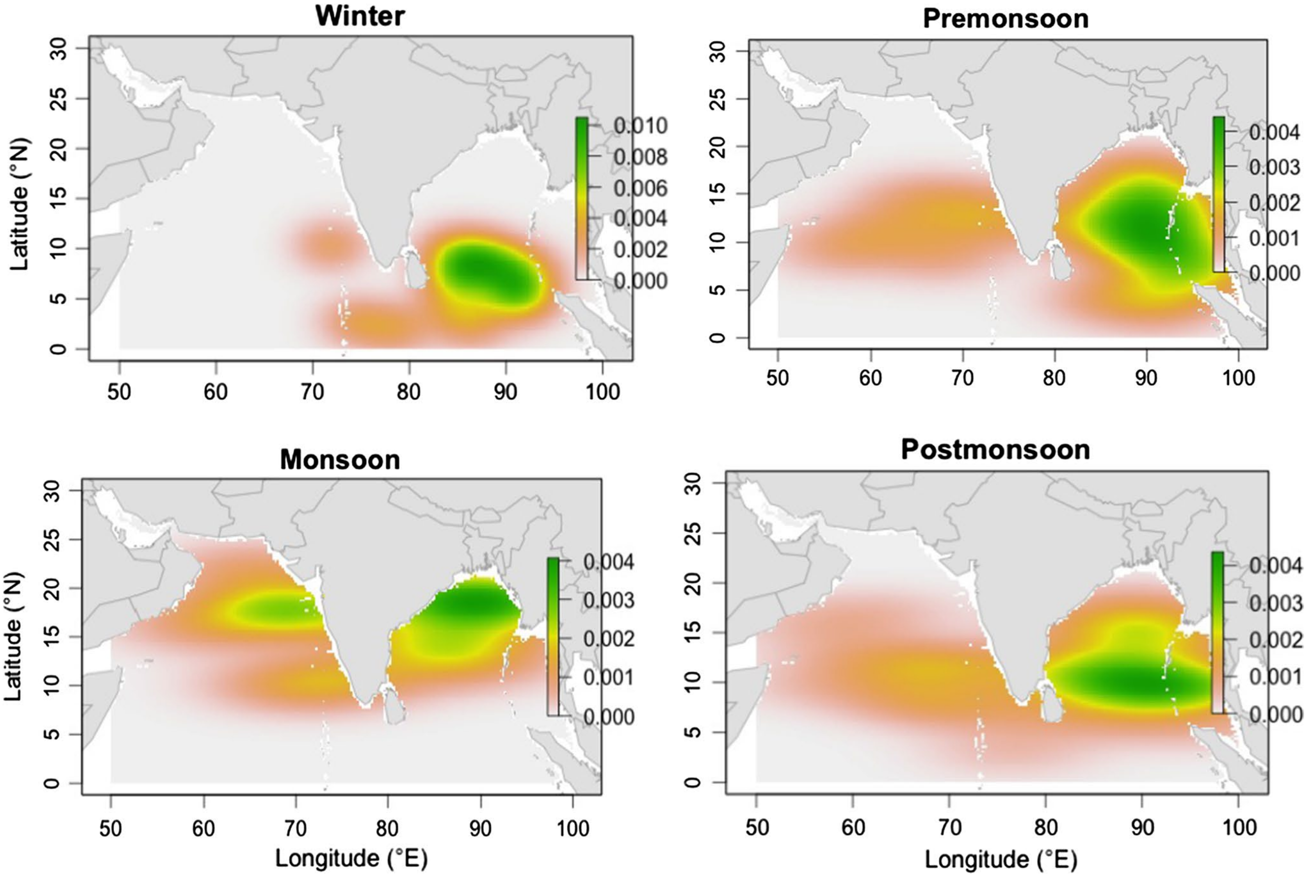

Fig. 4 Modelled seasonal distributions of tropical cyclone genesis based on kernel density estimates across the observations in the North Indian Ocean region (0°–30°N and 50°–100°E) based on the 35-year JTWC data record from 1979 to 2013. *Green colour* shows the highest density (concentration of TC numbers/km^2) area of genesis

Table 1 For each country around the NIO rim, we show here the observed total numbers (and percentage) of tropical cyclones that make landfall as a function of season

Country	Winter			Premonsoon			Monsoon			Postmonsoon			Total	
	No	%*	%#	No	%*	%#	No	%*	%#	No	%*	%#	No	%*
IND	4	10	**26.6**	5	12.5	20.8	10	25	**52.6**	21	**52.5**	**44.7**	**40**	100
MYN	1	10	6.7	5	**50**	20.8	1	10	5.3	3	30	6.4	10	100
BAN	0	0	0	8	**47.1**	**33.4**	1	5.8	5.3	8	**47.1**	17	17	100
SLN	3	**60**	20	0	0	0	0	0	0	2	40	4.3	5	100
YEM	0	0	0	0	0	0	0	0	0	1	**100**	2.1	1	100
OMAN	1	16.7	6.7	2	**33.3**	8.3	2	**33.3**	10.5	1	16.7	2.1	6	100
Landfall	9		60	20		83.3	14		73.7	36		76.6	79	
NON	6	–	40	4	–	16.7	5	–	26.3	11	–	23.4	26	–
TOTAL	15	–	100	24	–	100	19	–	100	**47**	–	100	**105**	–

For each country, the season with highest landfalls in that country is indicated by bold percentages. For each season, the country with highest landfalls in that season is indicated by bold with underlined percentages

IND India, *MYN* Myanmar, *BAN* Bangladesh, *SLN* Sri Lanka, *YEM* Yemen, *OMN* OMAN, *NON* no landfall recorded

* (Across season) and # (across countries)

Among the 6 countries, the highest percentage of landfall occurrences are in the pre-monsoon season (20 out of 24 TCs, ~83%), followed by the post-monsoon season (36 out of 47 TCs, ~76%); whereas 60% (9 TCs make landfall out of 15) of cyclones make landfall in winter and ~74% (14 TCs make landfall out of total 19) make landfall in the

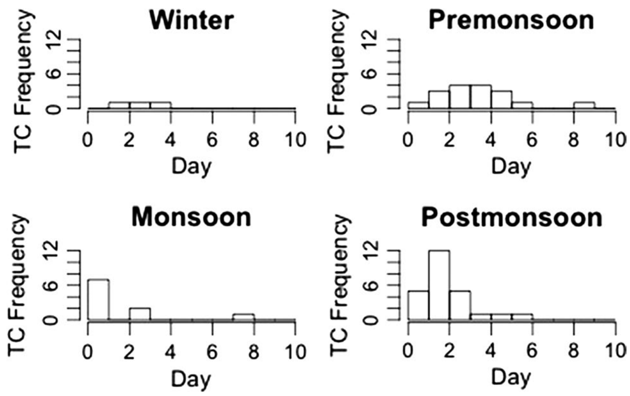

Fig. 5 Observed TC landfall frequency for each season, measured as a function of the number of days from genesis to landfall, within the North Indian Ocean region (0°–30°N and 50°–100°E) over the 35-year record (1979–2013)

monsoon season. Among the 6 NIO countries, almost 45% of cyclone landfalls occur in India during the post-monsoon season whereas Bangladesh (34%) is the favourable zone for TC landfalls in the pre-monsoon season.

We next look at how landfall is distributed with time since genesis, for each TC season (Fig. 5). Landfall tends to occur with greatest likelihood around 1–4 days after genesis in winter (100% of the time) and 2–4 days after genesis during the pre-monsoon (47%). During the monsoon, it tends to occur 0–1 day after genesis (70%) and during the post-monsoon 0–2 days after genesis (68%).

The GAM fit to the observed TC tracks acts to smooth estimates of the TC velocities as a function of space, independently for each season (Fig. 6). In winter, the GAM velocity field indicates a westward or north-westward movement (10°−20°N) of the fitted TC tracks. In the pre-monsoon season, the GAM fit shows a tendency for TCs to move towards the northwest or north (10–20°N), and then recurve towards the northeast (20°N above) in the Bay of Bengal or northwest/north (15–20°N) in the Arabian Sea. During the monsoon season, movement is towards the northwest (8–22°N) in the Bay of Bengal and west and north-westward (12–22°N) over the Arabian Sea. In the post-monsoon season, the initial movement of storms is north-westward (10°–18°N) followed by north-eastward curvature (18°N above).

3.2 Simulated cyclone tracks and landfall locations

Simulations of TC tracks (Fig. 7) indicate how the modelled storms tend to track and where they tend to make landfall. In winter, the modelled TCs tend to track in a westward or north-westward direction over the Bay of Bengal and most of them make landfall along the northern Tamil Nadu coast or the eastern coast of Sri Lanka. Over the Arabian Sea, the storms tend to move towards the west, although some of them move towards the north and northeast. This is generally consistent with observations. During the pre-monsoon season, storms initially move northwest or north over the Bay of Bengal and later recurve towards the northeast and

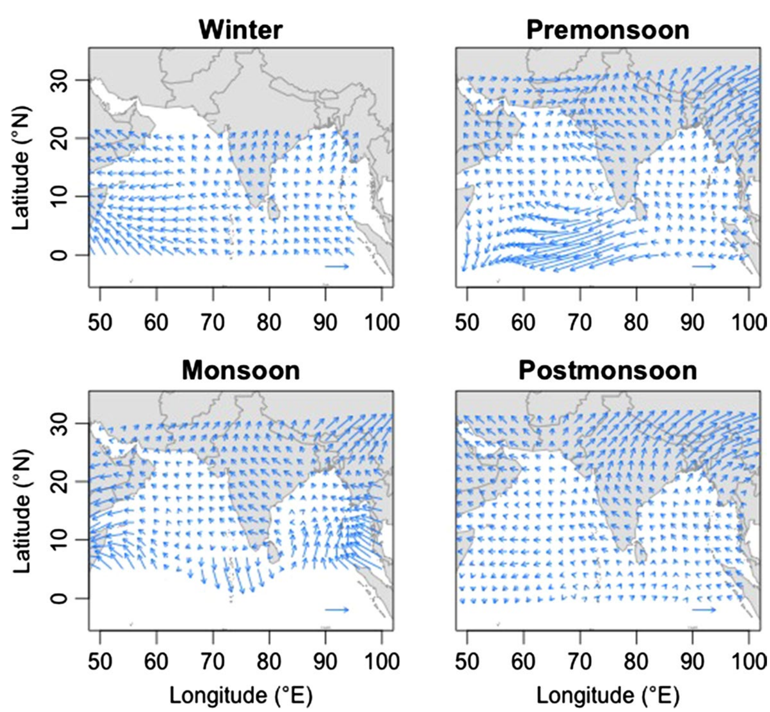

Fig. 6 Seasonal distribution of fitted tropical cyclone track velocities (data from 1979 to 2013) across the North Indian Ocean estimated using a generalised additive model approach. The length of the reference arrow corresponds to velocity magnitudes of 10 m/s

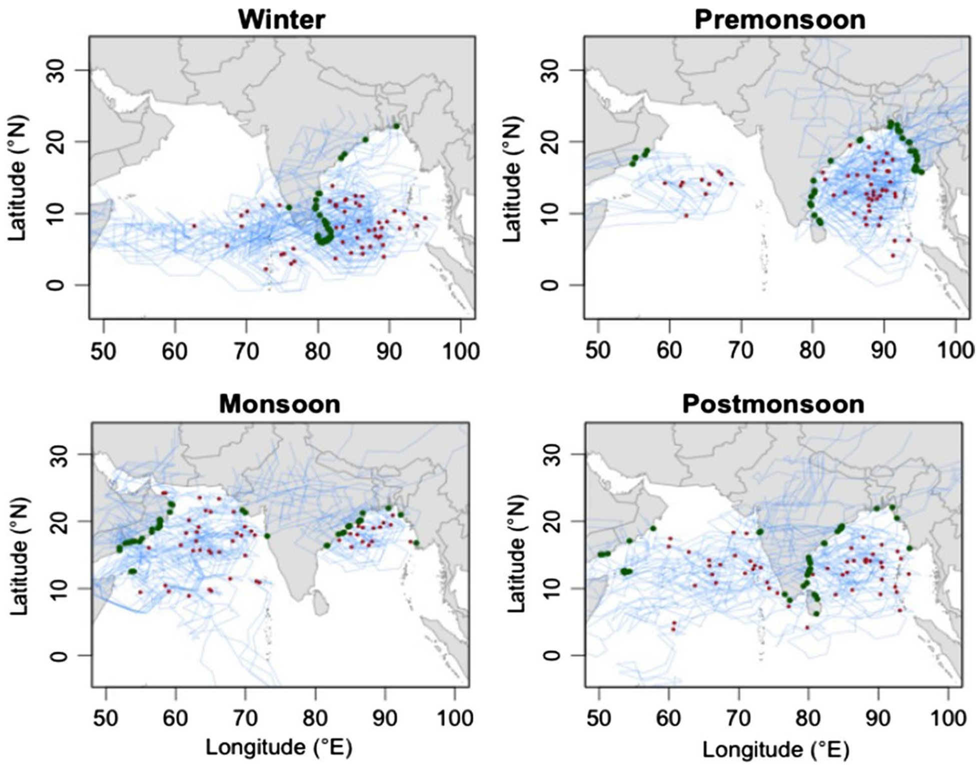

Fig. 7 Simulation of tropical cyclones across the North Indian Ocean region as a function of season. A total of 50 tropical cyclone genesis points were randomly selected from the modelled kernel density and the tropical cyclone tracks were simulated using the fitted generalised additive model combined with the random innovations. *Red dots* show the TC genesis points, *blue lines* indicate the tropical cyclone tracks, and *green dots* indicate landfall locations

Table 2 For each country around the NIO rim, we show here ***<u>the GAM fitted</u>*** total numbers (and percentage) of tropical cyclones that make landfall as a function of season

Country	Winter			Pre-monsoon			Monsoon			Post-monsoon			Total	
	No	%*	%#	No	%*	%#	No	%*	%#	No	%*	%#	No	%*
IND	21	24.1	**<u>42</u>**	21	24.1	**<u>42</u>**	19	21.8	**<u>38</u>**	26	29.9	**<u>52</u>**	**87**	100
MYN	1	4.3	2	13	**56.5**	26	6	26.1	12	3	13.1	6	23	100
BAN	0	0	0	6	**50**	12	1	12.5	2	5	**37.5**	10	12	100
SLN	7	**63.6**	14	4	36.4	8	0	0	0	0	40	0	11	100
YEM	1	16.7	2	1	16.7	2	0	0	0	4	**66.6**	8	6	100
OMAN	0	0	0	3	14.3	6	12	**57.1**	24	6	28.6	12	21	100
Landfall	30		60	48		96	38		76	44		88	160	
NON	20	–	40	2	–	4	12	–	24	6	–	12	40	–
TOTAL	50	–	100	50	–	100	50	–	100	50	–	100	200	–

Maxima are in bold and underlined. Format is the same as in Table 1

IND India, *MYN* Myanmar, *BAN* Bangladesh, *SLN* Sri Lanka, *YEM* Yemen, *OMN* OMAN, *NON* no landfall recorded

* (Across season) and # (across countries)

make landfall along the Arakan coast of Myanmar. Over the Arabian Sea, the simulated storms move initially towards the northwest or north, and then after recurving they move towards the Gujrat-Sind Mekran coast of India (consistent with observations). A greater propensity of storms is seen in the Arabian Sea during the monsoon where a majority of the simulated storms move towards the northwest (consistent with observations) and then recurve towards the north or north-northeast, resulting in increased landfall occurrences in the Middle East, which is also evident to a lesser degree in the post-monsoon season. Finally during the post monsoon season, simulated storms move in a north-westward direction and then recurve towards the northeast (in contrast to the observations in which cyclones move both in north-westward and eastward directions).

The simulated total number (and percentage) of TCs that make landfall as a function of season are provided in Table 2 for each country around the NIO rim. Looking along rows (and thus across columns) we can see how the landfall rates vary by season for a single country (%*). It is seen that 80%

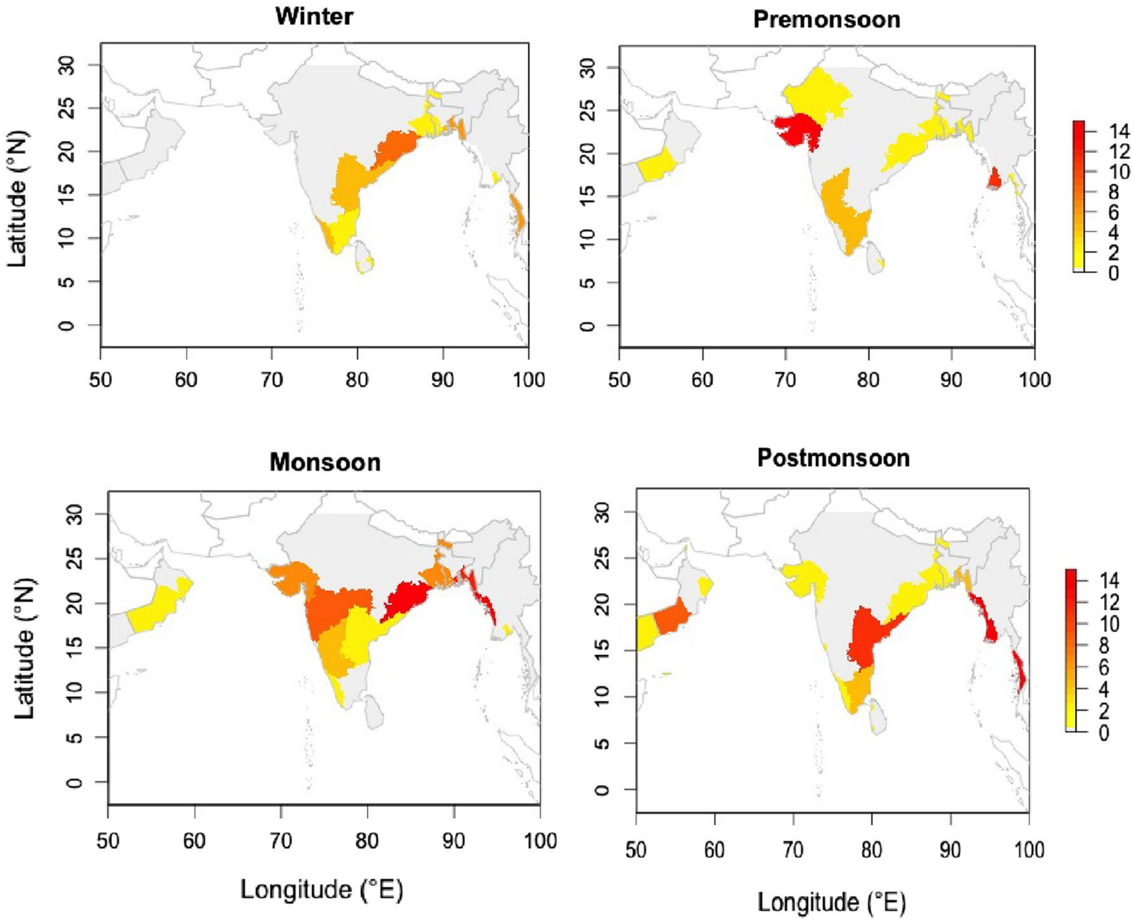

Fig. 8 Geographic distribution of model simulated percentage probabilities of tropical cyclone landfall by state across the NIO rim country boundaries. Colours range from *yellow to red* (according to percentage) corresponding from the lowest to highest model simulated percentage probabilities by TC landfall. *Grey* indicates a percentage of zero landfalls

of the simulated TCs make landfall across the NIO rim, similar to the observed value of 75%. Myanmar (56.5%) and Bangladesh (50%) were predicted to have the largest number of tropical cyclone landfalls during the pre-monsoon season, consistent with observations. The most active period for India is apparent during the post-monsoon season when 29.9% of the country's annual tropical cyclone landfalls occurred. For Yemen, a highest number of 4 TCs (~66.7%) make landfall during the post-monsoon. For Oman, the maximum of 57.1% of annual landfalls occur during the monsoon and the most landfall was in winter for Sri Lanka (63.6%).

By looking down columns (and thus across rows) we can see how landfall rates vary by country within a season (%#). The model simulations predict that India would have the highest proportion (52%) of tropical cyclones that make landfall during the post monsoon season, consistent with the observations. 60% (30 TCs make landfall out of 50) of cyclones make landfall in winter and 76% (38 TCs make landfall out of 50) in the monsoon season. Among the 6 NIO countries, Myanmar (56.5%) and Bangladesh (50%) is the favourable zone for landfalls in the pre-monsoon season.

Geographically, the simulated percentages of TC landfall as a function of season across the NIO rim countries are presented in Fig. 8. During winter, the model simulates the highest proportion of TC landfall occurrences for the states of Tamil Nadu and Andra Pardesh in India and across a few states in Bangladesh. In the pre-monsoon season, the states of Ayeyarwady, Rakhine and Tanintharyi in Myanmar, states of Oman, and the western Indian state of Orissa, are all predicted to have the largest proportion of TC landfalls relative to other NIO rim countries. During the post-monsoon, the Indian states of Andra Pradesh, Orissa and Tamil Nadu, and the coastal states of Myanmar, are predicted to have the largest proportion of TC landfalls during that season. In the monsoon season, Orissa in India, Chittagong in Bangladesh, and Myanmar states are predicted to have the highest proportion of TC simulated landfall occurrences,

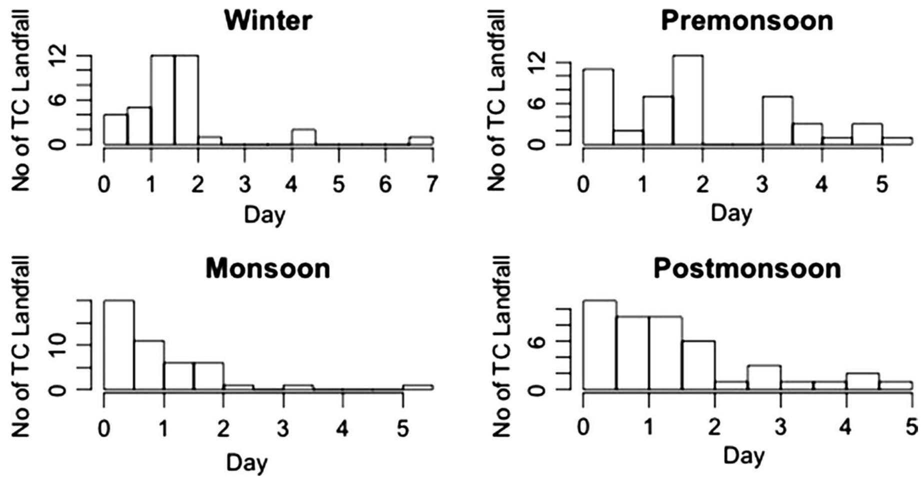

Fig. 9 Model simulated TC landfall frequency for each season, measured as time estimated in the number of days from genesis to landfall, for the North Indian Ocean region (0°–30°N and 50°–100°E)

Table 3 Number and percentages of observed and simulated annual probabilities of North Indian Ocean region tropical cyclones that landfall across each country coastline

Country	Observation		Simulation	
	No	%	No	%
India	40	38.1	57	57
Myanmar	10	9.5	10	10
Bangladesh	17	16.2	0	0
Sri Lanka	5	4.8	4	4
Oman	6	5.7	7	7
Yemen	1	0.9	4	4
Total landfall	79	75.2	82	82
No landfall	26	24.8	18	18
Total	105	100	100	100

Table 4 Number and percentages of annual total North Indian Ocean region observed and simulated tropical cyclones that landfall across each country coastline

Country	Observation		Simulation	
	No	%	No	%
India	40	38.1	61	61
Myanmar	10	9.5	7	7
Bangladesh	17	16.2	0	0
Sri Lanka	5	4.8	3	3
Oman	6	5.7	8	7
Yemen	1	0.9	1	1
Total landfall	79	75.2	80	80
No landfall	26	24.8	20	20
Total	105	100	100	100

In this analysis, vector fields depend only on season

with elevated numbers also predicted for Al Wusta, Ash Sharqiyah and Dhofar in Oman.

The highest percentage of modelled landfall occurrences are seen in 0–2 days after genesis in the winter (80%) and pre-monsoon (69.5%) seasons, whereas for the monsoon season (62.5%) it occurs in 0–1 day and within 0–2 days during the post monsoon season (74.4%) (Fig. 9). These simulated landfall times match well with the observations, except in the pre-monsoon season where the highest landfall rates occur in 2–4 days after genesis.

3.3 Prediction skill of model measured against observations

Leave-one-out cross validation (LOOCV) is performed by removing each track in turn and then simulating the trajectories corresponding to the deleted tracks based on the reduced data set, but from the observed genesis point corresponding to the removed track. For each simulated track, we predict the most likely country and point of landfall, by the majority vote from the simulated tracks. Based on comparison of the percentages of country-based statistics of cyclone landfalls between the observations and the model simulations, it was found that the model performs very well overall (except Bangladesh) according to the majority vote approach (Table 3).

The highest percentages of both observed and simulated landfall totals are in India—38 and 57%, respectively. The observed percentage (9.5%) for Myanmar is remarkably well simulated by the model (10%). The percentage difference in total landfall occurrences (Table 3) between the observations and simulated values is smallest for Myanmar (0%) and largest for Bangladesh (100%). The only exception to overall good model performance was for Bangladesh where the model fails to simulate landfall (Table 3 and 4). With Bangladesh's coastline being a relatively small target for tropical cyclones to strike (239 km long coastline, as compared with 7517 km for India), using a majority vote approach, where we decide on a single state/country that

Table 5 Results of the second model validation technique. Percentage frequency of total observed and predicted tropical cyclone landfall occurrences across country coastlines a) for all months, and b) for each season. Errors represent the standard deviation about the mean

(a) Country	Observation		Prediction mean with error
	No	%	%
India	40	**38.1**	**37.4** ± 17.1
Myanmar	10	9.5	13.4 ± 13.5
Bangladesh	17	16.2	4.7 ± 4.9
Sri Lanka	5	4.8	7.2 ± 6.4
Oman	6	5.7	12.8 ± 9.6
Yemen	1	0.9	3.5 ± 3.5
Total landfall	79	**75.2**	**79**
No landfall	26	24.8	21
Total	105	100	100

(b) Country	Observation				Prediction mean with error			
	Win	Pre	Mon	Post	Win	Pre	Mon	Post
	No (%)	No (%)	No (%)	No (%)	%	%	%	%
IND	4 (26.6)	5 (20.8)	10 (52.6)	21 (44.7)	32.7 ± 20.9	33.2 ± 14.4	39.6 ± 17.8	43.9 ± 15.4
MYN	1 (6.7)	5 (20.8)	1 (5.3)	3 (6.4)	8.9 ± 14.8	25.9 ± 18.4	4.9 ± 6.1	13.8 ± 14.7
BAN	0 (0)	8 (33.4)	1 (5.3)	8 (17)	2.2 ± 4.8	6.6 ± 5.3	3.6 ± 3.7	6.4 ± 5.6
SLN	3 (20)	0 (0)	0 (0)	2 (4.3)	20.9 ± 16.3	3.3 ± 3.8	0.9 ± 1.6	3.8 ± 4
OMN	1 (6.7)	2 (8.3)	2 (10.5)	1 (2.1)	0.9 ± 1.7	10.8 ± 7.1	30.7 ± 20.4	8.9 ± 9.2
YEM	0 (0)	0 (0)	0 (0)	1 (2.1)	1.9 ± 2.5	4.4 ± 3.6	3 ± 4.4	4.5 ± 3.4
Landfall	9 (60)	20 (**83.3**)	14 (73.7)	36 (**76.6**)	67.5	**84.2**	82.7	**81.3**
NON	6 (40)	4 (16.7)	5 (26.3)	11 (23.4)	32.5	15.8	17.3	18.7
Total	15 (100)	24 (100)	19 (100)	47 (100)	100	100	100	100

Bold numbers highlight the best match for the NIO rim

represents where most of the tracks make landfall, appears to be an inadequate approach.

Assuming the vector fields depend only on season, the Table 3 contents analysis (vector fields depend on smooth function of longitude and latitude and the function varies with season) is extended to produce Table 4 where we see the highest observation (40%) and simulation percentages (61%) are shown in India and the lowest in Yemen (0.9%) across the NIO rim countries—except for Bangladesh which again fails due to the majority vote approach. The percentage difference in total landfall occurrences (Table 4) between the observations and simulated values is smallest for Yemen (0%) and largest for Bangladesh (100%).

Following the second model validation approach (introduced in Sect. 2.2.4.2), we examined the probability distribution of simulated landfall occurrences against the observed landfall occurrences. The model was run 40 times and the statistical mean plus error was compared against the observations (Table 5a, b). This means that we have an ensemble of tracks, and therefore landfall locations, from which to build a distribution and compare its mean and spread against the observations. It is seen (Table 5a) that the percentage of simulated landfall occurrences (37.4%) matches well with the observed percentage (38.1%) of landfall occurrences for India. The observed values for all the other countries lie within the spread of the simulated ensemble. The only exception is Bangladesh, where the simulated probability is 0–9.6% but the observed percentage is 16.2%. When examining the results based on seasonal model fits (Table 5b) it is found that the predicted landfall probability with mean error for Bangladesh in the monsoon season (3.6%) is similar to the observed. However, as a whole, the model performance for Bangladesh is poor across the other TC seasons. For example, the simulated landfall probability in the pre-monsoon season is only 6.6% compared to 33.4% for the observations. The model also over predicts landfall rates for Yemen in all seasons and for Myanmar in the post-monsoon. Overall the percentage of simulated landfall occurrences against the observed landfall occurrences matches well for all seasons, especially the pre-monsoon and post-monsoon periods (Table 5b).

In the third model validation approach (introduced in Sect. 2.2.4.3), we calculated the distance between observed

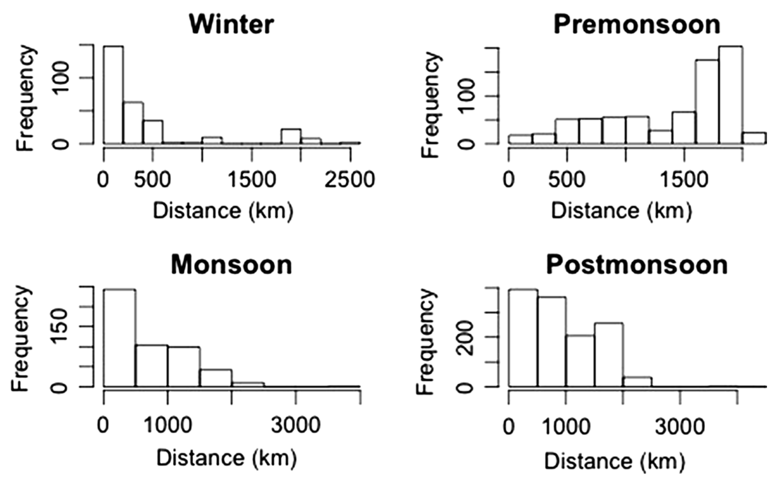

Fig. 10 Distribution of distances between observed and simulated landfall location over the NIO as per season

landfall and simulated landfall occurrences to evaluate model performance. The distribution of distances between the observed and simulated landfall points across all simulated tracks is presented in Fig. 10. It is found that the majority of simulated TC landfall occurrences (54.6%) occur within 0–500 km of the observed landfall locations, with the best model performances found in the winter, monsoon and post-monsoon seasons. For the pre-monsoon season, there is a greater spread of differences between modelled landfall locations and those observed, with the majority difference being at just over 1500 km from the observed. Given the spatial scale of tropical cyclones is typically O(1000 km), modelled landfall occurrences within ~500 km of the observations would suggest very good model performance, while those falling within ~1000 km should still be deemed as beneficial and effectively skilful.

4 Discussion

This study has described a new climatological statistical model of tropical cyclone genesis, tracks and landfall for the North Indian Ocean region. This climatological model is intended to represent a model baseline against which future TC forecast models might be developed and assessed for their seasonal forecast skill on year-to-year or longer time scales. The model presented here simulates tropical cyclone genesis, tracks and landfall as a function of season only, and takes no account of the influence of possible climatic prediction, or predictor, influences on year-to-year or longer time scales, such as the possible effects of El Niño—Southern Oscillation or the Indian Ocean Dipole. Hence, the model represents a climatological model for TCs in the North Indian Ocean region.

So far, little skill has been shown in dynamical (climate) model seasonal forecasts of tropical cyclones for the NIO region (Camargo et al. 2007; Camp et al. 2015). Specifically, Camargo et al. (2005) and Shaevitz et al. (2014) noted that both low and high-resolution models are generally unable to properly simulate the seasonal cycle. We have instead applied a statistical modelling approach here, where the tropical observed cyclone genesis is modelled by kernel density estimation, producing a probability density function (PDF) from which individual storms can be randomly sampled. Tracks are fitted using a generalised additive model, and model track simulations use random innovations at spatial increments to produce an ensemble of simulated tracks. The model is cross-validated in three separate ways to assess and ensure reliability of the results.

The simulation of a large number of tracks with an implementation of this model enables the calculation of landfall probabilities at any location of interest in the North Indian Ocean region affected by tropical cyclones. In our 35-year study period (1979–2013), a total number of 105 tropical cyclones (TCs) have been reported and are recorded in the JTWC database (within IBTrACS). Of these, 79 (75%) made landfall across six North Indian Ocean (NIO) rim countries, namely Bangladesh, India, Myanmar, Sri Lanka, Oman and Yemen. Almost half of the TCs that made landfall, have done so during the post-monsoon season, while 25.3% (20) made landfall during the pre-monsoon season—a total of 71% making landfall during only these two of the four seasons annually. Specifically, 50% of annual TC landfall events into Myanmar have occurred during the pre-monsoon season, while 47% of total annual TC landfall events also occurred during this same season in Bangladesh. Conversely, 52% of landfall events into India have occurred during the post-monsoon season. In terms of time from genesis to landfall, it was found that it takes around 2–4 days since genesis for 47% of events to make landfall during the pre-monsoon season and from 0 to 2 days (68% of events) during the post-monsoon season. Correspondingly, the model also predicts that the highest percentage of TC landfall events should occur in the pre-monsoon season for Myanmar and in the post-monsoon season for India. However, the model failed to predict the highest number of landfall occurrences in the pre-monsoon season for Bangladesh.

We have chosen three methods to cross-validate this statistical model. First, leave-one-out cross-validation was applied. We have shown that the percentages of annual total NIO region observed and simulated TCs that made landfall across each country coastline compare well. It was found that the model performs best in this sense for Myanmar, with the smallest prediction error of 0.5% difference. On the other hand, the worst performance is identified for Bangladesh where the model fails to predict. The cross validation analysis was repeated by assuming that the vector field depends on only season and then the model predicts well for Yemen and Myanmar, but again failed for Bangladesh due to the majority vote approach.

The second cross-validation approach considers the probability distribution of the simulated tracks, where the model is run several times and the statistical mean and error (standard deviation) are generated to assess whether the model results capture the percentage of observed landfalls within an error bound (standard deviation). The standard deviation represents the model error ($\pm$) around the model-predicted mean value, which encompasses the observed percentages of landfall occurrences for most of the six countries in the NIO rim. Exceptions are for Myanmar during the post-monsoon season (the model predicts 13.8% whereas the observations correspond to 6.4%) and Bangladesh in the pre-monsoon season (the model predicts 6.6% whereas the observations correspond to 33.4%).

The third and final cross-validation approach calculates the distance between observed and simulated landfall points, and reveals that the majority of modelled TC landfall occurrences are within 500 km of the observed landfall locations. This confirms the model skill, given that the typical spatial scale of tropical cyclones [O(1000 km)] is much larger than the typical error in landfall location. Importantly, a forecast with an error of 500 km is useful information as compared with no information, which would be the case in the absence of a forecast model. Further, it is notable that the NIO basin is difficult to model due to the double peak in the seasonal cycle (Camargo et al. 2005; Shaevitz et al. 2014; Camp et al. 2015). An error in predicted landfall location of 500 km is similar to the errors of dynamical models for the region, which have errors on the order of 100 s of km (Mallik et al. 2015; Rayhun et al. 2015).

One limitation of the statistical modelling approach is the lower sharpness (the tendency of the forecast to predict extreme values) that comes with such models (Vitart et al. 2010). Statistical models rarely predict very low or very high probabilities largely because of the constraints provided by climatology (Slade and Maloney 2013). In other words, we are examining the expected seasonal state rather than the extreme events. There is also room for the TC genesis kernel density estimation to be better tuned to the observations. Specifically, the kernel does not capture the genesis locations particularly well near Sri Lanka. Future improvements to the genesis model may translate to a bias reduction in the track simulations and thus landfall estimates.

Our model presented here provides the climatological basis against which future developments of statistical forecasting applications, given climate predictors on interannual to decadal time scales, can be tested for their skill. In particular, the Quasi-Biennial Oscillation and El Niño—Southern Oscillation are potentially relevant climate modes of variability to the North Indian region tropical cyclone activity (Girishkumar and Ravichandran 2012; Fadnavis et al. 2014). These climate modes and skill in seasonally forecasting tropical cyclone activity afforded by the model developed on other relevant predictor variables will be investigated and discussed in separate studies.

5 Conclusions

The key aim of this paper has been to develop a climatological statistical model of tropical cyclone genesis, tracks and landfall for the North Indian Ocean region and its neighbouring rim countries that will be beneficial for forecast models to be compared against. The main findings are as follows:

1. An effective method has been developed to estimate the distribution of tropical cyclone genesis points by kernel density estimation. This paper demonstrates that the masked kernel density estimates well the observed genesis points within each season.
2. A novel generalised additive model (GAM) has been introduced and fitted to track increments and used as a baseline for predicting the tropical cyclone velocity field in each season. In the winter and monsoon seasons, the GAM-fitted velocity field highlights the westward or north-westward cyclone movement during this time. Conversely, during the pre-monsoon season, the model fit highlights the north-westward/northward movement, and north-eastward recurvature. In the post-monsoon season, the GAM shows the cyclone movement is typically in a north-westward direction which later recurves towards the northeast.
3. The observations indicate the highest percentage of landfall occurrences in the pre-monsoon seasons are

found in Myanmar (50%) and Bangladesh (47%), while for the post-monsoon season the largest percentage of landfall occurs in India (52%). Our model shows characteristically similar results with simulated highest percentage landfall occurrences for Myanmar (56.5%) and Bangladesh (50%) in the pre-monsoon season, and for India (30%) in the post-monsoon.

4. The characteristic number of days from tropical cyclone genesis to landfall can provide useful time scale information for coastal planning and emergency response. We determined the time since genesis that landfall occurs, and found the percentage in the observations and simulations are characteristically 2–4 days for the pre-monsoon season and 0–2 days for the post-monsoon season.
5. From analysis of the results by season, we find the highest number of model simulated cyclones move eastward during the pre-monsoon season and make landfall around the eastern part of Bay of Bengal, whereas they move westward making landfall in the western part of Bay of Bengal/eastern coast of India during the post-monsoon.
6. To validate the quality of the model hindcasts against the observed landfall occurrences across the NIO countries, three validation methods were introduced [(1) leave-one-out cross-validation, (2) probability distribution of country of landfall, and (3) distance between observed and simulated landfall locations], collectively demonstrating that the model performs well as a climatological baseline.

In summary, we have developed a new and skilful seasonal climatological model of tropical cyclone genesis, tracks, and landfall for the North Indian Ocean region. It is found that our model produces skilful hindcasts of these tropical cyclone characteristics. An important outcome of this paper is the development of seasonal TC genesis distributions using kernel density estimation, and trajectory simulation and landfall detection using a generalised additive model with stochastic innovations that impinge on a country mask. Efforts are currently underway to adapt the presented model approach to develop a statistical seasonal forecasting model that incorporates interannual climate predictors, including simple metrics for the stratospheric Quasi-Biennial Oscillation or El Niño—Southern Oscillation.

Acknowledgements We would like to sincerely thank the two anonymous reviewers for their insightful comments that helped us to significantly improve the quality of this manuscript. Mohammad Wahiduzzaman was supported by a Tasmania Graduate Research Scholarship (TGRS) for this PhD research undertaken at the University of Tasmania, Hobart, Tasmania, Australia. This paper makes a contribution to the objectives of the Australian Research Council Centre of Excellence for Climate System Science (ARCCCS).

References

Alam E, Collins AE (2010) Cyclone disaster vulnerability and response experiences in coastal Bangladesh. Disasters 34(4):931–954

Alam E, Dominey-Howes D (2015) A new catalogue of tropical cyclones of the northern Bay of Bengal and the distribution and effects of selected landfalling events in Bangladesh. Int J Climatol 35(6):801–835

Alam MM, Hossain MA, Shafee S (2003) Frequency of Bay of Bengal cyclonic storms and depressions crossing different coastal zones. Int J Climatol 23(9):1119–1125

Balaguru K, Taraphdar S, Leung LR, Foltz GR (2014) Increase in the intensity of postmonsoon Bay of Bengal tropical cyclones. Geophys Res Lett 41(10):3594–3601

Bashtannyk DM, Hyndman RJ (2001) Bandwidth selection for kernel conditional density estimation. Comput Stat Data Anal 36:279–298

Belanger JI, Webster PJ, Curry JA, Jelinek MT (2012) Extended prediction of North Indian Ocean tropical cyclones. Weather Forecast 27(3):757–769

Bengtsson L, Hodges KI, Esch M (2007) Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analyses. Tellus Ser A Dyn Meteorol Oceanogr 59A(4):396–416

Camargo SJ (2013) Global and regional aspects of tropical cyclone activity in the CMIP5 models. J Clim 26(24):9880–9902

Camargo SJ, Barnston AG (2009) Experimental dynamical seasonal forecasts of tropical cyclone activity at IRI. Weather Forecast 24(2):472–491

Camargo SJ, Barnston AG, Zebiak SE (2005) A statistical assessment of tropical cyclone activity in atmospheric general circulation models. Tellus Ser A Dyn Meteorol Oceanogr 57(4):589–604

Camargo SJ, Barnston AG, Klotzbatch PJ, Landsea CW (2007) Seasonal tropical cyclone forecasts. Bull World Meterol Org 56:297–307

Camp J, Roberts M, Maclachlan C, Wallace E, Hermanson L, Brookshaw A, Arribas A, Scaife AA (2015) Seasonal forecasting of tropical storms using the Met Office GloSea5 seasonal forecast system. Q J R Meteorol Soc 141(691):2206–2219

Casson E, Coles S (2000) Simulation and extremal analysis of hurricane events. J R Stat Soc Ser C Appl Stat 49(2):227–245

Elsner JB, Jagger TH (2004) A hierarchical Bayesian approach to seasonal hurricane modeling. J Clim 17(14):2813–2827

Elsner JB, Jagger TH (2006) Prediction models for annual U.S. hurricane counts. J Clim 19(12):2935–2952

Elsner JB, Jagger TH (2010) Hurricanes and climate change. Springer, US, pp 1–419

Elsner JB, Schmertmann CP (1993) Improving extendedrange seasonal predictions of intense Atlantic hurricane activity. Weather Forecast 8:345–351

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

Evan AT, Camargo SJ (2010) A climatology of Arabian sea cyclonic storms. J Clim 24(1):140–158

Fadnavis S, Ernest Raj P, Buchunde P, Goswami BN (2014) In search of influence of stratospheric Quasi-Biennial Oscillation on tropical cyclones tracks over the Bay of Bengal region. Int J Climatol 34(3):567–580

Girishkumar MS, Ravichandran M (2012) The influences of ENSO on tropical cyclone activity in the Bay of Bengal during October–December. J Geophys Res Oceans 117:C02033

Gray WM (1984) Atlantic seasonal hurricane frequency, Part I: El Nino and 30 mb Quasi-Biennial Oscillation influences. Mont Weather Rev 112(9):1669–1683

Guisan A, Edwards TC Jr, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol Model 157(2–3):89–100

Haggag M, Badry H (2012) Hydrometeorological modeling study of tropical cyclone Phet in the Arabian Sea in 2010. Atmosp Clim Sci 2:174–190

Hall TM, Jewson S (2007) Statistical modelling of North Atlantic tropical cyclone tracks. Tellus Ser. A Dyn Meteorol Oceanogr 59A(4):486–498

Hall T, Yonekura E (2013) North American tropical cyclone landfall and SST: a statistical model study. J Clim 26(21):8422–8439

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd Edition. Springer, pp 295–333

Hess JC, Elsner JB (1994) Extended-range hindcasts of tropical-origin Atlantic hurricane activity. Geophys Res Lett 21(5):365–368

Holland GJ, Webster PJ (2007) Heightened tropical cyclone activity in the North Atlantic: natural variability or climate trend? Philos Trans R Soc A Math Phys Eng Sci 365(1860):2695–2716

Islam T, Peterson RE (2009) Climatology of landfalling tropical cyclones in Bangladesh 1877–2003. Nat Hazards 48(1):115–135

Jagger TH, Elsner JB (2010) A consensus model for seasonal hurricane prediction. J Clim 23(22):6090–6099

James MK, Mason LB (2005) Synthetic tropical cyclone database. J Waterw Port Coast Ocean Eng 131(4):181–192

Kang NY, Lim MS, Elsner JB, Shin DH (2016) Bayesian updating of track-forecast uncertainty for tropical cyclones. Weather Forecast 31(2):621–626

Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ (2010) The international best track archive for climate stewardship (IBTrACS). Bull Am Meteorol Soc 91(3):363–376

Knutson TR, Zeng F, Wittenberg A, Kim SH, Sirutis J, Bender M, Zhao M, Tuleya R (2014) Recent research at GFDL on surface temperature trends and simulations of tropical cyclone activity in the Indian Ocean Region. In: Mohanty UC, Mohapatra M, Singh OP, Bandyopadhyay BK, Rathore LS (eds) Monitoring and prediction of tropical cyclones in the indian ocean and climate change. Springer, New York, pp 50–62

Kossin JP, Olander TL, Knapp KR (2013) Trend analysis with a new global record of tropical cyclone intensity. J Clim 26(24):9960–9976

Landsea CW (2007) Counting Atlantic tropical cyclones back to 1900. EOS 88(18):197–202

Leroy A, Wheeler MC (2008) Statistical prediction of weekly tropical cyclone activity in the southern hemisphere. Mon Weather Rev 136(10):3637–3654

Lin II, Chen CH, Pun IF, Liu WT, Wu CC (2009) Warm ocean anomaly, air sea fluxes, and the rapid intensification of tropical cyclone Nargis (2008). Geophys Res Lett 36(3):L03817

Loader CR (1999) Bandwidth Selection: classical or plug-in? Ann Stat 27(2):415–438

Mallik MAK, Ahsan MN, Chowdhury MAM (2015) Simulation of track and landfall of tropical cyclone Viyaru and its associated strom surges using NWP models. Am J Mar Sci 3(1):11–21

McDonnell KA, Holbrook NJ (2004) A poisson regression model of tropical cyclogenesis for the Australian-southwest Pacific Ocean region. Weather Forecast 19(2):440–455

McPhaden MJ, Foltz GR, Lee T, Murty VSN, Ravichandran M, Vecchi GA, Vialard J, Wiggert JD, Yu L (2009) Ocean-atmosphere interactions during cyclone nargis. EOS 90(7):53–54

Mestre O, Hallegatte S (2009) Predictors of tropical cyclone numbers and extreme hurricane intensities over the North Atlantic using generalized additive and linear models. J Clim 22(3):633–648

Mohapatra M, Bandyopadhyay BK, Tyagi A (2012) Best track parameters of tropical cyclones over the North Indian Ocean: a review. Nat Hazards 63(3):1285–1317

Mohapatra M, Bandyopadhyay BK, Tyagi A (2014) Status and plans for operational tropical cyclone forecasting and warning systems in the North Indian Ocean region. In: Mohanty UC, Mohapatra M, Singh OP, Bandyopadhyay BK, Rathore LS (eds) Monitoring and prediction of tropical cyclones in the Indian Ocean and climate change. Springer, New York, pp 149–162

Mydans S, Cowell A (2008) Myanmar mourns victims of cyclone. New York Times, New York

Nath S, Kotal SD, Kundu PK (2015) Seasonal prediction of tropical cyclone activity over the North Indian Ocean using the neural network model. Atmosfera 28(4):271–281

Ng EKW, Chan JCL (2012) Interannual variations of tropical cyclone activity over the north Indian Ocean. Int J Climatol 32(6):819–830

Nicholls N (1985) Predictability of interannual variations of Australian seasonal tropical cyclone activity. Mon Weather Rev 113(7):1144–1149

Nicholls N, Landsea C, Gill J (1998) Recent trends in Australian region tropical cyclone activity. Meteorol Atmos Phys 65(3–4):197–205

O'Hare G (2001) Hurricane 07B in the Godivari Delta, Andhra Pradesh, India: vulnerability, mitigation and the spatial impact. Geogr J 167(1):23–38

Paliwal M, Patwardhan A, Sarda NL (2011) Analyzing tropical cyclone tracks of North Indian Ocean. In: COM. Geo 11 proceedings of the 2nd international conference on computing for geospatial research & applications, Article No.18. ACM, New York, USA

Pattanaik DR, Mohapatra M (2016) Seasonal forecasting of tropical cyclogenesis over the North Indian Ocean. J Earth Syst Sci 125(2):231–250

Rajasekhar M, Kishtawal CM, Prasad MYS, Seshagiri Rao V, Rajeevan M (2014) Extended range tropical cyclone predictions for East Coast of India. In: Mohanty UC, Mohapatra M, Singh BK, Rathore LS (eds) Monitoring and prediction of tropical cyclones in the indian ocean and climate change. Springer, Netherlands, pp 137–148

Rajeevan M, Srinivasan J, Niranjan Kumar K, Gnanaseelan C, Ali MM (2013) On the epochal variation of intensity of tropical cyclones in the Arabian Sea. Atmos Sci Lett 14(4):249–255

Ray A, Jaiswal N, Kishtawal CM (2012) Cyclone tracking over North Indian Ocean using artificial neural network models and implementation of regional and seasonal stratification, Second WMO international conference on Indian Ocean tropical and cyclone change (IOTCCC-2012). New Delhi

Rayhun KMZ, Quadir DA, Chowdhury MAM, Ahsan MN, Haque MS (2015) Simulation of structure, track and landfall of tropical cyclone Bijli using WRF-ARW model. J. Bangladesh Acad Sci 39(2):157–167

Rigollet P, Vert R (2009) Optimal rates for plug-in estimators of density level sets. Bernoulli 15(4):1154–1178

Rumpf J, Weindl H, Höppe P, Rauch E, Schmidt V (2007) Stochastic modelling of tropical cyclone tracks. Math Methods Oper Res 66(3):475–490

Rumpf J, Weindl H, Faust E, Schmidt V (2010) Structural variation in genesis and landfall locations of North Atlantic tropical cyclones related to SST. Tellus, Ser A Dyn Meteorol Oceanogr 62(3):243–255

Sahoo B, Bhaskaran PK (2016) Assessment on historical cyclone tracks in the Bay of Bengal, east coast of India. Int J Climatol 36(1):95–109

Saunders MA, Lea AS (2005) Seasonal prediction of hurricane activity reaching the coast of the United States. Nature 434(7036):1005–1008

Shaevitz DA, Camargo SJ, Sobel AH, Jonas JA, Kim D, Kumar A, Larow TE, Lim YK, Murakami H, Reed KA, Roberts MJ, Scoccimarro E, Vidale PL, Wang H, Wehner MF, Zhao M, Henderson

N (2014) Characteristics of tropical cyclones in high-resolution models in the present climate. J Adv Model Earth Syst 6(4):1154–1172

Shaji C, Kar SK, Vishal T (2014) Storm surge studies in the North Indian Ocean: a review. Indian J Mar Sci 43(2):125–147

Slade SA, Maloney ED (2013) An intraseasonal prediction model of atlantic and east Pacific tropical cyclone genesis. Mon Weather Rev 141(6):1925–1942

Turlach BA (1993) Bandwidth selection in kernel density estimation: a review. CORE Inst Stat 19:1–33

Tyagi A, Bandyopadhyay BK, Mohapatra M (2010) Monitoring and prediction of cyclonic disturbances over North Indian ocean by regional specialised meteorological centre, New Delhi (India): problems and prospective. In: Charabi Y (ed) Indian Ocean tropical cyclones and climate change. Springer, Netherlands, pp 93–103

Vickery PJ, Skerjl P, Steckley AC, Twinsdale L (2000) Simulation of hurricane risk in the united states using an empirical storm track modeling technique. J Struct Eng 126:1222–1237

Vissa NK, Satyanarayana ANV, Prasad Kumar B (2013) Intensity of tropical cyclones during pre- and post-monsoon seasons in relation to accumulated tropical cyclone heat potential over Bay of Bengal. Nat Hazards 68(2):351–371

Vitart F, Leroy A, Wheeler MC (2010) A comparison of dynamical and statistical predictions of weekly tropical cyclone activity in the Southern Hemisphere. Mon Weather Rev 138(9):3671–3682

Warrick RA, MAccines KL, Pittock AB, Kench PS (2000) Climate change, severe storms and sea level. In: Dennis J (ed) Floods. Parker Taylor and Francis group, London, p 139

Webster PJ (2008) Myanmar's deadly daffodil. Nat Geosci 1(8):488–490

Weinkle J, Maue R, Pielke R (2012) Historical global tropical cyclone landfalls. J Clim 25(13):4729–4735

Wilks DS (1995) Statistical methods in the atmospheric sciences. Academic Press, New York, p 467

WMO (1997) Tropical cyclone operational plan for the Bay of Bengal and the Arabian Sea,. WMO report 1997

Yahyai SSA (2014) NWP forecast Guidance during Phet at Oman meteorological service. In: Mohanty UC, Mohapatra M, Singh OP, Bandyopadhyay BK, Rathore LS (eds) Monitoring and prediction of tropical cyclones in the Indian Ocean and climate change. Springer, New York, pp 240–262

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

Zhao M, Held IM, Lin SJ, Vecchi GA (2009) Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. J Clim 22(24):6653–6678