



Bayesian optimization of typhoon full-track simulation on the Northwestern Pacific segmented by QuadTree decomposition

Wei Cui ^{a,b,c}, Lin Zhao ^{a,b,c,*}, Shuyang Cao ^{a,b,c}, Yaojun Ge ^{a,b,c}

^a State Key Lab of Disaster Reduction in Civil Engineering, Tongji University, Shanghai, 200092, China

^b Department of Bridge Engineering, College of Civil Engineering, Tongji University, Shanghai, 200092, China

^c Key Laboratory of Transport Industry of Bridge Wind Resistance Technologies, Tongji University, Shanghai, 200092, China



ARTICLE INFO

Keywords:

Tropical cyclone
Typhoon simulation
Quadtree
Bayesian optimization

ABSTRACT

The tropical cyclone is one of the most destructive weather phenomena for several coastal countries, including China, the United States of America, India, Japan, and Australia. Accurate modeling of their trajectories is essential for public safety. Current research has provided several principle methods for simulating typhoon tracks and intensity development from genesis to landing and decaying. However, simulation performance still needs to be improved. This paper first presents a new ocean segmentation algorithm based on QuadTree to divide the analysis region in the Northwestern Pacific adaptively according to data sample density. The regression analysis area can be automatically adjusted to produce a significant enough data sample for fine-grained modeling. Second, it proposes a Bayesian optimization for parameter tuning. Because Month Carlo typhoon simulation results unavoidably incorporate uncertainties and need long computing time, Bayesian optimization is suitable for typhoon simulation parameter adjustments. With optimized parameters, the simulated typhoon activities demonstrate better agreement with historical records.

1. Introduction

The tropical cyclone is one of the most destructive weather phenomena in the Pacific and Atlantic coastal areas. Their strong winds, high surge, and big rains cause severe property damage and loss of life. In 2005, 1833 fatalities and 108 billion USD of property damage were caused to the USA due to Hurricane Katrina landed on the Gulf of Mexico. In 2018, typhoon Mangkhut landed on the Pearl-River-Delta Area of Southern China and then invaded inland China, strongly affecting several major cities, including Hong Kong, Shenzhen, and Guangzhou. This super Typhoon killed five people and caused nearly 1200 houses to collapse. In 2019, typhoon Lekima initially landed on the southeastern coast of China, traveled north, and then landed once again on the Shandong Peninsula. These areas are not usually considered as typhoon-prone areas.

Motivated by repeated severe tropical cyclone damage, several studies were carried out to accurately forecast typhoon/hurricane activity in both the near and distant future. The following provides a short review of essential models currently employed for tropical cyclone simulation in wind engineering.

In the early stages, most researchers focused on the wind field model

of tropical cyclones. A pioneering method to numerically model a hurricane wind field was first introduced by Georgiou, which was used to predict hurricane wind speeds along the US Atlantic coastline (Georgiou, 1986). Later, Meng provided a height-resolving model based on several simplified semi-analytical algorithms (Meng et al., 1995). This novel 3-dimensional wind-field model can simulate the height-dependent wind-flow characteristics in typhoon climates and terrain effects on the atmospheric boundary layer. Meng's model has been widely used in the Pacific region and has inspired several improved typhoon wind field models emphasizing the height-dependence of wind speeds and radial wind pressure distribution (Fang et al., 2018; Hong et al., 2019).

In later stages, the frequency of typhoon landing and related probabilistic typhoon intensity was introduced to tropical cyclone damage modeling and structural design wind speed prediction. Simiu proposed a "single-site" probabilistic Monte Carlo model for deriving hurricane wind speed, in which five random parameters are considered (Simiu and Scanlan, 1996; Simiu and Yeo, 2019). Subsequently, Vickery proposed an empirical hurricane full-track model in Vickery et al. (2000b,a), employing linear statistical regression of a large hurricane database, known as HURDAT (Landsea and Franklin, 2013), in the Atlantic Ocean. This model was the milestone for tropical cyclone modeling in civil

* Corresponding author. 305 Wind Engineering Building, Tongji University, 1239 Siping Road, Shanghai, 200092, China.

E-mail address: zhaolin@tongji.edu.cn (L. Zhao).

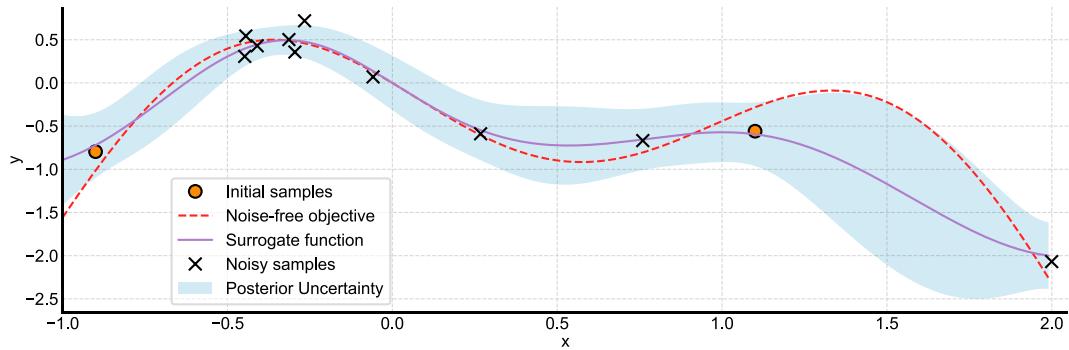


Fig. 1. Illustration of Bayesian optimization procedure with stochastic function evaluations.

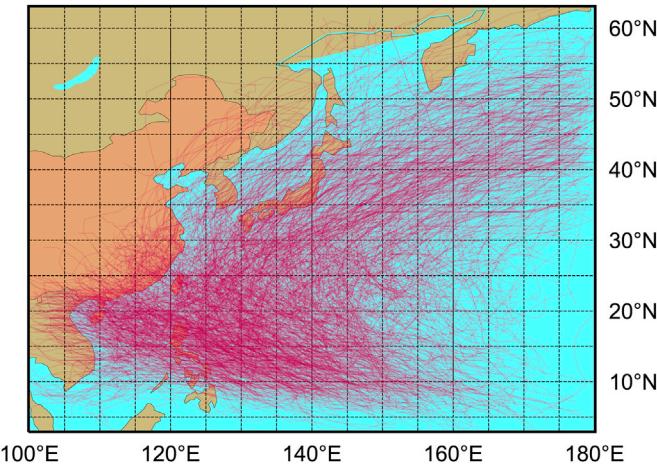


Fig. 2. Typhoon trajectory paths in the Northwestern Pacific Ocean recorded by the International Best Track Archive for Climate Stewardship (IBTrACS).

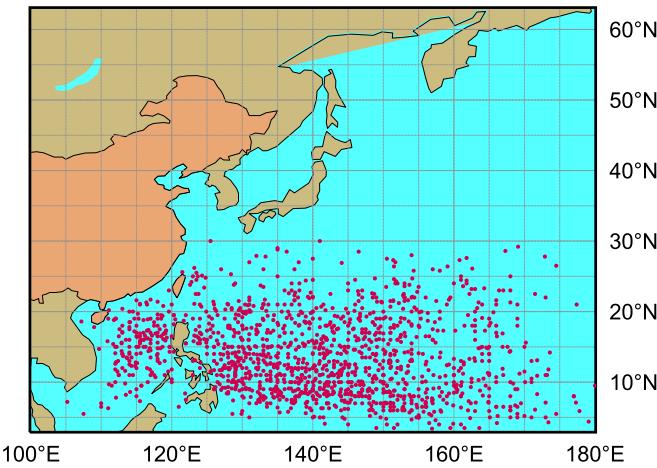


Fig. 3. Typhoon genesis locations in the Northwestern Pacific Ocean.

engineering and has been widely used for long-term hurricane predictions. This seminal work led to many subsequent studies on tropical cyclone wind speed forecast both for the Pacific ocean (Hong et al., 2016; Li and Hong, 2016; Chen and Duan, 2018) and the Atlantic Ocean (Emanuel et al., 2006; Lee and Rosowsky, 2007; Snaiki and Wu, 2019; He et al., 2019).

Currently, Vickery method (Vickery et al., 2000b) has been widely used to simulate tropical cyclones for the different coastal regions over the world. For example, this model has been adopted by the ASCE 7 design standard to provide basic design wind speeds in wind hazard maps for hurricane-prone regions of the USA (ASCE, 2016). Based on Vickery pioneering hurricane model, several researchers from different perspectives have made further contributions to various application fields. For example, Vickery method has been adapted to the northwestern Pacific coast of Asia to predict basic design wind speeds, including typhoon effects (Hong et al., 2016; Li and Hong, 2016). It has also been used to examine the potential influence of “global warming” on hurricane activities (Cui and Caracoglia, 2015, 2019) and consequent effects on structural damage probability and building function integrity and for skyscrapers (Cui and Caracoglia, 2018).

However, Vickery method and other derivative tropical cyclone simulation methods in Hong et al. (2016) and Chen and Duan (2018) have two difficulties, which often require extra effort to deal with. First, because of tropical cyclone activities’ heterogeneity, the ocean basin needs to be segmented for spatial-dependent regression analysis. Furthermore, due to the non-uniformity of tropical cyclone density over the ocean basin, uniform ocean basin segmentation means that some regions lack sufficient data to perform reliable regression analysis. In contrast, other regions have too many data records, and region size cannot be refined. The second difficulty is that Vickery method is purely data-driven and lacks verification and correction based on physical knowledge. Model parameters from original regression analysis cannot achieve adequate simulation accuracy and require extra calibration. Therefore, “regional calibration factors were introduced in the form of an artificial loop current” to improve the simulation performance in Vickery et al. (2009). Other studies follow a similar procedure (Hong et al., 2016). However, this parameter calibration does not have a standardized procedure that ensures consistent performance when applied to multiple ocean basins by different researchers.

This study proposes a procedure for dividing the ocean basin non-uniformly according to tropical cyclone data density to adjust the area size for regression analysis in each segmented analysis domain. Bayesian optimization is then performed to search regional calibration factors that can be employed as a standardized methodology in a data-driven tropical cyclone simulation. In the first part of this paper, the original Vickery method is presented, and the tropical cyclone data on the northwestern Pacific is utilized for training the tropical cyclone simulation model. In the second part, the proposed procedure is presented and used to simulate a tropical cyclone on the northwestern Pacific. Performances are compared in the last part. It should be noted that, because this study was carried out near the northwestern Pacific coast, “typhoon” is used instead of “tropical cyclone” throughout this paper for consistency.

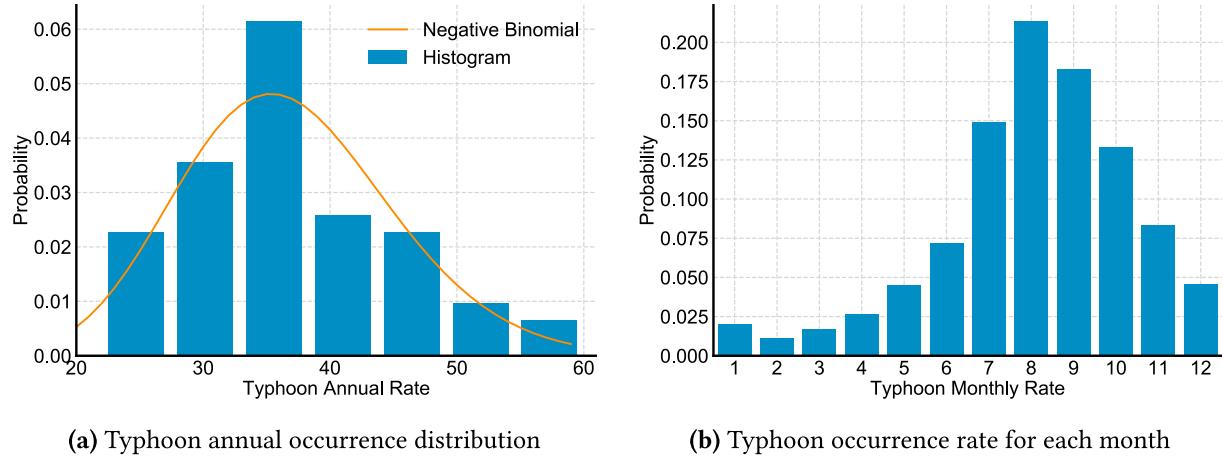


Fig. 4. Typhoon yearly and monthly genesis occurrence probability in the Northwestern Pacific Ocean.

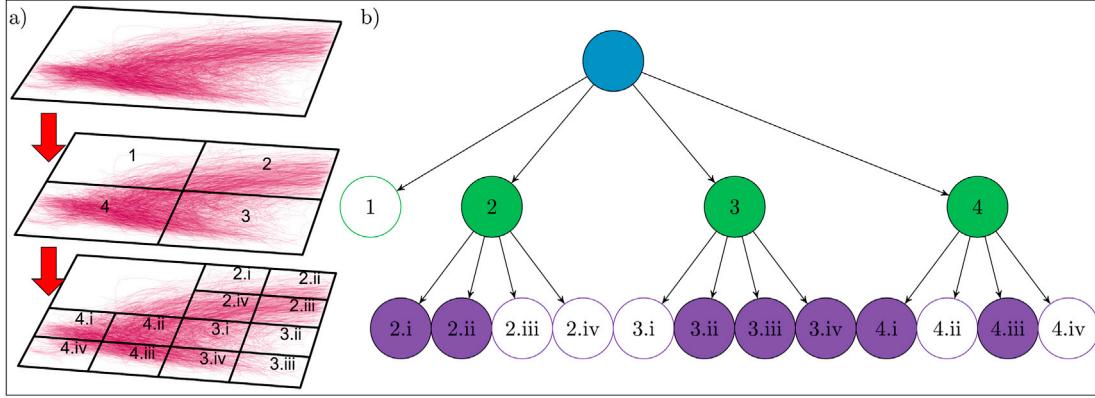


Fig. 5. Schematic Visualization of QuadTree segmentation of the Northwestern Pacific Ocean (a, Visual segmentation of the Northwestern Pacific; b, QuadTree structure for the first three levels).

2. Theoretical background

2.1. Typhoon simulation equations

Vickery model (Vickery et al., 2000b) is based on the following equations (Eqs. (1) and (2) below), which can be used to synthetically generate a sample of typhoon tracks:

$$\Delta \ln c_{i+1} = a_1 + a_2 \psi + a_3 \lambda + a_4 \ln c_i + a_5 \theta_i + e_c \quad (1a)$$

$$\Delta \theta_{i+1} = b_1 + b_2 \psi + b_3 \lambda + b_4 c_i + b_5 \theta_i + b_6 \theta_{i-1} + e_\theta \quad (1b)$$

$$\ln(I_{i+1}) = c_0 + c_1 \ln(I_i) + c_2 \ln(I_{i-1}) + c_3 \ln(I_{i-2}) + c_4 T_{s_i} + c_5 (\Delta T_s) + e_I \quad (2)$$

In Eq. (1a) and Eq. (1b), respectively, $a_j (j = 1, 2\cdots 5)$ and $b_j (j = 1, 2\cdots 6)$ are coefficients that can be estimated by linear regression. The quantities ψ and λ are the latitude and longitude of the typhoon eye position, c_i is typhoon translation speed at time step i , and θ_i is the typhoon heading (direction) at time step i . Quantities e_c , e_θ and e_I are linear regression residuals in Eqs. (1) and (2), respectively. The distribution of residuals is modeled using the method described in Cui and Caracoglia (2016). When a typhoon heading is North, $\theta = 0$, and the limits of θ are $-180 < \theta \leq 180$.

In Eq (2), the relative intensity I (Darling, 1991) (defined at various time steps i , $i - 1$, etc.) is used as a non-dimensional term that relates the actual typhoon pressure deficit Δp to the greatest possible central pressure deficit allowed by the “average climatology” of the typhoon season (Vickery et al., 2000b). ΔT_s is the SST (sea surface temperature)

difference between time step i and time step $i + 1$, which is calculated as $\Delta T_s = T_{s_{i+1}} - T_{s_i}$. Introducing SST into the model reduces some of the unexplained variability in the central pressure modeling. The variations of heading and translation velocity between time step i and time step $i + 1$ are calculated from $\Delta \theta$ and $\Delta \ln c$.

All the model coefficients in Eqs. (1) and (2) are estimated by linear regression, using a database of historical track records (HurDat provided by Landsea and Franklin (2013)) that is maintained by the National Oceanic and Atmospheric Administration (NOAA). The coefficients are independently derived for each rectangular non-overlapping cell of extension $5^\circ \times 5^\circ$, subdividing the Atlantic Ocean region (Vickery et al., 2000b). Moreover, the coefficients are different for east-heading and west-heading typhoons. For cells with insufficient historical data, a nearby cell's coefficients can be used during typhoon simulation.

2.2. Introduction to engineering application of Bayesian optimization

Bayesian optimization (Mockus, 1975) is a probability model of an objective function used to select the most promising parameters to maximize or minimize the true objective function. It is currently widely used in the machine learning field (Snoek et al., 2012) for tuning hyper-parameters of machine learning models.

The basic Bayesian strategy is to treat an objective function unknown at the beginning, as a random function, and place a prior over it to approximate the objective function. After evaluating the function, which is treated as data, the prior is updated to form a posterior distribution over the objective function. The posterior distribution can then construct

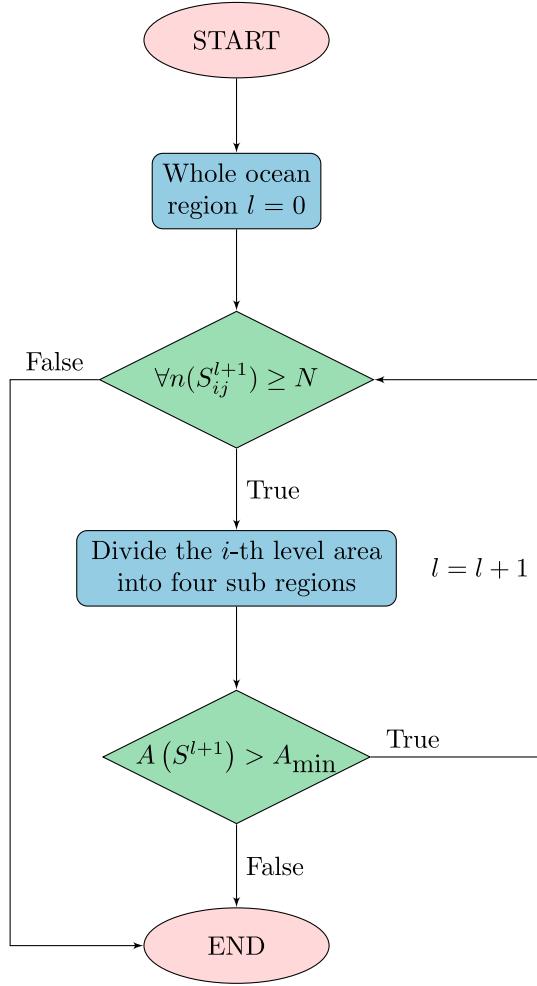


Fig. 6. Flowchart of QuadTree segmentation algorithm.

an acquisition function that determines the next function's evaluation point. This procedure is illustrated in Fig. 1. The dashed line is the real objective function, and the two orange dots are the initial data as priors and the posteriors, including the surrogate function and related uncertainty, keep being updated as more sample data are included.

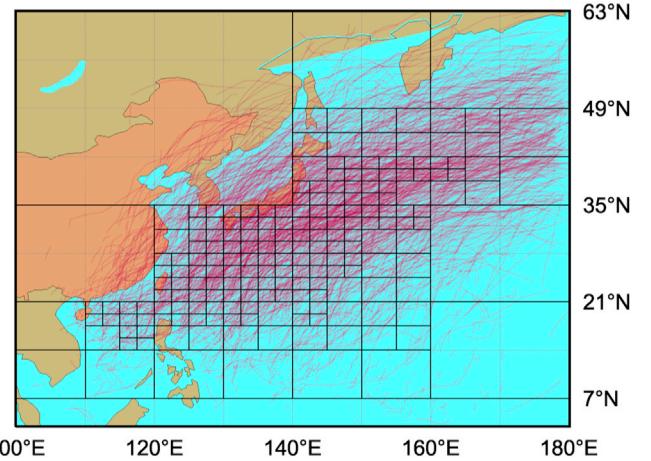
From the above introduction, Bayesian optimization is a gradient-free optimization algorithm, and a function evaluation with particular uncertainty is also acceptable. Another advantage is that posterior in Bayesian optimization can keep updating as the data samples increase, which means only a small number of data samples are required to find an optimal solution. In contrast, other global optimization algorithms need many sample points to search for multiple possible basins, such as genetic algorithm (Eiben et al., 1994) and particle swarm (Shi and Eberhart, 1998) algorithm.

Because of Bayesian optimization's two advantages, the typhoon simulation algorithm's parameters can be automatically configured through it. The detailed Bayesian optimization setup will be explained in the following section.

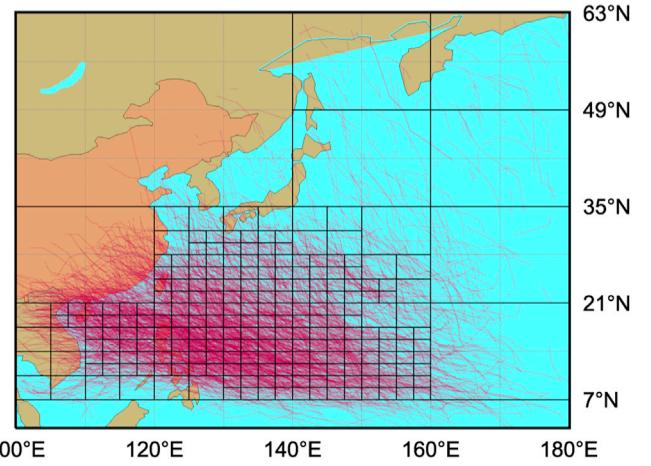
3. Typhoon simulation on the Northwestern Pacific using QuadTree segmentation

3.1. Typhoon best-track dataset on the northwestern Pacific

The International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010, 2018) is used in this study because it has been endorsed by the World Meteorological Organization Tropical



(a) Eastern heading tracks



(b) Western heading tracks

Fig. 7. QuadTree segmentation for eastern and western heading typhoon tracks.

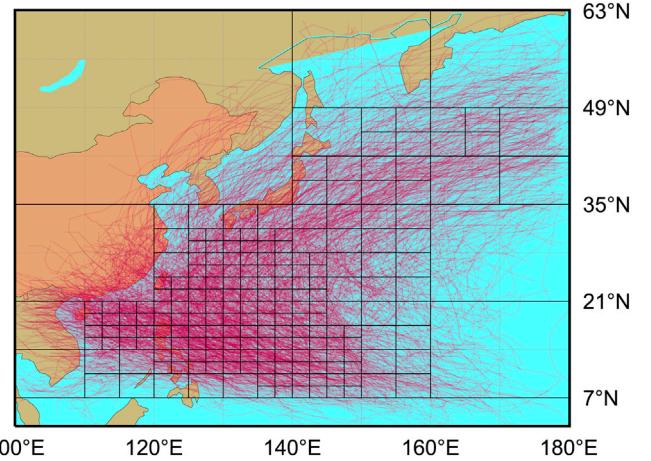


Fig. 8. QuadTree segmentation for typhoon intensity.

Cyclone Programme as an official archiving and distribution resource for tropical cyclone best track data. To maintain the data consistency, the early year typhoon data are discarded, and typhoon data from 1960 to 2018 identified through satellite images are used since the first

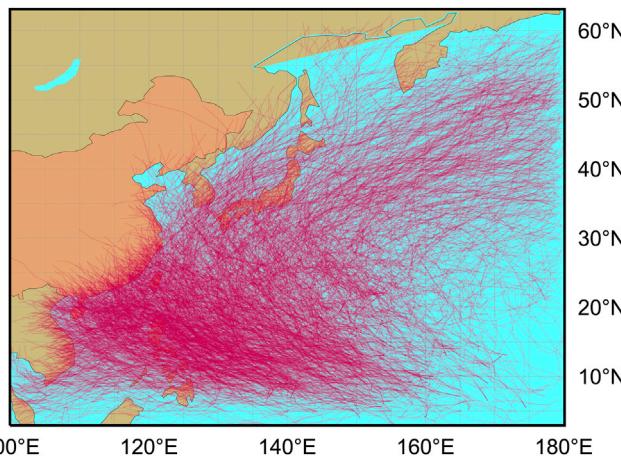


Fig. 9. Simulated typhoon tracks in Northwestern Pacific Ocean using original Vickery's method.

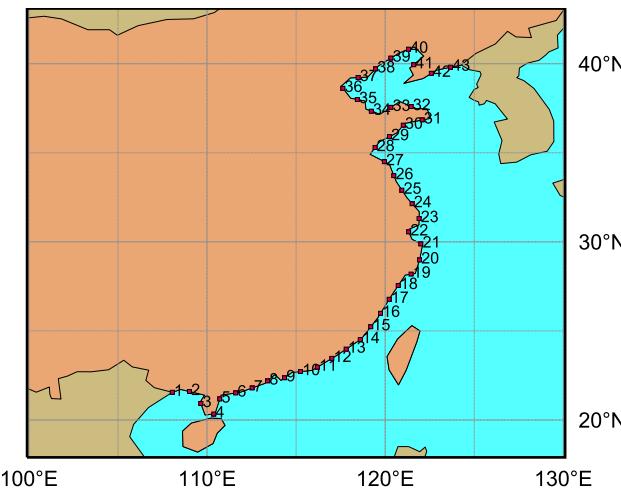


Fig. 10. Locations of mileposts along the China Pacific coastline.

meteorology satellite was launched in 1959. There were 2096 typhoon tracks from 1960~2018, which are plotted in Fig. 2.

All the coefficients of the model equations in Eqs. (1) and (2) can be estimated by linear regression using historical typhoon track records. The coefficients must be independently derived for each rectangular non-overlapping cell of extension $5^\circ \times 5^\circ$, uniformly subdividing the Northwestern Pacific Ocean region. Moreover, two sets of coefficients are achieved separately for east-heading and west-heading typhoon. For cells with insufficient historical data, a nearby cell's coefficients can be used in the typhoon track simulation. During typhoon propagation and simulation, the sea surface temperature field data were obtained from the Hadley Centre Global Sea Ice and Sea Surface Temperature dataset (HadISST) (Rayner et al., 2003). After landing, the typhoon intensity decay model proposed in Vickery et al. (2000b) was adopted in this study.

3.2. Typhoon genesis

The initial “birthplaces” of typhoons are randomly sampled from the historical typhoon database, as shown in Fig. 3. In the Northwestern Pacific Ocean basin, the overall annual typhoon frequency is represented as a Negative Binomial process (Vickery et al., 2000b), which is plotted in Fig. 4a against the histogram of historical typhoon annual occurrence. The typhoon “birth” month is also sampled from the monthly rate from

January to December, shown in Fig. 4b. This demonstrates that the typhoon prone period usually is from July to October each year.

3.3. Ocean basin adaptive segmentation method based on QuadTree

QuadTree (Finkel and Bentley, 1974) is a data structure in which each internal node has exactly four sub-nodes. This structure is most often used to partition a two-dimensional area by recursively subdividing it into four quadrants. One widely-used application example of QuadTree is image compression and image representation in the computational vision field (Sonka et al., 2014). A similar data structure, Octree, is often used to generate mesh for 3-dimensional finite element methods (Bielak et al., 2005).

Considering the non-uniform typhoon tracks data density, as shown in Fig. 2, QuadTree can be used to adaptively divide the ocean area into non-uniform segmentation to ensure that each grid has enough data to perform regression analysis. The ocean area with dense tracks data can be divided more finely to present different spatial characteristics.

Each node in QuadTree can represent one area of the Pacific ocean. As long as all four associated quadrants have enough typhoon track data samples, their parent node will be divided equally into four child nodes. This procedure continues until the minimum area requirement is reached or one of the possible sub-areas does not have enough typhoon track data samples. This algorithm is illustrated in Fig. 6, in which l is the level indicator in the QuadTree structure, S_{ij}^l is the divided ocean area at l -th level, S_{ij}^{l+1} means the four quadrants below the area at level l ($i, j \in 0, 1$), $n()$ is the counting operator, and $A()$ is the area size operator.

Fig. 5 shows the procedure of the first three-level QuadTree segmentation of the Northwestern Pacific ocean. S^0 is the uppermost parent node, indicating the whole analyzed Northwestern Pacific region ($100\text{--}180^\circ\text{E}, 7\text{--}63^\circ\text{N}$). By counting the number of typhoon tracks number from the four quadrants of the parent node, all the child nodes $S_{00}^1, S_{01}^1, S_{10}^1, S_{11}^1$ have enough typhoon track samples, which indicates that the recursive function $\forall n(S_{ij}^l) \geq N$ is true and, thus, the whole ocean area is divided into four equal parts at level $l = 1$. The same procedure will be applied to the four areas at level $l = 1$. Then, at level $l = 2$, three subsequent areas, excluding the northwestern one, are divided into four further parts. Fig. 5 shows the corresponding tree structure for the ocean segmentation. The solid nodes mean that areas can be divided, and the hollow nodes mean that the recursive function is false and cannot be divided. The recursive division will stop when the area reaches the minimum area restriction A_{\min} .

In this study, according to the original Vickery's method in Vickery et al. (2000b), the ocean segmentation is also separated for east-heading and west-heading typhoon tracks. Setting $N = 100$ (Van Voorhis and Morgan, 2007) and $A_{\min} = 2.5^\circ \times 1.75^\circ$, the Pacific ocean Quadtree segmentations for eastern and western heading typhoon tracks are illustrated in Fig. 7a and Fig. 7b, respectively. For the typhoon intensity simulation, it is not necessary to separate different typhoon heading directions, so N is reseted as 150 for intensity analysis. Fig. 8 shows the ocean segmentation for typhoon intensity regression analysis.

Figs. 7~8 show that the QuadTree segmentation method is very effective for adaptively dividing the ocean area according to the data sample density. For the typhoon tracks and intensity regression analysis and simulation, the QuadTree segmentation networks are different because the three categories' data density also has considerable spatial dissimilarity.

The regression analysis according to Eqs. (1)~(2) are performed on the QuadTree segmentations in Figs. 7~8, and then the synthetic typhoon data can be simulated afterward.

However, because of the complex nature of typhoon activities, the regression residuals e_c , e_θ and e_I from Eqs. (1)~(2) cannot form well-shaped normal distributions. Cui and Caracoglia (2016) proposes to employ the student- t distribution to model the regression residuals, and the parameters in student- t distribution need to be re-calibrated to

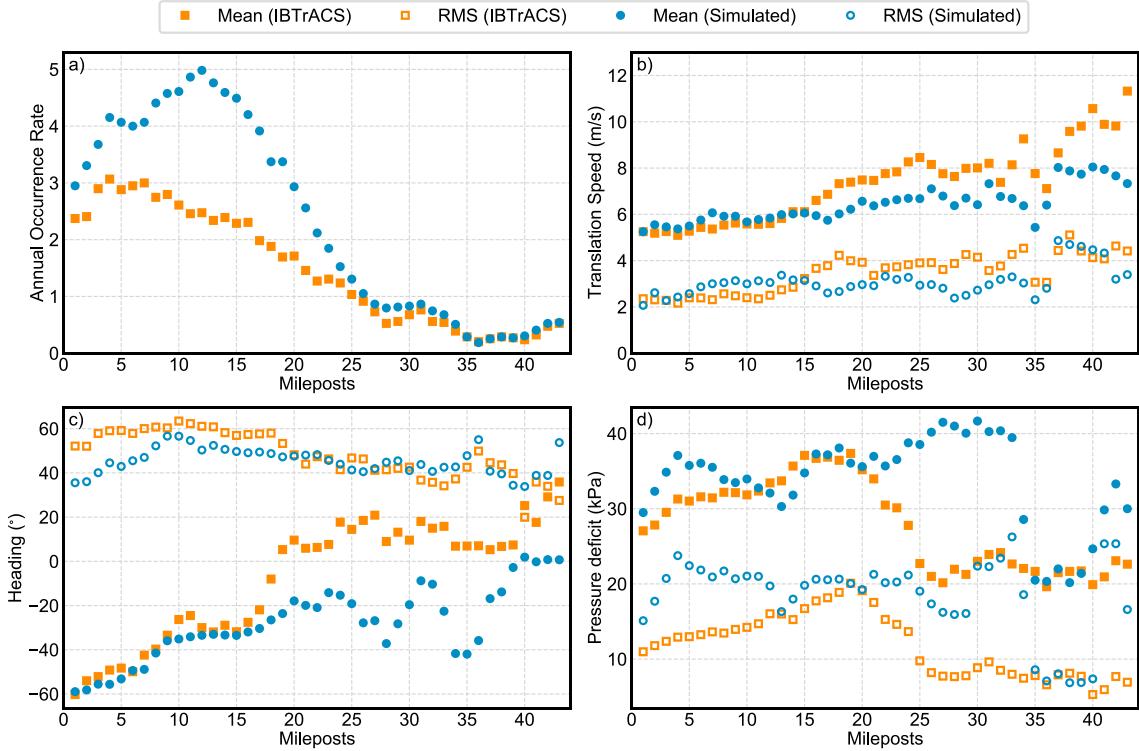


Fig. 11. Validation of simulated typhoon characteristics at landfall along the China Pacific coastline.

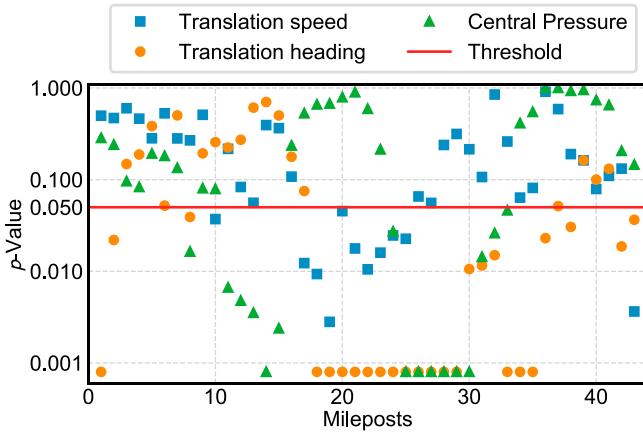


Fig. 12. Kolmogorov-Smirnov tests on similarity between historical and simulated typhoon.

improve the typhoon simulation performance. In the next part of this study, a parameters-calibration method based on Bayesian optimization is proposed.

3.4. Simulation results and discussion

Because of the unavoidable errors and uncertainties in the historical typhoon database, the outliers in the regression analysis will significantly affect the accuracy of typhoon simulation parameters in Eqs. (1) and (2). Therefore, the Robust linear regression method (Gross, 1977) is employed to remove outliers from the regression analysis automatically.

Using the coefficients derived through robust regression analysis of the historical typhoon database and the initial position, also sampled from the historical typhoon database, typhoon tracks can be numerically generated by Eq. (1). As an example, 59 years of synthetic typhoon tracks

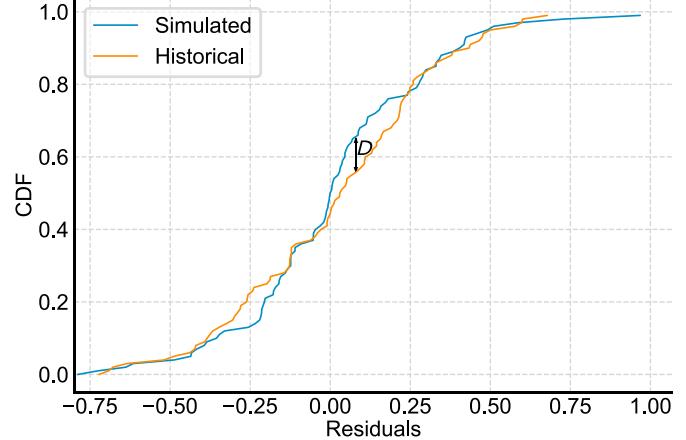


Fig. 13. Illustration of Kolmogorov-Smirnov statistic D_{ij} for two samples.

are simulated and plotted in Fig. 9. Visual inspection indicates that the simulation results obtained from the original algorithm in Eqs. (1) and (2) are disparate compared with historical typhoon tracks.

The simulation results in Fig. 9 tend to have more typhoon tracks in the southern Pacific ocean than those in Fig. 2. The typhoon tracks in Fig. 9 show fewer typhoons turning north at the middle latitude between 20°N and 30°N , and the track density in the Bohai Sea ($117\text{-}121^{\circ}\text{E}$, $37\text{-}41^{\circ}\text{N}$) is less than the historical data. Thus, in the simulation results, the typhoon landing rate on the southern China coastline is larger than the historical values.

To further validate this method, comparisons are given for the statistics of simulated and historical typhoons landing around pre-defined mileposts (MPs) along the China Pacific coastline. This method was initially proposed by Georgiou (1986) and was also adopted by Vickery et al. (2000b). This study defines MPs along the China coastline, starting from the border between China and Vietnam near the South China Sea.

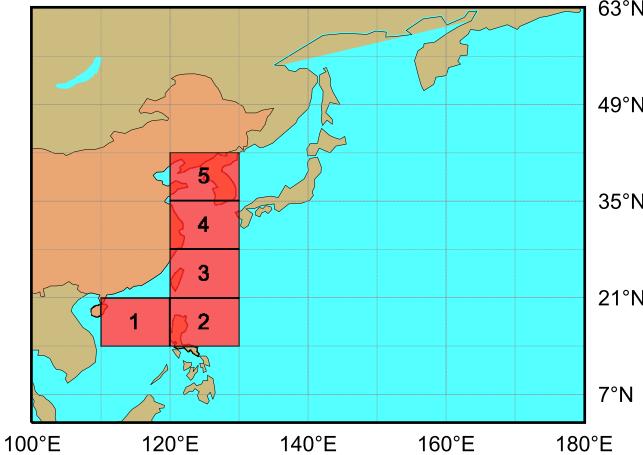


Fig. 14. Consolidated regions with parameters to be optimized.

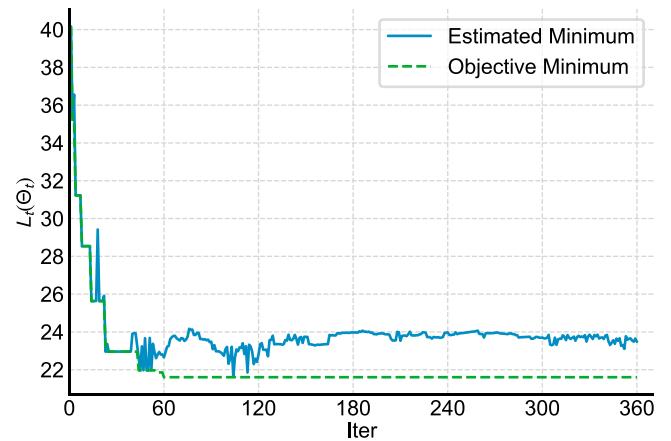
Table 1
Optimization parameters and their ranges.

	Variables	Ocean Region	Range
Θ_t	$\mu_{c,1}$	110-120°E, 14-21°N	[-0.1, 0.1]
	$\mu_{c,2}$	120-130°E, 14-21°N	[-0.1, 0.1]
	$\mu_{c,3}$	120-130°E, 21-28°N	[-0.1, 0.1]
	$\mu_{c,4}$	120-130°E, 28-35°N	[-0.1, 0.1]
	$\mu_{c,5}$	120-130°E, 35-42°N	[-0.1, 0.1]
	$\mu_{o,1}$	110-120°E, 14-21°N	[0, 5]
	$\mu_{o,2}$	120-130°E, 14-21°N	[-5, 10]
	$\mu_{o,3}$	120-130°E, 21-28°N	[-5, 10]
	$\mu_{o,4}$	120-130°E, 28-35°N	[0, 20]
	$\mu_{o,5}$	120-130°E, 35-42°N	[0, 20]
	$\mu_{I,1}$	110-120°E, 14-21°N	[-0.1, 0.1]
	$\mu_{I,2}$	120-130°E, 14-21°N	[-0.1, 0.1]
	$\mu_{I,3}$	120-130°E, 21-28°N	[-0.1, 0.1]
	$\mu_{I,4}$	120-130°E, 28-35°N	[-0.1, 0.1]
	$\mu_{I,5}$	120-130°E, 35-42°N	[-0.1, 0.1]
α_I	$\alpha_{I,1}$	110-120°E, 14-21°N	[0.5, 1.5]
	$\alpha_{I,2}$	120-130°E, 14-21°N	[0.5, 1.5]
	$\alpha_{I,3}$	120-130°E, 21-28°N	[0.5, 1.5]
	$\alpha_{I,4}$	120-130°E, 28-35°N	[0.5, 1.5]
	$\alpha_{I,5}$	120-130°E, 35-42°N	[0.5, 1.5]

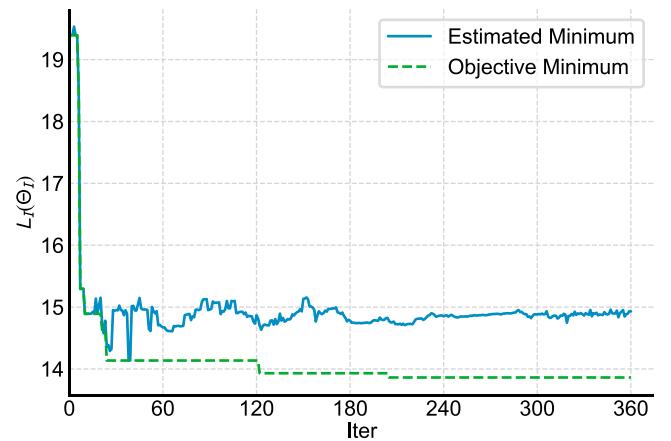
The distance between two nearby adjacent MPs is 500 km. All the locations of the MPs are shown in Fig. 10.

Detailed statistical comparisons of mean values and root-mean-square (RMS) are presented in Fig. 11 at various MP locations. When a simulated typhoon enters the 250 km region of coastal MPs, translation speeds, headings (translation direction), and pressure deficit are recorded. Both simulated and real typhoon activities are random processes, so certain differences between them are expected. However, clear and substantial discrepancies between simulation results and historical typhoon data are observed in Fig. 11. Corresponding to the simulation tracks in Fig. 9, the landing rate for MP 1~20 is much larger than the historical records. The mean translation wind speed from the simulation shows a closer agreement with historical records, but the mean typhoon heading is completely different. Most tracks keep moving northwesterly in the simulated typhoon records as in their initial direction rather than turning northeast. However, in Fig. 2, around half of them turned from northwest to northeast due to the Coriolis effect. For intensity simulation results, only typhoons with central pressures smaller than 1000 mBar have been included in Fig. 11. The simulation results also generally show larger values than historical ones, since they travel further over warm water and absorb more energy.

In order to qualitatively elaborate on the large discrepancies between historical and simulated results, Kolmogorov-Smirnov (K-S) tests are



(a) Tracking parameters



(b) Intensity parameters

Fig. 15. Decreasing values of Optimization objective function for each iteration.

performed to show the simulation performance about three factors: typhoon translation speeds, heading, and central pressure. The *p*-values derived from K-S tests are plotted in Fig. 12. Similar to the statistical comparison, Fig. 12 also demonstrates that the simulation results about translation speeds have good accuracy. The translation heading results have good performance for low latitude milepost, but, for middle and high latitude mileposts, the simulated results have large differences because of very low *p*-values. Also, a similar observation can be found for central pressure simulation results. Generally, the typhoon translation heading and central pressure simulation results on a considerable part of mileposts fail to pass the hypothesis tests that simulation results and historical typhoon data have similar distributions.

As discussed in the introduction section, the major reason for these large discrepancies is that the coefficients in Eqs. (1)~(2), derived from regression analysis of typhoon tracks data, require extra calibrations. In the next part of this study, a Bayesian optimization method is provided as a standard coefficients calibration procedure.

4. Bayesian optimization of typhoon simulation parameters

4.1. Optimization objective function

Generally, the first step in an optimization framework is to define the objective function. Unlike structural optimization problems either for construction cost or structural performance, the objective function is obvious and can be derived directly. For typhoon simulation, the object is to simulate the typhoon's activities as closely as possible to the historical

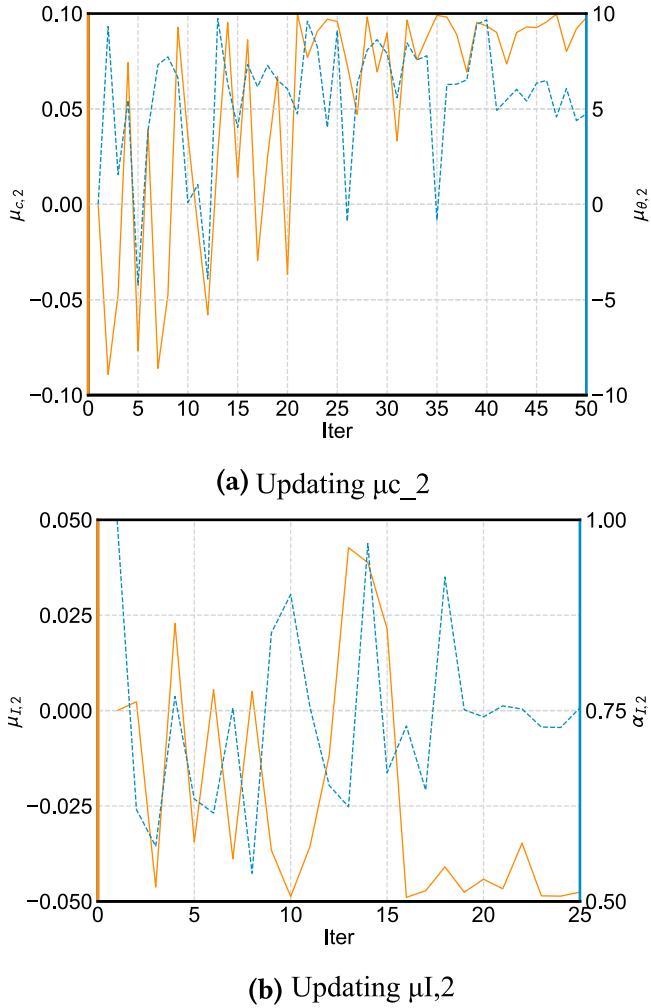


Fig. 16. Updating parameter values for each iteration.

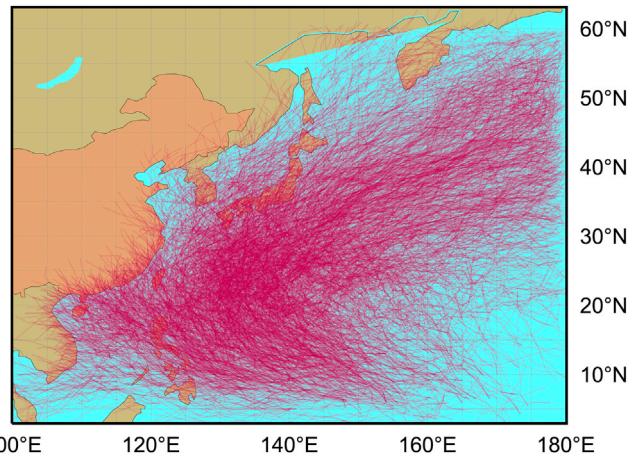


Fig. 17. Simulated Typhoon trajectory paths after optimization.

$$D_{ij} = \sup_x |F_{ij}^s(x) - F_{ij}^h(x)| \quad (3)$$

in which $F_{ij}^s(x)$ and $F_{ij}^h(x)$ are the cumulative distribution function for simulated typhoon (s) and historical typhoon (h) at i -th milestone for j simulated typhoon evaluation criteria. j can be typhoon translation speed c or heading θ or intensity I . $\sup_x |\cdot|$ is the supremum of the set of distances, which can be explained as the maximum distance between F^s and F^h .

Besides the typhoon statistical comparisons, another important factor is the typhoon landing rate at the different milestones along the coastline, as shown in Fig. 11a). The landing rate directly influences the extreme wind hazards occurrences frequency and wind speed distribution.

According to the typhoon simulation formulas in Eqs. (1) and (2), typhoon trajectory and intensity simulation can be considered as independent of each other. Therefore, to reduce the number of optimization parameters and computation time, the parameters optimization for typhoon trajectory and intensity are separated. Thus the optimization objective functions are also separated. Combing the typhoon landing rate and statistics on typhoon translation speeds and heading, the overall optimization objective function for the typhoon's trajectory is defined as:

$$L_t(\Theta_t) = w_r |r_{s,i} - r_{r,i}|_1 + w_t (|D_{i,c}|_1 + |D_{i,\theta}|_1) \quad (4)$$

where Θ_t is the array of typhoon simulation parameters related to the trajectory to be optimized; $|\cdot|_1$ is the L1 norm operator; $r_{s,i}$ and $r_{r,i}$ are the landing rate at the i -th milestone for simulated typhoon and historical records, respectively; and w_r and w_t are the weighting factors for landing rate and typhoon landing status, respectively, which are set as 2 and 1.

Similarly, the optimization objective function for typhoon intensity is defined as:

$$L_I(\Theta_t) = |D_{i,I}|_1 \quad (5)$$

in which Θ_t is the optimized parameters array related to typhoon intensity.

4.2. Optimized parameters

The selection of parameters is essential for typhoon simulation optimization. The total number of optimization parameters should be limited to ensure optimization efficiency. a Large number of parameters for derivate-free optimization requires long calculating time and expensive computation power, which may make the optimization became impractical or too expensive to be widely employed. As indicated in Cui and Caracoglia (2016), student-t distribution is used to model the linear regression residuals ε_c , ε_θ , ε_I in Eqs. (1)–(2). The probability density

Table 2
Optimized parameter values.

	Variables	Value	Variables	Value
Θ_t	$\mu_{c,1}$	-0.036	Θ_I	0.0206
	$\mu_{c,2}$	-0.023		0.0152
	$\mu_{c,3}$	-0.001		0.0123
	$\mu_{c,4}$	0.099		-0.0456
	$\mu_{c,5}$	0.026		-0.0256
	$\mu_{\theta,1}$	1.241		0.783
	$\mu_{\theta,2}$	4.952		0.511
	$\mu_{\theta,3}$	9.697		0.603
	$\mu_{\theta,4}$	4.298		0.714
	$\mu_{\theta,5}$	19.934		0.564

records, including both the trajectories and intensities through the typhoon life-time. However, there is no a direct indicator to evaluate the similarity between simulation and historical records. In Vickery et al. (2000b), mean and RMS values are compared used to assess the typhoon simulation accuracy when landing around the predefined equally-spaced milestones, which is also plotted in Fig. 11. Besides the comparison of mean and RMS values, two sample Kolmogorov-Smirnov tests (Massey, 1951) are also widely used to verify the typhoon simulation accuracy (Cui and Caracoglia, 2019).

Inspired by the Kolmogorov-Smirnov test, the Kolmogorov-Smirnov statistic D_{ij} , shown in Fig. 13, for two samples is employed as the optimization object.

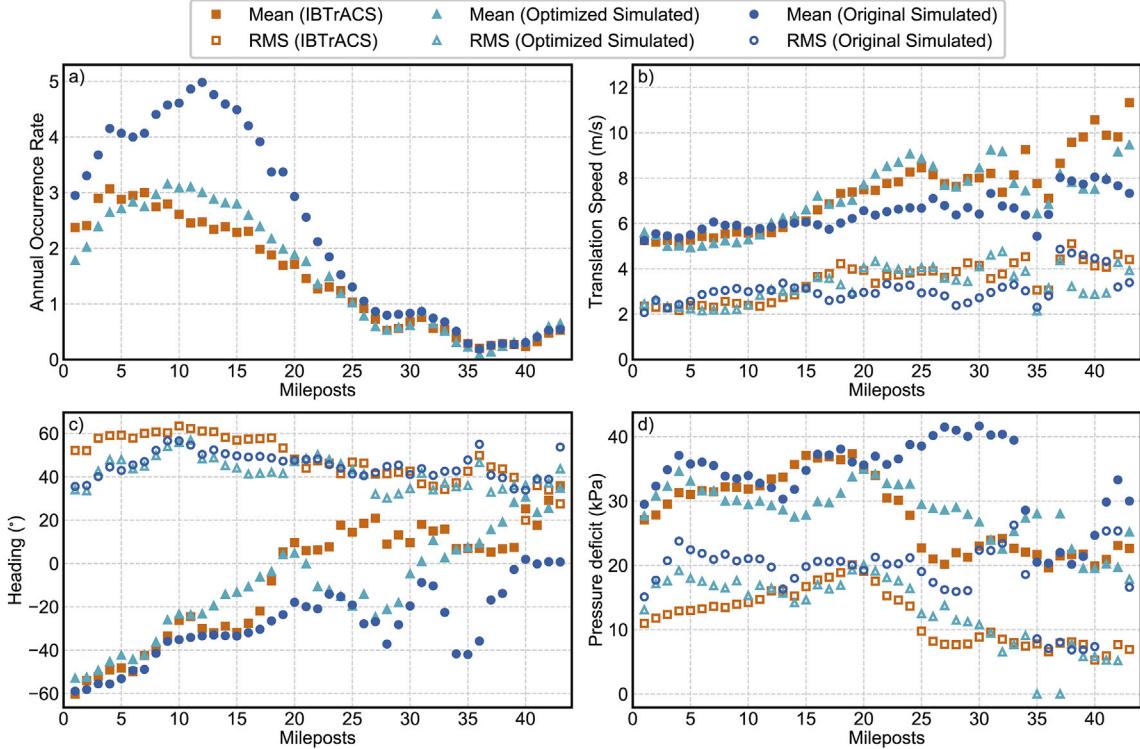


Fig. 18. Validation of optimized typhoon occurrence rate and translation speeds at landfall along the China Pacific coastline.

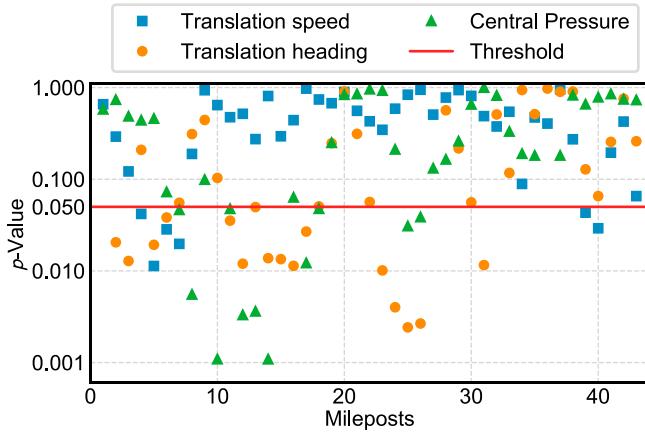


Fig. 19. Kolmogorov-Smirnov tests on similarity between historical and simulated typhoon with optimized parameters.

function of student-*t* distribution is:

$$f(t) = \frac{1}{\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad (6a)$$

$$t = \frac{\epsilon - \mu}{\sigma} \quad (6b)$$

where ν is the degree of freedom of student-*t* distribution, $B()$ is Beta function, and t is regularized random number normalized by mean μ and standard deviation σ .

As shown in Figs. 7–8, typhoon simulation parameters vary for a large number of grids. Considering that the optimization target is the similarity of simulated landing typhoon statistics against typhoon historical records, only the typhoon simulation parameters in the region

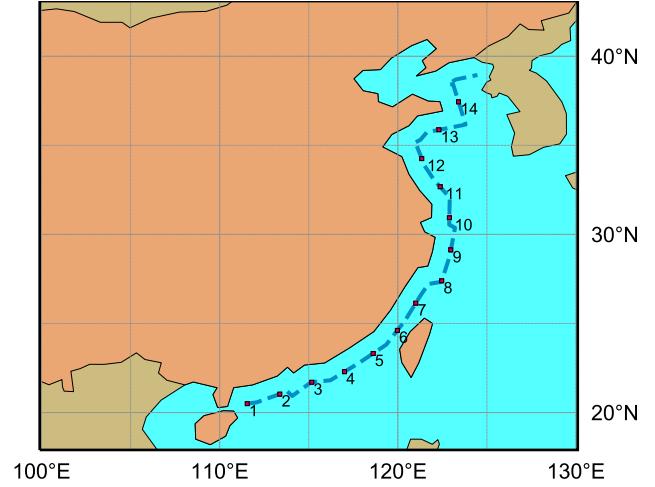


Fig. 20. Locations of mileposts along the 100 km offset coastline.

close to China's coastline are selected to be optimized. To further reduce the number of optimization variables, several nearby grids will share the same optimization variable. Finally, the tracking parameters μ_c and μ_θ for typhoon translation speeds and headings in 5 consolidated regions are selected as variables to be optimized, and are plotted in Fig. 14. For typhoon translation simulation, μ is selected as the optimized variable, while σ and ν are kept constant.

Thus, in the first step in typhoon tracking optimization, there are 10 optimization variables. Their initial values are all set at 0. For Bayesian optimization, the variable range should be predetermined. Comparing the typhoon simulation results with the parameters originally derived from regression analysis against historical records, the predefined optimization range are listed in Table 1.

Both the mean and RMS of simulated typhoons deviate from historical records, as shown in Fig. 9. Therefore, in the second step about typhoon

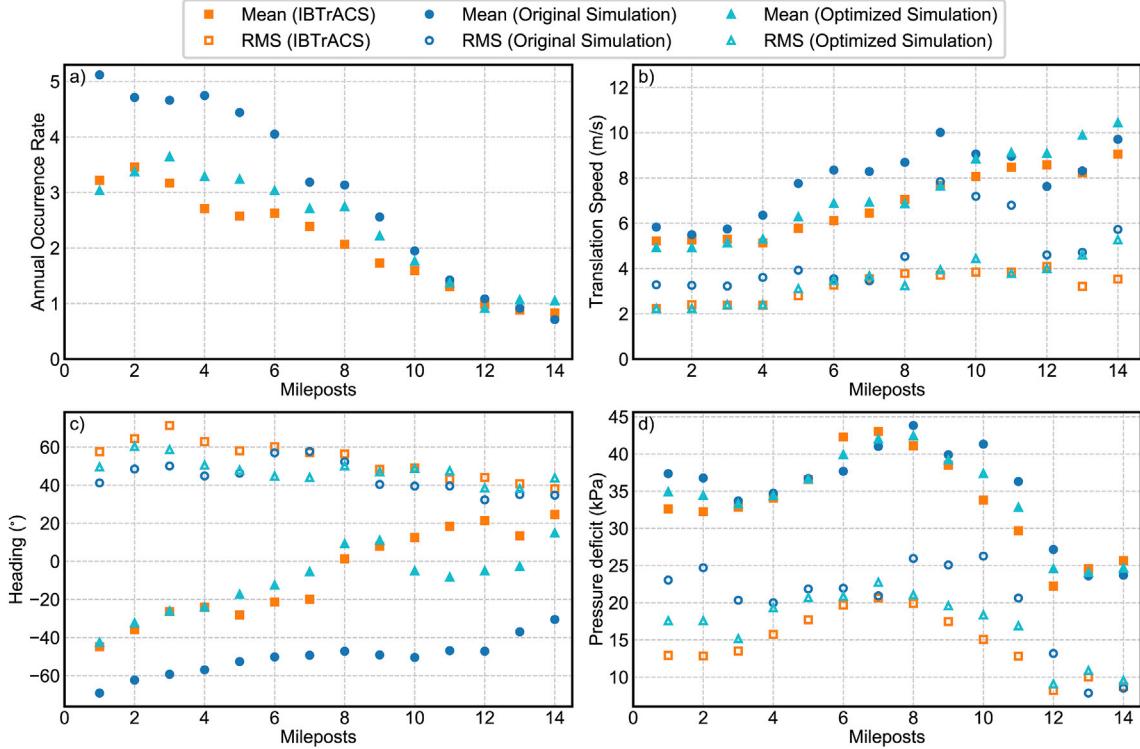


Fig. 21. Independent examination of optimized typhoon occurrence rate and translation speeds at offshore points.

intensity optimization, μ and σ will be optimized for intensity simulation, while ν is still kept constant. Another challenge is that the standard deviation σ varies significantly for different regions, and it is difficult to assign a uniform range to the five regions in Fig. 14. Therefore, an extra factor α is introduced to modify the original σ . Thus, for intensity simulation, the t value for modeling the probabilistic distribution of residuals is:

$$t = \frac{\varepsilon - \mu}{\alpha\sigma} \quad (7)$$

α is selected as the optimization variable rather than σ . All typhoon intensities related to optimization variables and their ranges are also listed in Table 1.

4.3. Two-step optimization procedure

The Bayesian optimization toolbox in the MATLAB platform is employed in this study to perform parameter tuning for typhoon simulation. Twelve initial points are selected as the seed, and “expected improvement plus” algorithm (Bull, 2011) is used.

As explained previously, to reduce the dimensions of the optimization searching space, parameters for typhoon trajectory and intensity are separated. In the first step, 10 variables in the array Θ_t shown in Table 1 are optimized. Then, based on the tuned trajectory variables, 10 variables in the array Θ_I are optimized by minimizing the optimization objective function $L_t(\Theta_t)$ and $L_I(\Theta_I)$. The maximum iteration for both steps is 360.

Fig. 15 shows the objective minimum and expected minimum (Snoek et al., 2012) evolution with iterations for both step one for typhoon trajectory and step two for typhoon intensity. Both $L_t(\Theta_t)$ and $L_I(\Theta_I)$ drop very fast for beginning iterations. Afterward, the expected minimum from the Bayesian optimization algorithm keeps stable. The objective minimum from the objective function evaluations can be marginally reduced, largely because of the Monte Carlo simulation’s evaluation randomness.

Fig. 16 shows the typhoon simulation optimization parameters

updating history for the first several iterations in the ocean region $120^\circ\text{--}130^\circ\text{E}$, $14^\circ\text{--}21^\circ\text{N}$ (region 2 in Fig. 14). For typhoon trajectory simulation, both $\mu_{c,2}$ and $\mu_{\theta,2}$ initially fluctuate greatly over the whole optimization searching space, then the two variable converge into smaller regions. After 360 iterations, the final optimized values of typhoon simulation variables are as listed in Table 2.

4.4. Optimization results

Using the optimized typhoon simulation parameters, 59 years of typhoon activities are reimplemented, and the updated typhoon tracks are plotted in Fig. 17. Comparison of the typhoon history shown in Fig. 2 and the original simulated typhoon records in Fig. 9 demonstrates that the typhoon simulation algorithm performance has been greatly improved. Unlike the initial simulation results, the Coriolis effect can affect the trajectories of a partial of simulated typhoons, so that their headings turn from northwestern to northeastern. However, in the middle latitude region ($25^\circ\text{--}35^\circ\text{N}$), the typhoon paths are much denser than the historical records. This is because, in historical records, a lot of typhoon tracks stopped when typhoon intensity had been weakened by the cold water. In contrast, typhoon trajectory simulation will be terminated only when the central pressure deficiency is 0. As a result, the simulated typhoons have more tracks, but with weaker intensities, in the middle latitude region. Therefore, those typhoon wind speeds with negligible intensity can be easily identified and removed in typhoon hazards estimation.

Like the validation procedure shown in Fig. 11, the typhoon landing statistics along the coastline at different MPs after optimization are also plotted in Fig. 18. In general, simulated typhoons after optimization show closer similarity to the historical typhoon records. In Fig. 18a), the optimized typhoon occurrence rate for MP1~24 has been reduced and more closely match with historical records. Fig. 18c) shows that the mean typhoon heading has increased, and the typhoon moving direction has been turned from northeastern to northwestern, as discussed previously. For typhoon intensity, the RMS values of the typhoon central pressure deficit are also decreased to match the historical value. Also, the

Kolmogorov-Smirnov tests results shown in Fig. 19 also prove that typhoon simulation with optimized parameters has been greatly improved. Most *p*-values are above 0.05 threshold, and only a small portion *p*-values are less than 0.01.

4.5. Independent examination of simulated typhoon tracks at offshore locations

Besides examining typhoon simulation performance at landing points along the coastline, simulated typhoon characteristics are also investigated at several offshore locations in this section. The offshore locations are shown in Fig. 20 are selected at points 100 km off from the coastline, and the distance between neighboring points is 200 km. Unlike the mileposts on the coastline, which are used for both parameter optimization and result comparison, the offshore locations in Fig. 20 are not included in the optimization process, which can be considered as independent of parameter tuning.

The simulated typhoons characteristics at these offshore locations are plotted in Fig. 21. Like the mileposts along the coastline, typhoon simulation with parameters derived from regression without tuning show large discrepancies from historical typhoon records for both occurrence rate points and typhoon translation headings. In contrast, after parameter tuning through Bayesian optimization, the typhoon simulations results show closer agreement. Comparing simulated and historical typhoons at independent offshore points demonstrates the effectiveness of the optimization procedure proposed in this study.

5. Conclusions

This paper has summarized the current research status of typhoon simulation for engineering applications. Implementation of classic typhoon simulation methods shows that the original algorithm cannot achieve the required simulation performance compared with historical typhoon records from the IBTrACS database. This paper first proposed to use QuadTree to divide the Northwestern Pacific adaptively according to data sample density so the regression analysis could be performed for a refined area, and enough data samples could be gathered from an ocean region with sparse data density. After ocean segmentation, this paper proposed a parameter tuning methodology based on Bayesian optimization. Because the Monte Carlo principle is employed in the typhoon simulation, one implementation of typhoon simulation for a large period has certain randomness and requires long computing time. Bayesian optimization, which is normally used for hyper-parameter tuning in machine learning, is appropriate for typhoon stimulation parameter adjustment. With optimized parameters, by comparing the statistics of landing typhoons and also typhoon characteristics on selected offshore locations, the simulation results demonstrate better agreement with historical records.

CRediT authorship contribution statement

Wei Cui: Writing - original draft, Conceptualization, Validation, Visualization. **Lin Zhao:** Supervision, Writing - review & editing. **Shuyang Cao:** Project administration. **Yaojun Ge:** Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the support of National Natural Science Foundation of China (52008314, 51678451), Shanghai Pujiang

Program (No. 19PJ1409800) and National Key Research and Development Program of China (2018YFC0809600, 2018YFC0809604). Any opinions, findings and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the above agencies.

References

- ASCE, 2016. Minimum Design Loads for Building and Other Structures (ASCE 7-16).
- Bielak, J., Ghattas, O., Kim, E., 2005. Parallel octree-based finite element method for large-scale earthquake ground motion simulation. *Comput. Model. Eng. Sci.* 10, 99.
- Bull, A.D., 2011. Convergence rates of efficient global optimization algorithms. *J. Mach. Learn. Res.* 12, 2879–2904.
- Chen, Y., Duan, Z., 2018. A statistical dynamics track model of tropical cyclones for assessing typhoon wind hazard in the coast of southeast China. *J. Wind Eng. Ind. Aerod.* 172, 325–340.
- Cui, W., Caracoglia, L., 2015. Simulation and analysis of intervention costs due to wind-induced damage on tall buildings. *Eng. Struct.* 87, 183–197.
- Cui, W., Caracoglia, L., 2016. Exploring hurricane wind speed along US Atlantic coast in warming climate and effects on predictions of structural damage and intervention costs. *Eng. Struct.* 122, 209–225.
- Cui, W., Caracoglia, L., 2018. A unified framework for performance-based wind engineering of tall buildings in hurricane-prone regions based on lifetime intervention-cost estimation. *Struct. Saf.* 73, 75–86.
- Cui, W., Caracoglia, L., 2019. A new stochastic formulation for synthetic hurricane simulation over the North Atlantic Ocean. *Eng. Struct.* 199, 109597.
- Darling, R., 1991. Estimating probabilities of hurricane wind speeds using a large-scale empirical model. *J. Clim.* 4, 1035–1046.
- Eiben, A.E., Raué, P.E., Ruttkay, Z., 1994. Genetic algorithms with multi-parent recombination. In: International Conference on Parallel Problem Solving from Nature. Springer, pp. 78–87.
- Emanuel, K., Ravela, S., Vivant, E., Risi, C., 2006. A statistical deterministic approach to hurricane risk assessment. *Bull. Am. Meteorol. Soc.* 87, 299–314.
- Fang, G., Zhao, L., Cao, S., Ge, Y., Pang, W., 2018. A novel analytical model for wind field simulation under typhoon boundary layer considering multi-field correlation and height-dependency. *J. Wind Eng. Ind. Aerod.* 175, 77–89.
- Finkel, R.A., Bentley, J.L., 1974. Quad trees a data structure for retrieval on composite keys. *Acta Inf.* 4, 1–9.
- Georgiou, P.N., 1986. Design Wind Speeds in Tropical Cyclone-Prone Regions. Ph.D. Dissertation, University of Western Ontario, London, Ontario, Canada.
- Gross, A.M., 1977. Confidence intervals for bisquare regression estimates. *J. Am. Stat. Assoc.* 72, 341–354.
- He, Y., Li, Y., Chan, P., Fu, J., Wu, J., Li, Q., 2019. A height-resolving model of tropical cyclone pressure field. *J. Wind Eng. Ind. Aerod.* 186, 84–93.
- Hong, H.P., Li, S.H., Duan, Z.D., 2016. Typhoon wind hazard estimation and mapping for coastal region in mainland China. *Nat. Hazards Rev.* 17, 04016001.
- Hong, X., Hong, H., Li, J., 2019. Solution and validation of a three dimensional tropical cyclone boundary layer wind field model. *J. Wind Eng. Ind. Aerod.* 193, 103973.
- Knapp, K.R., Diamond, H.J., Kossin, J.P., Kruk, M.C., J. S., 2018. International Best Track Archive for Climate Stewardship (IBTrACS) Project version 4. [Online; accessed 25-September-2019].
- Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J., Neumann, C.J., 2010. The international best track archive for climate stewardship (ibtracs): unifying tropical cyclone data. *Bull. Am. Meteorol. Soc.* 91, 363–376.
- Landsea, C.W., Franklin, J.L., 2013. Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Weather Rev.* 141, 3576–3592.
- Lee, K.H., Rosowsky, D.V., 2007. Synthetic hurricane wind speed records: development of a database for hazard analyses and risk studies. *Nat. Hazards Rev.* 8, 23–34.
- Li, S.H., Hong, H.P., 2016. Typhoon wind hazard estimation for China using an empirical track model. *Nat. Hazards* 82, 1009–1029.
- Massey Jr., F.J., 1951. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* 46, 68–78.
- Meng, Y., Matsui, M., Hibi, K., 1995. An analytical model for simulation of the wind field in a typhoon boundary layer. *J. Wind Eng. Ind. Aerod.* 56, 291–310.
- Mockus, J., 1975. On bayesian methods for seeking the extremum. In: Optimization Techniques IFIP Technical Conference. Springer, pp. 400–404.
- Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C., Kaplan, A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.: Atmosphere* 108, 4407.
- Shi, Y., Eberhart, R., 1998. A modified particle swarm optimizer. In: 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98TH8360). IEEE, pp. 69–73.
- Simiu, E., Scanlan, R.H., 1996. Wind Effects on Structures: Fundamentals and Applications to Design, 3 ed. John Wiley & Sons, New Jersey, USA.
- Simiu, E., Yeo, D., 2019. Wind Effects on Structures: Modern Structural Design for Wind, 4 ed. Wiley-Blackwell, New Jersey, USA.
- Snaiki, R., Wu, T., 2019. Knowledge-enhanced deep learning for simulation of tropical cyclone boundary-layer winds. *J. Wind Eng. Ind. Aerod.* 194, 103983.
- Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. In: Advances in Neural Information Processing Systems, pp. 2951–2959.
- Sonka, M., Hlavac, V., Boyle, R., 2014. Image Processing, Analysis, and Machine Vision. Cengage Learning.

- Van Voorhis, C.W., Morgan, B.L., 2007. Understanding power and rules of thumb for determining sample sizes. *Tutorials Quantitative Methods Psychol.* 3, 43–50.
- Vickery, P.J., Skerlj, P.F., Steckley, A.C., Twisdale Jr., L.A., 2000a. Hurricane wind field model for use in hurricane simulations. *J. Struct. Eng.* 126, 1203–1221.
- Vickery, P.J., Skerlj, P.F., Twisdale, L.A., 2000b. Simulation of hurricane risk in the U.S. using empirical track model. *J. Struct. Eng.* 126, 1222–1237.
- Vickery, P.J., Wadhera, D., Twisdale Jr., L.A., Lavelle, F.M., 2009. US hurricane wind speed risk and uncertainty. *J. Struct. Eng.* 135, 301–320.