August 27-31, 2023
FLORENCE - ITALY
16th International Conference on Wind Engineering

Typhoon genesis model based on Support Vector Machine considering effects of multiple meteorological parameters

Miaomiao Wei [1], Genshen Fang [2], Yaojun Ge [3]

[1]*State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai, China, wmm@tongji.edu.cn*

[2]*Key Laboratory of Transport Industry of Bridge Wind Resistance Technologies, Tongji University, Shanghai China, 2222tjfgs@tongji.edu.cn*

[3]*Key Laboratory of Transport Industry of Bridge Wind Resistance Technologies, Tongji University, Shanghai China, yaojunge@tongji.edu.cn*

SUMMARY:
In the existing typhoon simulation methods, the number and location of typhoons are mostly simulated based on the statistical characteristics of historical data. The annual frequency of typhoons is mostly defined by negative binomial or Poisson distribution, but there is no conclusive evidence that the annual frequency of typhoon perfectly conforms to a particular statistical distribution model. This paper explores to use support vector machine (SVM) to simulate typhoons. The classification results are achieved by performing high-dimensional mapping between the meteorological parameters and the recorded generated results. This paper also studies the optimal proportion of positive and negative examples during training, as well as the optimal ratio of training sets and test sets. Compared with the traditional method, SVM method improve the calculation efficiency greatly. In addition, the real meteorological data is more fully utilized, taking into account real physical mechanisms. In the case of considering the impact of climate change in the future, the predicted meteorological data can be directly input into this model, and the typhoon generation pattern considering the impact of climate change can be obtained.

Keywords: Typhoon genesis; Support vector machine; Machine learning

1. MOTIVATIONS

The Northwest Pacific region (WNP) is one of the areas that suffer the most typhoon disasters every year, and typhoons (tropical cyclones, TC) are also the top ten natural disasters in the world (Cao et al., 2022; Chen et al., 2022). Existing methods for simulating the number and location of typhoon rely mostly on parametric and non-parametric estimation methods based on statistical dynamics. The parameter estimation method needs to assume a certain distribution in advance, but there is no conclusive evidence that the typhoon spawn rate perfectly conforms to a definite distribution. The spawn position usually uses a non-parametric estimation method, and the Gaussian kernel density estimation (KDE) method is widely used. Traditional statistical methods do not take into account the influence of meteorological parameters, making it difficult to predict typhoon disasters under climate change conditions. However, with the improvement of historical reanalysis data of meteorological parameters recently, it has become possible to simulate typhoon considering the actual physical mechanism, which provides a new idea for typhoon simulation. At the same time, traditional statistical methods are difficult to comprehensively consider many meteorological parameters related to typhoon genesis, so it is necessary to find a new method that can directly link meteorological parameters and typhoon genesis.

Machine learning (ML) is a method that is faithful to all measured data, as well as a mapping method that minimizes errors by training massive data. Support vector machine, one of ML methods, can take into account a variety of meteorological parameters. Its simple classification of high-dimensional mapping can well solve the mapping problem between meteorological parameters and typhoon generation results. And in the context of considering future climate change, as long as the meteorological parameters under future conditions are input into the model, the genesis of typhoons in the future can be predicted, including the starting position and number of typhoons.

2. METHODS

This study proposes the method of using SVM (Piccialli and Sciandrone, 2022; Ribeiro Mendes Junior et al., 2022) combined with AdaBoost (Li et al., 2022) to directly establish the mapping between the typhoon generation result (1 if generated, -1 if not generated) and meteorological data. Meteorological parameters used herein include absolute vorticity (AV) at 850hpa, relative humidity (RH) at 600hpa, vertical velocity (VV), relative sea surface temperature (SST), and vertical wind shear (WS) at 850-200hpa (Tippett et al., 2011).

2.1. Data Division and Evaluation Index

In order to consider the uncertainty caused by the different meteorological agencies, the historical typhoon data are compared using the observations of JMA, CMA, and JTWC. As can be seen in Table 1, whether it is $Accuracy$ and $F1\_score$, Tra_{num}:Tes_{num}=7:3 can always make these two indicators reach the maximum at the same time. When O_{num}:N_{num}≈1, the two planes in Fig. 1 of $Accuracy$ and $F1\_score$ intersect, and both indicators have a large value at this ratio. Therefore, under the condition that Tra_{num}:Tes_{num}=7:3 and O_{num}:N_{num} = 1, the classification results of ten working conditions considering the combination of five meteorological parameters are obtained, as shown in Fig. 2.

Table 1. The impact of dataset ratio on $Accuracy$ and $F1\_score$(%).

Tra_{num}:Tes_{num}	7:1		6:1		5:1		4:1		3:1		7:3		2:1		1:1	
O_{num}:N_{num}	A	F	A	F	A	F	A	F	A	F	A	F	A	F	A	F
1:5	91.89	91.89	91.55	91.55	91.80	91.80	91.91	91.91	92.20	92.20	92.64	92.64	92.48	92.48	92.39	92.39
1:4	91.57	91.57	91.06	91.06	91.31	91.31	91.46	91.46	92.09	92.09	92.55	92.55	92.30	92.30	92.26	92.26
1:3	90.87	90.87	90.71	90.71	90.91	90.91	90.92	90.92	91.67	91.67	92.02	92.02	91.77	91.77	91.43	91.43
1:2	89.55	89.55	89.49	89.49	89.55	89.55	89.60	89.60	90.52	90.52	90.79	90.79	90.58	90.58	90.29	90.29
1:1	89.26	89.26	89.35	89.35	89.25	89.25	89.43	89.43	89.86	89.86	89.75	89.75	89.64	89.64	89.10	89.10
2:1	90.29	90.29	90.45	90.45	90.46	90.46	90.27	90.27	90.68	90.68	90.44	90.44	90.22	90.22	90.11	90.11

Tips: Tra_num refers to Train Number; Tes_num refers to Test Number; O_num refers to Occurrence Number; N_num refers to Nonoccurrence Number, A refers to $Accuracy$, F means $F1\_score$.

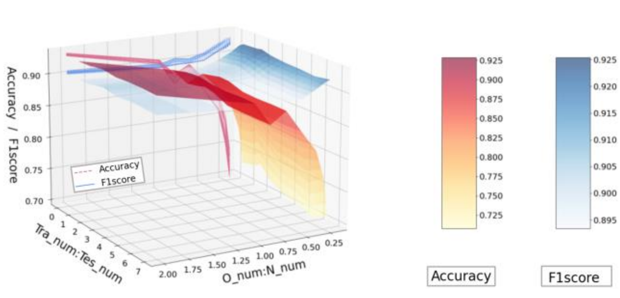


Figure 1. Optimized three-dimensional schematic diagram of $Accuracy$ and F1_score.

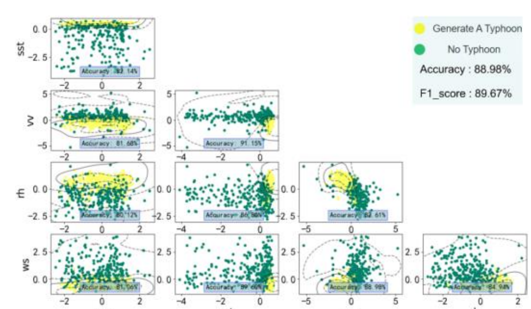


Figure 2. Classification results of test set (Tra_num:$Tes\_num = 7$:3，O_num:$N\_num = 1$).

2.2. Parameter Selection and Optimization

Fig. 2 shows that the result of classification accuracy varies with the combination of parameters, and some of these parameter combinations must have positive or negative effects on the accuracy.

In detail, the row and column where AV and RH are located have the lowest accuracy, so it is only necessary to eliminate these two parameters and recombine the remaining parameters. Table 2 presents the results of only using the Voting method and using AdaBoost and Voting at the same time respectively. It can be found that the improvement of the results is the most obvious after removing AV and RH at the same time under both situations. It can be found using AdaBoost and Voting at the same time increases the $Accuracy$ and $F1\_score$ by 3.5%~5% compared to Voting alone.

Table 2. Influence of meteorological parameter on $Accuracy$ and $F1\_score$(%).

Meteorological Parameter / Evaluation Index	All 5 parameter		Remove AV		Remove RH		Remove AV and RH	
	Voting	AdaBoost +Voting	Voting	AdaBoost +Voting	Voting	AdaBoost +Voting	Voting	AdaBoost +Voting
$Accuracy$	88.76	92.71	88.39	92.19	89.51	94.58	90.64	95.41
$F1\_score$	89.36	92.88	89.05	92.30	89.06	94.39	90.57	95.26

3. RESULTS

The historical reanalysis data of ERA5 from 1979 to 2020 and the typhoon generation data of JMA, JTWC and CMA are used for training respectively. Fig. 3 only shows the simulation results of historical typhoon generation from 2018 to 2020 of JMA. The prediction results of other years are in the same form with a grid resolution of 0.25°. The black circles are the historical typhoon generation positions of JMA, and the heat map represents typhoon-prone areas, which shows a good agreement between forecasts and historical typhoon locations. Fig. 4 shows the results of the interannual distribution regularity obtained by using typhoon records from different meteorological agencies. Fig. 4(a), (b), and (c) respectively represent the comparison of the sliding averages of the annual generation frequency using the typhoon generation data of JMA, JTWC and CMA, and these values have been normalized. Fig. 5 shows the KDE comparison between the historical data and the records from 1979 to 2020.

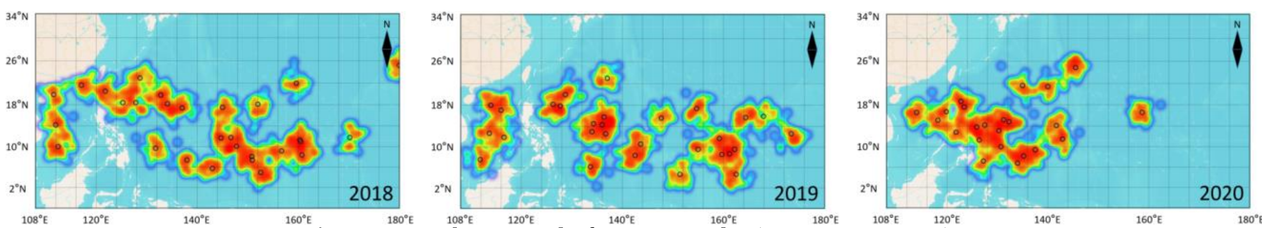
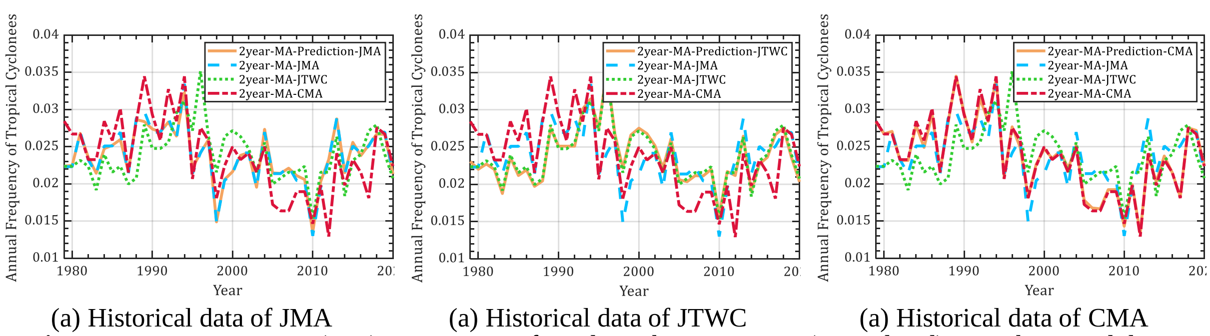


Figure 3. Typhoon yearly forecast results (2017-2020, JMA).



(a) Historical data of JMA (a) Historical data of JTWC (a) Historical data of CMA

Figure 4. Moving average (MA) comparison of yearly typhoons spawns (normalized) using historical data of different agencies (1979-2020).

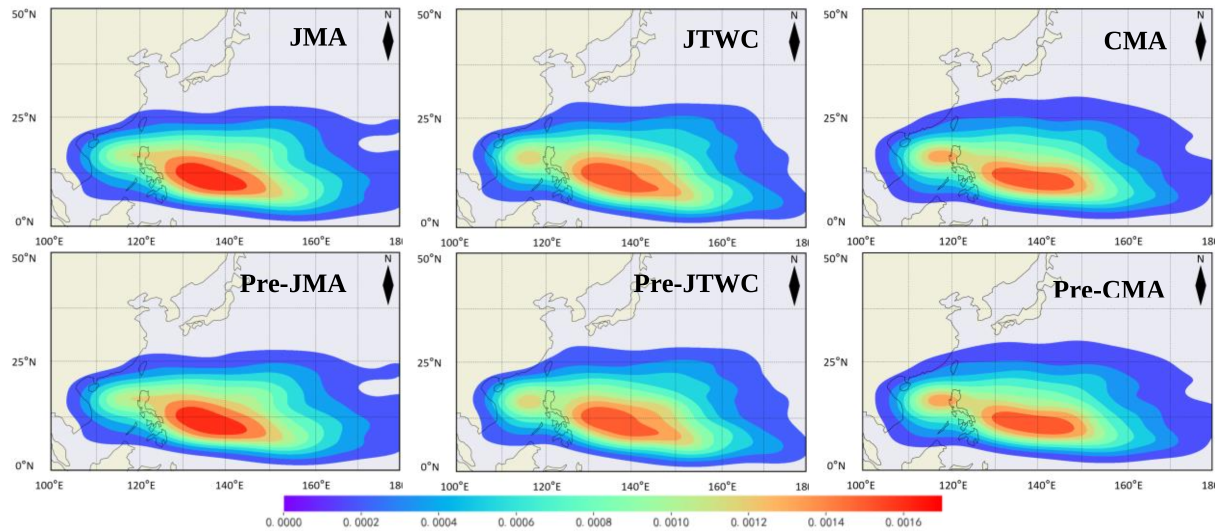


Figure 5. KDE comparison between historical records of different agencies and forecasting results (1979-2020).

4. CONCLUSIONS

In summary, this paper uses machine learning methods to build a typhoon generation model in the WNP region, and the following conclusions are obtained: (1) The distribution of KDE in a single year of different meteorological agencies is obviously different, but with the increase of the number of statistical years, the distribution of KDE gradually converges which shows only partial inconsistencies in special sea areas (12°N-25°N, 110°E-120°E). (2) The data of different meteorological agencies show different pattern in the time course, but the SVM method can always accurately represent this changing pattern. (3) It is found that for the problem of typhoon generation, Tra_{num}:Tes_{num}=7:3， O_{num}: N_{num}=1:1 will obtain the best training results. (4) When only the three meteorological parameters of SST, WS and VV are used to generate predictions, the $Accuracy$ is 2% higher than that of all five meteorological parameters, and the $F1\_score$ is 1% higher. (5) The method of using AdaBoost will increase $Accuracy$ and $F1\_score$ by 3.5%~5%.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the National Natural Science Foundation of China (52108469, 52278520).

REFERENCES

Cao, X., Wu, R., Xu, J., Dai, Y., Bi, M., Lan, X., Zhang, X., 2022. A comparison of tropical cyclone formation over the western North Pacific in August between 1996 and 2014. Atmospheric Research 266, 105952.

Chen, S., Zhang, Y., Xu, J., Shen, W., Ye, G., Lu, Z., 2022. Data assimilation of adaptive observation and application for typhoon forecasts over the Western North Pacific. Atmospheric Research 276, 106274.

Li, J., Lv, J., Wan, A.T.K., Liao, J., 2022. AdaBoost Semiparametric Model Averaging Prediction for Multiple Categories. Journal of the American Statistical Association 117, 495–509.

Piccialli, V., Sciandrone, M., 2022. Nonlinear optimization and support vector machines. Ann. Oper. Res. 314, 15–47.

Ribeiro Mendes Junior, P., Boult, T.E., Wainer, J., Rocha, A., 2022. Open-Set Support Vector Machines. IEEE Trans. Syst. Man Cybern. Syst. 52, 3785–3798.

Tippett, M.K., Camargo, S.J., Sobel, A.H., 2011. A Poisson Regression Index for Tropical Cyclone Genesis and the Role of Large-Scale Vorticity in Genesis. J. Clim. 24, 2335–2357.