Adaptive Wind Gust and Associated Gust-factor Model for the Gust-producing Weather over the Northern South China Sea

HUANG Ling (黄 龄), LIU Chun-xia (刘春霞), LIU Qian (刘 骅)
(1. Guangdong Province Key Laboratory of Regional Numerical Weather Prediction, Guangzhou Institute of Tropical and Marine Meteorology, CMA, Guangzhou 510641 China; 2. School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai, Guangdong 519082 China)

Abstract: Wind gusts are common environmental hazards that can damage buildings, bridges, aircraft, and cruise ships and interrupt electric power distribution, air traffic, waterway transport and port operations. Accurately predicting peak wind gusts in numerical models is essential for saving lives and preventing economic losses. This study investigates the climatology of peak wind gusts and their associated gust factors (GFs) using observations in the coastal and open ocean of the northern South China Sea (NSCS), where severe gust-producing weather occurs throughout the year. The stratified climatology demonstrates that the peak wind gust and GF vary with seasons and particularly with weather types. Based on the inversely proportional relationship between the GF and mean wind speed (MWS), a variety of GF models are constructed through least squares regression analysis. Peak gust speed (PGS) forecasts are obtained through the GF models by multiplying the GFs by observed wind speeds rather than forecasted wind speeds. The errors are thus entirely due to the representation of the GF models. The GF models are improved with weather-adaptive GFs, as evaluated by the stratified MWS. Nevertheless, these weather-adaptive GF models show negative bias for predicting stronger PGSs due to insufficient data representation of the extreme wind gusts. The evaluation of the above models provides insight into maximizing the performance of GF models. This study further proposes a stratified process for forecasting peak wind gusts for routine operations.

Key words: peak wind gust; gust-factor model; weather adaptive; northern South China Sea


1 INTRODUCTION

Wind gusts are short-lived extremes within the spectrum of wind variations that are typically responsible for the worst damage caused by winds\(^1\). Strong wind gusts can damage buildings, bridges, aircraft, and trains and interrupt electric power distribution, air traffic, waterway transport, and port operations\(^2-5\). Accurate forecasts of peak wind gusts are crucial for saving lives, reducing economic losses, and preserving construction.

Wind gusts are complex, small-scale weather phenomena that occur when high-momentum air is brought to the surface. Previous research has suggested that such wind fluctuations are primarily governed by mechanical mixing that depends on factors such as surface roughness, landscape patchiness, and topographical complexity\(^6-8\). Several studies have examined wind gust characteristics in inhomogeneous terrain and land use/land cover, suggesting that wind gusts are strongly associated with wind direction, which partially results from terrain and land cover features\(^9-12\). Limited research has focused on wind gust features over open water with few surrounding obstacles. In addition to the local underlying surface, the characteristics of wind gusts are also highly related to the mesoscale conditions and local climate. Wind gust features among most intensive turbulent weather events, such as severe tropical cyclones\(^13-15\), convective rainstorms\(^16\), and winter storms\(^17\), exhibit visible differences. Therefore, it is necessary to distinguish such gust-producing weather events to accurately describe wind gust characteristics.

Developing an accurate method to forecast the peak gust speed (PGS) is important in weather forecasting services. Forecasting PGSs can be conducted using physically based methods and statistically based methods, each of which has advantages and disadvantages. Current gust parameterizations are based on a physical understanding of boundary layer turbulence and the variables within a numerical weather prediction (NWP) model; thus, an inadequate understanding of boundary turbulent processes could lead to errors in forecasting gusts\(^18\). On the other hand, the statistically-based methods
are data-driven and explore only the statistical relationship between the gust parameters and the common average wind.\cite{19, 20} Such statistical methods can continue to offer new ways of efficiently providing accurate estimates of gust risk by extending observational datasets.

One statistical method, the gust factor (GF) model, employs a combination of climatological measures of local gustiness along with average wind speed forecasting. The GF, affected by various elements, such as wind direction, upstream terrain conditions (roughness length), and anemometer height\cite{9, 21, 22}, is simply calculated as the ratio of the PGS and mean wind speed (MWS) and thus is a climatological measure of gustiness. Previous studies have suggested that the GF model is a viable and skillful method of forecasting wind gusts when the gust factors are stratified by wind speed and wind direction\cite{23}. They further advanced gust research by evaluating the peak wind gust predictions during several types of gust-producing weather phenomena.\cite{24}. Their results suggested that the meteorological stratified GF model performed best during high pressure and nocturnal conditions, was skillful during conditions involving snow, but failed during “rain with thunder” weather. Their findings suggested that gust forecasting models are necessary for considering the different meteorological conditions of gust-producing weather.

The northern South China Sea (NSCS) is located in the monsoonal region of the subtropics. Dominated by monsoonal flow during the warm seasons, the NSCS suffers from heavy monsoonal rainfall (hereafter RAIN) that is associated with mesoscale processes that could produce strong downbursts\cite{24, 25} and landfalling TCs with large-scale rotating circulations\cite{26}. During the cold seasons, the East Asian winter monsoon frequently brings cold air masses to the NSCS. Such cold air outbreaks (CAOs) cause a sharp drop in air temperature\cite{27} and strong wind gusts\cite{28} that deviate from mean climatic states. Existing investigations of the wind gusts over the NSCS seem to be inadequate due to the lack of oceanic measurements, as are the studies focusing on the distinctive gust-producing weather. Thus, an analysis of the wind gust over the NSCS that is stratified by seasons and weather, such as TCs, CAOs, and RAIN, is analyzed in section 3. The summary and discussion are presented in section 4.

2 DATA AND METHODS

2.1 Meteorological data

An enhanced observation network for real-time wind and wind gust measurements during the period of 2015–2021 consists of 14 buoys and oil platforms over the NSCS and 55 automatic weather stations on the coastal island (Fig. 1). The hourly MWS and wind directions reported in this database represent 2-m averages (minutes 58–60 of the hour) of 3-s averages before each standard hour, while the hourly wind gust (PGS) represents the maximum 3-s average occurring each minute within the hour. The 2-m averages are remarkably analogous to the mean wind of the hour (60-min average) in the frequency distribution\cite{22}. These data are subjected to strict quality-control procedures, including the climatological limit value test, internal consistency test and space/time continuity test. The reported gust used to construct the GF model must satisfy two criteria: 1) the reported MWS must be greater than 1 m s\(^{-1}\), and 2) the difference between the reported PGS and the MWS must be \(\geq 1.5 \text{ m s}^{-1}\) to eliminate persistent strong winds\cite{29}. Notably, there is no restriction on the duration of the PGS in the current study. Only the wind gust that met the above criteria (consisting of the gust dataset) and its corresponding GF are analyzed in this study. A total of 1531868 pairs of MWS and PGS meet the above criteria during the 7 years. The corresponding rain-gauge observations provided by the above stations are used to select the MWS and PGS that are impacted by heavy rainfall.

The ERA5 reanalysis data were utilized to present the climatological surface wind field, surface temperature, wind fields and the 24 h temperature drop\cite{30}. ERA5 provides 1-hourly reanalysis at a spatial resolution of 0.25° with 37 vertical pressure levels. The fine horizontal resolution of the ERA5 dataset provides a better
representation of surface conditions, especially over South China, which has complex terrain (Fig. 1). Moreover, the ERA5 for the period covering 7 cold seasons (October to March) from 2015 to 2021 are used to define the arrival of the cold air mass. The climatological distribution of heavy rainfall events is presented by estimated precipitation data from the integrated multisatellite retrievals for global precipitation measurement (IMERG) [31]. The IMERG data have a spatial resolution of 0.1° and a temporal resolution of 30 min, with good performance in land and marine areas [32-34]. This quality-controlled rainfall product has been widely used in rainfall analysis [35, 36].

The TC tracks, and the radius of gale force wind (34 kt or 17 m s⁻¹; R34) that are used to identify the measurements affected by TCs are obtained from the best track archive from the Joint Typhoon Warning Center (JTWC) [37]. This dataset provides not only the TC track information but also some other potentially useful variables, such as wind radii. Information from more than 6 meteorological agencies, such as the Chinese Meteorological Agency (CMA), Hong Kong Observatory (HKO), Meteorological Agency of the United States of America (USA), and Tokyo Meteorological Agency (TOKYO), are differentiates by different techniques used to designate the TC center, wind radii, or TC intensity. Specifically, three kinds of wind radii are provided by the JTWC: 1) When the wind circle is presumed to be round, only one value is provided as the radius of gale force wind (e.g., CMA). 2) When the wind circle is presumed to be an oval, the long radius and the short radius are provided (e.g., TOKYO). 3) When the wind circle is presumed to be irregular, the radii from four quadrants are provided (e.g., USA). The recorded time interval varies from 1 h (CMA when TC passed the 24-h cordon) to 6 h. The R34 value of the CMA and USA are further used to select the TC-impacted gust due to its data integrity.

2.2 Construction and evaluation of the GF model

The GF is the ratio of the PGS to the MWS [38]. If the reporting criteria are met, the GF is defined as:

\[
GF = \frac{PGS}{MWS} \quad (1)
\]

Such a definition of the GF using hourly observations or their equivalents has been used in many other GF studies [39-41].

In the current study, we consider the GF to be mainly impacted by the MWS. Thus, with more than 1,500,000 available observations, a GF model in which the GF is a function of MWS is constructed by performing unary regressive analysis:

\[
GF = \text{func}(MWS) \quad (2)
\]

The forecasted PGS in this study is determined by a simple statistical technique in which a known GF is multiplied by a forecasted MWS:

\[
PGS_{\text{fcst}} = GF \times MWS_{\text{fcst}} \quad (3)
\]

Such a gust forecast (PGS_{fcst}) is a function of the forecast mean wind speed (MWS_{fcst}) and a known GF, which is also a function of the MWS (Eq. 2). Any error in the gust forecast thus includes a contribution from the error associated with the MWS_{fcst} and the representation of the GF. In this study, we replace MWS_{fcst} with an observed MWS (MWS_{obs}) corresponding to the forecast hour, which creates a perfect wind speed forecast by eliminating any errors associated with wspd_{fcst}. With this strategy, the PGS_{fcst} errors are solely due to the representativeness of the GF model that is used. PGS_{fcst} is then verified against the observed gust (PGS_{obs}) regarding the strength of the mean wind by verification metrics, such as bias (PGS_{fcst} – PGS_{obs}) and absolute error (|bias|), following Harris and Kahl [22]. Absolute-error distributions for selected pairs of models are compared via Student’s 𝑡-test and are statistically significant when the null hypothesis (of no difference) is rejected at the 95% confidence level. The final mean bias and absolute error are obtained from the arithmetic means of their corresponding hourly selected pairs.

3 RESULTS

3.1 Overview of the PGS and the associated GF

3.1.1 Climatology of PGS, MWS, and the associated GF

Figure 2 summarizes the distributional fits for the MWS, PGS and GF. The lognormal distribution is the best fit for the MWS and PGS (Fig. 2a), although the Weibull distribution is frequently applied to describe the wind speed distribution [42]. Such lognormal distributions of MWS and PGS have also been reported in recent studies [8, 22, 43]. The maximum probable distribution (MPD) of the MWS in the lognormal distribution reaches 2.11 m s⁻¹, and 80% of the values fall in the area from 1 m s⁻¹ to 6.5 m s⁻¹ (Fig. 2a). Not surprisingly, gusts strongly bias the distribution toward higher speeds, with an MPD of 5.7 m s⁻¹ and a wider distribution range of 5–10 m s⁻¹. Seventy-six percent (67.6%) of the MWS (PGS) values are greater than the MPD. Such strongly right-skewed lognormal distributions of the PGS and MWS require a different dynamic response of wind turbine systems under high loading.

As the ratio of PGS and MWS, the distribution of GF is more similar to that of the MWS (Fig. 2b). The GFs are concentrated in the area of 1–2, with nearly 80% of the GFs smaller than 2 (Fig. 2b). Climatologically, the mean GF over the NSCS during 2015–2021 is 2.12, which is greater than what was found over Lake Michigan (1.74; Harris and Kahl [22]) and New York State in the USA (1.3; Hu et al. [9]), Germany (1.40; Hoffner and Kunz [19]), and coastal South China at 160 m above ground level (1.3; Shu et al. [44]). Previous studies have suggested that the GF tends to decrease with MWS [44]. Thus, the greater GF over the NSCS could be attributed to the weaker MWS.

To accurately construct the GF model, the PGS and the associated GF, stratified by hour and season, are first
investigated (Fig. 3). The diurnal cycle of PGS exhibits seasonal variations. During the warm season (March to August; MAM and JJA), the prevailing wind over the NSCS is the southerly wind from the South China Sea (Figs. 4a-b) under the control of the southerly monsoon with occasionally high impact systems, such as subtropical highs and tropical cyclones. During this period, the diurnal peak of the PGS occurs at 1500 LST in the afternoon, with a climatological mean of 7.24 m s$^{-1}$. In contrast, dominated by the northerly wind in the cold season (Figs. 4c-d; September to February; SON and DJF), the diurnal peak of the PGS shifts toward early morning (0600-0800 LST),

![Figure 2](image)

**Figure 2.** (a) The probability distribution of the mean wind speed (red bars) and peak wind gust (blue bars). The red (blue) contour denotes the lognormal distribution. (b) The probability distribution function of the gust factor (gray bars; left axis) and the accumulative frequency (black line; right axis).

![Figure 3](image)

**Figure 3.** (a) The diurnal variation in the peak gust speed (PGS) from 2015–2021, in DJF (blue), MAM (green), JJA (black), and SON (red). (b) is similar to (a) but for the gust factor.
with a greater climatological mean reaching 8.25 m s\(^{-1}\) (Fig. 3a). Previous studies have suggested that the enhanced sea surface gustiness is strongly related to the growth of the marine boundary layer that is triggered by the monsoon cold air and warmer sea surface temperature that leads to the strong downdraft within the developed convection\(^{[45]}\). Interestingly, the antiphase diurnal variation in PGS between the warm and cold seasons is different from those in previous studies because their diurnal variations are homophases, with the diurnal peak occurring in the early evening (1800 LST; Kahl\(^{[46]}\)). Such diurnal features of PGS are speculated to be associated with the unique geographical condition of the monsoonal region over the NSCS that impacts the seasonal wind direction variation and the seasonal controlling weather system.

The GFs, on the other hand, exhibit less diurnal variation (Fig. 3b). The seasonal GF remains almost constant, with a diurnal amplitude (daily maximum minus daily minimum) of less than 0.1. Nevertheless, the differences in the GF among the four seasons reach 0.12, resulting in a difference of more than 0.9 m s\(^{-1}\) when roughly estimating the gust speed. Specifically, the mean GFs for the four seasons are 2.10 for MAM, 2.20 for JJA, 2.12 for SON, and 2.08 for DJF. Such differences could be attributed to the surface roughness proxied by the wind direction over the highly heterogeneous terrain over the NSCS coast (Figs. 1, 4)\(^{[6, 47]}\). Thus, the difference among the seasonal GFs is highly associated with the wind direction varying with the change in the seasons. Such seasonally-dependent PGS and GF values suggest that the GF model should be formulated based on seasonal conditions characterized by different large-scale meteorological features.

### 3.1.2 CONSTRUCTION OF THE GF MODELS

The GF-MWS relationship is first constructed based on the MWS and GF pairs that meet the gust criteria (section 2.1). As indicated by Fig. 5, the variation in GF with MWS is approximated by an inverse proportional function. Thus, the power function is used to perform the regression analysis:

\[
g_{\text{GF}} = a \times \text{MWS}^b + c \tag{4}
\]

Least squares regression is performed to quantify the GF at a given MWS based on Eq. (4). Several variants of the GF models in Eq. (4) are constructed using different subdatasets described as follows. The general GF model (hereafter “General GF”) is constructed using the PGS and MWS of the entire period (Jan.-Dec., 2015–2021). Four adaptive GF models (hereafter “Adaptive GF”) are constructed based on the subdatasets from the four seasons (MAM, JJA, SON, and DJF). The parameters of each GF model are shown in Table 1. In addition to the parameterized GF, the mean GFs (hereafter “mean GF”) derived from the averaged GF corresponding to the entire year and the four seasons are also examined for their representativeness.

Notably, the current constructed GF model considers
the GF as a function of only the MWS, which shows a strong correlation with the GF (Fig. 5). Multivariate regression analysis, which includes multiple variables that could impact the GF, such as wind direction and roughness length, is beyond the scope of the current study. Such a multivariate GF model is further established in a subsequent study based on machine learning techniques.

Of note is that the construction of the GF model is established based on the pairs of MWS-GF in which MWS is greater than 1 m s\(^{-1}\). It is well known that the GF has significant variations at low wind speed, as shown in Fig. 5. Some previous studies have constructed the GF model based on the MWS greater than 5 m s\(^{-1}\) by the one-order polynomial. The follow-up verification suggested that the GF model fitted by the subdataset with greater MWS significantly underestimated the GF at higher MWS, showing the greater underestimation for PGS when the MWS\( \geq 20 \) m s\(^{-1}\). Therefore, to gain a better representation of PGS at higher MWS, the dataset with MWS greater than 1 m s\(^{-1}\) is still used to construct the GF model.

### 3.1.3 VERIFICATION OF THE GF MODELS

In this study, we replace \( \text{wspd}_{\text{est}} \) in the GF model with MWS\(_{\text{obs}}\) at the forecast hour. Thus, the performance of the GF model is purely impacted by its representativeness in forecasting the PGS. The PGS\(_{\text{fact}}\) derived from Eq. (3) are evaluated by stratified MWS values, which are categorized into four groups, namely, weaker MWS (1 \( \leq \) wind speed < 5 m s\(^{-1}\)) moderate MWS (1 \( \leq \) wind speed < 10 m s\(^{-1}\)) strong MWS (1 \( \leq \) wind speed < 20 m s\(^{-1}\)) and intensive MWS (wind speed \( \geq 20 \) m s\(^{-1}\)) Such category covers the CMA wind classification of grade 1–3, grade 4–5, grade 6–8, and grade larger than 9. The General GF, Adaptive GF, and mean GF models are first compared by the absolute bias (Fig. 6) of the entire dataset and the seasonal subdatasets. The mean GF models show greater absolute bias than the other two models, especially at higher wind speed scenarios. The General GF and Adaptive GF models are significantly different even though they have similar absolute errors (passing the 95% significance test).

Specifically, the absolute errors of the mean GF model for weak MWSs are less than 1.5 m s\(^{-1}\) (Fig. 6a). The errors increase to over 4 m s\(^{-1}\), 10 m s\(^{-1}\), and even 20 m s\(^{-1}\) when the mean winds intensify (Figs. 6b–d). It is not surprising that the constant GF (~2) can easily underestimate the PGS\(_{\text{fact}}\) when the MWS is lower (< 15 m s\(^{-1}\)), while it overestimates the PGS\(_{\text{fact}}\) when the MWS is higher (> 15 m s\(^{-1}\); Fig. 5). In contrast, the GF varies with the MWS in the parameterized GF models (General GF and Adaptive GF). Thus, their inversions to PGS\(_{\text{fact}}\) are less deviated (Fig. 6).

The General GF and Adaptive GF have comparable mean absolute bias values but are significantly different in their distributions according to Student’s \( t \)-test (Fig. 6, 7), except for in MAM and JJA in the intensive MWS category. Thus, the General GF and Adaptive GF are further investigated for their forecasted gust bias in each season at various MWSs (Fig. 7). The performances of the General GF and Adaptive GF vary by season and mean wind intensity. Specifically, for MAM and JJA, when the NSCS is under prevailing weaker southerly winds with frequently occurring summer rainfall (Figs. 3a, 4a–b), the General GF tends to overestimate the PGS at a weaker MWS (<10 m s\(^{-1}\)), while the forecast gust error from the Adaptive GF is almost negligible but shows overestimation at a higher wind speed scenario (>10 m s\(^{-1}\); Figs. 7a, b). For SON, when the NSCS is occasionally impacted by high wind weather (e.g., CAO), the bias of the General GF is significantly higher than that of the Adaptive GF (Figs. 7c, d), while the bias of the General GF in DJF is smaller than the Adaptive GF due to the inadequate of the Adaptive GF samples. When the MWS is greater, the GF models tend to underestimate the PGS, even for the Adaptive GF (Figs. 7c, d). Such underestimation might be due to the failure of the seasonally stratified model to represent high-impact weather. The inconsistency of underestimation or overestimation between the General GF and Adaptive GF suggests that the construction of the GF model should consider seasonal variations in meteorological conditions, such as prevailing wind and dominant weather types. The Adaptive GF might still exhibit great bias when high-impact weather is loaded. Thus, the construction and evaluation of GF models at high wind speed events, such...
Figure 6. The mean absolute bias for the General GF models (dark gray), adaptive GF models (gray), and mean GF models (light gray) for the whole year, MAM, JJA, SON, and DJF, stratified by (a) weaker mean wind speed, (b) moderate mean wind speed, (c) stronger mean wind speed, and (d) intensive mean wind speed. The numbers over the bars indicate the number of observations within each category. The dots indicate that the absolute biases between the General and adaptive GF models are significantly different at a 95% confidence level.

Figure 7. Similar to Fig. 6 but for mean biases.
as TCs, CAOs, and RAINs (Fig. 8), are considered accordingly in the following sections.

3.2 Wind gusts on TC days

This section investigates the wind gust characteristics and the performance of the GF model on TC-impacted days. Observations that are impacted by TCs are included if they fall within the circle of R34. These identified observations are then used to perform regression analysis to construct the formula of the GF models. The TC observation-based GF model is then further evaluated.

3.2.1 Wind gust parameters in four quadrants

Figure 8a shows the tracks of the TCs that influenced the NSCS. A total of 46 TCs impacted the NSCS during 2015–2021, including Super Typhoon Mujigae (2015), Hato (2017), and Mangkhut (2018). The observational network monitors the variation in meteorological conditions when TCs make landfall in South China. Such quality-controlled oceanic observations capture the sharp increase in the wind speed (Fig. 9a), the change in the wind direction (Fig. 9b), and the sudden drop in surface pressure (Fig. 9c) when the TC approaches the station (Fig. 9d).

The R34 values from the CMA (hereafter CMA-R34) and USA (hereafter USA-R34) are used to select the observations that are influenced by landfalling TCs. To simplify the selection, the TC tracks provided by the CMA are used to locate the real-time positions of the TCs. Given that the root mean square error (RMSE) of the latitudes and longitudes between the USA and CMA only reaches 0.2°, the significant differences in the TC gust defined by the CMA-R34 and USA-R34 are solely due to their recorded radii.

The area circled by R34 is divided into four quadrants based on the movement direction of the TC. This area dynamically varies with the TC moving direction at a recorded hour \( T_0 \) that is defined by the locations of the TC at \( T_0 \pm 6h \). Four quadrants are divided by the parallel and vertical axes to the movement direction. The quarter areas are defined as the northeastern quadrant (NE), northwestern quadrant (NW), southwestern quadrant (SW), and southeastern quadrant (SE). The observations at \( T_0 \) that fall into each quadrant with a distance from the TC center of less than R34 are categorized accordingly.

The R34 values provided by CMA and the USA are used to define the observations that are impacted by TCs. Due to the statistically shorter R34 of the USA (not shown), the number of observations in each quadrant of the CMA is almost twice that of the USA (Fig. 10).

The comparison of the gust statistics from the USA-R34 and CMA-R34 in the four quadrants is shown in Fig. 10. The PGS from CMA-R34 is generally greater than that from USA-R34 (Figs. 10a, b). For instance, the median of the PGS of the CMA in the NE quadrant is 17.7 m s\(^{-1}\), which is closer to 34 knots (17.5 m s\(^{-1}\)) than that from the USA (14.3 m s\(^{-1}\)). The closer the PGS is to the 34 knots defined by CMA-R34, the more accurate the R34 obtained by the CMA. In both the CMA-R34 and USA-R34 groups, the strongest PGS occurred in the NE quadrant, followed by that in the NW, SE and SW quadrants (Fig. 10a). This finding is consistent with reports from previous studies that the strongest gusts frequently occurred in the northeastern quadrant\(^{[48, 49]}\). Note that the smaller number of observations in the SW quadrant partly contributes to its weaker estimated speed since insufficient

Figure 8. (a) TC tracks impacting the key region from 2015–2021. The red line denotes the track of Super Typhoon Mujigae (2015) used in Fig. 9. The red dot in (a) denotes station G2424 in Fig. 9. (b) The average surface temperature anomaly on the CAO-impacted days from 2015–2021. (c) The hourly average rainfall distribution when the observed hourly heavy rainfall (> 5 mm h\(^{-1}\)) from the AWS impacts South China. The black dots indicate the distribution of the oceanic observations.
Figure 9. An example of oceanic observations documenting TC transit (Super Typhoon Mujigae. Oct. 02–06, 2015. G2424 in Fig. 8).

Figure 10. (a) Violin plot of the peak gust speed (PGS) of CMA-R34 (pink) and USA-R34 (light blue) in four quadrants. (b) is similar to Fig. 2a, but for the PGS impacted by the TCs, using CMA-R34 (pink) and USA-R34 (light blue). (c) is similar to (a) but for the gust factor (GF). (d) is similar to Fig. 2b, but for the GF impacted by the TCs, using CMA-R34 (pink) and USA-R34 (light blue). The numbers in (a) denote the numbers of observations in the four quadrants identified by the CMA-R34 and USA-R34. The long dashed (short dashed) lines in the violin plot denotes the quartiles.
data representativeness undermines the possibility of capturing extremes. The weaker gust speed of USA-R34 (Fig. 10b) might also be affected by the same statistical errors.

In contrast, the probability distribution functions of the GF from the USA and CMA are comparable in the four quadrants (Fig. 10c), with similar medians (1.8), means (2.1), and MPDs (1.5) for both CMA-R34 and USA-R34. Such GF similarities in the four quadrants suggest the construction of a uniform GF model for the TC dataset. For the observations that were impacted by TCs, approximately 60% of the GFs for the entire TC dataset range from 1 to 2 (Fig. 10d). This proportion is less than that of the entire dataset (Fig. 2b), indicating a greater GF in the TC events. The different probability distribution functions between the TC dataset and the entire dataset (Figs. 2b, 10b) suggest the necessity of constructing a weather-adaptive GF model for TC events.

3.2.2 CONSTRUCTION AND EVALUATION OF THE GF MODEL FOR TC EVENTS

The observations that were impacted by the approaching TC are collected as the TC dataset to construct the GF-MWS relationship in TC events (hereafter TC-GF) by the least-squares-regression method. The GF model also follows the rule of the power function:

$$\text{GF} = 4.275 \times \text{MWS}^{0.677} + 0.833 \quad (5)$$

The PGS_{fact} to be evaluated is obtained by multiplying the GF and MWS. Fig. 11 shows the stratified evaluations of the General GF and TC-GF models for representing the GF-MWS relationship. The General GF tends to greatly underestimate the PGS under all scenarios (Fig. 11), with the largest mean bias reaching $-7 \text{ m s}^{-1}$. This underestimation is due to the larger sample of the General GF that could compromise its representation of the higher wind speed scenarios of TC events. The TC-GF model, on the other hand, shows a better representation of TC-impacted PGS. Specifically, the mean bias of the PGS_{fact} derived from the TC-GF falls within a range of $[-2, 2] \text{ m s}^{-1}$, except for the PGS in the SE quadrant in the intensive wind speed scenario (Fig. 11d). The overestimation or underestimation of the PGS_{fact} that is obtained by the TC-GF model is dependent on the location relative to the TC center. The results suggest that the TC-GF model tends to overestimate the PGS in the NW and SE quadrants stratified by weaker, moderate and stronger mean wind speeds. When it comes to the intensive MWS, the TC-GF model significantly overestimates the PGS. Although the data representation is speculated to be associated with the GF model performance, we can still refer to such error characteristics in further operational forecasting.

3.3 WIND GUSTS ON CAO DAYS

The current section investigates the characteristics of the wind gust on CAO days and the performance of the associated GF model. Observations that were influenced by the cold air approaching South China are identified to perform the regression analysis. The constructed CAO observation-based GF model is further evaluated.

3.3.1 DEFINITION OF THE CAO

The observations on days when the CAO influenced South China are included in the CAO dataset. The days

![Figure 11](https://example.com/figure11.png)

Figure 11. Similar to Figure 7 but for TC events.
impacted by the CAOs are defined as the days on which cold airmass fluxes reach South China. Cold air masses and their southward fluxes are quantitatively described using the isentropic analysis method\(^{[50]}\). Following Liu et al.\(^{[51]}\), a cold airmass flux \((F)\) is calculated by vertically integrating the horizontal wind \((v)\) within the cold airmass:

\[
F = \int_{\phi_1}^{\phi_2} P(\theta_F) v \, dp
\]

where \(P\) is the ground surface pressure and \(P(\theta_F)\) is the pressure where the potential temperature \((\theta_F)\) is 290 K. Furthermore, the southward flow of the cold airmass at a particular latitude \((\phi)\) can be defined as:

\[
F_\phi = \frac{\cos \phi}{g} \int_{\lambda_1}^{\lambda_2} F_v \, d\lambda
\]

where \(v\), \(a\), \(g\), and \(\lambda\) indicate the meridional wind, the radius of Earth, gravitational acceleration, and the distance in longitude between \(\lambda_1\) and \(\lambda_2\) at latitude \(\phi\), respectively.\(^{[52]}\) In this study, the specific latitude is 22° N, and the longitude ranges from 109°E to 118°E (Fig. 12). The CAO-impacted days exhibited significant negative surface temperature anomalies relative to the climatological mean of the cold seasons (SON and DJF) from 2015–2021 (Fig. 8b).

An index measuring the intensity of the CAO (CAOI) is defined by the daily deviation of \(F_{22^\circ N}\) normalized by the standard deviation:

\[
\text{CAOI}_\phi = \frac{F_\phi - F_{\text{avg}}}{\sigma_{F_\phi}}
\]

The overbar in the equation indicates the average of the study period from 2015 to 2021. Cold air outbreak (CAO) that causes intensive gusts are identified when the CAOIs are greater than 1.5 \(^{[51]}\). Events with CAOI values exceeding 1.5 are further categorized into weak CAO (1.5 \(\leq\) CAOI < 3.0), moderate CAO (3.0 \(\leq\) CAOI < 4.5), and strong CAO (CAOI \(\geq\) 4.5) events to investigate the relationship between the PGS and the intensity of the CAO, which is strongly related to the 24-h change rate of the surface air temperature over South China (Fig. 12).

### 3.3.2 Statistics of the Gust Parameters on CAO Days

A total of 74766 observations during 41 CAO events within 94 days from November to March 2015–2021 are included for further analysis. These CAO events consist of 68% weak CAOs, 22% moderate CAOs, and 10% strong CAOs. The occurrence of CAOs in each category shows seasonal variations in which moderate and strong CAOs are found to occur in DJF, while the peak occurrence of weak CAOs occurs in December. Most (89%) of the CAOs with a weak CAOI have a duration of less than 2 days. The durations of moderate CAOs are uniformly distributed from 1 to 6 days, while those of strong CAOs tend to be long-lasting (>3 days).

The characteristics of the PGS and GF that are associated with the CAOI are further examined in Fig. 13. The PGS shows a good correspondence with the CAOI, with a higher mean and median as the CAOI increases (Fig. 13a). The mean PGS on CAO days (10.19 m s\(^{-1}\)) is weaker than that on TC days (15.14 m s\(^{-1}\)) but stronger than the average gust speed in DJF from 2015 to 2021 (8.25 m s\(^{-1}\)). In contrast, the GF in all the categories appears similar, with comparable median and mean values (Fig. 13b). Such similarity suggests that the unified GF model should be applied to all CAO events despite their different CAO intensities.

### 3.3.3 Performance of Gust Parameterizations on CAO Events

Regression analysis is performed using the CAO dataset obtained in section 3.3b. The formula (CAO-GF) resembles that of the entire dataset but with different
parameters:

\[ GF = 3.817 \times MWS^{-0.686} + 0.662, \quad (9) \]

Stratified evaluations of the General GF and CAO-GF model representations of the CAO dataset are further compared (Fig. 14). The General GF tends to underestimate the PGS in all CAOI and all MWS scenarios, with mean biases varying from \(-2.4 \text{ m s}^{-1}\) to \(-0.4 \text{ m s}^{-1}\). In contrast, the CAO-GF shows less bias, especially when the MWS is weaker than \(20 \text{ m s}^{-1}\) (Figs. 14a-c). Notably, the intensive wind samples are inadequate (Fig. 14d); thus, the mean bias of the forecasted PGS has a lower confidence level.

### 3.4 Wind gust and the GF model on rainy days

The observations impacted by heavy rainfall are selected from the entire dataset when the corresponding hourly rainfall exceeds \(5 \text{ mm h}^{-1}\), excluding days when South China was influenced by the TC and CAO. A total of \(56911\) PGS and MWS pairs are included in the RAIN dataset. The RAIN and entire dataset shared similar diurnal variations and probability distributions (Fig. 15a), while the GF of the RAIN dataset biased toward greater values (Fig. 15b). Furthermore, the \(56911\) PGS and MWS pairs are divided into three groups based on the hourly rainfall amount \((r)\): weak rainfall \((5 \leq r < 15 \text{ mm})\), moderate rainfall \((15 \leq r < 30 \text{ mm})\), and intensive rainfall \((r \geq 30 \text{ mm})\). Further analysis suggests indistinguishable distributions of the PGS and GF among the above categories (not shown).

Given the substantial bias of the RAIN-GF model (Fig. 15b), the weather-adaptive GF model should be

![Figure 13](image) Boxplots of (a) peak gust speed (PGS) and (b) gust factor (GF) for all, weak, moderate, and strong CAO events. The red lines indicate the median PGS (GF). The hollow circles indicate the mean PGS (GF). The upper limbs (lower limbs) indicate values that are >1.5 times the IQR above the 75\(^{th}\) percentile (below the 25\(^{th}\) percentile).

![Figure 14](image) Similar to Fig. 11 but for CAO events for categories of weak CAOs, moderate CAOs and strong CAOs.
considered. Using regression analysis, the GF model obtained from the RAIN dataset is shown as follows:

\[
GF = 3.218 \times \text{MWS}^{-0.861} + 1.007 \tag{10}
\]

This GF model (RAIN-GF) is approximately unbiased in representing the PGS when the MWS is weaker than 10 m s\(^{-1}\) but overestimates the gust speed by 0.2 m s\(^{-1}\) in the stronger MWS scenario. In contrast, the General GF shows a negative bias, which increases as the mean wind intensifies (-0.4 m s\(^{-1}\) to -0.8 m s\(^{-1}\); Fig. 16). In comparison to the performance of the General GF in TC and CAO events, the General GF better represents RAIN events (Figs. 11, 14), corresponding to the indistinguishable distribution of the PGS of the climatology (Fig. 15a).

4 SUMMARY AND DISCUSSION

Peak wind gusts pose a potential threat to construction sites, transportation, and outdoor activities. Understanding and accurately forecasting the peak gust speed (PGS) is essential for saving lives and preventing economic losses. The gust factor (GF), as the ratio of the PGS and mean wind speed (MWS), is usually used as the parameter for forecasting the PGS, which is not a direct output of the numerical model. The current study investigates the climatology of the PGS and the associated GF of oceanic measurements over the northern South China Sea (NSCS) from 2015 to 2021. Characteristics of the hourly PGS and GF in stratified meteorological conditions are investigated, including seasons and gust-producing weather such as tropical cyclones (TCs), cold air outbreaks (CAOs), and heavy rainfall (RAIN). The findings are as follows:

1) The climatological mean of the PGS over the NSCS varies with seasonal meteorological conditions, ranging from 7.18 to 8.44 m s\(^{-1}\). The PGS exhibits apparent diurnal variability, with the diurnal peaks in MAM and JJA occurring in the afternoon and those in SON and DJF occurring in the early morning. This phase diversity across seasons is likely associated with changes in wind direction.

2) The PGS in gust-producing weather exhibits

![Figure 15](image1)

![Figure 16](image2)

**Figure 15.** (a) Probability distribution function of the PGS of the entire climatology (black), TC events (blue), CAO events (green), and heavy rainfall events (red). (b) is similar to (a) but for the GF.

**Figure 16.** Similar to Fig. 11 but for RAIN events for categories of weak rainfall, moderate rainfall, heavy rainfall and intensive rainfall.
followed by that on CAO days, all days and RAIN days. Heavy rainfall days shared a PGS distribution comparable to that of the climatology (Fig. 15a).

3) The GF is more stable and constant than the PGS. The climatology of the GF is 2.12, but it slightly varies seasonally (2.10 for MAM, 2.20 for JJA, 2.12 for SON, and 2.08 for DJF), which might be attributable to the different strengths and directions of the prevailing winds as the seasons vary. The distribution of the GF in different gust-producing weather is less distinguishable than that of the PGS, with the RAIN dataset showing greater GF values (Fig. 15b).

A variety of GF models are constructed for different gust-producing weather types or seasons. Climatology GF models using the climatological mean of the GF, the General GF model and the adaptive GF models (seasonally-adaptive and weather-adaptive) using the empirical regression relationship between the GF and MWS are evaluated and compared by stratified MWS in different scenarios (seasons, TC, CAO, rainfall). In such models, the gust speed is normally estimated by multiplying the GF by a wind speed forecast. However, we replace the mean wind speed forecast with the observed mean wind speed during the forecast hour. The derived gust speed represents the best-case scenario for the GF models, whose forecast errors are entirely attributable to the representation of the GF model itself. The errors are therefore small but provide the lower bound of forecast errors in operational applications. The results are still relatively meaningful, providing insight into how to best maximize the performance of GF models by revealing the following:

1) The adaptive GF models skillfully forecast the PGS compared with the climatology GF models. These adaptive GF models show an even higher skill when the MWS < 20 m s\(^{-1}\).

2) The General GF model tends to underestimate the PGS in all MWS scenarios for gust-producing weather types (TC, CAO, rainfall). The model performance improves when adaptive GFs are used, with the most improved GFs on TC-impacted days.

3) The General GF model shows comparable skills in representing gust-producing weather that is stratified by weather features (i.e., TC quadrant, CAO intensity, rainfall intensity), while distinguishable differences occur when the MWS is stratified. Such findings indicate that the application of adaptive GF model skills should consider the MWS more than specific weather features.

Notably, constructing and evaluating the Adaptive GF model using the same dataset across weather types inevitably leads to better PGS estimates than the General GF model does. The comparison between the performance of the two models aims to show the extent of representing errors when the uniform GF model is used. Such a simple statistic model concerning only one variable is suggested to be comparable to other modeling approaches\[^{[53, 54]}\]. Our analysis also suggests that the Adaptive GF models are viable means of forecasting the gust speed under different meteorological conditions. Given the limitations of the current GF model, further work will 1) construct a GF model by machine learning techniques, where the GF is predicted as a function of multiple variables, such as temperature, wind direction, sea surface pressure, and friction velocity, 2) construct a station-by-station GF model to account the impact of the local construction and terrain on wind gust forecast, and 3) construct a stratified operational model for forecasting gust speed using the actual forecast mean wind (Fig. 17). The errors of such operational forecasted gust speeds will therefore be attributable to the forecast error of the mean wind speed and the representativeness of the GF models. Such errors will be further investigated to finally maximize the performance of the GF model and provide a better gust speed forecast for weather service operations.

**Figure 17.** Schematic diagram of the process of forecasting peak wind gusts in real-time routine operations.
REFERENCES


[18] LIU X, CHENG X, WU Q, et al. Some characteristics of the

HUANG Ling, LIU Chun-xia, LIU Qian. Adaptive Wind Gust and Associated Gust-factor Model for the Gust-producing

Citation:


