Hindawi
Advances in Meteorology
Volume 2022, Article ID 3298501, 11 pages
https://doi.org/10.1155/2022/3298501

*Research Article*

# Sources of Forecast Errors for Rainstorms in the South China Monsoon Region

**Lin Lin**,[1,2] **Chuhan Lu**,[1] **and Feifan Zhou**[2,3]

[1]*Key Laboratory of Meteorological Disaster, Ministry of Education/Joint International Research Laboratory of Climate and Environment Change/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[2]*Laboratory of Cloud–Precipitation Physics and Severe Storms (LACS), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*
[3]*University of Chinese Academy of Sciences, Beijing 100049, China*

Correspondence should be addressed to Feifan Zhou; zhouff04@163.com

Received 3 September 2021; Accepted 22 January 2022; Published 26 March 2022

Academic Editor: Ismail gultepe



The possible sources of forecast errors associated with rainstorms in the South China monsoon region were investigated based on Weather Research and Forecasting (WRF) model forecasts for 19 rainstorm cases that occurred in the past 13 years. Two datasets were separately selected as the initial fields of WRF with the same physical parameterization schemes. By investigating the improvement rate of the forecast when using one set of data rather than the other, the important degree of the initial conditions with respect to the forecasts for each case has been obtained. For those initial errors are the important sources of forecast errors, we further explored the source of the initial errors by comparing the two initial conditions. It was found that, interestingly, the significant differences between two initial conditions are all located upstream the rainfall area, with a distance of 5° of longitude away and an area of about $4^\circ \times 4^\circ$. Based on this, we developed a new method (which we refer to as the "guide flow method") to identify the sensitive area for rainstorm forecasts in the South China monsoon region and then examined the efficiency of the sensitive areas. It was found that reducing the initial errors in the sensitive areas leads to better forecast results than doing the same in other areas. Thus, the sensitive areas are the source areas of forecast errors for rainstorms in the South China monsoon region.

## 1. Introduction

South China is affected by both the tropical and subtropical monsoon, in which the rainstorm process is closely related to the circulation of the summer monsoon [1, 2]. Rainstorms are one of the most serious types of natural disasters in South China [3–5]. However, currently, most global numerical models still carry considerable uncertainty in their forecasting of the location and precipitation amount of rainstorms [6, 7]. In particular, the forecasting of rainstorms in the South China monsoon region has always been a difficult and challenging issue in China's operational forecasting sector [8–10].

In recent years, most researchers have tended to focus on the underlying physical mechanisms of rainstorms in the South China monsoon region. Some of these studies have found that, during the active period of the summer monsoon, the South China Sea summer monsoon can extend to the rainstorm area in the form of low-frequency oscillations, and the rainstorm process has a good correlation with the pulsation or strengthening period of the monsoon [2]. After the onset of the summer monsoon, the Bay of Bengal and the Indian Ocean contribute significantly to the quantity of water vapor in South China [11]. Moreover, both the average precipitation and the convection intensity in South China have generally increased, and the convection intensity is basically consistent with the subseasonal changes in atmospheric thermodynamic conditions, which probably leads to the occurrence of regional extreme precipitation [12, 13]. For example, two strong rainstorms in Dongguan in July

2011 were caused by the onset of the southwest monsoon, the northward uplift of the ITCZ, and the pressure of the high-altitude East Asian trough [14].

In addition to the direct influence of summer monsoon circulation on precipitation, the rainstorm process in the South China monsoon region is also closely related to the prevailing mesoscale system under the influence of the summer monsoon [15]. For instance, the rainstorms on 11 May 2014, 19 May 2015, and 27 August 2018 were all caused by one or more mesoscale convective systems [16–19]. The evolution of the monsoon low-pressure intensity is synchronized with the daily distribution of the rainstorm area's size but not completely synchronized with the maximum daily precipitation's day-to-day evolution [17].

To study the predictability of rainstorms in the South China monsoon region, it is not enough just to understand the underlying physical mechanism of the rainstorm process; studying the forecast error of the numerical model is also necessary [20–23]. With continuous improvement in model accuracy, increasing the horizontal resolution of the model may improve the effects of the precipitation forecast [24–26]. Nevertheless, high-resolution models are still unable to reasonably reproduce the characteristics of the precipitation's distribution owing to the inherent unpredictability of the atmosphere, the imperfection of the numerical method in the dynamic framework of the physical processes, and so on, meaning current numerical precipitation forecasts usually carry large uncertainty [27].

In 2013, the China Meteorological Administration initiated the South China Monsoon Precipitation Experiment, which included studies on improving models and their initial fields [28–30]. Several researchers have found that improving the initial value assimilation technology of wind profile data can better describe the development of the convective system. Additionally, improvement of the low-level water vapor and wind field in the Weather Research and Forecasting (WRF) model data assimilation system can reduce the forecast error of heavy rain, thereby improving the prediction skill [31]. With the occurrence and development of precipitation, the evolution of errors gradually develops from local growth to global propagation, and the initial error in the precipitation area makes an important contribution to the precipitation forecast error [32]. A number of researchers have also studied the sensitivity of physical parameters and related conditional nonlinear optimal disturbances in the Global/Regional Assimilation and Prediction System for the forecasting of heavy rain in South China [33]. For instance, Lu et al. [34] studied a case of heavy rain in the South China monsoon region in 2015 and found that the initial error was an important source of forecast error, and the sensitive area could be found more accurately by the moist energy of perturbations.

Despite the clear need to explore the sources of error in rainfall forecasts in the South China monsoon region, there have been relatively few studies in this regard and so further exploration is still required. The present reported research sought to address this knowledge gap by developing a guide flow method to identify the sensitive area for rainstorm forecasts in this region. The overarching aim was to strengthen our understanding of the source areas of forecast error and improve the forecasting ability in the South China monsoon region.

## 2. Materials and Methods

*2.1. Data and Study Area.* Nineteen heavy rainstorm cases in South China in the past thirteen years that were mainly a result of the summer monsoon (rather than typhoons, etc.) were selected. In terms of observational data, the hourly precipitation gridded dataset (version 1.0) of the Chinese automatic station combined with the CMORPH (Climate Prediction Centre morphing technique) precipitation product (https://data.cma.cn/data/detail/dataCode/SEVP_CLI_CHN_MERGE_CMP_PRE_HOUR_GRID_0.10/keywords/CMORPH.html) was selected (hereinafter referred to as OBS). Besides, China Ground International Exchange Station daily climate data (http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_MUL_DAY_CES_V3.0.html) were used to assist in screening the precipitation cases. The observed cumulative precipitation and distribution during the precipitation period of these cases are given in Table 1. From the data in Table 1, the daily mean maximum site precipitation of the selected 19 cases is 83.35 mm. Details of the different datasets used in the present study are given below.

NCEP_FNL (Final Global Data Assimilation System of the National Centers for Environmental Prediction) analysis data with a spatial resolution of 1.0° were used to generate the input field of the WRF model. According to Huang and Luo's [13] analysis of the five-day precipitation forecasts in three seasons (2013–2015) in South China, the European Centre for Medium-Range Weather Forecasts (ECMWF) model generally has high forecasting skill. Therefore, it can be assumed that both the initial field and the model of ECMWF have high accuracy. The ECMWF output (obtained from the TIGGE (THORPEX Interactive Grand Global Ensemble) dataset of the ECMWF center, hereinafter referred to as TIGGE_EC) was used as a control group with a spatial resolution of 0.5° latitude by 0.5° longitude. The initial conditions of TIGGE_EC were used to generate accurate initial conditions for the WRF forecasts.

*2.2. WRF.* Version 3.6.1 of WRF was employed in the present study. Both the initial fields and boundary fields were generated using FNL analysis data with a time interval of six hours. The horizontal resolution of the forecast was 3 km, with $1100 \times 700$ grid points, covering the entire South China region. Moreover, there were 60 layers in the vertical direction and the integration time step was 15 s without nesting. The physical parameterizations were as follows: Thompson scheme for the microphysical scheme; Goddard shortwave scheme for the shortwave radiation scheme; Rapid Radiative Transfer Model scheme for the longwave radiation scheme; Eta Mellor-Yamada-Janjic TKE (turbulent kinetic energy) scheme for the boundary layer; and the land surface process scheme adopted the Noah land surface model scheme, without cumulus parameterization [16, 34].

Table 1: Details of the cases used, including the time period, distribution, and cumulative amount of precipitation.

| Case | Precipitation period | Precipitation distribution | Maximum site precipitation (mm) | Measured cumulative precipitation (mm) |
|---|---|---|---|---|
| 1 | 2008.6.8(00 : 00)–6.9(00 : 00) | Central and western south China | 164.9 | 582967.4 |
| 2 | 2008.6.11(00 : 00)–6.13(00 : 00) | Central and eastern south China | 297.9 | 698194.2 |
| 3 | 2010.5.31(00 : 00)–6.2(00 : 00) | Central south China | 87.0 | 983323.4 |
| 4 | 2011.6.28(00 : 00)–6.30(00 : 00) | Central and southern south China | 448.5 | 506162 |
| 5 | 2011.7.15(00 : 00)–7.17(00 : 00) | Eastern south China | 134.2 | 477601.5 |
| 6 | 2013.5.14(00 : 00)–5.17(00 : 00) | Central and eastern south China | 195.3 | 705305.8 |
| 7 | 2013.7.26(00 : 00)–7.28(00 : 00) | Southern south China | 189.3 | 567458.6 |
| 8 | 2014.5.21(12 : 00)–5.23(12 : 00) | Eastern south China | 115.7 | 775429.6 |
| 9 | 2015.5.16(00 : 00)–5.17(00 : 00) | Southern and eastern south China | 140.3 | 155520.7 |
| 10 | 2015.5.22(00 : 00)–5.23(00 : 00) | Southern south China | 62.9 | 470785.5 |
| 11 | 2016.6.14(00 : 00)–6.15(00 : 00) | Eastern south China | 68.7 | 238642 |
| 12 | 2017.5.14(00 : 00)–5.16(00 : 00) | Southeast of south China | 106.5 | 277531.8 |
| 13 | 2017.6.4(00 : 00)–6.6(00 : 00) | Northeast of south China | 106.7 | 293025.2 |
| 14 | 2017.7.2(00 : 00)–7.4(00 : 00) | Southern south China | 246.2 | 294511.5 |
| 15 | 2018.4.15(00 : 00)–4.17(00 : 00) | Southern and eastern south China | — | 45559.25 |
| 16 | 2018.5.7(00 : 00)–5.8(00 : 00) | Southern and eastern south China | 115.1 | 106109.9 |
| 17 | 2018.5.9(00 : 00)–5.10(00 : 00) | Central south China | 94.4 | 92764.41 |
| 18 | 2019.4.11(00 : 00)–4.12(12 : 00) | Southeast of south China | 87.8 | 82066.8 |
| 19 | 2019.4.23(12 : 00)–4.25(00 : 00) | Central and eastern south China | 43.9 | 86832.09 |

The forecasts using the WRF model initialized by NCEP_FNL data are denoted as WRF_FNL, while TIGGE_EC is used as a control group to be compared with WRF_FNL and OBS. Besides, we used accurate initial conditions (namely, the initial conditions of the TIGGE_EC forecasts) as the input field to generate the WRF forecasts (hereinafter referred to as WRF_EC) under the same physical parameterization schemes as WRF_FNL.

*2.3. Methods.* In order to comprehensively investigate forecast accuracy, namely, the similarity in the location and intensity between the forecast and observed cumulative precipitation in South China, the correlation coefficient was calculated as follows [35]:

$$r(F,O) = \frac{\operatorname{cov}(F,O)}{\sqrt{\operatorname{Var}[F] + \operatorname{Var}[O]}}, \quad (1)$$

where $F$ is the forecast cumulative precipitation (TIGGE_EC, WRF_FNL or WRF_EC) in South China, $O$ is the observed cumulative precipitation in South China, $r(F,O)$ is the correlation coefficient between $F$ and $O$, cov($F$, $O$) is the covariance of $F$ and $O$, Var[$F$] is the variance of $F$, and Var [$O$] is the variance of $O$.

Besides, the grid-to-grid threat score ($T_S$) was calculated to investigate the forecasting of rainstorm events for different levels of precipitation, namely, light rain and heavy rain. $T_S$ was calculated as follows [36]:

$$T_S = \frac{N_A}{N_A + N_B + N_C}, \quad (2)$$

where $N_A$ represents the number of grid points where the forecast and observed precipitation are at the same level as light rain or (or heavy rain) and above, $N_B$ is the number of grid points where the forecast precipitation level is light rain (or heavy rain) and above but observed precipitation is not at this level, and $N_C$ is the number of grid points where light rain (or heavy rain) has not been forecast.

Since we found that improving the accuracy of the initial conditions could significantly improve the WRF model's forecasts, it is interesting to further explore the sources of initial errors. In this aspect, we compared the differences between the two kinds of initial conditions (FNL and EC) and calculated the vertical integration of moist energy of the differences. The idea is that the area of large moist energy will indicate where the large differences exist, and thus it is taken as the source area of the initial error [34, 37, 38].

The formula of moist energy is as follows [34]:

$$J = \int_0^1 \left[ u_0'^2 + v_0'^2 + \frac{c_p}{T_r} T_0'^2 + R_a T_r \left( \frac{p_{S0}'}{pr} \right)^2 + \frac{L^2}{c_{pT_r}} q_0'^2 \right] \mathrm{d}\sigma, \quad (3)$$

where $\sigma$ is the vertical coordinate; $C_p$ is the specific heat at constant pressure (= 1005.7 J kg$^{-1}$·K$^{-1}$); $R_a$ is the dry air gas constant (=287.04 J·kg$^{-1}$·K$^{-1}$); $p_r$ = 1000 hPa; $T_r$ = 270 K; and $u_0'$, $v_0'$, $T_0'$, $p_{s0}'$, and $q_0'$ are the differences in meridional wind, zonal wind, temperature, surface pressure, and water vapor mixing ratio between the EC and FNL initial fields, respectively. Compared with a single variable, moist energy indicates the comprehensive influence of all variables. Lu et al. [34] found that moist energy could help screen out the source area of the initial error, and the improvement of the initial conditions in that source area could achieve the greatest benefits compared to the improvement of initial conditions in other areas. Thus, hereafter, we refer to this method as the "moist energy method," and the source area as the "sensitive area."

## 3. Results

In 73.7% (14 cases) of the 19 cases, the correlation coefficient between TIGGE_EC and OBS is higher than that between WRF_FNL and OBS, and the average correlation between TIGGE_EC and OBS of the 19 cases is 0.428, which is significantly higher than that between WRF_FNL and OBS (0.328). Besides, the shaded area of the correlation coefficient in Figure 1 shows the extent to which TIGGE_EC is better than WRF_FNL.

Comparison of the cumulative precipitation distribution of WRF_FNL, TIGGE_EC, and OBS in cases 4 and 9 shows that WRF_FNL and TIGGE_EC differ in precipitation intensity and location compared with OBS (Figure 2). However, both the intensity and location of the precipitation forecast by TIGGE_EC are more similar than those of WRF_FNL to OBS. TIGGE_EC has a better effect on the forecasting of the precipitation center in southwest Guangdong and southern Guangxi in both case 4 and case 9. Combined with Figure 1, it can be seen that TIGGE_EC is closer to OBS than WRF_FNL and has a better forecast result, which is consistent with the findings of Huang and Luo [13]. Therefore, in these 14 cases, we assume that both the initial field and the model (i.e., the ECMWF model) have high accuracy, and we only examine these 14 cases in the following parts.

Next, we used accurate initial conditions (namely, those of TIGGE_EC forecasts) to generate WRF forecasts (hereafter referred to as WRF_EC) for the 14 cases in which the forecast of TIGGE_EC was better than WRF_FNL and quantitatively compared them with WRF_FNL. First, the TIGGE_EC forecasts were taken as the true values to examine the improvements in forecasts when using accurate initial conditions. It was found that the average correlation between the cumulative precipitation in South China forecast by WRF_EC and TIGGE_EC is 0.430, which is higher than that between the forecasts of WRF_FNL and TIGGE_EC (0.390).

Meanwhile, 71.4% of the 14 cases have better forecast skill with WRF_EC than with WRF_FNL. Taking case 9 as an example, it can be seen that both the location and the intensity of the precipitation forecasted by WRF_EC are closer to those of TIGGE_EC than WRF_FNL (Figure 2). Arguably, however, although the TIGGE_EC forecasts appear to be more accurate than the WRF_FNL forecasts, they still contain errors. Thus, we further used OBS as the true values to evaluate the improvements of WRF_EC compared to WRF_FNL, and the results turned out to be similar. That is, the cumulative precipitation in South China of OBS generally has higher correlation coefficients with WRF_EC than with WRF_FNL (Table 2). In other words, WRF_EC generates precipitation patterns that are more similar to OBS compared to WRF_FNL.

Besides, we also checked the $T_s$ scores of WRF_EC and WRF_FNL; here, the OBS are used as true values. From Table 3, we can see that, of the 14 cases, there are, respectively, 8 cases and 10 cases for which WRF_EC has better $T_s$ than WRF_FNL for "light rain and above" and "heavy rain and above" (Table 3). This indicates that WRF_EC generates precipitation amounts that are more similar to OBS than WRF_FNL.

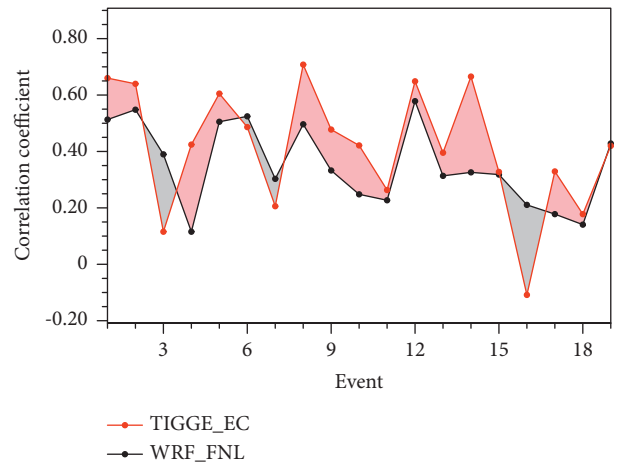


Figure 1: Correlation with OBS of the cumulative precipitation forecast by WRF_FNL (black) and TIGGE_EC (red).

From the above results, it is seen that forecasts of WRF_EC are better than those of WRF_FNL as regards both the precipitation patterns and precipitation amounts. In short, the WRF_EC forecasts are better than the WRF_FNL forecasts.

By quantitatively investigating the improvement in the forecast, we can examine the importance of the accuracy of the initial conditions in the forecasts and thus judge the source of the forecast errors for summer rainstorms in South China. Therefore, we define the following parameters:

$$\mathrm{IM}_1 = \frac{O_{\mathrm{EC}} - O_{\mathrm{FNL}}}{O_{\mathrm{FNL}}},$$
$$\mathrm{IM}_2 = \frac{T_{\mathrm{EC}} - T_{\mathrm{FNL}}}{T_{\mathrm{FNL}}}. \tag{4}$$

Here, $O_{\mathrm{FNL}}$ is the correlation between the cumulative precipitation in South China of WRF_FNL and OBS, while $O_{\mathrm{EC}}$ is the same but between WRF_EC and OBS. Similarly, $T_{\mathrm{FNL}}$ represents the correlation between WRF_FNL and TIGGE_EC, while $T_{\mathrm{EC}}$ represents the correlation between WRF_EC and TIGGE_EC. Then, we define the degree of improvement with $\mathrm{IM}_1$ and $\mathrm{IM}_2$ for each case. If $\mathrm{IM}_1$ ($\mathrm{IM}_2$) $\geq 50\%$, improving the initial value can significantly improve the forecast, and therefore the initial error is the main source of forecast errors. We define such a degree of improvement as "significantly improved." In this category, there are six cases, accounting for 31.6% of all cases. Likewise, if $0 < \mathrm{IM}_1$ ($\mathrm{IM}_2$) $< 50\%$, then improving the initial value can slightly improve the forecast (thus defined as "slightly improved"), for which there are four cases, accounting for 21.1% of all cases. Lastly, if $\mathrm{IM}_1$ ($\mathrm{IM}_2$) $< 0$, improving the initial value can barely improve the forecast results, and so the degree of improvement is defined as "not improved," for which there are also four cases. The $\mathrm{IM}_1$ and $\mathrm{IM}_2$ values for the above 14 cases are shown in Table 4.

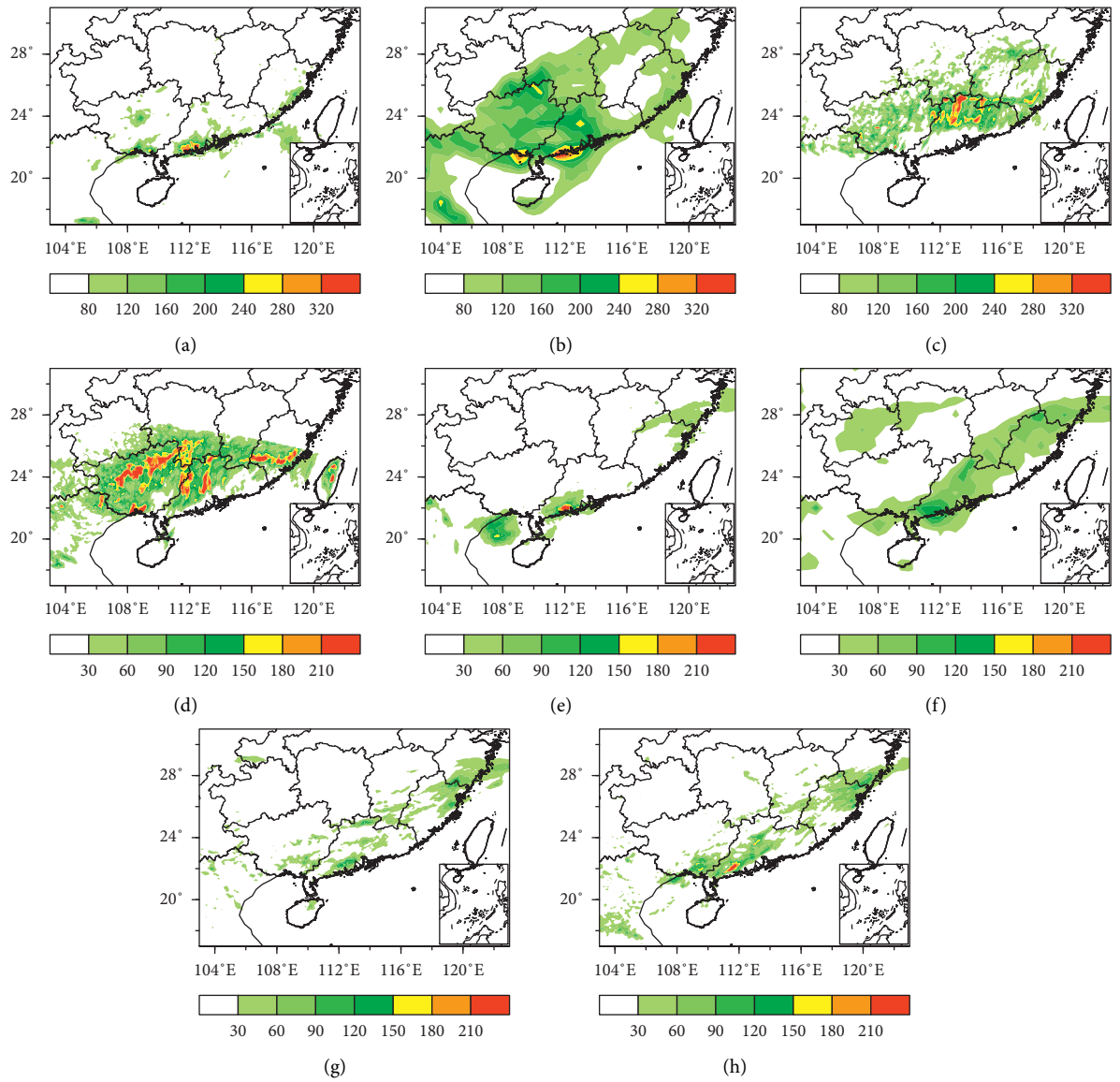


Figure 2: Cumulative precipitation (unit: mm) distribution of ((a) and (e)) OBS, ((b) and (f)) WRF_FNL, ((c) and (g)) TIGGE_EC, and ((d) and (h)) WRF_EC in ((a)–(d)) case 4 and ((e)–(h)) case 9.

According to Table 4, the 14 cases can be divided into three types: "significantly improved," "slightly improved," and "not improved." It is seen that, among the 14 cases, the forecasts of 6 cases are significantly improved after improving the initial field, For those slightly improved cases, the initial errors also play some parts in the forecast errors, so it is also meaningful to explore the source of the initial errors for those cases. Thus, next step, we explore the source of the initial errors for the "significantly improved" group and the "slightly improved" group (a total of 10 cases).

The differences between two kinds of initial conditions are explored, and the area of largest differences is taken as the source area (also called sensitive areas hereafter) of the initial errors. By analyzing the basic flows at 700 hPa, we found that the sensitive areas are located in the basic flows that are directed towards the precipitation area. This means that the initial errors mainly come from upstream the precipitation area. To quantitatively find out the relationship between the precipitation area and the sensitive area, we selected both their sizes as $4^\circ \times 4^\circ$ (Figure 3) and found that their distances are about $5^\circ$ of longitude. Based on the above results, we developed a method to identify the sensitive area by combining the precipitation area and the basic state wind (hereafter referred to as the "guide flow method"). The precipitation area is centered on the maximum precipitation point. First, we define the direction of the guide flow at 700 hPa as follows:

$$\tan(\theta) = \frac{v}{u} = \frac{\text{lat}}{\text{lon}}, \tag{5}$$

where $u$ and $v$ are the zonal and meridional wind components on 700 hPa, respectively; $\theta$ is the angle between the wind direction at 700 hPa and the horizontal direction. lat is

Table 2: Correlation coefficients of WRF_FNL and WRF_EC with both OBS and TIGGE_EC for cumulative precipitation in South China.

| Case | Correlation coefficient (with OBS) | | Correlation coefficient (with TIGGE_EC) | |
|---|---|---|---|---|
| | WRF_FNL | WRF_EC | WRF_FNL | WRF_EC |
| 1 | 0.512974 | 0.607617 | 0.486501 | 0.643501 |
| 2 | 0.548051 | 0.174862 | 0.5627 | 0.177614 |
| 4 | 0.115652 | 0.42435 | 0.28005 | 0.369751 |
| 5 | 0.505295 | 0.537231 | 0.560114 | 0.597676 |
| 8 | 0.496311 | 0.46829 | 0.647559 | 0.540012 |
| 9 | 0.332558 | 0.363309 | 0.437344 | 0.487258 |
| 10 | 0.062397 | 0.248002 | 0.279877 | 0.290106 |
| 11 | 0.226967 | 0.305981 | 0.188803 | 0.340021 |
| 12 | 0.578311 | 0.532898 | 0.444847 | 0.349342 |
| 13 | 0.313502 | 0.326267 | 0.250141 | 0.271314 |
| 14 | 0.325913 | 0.393516 | 0.325419 | 0.499911 |
| 15 | 0.318089 | 0.447602 | 0.204124 | 0.408888 |
| 17 | 0.177811 | 0.329199 | 0.321906 | 0.501076 |
| 18 | 0.140593 | 0.070364 | 0.140593 | 0.393925 |

Table 3: $T_S$ of WRF_FNL and WRF_EC with respect to OBS in the 14 cases ($T_S$ of WRF_EC marked in red has better forecast).

| Case | $T_S$ (light rain and above) | | $T_S$ (heavy rain and above) | |
|---|---|---|---|---|
| | WRF_FNL | WRF_EC | WRF_FNL | WRF_EC |
| 1 | 0.212 | 0.405 | 0.189 | 0.311 |
| 2 | 0.417 | 0.170 | 0.356 | 0.123 |
| 4 | 0.166 | 0.200 | 0.088 | 0.094 |
| 5 | 0.292 | 0.318 | 0.088 | 0.266 |
| 8 | 0.296 | 0.286 | 0.114 | 0.134 |
| 9 | 0.241 | 0.350 | 0.088 | 0.014 |
| 10 | 0.29 | 0.286 | 0.28 | 0.223 |
| 11 | 0.193 | 0.200 | 0 | 0 |
| 12 | 0.202 | 0.210 | 0.045 | 0.054 |
| 13 | 0.346 | 0.310 | 0.061 | 0.091 |
| 14 | 0.230 | 0.242 | 0.161 | 0.202 |
| 15 | 0.411 | 0.377 | 0.063 | 0.071 |
| 17 | 0.210 | 0.271 | 0.008 | 0.015 |
| 18 | 0.065 | 0.032 | 0.010 | 0.012 |

Table 4: Classification of $IM_1$ and $IM_2$ and the degree of improvement corresponding to each case.

| Case | $IM_1$ | $IM_2$ | Degree of improvement |
|---|---|---|---|
| 1 | 0.184498 | 0.322713 | Slightly improved |
| 2 | −0.68094 | −0.68435 | Not improved |
| 4 | 0.906682 | 0.320304 | Significantly improved |
| 5 | 0.063203 | 0.067061 | Slightly improved |
| 8 | −0.05646 | −0.16608 | Not improved |
| 9 | 0.092468 | 0.11413 | Slightly improved |
| 10 | 2.974582 | 0.036548 | Significantly improved |
| 11 | 0.348127 | 0.80093 | Significantly improved |
| 12 | −0.07853 | −0.21469 | Not improved |
| 13 | 0.04072 | 0.084644 | Slightly improved |
| 14 | 0.207425 | 0.536207 | Significantly improved |
| 15 | 0.407161 | 1.003135 | Significantly improved |
| 17 | 0.851398 | 0.556591 | Significantly improved |
| 18 | −0.998083 | 0.643097 | Not improved |

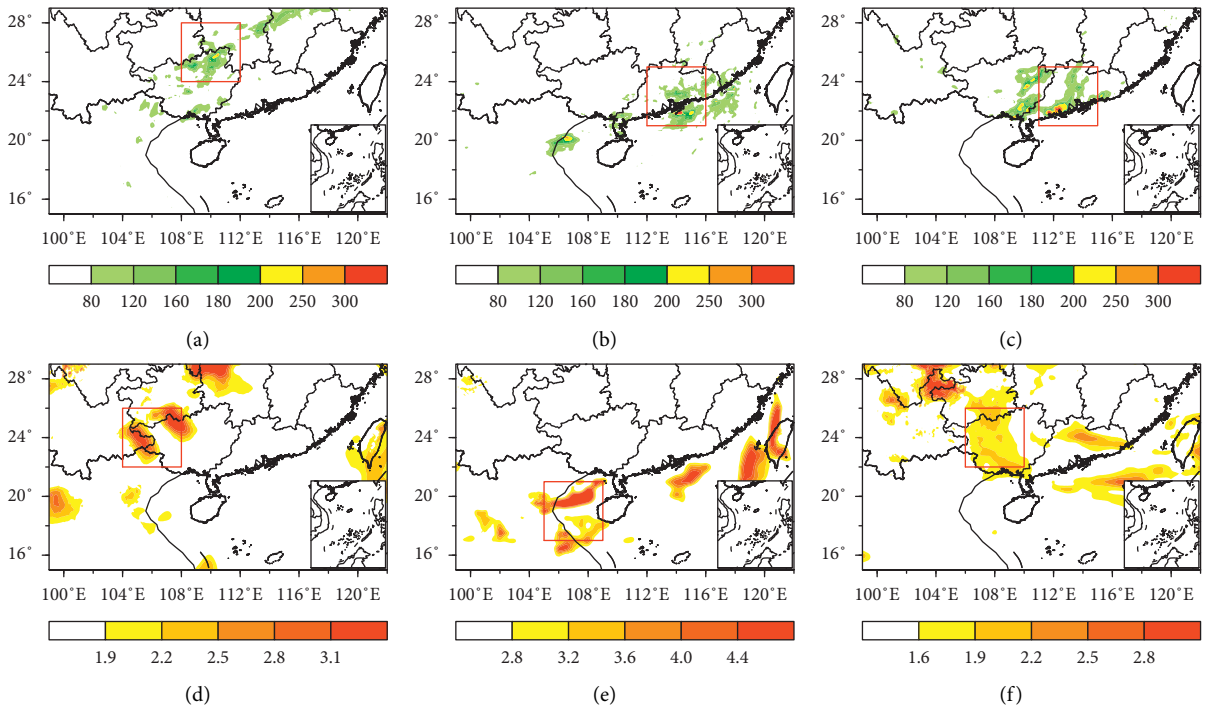


Figure 3: ((a)–(c)) Distribution of cumulative precipitation (unit: mm) in cases 1, 5, and 14, respectively (the red box is the range of the precipitation area selected). ((d)–(f)) The sensitive areas selected with the moist energy (unit: $1 \times 10^5$ J) method for cases 1, 5, and 14, respectively (the red box is the range of the sensitive area selected).

the latitudinal difference between the maximum precipitation point and the selected maximum point of the sensitive area; lon = 5°, which is the longitudinal difference between the maximum precipitation point and the selected maximum point of the sensitive area. From the description above, $u$, $v$, and lon are all known, and then lat can be derived. Thus, the sensitive area can be determined according to the precipitation area, $\tan(\theta)$, and the lat and lon values.

If the guide flow in the precipitation area is relatively straight, then $u$ and $v$ are the zonal and meridional wind components of the maximum precipitation point. With lon known and $\tan(\theta)$ obtained by linear backward deduction of the wind ($u$ and $v$) in the precipitation area, lat can be derived, so the sensitive area can be determined (Figure 4(a)). If there are obvious troughs or cyclonic circulations in the precipitation area, $u$ and $v$ are the average zonal and meridional wind components of the precipitation area. Similarly, the location of the sensitive area can be deduced according to the average guide flow direction in the precipitation area (Figure 4(b)). Whether the guide flow is relatively straight or there are obvious troughs or cyclonic circulations in the precipitation area, the sensitive area selected by the guide flow method is relatively consistent with the location of the large-value area of moist energy (Figure 4).

Table 5 shows the locations of the sensitive areas determined by the guide flow method and the moist energy method. For 80% (8 cases) of the 10 cases, the positions of the sensitive areas selected by the two methods are relatively consistent, with the distances between them being less than 2°. Thus, the sensitive area selected by the guide flow method is similar to that identified by the moist energy method. To confirm this point, we also use both the guide flow method and the moist energy method to identify the sensitive areas of those cases that belong to "not improved" group (cases 2, 8, 12, and 18) in Table 4 and those cases where TIGGE_EC has lower skills than WRF_FNL (cases 3, 6, 7, 16, and 19) in Figure 1. Results showed that, for 8 out of 9 cases, the positions of the sensitive areas selected by the two methods are similar (Table 6). This confirms that the guide flow method can be used to identify sensitive areas associated with rainstorm forecasts. Since the guide flow method is easy to use, it thus may be helpful for quickly selecting perturbation areas for ensemble forecasts or carrying out supplemental observations for adaptive observations.

To verify the accuracy of the sensitive areas, namely, to demonstrate that they are the source areas of the initial errors which lead to large forecast errors, we chose three other areas for comparison and carried out sensitivity experiments. These three areas had the same size as sensitive areas, and they were, respectively, located to the west, south, and north of the precipitation areas with distances of 10° longitude, 5° latitude, and 5° latitude. Then, the initial conditions of WRF_FNL were replaced with those of WRF_EC in the sensitive area to form new ones. Similarly, three other sets of new initial conditions were generated by replacing those of WRF_FNL with the initial conditions of WRF_EC in those three areas.

Keeping the model configuration unchanged, four new forecasts were produced with the above four sets of new initial conditions, and these are, respectively, referred to here

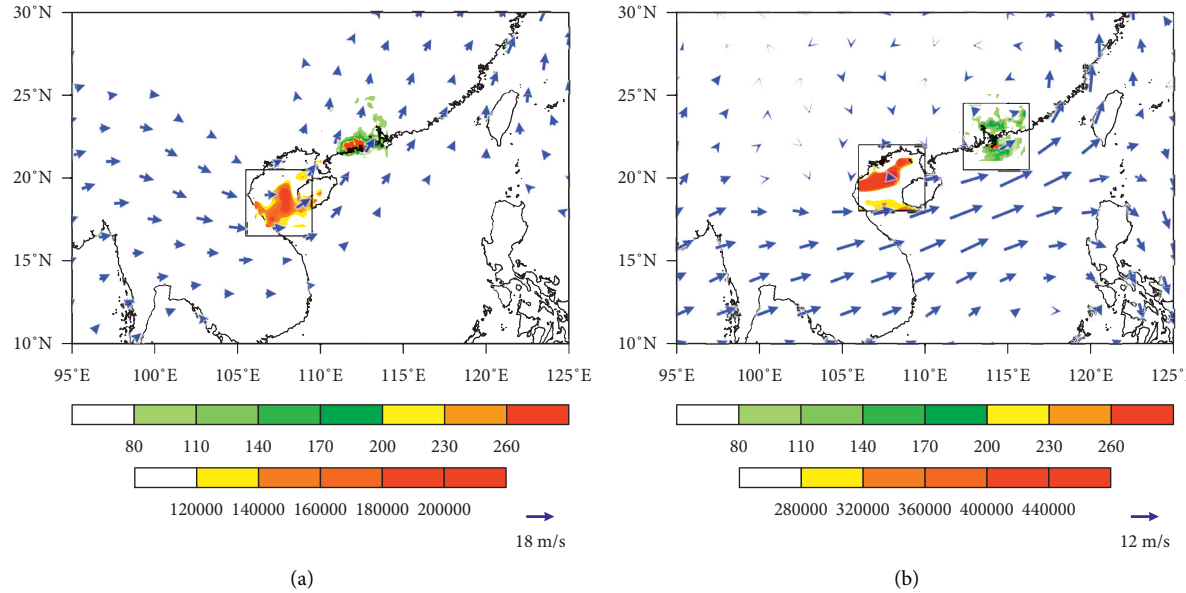


Figure 4: Sensitive area selected with the guide flow method (rectangles in panel (a) and left rectangles in panel (b)), sensitive areas identified by moist energy (unit: $1 \times 10^5$ J the left parts of shadings in both two panels), and the precipitation area (unit: mm, the right parts of shadings in both two panels) overlayed with the 700 hPa wind field in (a) case 4 and (b) case 5.

Table 5: Comparison of sensitive areas selected with moist energy method and guide flow method.

| Case | Sensitive areas selected with the moist energy method | Sensitive areas selected with guide flow method |
|---|---|---|
| 1 | (22°–26°N, 104°–108°E) | (21°–25°N, 103°–107°E) |
| 4 | (16°–20°N, 107°–111°E) | (16°–20°N, 106°–110°E) |
| 5 | (18°–22°N, 105°–109°E) | (18°–22°N, 106°–110°E) |
| 9 | (20°–24°N, 104°–108°E) | (18°–22°N, 105°–109°E) |
| 10 | (18°–22°N, 106°–110°E) | (18°–22°N, 107°–111°E) |
| 11 | (12°–16°N, 102°–106°E) | (21°–25°N, 102°–106°E) |
| 13 | (17°–21°N, 104°–108°E) | (18°–22°N, 102°–106°E) |
| 14 | (20°–24°N, 105°–109°E) | (16°–20°N, 105°–109°E) |
| 15 | (20°–24°N, 107°–111°E) | (21°–25°N, 106°–110°E) |
| 17 | (20°–24°N, 106°–110°E) | (19°–23°N, 105°–109°E) |

Table 6: Same as Table 5 but for verifying cases.

| Case | Sensitive areas selected with the moist energy method | Sensitive areas selected with guide flow method |
|---|---|---|
| 2 | (17°–21°N, 103°–107°E) | (20°–24°N, 103°–107°E) |
| 3 | (15°–19°N, 100°–104°E) | (13°–17°N, 102°–106°E) |
| 6 | (20°–24°N, 106°–110°E) | (21°–25°N, 107°–111°E) |
| 7 | (14°–18°N, 99°–103°E) | (19°–23°N, 102°–106°E) |
| 8 | (20°–24°N, 105°–109°E) | (18°–22°N, 106°–110°E) |
| 12 | (21°–25°N, 107°–111°E) | (19°–23°N, 106°–110°E) |
| 16 | (21°–25°N, 106°–110°E) | (21°–25°N, 109°–113°E) |
| 18 | (22°–26°N, 105°–109°E) | (22°–26°N, 106°–110°E) |
| 19 | (24°–28°N, 105°–109°E) | (23°–27°N, 104°–108°E) |

as F-sens, F-north, F-west, and F-south. The cumulative precipitation of these forecasts was compared with the observed precipitation in South China, and their correlation coefficients were calculated. It can be seen from the results (Figure 5) that, for the 10 improved cases in Table 4, there are 9 cases where the correlations between F-sens and OBS are

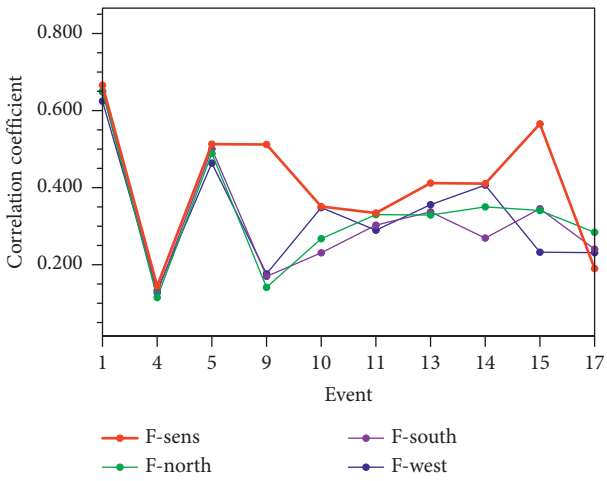


Figure 5: Sensitivity experiment results of 10 improved cases in Table 4, in which the red, green, purple, and blue lines plot the correlation coefficients between the cumulative precipitation results after replacing the sensitive area, northern control area, southern control area, western control area, and OBS in South China, respectively.

higher than other correlations (F-north and OBS, F-west and OBS, and F-south and OBS). This means that improving the initial conditions in sensitive areas can improve the forecast to a greater extent than by improving them in other areas, and again this verifies that the sensitive areas are the source areas of initial errors as well as the forecast errors.

## 4. Conclusion

Based on the WRF model, this paper has investigated the possible sources of forecast errors with respect to rainstorms in the South China monsoon region. First it is demonstrated that the initial error is the important source of the forecast errors, and then the source area (sensitive area) of initial errors is explored. Next, the relationship between the sensitive area and the precipitation area is analyzed. Based on the above results, a new method is developed (which we call the "guide flow method") to identify the sensitive area. Finally, the sensitive areas were examined through sensitivity experiments, which confirmed the accuracy of the sources of initial errors as well as the forecast errors.

By investigating the improvement rate of the forecast when using one set of data rather than the other, the important degree of the initial conditions with respect to the forecasts for each case has been obtained. The results showed that forecasts of 6 cases have been significantly improved (the improvement rate is larger than 0.5), which means that, for these cases, the initial condition plays a main role in the forecasts. While the forecasts of 4 cases have been slightly improved, this means the initial condition also has some effects on the forecasts. For these 10 cases, the initial errors are the important sources of forecast errors; then we further explored the source of the initial errors by comparing the two initial conditions. The results showed that the initial errors mainly came from an area located upstream of the rainfall area (about 5° of longitude away from the maximum precipitation area) and we called that area the sensitive area. By studying the relationship between the sensitive area and the precipitation area, we found out the rules behind it and then put forward a "guide flow" method to identify the sensitive areas. The sensitive areas identified by the guide flow method were found to be generally consistent with those identified by the moist energy method. Since the guide flow method is easy to use, it thus may be helpful for quickly selecting perturbation areas for ensemble forecasts or carrying out supplemental observations for adaptive observations. Finally, sensitivity experiments demonstrated that improving the initial conditions in the sensitive areas leads to more benefits than improving them in other same-sized areas. This verifies the accuracy of the sensitive areas and confirms the source of the forecast errors.

Statistically, improving the initial conditions may improve the forecasting of rainstorms in the South China monsoon region. However, there are some cases (such as the 2019.4.12 case) in which improving the initial conditions has no benefit on the forecasts. Thus, for these cases, the model error has an important impact on the rainstorm forecast. Besides, there are some cases where we fail to find a better initial condition (cases 3, 6, 7, 16, and 19); for these cases, it is necessary to find out other ways to evaluate the importance of the initial conditions. In a word, in order to further improve the prediction skill for rainstorms in the South China monsoon region, further in-depth research is still needed.

## Data Availability

The data used in this paper can be obtained from Lin Lin (20211101017@nuist.edu.cn) upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The authors express their appreciation to the staff at the China Meteorological Data Service Centre and those involved in the Final Global Data Assimilation System and the THORPEX Interactive Grand Global Ensemble from the ECMWF. This research was funded by the National Key Research and Development Program of China (Grants 2018YFC1507405 and 2017YFC1501601).

## References

[1] T. J. Zhou and R. C. Yu, "Atmospheric water vapor transport associated with typical anomalous summer rainfall patterns in China," *Journal of Geophysical Research: Atmospheres*, vol. 110, 2005.

[2] R. X. Liu, J. H. Sun, and B. F. Chen, "Selection and classification of warm-sector heavy rainfall events over South

China," *Chinese Journal of Atmospheric Sciences*, vol. 43, no. 1, pp. 122–133, 2019, in Chinese.

[3] M. Zhang and Z. Meng, "Impact of synoptic-scale factors on rainfall forecast in different stages of a persistent heavy rainfall event in south China," *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 7, pp. 3574–3593, 2018.

[4] Q. Q. Li, G. H. Wang, S. Yang, and F. Wang, "Sub-seasonal prediction of rainfall over the South China Sea and its surrounding areas during spring–summer transitional season," *International Journal of Climatology*, vol. 40, no. 10, pp. 4326–4346, 2019.

[5] S. X. Zhong, S. Yang, C. Y. Guo, and Z. T. Chen, "Capabilities and limitations of GRAPES simulations of extreme precipitation in the warm sector over a complex orography," *Journal of Tropical Meteorology*, vol. 25, no. 2, pp. 42–53, 2019.

[6] Z. Li, Y. Luo, Y. Du, and J. C. L. Chan, "Statistical characteristics of pre-summer rainfall over south China and associated synoptic conditions," *Journal of the Meteorological Society of Japan. Series II*, vol. 98, no. 1, pp. 213–233, 2020.

[7] J. Du and J. Li, "Application of ensemble forecast method in the study and forecast of heavy rain," *Advances in Meteorological Science and Technology*, vol. 4, no. 5, pp. 6–20, 2014.

[8] T. Chen, J. Sun, Y. Chen, Y. C. Guo, and J. Xu, "Study on the numerical predictivity of localized severe mesoscale rainstorm in Guangzhou on 7 may 2017," *Meteorological Monthly*, vol. 45, no. 9, pp. 1199–1212, 2019, in Chinese.

[9] Y. Luo, R. Zhang, Q. Wan, B. Wang, and Y. Xiao, "The southern China monsoon rainfall experiment (SCMREX)," *Bulletin of the American Meteorological Society*, vol. 98, no. 5, 2016.

[10] Y. Luo, J. Sun, Y. Li et al., "Science and prediction of heavy rainfall over China: research progress since the reform and opening-up of new China," *Journal of Meteorological Research*, vol. 34, no. 3, pp. 427–459, 2020.

[11] Y. Chen and Y. Luo, "Analysis of paths and sources of moisture for the south China rainfall during the presummer rainy season of 1979-2014," *Journal of Meteorological Research*, vol. 32, no. 5, pp. 744–757, 2018.

[12] Y. Luo, H. Wang, R. Zhang, W. Qian, and Z. Luo, "Comparison of rainfall characteristics and convective properties of monsoon precipitation systems over south China and the yangtze and huai river basin," *Journal of Climate*, vol. 26, no. 1, pp. 110–132, 2013.

[13] L. Huang and Y. Luo, "Evaluation of quantitative precipitation forecasts by TIGGE ensembles for south China during the presummer rainy season," *Journal of Geophysical Research: Atmospheres*, vol. 122, no. 16, pp. 8494–8516, 2017.

[14] J. M. Zhang, W. Q. Mo, and J. Y. Yan, "Analysis of characteristics of two rounds of heavy rainfall in Dongguan under the background of southwest Monsoon onset," *Guangdong Meteorology*, vol. 36, no. 1, pp. 40–44+48, 2014, in Chinese.

[15] J. B. Lin, Y. T. Li, Y. N. Zheng, and H. Chen, "Analysis of a continuous rainstorm process of South China Monsoon low pressure," *Guangdong Meteorology*, vol. 41, no. 6, pp. 29–33, 2019, in Chinese.

[16] X. Liu, Y. Luo, Z. Guan, and D.-L. Zhang, "An extreme rainfall event in coastal south China during SCMREX-2014: formation and roles of rainband and echo trainings," *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 17, pp. 9256–9278, 2018.

[17] J. J. Cai, Z. F. Wu, X. Q. Chen, Y. Lan, Z. Y. Guo, and C. Y. Guo, "An analysis of the causes of the "18·8" Guangdong monsoon low and sustained heavy rain," *Terminal Rain and Disasters*, vol. 38, no. 6, pp. 576–586, 2019, in Chinese.

[18] Z. Y. Guo, Z. F. Wu, J. J. Cai, H. L. Zhang, and X. Y. Chen, "Analysis of water vapor transport characteristics of a monsoon low-pressure continuous heavy rain event at the end of August 2018 in Guangdong area," *Torrential Rain and Disasters*, vol. 38, no. 6, pp. 587–596, 2019, in Chinese.

[19] Y. Shen, Y. Du, and G. X. Chen, "Ensemble sensitivity analysis of heavy rainfall associated with three MCSs coexisting over southern China," *Journal of Geophysical Research: Atmospheres*, vol. 125, no. 2, Article ID e2019JD031266, 2020.

[20] M. Mu, D. Wansuo, and C. Jifan, "Recent advances in predictability studies in China (1999-2002)," *Advances in Atmospheric Sciences*, vol. 21, no. 3, pp. 437–443, 2004.

[21] M. Mu, W. Duan, Q. Wang, and R. Zhang, "An extension of conditional nonlinear optimal perturbation approach and its applications," *Nonlinear Processes in Geophysics*, vol. 17, no. 2, pp. 211–220, 2010.

[22] Q. Wang, M. Mu, and H. A. Dijkstra, "Application of the conditional nonlinear optimal perturbation method to the predictability study of the kuroshio large meander," *Advances in Atmospheric Sciences*, vol. 29, no. 1, pp. 118–134, 2012.

[23] M. Mu and W. Duan, "Conditional nonlinear optimal perturbation and its applications to the studies of weather and climate predictability," *Chinese Science Bulletin*, vol. 50, no. 21, pp. 2401–2407, 2005.

[24] N. N Li, "Hourly refined evaluation of precipitation simulation in China," *Chinese Academy of Meteorological Sciences*, in Chinese, 2020.

[25] X. Li, L. F. Zhang, and J. N. Wang, "The sensitivity of the prediction of Meiyu torrential rainfall to model resolution and cumulus parameterization," *Torrential Rain and Disasters*, vol. 39, no. 6, in Chinese, Article ID 10, 2020.

[26] A. Fernández, V. Schumacher, I. Ciocca, A. Rifo, and F. Justino, "Validation of a 9-km WRF dynamical downscaling of temperature and precipitation for the period 1980–2005 over Central South Chile," *Theoretical and Applied Climatology*, vol. 143, no. 1-2, pp. 1–18, 2021.

[27] Z. H. Li and Y. L. Luo, "Statistical characteristics of presummer rainfall over South China before and after South China Sea monsoon onset from 1980 to 2017," *Torrential Rain and Disasters*, vol. 40, no. 2, Article ID 10, 2021.

[28] R. Zhang, Y. Ni, L. Liu, Y. Luo, and Y. Wang, "South China heavy rainfall experiments (SCHeREX)," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 89A, no. 2, pp. 153–166, 2011.

[29] S. W. Li, S. Yang, and H. W. Liu, "Sensitivity of warm-sector heavy precipitation to the impact of anthropogenic heating in South China," *Atmospheric and Oceanic Science Letters*, vol. 11, no. 3, pp. 236–245, 2018.

[30] Y. Luo, L. Li, R. H. Johnson et al., "Science and prediction of monsoon heavy rainfall," *Science Bulletin*, vol. 64, no. 21, pp. 1557–1561, 2019.

[31] X. Zhang, Y. Luo, Q. Wan, W. Ding, and J. Sun, "Impact of assimilating wind profiling radar observations on convection-permitting quantitative precipitation forecasts during SCMREX," *Weather and Forecasting*, vol. 31, no. 4, pp. 1271–1292, 2016.

[32] Y. Luo and L. F. Zhang, "A case study of the error growth evolution in a meiyu front heavy precipitation forecast and an analysis of the predictability," *Acta Meteorologica Sinica*, vol. 68, no. 3, pp. 411–420, 2010.

[33] L. Wang, X. Shen, J. Liu, and B. Wang, "Model uncertainty representation for a convection-allowing ensemble prediction system based on CNOP-P," *Advances in Atmospheric Sciences*, vol. 37, no. 8, Article ID 15, 2020.

[34] C. H. Lu, L. Lin, and F. F. Zhou, "Analysis of the source of forecast errors for a heavy precipitation in the southwest of Guangdong province," *Chinese Journal of Atmospheric Sciences*, vol. 44, no. 6, in Chinese, Article ID 12, 2020.

[35] F. Y. Wei, *Modern Climate Statistics Diagnosis and Prediction Technology*, Second edition, Weather Press, Beijing, China, 2007, in Chinese.

[36] M. Y. Ding, L. L. Wang, Y. Xin et al., "TS score analysis of WRF cloud microphysical parameterization scheme on Xinjiang's rainstorm simulation ability," *Arid Zone Research*, vol. 36, no. 6, pp. 1411–1418, 2019.

[37] F. Zhou, M. Yamaguchi, and X. Qin, "Possible sources of forecast errors generated by the global/regional assimilation and prediction system for landfalling tropical cyclones. Part I: initial uncertainties," *Advances in Atmospheric Sciences*, vol. 33, no. 7, pp. 841–851, 2016.

[38] F. F. Zhou, W. Duan, H. Zhang, and M. Yamaguchi, "Possible sources of forecast errors generated by the global/regional assimilation and prediction system for landfalling tropical cyclones. Part II: model uncertainty," *Advances in Atmospheric Sciences*, vol. 33, no. 7, pp. 841–851, 2018.