Precipitation Biases in the ECMWF Integrated Forecasting System

DAVID A. LAVERS,[a,b] SHAUN HARRIGAN,[a] AND CHRISTEL PRUDHOMME,[a,c,d]

[a] *European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom*
[b] *School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom*
[c] *UK Centre for Ecology & Hydrology, Wallingford, United Kingdom*
[d] *Geography Department, Loughborough University, Loughborough, United Kingdom*

(Manuscript received 18 December 2020, in final form 18 February 2021)

ABSTRACT: Precipitation is a key component of the global water cycle and plays a crucial role in flooding, droughts, and water supply. One way to manage its socioeconomic effects is based on precipitation forecasts from numerical weather prediction (NWP) models, and an important step to improve precipitation forecasts is by diagnosing NWP biases. In this study, we investigate the biases in precipitation forecasts from the European Centre for Medium-Range Weather Forecasts Integrated Forecasting System (IFS). Using the IFS control forecast from 12 June 2019 to 11 June 2020 at 5219 stations globally, we show that in each of the boreal winter and summer half years, the IFS 1) has an average global wet bias and 2) displays similar bias patterns for forecasts starting at 0000 and 1200 UTC and across forecast days 1–5. These biases are dependent on observed (climatological) precipitation; stations with low observed precipitation have an IFS wet bias, while stations with high observed precipitation have an IFS dry bias. Southeast Asia has a wet bias of 1.61 mm day^{-1} (in boreal summer) and over the study period the precipitation is overestimated by 31.0% on forecast day 3. This is the hydrological signature of several hypothesized processes including issues specifying the IFS snowpack over the Tibetan Plateau, which may affect the mei-yu front. These biases have implications for IFS land–atmosphere feedbacks, river discharge, and for ocean circulation in the Southeast Asia region. Reducing these biases could lead to more accurate forecasts of the global water cycle.

KEYWORDS: Hydrologic cycle; Mei-yu fronts; Precipitation; In situ atmospheric observations; Forecast verification/skill; Numerical weather prediction/forecasting

1. Introduction

Precipitation is one of the most important climate variables directly affecting humans and is a key part of the global water cycle (WMO 2020). It is fundamental for public water supply, the health of the natural environment, such as river ecosystems, and too much or too little can cause floods or droughts, respectively, both of which lead to socioeconomic damages. The ability to manage it and its effects are therefore of great importance to society. An important tool to do this comes in the form of precipitation forecasts from numerical weather prediction (NWP) models. Precipitation forecasts at a range of spatiotemporal scales can be assessed directly from an NWP model, following the application of a precipitation bias-correction procedure (e.g., Hamill et al. 2017; Hewson and Pillosu 2020), or by driving a hydrological model to calculate forecasts of river discharge in space and time over the model domain (Lavers et al. 2020). These outputs can then be used, for example, to issue warnings of upcoming flooding events.

Hydrological models form a key part of the forecast systems used to provide early warning of hydrological extremes. There are many hydrological models available (Pechlivanidis et al. 2011; Gayathri et al. 2015), and these models, in varying ways, represent the physical characteristics of river catchment topography, soil types, land use, and storages. Using hydrometeorological observations to determine the initial hydrological conditions, NWP output (e.g., precipitation, temperature, snow, and/or surface runoff) then drives a hydrological model to calculate river discharge. In April 2018, one such system, the Global Flood Awareness System (GloFAS; Alfieri et al. 2013; Zsótér et al. 2019a), became operational under the European Commission Copernicus Emergency Management Service. Every day GloFAS produces forecasts of river discharge out to 30 days and its usefulness in supporting humanitarian efforts was highlighted in 2019 during Tropical Cyclones Idai and Kenneth in eastern Africa (Magnusson et al. 2019; Emerton et al. 2020). One major challenge for global hydrological forecasting, however, is the water balance errors introduced from the NWP forcing data. A study by Harrigan et al. (2020a) used a network of 1801 in situ river discharge observation stations to evaluate the new GloFAS-ERA5 river discharge reanalysis. The GloFAS-ERA5 product is a model estimate for observed discharge and is created by driving GloFAS with the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 atmospheric reanalysis (Hersbach et al. 2020), which uses an older version (Cycle 41r2) of the ECMWF Integrated Forecasting System (IFS). Results showed that while 86% of stations had skillful river discharge simulations, the skill was significantly deteriorated due to large biases in the river discharge which were introduced by the atmospheric forcing or the data assimilation procedure. Catchments with the worst skill were characterized mainly with large positive

Denotes content that is immediately available upon publication as open access.

Corresponding author: David Lavers, david.lavers@ecmwf.int

DOI: 10.1175/JHM-D-20-0308.1

river discharge biases located in the drier rivers of the central United States, Africa, eastern Brazil, the western coast of South America, and Southeast Asia. Research by Zsótér et al. (2019b) has shown that the ECMWF land data assimilation system contributes to negative ERA5 runoff biases in the high northern latitudes and positive ERA5 runoff biases in lower latitudes. The possible biases introduced by NWP precipitation, however, have not yet been explored.

In an Earth system model, skillful precipitation forecasts are also important for soil moisture, evaporation, temperature, and land–atmosphere feedbacks, all of which can influence Earth system predictability. This means it is important to identify and investigate any issues with NWP precipitation to allow for a broad range of future NWP improvements to be realized. The aim of this paper is therefore to evaluate precipitation forecasts from the ECMWF IFS to elucidate precipitation biases globally. Furthermore, following the identification of the most error-prone region, namely, Southeast Asia in boreal summer, we present an investigation on the causes of this poor forecast performance. It is thought that it will then be possible to focus research on improving processes in the most erroneous areas.

2. Data and methods

a. Precipitation observations

Observations of 24-h gauged precipitation totals were retrieved from 13 June 2019 to 16 June 2020 from the ECMWF archive; the original source was the World Meteorological Organization Global Telecommunications System. Six reporting times (0000, 0100, 0300, 1200, 1300, and 1500 UTC) were considered and for stations to be used herein, two criteria were applied: 1) their availability at 0000 and 1200 UTC (or 0100 and 1300 UTC, or 0300 and 1500 UTC) to capture daily precipitation variation at a station due to the accumulation window; and 2) a 50% daily availability across the whole study period and within each meteorological season (June–August, September–November, December–February, March–May). This resulted in 5219 stations in total, with 4656 stations at 0000/1200 UTC, 77 stations at 0100/1300 UTC mostly in western Australia, and 486 stations at 0300/1500 UTC mostly in western Russia.

A quality-control procedure was applied to the precipitation observations to remove any negative or 999 values. We note, however, that this study uses the raw reported precipitation values, and as such, the observations may contain both random and systematic errors (Muchan and Dixon 2019). Random errors can result, for example, from blockages in a rain gauge, while an example of a systematic error is the undercatch of precipitation due to the deflection of raindrops (by the wind) away from the gauge.

b. ECMWF forecasts and evaluation

In this study, the total precipitation in the control member of the IFS ensemble (Cycle 46r1)—the control member is generated from the most accurate estimate of the state of the Earth system—is evaluated from 12 June 2019 to 11 June 2020. This time period is used to have a consistent IFS model cycle (and thus not to mix different model versions) and to sample precipitation characteristics throughout the year. From the ECMWF archive, we retrieved the 24-h total precipitation on forecast days 1–5 on the Gaussian octahedral (O640) native model grid from the 0000 UTC and 1200 UTC forecast initializations. The horizontal resolution of the octahedral grid varies from about 15.6 km at the equator to 18.5 km in the midlatitudes, with each grid point approximately having an equal area. The 0000 UTC forecasts were verified against the 0000, 0100, and 0300 UTC observations and the 1200 UTC forecasts against the 1200, 1300, and 1500 UTC observations; and the forecast accumulation periods for the different observation times are provided in Table 1. The nearest neighbor approach was employed to match the closest model grid point to a station observation, and the forecast evaluation then entailed calculating the forecast-minus-observation bias at different forecast lead times at each station for days in the boreal winter half year (October–March) and summer half year (April–September).

To study the large-scale atmospheric circulation at different forecast lead times, the control forecast of the 500-hPa geopotential height and the total water vapor flux (integrated from the 91 model levels) were also retrieved from the ECMWF archive. These fields were used in composite analysis, a method which considers nonlinear associations and is easy to interpret (Kingston et al. 2006).

3. Results

a. Evaluation of precipitation biases at 0000 and 1200 UTC and with lead time

Figures 1 and 2 show at 0000 UTC and 1200 UTC, respectively, the 5219 station locations and their precipitation biases on forecast days 1, 3, and 5 in boreal winter and boreal summer. The majority (3538 or 67.8%) of these stations are in the mid latitudes (35°–65°N and 35°–65°S) while the polar regions (>65°N and >65°S) have the fewest stations (180 or 3.4%), and across all stations there is a mean wet bias ranging from 0.13 (Figs. 1f and 2f) to 0.27 mm day^{-1} (Fig. 1a). In each half year, the biases for forecasts starting at 0000 and 1200 UTC and for the different forecast days have similar patterns and values (cf. Figs. 1 and 2). These results imply that the bias does not have a large dependence on the initial forecast time (i.e., 0000 or 1200 UTC). A further example of similar biases across lead times is in the boreal summer in Southeast Asia, where a spatially consistent model wet bias, as shown by dark blue markers, is found (Figs. 1b,d,f and 2b,d,f). Also, the sparsity of stations in certain areas, such as Africa, is noteworthy and highlights a major obstacle to verification studies, and in turn, improving forecast systems.

We summarize the precipitation biases on forecast days 1–5 using boxplots, and these are shown in Fig. 3. Using all 5219 global stations, the mean wet bias is again seen in both boreal winter (Fig. 3a) and boreal summer (Fig. 3b) for forecasts

TABLE 1. The forecast accumulation periods for total precipitation on forecast days 1–5 for the different observation times.

Observation time (UTC)	Forecast day 1	Forecast day 2	Forecast day 3	Forecast day 4	Forecast day 5
0000/1200	0–24 h	24–48 h	48–72 h	72–96 h	96–120 h
0100/1300	1–25 h	25–49 h	49–73 h	72–96 h	96–120 h
0300/1500	3–27 h	27–51 h	51–75 h	75–99 h	99–123 h

starting at 0000 UTC and 1200 UTC across all forecast days. In winter, the mean wet bias is almost constant and approximately 75% of the stations have a wet bias across the forecast days. Conversely, in summer, there is a suggestion that the percentage of stations with negative biases increases with lead time. This is exemplified at 0000 UTC when the percentage rises from 28.5% on forecast day 1 to 36.8% on forecast day 5 (Fig. 3b). Given the large wet bias uncovered in Southeast Asia in boreal summer (Figs. 1 and 2), we now evaluate boxplots of the precipitation biases at the 122 stations located in Southeast Asia (20°–30°N and 100°–120°E). Figures 3c and 3d corroborate the wet bias and highlight that its magnitude and variability are largest in summer; and the mean bias is also larger at 1200 UTC, a finding largely arising because the 0000 UTC accumulation window has larger observed precipitation totals. Moreover, at least 75% of the Southeast Asia stations have a wet bias across forecast days 1–5. Last, further evidence of the wet bias in Fig. 3 is provided by inspecting the 95% confidence intervals (notches) around the boxplot medians, all of which are >0.

In summary, as for each half year (boreal winter and summer) there are similar precipitation biases for forecasts starting at 0000 UTC and 1200 UTC and across the forecast days, hereafter we only investigate the biases at 0000 UTC on

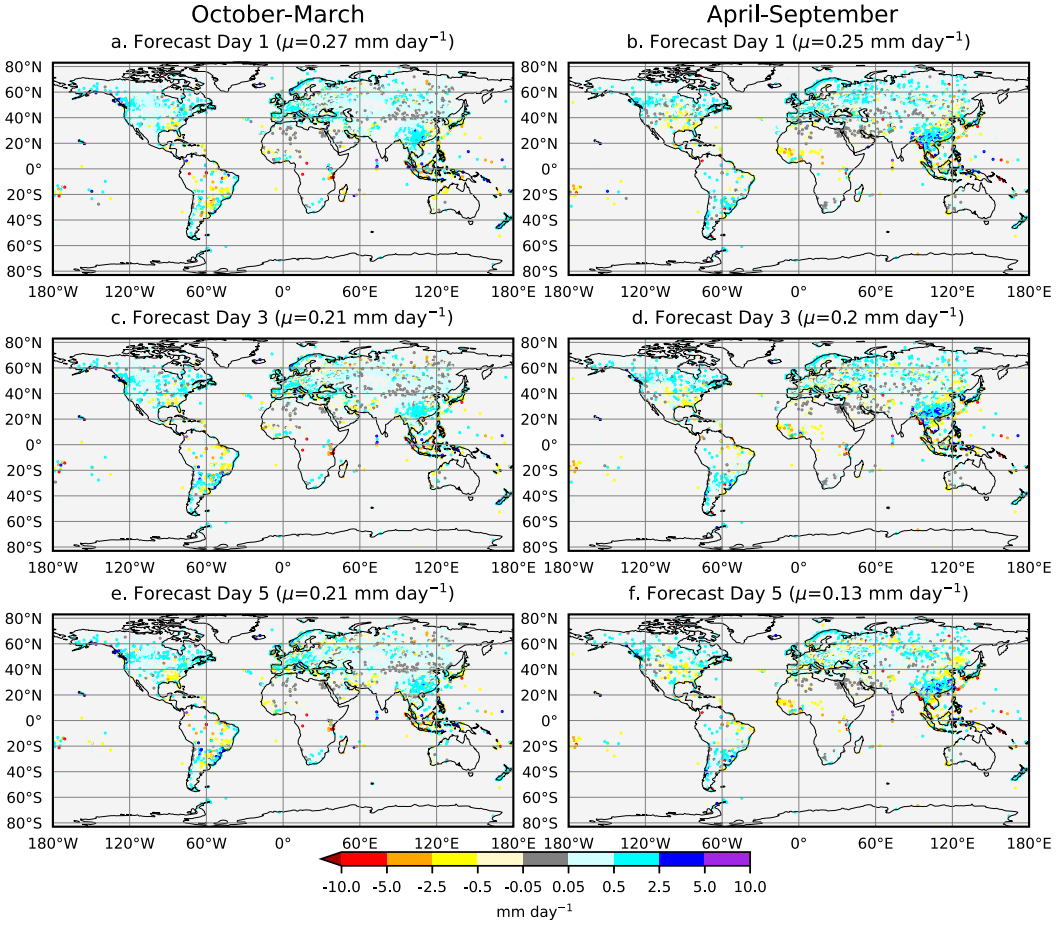

FIG. 1. The precipitation biases (mm day^{-1}) at 0000 UTC at the 5219 stations on forecast days 1, 3, and 5 in (a),(c),(e) boreal winter and (b),(d),(f) boreal summer. The average bias across the 5219 stations is shown in the title of each panel.

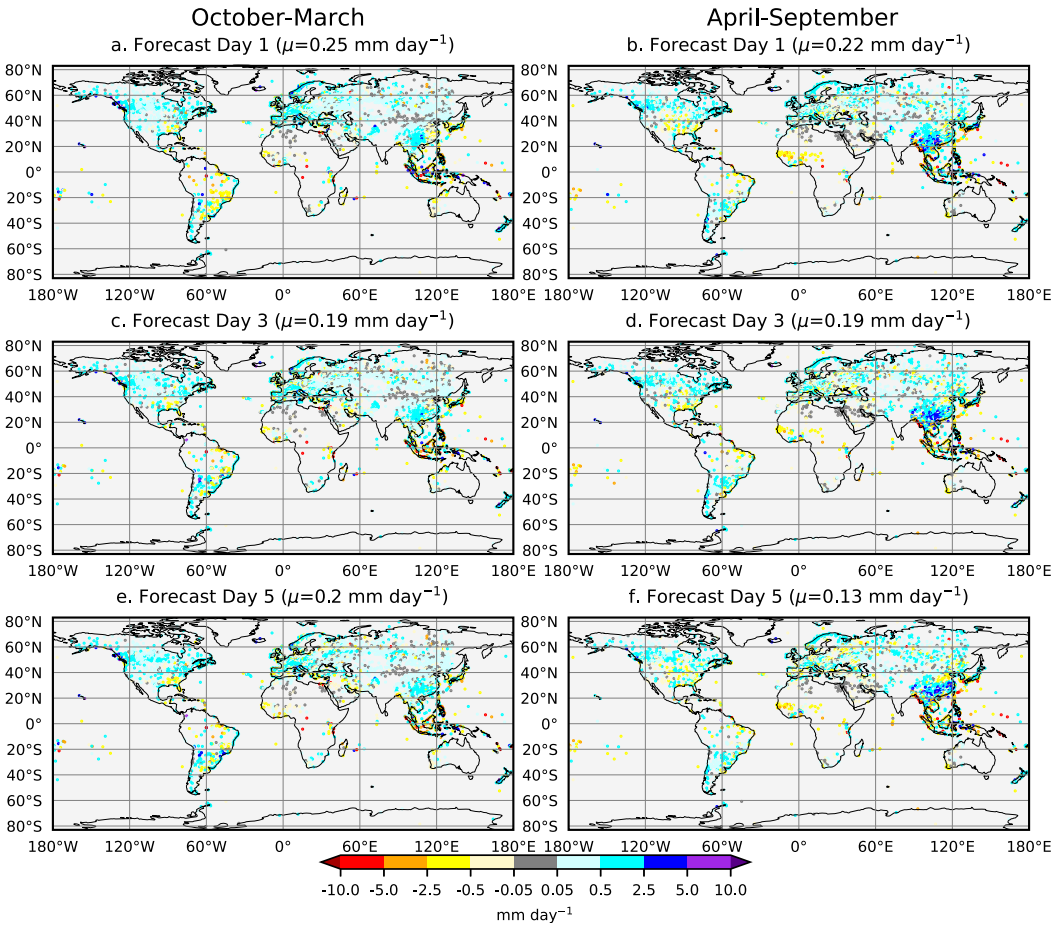

FIG. 2. The precipitation biases (mm day^{-1}) at 1200 UTC at the 5219 stations on forecast days 1, 3, and 5 in (a),(c),(e) boreal winter and (b),(d),(f) boreal summer. The average bias across the 5219 stations is shown in the title of each panel.

forecast day 3 because in general these are representative of the biases found.

b. *Dependence of the biases on observed precipitation*

The dependence of the bias on the average observed (climatological) precipitation at each station is now investigated on forecast day 3 for boreal winter in Fig. 4. It is found that the model has a wet bias for stations with low observed precipitation totals (i.e., the points are mostly above the $y = 0$ line), and a dry bias for stations with high observed precipitation totals (i.e., the points are mostly below the $y = 0$ line). This is exemplified by assessing the conditional distribution of the biases given the 10th and 90th percentiles of the observed precipitation from all stations; and these biases are 0.31 and -1.13 mm day^{-1} (Fig. 4c), respectively. There is also an indication that it is the stations in the tropics (red dots in Fig. 4a) that have the highest observed precipitation totals and dry biases of 1.8 (Fig. 4b) and -0.33 mm day^{-1} (Fig. 4c), respectively. This reflects the heavy convective-type precipitation that occurs there and the difficulty the model can have in resolving the small spatial scales associated with convective precipitation. Figure 5 shows similar findings in boreal summer, albeit with two main differences. First, there is more precipitation observed in boreal summer, with for example, a station average in the tropics of 2.1 mm day^{-1} compared to 1.8 mm day^{-1} in winter (cf. Figs. 4b and 5b). Second, except for the subtropical and polar regions, the average biases presented in Fig. 5c are smaller, a result most noticeable at the (climatologically) wettest 10% of stations where the bias reduces from -1.13 mm day^{-1} in boreal winter to -0.91 mm day^{-1} in boreal summer.

c. *Evolution of the biases over the study period*

The temporal variability of the bias on forecast day 3 using all global stations is now evaluated in Fig. 6a. By calculating on each day, the average forecast value from all available stations and subtracting the corresponding average observed value, this shows that on average the model wet bias is 0.21 mm day^{-1} across the study period. There is also a suggestion of larger variability in boreal summer, a time when there is a standard deviation of 0.26 mm day^{-1} compared to 0.22 mm day^{-1} in boreal winter. The wet bias is confirmed in Fig. 6b by assessing

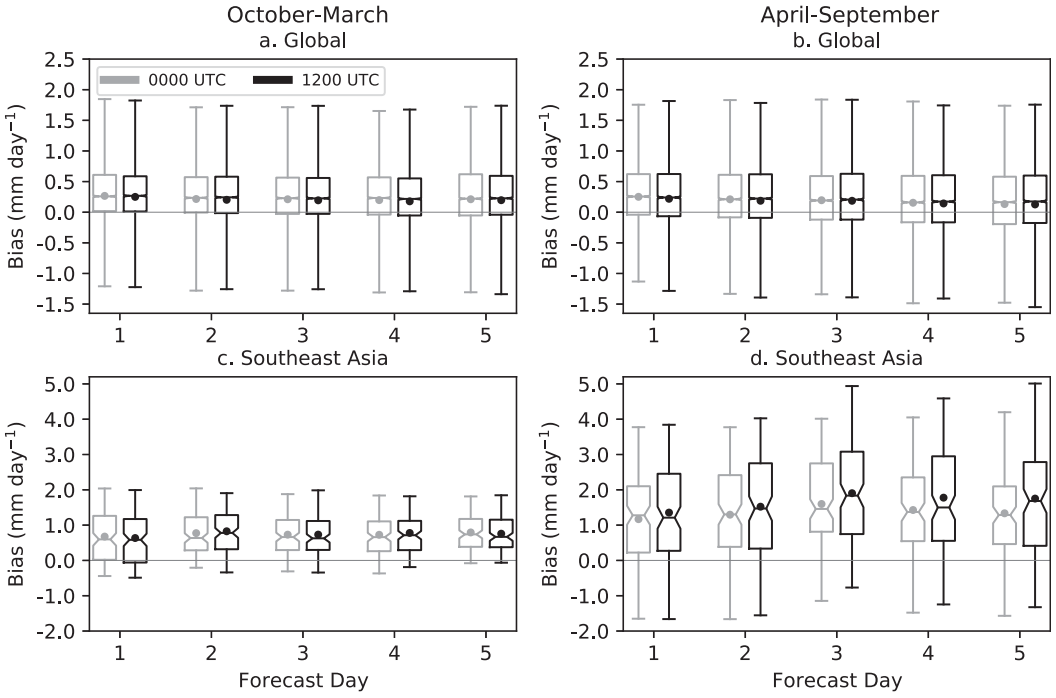

FIG. 3. The evolution of the bias with lead time in October–March (boreal winter) and April–September (boreal summer) for (a),(b) 5219 global stations and (c),(d) 122 stations in Southeast Asia (20°–30°N and 100°–120°E). The bottom and top of the boxes are the 25th and 75th percentiles, respectively, the line in the box is the median, the dot in the box is the mean, and the whiskers are the 5th and 95th percentiles. The notches in the boxplots show the 95% confidence interval around the median calculated from a 1000 bootstrapped sample.

the accumulated forecast and observed precipitation (double mass plots) over the period 15 June 2019–14 June 2020. Here it is shown that the forecast precipitation of 969.4 mm is 8.6% larger than the observed precipitation of 892.5 mm. This global wet bias is hypothesized in part to be due to the representativeness of the model grid in capturing a point (station) observation. Of the 5219 stations evaluated, the station height is available at 5015 locations, and these have an average height of 375.7 m. By taking the IFS model height from the nearest neighbor grid points, the average IFS model height is 439.5 m. This suggests that the model grid points are generally at a higher altitude compared to the stations which tend to be situated in relatively low-lying areas; hence, higher precipitation totals may occur, for example, from orographic enhancement. In the future, the development and implementation of NWP models with finer orography and more realistic land surface characteristics will improve the IFS grid point representativeness of station observations, which could in turn, decrease the bias.

We also examine the bias with time for stations in Southeast Asia because of the large model wet bias in boreal summer (Figs. 1b,d,f, 2b,d,f, and 3d). Figure 6c shows its temporal variability and it confirms that the boreal summer has the largest bias with an average value of 1.61 mm day^{-1} (compared with 0.74 mm day^{-1} in boreal winter). There is also more variance in the bias in the summer, as quantified by a standard deviation of 2.13 mm day^{-1} (0.98 mm day^{-1} in winter). The magnitude of the bias is further illustrated in the double mass plots in Fig. 6d, where strikingly, the forecast precipitation of 1815.1 mm is 31.0% larger than the observed precipitation of 1385.4 mm. Given the size of this wet bias in Southeast Asia, in the next section we examine the large-scale atmospheric patterns responsible for the most erroneous forecasts.

d. Large-scale atmospheric patterns across Southeast Asia

To investigate the large-scale atmospheric patterns, we first assess the control forecast of precipitation, water vapor flux, and 500-hPa geopotential height on forecast day 3 for the top two bias events (these are marked by numbers in Fig. 6c). The largest average wet bias of 9.55 mm day^{-1} occurred for the 24-h period up to 0000 UTC 26 June 2019, a day when forecast precipitation values of >150 mm contrasted with little precipitation observed (open circles in Fig. 7a). The forecast rainband across the region in Fig. 7a is known as the mei-yu front (e.g., Sampe and Xie 2010), and this feature is also clear in the water vapor flux at 1200 UTC 25 June 2019 in Fig. 7b. A possible reason for the large wet bias becomes clear by subtracting the $T + 0$ analysis fields of water vapor flux and 500 hPa geopotential height valid at 1200 UTC 25 June 2019 from the $T + 60$ forecast fields valid at the same time. This reveals a strong anomalous water vapor flux of >250 kg m^{-1} s^{-1} (thick black contour) along the coast (Fig. 7b) and shows a broad area (red filled contours) of negative geopotential height anomalies (Fig. 7c). These plots suggest that the atmospheric circulation and mei-yu front developed erroneously in the model, which indicates an area for future model improvement. Previous

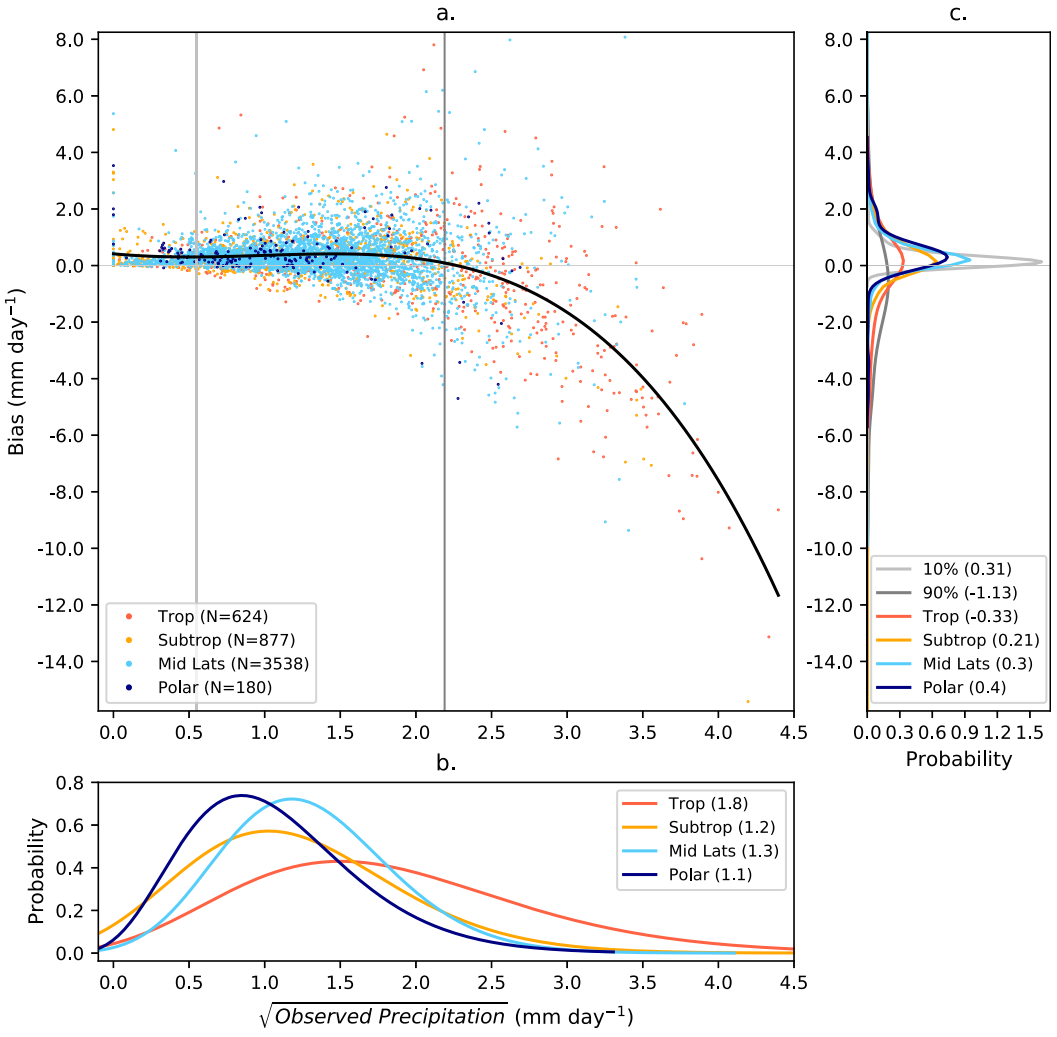

FIG. 4. (a) The average observed precipitation in square root space (mm day^{-1}) vs the forecast-minus-observation bias (mm day^{-1}) at 0000 UTC on forecast day 3 at the 5219 stations in boreal winter. The black line is the fourth-order polynomial, and the legend shows the number of stations in the tropics (20°S–20°N), subtropics (20°–35°N and 20°–35°S), midlatitudes (35°–65°N and 35°–65°S), and polar regions (>65°N and >65°S). (b) The gamma-estimated probability distribution function of average observed precipitation at stations in the four regions; the average values are given in the legend. (c) The Gaussian-estimated probability distribution function of the bias given the 10th and 90th percentiles of the observed precipitation and at stations in the four regions; the average values are given in the legend.

research has also suggested a northward bias in the positioning of the mei-yu front in the IFS (Ma et al. 2019). For the 24-h accumulated precipitation up to 0000 UTC 22 May 2020, the second largest wet bias of 7.75 mm day^{-1} occurred, and Fig. 7d shows the areas where the forecast precipitation did not match the observed values (e.g., around 26°N, 115°E). It is near 26°N, 115°E that at $T + 60$ valid at 1200 UTC 21 May 2020 the water vapor flux is overestimated by >250 kg m^{-1} s^{-1} (Fig. 7e) and the 500-hPa geopotential height had a trough (Fig. 7f) that did not materialize.

We secondly employ composite analysis to evaluate during the top 10% ($n = 36$) of wet bias days the difference between the forecast day 3 large-scale atmospheric patterns and those patterns that were later produced as the IFS analysis. (Note that 32 of the 36 events assessed were in the boreal summer half year.) To do this, we subtract the $T + 0$ analysis field at 1200 UTC on each day from the $T + 60$ forecast field valid at the same time, and then compute a composite mean error by averaging the 36 error fields. A 90% confidence interval around the composite mean error at each grid point is also calculated as $1.645 \times (\sigma/\sqrt{n})$, where 1.645 is the critical value used to calculate the two-sided confidence interval at the 90% level, σ is the standard deviation, and n is the sample size. Figures 8a and 8b display the composite patterns for the 500-hPa geopotential height and water vapor flux, respectively. For the geopotential height, there are significant error patterns

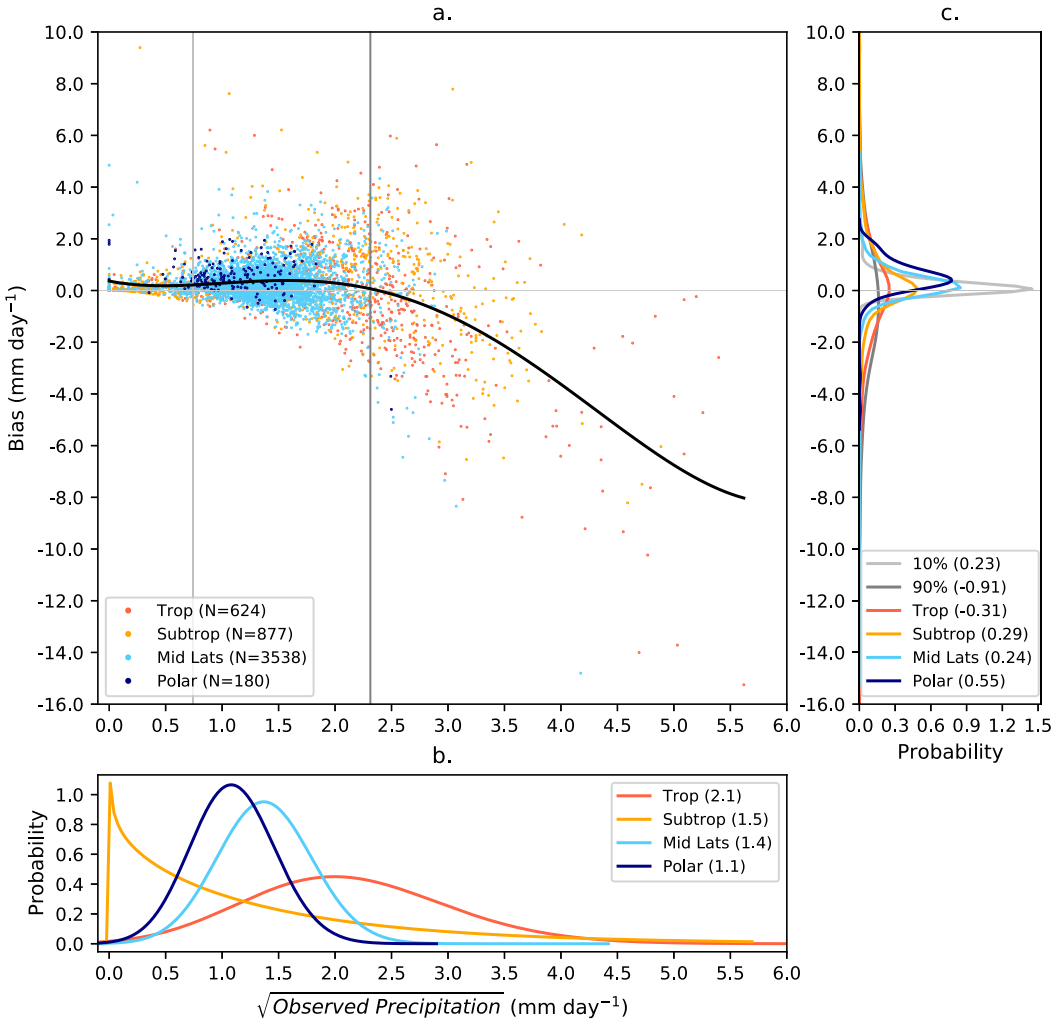

FIG. 5. (a) The average observed precipitation in square root space (mm day^{-1}) vs the forecast-minus-observation bias (mm day^{-1}) at 0000 UTC on forecast day 3 at the 5219 stations in boreal summer. The black line is the fourth-order polynomial, and the legend shows the number of stations in the four regions (the region definitions are given in the caption of Fig. 2). (b) The gamma-estimated probability distribution function of average observed precipitation at stations in the four regions; the average values are given in the legend. (c) The Gaussian-estimated probability distribution function of the bias given the 10th and 90th percentiles of the observed precipitation and at stations in the four regions; the average values are given in the legend.

surrounding the Tibetan Plateau, with negative errors (or the $T + 60$ forecasts over deepening the circulation or having too low heights) to the southeast (over Southeast Asia) and to the northwest of the Tibetan Plateau and positive errors (or erroneously high heights at $T + 60$) to the northeast near the Korean Peninsula. These positive errors also extend to the southwest toward the northern Philippines, which together with the erroneous negative heights over Southeast Asia, may indicate a misplacement and northward bias of the mei-yu front in the IFS (Ma et al. 2019). There is also an area of negative errors centered on southern India. The water vapor flux (Fig. 8b) has positive errors $> 25\,\mathrm{kg\,m^{-1}\,s^{-1}}$ (blue colors) in a southwest–northeast orientated corridor across Southeast Asia, which provides evidence that anomalous water vapor flux in the forecasts is associated with the model wet bias. A region of negative flux errors $< -25\,\mathrm{kg\,m^{-1}\,s^{-1}}$ (red colors) over the Philippine Sea from the north of the Philippines to the south of Japan suggests the misplacement of the water vapor flux in the forecasts and thus may corroborate the suggestion of a northward bias of the mei-yu front.

4. Discussion

There are three main implications of the uncovered precipitation biases. The biases will first cause errors in the water balance of the IFS, which will affect the soil moisture, evaporation, temperature, and snowpack at the surface. This will then degrade the land–atmosphere feedback and hence deteriorate the IFS forecast skill. Second, any river discharge based on raw IFS precipitation forecasts will be biased, thus lowering

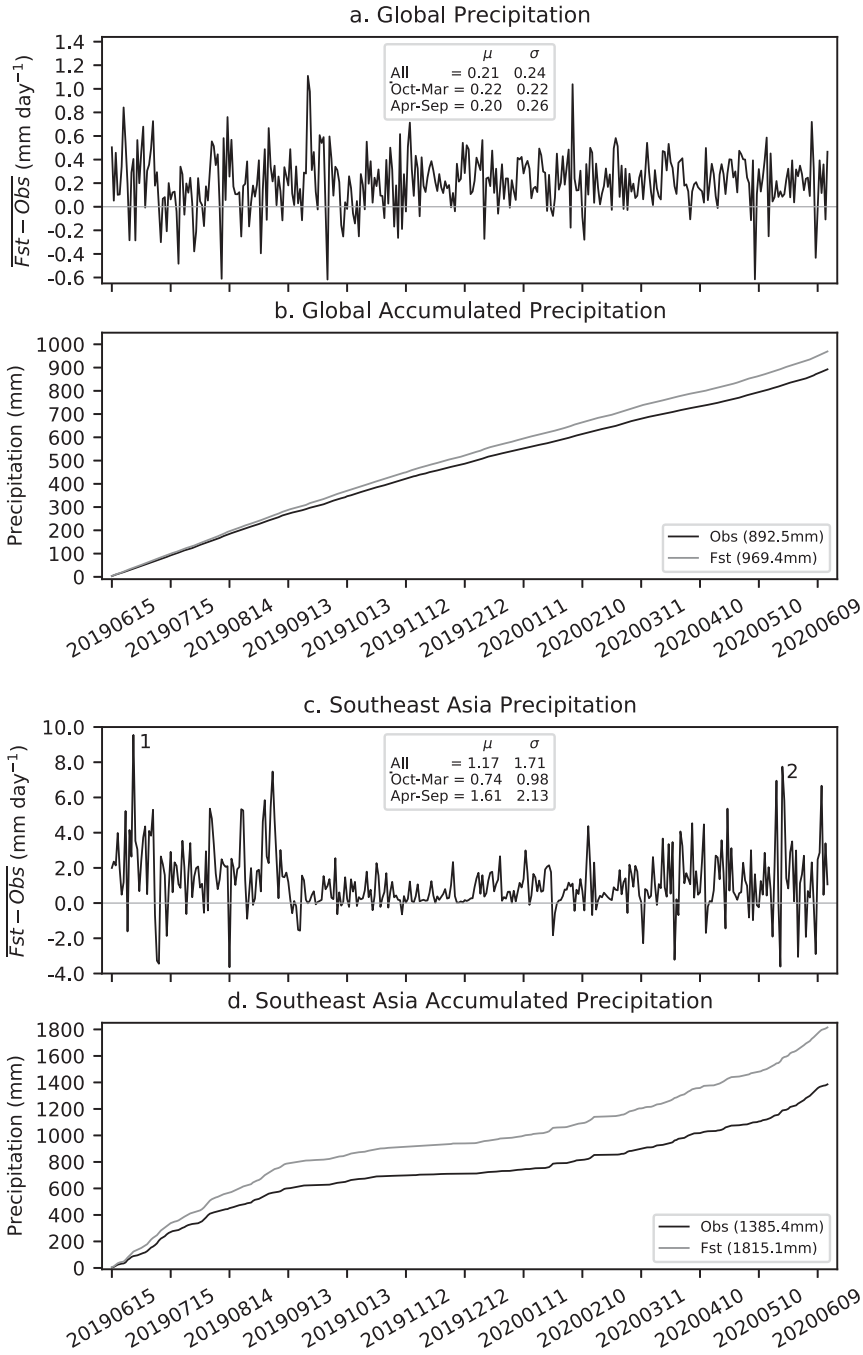

FIG. 6. (a) The bias (mm day^{-1}) using all global stations at 0000 UTC on forecast day 3 from 15 Jun 2019 to 14 Jun 2020; the median number of stations across all days was 4969. (b) Accumulated precipitation (double mass plots) using average forecast and observed values from all stations. (c) The bias (mm day^{-1}) using stations in Southeast Asia (20°–30°N and 100°–120°E) at 0000 UTC on forecast day 3 from 15 Jun 2019 to 14 Jun 2020; the median number of stations across all days was 120. The numbers 1 and 2 mark the two largest wet bias precipitation events. (d) Accumulated precipitation (double mass plots) using average forecast and observed values from stations in Southeast Asia. The legends in (a) and (c) provide the average and standard deviation of the bias for all days, October–March, and April–September periods; the legends in (b) and (d) give the forecast and observed precipitation accumulations.

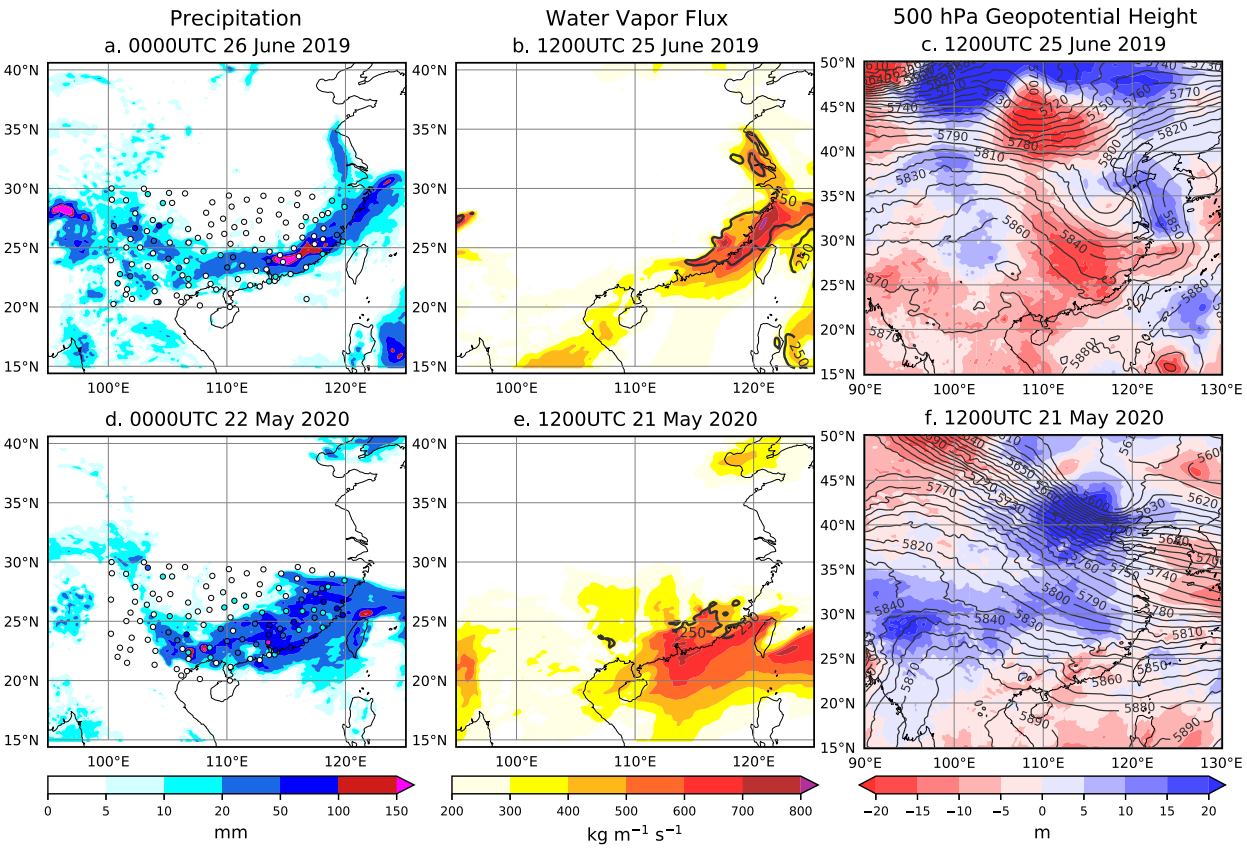

FIG. 7. (top) The wet bias event from 0000 UTC 25 Jun 2019 to 0000 UTC 26 Jun 2019. (a) The 24-h accumulated precipitation on forecast day 3 (from $T + 48$ to $T + 72$ initialized at 0000 UTC 23 Jun 2019; filled contours) and observed station precipitation (circles, which are also filled with colors). (b) At 1200 UTC 25 Jun 2019 the water vapor flux at $T + 60$ (filled contours; forecast initialization at 0000 UTC 23 Jun 2019) and the $T + 60$ minus $T + 0$ error field (thick black 250 kg m^{-1} s^{-1} line contour). (c) At 1200 UTC 25 Jun 2019 the 500-hPa geopotential height at $T + 60$ (line contours; forecast initialization at 0000 UTC 23 Jun 2019) and the $T + 60$ minus $T + 0$ error field (filled contours). (bottom) The wet bias event from 0000 UTC 21 May 2020 to 0000 UTC 22 May 2020. (d) The 24-h accumulated precipitation on forecast day 3 (from $T + 48$ to $T + 72$ initialized at 0000 UTC 19 May 2020; filled contours) and observed station precipitation (circles, which are also filled with colors). (e) At 1200 UTC 21 May 2020 the water vapor flux at $T + 60$ (filled contours; forecast initialization at 0000 UTC 19 May 2020) and the $T + 60$ minus $T + 0$ error field (thick black 250 kg m^{-1} s^{-1} line contour). (f) At 1200 UTC 21 May 2020 the 500-hPa geopotential height at $T + 60$ (line contours; forecast initialization at 0000 UTC 19 May 2020) and the $T + 60$ minus $T + 0$ error field (filled contours). Note that (c) and (f) have a larger domain.

skill in hydrological forecast systems, such as GloFAS. Results in Harrigan et al. (2020a) based on the GloFAS-ERA5 discharge reanalysis found a discharge bias that corresponds well to the wet bias in Southeast Asia herein. This suggests the propagation of the precipitation bias into the hydrological system, and as the GloFAS-ERA5 discharge reanalysis is based on ERA5, which is itself based on an earlier version of the IFS, this indicates the wet bias in Southeast Asia is a longstanding IFS issue. A third consequence is that the precipitation wet bias would lead to an overestimation of GloFAS river discharge and hence too high a modeled freshwater flux from the land to the ocean. The resulting reduced ocean salinity then in turn could affect the ocean circulation and degrade ocean–atmosphere interactions. It has been previously found in a project—named the Benefits of dynamically modeled river discharge input for ocean and coupled atmosphere–land–ocean systems (Mercator Ocean 2020)—that when coupling the land and the ocean with the GloFAS-ERA5 river discharge reanalysis, a large degradation was seen in ocean modeling skill around Southeast Asia and the Maritime Continent. The precipitation biases herein may now help to explain this finding. Taking an Earth system modeling approach (e.g., Harrigan et al. 2020b) means that reducing precipitation biases has enormous potential to improve many aspects of the forecast chain. In summary, correcting precipitation biases and the closing of the water balance could provide more realistic land–atmosphere feedbacks, better river discharge, better land–ocean coupling, and thus more skillful weather forecasts.

The mei-yu front discussed in section 3d is part of the East Asian summer monsoon (Sampe and Xie 2010), which itself is one subsystem of the larger Asian monsoon system (Yihui and Chan 2005). Our results showing that the model has trouble in handling the atmospheric circulation and regional water cycle in Southeast Asia are not only an important finding for the

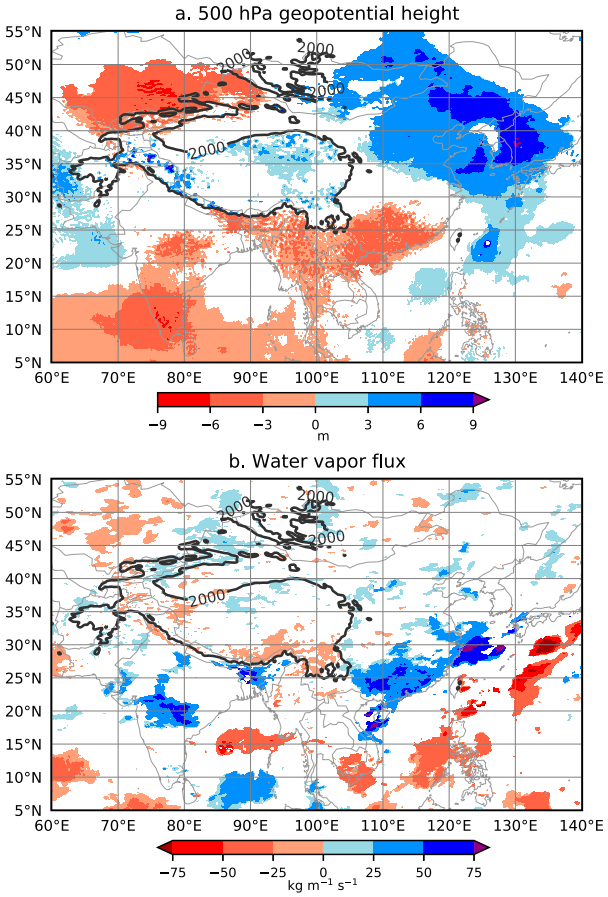

FIG. 8. Composite mean error patterns of the top 10% ($n = 36$) wet bias days on forecast day 3 for (a) 500-hPa geopotential height and (b) water vapor flux. Colored regions identify areas where the composite mean is different from zero at the 90% confidence level. The thick black contours in each panel show where the IFS orography (on the O640 grid) is 2000 m.

forecast skill in that region. These forecast errors will also have consequences for the downstream development and hence predictability in the North Pacific storm track via, for example, the incorrect latent heat release that arises from the erroneous forecast water vapor flux. There are multiple possible reasons as to why this region is prone to the uncovered circulation errors and precipitation biases, and these are as follows: 1) the difficulty in specifying the snowpack in the IFS in the complex terrain of the Tibetan Plateau; 2) uncertainties in representing convection and cloud and moist processes in this region; 3) errors related to synoptic-scale systems called Tibetan Plateau vortices that develop over the Tibetan Plateau (Li et al. 2019, 2020); 4) tropical variability, such as the Madden–Julian oscillation, and the modeled relationship between the tropics and extratropics; and 5) limitations associated with aerosol parameterizations in the IFS (Rémy et al. 2019). Furthermore, a model study by Lau and Kim (2018) expounds a hypothesis related to the settling of light-absorbing aerosols (e.g., from the Middle East and west Asia) on the snowpack of the Tibetan Plateau during April–June. It is suggested that these aerosols can reduce the surface albedo, increase snowmelt and surface heating, and then be associated with a Rossby wave train into East Asia which weakens the subtropical westerly jet and strengthens (and displaces to the north) the mei-yu front. Although out of the scope of this paper, these topics warrant further investigation.

5. Conclusions

This study has used precipitation observations from 5219 stations over a whole year to assess precipitation biases in the ECMWF IFS. Results show that in each half year (boreal winter and summer), the biases have similar patterns and values for forecasts starting at 0000 and 1200 UTC and across forecast days 1–5, and on average globally there is a wet bias of up to 0.27 mm day^{-1} at 0000 UTC in the boreal winter half year. These precipitation biases are, in general, dependent on the observed (or climatological) precipitation at the stations, with the stations with low observed precipitation having an IFS wet bias and stations with high observed precipitation having an IFS dry bias. These results agree with previous findings using the GloFAS-ERA5 river discharge reanalysis that showed that the largest positive biases occur in climatologically drier catchments (Harrigan et al. 2020a). A significant model wet bias has also been identified in Southeast Asia, a region that has an average bias of 1.61 mm day^{-1} in the boreal summer half year, and over the year studied, a precipitation overestimation on forecast day 3 of 31.0%. This issue is the hydrological signature of several hypothesized processes, namely, the difficulty in specifying the snowpack on the Tibetan Plateau in the IFS, possible uncertainties in representing convection and cloud and moist processes, potential errors with synoptic-scale systems called Tibetan Plateau vortices, tropical variability, and limitations with the aerosol parameterizations in the IFS. These factors are thought to affect the IFS's handling of the mei-yu front resulting in an over deepening of the circulation (as shown herein with the 500-hPa geopotential height), an anomalously large water vapor flux over Southeast Asia, and an overestimation of precipitation. The hypothesized causes of these erroneous forecasts presented above require further investigation.

The uncovered precipitation biases have implications for the soil moisture, evaporation, temperature, and snowpack in the IFS, which will then affect land–atmosphere feedbacks in the model. River discharge forecasts generated from IFS precipitation will also exhibit biases, thus lowering skill in hydrological forecast systems (e.g., GloFAS), and we have suggested that this biased freshwater flux from Southeast Asia could be a key contributor to ocean model salinity and circulation issues in the region and across the Maritime Continent. To address systematic issues in clouds, precipitation, and radiation, a major moist physics upgrade is planned in an upcoming IFS cycle (Bechtold et al. 2020); but we note that these model improvements do not fully address the biases presented herein, such as the IFS wet bias at climatologically dry stations. This study is also potentially subject to the representativeness issue that arises when verifying a model grid against a point (station) observation (Haiden et al. 2018). Although the results

in Haiden et al. (2018) indicate that the representativeness will not significantly change our findings, this is a topic that could be investigated in future by using gridded precipitation datasets (e.g., from satellite) or by characterizing the representativeness errors using parametric approaches (Ben Bouallègue 2020). Finally, the issue of precipitation gauges predominantly being situated in relatively low-lying areas is symptomatic of a broader problem, that is the lack of readily available precipitation (and river discharge) observations. We therefore make another call here, as in Lavers et al. (2019), for an international effort to be made toward the wider availability and sharing of hydrological observations. This would allow for the identification of further model issues which, in turn, would lead to improvements in modeling the global water cycle and would bring socioeconomic benefits.

Acknowledgments. The authors acknowledge financial support from the European Union Copernicus Emergency Management Service (Framework Contract 198702). We also thank David Richardson, Richard Forbes, and Zied Ben Bouallègue for discussions on this research, and the two anonymous reviewers whose comments helped to improve the paper.

Data availability statement. The data used are available through the ECMWF archive (https://www.ecmwf.int/en/forecasts/datasets/archive-datasets).

REFERENCES

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013: GloFAS - Global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.*, **17**, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013.

Bechtold, P., R. Forbes, I. Sandu, S. Lang, and M. Ahlgrimm, 2020: A major moist physics upgrade for the IFS. *ECMWF Newsletter*, No. 164, ECMWF, Reading, United Kingdom, 24–32, https://www.ecmwf.int/node/19720.

Ben Bouallègue, Z., 2020: Accounting for representativeness in the verification of ensemble forecasts. ECMWF Tech. Memo. 865, ECMWF, 28 pp., https://www.ecmwf.int/node/19544.

Emerton, R., and Coauthors, 2020: Emergency flood bulletins for cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response. *Int. J. Disaster Risk Reduct.*, **50**, 101811, https://doi.org/10.1016/j.ijdrr.2020.101811.

Gayathri, K. D., B. P. Ganasri, and G. S. Dwarakish, 2015: A review on hydrological models. *Aquat. Procedia*, **4**, 1001–1007, https://doi.org/10.1016/j.aqpro.2015.02.126.

Haiden, T., and Coauthors, 2018: Use of in situ surface observations at ECMWF. ECMWF Tech. Memo. 834, ECMWF, 28 pp., https://www.ecmwf.int/node/18748.

Hamill, T. M., E. Engle, D. Myrick, M. Peroutka, C. Finan, and M. Scheuerer, 2017: The U.S. National Blend of models for statistical postprocessing of probability of precipitation and deterministic precipitation amount. *Mon. Wea. Rev.*, **145**, 3441–3463, https://doi.org/10.1175/MWR-D-16-0331.1.

Harrigan, S., and Coauthors, 2020a: GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth Syst. Sci. Data*, **12**, 2043–2060, https://doi.org/10.5194/essd-12-2043-2020.

——, H. Cloke, and F. Pappenberger, 2020b: Innovating global hydrological prediction through an Earth system approach. *WMO Bulletin*, Vol. 69, World Meteorological Organization, Geneva, Switzerland, https://public.wmo.int/en/resources/bulletin/innovating-global-hydrological-prediction-through-earth-system-approach.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hewson, T. D., and F. M. Pillosu, 2020: A new low-cost technique improves weather forecasts across the world. arXiv, 27 pp., https://arxiv.org/abs/2003.14397.

Kingston, D. G., G. R. McGregor, D. M. Hannah, and D. M. Lawler, 2006: River flow teleconnections across the northern North Atlantic region. *Geophys. Res. Lett.*, **33**, L14705, https://doi.org/10.1029/2006GL026574.

Lau, W. K. M., and K.-M. Kim, 2018: Impact of snow darkening by deposition of light-absorbing aerosols on snow cover in the Himalayas–Tibetan Plateau and influence on the Asian summer monsoon: A possible mechanism for the Blanford hypothesis. *Atmosphere*, **9**, 438, https://doi.org/10.3390/atmos9110438.

Lavers, D. A., S. Harrigan, E. Andersson, D. S. Richardson, C. Prudhomme, and F. Pappenberger, 2019: A vision for improving global flood forecasting. *Environ. Res. Lett.*, **14**, 121002, https://doi.org/10.1088/1748-9326/ab52b2.

——, and Coauthors, 2020: A vision for hydrological prediction. *Atmosphere*, **11**, 237, https://doi.org/10.3390/atmos11030237.

Li, L., R. H. Zhang, M. Wen, J. P. Duan, and Y. J. Qi, 2019: Characteristics of the Tibetan plateau vortices and the related large-scale circulations causing different precipitation intensity. *Theor. Appl. Climatol.*, **138**, 849–860, https://doi.org/10.1007/s00704-019-02870-4.

——, C. Zhu, R. Zhang, and B. Liu, 2020: Roles of the Tibetan Plateau vortices in the record Meiyu rainfall in 2020. *Atmos. Sci. Lett.*, **2020**, e1017, https://doi.org/10.1002/asl.1017.

Ma, J., K. A. Bowley, and F. Zhang, 2019: Evaluating the forecast performance of the Meiyu front rainbelt position: A case study of the 30 June to 4 July 2016 extreme rainfall event. *Atmosphere*, **10**, 648, https://doi.org/10.3390/atmos10110648.

Magnusson, L., and Coauthors, 2019: ECMWF works with universities to support response to tropical cyclone Idai. *ECMWF Newsletter*, No. 160, ECMWF, Reading, United Kingdom, 2–3, https://www.ecmwf.int/en/newsletter/160/news/ecmwf-works-universities-support-response-tropical-cyclone-idai.

Mercator Ocean, 2020: Benefits of dynamically modelled river discharge input for ocean and coupled atmosphere-land-ocean systems. Accessed 15 December 2020, https://www.mercator-ocean.fr/en/portfolio/bronco-2/.

Muchan, K., and H. Dixon, 2019: Insights into rainfall undercatch for differing raingauge rim heights. *Hydrol. Res.*, **50**, 1564–1576, https://doi.org/10.2166/nh.2019.024.

Pechlivanidis, I. G., B. M. Jackson, N. R. McIntyre, and H. S. Wheater, 2011: Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST J.*, **13**, 193–214.

Rémy, S., and Coauthors, 2019: Description and evaluation of the tropospheric aerosol scheme in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS-AER, cycle 45R1). *Geosci. Model Dev.*, **12**, 4627–4659, https://doi.org/10.5194/gmd-12-4627-2019.

Sampe, T., and S. Xie, 2010: Large-scale dynamics of the Meiyu-Baiu rainband: Environmental forcing by the westerly jet. *J. Climate*, **23**, 113–134, https://doi.org/10.1175/2009JCLI3128.1.

WMO, 2020: Precipitation Essential Climate Variable (ECV) factsheet. Accessed 30 November 2020, https://gcos.wmo.int/en/essential-climate-variables/precipitation.

Yihui, D., and J. Chan, 2005: The East Asian summer monsoon: An overview. *Meteor. Atmos. Phys.*, **89**, 117–142, https://doi.org/10.1007/s00703-005-0125-z.

Zsótér, E., C. Prudhomme, and S. Harrigan, 2019a: Major upgrade for global flood forecasts. *ECMWF Newsletter*, No. 158, ECMWF, Reading, United Kingdom, 8, https://www.ecmwf.int/en/newsletter/158/news/major-upgrade-global-flood-forecasts.

——, H. Cloke, E. Stephens, P. de Rosnay, J. Muñoz Sabater, C. Prudhomme, and F. Pappenberger, 2019b: How well do operational numerical weather prediction configurations represent hydrology? *J. Hydrometeor.*, **20**, 1533–1552, https://doi.org/10.1175/JHM-D-18-0086.1.