

Subseasonal prediction of tropical cyclone precipitation

Jorge L. García-Franco ^{a,b}, Chia-Ying Lee ^a, Michael K. Tippett ^c, Suzana J. Camargo ^a,
Daehyun Kim ^{d,e}, Andrea Molod ^f and Young-Kwon Lim ^{f,g}

^a *Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY*

^b *Escuela Nacional de Ciencias de la Tierra, Universidad Nacional Autónoma de México, Mexico
City*

^c *Department of Applied Physics and Applied Mathematics, Columbia University, NY*

^d *School of Earth and Environmental Sciences, Seoul National University, Korea.*

^e *Department of Atmospheric Sciences, University of Washington, Seattle, Washington*

^f *Global Modeling and Assimilation Office, Goddard Space Flight Center, NASA, Greenbelt, MD,
USA*

^g *University of Maryland, Baltimore County, Baltimore, MD, USA*

Corresponding author: Jorge L. García-Franco, jorgegf@ldeo.columbia.edu

14 ABSTRACT: The accurate prediction of tropical cyclone precipitation (TCP) at an extended-
15 range could be crucial to mitigate the impacts of TC-related flooding. This study examines
16 probabilistic predictions of weekly-accumulated TCP and total precipitation using 11 subseasonal
17 forecast systems. Raw, uncalibrated, categorical forecasts of basin-wide TCP are only skillful in
18 the ECMWF model and only up to 15 days in advance and except in the northern Indian Ocean
19 and the South Pacific. Calibration, through linear regression, improves forecasts and makes several
20 forecast systems (GEOS, UKMO) skillful up to 15 days in advance but only in some basins. In
21 most models and basins, such as the GEOS model in the Atlantic basin, the bias in the forecast
22 probability of TC occurrence is the main factor driving biases in TCP and decreasing forecast skill.
23 At the regional-scale, calibrated ECMWF forecasts are skillful beyond 15 day leads and globally.
24 The poor prediction of TCP in raw forecasts is shown to affect total precipitation prediction skill.
25 Therefore, biases in the TC occurrence probability forecast is the leading cause of low skill of TCP
26 and may play a role in the skill of total precipitation.

27 SIGNIFICANCE STATEMENT: Tropical cyclone (TC)-related flooding is a significant hazard,
28 and accurate extended-range predictions of tropical cyclone precipitation (TCP) are vital for risk
29 mitigation efforts. This study assesses 11 subseasonal forecast systems, finding that uncalibrated
30 models struggle with TCP prediction, with only the ECMWF model being skillful beyond 15
31 days in some basins. Calibration improves forecast skill, particularly for models like GEOS and
32 UKMO, but biases in TC occurrence forecasts remain a major challenge. These biases reduce
33 TCP prediction skill and affect total precipitation forecasts. Improving TCP prediction could
34 significantly enhance preparedness for TC-related flooding events in key regions of the tropics.

35 1. Introduction

36 Landfalling tropical cyclones (TCs) cause severe impacts such as loss of life, property damage
37 and economic disruption due to storm surge, strong winds and the intense precipitation rates found
38 around a TC (Villarini et al. 2014; Needham et al. 2015; Lenzen et al. 2019). TC precipitation
39 (TCP) is a major influence on flood frequency and flood-related hazards, such as flash floods and
40 landslides, in several tropical coastal regions (Hu et al. 2018; Smith et al. 2023). For example, the
41 TC-related rainfall is estimated to account for about 30% of TC-related deaths in the United States
42 (Rappaport 2014). TCP impacts are exacerbated by local vulnerabilities (Dominguez et al. 2021)
43 which in marginalized populations can lead to population displacement and migration (Black et al.
44 2013; Heslin et al. 2019). The destructive potential of TCs and TCP makes the skillful prediction
45 of both quantities a key element of disaster preparedness and early alert systems.

46 For this purpose, numerical weather prediction (up to 72-h) of TCP has steadily improved in
47 recent years (Luitel et al. 2018), partially due to reduced errors in track forecasts and increased
48 model resolution (Cangialosi et al. 2020). Reliable information at longer leads, such as extended
49 range forecasts (e.g., 6 to 20 days) on TCP could be useful to guide decision-making and inform
50 users on how to manage risk and mitigate TC rainfall-related impacts (Beatty et al. 2019; Nkiaka
51 et al. 2019; White et al. 2022). However, our understanding of the limitations and capabilities
52 of subseasonal-to-seasonal (2-6 weeks, hereafter S2S) predictions, especially of TCP and total
53 precipitation (P), is still at its infancy compared to seasonal and weather forecasts (Camargo et al.
54 2019; Domeisen et al. 2022).

55 In the context of S2S TC forecasting, case studies have shown skillful forecasts of genesis at lead-
56 times of 10-15 days (Xiang et al. 2015; Gregory et al. 2019). However, verification studies suggest
57 that few models, and only in select basins, are consistently skillful at predicting TC occurrence or
58 genesis beyond week 1 (Lee et al. 2018, 2020; Li et al. 2022; García-Franco et al. 2024). Model
59 skill at predicting other aspects of TC activity such as track clusters (Camargo et al. 2021) or the
60 modulation of TC activity by the Madden-Julian Oscillation (MJO Lee et al. 2018; Domeisen et al.
61 2022) is also limited beyond week 2, at least based on raw dynamical forecasts.

62 On the other hand, subseasonal prediction of total precipitation (i.e, not limited to TCP) has
63 been investigated extensively (see e.g. Li and Robertson 2015; Vigaud et al. 2018; De Andrade
64 et al. 2019; Specq and Batté 2020; Moron and Robertson 2020). Forecast skill from dynamical
65 models appears limited when predicting tropical rainfall beyond a 1-week lead time, regardless of
66 whether skill is diagnosed using deterministic (Li and Robertson 2015; De Andrade et al. 2019;
67 Specq and Batté 2020) or probabilistic skill scores (Vigaud et al. 2017; Klingaman et al. 2021).
68 The precipitation skill at S2S scales is lower over land than over the ocean (Moron and Robertson
69 2020). These low skill scores are due to noteworthy biases in the mean (García-Franco et al. 2023)
70 and variance of precipitation (De Andrade et al. 2021) which can hamper forecast reliability and
71 accuracy (Tippett et al. 2015; Specq and Batté 2020; Klingaman et al. 2021).

72 Forecast calibration has been shown to improve probabilistic TC occurrence and genesis forecasts,
73 extending skill to weeks 2 and 3 in some basins (Camargo et al. 2019; Lee et al. 2020; Schreck
74 et al. 2023). For example, in the context of TC occurrence, a linear regression-based calibration
75 can extend forecast skill beyond week 2, depending on the model and basin (Lee et al. 2020).
76 Presumably, TCP forecasts might benefit from calibration since TCP biases are primarily driven
77 by TC occurrence biases (García-Franco et al. 2023).

78 This work builds upon previous studies, particularly García-Franco et al. (2023, 2024). García-
79 Franco et al. (2023) analyzed TCP climatology in the World Meteorological Organization (WMO)
80 S2S models, identifying TC occurrence biases as the primary driver of TCP biases. García-Franco
81 et al. (2024) examined TCs in NASA's Goddard Earth Observing System (GEOS) S2S forecast
82 system version 2, evaluating S2S forecast skill for TC occurrence and MJO-related predictability.
83 However, the subseasonal prediction of TCP in dynamical models remains largely unexplored,

84 leaving a critical gap in understanding the practical capabilities and limitations of subseasonal TC
85 forecasts beyond occurrence.

86 This study aims to evaluate subseasonal TCP prediction skill in both the WMO S2S models
87 and GEOS, addressing challenges in raw forecasts and quantifying the impact of bias correction
88 methods on improving skill. Additionally, we explore the relationship between forecast skill for
89 TCP and total P, emphasizing the potential benefits of accurate TCP prediction for improving total
90 P forecasts.

91 The paper is structured as follows. Section 2 describes the observational and forecast datasets
92 as well as the verification methodology. Section 3 presents a brief investigation into the biases
93 of total P and TCP in the GEOS model to fill in the gaps between our knowledge of the WMO
94 S2S models and GEOS. Then, section 4 analyzes the prediction skill of TCP in all models, both
95 basin-wide and regionally. Section 5 considers the question of whether a better prediction of TCP
96 relates to a better prediction of total P. Finally, section 6 presents a summary and discussion of the
97 main results.

98 **2. Data and methods**

99 *a. Observational datasets*

100 TC tracks are obtained from the International Best Track Archive for Climate Stewardship
101 (IBTrACS; Knapp et al. 2018) dataset, version 4 (v04r00). The IBTrACS dataset is subsampled to
102 retain only the records at 00:00, 06:00, 12:00 and 18:00 UTC for consistency with the precipitation
103 datasets. Moreover, only storms that reached a lifetime maximum intensity (LMI) threshold of 34
104 kt are included in the analysis. Cases classified as tropical disturbances, waves or extra-tropical
105 are removed. Finally, only TCs equatorward of 40°N/S are considered.

106 Observed precipitation is taken from two satellite-based datasets: the Climate Prediction Center
107 (CPC) morphing technique (CMORPH; Xie et al. 2017; NOAA CDR Program 2022) and the Global
108 Precipitation Climatology Project (GPCP), version 1.2 (Huffman et al. 2001). GPCP data is used at
109 1° resolution and daily frequency and CMORPH data at 0.25° resolution and 6-hourly frequency.
110 Both datasets have comparable total precipitation and TCP climatologies (García-Franco et al.
111 2023), and were analyzed from 1998-2021 to match the hindcast period of the models.

TABLE 1. Summary of the models used in this study, including their full names, abbreviations, native resolution (horizontal resolution and the number of vertical levels), the coverage period used in this study, the ensemble size, reforecast frequency, and the LMI threshold. Models were classified into low (LR), medium (MR), and high (HR) resolutions.

Model	Version	Native resolution	Coverage Period	Ens. size	Rfc. frequency	LMI thr. [kt]
BoM	POAMA P24; 2014-01-01	2° [LR], L17	1998-2013	33	6/month	34
CNRM	CNRM-CM 6.0; 2014-12-01	1.4°, [LR], L91	1998-2014	15	4/month	28
CNR-ISAC	GLOBO; 2017-06-08	0.75° [MR], L54	1998-2010	5	5 days	22
CMA	BCC-CPS-S2Sv2; Jan-Dec 2021	0.5° [HR], L56	2006-2020	4	2/week	23
ECCC	GEPS 6; Jan-Dec 2021	0.35° [HR], L45	1998-2017	4	weekly	25
ECMWF	CY47R3; Jan-Dec 2022	0.15° [0-15 d], 0.3° [16-d] [HR], L137	2001-2021	11	2/week	25
GEOS	version 2; Jan-Dec 2020	0.5° [HR], L78	1999-2020	4	5 days	24
JMA	GEPS2003; 2017-01-31	0.5° [HR], L60	1998-2012	5	3/month	22
KMA	GloSea5-GC2; Jan-Dec 2021	0.75° [MR], L85	1998-2017	3	4/month	24
NCEP	CFSv2; 2011-03-01	1° [LR], L64	1999-2010	4	daily	26
UKMO	GloSea6; Jan-Dec 2022	0.75° [MR], L85	1998-2016	7	4/month	24

b. Forecast datasets

The GEOS S2S subseasonal forecast system version 2 (hereafter GEOS) has a 4-member ensemble and a 1/5-day reforecast frequency (Molod et al. 2020; Lim et al. 2021). GEOS is a fully coupled Earth System model with an atmospheric model that runs at a resolution of 0.5° and 72 vertical levels (Molod et al. 2015) and an ocean model, MOM5, that uses a tripolar grid at approximately 50 km horizontal resolution (Griffies et al. 2012). The GEOS reforecasts are initialized from a weakly coupled atmosphere-ocean data assimilation system and each forecast is run for 45 days (Molod et al. 2020).

In addition to GEOS, we use reforecasts from models participating in the WMO S2S project (Vitart 2017; Vitart and Robertson 2018; Lee et al. 2018). We use the most recent model versions that include at least one year of real-time forecast dates with corresponding complete hindcasts, ensuring a consistent reforecast dataset, updated from the model versions used by García-Franco et al. (2023).

The S2S reforecast dataset is available on a common 1.5°×1.5° horizontal grid and TC tracks are available at a daily frequency. A summary of the reforecast data used in this study is given in Table 1. Several models in the S2S database have been updated operationally, such as ACCESS-S2

132 (Camp et al. 2023) from BoM and GEFSv12 (Pegion et al. 2019) from NCEP. However, these
133 forecasts are not available in the S2S database, which is the primary dataset for this study.

134 *c. TC precipitation*

135 Precipitation from TCs is estimated using a fixed radius of 500 km. All precipitation within 500
136 km from the storm center is considered as TCP. This threshold is a common choice in observational
137 (Khouakhi et al. 2017) and modelling studies (Vannière et al. 2020) and produces similar results
138 to more sophisticated object-tracking methodologies (Lavender and McBride 2021). In the S2S
139 models, TCP is accumulated on daily frequencies, due to the fact that the TC tracks are only
140 available at this frequency. In the case of GEOS, TCP is calculated using the 6-hourly track and
141 6-hourly averaged precipitation, and this TCP is then accumulated daily.

142 For the observations, two separate calculations of TCP were conducted. First, the track from
143 IBTrACS was interpolated to daily frequency to fit the GPCP frequency, whereas for CMORPH the
144 TCP is defined using IBTrACS at the original 6-hourly frequency and then accumulated to daily
145 values of TCP. Similarly, CMORPH was interpolated to daily and 1° resolution, matching GPCP,
146 and the TCP calculations were repeated. Our analysis showed that skill scores are not sensitive to
147 these choices. Unless otherwise stated, the observed estimates of total and TCP correspond to the
148 analysis of CMORPH at the 1° resolution resolution and daily frequency.

149 *d. Defining forecasts and forecast verification*

150 As in previous studies of S2S TC prediction (Lee et al. 2018, 2020), forecasts are verified using
151 weekly averages, or more precisely, moving averages with a 7-day window. Lead-time is defined
152 as the number of days between a forecast start date and the valid time of the forecast. Forecasts
153 are verified from day 4 lead-time up to day 35 using the weekly periods with a moving window of
154 1 day. This means that the verification is done first for day 4, using forecast target lead-times from
155 day 1 to day 7. Then, verification for day 5 uses, either as averages or counts, lead-times from day 2
156 to 8. The term “week” is used to describe the forecasts, in which, for example, “week 1 forecasts”
157 refers to forecasts containing data from days 1 to 7. In the moving average verification, this same
158 period would correspond to day 4 lead.

159 Forecasts are verified for basin-wide, weekly-accumulated, TCP categories (defined in the fol-
 160 lowing section). Following Camp et al. (2018) and Lee et al. (2020), skill was also investigated
 161 regionally using verification areas which subdivide the global TC basins. These verification boxes
 162 span an area of 20° longitude by 15° latitude and overlap with each other by 5° in longitude and
 163 latitude (see Fig. 1 of Lee et al. 2020).

164 Forecast verification of TC activity, for the analysis of the case studies, is done using the
 165 probability of detection (PoD) and the false alarm ratio (FAR) with TC occurrence as the quantity
 166 to verify (0 or 1) (Coelho et al. 2019). The PoD, also labeled H , is defined as:

$$H = \frac{a}{a + b} \quad (1)$$

167 where a is the number of hits and b the number of false alarms. Similarly, the FAR is defined as:

$$FAR = \frac{b}{a + b}. \quad (2)$$

168 These quantities are oriented such that good forecasts have a high PoD and a low FAR.

169 1) THE RANKED PROBABILITY SKILL SCORE

170 Probabilistic forecasts of basin-wide and regional boxes of total precipitation and TCP are verified
 171 using the Ranked Probability Skill Score (RPSS). The Ranked Probability Score (RPS) is a measure
 172 of forecast reliability and resolution of discrete probability categories (Murphy 1973; Weigel et al.
 173 2007; Tippett 2008). The RPS is defined as the sum of the squared differences between the forecast
 174 probability P_i of a category and the observed (O_i) value, either 0 or 1, across categories, i.e.:

$$RPS = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C (P_{i,j} - O_{i,j})^2 \quad (3)$$

175 where C is the number of categories, N is the number of valid target weeks, i is each forecast time,
 176 and j is each forecast category. The RPSS can be defined using the RPS of a reference forecast
 177 (RPS_{ref}) as:

$$RPSS = 1 - \frac{RPS}{RPS_{\text{ref}}}. \quad (4)$$

178 The RPSS is commonly used in verification of probabilistic subseasonal and seasonal precipita-
179 tion forecasts (Tippett et al. 2014; Vignaud et al. 2017; Klingaman et al. 2021) as well as for various
180 metrics of TC activity (Lee et al. 2020). The RPSS is oriented such that positive values mean that
181 forecasts are skillful compared to the reference forecast.

182 In the case of TCP, 4 categories are used for verification: no precipitation or zero TCP, above-
183 normal, near-normal and below normal. The zero TCP category corresponds to days or weeks
184 where no TCP was diagnosed within a grid-point or basin. The non-zero TCP categories are
185 defined using tercile values from the forecast or observed distribution of non-zero precipitation
186 values (Tippett et al. 2007; Vignaud et al. 2018). In other words, the “above-normal” category
187 corresponds to forecast or observed weeks that exceeded the 66th percentile or second tercile,
188 weeks in the “below-normal” category are defined where the accumulated TCP is lower than the
189 first tercile and the “near-normal” category corresponds to accumulations found within the interval
190 between the 33th and the 66th percentile. The TCP categories differ from the usual equally likely
191 terciles used in the RPSS because the zero TCP category is strongly influenced by TC occurrence
192 probability, which varies by basin and model/observations.

193 The reference climatology is calculated using the “one-year-out” approach (Risbey et al. 2021).
194 This approach calculates forecast and observed tercile values from a climatological distribution for
195 a given target forecast computed using data from every year except data from the target forecast year.
196 In the case of total P, only three, equally likely, categories were used: above-normal, near-normal
197 and below-normal. The total P categories are defined using terciles in a similar manner to the TCP
198 categories except that the zero category is not considered. The reference forecast used for total
199 precipitation verification is the “equal chance” forecast, whereas for TCP, the reference forecast
200 is constructed using the monthly-varying climatological probabilities ($RPSS_m$) or the constant
201 climatology ($RPSS_c$) for each category.

202 2) CALIBRATION

203 Linear regression can improve forecast probabilities by reducing the mean bias while also
204 reducing the conditional bias (Tippett et al. 2014; Lee et al. 2020). Forecasts are calibrated using
205 linear regression-only since García-Franco et al. (2024) found that logistic and a neural network
206 regression render similar results as linear regression. In the case of TCP, each of the four categories

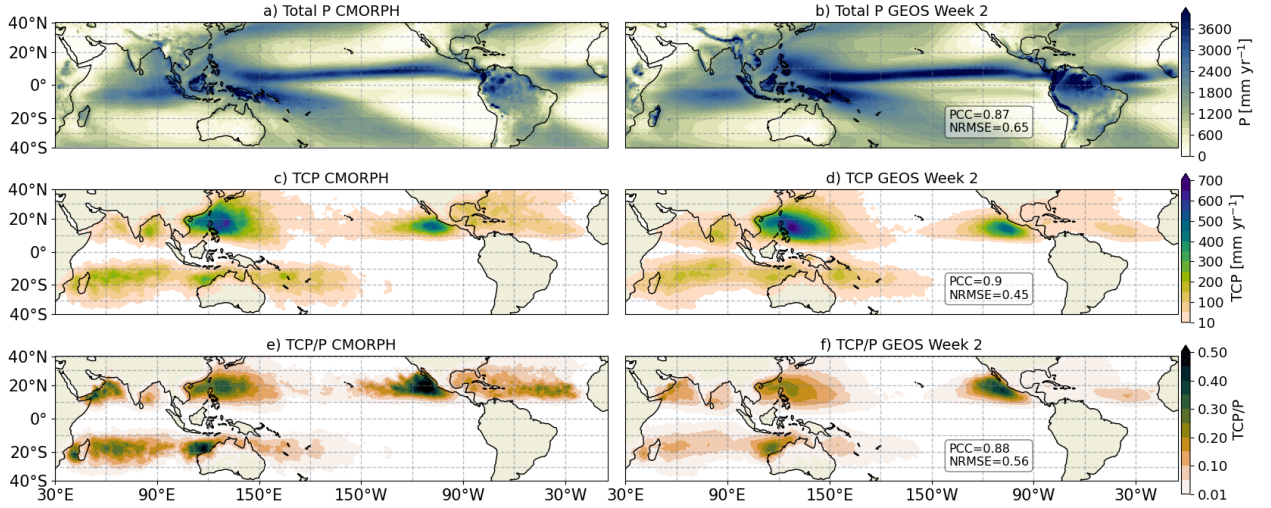


FIG. 1. Climatology of (a, b) total precipitation, (c, d) TCP and (e, f) TCP/P in (a, c, e) observations (CMORPH) and (b, d, f) week-2 GEOS forecasts. The pattern correlation coefficient and normalized root-mean square error (NRMSE) are shown in the GEOS panels.

for the RPSS is calibrated individually. The regression coefficients, slope and intercept, are found by fitting observed versus forecast probabilities individually for each model, basin and lead-time, after which the calibrated probabilities are calculated for each forecast.

3. Results

The climatology of TCP and total P in GEOS is analysed first, prior to the analysis of the skill, because the model evaluation by García-Franco et al. (2024) did not include the spatial climatology of TCP or total P. In contrast, the climatology of TCP and TC activity in the WMO S2S models has been described in previous studies (Lee et al. 2018; García-Franco et al. 2023), which used, for the most part, the same model versions as this study.

a. Total and TC precipitation biases in GEOS

The climatology of TCP and P in GEOS week-2 forecasts, compared to CMORPH, is shown in Figure 1. The forecast climatologies are evaluated against observations using the pattern correlation coefficient (PCC) which measures the spatial correlation between the two maps, and the normalized root-mean square error (NRMSE) which is the standard RMSE normalized by the observed-mean.

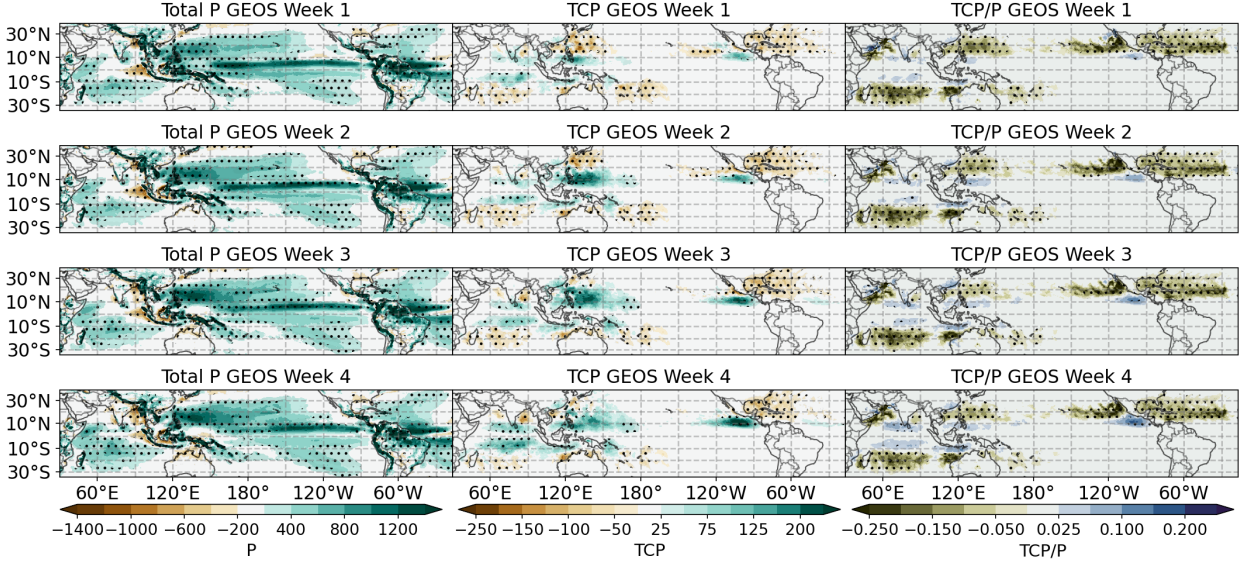


FIG. 2. Biases in (left) total precipitation, (middle) TCP and (right) TCP/P in GEOS reforecasts as a function of lead-time. Stippling indicates regions where the biases are larger than the observed interannual standard deviation.

The total P and TCP patterns ($PCC \sim 0.9$) are very similar to observations. However, GEOS forecasts of total P are wetter in the Pacific and Atlantic intertropical Convergence Zones (ITCZs), northern South America and South Pacific Convergence Zone (SPCZ) compared to CMORPH.

The most noteworthy bias in TCP is the low TCP in the Atlantic (ATL) due to the low TC occurrence bias (see Figure 2 in García-Franco et al. 2024). Based on the two evaluation metrics (PCC and NRMSE), the GEOS model is the best ranked model for the spatial pattern of TCP and the third best model for total P, compared to the WMO S2S model cohort (García-Franco et al. 2023).

One quantity that has been used to describe the interaction between TCP and total P is TCP/P, which measures the contribution of TCs to the total rainfall in a given region (Vannière et al. 2020; García-Franco et al. 2023). The comparison of TCP/P suggests that despite a good representation of TCP in GEOS, the fraction of TCP to total P may be biased. In several regions such as the western north Pacific (WNP) and the western coast of Australia the high bias of total P means that overall TCP/P appears lower in GEOS than in observations.

In Figure 2, the biases in TCP and total P in GEOS are shown as a function of lead-time. Total P biases in GEOS are mostly wet and are significant across most of the tropics. In particular,

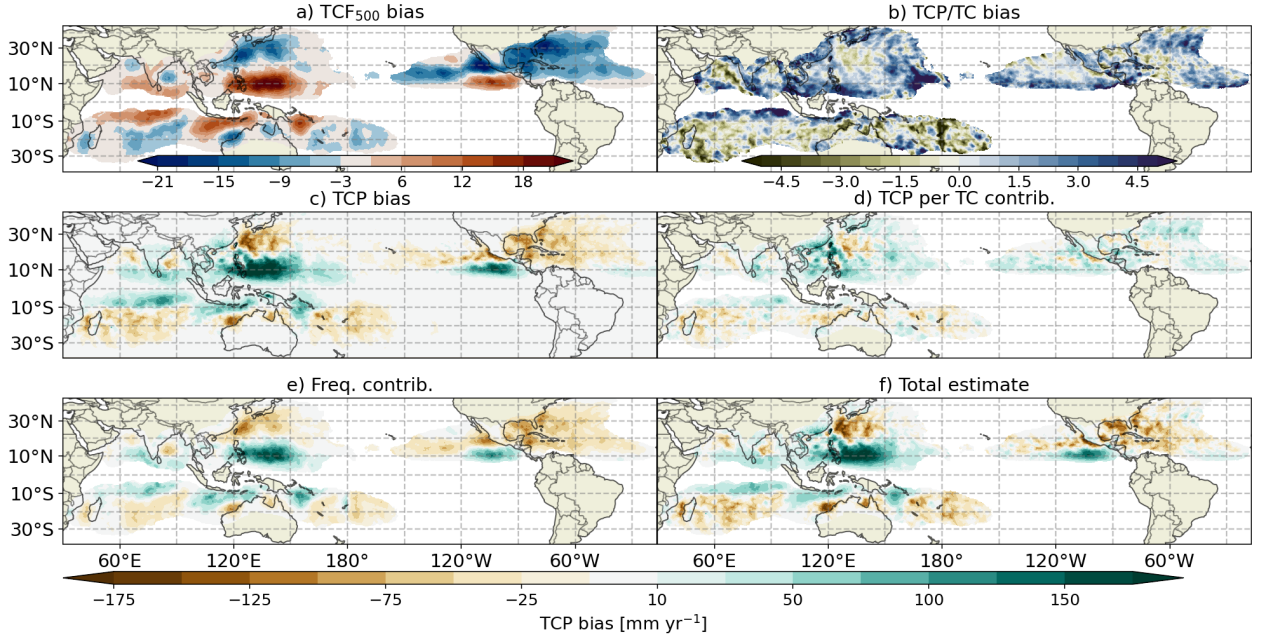


FIG. 3. Decomposition of week-2 TCP biases stemming from biases in a) TC occurrence TCF_{500} [counts yr^{-1}] and b) TCP/TC [$mm\ NTC^{-1}$]. The c) total bias is decomposed into individual contributions (d and e, see text). The total estimate from the sum of d+e is shown in panel f.

the Pacific and Atlantic ITCZs have the largest biases in total P. TCP biases are wet at latitudes equatorward of 20° and dry at higher latitudes, except for the low TCP bias across all of the ATL basin.

The TCP/P biases reveal that TCs contribute less to total rainfall in GEOS than in observations, specially in the southern Indian (SIN) Ocean, ATL and west of the Baja California Peninsula. The spatial pattern of the biases of all 3 quantities vary little with lead-time. The main biases, their patterns, as well as the lack of change with lead-time are similar and comparable to the models from WMO S2S project (García-Franco et al. 2023).

Biases in TCP may stem from biases in the TC occurrence frequency (TCF) or the average precipitation from each storm (Vannière et al. 2020; García-Franco et al. 2023) which separates the processes in the model that control TC activity and those that determine the storm-structure once a TC exists. One can separate these contributions by holding one quantity fixed to observations (e.g. TCP/TCF) and multiplying it by the bias in the other quantity such as the TCP occurrence or

the frequency of TCs striking within 500 km (TCF_{500}). For further details on this calculation, see García-Franco et al. (2023).

Figure 3 shows the decomposition of biases in GEOS into contributions from TC occurrence biases (c, d) and rainfall per TC biases (panels b, e). The results show that the TCP bias pattern is mainly due to the TC occurrence bias. However, the TCP/TC contributes to some extent to the TCP bias magnitude, since the TCP/TC contribution appears relevant in the WP and sparsely in other basins such as SIN and SPC. However, and in agreement with the S2S models, the biases of TCP in GEOS are largely driven by biases in TC occurrence.

In summary, the climatology of TCP and P in GEOS is arguably similar to observations and comparable to the majority of the WMO S2S models. There are several common biases in TCP and total P model biases are common in the WMO and GEOS SS2 models in TCP and total P such as the overall wet biases in total P and the low bias in TCP over the Atlantic.

b. Tropical cyclone prediction skill

This section examines the probabilistic prediction of TCP and total P using case studies, raw and calibrated, and basin-wide and regional forecasts.

1) CASE STUDIES

Case studies are useful to demonstrate the prediction capabilities of current dynamical models (Kolstad 2021). This section examines two well known cases, Hurricanes Katrina (2005) and Harvey (2017), to illustrate how TC activity forecasts relate to TCP and total P predictions. The two forecast systems, ECMWF for the Harvey case and GEOS for the Katrina case, differ in several aspects, including ensemble size (Table 1), resolution, and physical parameterizations.

Figures 4 and 5 show the observed track and precipitation fields the prediction of TC occurrence, TCP and total precipitation for two cases, Hurricane Harvey (2017; Fig. 4) and Hurricane Katrina (2005; Fig. 5). The FAR and PoD metrics, shown in each panel, were computed using an 800 km radius. This value is a slightly looser threshold compared to previous studies employing these metrics (Halperin et al. 2013; Yamaguchi and Koide 2017) but those studies typically limited their analysis to 15-day lead-time forecasts, whereas our work extends beyond that range. The 800 km radius for verifying TC strike probability (FAR) differs from the 500 km radius used to estimate

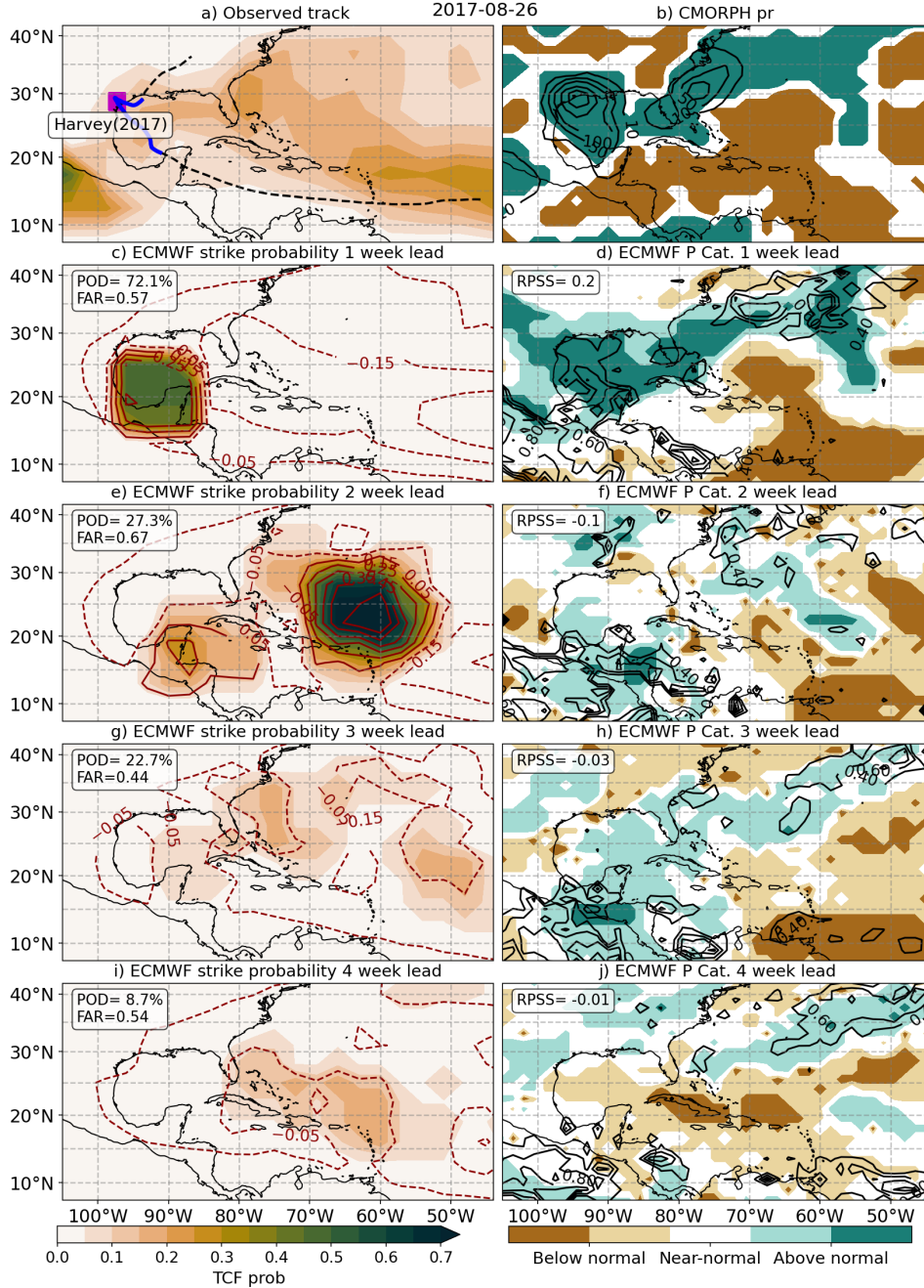


FIG. 4. (a-b) Observations during the week of August 22-28, 2017 for (a) TC tracks and (b) total precipitation categories (Above-normal, below-normal or near-normal) compared to climatology. The TC occurrence probability climatology is shown in a) (shading) and the observed TCP is shown in b) (contours in mm week^{-1}). (c-j) ECMWF forecasts for the same target week showing the (left) strike probability (shading) and anomaly from climatology (contours) and (right) total precipitation categorical forecast (shading) and the probability of above average TCP (contours). The probability of detection of Hurricane Harvey is shown for each forecast week on the left panels and the spatial RPSS between forecast and observation is shown on the right panel.

294 TCP and the $20^{\circ} \times 15^{\circ}$ verification area, because the former measures the probability of a TC being
295 in the vicinity at any point, while the latter is employed to calculate accumulated rainfall within a
296 regional box.

297 Figure 4 shows that the track of Hurricane Harvey is collocated with the category of “above-
298 normal” total P and large weekly accumulations of TCP (up to 200 mm). The forecasts from
299 ECMWF in week 1 suggest a high strike probability in the Gulf of Mexico in addition to high
300 probabilities of above average TCP collocated with categorical total P forecasts for above-normal
301 total P and high TC occurrence strike probabilities.

302 Beyond week 1, ECMWF forecasts show a low strike probability for a Harvey-like track (PoD
303 below 30% in all cases). Similarly, the probabilistic forecasts for TCP and total P around the landfall
304 area and the wider Gulf of Mexico were calling both TCP and total P to be in the “below-normal”
305 category. For this reason, the spatial RPSS value decreases notably from week 1 to weeks 2 and 3.
306 This is an example of forecasts where total P and TCP forecasts were poor beyond week 1, since
307 the lack of a TC in the forecasts negatively impacted the precipitation forecasts.

308 Figure 5 similarly shows that the track of Hurricane Katrina is collocated with “above-normal”
309 total P in the observations. At lead-times of weeks 1 and 3, GEOS forecasts have a high strike
310 probability for a TC near the track of Katrina. Along with these TC strike forecasts, categorical
311 precipitation forecasts in the Caribbean or the Gulf of Mexico called for an “above-normal” week
312 of TCP and total P. In contrast, for weeks 2 and 4, the strike probability decreases significantly,
313 suggesting a below-normal probability of TC occurrence. The spatial RPSS is highly related to
314 whether the forecasts were correct in predicting the existence of a Katrina-like track, since week 1
315 and week 3 RPSS values are higher than the weeks 2 and 4 scores.

316 The GEOS week 3 strike probability and precipitation forecasts appear to outperform those of
317 week 2. This improvement may be influenced by intraseasonal oscillations present at the time of
318 forecast initialization. For example, García-Franco et al. (2024) demonstrated that prediction skill
319 in both GEOS and ECMWF can vary based on the phase or amplitude of the MJO at initialization
320 and Week 3 forecasts were initialized under strong MJO conditions in phases 6-7, whereas Week
321 2 forecasts were initialized in phases 1-2 which may have less predictability.

322 Despite a qualitatively good forecast of Katrina, GEOS week 3 precipitation forecasts show
323 negative RPSS values. This results highlights how errors elsewhere of the Katrina track, e.g., in

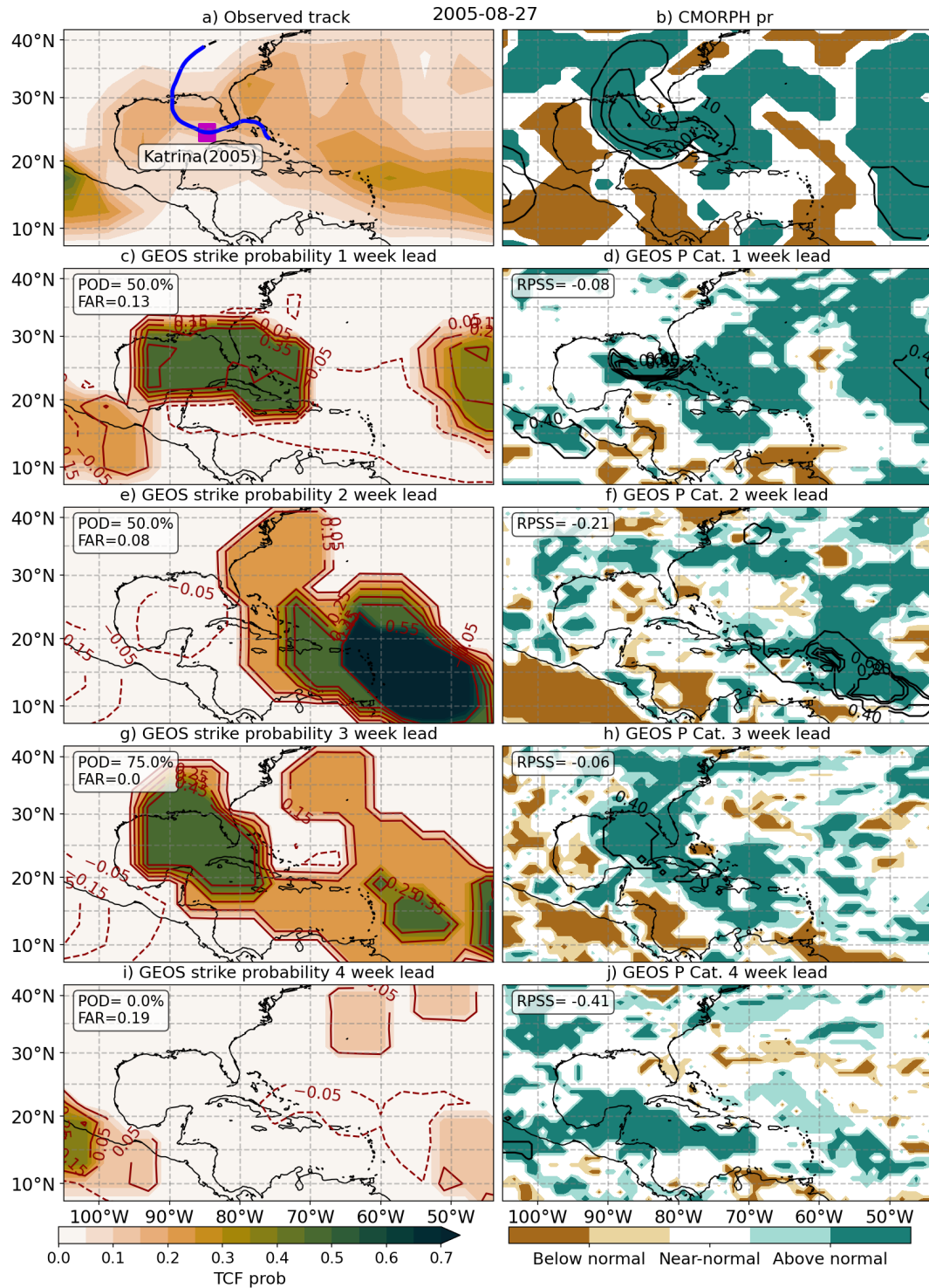


FIG. 5. As in Fig. 4, but for the case of Hurricane Katrina (2005) using GEOS forecasts.

the North Atlantic in Week-3, are penalized by the RPSS. The RPSS is decreased by biases and reliability issues which are affected by a low ensemble size as is the case in GEOS. This analysis,

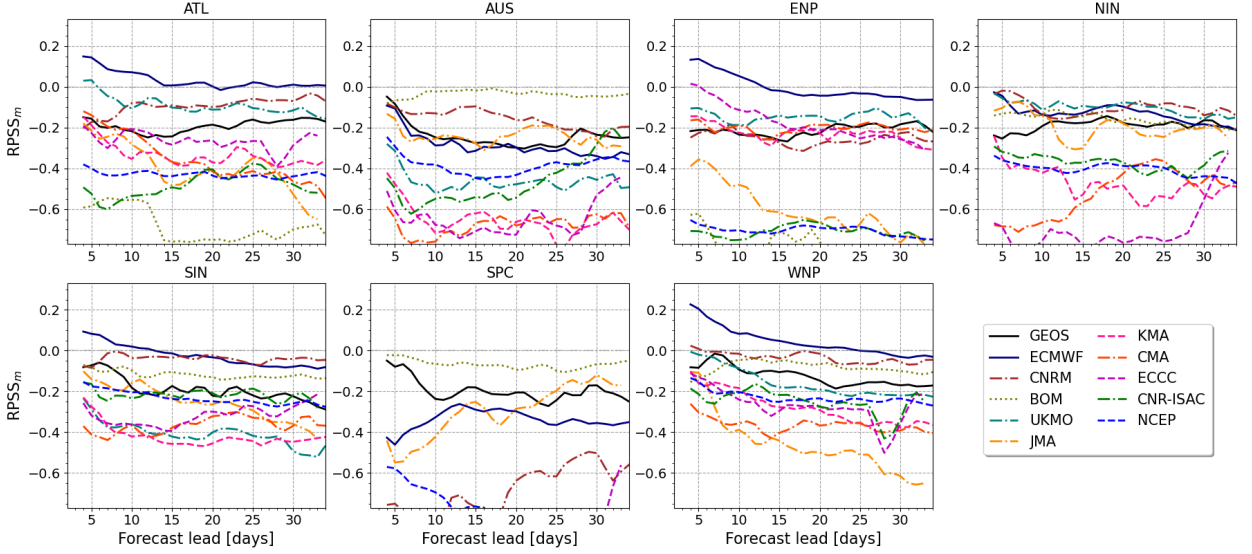


FIG. 6. Basin-wide TCP RPSS using a monthly-varying reference climatology ($RPSS_m$).

however, is limited to just two cases raising the need for a systematic basin-wide evaluation of TCP and total P.

2) BASIN-WIDE TCP SKILL

The previous section illustrated how, in some cases, predicting a TC correctly or incorrectly can greatly influence the total P categorical forecasts. In this section, results from the verification of the complete record of reforecasts of TCP are shown using basin-wide weekly accumulated estimates of TCP with 4 categories, measured by the RPSS.

Figure 6 shows the $RPSS_m$, which is obtained using the monthly-varying reference climatology, as a function of lead-time. As expected, skill at predicting TCP decreases with lead-time but more importantly, most forecast systems are not skillful at predicting TCP at any lead time. Only ECMWF is skillful in the ATL, ENP, SIN, and WNP at lead-times of 4-15 days. Beyond week 2, or day 15, no model is skillful at any lead-time.

In the SPC, the AUS or the North Indian (NIN) Ocean basins, no model is skillful at any lead-times, not even at the shortest of lead-times. These results largely agree with similar weekly-based skills scores for the Accumulated Cyclone Energy (ACE) analyzed by Lee et al. (2020). The implication of this result is that raw dynamical forecasts of TCP and ACE at subseasonal lead-times (2-6 weeks) have no skill.

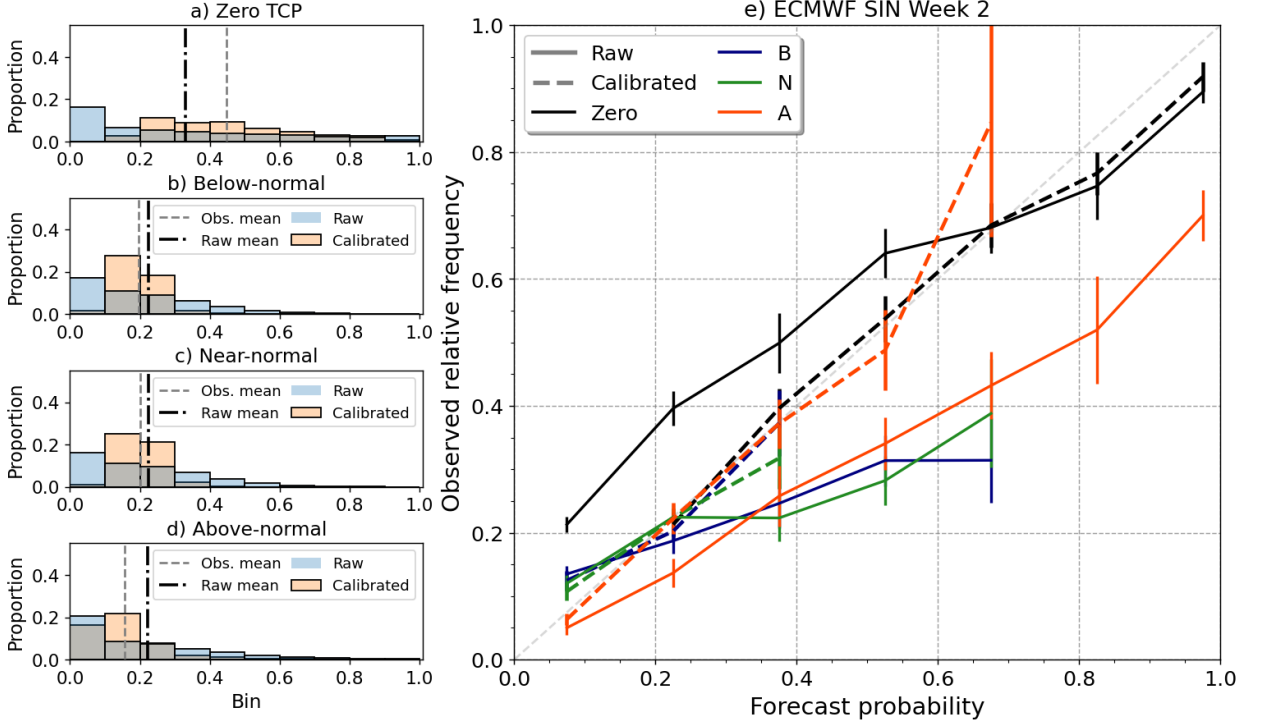


FIG. 7. (a-d) Histogram of proportion of ECMWF week-2 forecast probabilities for the categories of (a) zero, (b) below-normal, (c) near-normal and (d) above-normal TCP in the SIN. (e) Reliability diagrams of raw and calibrated TCP forecasts in the SIN basin for Raw and Calibrated forecasts split by the categories of Zero TCP, Below-normal (B), Near-normal (N) and above-normal (A). The error bars in e) indicates the standard error of the observed relative frequency estimate.

3) THE IMPACT OF CALIBRATION ON TCP FORECASTS

Biases in the forecast distribution can decrease skill scores, such as the RPSS, but calibrating forecasts may improve skill by fixing conditional and mean biases (Moron and Robertson 2020; Lee et al. 2020; García-Franco et al. 2024). Due to the fact that TCP was verified using 4 not equally likely categories that were defined using the forecast distributions, models exhibit biases in the forecast probability of all categories. One way to illustrate biases in the forecast distribution is the reliability diagram which simultaneously shows the need for calibration, its impact and how it improves forecasts. Figure 7 shows the forecast distributions for each category of TCP, as well as the corresponding reliability diagram, for a select case of the ECMWF forecasts in the SIN basin.

Figure 7a shows that the zero TCP occurrence probability is biased low in the model, meaning the forecasts predict fewer zero-TCP cases than observed. This bias in TC occurrence penalizes

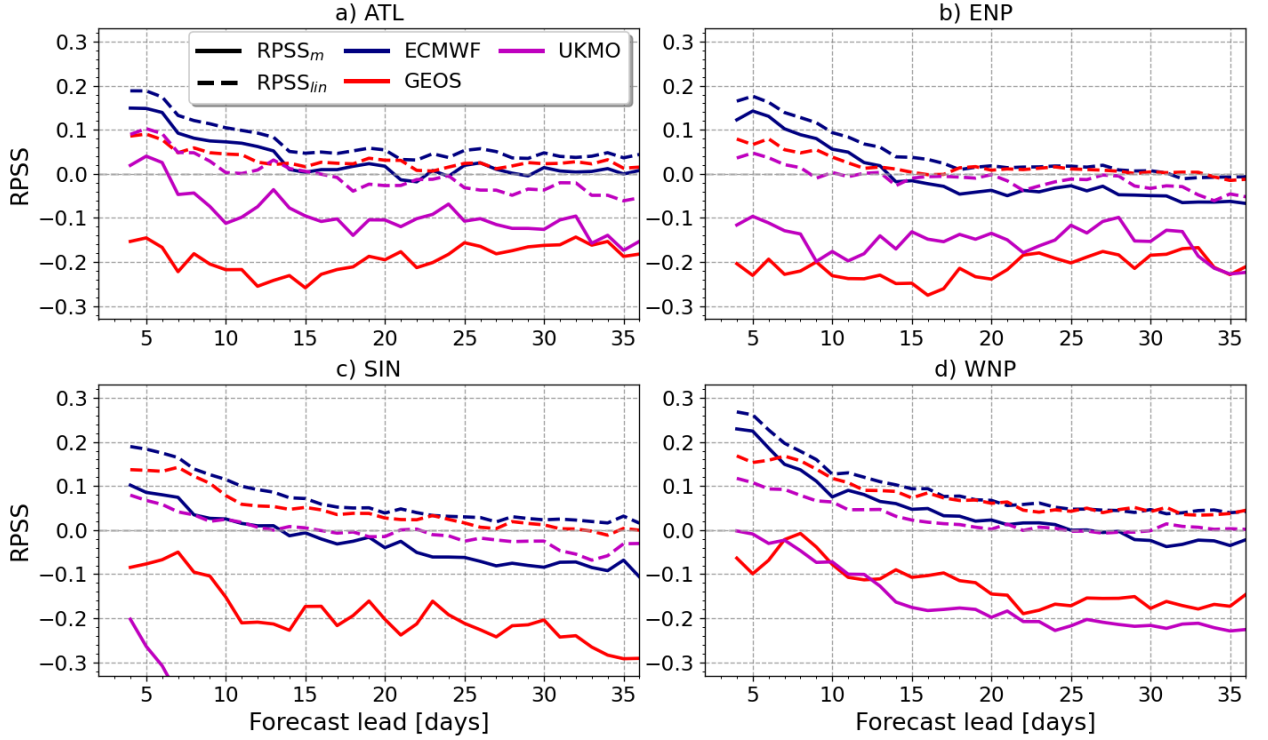


FIG. 8. As in Figure 6 but showing the impact of calibration on the RPSS through the $RPSS_{lin}$ for GEOS, UKMO and ECMWF forecasts in the (a) ATL, (b) ENP, (c) SIN and (d) WNP basins.

the RPSS. After linear regression calibration, the forecast distributions align more closely with the observed mean probability, as expected by design. The non-zero TCP categories (above, below, and near-normal) are closer to observations in the raw forecasts than the zero category, but the calibrated distributions are narrower and better centered around the observed mean. This narrowing reflects the correction of both conditional and unconditional biases.

The reliability diagram indicates that raw forecasts are overconfident for non-zero TCP categories but underconfident for the zero category, underestimating weeks with no TCP and overestimating the other categories. Linear calibration improves reliability, notably reducing mean and conditional biases. In the above-average category, the model becomes slightly underconfident at high probabilities but overall shows improved performance.

Calibration minimizes the Brier Score (BS), reducing mean square probability errors for each category (Bradley and Schwartz 2011). Consequently, the RPSS values increase with calibrated forecasts ($RPSS_{lin}$) compared to raw forecasts ($RPSS_m$).

374 Figure 8 shows that the calibration of the forecasts leads to better skill in all cases. The method
375 improves GEOS forecasts more than ECMWF since GEOS has larger biases (such as TC mean
376 occurrence biases (García-Franco et al. 2024)) and, therefore, there is more room for improvement
377 for the RPSS values. After calibration, forecasts are skillful in all basins in the 5-15 day lead time
378 range.

379 Although GEOS raw forecasts had the lowest RPSS out of the three models considered in the
380 ATL, the RPSS of bias-corrected GEOS forecasts is higher than that of the UKMO bias-corrected
381 forecasts. This result illustrates that the impact of calibration is highly dependent on the nature
382 of the biases in the forecast considered. In the WNP, forecast calibration appears to extend the
383 prediction skill of TCP to week 3 and beyond for GEOS and ECMWF, and to at least week 2
384 for the UKMO forecasts. This result is consistent with previous findings that calibration extends
385 prediction skill for TC occurrence to week 3 in this basin and these models (García-Franco et al.
386 2024).

390 Figure 9 shows that the calibration improves the forecast skill for all categories across all basins.
391 In addition, to diagnose the contribution from each category to the RPSS, the Brier Skill Score
392 (BSS) is used to measure skill in each individual category, knowing that the RPS is the average
393 BS (Bradley and Schwartz 2011). These results show that some categories contribute more than
394 others to the overall improvement of the RPSS. The zero TCP and above-normal categories have
395 generally higher scores than the other categories. This means that correcting the forecast mean
396 probability and making reliable categorical forecasts explain the relatively large improvements in
397 RPSS (Fig. 8). The skill mostly arises from periods with copious TCP (above-normal) or no TCP,
398 which suggests that the prediction of TC occurrence is key for a good prediction of TCP categories.

403 In order to show the impact of calibration beyond basin-wide estimates, Figure 10 shows the
404 RPSS calculated using the regional boxes and comparing raw versus calibrated forecasts. Based
405 on raw forecasts, the ECMWF system is noticeably more skillful than GEOS, except in the AUS
406 and NIN basins where neither model is skillful.

407 Figure 10 also shows that the calibration increases skill globally and for both models. The
408 ECMWF forecasts are skillful globally, with particularly large skill in the ENP, SIN and SPC. In
409 contrast, GEOS, even when calibrated, shows little skill in the ENP and ATL basins. Nevertheless,
410 the improvement from the raw forecasts is also notable in GEOS.

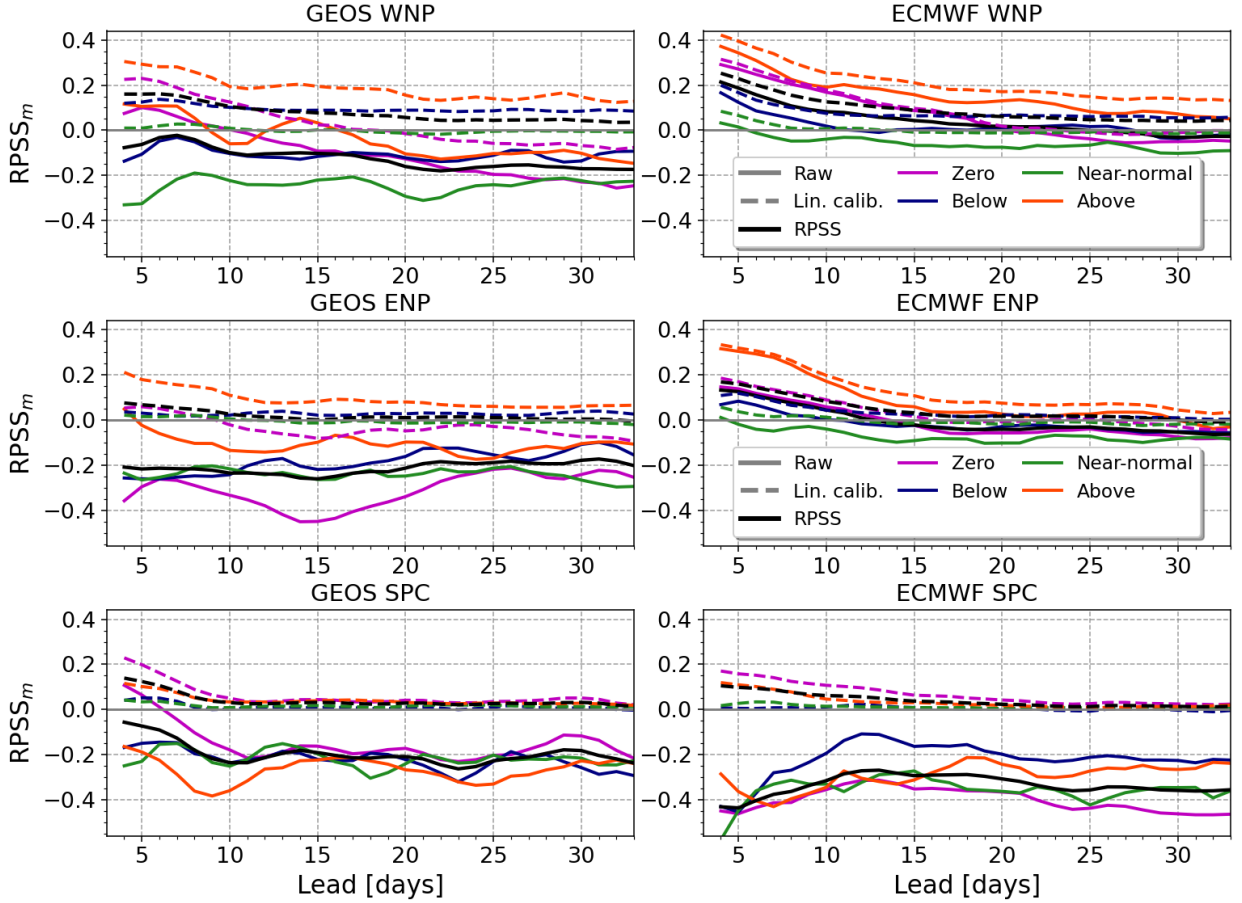


FIG. 9. RPSS (black) and Brier Skill Score (BSS) values of the categorical forecasts for the zero TCP (magenta), below-normal (blue), near-normal (green) and above-normal (red) categories. Raw forecasts are shown as solid lines and calibrated forecasts as dashed lines.

The differences in skill between models, raw or calibrated, may stem from different biases in the forecast distribution which can be due to different biases in TC activity (Lee et al. 2020; Li et al. 2022; García-Franco et al. 2024). Another reason for the differences in skill could be due to the smaller ensemble size which can hamper forecast skill in probabilistic forecasts regardless of model biases (García-Franco et al. 2024).

c. The impact of TCP on total precipitation forecast skill

This study has so far showed that S2S models struggle to forecast TCP. It is also true that TCs can contribute a large fraction of total precipitation (Fig. 1) and are an important source of total

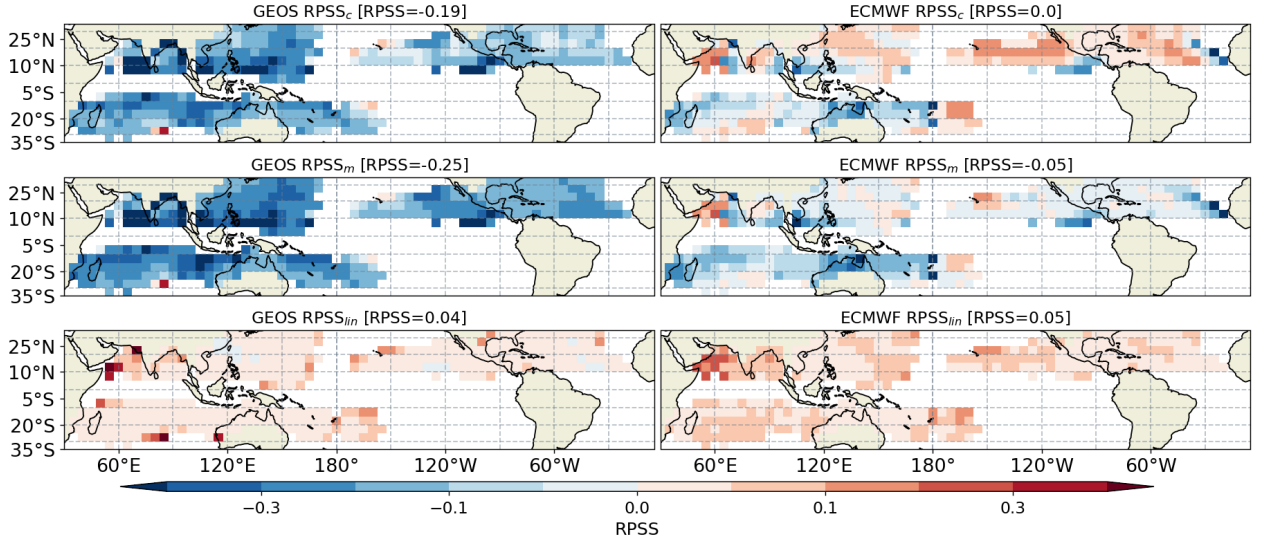


FIG. 10. Regional skill of TCP using RPSS. The RPSS was calculated using week 2 forecasts verified in boxes spanning 15° in latitude $\times$ 20° in longitude separated by a 5° buffer. Results are shown for (a, c, e) GEOS and (b, d, f) ECMWF for the $RPSS_c$, $RPSS_m$ and the $RPSS_{lin}$. Boxes were verified if, in observations, at least 80% of years there is a TC pass in a 500 km radius of the center of the box.

P model biases (García-Franco et al. 2023). One could reasonably wonder the extent to which forecast errors in TCP are related to total P errors.

For example, Figure 11 shows a qq-plot of precipitation distributions in the WNP, which compares two distributions by matching sets of quantile values in each distribution. This means the 5th percentile of the observed distribution is matched with the 5th percentile of the forecast distribution of total P, TCP and non-TCP. In this case, bins of 5 percentiles were used. Biases in the distribution of the weekly-accumulated total P (All), TCP and non-TCP show whether forecast errors are distinct depending on the source or process causing the precipitation. Note that the quantiles of total P (All) are not a simple sum or weighted average of TCP and non-TCP quantiles since these depend on the joint distribution of the two components and their relative contributions across percentiles.

The biases in non-TCP quantiles differ significantly from TCP biases, as shown in Fig. 11. Firstly, the slopes of the TCP and non-TCP curves differ, reflecting key differences in their forecast distributions. For example, in ECMWF, CNRM, and UKMO, the forecasted TCP percentile values are consistently lower than observed at high percentiles, indicating an underestimation of extreme TCP values. In contrast, GEOS overestimates TCP across most percentiles, with particularly large

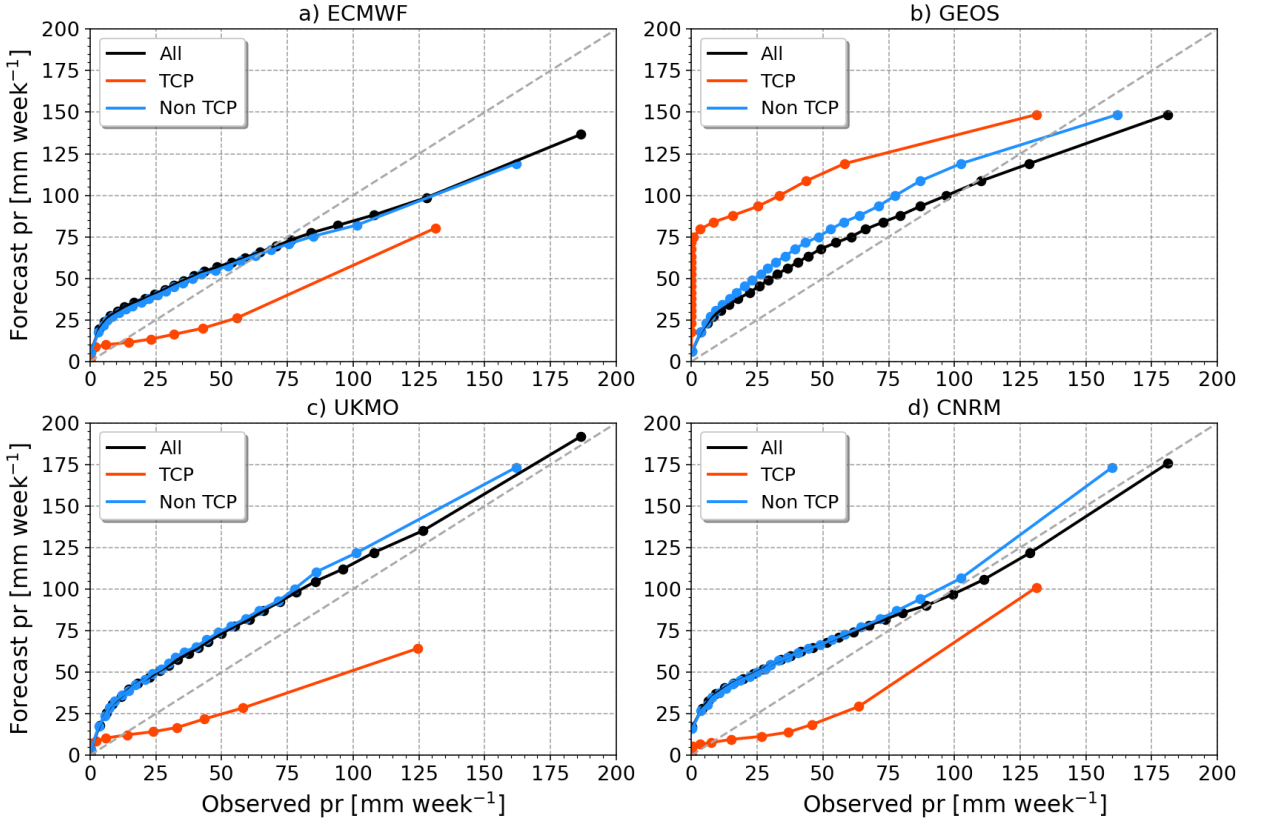


FIG. 11. Quantile-quantile (QQ) plot of the observed versus forecast distributions of total P, non-TCP and TCP in four S2S models. The scatter is done by splitting the distributions into 5th percentile bins. Results are shown for a grid-point [115°E, 15°N] in the WNP basin where the frequency of TCP in observations is higher than 12 days per year.

overestimations at low percentiles, which are much greater than those for non-TCP, confirming that GEOS overestimates TCP in the WNP (see also Fig. 2).

The TCP contribution to the total precipitation (Total P) distribution is more pronounced at higher percentiles. This contribution, measured as the difference between the Total and non-TCP curves, increases with percentile and can reach as much as 30 mm week⁻¹. In other words, Fig. 11 suggests that TCP is a relevant contributor to the high end of the Total P distribution while also illustrating how forecast errors differ for the TCP and non-TCP portions of precipitation. This result makes the case for separate calibration of TCP and non-TCP to improve Total P predictions, particularly for higher-than-average precipitation categories that drive total RPSS (e.g. Tippett et al. 2015), especially in TC-active regions.

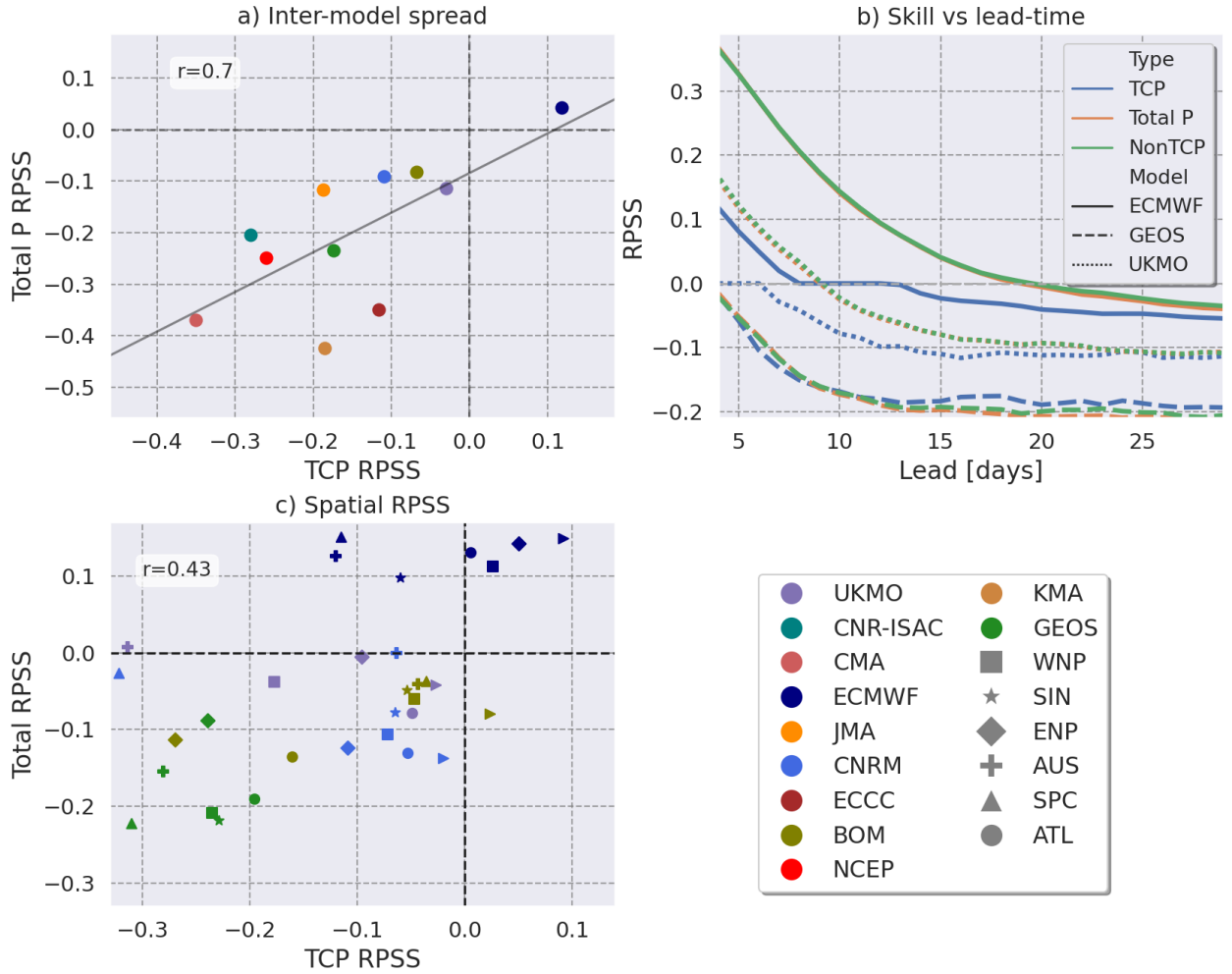


FIG. 12. (a) Scatter plot of model-mean of total P versus TCP RPSS using week-2 forecasts and averaged across tropical regions. b) shows the RPSS scores for total P, TCP and non-TCP as a function of lead-time. c) shows the basin-averaged skill, total P versus TCP RPSS, for select models. In all cases, only grid-points where the observed mean occurrence rate of TCP is at least 4 times a year were used in the calculations.

Figure 12 illustrates the relationship between TCP and total P skill. The average of RPSS values in the tropics for TCP and total P in the 11 S2S models reveals a linear and positive relationship between the forecast skill of TCP and that of total P (Fig. 12a). This means that the models with higher skill in TCP are also the models with higher skill at predicting total P. The model with the highest TCP skill (ECMWF) is also the model with highest skill in total P. The lowest scores for TCP, however, found in the CMA model do not correspond to the model with the lowest total P RPSS.

Models that are skillful at predicting TCP, e.g. ECMWF and UKMO, find it harder to predict TCP than total P and non-TCP (Figure 12b). At short lead-times, total and non-TCP precipitation appear skillful in the ECMWF model up to 15 days, but TCP is only marginally skillful 4-5 days after initialization. The non-TCP skill tracks very closely with the total P skill, indicating that predicting non-TCP is more feasible for these forecast systems than TCP. This means that the TCP portion of the total P is harder to predict and fixing TCP biases could have relevant consequences not just for the prediction of TCP but also for the prediction of total P. However, in models with generally low predictive skill, such as GEOS, there appears to be no clear relationship between the ability to predict TCP and the skill in predicting total P.

Figure 12c shows the spatial relationship between TCP and total P RPSS values, using the top 5 most skillful models at TCP. The scatter plots illustrates whether basins with better prediction skill for TCP are also the basins with a better prediction skill for total P. These results suggest that there is a positive relationship between these quantities, but the strength of the relationship is weaker than in the inter-model spread. To some extent, basins with higher prediction skill of TCP are also basins with a higher prediction skill of total P but in individual models this relationship can be weak (e.g. GEOS), non-existent (ECMWF) or negative (CNRM).

4. Summary and discussion

This study analyzed the prediction skill of rainfall associated with TCs at the subseasonal range. Forecast models, such as GEOS, reasonably reproduce the observed climatology of total P, TCP and TCP/P despite some basin and model-dependent biases. For example, GEOS has noteworthy dry biases of TCP in the ATL due to the low TC occurrence bias when compared to observations (García-Franco et al. 2024). The TC frequency bias is the main factor driving the TCP biases in these models (Fig. 3 and Figure 13 from García-Franco et al. 2023).

Forecast verification, through the probabilistic score of RPSS, showed that only the ECMWF model is skillful at predicting TCP. ECMWF is skillful up to day 15 in the ATL, WNP, ENP and SIN basins. This means that none of the forecast systems analysed in this study are able to provide reliable information about TCP more than 2 weeks in advance based on their raw forecasts. A similar result was found by Lee et al. (2020) regarding ACE predictions.

487 Calibrating forecasts is necessary to minimize forecast errors. In particular, linear regression
488 fixes several biases in the forecast probability (Tippett et al. 2014; García-Franco et al. 2024). In the
489 case of the categorical TCP forecasts, calibration fixes the conditional and unconditional biases in
490 the probability of TCP occurring (the zero-TCP category) thereby fixing the mean TC occurrence
491 rate (Lee et al. 2020). After calibration, three models, GEOS, UKMO and CNRM, provide skillful
492 forecasts in most basins up to day 15, and in the case of ECMWF, positive skill is extended to week
493 3. Moreover, at the regional scale, the calibration is also able to make ECMWF forecasts skillful.

494 The decomposition of TCP RPSS into the BSS of individual categories shows that calibration
495 improves all categories, especially the above-average and zero TCP categories.. Since the best
496 possible BSS for a forecast is the squared correlation between forecast probability and occurrence,
497 our calibration results show that some model (UKMO, ECMWF) probabilities in specific basins
498 (GEOS in the WNP) are well correlated with observations. This means that there are linear relations
499 between forecast probability and observed occurrence of TCP, but model biases in the occurrence
500 of TCP are greatly hampering skill. Therefore, biases in TC occurrence are a key limitation for
501 the prediction skill of TCP which emphasizes the impact of calibration in probabilistic predictions
502 such that fixing the probabilities for the mean occurrence rate of TCP aids overall forecast skill.

503 Even calibrated forecasts of several models in select basins are not skillful at any lead time. Such
504 is the case of ATL TCP predictions from the GEOS forecast system, likely due to its very low
505 number of ATL storms (García-Franco et al. 2024). Models with a low ensemble size (GEOS,
506 KMA, CMA, JMA, ECCC) may struggle to provide skillful probabilistic predictions given their
507 small ensemble size simply because scores such as the RPSS are penalized for low ensemble
508 sizes (Richardson 2001; Tippett et al. 2007). The growing number of models, and growing
509 ensemble sizes per model, participating in efforts such as the S2S project provides the opportunity
510 to investigate how a multi-model ensemble may perform at predicting total P and TCP, especially
511 after calibration.

512 The results of this study also pointed to the potential role that forecast errors in TCP may play in
513 the forecast skill of total P. The analysis of forecast distributions and forecast skill of TCP, non-TCP
514 and total P shows that errors in predicting TCP are distinct from errors in non-TCP. Moreover,
515 models that are better at predicting TCP are also better at predicting total P. In individual models,
516 the relationship between TCP skill and total P skill appears weak when considering whether basins

517 with better TCP skill are also basins where models are better at predicting total P. One could
518 still argue, nonetheless, that if TCP forecasts are improved, regions where TC activity is a major
519 influence on precipitation may be benefited from increased forecast skill in total P.

520 The fact that TCP and non-TCP forecast errors appear distinct opens up the question of whether
521 a regression-based calibrated forecast that accounts for TCP and non-TCP biases separately, may
522 provide a significant improvement of forecast skill. In particular, this methodology may be most
523 useful in regions and season where TCP is a main factor modulating total P.

524 Further work should also examine the predictability of TCP and total P driven by the MJO or
525 other sources of S2S predictability. This study has mainly analyzed TCP skill, but estimates of
526 predictability may provide further clarity as to which processes link TCP and total P at the S2S
527 range, as well as manage our expectations of TCP skill.

528 *Acknowledgments.* This study was funded by the NASA MAP program (80NSSC21K1495).
529 JLGF was supported by Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica
530 (PAPIIT) grant number IA 101024. DK was supported by the New Faculty Startup Fund and
531 Creative-Pioneering Researchers Program from Seoul National University, the National Research
532 Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2024-00336160),
533 NOAA MAPP program (NA21OAR4310343), NOAA CVP program (NA22OAR4310608),
534 and KMA R&D program (KMI2022-01313). SJC acknowledges funding by NOAA CVP
535 (NA22OAR4310610).

536 *Data availability statement.* The S2S data is publicly available at [https://apps.ecmwf.](https://apps.ecmwf.int/datasets/data/)
537 [int/datasets/data/](https://apps.ecmwf.int/datasets/data/). S2S data and S2S TC tracks are available to research community
538 at <http://s2sprediction.net>. The gridded precipitation datasets are available from the
539 FROGS database at <ftp://ftp.climserv.ipsl.polytechnique.fr/FROGS/>. The IBTrACS
540 dataset is publicly available (Knapp et al. 2018) at [https://www.ncei.noaa.gov/products/](https://www.ncei.noaa.gov/products/international-best-track-archive)
541 [international-best-track-archive](https://www.ncei.noaa.gov/products/international-best-track-archive).

References

- Beatty, T. K. M., J. P. Shimshack, and R. J. Volpe, 2019: Disaster Preparedness and Disaster Response: Evidence from Sales of Emergency Supplies Before and After Hurricanes. *Journal of the Association of Environmental and Resource Economists*, **6** (4), 633–668, <https://doi.org/10.1086/703379>, URL <https://www.journals.uchicago.edu/doi/10.1086/703379>.
- Black, R., N. W. Arnell, W. N. Adger, D. Thomas, and A. Geddes, 2013: Migration, immobility and displacement outcomes following extreme events. *Environmental Science & Policy*, **27**, S32–S43.
- Bradley, A. A., and S. S. Schwartz, 2011: Summary verification measures and their interpretation for ensemble forecasts. *Monthly weather review*, **139** (9), 3075–3089.
- Camargo, S. J., F. Vitart, C.-Y. Lee, and M. K. Tippett, 2021: Skill, Predictability, and Cluster Analysis of Atlantic Tropical Storms and Hurricanes in the ECMWF Monthly Forecasts. *Monthly Weather Review*, **149** (11), 3781–3802, <https://doi.org/10.1175/MWR-D-21-0075.1>, URL <https://journals.ametsoc.org/view/journals/mwre/149/11/MWR-D-21-0075.1.xml>.
- Camargo, S. J., and Coauthors, 2019: Tropical Cyclone Prediction on Subseasonal Time-Scales. *Tropical Cyclone Research and Review*, **8** (3), 150–165, <https://doi.org/10.1016/j.tcrr.2019.10.004>, URL <https://linkinghub.elsevier.com/retrieve/pii/S2225603219300712>.
- Camp, J., P. Gregory, A. Marshall, J. Greenslade, and M. Wheeler, 2023: Multiweek tropical cyclone prediction for the southern hemisphere in access-s2: Maintaining operational skill and continuity of service. *Quarterly Journal of the Royal Meteorological Society*, **149** (757), 3401–3422.
- Camp, J., and Coauthors, 2018: Skilful multiweek tropical cyclone prediction in ACCESS-S1 and the role of the MJO. *Quarterly Journal of the Royal Meteorological Society*, **144** (714), 1337–1351, <https://doi.org/10.1002/qj.3260>, URL <https://onlinelibrary.wiley.com/doi/10.1002/qj.3260>.
- Cangialosi, J. P., E. Blake, M. DeMaria, A. Penny, A. Latta, E. Rappaport, and V. Tallapragada, 2020: Recent progress in tropical cyclone intensity forecasting at the national hurricane center. *Weather and Forecasting*, **35** (5), 1913–1922.

- Coelho, C. A., B. Brown, L. Wilson, M. Mittermaier, and B. Casati, 2019: Forecast verification for s2s timescales. *Sub-Seasonal to Seasonal Prediction*, Elsevier, 337–361.
- De Andrade, F. M., C. A. S. Coelho, and I. F. A. Cavalcanti, 2019: Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Climate Dynamics*, **52** (9-10), 5451–5475, <https://doi.org/10.1007/s00382-018-4457-z>, URL <http://link.springer.com/10.1007/s00382-018-4457-z>.
- De Andrade, F. M., M. P. Young, D. MacLeod, L. C. Hirons, S. J. Woolnough, and E. Black, 2021: Subseasonal Precipitation Prediction for Africa: Forecast Evaluation and Sources of Predictability. *Weather and Forecasting*, **36** (1), 265–284, <https://doi.org/10.1175/WAF-D-20-0054.1>, URL <https://journals.ametsoc.org/view/journals/wefo/36/1/WAF-D-20-0054.1.xml>.
- Domeisen, D. I., and Coauthors, 2022: Advances in the subseasonal prediction of extreme events: Relevant case studies across the globe. *Bulletin of the American Meteorological Society*.
- Dominguez, C., A. Jaramillo, and P. Cuéllar, 2021: Are the socioeconomic impacts associated with tropical cyclones in Mexico exacerbated by local vulnerability and ENSO conditions? *International Journal of Climatology*, **41** (S1), <https://doi.org/10.1002/joc.6927>, URL <https://onlinelibrary.wiley.com/doi/10.1002/joc.6927>.
- García-Franco, J. L., C.-Y. Lee, S. J. Camargo, M. K. Tippett, G. N. Emlaw, D. Kim, Y.-K. Lim, and A. Molod, 2024: Tropical cyclones in the GEOS-S2S-2 subseasonal forecasts. *Weather and Forecasting*, **39** (9), 1297 – 1318, <https://doi.org/10.1175/WAF-D-23-0208.1>, URL <https://journals.ametsoc.org/view/journals/wefo/39/9/WAF-D-23-0208.1.xml>.
- García-Franco, J. L., C.-Y. Lee, S. J. Camargo, M. K. Tippett, D. Kim, A. Molod, and Y.-K. Lim, 2023: Climatology of Tropical Cyclone Precipitation in the S2S Models. *Weather and Forecasting*, **38** (9), 1759–1776, <https://doi.org/10.1175/WAF-D-23-0029.1>, URL <https://journals.ametsoc.org/view/journals/wefo/38/9/WAF-D-23-0029.1.xml>.
- Gregory, P. A., J. Camp, K. Bigelow, and A. Brown, 2019: Sub-seasonal predictability of the 2017–2018 Southern Hemisphere tropical cyclone season. *Atmospheric Science Letters*, **20** (4), e886, <https://doi.org/10.1002/asl.886>, URL <https://onlinelibrary.wiley.com/doi/10.1002/asl.886>.

- 597 Griffies, S. M., and Coauthors, 2012: Elements of the modular ocean model (MOM). *GFDL Ocean*
598 *Group Tech. Rep.*, **7 (620)**, 47.
- 599 Halperin, D. J., H. E. Fuelberg, R. E. Hart, J. H. Cossuth, P. Sura, and R. J. Pasch, 2013: An
600 evaluation of tropical cyclone genesis forecasts from global numerical models. *Weather and*
601 *Forecasting*, **28 (6)**, 1423–1445.
- 602 Heslin, A., N. D. Deckard, R. Oakes, and A. Montero-Colbert, 2019: Displacement and resettle-
603 ment: understanding the role of climate change in contemporary migration. *Loss and damage*
604 *from climate change: Concepts, methods and policy options*, 237–258.
- 605 Hu, P., Q. Zhang, P. Shi, B. Chen, and J. Fang, 2018: Flood-induced mortality across the globe:
606 Spatiotemporal pattern and influencing factors. *Science of the Total Environment*, **643**, 171–182.
- 607 Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock,
608 and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite
609 observations. *Journal of hydrometeorology*, **2 (1)**, 36–50.
- 610 Khouakhi, A., G. Villarini, and G. A. Vecchi, 2017: Contribution of Tropical Cyclones to Rainfall at
611 the Global Scale. *Journal of Climate*, **30 (1)**, 359–372, <https://doi.org/10.1175/JCLI-D-16-0298.1>,
612 URL <https://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0298.1>.
- 613 Klingaman, N. P., and Coauthors, 2021: Subseasonal Prediction Performance for Aus-
614 tral Summer South American Rainfall. *Weather and Forecasting*, **36 (1)**, 147–169,
615 <https://doi.org/10.1175/WAF-D-19-0203.1>, URL <https://journals.ametsoc.org/view/journals/wefo/36/1/waf-d-19-0203.1.xml>.
- 617 Knapp, K. R., H. J. Diamond, J. P. Kossin, M. C. Kruk, and C. J. Schreck, 2018: International Best
618 Track Archive for Climate Stewardship (IBTrACS) Project, Version 4. NOAA National Centers
619 for Environmental Information, URL [https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:](https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C01552)
620 [C01552](https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C01552), type: dataset, <https://doi.org/10.25921/82TY-9E16>.
- 621 Kolstad, E. W., 2021: Prediction and precursors of *Idai* and 38 other tropical cyclones and storms in
622 the Mozambique Channel. *Quarterly Journal of the Royal Meteorological Society*, **147 (734)**, 45–
623 57, <https://doi.org/10.1002/qj.3903>, URL <https://onlinelibrary.wiley.com/doi/10.1002/qj.3903>.

624 Lavender, S. L., and J. L. McBride, 2021: Global climatology of rainfall rates and lifetime
625 accumulated rainfall in tropical cyclones: Influence of cyclone basin, cyclone intensity and
626 cyclone size. *International Journal of Climatology*, **41** (S1), <https://doi.org/10.1002/joc.6763>,
627 URL <https://onlinelibrary.wiley.com/doi/10.1002/joc.6763>.

628 Lee, C.-Y., S. J. Camargo, F. Vitart, A. H. Sobel, J. Camp, S. Wang, M. K. Tippett, and Q. Yang,
629 2020: Subseasonal Predictions of Tropical Cyclone Occurrence and ACE in the S2S Dataset.
630 *Weather and Forecasting*, **35** (3), 921–938, <https://doi.org/10.1175/WAF-D-19-0217.1>, URL
631 <https://journals.ametsoc.org/doi/10.1175/WAF-D-19-0217.1>.

632 Lee, C.-Y., S. J. Camargo, F. Vitart, A. H. Sobel, and M. K. Tippett, 2018: Subseasonal Tropical
633 Cyclone Genesis Prediction and MJO in the S2S Dataset. *Weather and Forecasting*, **33** (4),
634 967–988, <https://doi.org/10.1175/WAF-D-17-0165.1>, URL [https://journals.ametsoc.org/doi/10.](https://journals.ametsoc.org/doi/10.1175/WAF-D-17-0165.1)
635 [1175/WAF-D-17-0165.1](https://journals.ametsoc.org/doi/10.1175/WAF-D-17-0165.1).

636 Lenzen, M., A. Malik, S. Kenway, P. Daniels, K. L. Lam, and A. Geschke, 2019: Economic
637 damage and spillovers from a tropical cyclone. *Natural Hazards and Earth System Sciences*,
638 **19** (1), 137–151.

639 Li, H., J. H. Richter, C.-Y. Lee, and H. Kim, 2022: Subseasonal tropical cyclone prediction
640 and modulations by mjo and enso in cesm2. *Journal of Geophysical Research: Atmospheres*,
641 **127** (22), e2022JD036 986.

642 Li, S., and A. W. Robertson, 2015: Evaluation of submonthly precipitation forecast skill from
643 global ensemble prediction systems. *Monthly Weather Review*, **143** (7), 2871–2889.

644 Lim, Y.-K., N. P. Arnold, A. M. Molod, and S. Pawson, 2021: Seasonality in Prediction Skill of
645 the Madden-Julian Oscillation and Associated Dynamics in Version 2 of NASA’s GEOS-S2S
646 Forecast System. *Journal of Geophysical Research: Atmospheres*, **126** (18), e2021JD034 961.

647 Luitel, B., G. Villarini, and G. A. Vecchi, 2018: Verification of the skill of numerical weather pre-
648 diction models in forecasting rainfall from us landfalling tropical cyclones. *Journal of hydrology*,
649 **556**, 1026–1037.

- 650 Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the GEOS-5 atmo-
 651 spheric general circulation model: Evolution from MERRA to MERRA2. *Geoscientific Model*
 652 *Development*, **8** (5), 1339–1356.
- 653 Molod, A., and Coauthors, 2020: GEOS-S2S version 2: The GMAO high-resolution coupled model
 654 and assimilation system for seasonal prediction. *Journal of Geophysical Research: Atmospheres*,
 655 **125** (5), e2019JD031 767.
- 656 Moron, V., and A. W. Robertson, 2020: Tropical rainfall subseasonal-to-seasonal predictability
 657 types. *npj climate and atmospheric science*, **3** (1), 4.
- 658 Murphy, A. H., 1973: A new vector partition of the probability score. *Journal of Applied Meteo-*
 659 *rology and Climatology*, **12** (4), 595–600.
- 660 Needham, H. F., B. D. Keim, and D. Sathiaraj, 2015: A review of tropical cyclone-generated storm
 661 surges: Global data sources, observations, and impacts. *Reviews of Geophysics*, **53** (2), 545–591.
- 662 Nkiaka, E., and Coauthors, 2019: Identifying user needs for weather and climate services to
 663 enhance resilience to climate shocks in sub-Saharan Africa. *Environmental Research Letters*,
 664 **14** (12), 123 003, <https://doi.org/10.1088/1748-9326/ab4dfe>, URL [https://iopscience.iop.org/](https://iopscience.iop.org/article/10.1088/1748-9326/ab4dfe)
 665 [article/10.1088/1748-9326/ab4dfe](https://iopscience.iop.org/article/10.1088/1748-9326/ab4dfe).
- 666 NOAA CDR Program, 2022: NOAA Climate Data Record (CDR) of CPC Morphing Technique
 667 (CMORPH) High Resolution Global Precipitation Estimates, Version 1. NOAA National Centers
 668 for Environmental Information, URL [https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:](https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00948)
 669 [C00948](https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00948), type: dataset, <https://doi.org/10.25921/W9VA-Q159>.
- 670 Pegion, K., and Coauthors, 2019: The Subseasonal Experiment (SubX): A Multimodel Subseasonal
 671 Prediction Experiment. *Bulletin of the American Meteorological Society*, **100** (10), 2043–2060,
 672 <https://doi.org/10.1175/BAMS-D-18-0270.1>, URL [https://journals.ametsoc.org/view/journals/](https://journals.ametsoc.org/view/journals/bams/100/10/bams-d-18-0270.1.xml)
 673 [bams/100/10/bams-d-18-0270.1.xml](https://journals.ametsoc.org/view/journals/bams/100/10/bams-d-18-0270.1.xml).
- 674 Rappaport, E. N., 2014: Fatalities in the united states from atlantic tropical cyclones: New data
 675 and interpretation. *Bulletin of the American Meteorological Society*, **95** (3), 341–346.

- Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quarterly Journal of the Royal Meteorological Society*, **127** (577), 2473–2489.
- Risbey, J. S., and Coauthors, 2021: Standard assessments of climate forecast skill can be misleading. *Nature Communications*, **12** (1), 4346.
- Schreck, C. J., and Coauthors, 2023: Advances in tropical cyclone prediction on subseasonal time scales during 2019–2022. *Tropical Cyclone Research and Review*, **12** (2), 136–150, <https://doi.org/10.1016/j.tcr.2023.06.004>, URL <https://linkinghub.elsevier.com/retrieve/pii/S2225603223000255>.
- Smith, J. A., M. L. Baeck, Y. Su, M. Liu, and G. A. Vecchi, 2023: Strange storms: Rainfall extremes from the remnants of hurricane ida (2021) in the northeastern us. *Water resources research*, **59** (3), e2022WR033 934.
- Specq, D., and L. Batté, 2020: Improving subseasonal precipitation forecasts through a statistical–dynamical approach : application to the southwest tropical Pacific. *Climate Dynamics*, **55** (7-8), 1913–1927, <https://doi.org/10.1007/s00382-020-05355-7>, URL <https://link.springer.com/10.1007/s00382-020-05355-7>.
- Tippett, M. K., 2008: Comments on “the discrete brier and ranked probability skill scores”. *Monthly Weather Review*, **136** (9), 3629–3633.
- Tippett, M. K., M. Almazroui, and I.-S. Kang, 2015: Extended-range forecasts of areal-averaged rainfall over saudi arabia. *Weather and Forecasting*, **30** (4), 1090–1105.
- Tippett, M. K., A. G. Barnston, and A. W. Robertson, 2007: Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. *Journal of climate*, **20** (10), 2210–2228.
- Tippett, M. K., T. DelSole, and A. G. Barnston, 2014: Reliability of regression-corrected climate forecasts. *Journal of climate*, **27** (9), 3393–3404.
- Vannière, B., and Coauthors, 2020: The Moisture Budget of Tropical Cyclones in High-ResMIP Models: Large-Scale Environmental Balance and Sensitivity to Horizontal Resolution. *Journal of Climate*, **33** (19), 8457–8474, <https://doi.org/10.1175/JCLI-D-19-0999.1>, URL <https://journals.ametsoc.org/doi/10.1175/JCLI-D-19-0999.1>.

- Vignaud, N., A. W. Robertson, and M. K. Tippett, 2017: Multimodel Ensembling of Sub-seasonal Precipitation Forecasts over North America. *Monthly Weather Review*, **145** (10), 3913–3928, <https://doi.org/10.1175/MWR-D-17-0092.1>, URL <https://journals.ametsoc.org/doi/10.1175/MWR-D-17-0092.1>.
- Vignaud, N., M. K. Tippett, and A. W. Robertson, 2018: Probabilistic Skill of Subseasonal Precipitation Forecasts for the East Africa–West Asia Sector during September–May. *Weather and Forecasting*, **33** (6), 1513–1532, <https://doi.org/10.1175/WAF-D-18-0074.1>, URL <https://journals.ametsoc.org/doi/10.1175/WAF-D-18-0074.1>.
- Villarini, G., R. Goska, J. A. Smith, and G. A. Vecchi, 2014: North Atlantic Tropical Cyclones and U.S. Flooding. *Bulletin of the American Meteorological Society*, **95** (9), 1381–1388, <https://doi.org/10.1175/BAMS-D-13-00060.1>, URL <https://journals.ametsoc.org/view/journals/bams/95/9/bams-d-13-00060.1.xml>.
- Vitart, F., 2017: Madden–Julian Oscillation prediction and teleconnections in the S2S database. *Quarterly Journal of the Royal Meteorological Society*, **143** (706), 2210–2220, <https://doi.org/10.1002/qj.3079>, URL <https://onlinelibrary.wiley.com/doi/10.1002/qj.3079>.
- Vitart, F., and A. W. Robertson, 2018: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate and Atmospheric Science*, **1** (1), 3, <https://doi.org/10.1038/s41612-018-0013-0>, URL <http://www.nature.com/articles/s41612-018-0013-0>.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete brier and ranked probability skill scores. *Monthly Weather Review*, **135** (1), 118–124.
- White, C. J., and Coauthors, 2022: Advances in the application and utility of subseasonal-to-seasonal predictions. *Bulletin of the American Meteorological Society*, **103** (6), E1448–E1472.
- Xiang, B., and Coauthors, 2015: Beyond Weather Time-Scale Prediction for Hurricane Sandy and Super Typhoon Haiyan in a Global Climate Model. *Monthly Weather Review*, **143** (2), 524–535, <https://doi.org/10.1175/MWR-D-14-00227.1>, URL <http://journals.ametsoc.org/doi/10.1175/MWR-D-14-00227.1>.
- Xie, P., R. Joyce, S. Wu, S.-H. Yoo, Y. Yarosh, F. Sun, and R. Lin, 2017: Reprocessed, Bias-Corrected CMORPH Global High-Resolution Precipitation Estimates from 1998. *Journal of*

- 732 *Hydrometeorology*, **18 (6)**, 1617–1641, <https://doi.org/10.1175/JHM-D-16-0168.1>, URL [http:](http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0168.1)
733 [//journals.ametsoc.org/doi/10.1175/JHM-D-16-0168.1](http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0168.1).
- 734 Yamaguchi, M., and N. Koide, 2017: Tropical cyclone genesis guidance using the early stage
735 dvorak analysis and global ensembles. *Weather and Forecasting*, **32 (6)**, 2133–2141.