

# Retrospective Sub-Seasonal Forecasts of Extreme Precipitation Events In The Arabian Peninsula Using Convective-Permitting Modeling

Christoforus Bayu Risanto (✉ [cbrisanto@email.arizona.edu](mailto:cbrisanto@email.arizona.edu))

The University of Arizona <https://orcid.org/0000-0003-3225-4948>

Hsin-I Chang

The University of Arizona

Thang M. Luong

King Abdullah University of Science and Technology

Christopher L. Castro

The University of Arizona

Hari P. Dasari

King Abdullah University of Science and Technology

Ibrahim Hoteit

KAUST: King Abdullah University of Science and Technology

---

## Research Article

**Keywords:** Sub-seasonal forecasts, Extreme precipitation, Arabian Peninsula, Convective-permitting model, Ensemble forecasts

**Posted Date:** August 4th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-733656/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

This paper is to demonstrate the potential of extreme cool-season precipitation forecasts in the Arabian Peninsula (AP) at sub-seasonal time scales, identify the region and periods of forecast opportunity, and investigate the predictability of synoptic-scale forcing at sub-seasonal time scales. To this end, we simulate 18 extreme precipitation events using the convective-permitting weather research and forecasting (CP-WRF) model with lateral boundary forcing from the European Centre of Medium-range Weather Forecasts sub-seasonal to seasonal reforecasts (ECMWF S2S reforecasts). The simulations are initiated at one-, two-, and three-week lead times. At all lead times, the CP-WRF improved the mean accumulated precipitation in the extratropical synoptic regimes over the west coastal and central AP and the central Red Sea. Based on categorical statistics with a threshold of 20-mm accumulated precipitation over 7 days, the CP-WRF accurately forecasted the precipitation over Jeddah, the west coast of AP, and the central Red Sea up to three weeks lead time. The relative operating characteristic curve reconfirmed the high forecasting skill of the CP-WRF, with an area under the curve above 0.5 in most of the events at all lead times. Finally, the correlation coefficients between the ECMWF and ECMWF reanalysis interim 500-hPa geopotential heights were higher in the events associated with the extratropical synoptic regime than in those associated with the tropical synoptic regime, regardless of lead time. Therefore, the convective-permitting model can potentially improve the accuracy of extreme winter precipitation forecasts at two- and three-week lead times over Jeddah, the west coast of AP, and the central Red Sea in the extratropical synoptic regime.

## Introduction

More reliable forecast systems are needed for early warning of extreme weather events (Vitart and Robertson 2018) such as heavy precipitation, flash flooding, and heat waves (e.g., Liang and Lin 2018; Vuillaume et al. 2018; Grimm et al. 2021). The socioeconomic impacts of such events have become more profound in today's changing climate (e.g., Smith and Katz, 2013). We are interested in extreme precipitation events that cause flash flooding on the Arabian Peninsula (AP), one of the driest and most water-limited environments in the world (Köppen 1936). Flash flooding occurs when heavy precipitation falls on dry soils with low soil-infiltration capacity (Al Saud 2010; Almazroui, 2011), filling ephemeral stream channels (called *wadis* in the Middle East) within a matter of minutes (Haggag and El-Badry 2013; Deng et al. 2015). In recent decades, the AP has experienced rapid socioeconomic development, expansion of urbanization and agricultural activities, and high population growth (Luong et al. 2020b; Hoteit et al. 2021). Episodic extreme precipitation events and flash flooding have increased concurrently, causing hundreds of casualties and infrastructural damages costing more than one billion U.S. dollars (Al-khalaf and Basset 2013; Haggag and El-Badry 2013; de Vries et al. 2016).

Climatologically, most precipitation in the AP occurs during the cool season from November to April. On average, the southwestern part of the AP receives approximately 300 mm of rainfall (60% of the annual precipitation) from October to May, while the rest of the region receives very small amounts of precipitation (Almazroui 2011). During this warm season from May to October, the precipitation at the

southwestern tip of the AP is principally generated by southwesterly Indian monsoon winds and orographic lifting (Atlas 1984; Almazroui 2011; Viswanadhapalli et al. 2016). The present study focuses on extreme precipitation events over the AP during the cool season, which are often associated with mesoscale convective systems (MCSs) (Dasari et al. 2017; de Vries 2018). In general, MCSs occurring over land produce large areas of heavy precipitation (covering 100 km or more) and strong winds (e.g., Houze 1993; Schumacher and Rasmussen 2020). The driving cool-season synoptic patterns that favor MCS occurrence over the AP are detailed in Sect. 2.

Our objective is to improve the accuracy of retrospective sub-seasonal forecasts of extreme precipitation in the AP. The time period of sub-seasonal forecasts falls between those of medium-range numerical weather predictions (one to two weeks) and seasonal forecasts (three to six months) (Vitart et al. 2017; Robertson et al. 2020). Therefore, the sub-seasonal forecast window lies between two weeks and less than two months (Vitart et al. 2017). Within this window, the three- and four-week lead time forecasts are most important. Reliable sub-seasonal forecasts are needed for planning and management decisions in multiple sectors, including agriculture, energy, water resources, and emergency preparedness (e.g., Mariotti et al. 2020; Vitart and Robertson 2015; Vitart et al. 2017; Vitart and Robertson 2018). Particularly on the AP, improved sub-seasonal forecasts of extreme precipitation events would mitigate physical damage and loss of life from flash flooding, conferring significant socioeconomic benefits. Forecasting within this time scale has been challenging (e.g., Collier and Zhang 2007; Becker 2017), because the forecast lead time is sufficiently long to lose much of the memory of the atmospheric initial conditions, yet too short to be influenced by variability of the ocean boundary conditions (Vitart et al. 2017; Liang and Lin 2018).

Sub-seasonal predictability has been well evaluated in the context of existing forecast systems throughout the world with relatively coarse spatial resolutions ( $> 0.5^\circ$ ), as compared to convective-permitting short-term numerical weather prediction forecasts. Within Asia, these modeling systems have shown value in the sub-seasonal prediction of summer monsoon precipitation (Jie et al. 2018), extreme precipitation in Sri Lanka (Vuillaume et al. 2018), and temperature and precipitation in East Asia (Liang and Lin 2018), especially during periods of 'forecast opportunity' related to modes of climate variability (Mariotti et al. 2020; Merryfield et al. 2020). These modes include the boreal summer intra-seasonal oscillation, the Madden Julian Oscillation, and the El Niño Southern Oscillation.

The present study investigates whether there is value added by convective-permitting modeling at the sub-seasonal forecast timescale in a target region where the forecast is realized, in this case the AP. The grid spacing of convective-permitting models (CPMs) is nominally less than 4 km and the precipitation processes are explicitly represented without convective parameterization (e.g., Prein et al. 2015). CPMs have been widely applied in weather forecasting (e.g., Clark et al. 2016) and climate projection (e.g., Prein et al. 2015, Kendon et al. 2017). Included in the CPM literature are some studies simulating extreme weather events over the AP (e.g., Deng et al. 2015; Dasari et al. 2017; Luong et al. 2020b) and incorporation of data assimilation (Viswanadhapalli et al. 2016) to improve short-range forecasts, which inform the methodological approaches in this work. With respect to convective precipitation, CPMs add

substantial value in determining the amount, timing, and storm structure of convective precipitation (e.g., Prein et al. 2015; Ban et al. 2014; Kendon et al. 2012).

It is technically possible to dynamically downscale long-term reforecast products at S2S timescales to investigate potential value added in forecast skill. Such data are currently available from the S2S Project (Vitart et al. 2017) and the Sub-seasonal Experiment (Sub-X) project (Pegion et al. 2019). The value added by dynamical downscaling of global models for retrospective seasonal forecasts has been investigated through coordinated community efforts. An example of seasonal forecasting on meso- $\beta$  grid scales (tens of km) is the Multi-RCM Ensemble Downscaling of Multi-GCM Seasonal Forecasts (MRED) experiment (Shukla and Lettenmaier 2013; De Haan 2015). MRED and similar works have shown that regional models can improve the forecast skill where there is already some pre-existing skill in the global model, add the greatest value during periods of ‘forecast opportunity’ when large-scale climate variability projects strongly on the regional precipitation, and better represent mesoscale meteorological processes (e.g., Shukla and Lettenmaier 2013; De Haan 2015; Castro et al. 2012).

In this study, we apply CPM to extreme precipitation forecasting in the AP at sub-seasonal time scales, which has rarely been attempted previously. This work is important not only for improving the capability of extreme weather forecasting and enhancing the environmental sustainability of the AP (Hoteit et al. 2021), but for establishing a replicable methodological approach for sub-seasonal CPM in other parts of the world. The goals of this study are threefold: 1) to demonstrate the potential improvement of extreme cool-season precipitation forecasting of a CPM over the AP with one-, two- and three-week lead times, 2) to identify the regions and periods of forecast opportunity, and 3) to investigate the predictability of synoptic-scale forcing at these sub-seasonal time scales. The remainder of this manuscript is structured as follows. Section 2 discusses the synoptic patterns associated with extreme convective precipitation events over the AP during the cool season. Section 3 describes the data and methods of the analysis performed on 21 extreme precipitation events. Section 4 presents the results, emphasizing the value added by the CPM during periods of forecast opportunity. Section 5 and 6 conclude the work with a general discussion and a summary of the main conclusions, respectively.

## **Synoptic Environments Of Extreme Cool-season Precipitation In Arabian Peninsula**

Prior studies of convective precipitation events have described the favorable conditions for MCSs over the AP (e.g., Alkhalaf and Basset 2013; Haggag and El-Badry 2013; de Vries 2016, 2018; Luong et al. 2020a). On the synoptic scale (Fig. 1), a mid-latitude middle-to-upper level (850–500 hPa) trough in the eastern Mediterranean interacts with a Red Sea trough (RST) near the surface. This interaction is influenced by a stationary lower-to-middle level Arabian anticyclone (AA) centered over the southeastern AP and the Arabian Sea. As the mid-latitude trough intrudes into the subtropics and propagates eastward, it amplifies the RST, deflecting the moisture flow from the Red Sea and the Arabian Sea toward the AP (de Vries et al. 2018). The interaction between the mid-latitude trough and the AA at 850 to 700 hPa creates a Red Sea convergence zone (RSCZ; Pedgley 1966; Langodan et al. 2017; Viswanadhapalli et al. 2017) at

approximately 18°N. It also drives intense winds between the coastal mountain gaps along the African coast (Longadan et al. 2017), mostly prominently the Tokar jet (Longadan et al. 2014; Davis et al. 2015). These cross-Red Sea winds gather the moisture from the sea surface and transport it eastward (Hoteit et al. 2021). The low-level moisture transport combined with orographic lift and diabatic heating over the mountain ranges in the southwestern AP increases the atmospheric instability, facilitating convective organization and extreme precipitation (de Vries et al. 2016; Dasari et al. 2017).

Recently, the dynamic and thermodynamic aspects of MCS-driven extreme precipitation over the AP have also been reported. De Vries et al. (2013) demonstrated that extreme precipitation events over the western AP are generated by an active RST (ARST). During the ARST, the RST extends northward and the AA intensifies, increasing the southerly advection of warm moist air and enhancing the tropospheric instability, especially over the southwestern AP. The large amount of moisture falls as heavy precipitation. The severity of this precipitation is closely related to the potential vorticity (PV) intrusion toward the subtropics and the integrated water vapor transport (IVT; de Vries et al., 2018). As the stratospheric PV intrudes further south and the IVT magnitude increases, the precipitation becomes more extreme.

Extreme cool-season precipitation events, especially around the city of Jeddah, are also related to the synoptic-scale weather pattern over the AP. Luong et al. (2020a) classified these events into three synoptic-scale circulation regimes or clusters, namely, extratropical, transitional, and tropical regimes (Fig. 1). The cluster classification was based on self-organized map (SOM) analysis (hereafter referred to as Luong–SOM analysis) of 18 daily-averaged atmospheric variables derived from the ECMWF reanalysis interim (ERA-Interim), which was conducted over the whole AP (10–50° N, 0–70° E) during November to April of each year from 1979 to 2018. The variables included the zonal and meridional winds, geopotential heights, and PV at the 330 K potential temperature level. Luong et al. (2020a) associated the most intense and organized precipitation events with the RST and AA combined with the equatorward deep intrusion of the upper level (850 hPa) trough (Fig. 1a). The precipitation over Jeddah generally reduces when the synoptic pattern is strongly influenced by tropical air in the RST (Fig. 1c). The authors also hypothesized that the precipitation driven by the extratropical regime is more predictable than that driven by the tropical synoptic regime.

Dasari et al. (2018) suggested that synoptic-scale patterns are also influenced by the El Niño Southern Oscillation. During El Niño years, the RST extends farther north, shifting the RSCZ northward to around 20.5°N. This shift significantly increases the total winter precipitation in the northern Red Sea region. During La Niña years, the RST appears over the southern Red Sea, shifting the RSCZ southward to around 18.5°N. Our study adopts the Luong et al. (2020a) severe weather event classification to assess the forecast opportunity of our extreme event sub-seasonal forecast simulations.

## Data And Methods

### *a. WRF domain and configuration*

Our retrospective sub-seasonal simulations applied the Advanced Research version of the Weather Research and Forecasting (WRF-ARW; Skamarock et al. 2008; Powers et al. 2019) model version 3.8. Our model configuration implemented two two-way nested domains with horizontal resolutions of 20 km (d01) and 4 km (d02), each with 53 vertical levels in eta coordinates (see Fig. 2). As the inner domain has a convective-permitting horizontal grid spacing, it explicitly resolves convection at the meso- $\gamma$  scale (1–4 km) without a convective parameterization scheme. This convective-permitting configuration has been implemented for the AP region in previous studies. For example, Viswanadhapalli et al. (2016) adopted the Kain–Fritsch scheme, but our outer domain employed the Grell–Freitas cumulus parameterization scheme (Grell et al. 2013) following Gao et al. (2017) that demonstrated the Grell–Freitas scheme’s lower sensitivity to the model resolution than that of Kain–Fritsch scheme. Other physical models applied to both domains are shown in Table 1.

Table 1  
List of the physical schemes used in the WRF-ARW configuration.

Category	Scheme
Microphysics	Thompson (Thompson et al. 2008)
Longwave radiation	Rapid Radiative Transfer Model GCM (Iacono et al. 2008)
Shortwave radiation	RRTMG (Iacono et al. 2008)
Surface layer	Melior-Yamada Nakanishi and Niino (Nakanishi and Niino 2006)
Land surface	unified Noah (Tewari et al. 2004)
Planetary boundary layer	Melior-Yamada Nakanishi and Niino 3rd level TKE (Nakanishi and Niino 2006)
Cumulus Parameterization (d01)	Grell–Freitas (Grell et al. 2013)

We note that other studies (e.g., Carrillo et al. 2017) applied the sea surface temperature (SST) bias correction in a dynamically-downscaled climate forecast reanalysis. They chose this correction because the SST influences the convective organization. As our study investigates the potential of sub-seasonal forecasts, we decided not to apply the SST bias correction yet.

#### *b. ECMWF S2S reforecast datasets*

As the lateral boundary conditions and initial conditions of the simulations, we used the ECMWF S2S reforecast datasets (Vitart et al. 2017). The datasets (hereafter referred to as raw ECMWF) range from 0 to 46 days with a spatial resolution of  $0.25^\circ \times 0.25^\circ$  on days 0 to 10 and  $0.5^\circ \times 0.5^\circ$  after day 10. There are 91 vertical levels. The reforecast datasets are generated twice weekly (Mondays and Thursdays) with 11 ensemble members in each initialization. Using the actual forecast model, the reforecasts are run for the past several years on the same calendar day as the forecast. These reforecasts are used for calibrating the actual forecasts (Vitart et al. 2017). The reforecast years of the datasets now cover the

past 20 years. The S2S reforecast database can be accessed online at <http://apps.ecmwf.int/datasets/data/s2s>; <http://s2s.ecmwf.int>.

*c. Simulations and classification of extreme events over the AP*

Among the winter convective events from 1999 to 2018, we selected the top 21 extreme precipitation events with recorded rain gauge measurements over 20 mm day<sup>-1</sup> near Jeddah (see Table 2). All events were verified against the spatial precipitation distribution obtained from the Global Precipitation Measurement daily precipitation product (GPM\_3IMERGDF; Huffman 2018). The WRF S2S simulations of the individual convective extreme events were initialized at approximate lead times of three weeks (W3), two weeks (W2), and one week (W1) using all 11 ensemble members. The exact initialization date was based on the raw ECMWF forcing run on Mondays. For example, the simulation of the 3 November 2018 event was initiated at 0000 UTC on 14 October 2018 for W3, 21 October 2018 for W2, and 28 October 2018 for W1. The W3, W2, and W1 simulations were integrated over 21, 14, and 7 days respectively and outputs were generated every three hours. All simulations were ended several days after the event, giving a  $\pm 3$ -day accumulated precipitation centered on the day of the event for statistical verification purposes of W2 and W3.

Table 2

List of extreme precipitation events (cases) and dates of the simulation initiation at 1-, 2-, and 3-week lead times

<b>CASE</b>	<b>WEEK1</b>	<b>WEEK2</b>	<b>WEEK3</b>	<b>node number</b>	<b>Synoptic regime</b>
1999-01-08	1999-01-01	1998-12-24	1998-12-17	1	Extratropical
2000-11-15	2000-11-12	2000-11-05	2000-10-29	3	Tropical
2000-11-16	2000-11-12	2000-11-05	2000-10-29	3	Tropical
2001-12-31	2001-12-24	2001-12-17	2001-12-10	1	Extratropical
2002-03-05	2002-02-26	2002-02-19	2002-02-12	3	Tropical
2002-11-30	2002-11-26	2002-11-19	2002-11-12	2	Transition
2004-12-03	2004-11-26	2004-11-19	2004-11-12	3	Tropical
2005-01-22	2005-01-15	2005-01-08	2005-01-01	1	Extratropical
2009-11-25	2009-11-19	2009-11-12	2009-11-05	1	Extratropical
2009-12-22	2009-12-17	2009-12-10	2009-12-03	3	Tropical
2010-12-30	2010-12-24	2010-12-17	2010-12-10	1	Extratropical
2011-01-26	2011-01-18	2011-01-11	2011-01-04	1	Extratropical
2015-03-25	2015-03-19	2015-03-12	2015-03-05	2	Transition
2015-10-29	2015-10-22	2015-10-15	2015-10-08	3	Tropical
2015-11-17	2015-11-12	2015-11-05	2015-10-29	1	Extratropical
2015-11-25	2015-11-19	2015-11-12	2015-11-05	2	Transition
2016-12-02	2016-11-26	2016-11-19	2016-11-12	1	Extratropical
2017-02-15	2017-02-12	2017-02-05	2017-01-29	1	Extratropical
2017-11-21	2017-11-19	2017-11-12	2017-11-05	1	Extratropical
2018-10-28	2018-10-21	2018-10-14	2018-10-07	3	Tropical
2018-11-03	2018-10-28	2018-10-21	2018-10-14	3	Tropical

To classify the events, we evaluated the error of each event from each cluster node produced by the Luong–SOM analysis. The event was assigned to the cluster with the lowest error (Hewitson and Crane 2002). Ten and eight events were classified as dominantly influenced by extratropical and tropical drivers, respectively, and three events were transitional. Table 2 shows the classification of each event. As the transitional precipitation events were very few, they were excluded from our analysis and discussion.

#### *d. Ground reference*

To evaluate the predictability of the precipitation forecasts in each synoptic classification, we required a precipitation ground reference. For this purpose, we used the precipitation products from the Global Precipitation Measurement Final Precipitation L3 v06 (GPM\_3IMERGHHV06; hereafter referred to as GPMF; Huffman 2018) and the King Abdullah University Science and Technology – Reanalysis (hereafter referred to as KAUST-RA; Dasari et al. 2019). In observation-limited regions such as the AP, the observation datasets require sufficient spatial and temporal availability and must reasonably represent the ground-based measurements. Being a convection-permitting reanalysis dataset, KAUST-RA is a viable alternative option for regions lacking a sub-daily precipitation product based on rain gauges and/or radar-derived precipitation. The AP region fits this category.

Other selection criteria for the ground reference were the bias and root mean square error (RMSE) of the multiple-satellite products and the KAUST-RA relative to rain gauge measurements. The satellite products included the GPMF, the National Oceanic and Atmospheric Administration Climate Prediction Center Morphing technique (CMORPH; Joyce et al. 2004), the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Sorooshian et al. 2002), and the Global Satellite Mapping of Precipitation (GSMaP; Kubota et al. 2007). Details of these satellite products are given in Table 3. The biases and RMSEs were computed relative to the daily precipitation dataset collected from weather stations across Saudi Arabia by the Presidency of Meteorology and Environment between January of 2000 and December of 2012. From the stations with over 95% completion of their precipitation data, we extracted the cool-season precipitation events that accumulated 5 mm or more precipitation. Only 43 out of 208 stations met the data-completion threshold, and among the 12 cool seasons, 100 precipitation days met the 5-mm accumulated precipitation threshold. We computed the bias and RMSE at each station and averaged them across the 43 stations. The GPMF precipitation product, with the lowest bias and the second lowest RMSE (Fig. 3), was selected along with the KAUST-RA precipitation products as the ground reference for this study.

Table 3  
List of satellite precipitation products and a model reanalysis product (KAUST-RA)

Product	Spatial resolution	Temporal resolution	References
GPM Final v6	0.1°	Half-hourly	Huffman (2018)
CMORPH	8 km	Half-hourly	Joyce et al. (2004)
PERSIANN	0.25°	1 hourly	Sorooshian et al. (2002)
GSMaP	0.1°	daily	Kubota et al. (2007)
KAUST-RA	5 km	daily	Dasari et al. (2019)

#### e. Analysis Methods

##### i. Average accumulated precipitation

To evaluate how effectively the convective-permitting-WRF (CP-WRF) predicts the convective extremes over the AP at sub-seasonal time scales, we first examined the accumulated precipitation. The model-generated precipitation was evaluated at  $\pm 1$  days of the event (three days of accumulated precipitation) on weather forecasting time scale W1, and at  $\pm 3$  days of the event (seven days of accumulated precipitation) on W2 and W3. The ensemble mean of each simulated event and the average precipitation during events associated with the extratropical and tropical regimes were then calculated. This procedure was repeated for the raw ECMWF precipitation. The precipitation forecasts of the extratropical and tropical events were compared by subtracting the average precipitation of the raw ECMWF from that of the CP-WRF. The CP-WRF average precipitation biases relative to the GPMF and KAUST-RA precipitations were also calculated. The differences and biases indicate the quality of the CP-WRF relative to the coarser-resolution raw ECMWF and observations.

## *ii. Categorical statistics*

As the model simulations were evaluated at every grid point in the domain, the spatial resolution of the modeled precipitation field was up-scaled from 4 km to 5 km (the resolution of KAUST-RA) and  $0.1^\circ$  (the resolution of GPMF) using the `remapbil` function available on the Climatic Data Operators built by the Max Planck Institute für Meteorologie (<https://code.mpimet.mpg.de/projects/cdo/>). The raw ECMWF reforecast dataset was re-gridded to the resolutions of GPMF and KAUST-RA using the same function. Following Moker et al. (2018) and Risanto et al. (2021), the precipitation forecast skill was measured with three categorical statistics: the critical success index (CSI), probability of detection (POD), and false alarm ratio (FAR), as described in Wilks (2011). These verification metrics are intuitive because they are based on a  $2 \times 2$  forecast contingency table with four entries: hits, misses, false alarms, and correct negatives. The CSI, POD, and FAR verification metric are respectively calculated as

$$CSI = \frac{hits}{hits+misses+false\ alarms} \quad (1)$$

$$POD = \frac{hits}{hits+misses} \quad (2)$$

$$FAR = \frac{false\ alarms}{hits+false\ alarms} \quad (3)$$

We defined a precipitation event at a grid point with 20 mm of accumulated precipitation in three days for W1 forecasts and seven days for W2 and W3 forecasts. Considering the location uncertainty at these sub-seasonal time scales, we applied a neighborhood verification technique that considers the modeled events in  $\pm 4$  grid points, covering approximately  $90 \text{ km} \times 90 \text{ km}$ , approximately twice the raw ECMWF grid spacing. The precipitation verification in the CPM simulations was then approximately within the effective resolution of the raw ECMWF data, alleviating the double-penalty problem of missing precipitations in correct locations and getting precipitations in wrong locations. A precipitation event is included in the contingency table if it meets the precipitation threshold and falls into one of the 81 grid points. The CSI, POD, and FAR values range from 0 to 1. As the CSI and POD values approach 1 (0), the

precipitation forecast skills increase (decrease). In contrast, as the FAR values approach 1 (0), the precipitation forecast skills decrease (increase). Here we calculated  $1 - \text{FAR}$  to give consistent interpretations of the CSI, POD, and FAR.

To assess the modeled precipitation forecast skill, we subtracted the CSI, POD, and FAR values of each grid point in the raw ECMWF precipitation from those in the CP-WRF precipitation. At grid points with positive values, the CP-WRF demonstrated higher precipitation forecasting skill than the raw ECMWF. The statistical field significance was also computed. At each grid point, a two-tailed local significance test was established via a Monte Carlo resampling method with 1000 random permutations ( $p < 0.1$ ). The statistical field significance was computed with 1000 random resampling of the maps, as described in Livezey and Chen (1983). The critical value of the statistical field significance was set to 90%, the 900th value of the histogram. That is, only grid points containing at least 900 unique values were used. We emphasize that the methods in all of these analyses were previously demonstrated in convective-permitting NWP simulations in northwest Mexico (Moker et al. 2018, Risanto et al. 2021).

We also computed the frequency distribution of the domain-wide POD as a percentage of total number of grid points in the domains for the extratropical and tropical events at W1, W2, and W3. To calculate the statistical significance of the distribution, we generated 1000 random permutations of the POD distributions of the raw ECMWF and CP-WRF using the Monte Carlo technique. The statistical significance was set to 90%. Only the PODs relative to the GPMF are shown here, as the PODs relative to KAUST-RA are identical. The purpose is to show the forecast skill of the CP-WRF.

### *iii. Relative Operating Characteristic*

In addition to the previously mentioned categorical indices (CSI, POD and FAR), we constructed the relative (or receiver) operating characteristic (ROC; Mason and Graham 2002) for determining the probability of precipitation forecasts at the sub-seasonal time scale. Like the other standard probabilistic verification metrics, the ROC curve is a common verification index of operational forecasts and is suitable for ensemble forecast systems (Coelho et al. 2019). The methodology was first developed to classify the accuracy of radar signals during World War II, and its use has expanded to clinical diagnostic testing and screening (Zou et al. 2007). In forecast verification, the ROC curve classifies the forecasts into hits or misses based on the observations at a pre-determined threshold and is often plotted as POD versus FAR in Cartesian coordinates. Being conditioned by the observations, the ROC curve in our case can be analyzed only during convective events. When the curve starts from the bottom left corner, ends at the top right corner, and bends toward the top left corner, the skill of the forecasting system is high. When the curve is below the diagonal line joining the bottom left to the top right corner (the ROC curve of a random classifier), the forecasting system has no forecasting skill. The forecast probability is quantified by the area under the curve (AUC), which ranges from 0 (no forecasting skill) to 1 (perfect forecast ability). Any AUC below 0.5 means that the classifier performance is worse than random.

ROC curves are also suitable for forecasting evaluations on sub-seasonal timescales. ROC curve analyses were performed by Hudson et al. (2011), who evaluated the potential of sub-seasonal

forecasting of extreme temperatures in Australia in a T47 ( $3.7^\circ \times 3.7^\circ$ ) horizontal resolution model, and by Coelho et al. (2018), who validated sub-seasonal precipitation forecasts in South America, which implemented both the real-time forecast and reforecast ECMWF. Here, we computed the ROC curves of both the CP-WRF and raw ECMWF precipitation events. Similarly to the aforementioned categorical statistics, we calculated the hit and false alarm rates with  $\pm 4$ -gridpoint neighborhood verification and a 20-mm accumulated precipitation in 3 days for W1 and 7 days for W2 and W3 for each ensemble member. The ROC curve of each event was obtained by averaging the values across the ensemble members.

Following previous studies (e.g., Hafez and Almazroui, 2016; Christidis and Stott, 2015), which reported a relationship between geopotential heights and extreme weather and climate, we evaluated the raw ECMWF 500- and 850-hPa geopotential heights relative to those of ERA-Interim using the Pearson's correlation coefficient. Here, we applied the synoptic classification of Luong et al. (2020a). The spatial resolution of the ERA-Interim dataset is approximately 80 km on 60 vertical levels from the surface to 0.1 hPa (Berrisford et al. 2011). This evaluation quantifies the predictability of the sub-seasonal forecast based on the synoptic-scale patterns. At each grid point in the domain, the raw ECMWF geopotential height daily mean of the events was correlated with that of ERA-Interim. As the raw ECMWF reforecast is an 11-member ensemble dataset, the correlation relative to ERA-Interim was calculated for every ensemble member and averaged across the 11 ensemble members. The raw ECMWF geopotential height was qualitatively evaluated by plotting its daily mean and the synoptic-scale pattern determined in the Luong–SOM analysis.

## Results

### *a. Precipitation difference and biases*

When simulating the CP-WRF events, the average total precipitation in all synoptic-related events exceeded the raw ECMWF precipitation. Among the extratropical events (Fig. 4), the total precipitation notably increased at lead time W1 over the west coast of the AP, just north of Jeddah and the Zagros Mountains east of the Persian Gulf (also known as the Arabian Gulf) (see the difference plot in Fig. 4g). The precipitation was notably increased over the central AP where the raw ECMWF generates drier conditions. At W2, the CP-WRF precipitation was again higher over Jeddah, the central Red Sea, and the Zagros Mountains than the raw ECMWF precipitation (Fig. 4b and e) as shown in the difference plots (Fig. 4h). However, the CP-WRF produced drier conditions over the central AP than the raw ECMWF. Similarly to W2, the CP-WRF at W3 obtained higher total precipitation over the central Red Sea, near Jeddah, and the Zagros Mountains than the raw ECMWF (Fig. 4c, f, and i). Over the Red Sea, where the raw ECMWF barely generated precipitation at W3, the precipitation increase was notable.

Among the tropical events (Fig. 5), the average total precipitation of the CP-WRF at W1 was also higher than the raw ECMWF precipitation, especially over the central AP (see difference plot in Fig. 5g). Interestingly the CP-WRF produced two separate precipitation maxima: one southwest of Jeddah, the

other southeast of Jeddah. The city of Jeddah was drier in the CP-WRF model than in the raw ECMWF. At W2 in the raw ECMWF, the precipitation exceeded 8 mm over a large area of the central Red Sea (Fig. 5e), whereas in the CP-WRF, the precipitation was concentrated over the high terrain southeast of Jeddah to the central AP and the east coast of Sudan, exceeding 18 mm precipitation in total (Fig. 5b). In contrast, the central Red Sea was almost dry. Similar patterns were found at W3, but the precipitation signatures were weaker than at W2 (Fig. 5c). These high average precipitations at W1, W2, and W3 in the CP-WRF were likely attributable to convective organization at the meso- $\gamma$  scale, particularly over the complex terrain southeast of Jeddah, which could be resolved in this model but not in the raw ECMWF.

Figure 6 shows the average accumulated precipitations of GPMF and KAUST-RA and the precipitation biases of the CP-WRF at W1, W2, and W3 in the extratropical events. The GPMF clearly gave more precipitation ( $> 20$  mm) in the central AP than the KAUST-RA (Fig. 6a, b, c, g, h, i). The CP-WRF precipitation exhibited a dry bias (indicated by the blue shading) relative to the GPMF precipitation at W1, W2, and W3 (Fig. 6d to f) in the north of Jeddah, the central AP, and the Zagros Mountains. Conversely, it showed a wet bias ( $< 15$  mm) (indicated by the red shading) in the central Red Sea at W2 and W3. Compared to the KAUST-RA precipitation, the CP-WRF consistently exhibited a dry bias over Jeddah and the Zagros Mountains at all lead times (Fig. 6j to l).

Figure 7 shows the observed precipitation and the CP-WRF biases during the tropical events. The KAUST-RA generated more precipitation ( $> 18$  mm) than the GPMF, especially over the central AP (Fig. 7a, b, c, g, h, i). Compared to the GPMF, the CP-WRF exhibited a dry bias over Jeddah and the central AP that increased from W1 to W3 (Fig. 7d to f). Interestingly, wet biases were notable over the east coast of Sudan and the southwestern tip of the AP (Fig. 7d to f). The dry bias of the CP-WRF was more pronounced relative to the KAUST-RA precipitation than relative to the GPMF, especially over the central AP (Fig. 7k, l). This result was expected, because the KAUST-RA generated more precipitation than the GPMF.

In summary, the CP-WRF generated more precipitation in the extratropical events than the raw ECMWF at W1, W2, and W3. This is shown in the differences, which were dominated by positive values over the central Red Sea, the central AP, and the Zagros Mountains. In the tropical events, the CP-WRF generated notable precipitation only in the southwest of Jeddah, the east coast of Sudan, and the central AP at W1. The raw ECMWF generated more precipitation over the central Red Sea at W2 and W3 than the CP-WRF. In general, the CP-WRF precipitation was lower than the GPMF and KAUST-RA precipitations, but higher than the raw ECMWF precipitation.

### *b. Precipitation categorical statistics*

Figure 8 displays the differences between the CSI, POD, and FAR of the CP-WRF and raw ECMWF precipitations relative to the GPMF precipitation for the extratropical events. Consistent with the average CP-WRF precipitation, the precipitation forecast skills were consistently higher in the CP-WRF than in the raw ECMWF when domain-averaging was applied, irrespective of the forecast timescale. The CSI and POD differences were also field-significant ( $> 90\%$ ) from W1 to W3. At W1, the precipitation forecast skill

was high (blue shading) over the entire Red Sea, the central AP, and the Zagros Mountains. At W2 and W3, the forecast skill was notably high over the central Red Sea including Jeddah and areas along the west coast of the AP, and also over the Zagros Mountains. The FAR results (here plotted as  $1 - \text{FAR}$ ) were consistent with the CSI and POD results but were not field-significant.

Figure 9 displays the differences between the three verification metrics of the CP-WRF and raw ECMWF precipitations relative to the GPMF for the tropical events. These results confirm the average CP-WRF precipitations during the tropical events in the previous subsection. At W1, the CP-WRF better forecasted the precipitation than the raw ECMWF (blue shading) over the central AP and a small part of the Red Sea, but underperformed around Jeddah and the central Red Sea (red shading). At W2, the forecast skill of the CP-WRF was higher only over the areas east of Jeddah, the central and northern AP, and the east coast of Sudan. Over this area, the raw ECMWF generally outperformed the CP-WRF, possibly because the CP-WRF inadequately resolves the precipitation over the central Red Sea. As mentioned earlier, the persistent SST might reduce the forecasting ability of the CP-WRF. Finally, at W3, the CP-WRF better forecasted the precipitation than the raw ECMWF in the central Red Sea, Jeddah, and the Zagros Mountains.

Similar results were obtained using KAUST-RA as the ground reference. Figure 10 shows the CSI, POD, and FAR differences between the CP-WRF and raw ECMWF for the extratropical events. The CSI and POD differences were field-significant. At W1, the CP-WRF outperformed the raw ECMWF in the Red Sea, the central AP including the Jeddah area, and the Zagros Mountains. At W2 and W3, the enhanced precipitation forecast skill of the CP-WRF was notable in the central Red Sea, Jeddah, the west coast of AP, and the Zagros Mountains.

For the tropical events (see Fig. 11), the difference plots of the CP-WRF and raw ECMWF again indicate a higher precipitation forecast skill of the CP-WRF than of ECMWF, and field significance of the CSI and POD results. At W1, the CP-WRF outperformed the ECMWF over the area east of Jeddah and the central AP, and northwest of the Persian Gulf. At W2, the enhanced forecast skill was notable over east of Jeddah, some parts of the central and northern AP, and along the east coast of Sudan. In contrast, the forecast skill of CP-WRF was reduced in the central Red Sea. At W3, the CP-WRF outperformed the ECMWF in the central Red Sea and the west coast of AP, including Jeddah.

Figures 12 and 13 show the frequency distributions of the POD relative to the GPMF for all week lead times in the extratropical and tropical events, respectively. In both sets of events and at all lead times, PODs above 0.8 appeared more frequently in the CP-WRF than in the raw ECMWF. At W2, the differences between CP-WRF and ECMWF were statistically significant (marked by yellow stars) only for PODs greater than 0.9, but at W1 and W3, they were statistically significant for PODs greater than 0.8. Among the tropical events (Fig. 13), PODs greater than 0.3 appeared more frequently in the CP-WRF than in the raw ECMWF at W2 and W3. The same results were found in the frequency distributions of POD relative to KAUST-RA (not shown).

In summary, the CSI, POD, and FAR difference results were very similar for two different data sources of observed precipitation. Therefore, the value added by the CP-WRF is not a function of any particular

observational precipitation product. The CP-WRF can better predict both extratropical and tropical precipitation events than the driving raw ECMWF model, up to three weeks ahead of the events. More importantly, the CP-WRF demonstrated higher forecast skill than the raw ECMWF around the area of Jeddah, the central Red Sea, and the high terrain in the west coast of AP, suggesting that the CP-WRF can improve the forecast of extratropical-driven extreme precipitation events. Moreover, these extreme precipitation events occurred at sub-seasonal time scales at the original raw ECMWF resolution, but the overall precipitation amounts were lower than the observed amounts (see subsection 3.1).

### *c. ROC curves and AUC of precipitation*

A 20-mm precipitation threshold was applied to each event in the ROC analysis. Figure 14 shows the ROC curves of the individual extratropical event relative to the GPMF precipitation for W1, W2, and W3. At W1, the raw ECMWF exhibited a high forecast skill in most of the events, but was outperformed by CP-WRF in five events (with AUC values  $> 0.8$ ). The higher forecast skill of the CP-WRF is consistent with the verification metrics analyzed in the previous subsection. At W2, the AUCs of the CP-WRF exceeded 0.5 for nearly all events, again outperforming the raw ECMWF. Most importantly, the AUCs of more than 90% of the events at W3 were higher in the CP-WRF than in the raw ECMWF. Most of the AUCs in CP-WRF were slightly above 0.5. The AUCs below 0.5 in the CP-WRF (two events) were higher than their counterpart AUCs in the raw ECMWF.

Figure 15 shows the ROC curves of the individual tropical events for all lead times. Similarly to the extratropical events, the tropical events were well forecasted by the CP-WRF. At W1, the AUCs of all events were higher in CP-WRF than in the raw ECMWF, although the AUC of one ROC was lower than 0.5 in the CP-WRF model. At W2, the AUCs of the ROC curves of all events exceeded 0.5 in the CP-WRF, but were below 0.5 in the raw ECMWF. At W3, the CP-WRF and raw ECMWF again yielded notably different ROC curves. The AUCs of the ROCs exceeded 0.5 for almost all events in the CP-WRF (for the two exceptions, the AUC was  $\sim 0.4$ ). In contrast, all events in the raw ECMWF yielded AUC values around 0.4.

Figures 16 and 17 show the ROC curves of the individual extratropical and tropical events, respectively, relative to the KAUST-RA precipitation for all lead times. Again, the precipitation prediction accuracy was higher in CP-WRF (AUC  $> 0.5$ ) than in raw ECMWF at W1, W2, and W3. These results are consistent with the ROC curve analysis relative to the GPMF precipitation.

The average differences between the AUCs of the CP-WRF and raw ECMWF analyses relative to the GPMF and KAUST-RA precipitation data are shown in Tables 4 and 5, respectively. All differences are positive, confirming the higher precipitation forecast skill of CP-WRF than of raw ECMWF at W1, W2, and W3 in both extratropical and tropical events. Although the differences were slightly higher for the tropical events than for the extratropical events, whether the tropical events are more predictable than the extratropical events should not be decided from this ROC curve analysis alone, because the number of samples was statistically constrained and the ROC curve analysis covers the whole domain rather than regional areas such as Jeddah and the central Red Sea.

Table 4  
 Difference between the average AUCs of the CP-WRF precipitation events and those of the raw ECMWF precipitation events relative to GPMF precipitation

	W1	W2	W3
Extratropical events	0.13	0.23	0.12
Tropical events	0.16	0.26	0.21

Table 5  
 Difference between the average AUCs of the CP-WRF precipitation events and those of the raw ECMWF precipitation events relative to KAUST-RA precipitation

	W1	W2	W3
Extratropical events	0.10	0.19	0.12
Tropical events	0.17	0.23	0.20

In summary, the precipitation forecast skill of CP-WRF was high at W1, W2, and W3 for both extratropical and tropical events, whereas the driving raw ECMWF exhibited almost no skill at W2 and W3. At W1, the raw ECMWF showed some forecast skill for many events (i.e.,  $AUC > 0.5$ ), but was consistently outperformed by the CP-WRF (see Tables 4 and 5). Therefore, we concluded that the CP-WRF improves the predictability of extreme precipitation events ( $> 20$  mm), up to three weeks ahead of the events on sub-seasonal forecast time scales. According to the categorical statistics (CSI, POD, and FAR) in subsection 3.2, the precipitation forecast skill over Jeddah and the central Red Sea is higher in the extratropical events than in the tropical events. Therefore, the extratropical events provide the forecast opportunity of CP-WRF at these sub-seasonal time scales.

### *Geopotential heights*

Luong et al. (2020a) attempted to identify the dominant synoptic patterns underlying convective extremes over the AP. They found that the interaction between the 500- and 850-hPa geopotential height fields closely influences the positioning of the RST and AA and convective organization over the AP. Therefore, we assessed the predictability of the precipitation forecasts by evaluating the synoptic-scale geopotential heights in the ERA-Interim and raw ECMWF at different lead times of the forecasting. Figure 18 displays the Pearson's correlation-coefficient contours (color-coded shading) between the ERA-Interim and the raw ECMWF 500-hPa geopotential heights, overlaid with the 500-hPa geopotential heights from the Luong–SOM analysis (solid lines) and the daily means of the 500-hPa geopotential heights extracted from the raw ECMWF (dashed lines). The correlation coefficients at W1, W2, and W3 were consistently higher for the extratropical events (Fig. 18a, c, e) than for the tropical events, as shown by both the color-coded shading and the average correlation coefficients ( $R_s$ ) across the domain. The

difference was notable at W3 (Fig. 18e and f) over Jeddah and along the west coast of the AP, where  $R > 0.7$  for the extratropical events and  $R < 0.5$  for the tropical events.

The 500-hPa geopotential height contours were also consistent with the correlation coefficients. At W1 (Fig. 18a), the raw ECMWF afforded extratropical 500-hPa geopotential contours that almost matched those of the Luong–SOM analysis, but its tropical 500-hPa geopotential contours (Fig. 18b) notably differed from those of the Luong–SOM analysis. At W2 and W3, the differences between the contours of the tropical events became more pronounced because the trough over the east coast of Africa was not developed (Fig. 18d, f). In contrast, the extratropical 500-hPa geopotential contours of the raw ECMWF almost matched those of the Luong–SOM analysis. The raw ECMWF did forecast the trough at W2 and W3, but at a weaker magnitude than that in the ERA-Interim.

Interestingly, the 850-hPa geopotential height contours presented higher correlations at W2 and W3 in the tropical events than in the extratropical events (Fig. 19c, d, e, f), whereas at W1, the correlation coefficients of both types of events were almost identical. The correlation-coefficient difference was large at W3 (0.05 for the extratropical events versus 0.32 for the tropical events, also evidenced by the shading). In the tropical events, the raw ECMWF at W2 and W3 (Fig. 19d, f) apparently predicted the RST that closely matched both the ERA-Interim and the Luong–SOM analysis. The AA was also predicted (albeit weakly, with  $R < 0.5$ ) at W2 and W3. However, in the extratropical events, the raw ECMWF poorly predicted the dominant extratropical trough over the Red Sea at W2 and W3 (Fig. 19c, e). Instead, it presented a weak signature of RST over the east coast of Sudan, probably because the SST is warmer in the raw ECMWF than in the ERA-Interim.

The correlation coefficients of the extratropical and tropical 500-hPa geopotential heights were compared by Student's t-test. The differences at W1, W2, and W3 were statistically significant at the  $p$ -value  $>$  significance level. However, when the correlation coefficients of the 850-hPa geopotential heights results were analyzed by the same test, the differences were statistically significant only at W2 and W3. The non-significant difference at W1 was consistent with the similar correlation coefficients at W1.

Whether extratropical events over the AP are more predictable at sub-seasonal time scales than tropical events is difficult to conclude from the above results. We hypothesize that the 850-hPa geopotential height is more directly affected by the rapid changes of near-surface variables such as the SST, moisture fluxes, and near-surface temperature, than the 500-hPa geopotential heights. Therefore, at this height, the forecast contains more uncertainty at sub-seasonal time scales than at 500 hPa. Consequently, the predictability of the 850-hPa geopotential heights disagreed with those of the 500-hPa geopotential heights based on the event classification. An investigation with more extreme event samples is needed. However, previous studies (e.g., Hafez and Almazroui, 2016; Christidis and Stott, 2015) have heavily relied on the 500-hPa geopotential heights for diagnosing extreme events because 500 hPa is the level of non-divergence and is commonly assumed in extreme weather and climate diagnostics. Therefore, considering only the predictability of the 500 hPa geopotential heights, the categorical statistics (CSI,

POD, FAR), and the ROC analysis, we conclude that extreme precipitation at sub-seasonal time scales (W1 to W3) is more predictable in the extratropical synoptic regime than in the tropical synoptic regime.

## Discussion

Our results were constrained by the size of the ensemble members and the number of events. With only 11 members in the reforecast dataset, the statistical analyses are less robust than analyses with at least 30 members. Thirty is commonly considered as the threshold sample number of meaningful statistics (Wilks 2011). Increasing the ensemble member size is quite feasible because the ECMWF generates 51-member ensembles through twice-weekly real-time forecasts (Vitart et al. 2017). Increasing the ensemble members and number of events would certainly increase the statistical robustness of our findings, but risks exhausting the computing resources. In any event, our results demonstrate that the precipitation forecast skill at sub-seasonal time scales can be improved with CPM. Dynamic downscaling of the raw ECMWF clearly adds value to extreme precipitation forecasts by increasing the forecast skill and predictability to two- and three- week lead times, especially over the central Red Sea, Jeddah, and the west coast of AP. Increasing the ensemble member size is unlikely to change this essential conclusion.

As mentioned in subsection 3d, the rain gauge measurement dataset on which the satellite products are evaluated is limited both temporally (2000–2012) and spatially (43 sites sparsely distributed over the AP). This dataset includes only 11 events for our study, preventing a statistical analysis of the site-equivalent model precipitation relative to the rain gauge measurements. Moreover, the GPMF precipitation product has a large RMSE and bias relative to the rain gauge measurements, which poses an accuracy problem. Many studies (e.g., Kim et al. 2017; Tan et al. 2017; Zhang et al. 2018) have demonstrated that the GPMF better obtains the daily precipitation and its spatial distribution than the Tropical Rainfall Measuring Mission (TRMM) and other satellite precipitation products, but only for GPMF precipitation data collected after 2014. The dataset prior to 2014 is based on the TRMM Combined Ku Radar–Radiometer Algorithm (TRMM CORRA) product with fewer microwave channels and radar data than the GPM CORRA. The use of earlier GPMF data might lower the accuracy of the precipitation histogram and affect the mean rate (Huffman 2019). Therefore, our verification metrics and ROC curves might be slightly less accurate for the events prior to 2014 than for those after 2014. This problem reinforces the need for evaluating the model precipitation against the 5 km-resolution KAUST-RA product. As the results of the GPMF and KAUST-RA products were qualitatively the same, we are confident that the value added by our CPM is true, and is unaffected by the vagaries of the observed precipitation products used in the analysis.

In this work, the 20-mm threshold was considered as the extreme precipitation threshold for the AP, especially as desert soil does not easily soak up water (Almazroui 2011). We found that the forecast skill of CP-WRF over the central Red Sea and near Jeddah still outperformed that of raw ECMWF even when the precipitation thresholds are set to 30 and 35 mm (data not shown). Again, there is potential improvement for sub-seasonal precipitation forecasts by CPM over the AP with a three-week lead time. This improvement is possible for extratropical events.

SST plays an important modulating role on the weather in this region (Hoteit et al. 2021). The drier conditions in the central Red Sea at W2 and W3 during the tropical events were probably incurred by the specification of the modeled SST, which does not change significantly at sub-seasonal time scales. Low atmospheric instability is unfavorable for convective organization. The SST can influence extreme precipitation by up to 20% (Orth and Seneviratne 2017) and a cooler SST can suppress convective organization, especially in gulf-type water bodies (e.g., Mitchell et al. 2002). An SST bias correction might alleviate this problem, which affected the precipitation forecast over the Red Sea. The benefit of SST bias correction for convective organization has been demonstrated in a dynamically-downscaled climate forecast reanalysis (Carrillo et al. 2017). We note that the SST in our simulations was extracted from the ECMWF S2S reforecast dataset, and was not bias-corrected beyond the time scale of medium-range weather forecasts (15 days). Thus, during tropical events when the strong RST brings warm tropical air masses to the Red Sea region (Luong et al. 2020a), the CP-WRF probably adjusted the SST inadequately and suppressed the convective organization. A more dynamic treatment of the SST in the Red Sea should be investigated in future CPM simulation studies.

Finally, given its potential to improve extreme precipitation forecasts at sub-seasonal time scales, our dynamically-downscaled modeling approach can broaden the range of applications in the region. One potential application is MCS tracking based on the modeled cloud-top temperature and precipitation on sub-seasonal time scales. Feng et al. (2012) demonstrated MCS tracking with a precipitation-feature-based algorithm that inputs the satellite infrared-brightness temperature data and uses some pre-defined parameters. The same algorithm could be applied to MSC tracking over the AP using the CP-WRF model output at sub-seasonal time scales. The results would approximately identify the areas at highest risk of MCS-driven precipitation with one- to three-week lead times, providing early warnings of extreme weather events within an object-based context.

## Summary And Conclusions

Interest in sub-seasonal weather prediction has grown in recent years because it can potentially provide early extreme weather warnings beyond the weather forecast timescale. Ultimately, we hope to improve the preparedness capability of detecting extreme weather events over the AP, where occasional heavy precipitation in the cool season causes substantial loss of life and damage to infrastructure. Applying CPM, we here demonstrated that sub-seasonal forecasts of extreme precipitation over the AP can be potentially improved during the cool season (November–March). We identified the regions and periods of forecast opportunity and investigated the predictability of synoptic-scale forcing at sub-seasonal time scales.

Using the ECMWF S2S reforecast dataset as the lateral boundary condition of the CPM, we simulated 10 extreme precipitation events associated with an extratropical synoptic regime and eight extreme precipitation events associated with a tropical synoptic regime in Jeddah from 1999 to 2018. The results of the extratropical events were evaluated relative to the GPMF and KAUST-RA precipitation products based on three categorical statistics: CSI, POD, and FAR. The CP-WRF exhibited a high precipitation

forecast skill at lead time W1 over the Red Sea, Jeddah, the central AP, and the Persian Gulf. At W2 and W3, the high forecast skill was maintained mainly around Jeddah, the central Red Sea, and along the west coast of the AP. The CSI, POD, and some of the FAR results were field-significant. For the tropical events, the precipitation forecast skill of CP-WRF was low over the Jeddah area and the central Red Sea, but high over the central AP at W1. The precipitation forecast skill over this area remained low at W2, but was high in the central and northern AP. At W3, the forecast skill of CP-WRF was high near Jeddah and the central Red Sea.

This high forecast skill of the CP-WRF was consistently reflected in the ROC curve analysis. At W1, the ROC curves of the extratropical and tropical events obtained by CP-WRF (unlike those of the raw ECMWF) were above the no-skill diagonal (AUC = 0.5) at W1, with AUC values greater than 0.59. At W2 and W3, the forecast improvement of the CP-WRF was also notable. The ROC curves of more than 50% of the events simulated in the CP-WRF were above the no-skill diagonal. The average differences, obtained by subtracting the AUC of the raw ECMWF from the AUC of the CP-WRF, were all positive. Clearly, the CP-WRF can potentially improve the forecasting of extreme precipitation events associated with both extratropical and tropical synoptic regimes over Jeddah and the central Red Sea during the cool season, with lead times up to three weeks.

The correlation coefficients between the raw ECMWF and ERA-I at the 500-hPa geopotential heights were higher for the extratropical events than for the tropical events at W1, W2, and W3. Over Jeddah and the west coast of AP, the correlation coefficients were also higher for the extratropical than for the tropical events at W3. Moreover, the 500-hPa geopotential contours of the raw ECMWF more closely matched those of the ERA-Interim for the extratropical events than the tropical events. In contrast, the same analysis on the 850-hPa geopotential heights showed higher correlation coefficients at W2 and W3 for the tropical events than for the extratropical events. At W3, the correlations were much lower ( $< 0.1$ ) in the extratropical events than in the tropical events. Considering the high uncertainty in the 850-hPa geopotential height forecasts at sub-seasonal time scales and the analysis results of the categorical statistics and ROC, we concluded that the extreme precipitation events associated with the extratropical synoptic regime are more predictable than those associated with the tropical synoptic regime at sub-seasonal time scales.

Although this study was constrained by the low number of events and ensemble size, it demonstrates the potential improvement in forecasting extreme cool-season precipitation over the AP at sub-seasonal time scales. If this potential is realized, we could establish a forecast system for extreme weather warnings beyond the weather forecast timescales over the AP. Moreover, this S2S downscaling technique is expected to be applicable in other parts of the world, especially regions with similar climatic characteristics to the AP such as Chile, the southwest United States, and northwestern Mexico. It should also be useful in regions subjected to increasing frequency of extreme precipitation events driven by climate change. Issuing extreme weather warning with three- to four-week lead times could prevent casualties and socioeconomic loss.

A long-term dynamically-downscaled CPM reforecast is imminent. We will generate a 20-year database of sub-seasonal reforecasts from 1999 to 2020 at the same convective-permitting resolution. This database will include more cool-season precipitation events than the present dataset, and an additional four-week lead time for event forecasting. Increasing the number of precipitation events will improve the sample size in our statistical analyses. We expect to more robustly demonstrate the potential improvement of sub-seasonal forecasts, extending the lead time to four weeks, in a future analysis.

## Declarations

### Acknowledgment

This work is supported by a competitive research grant from the King Abdullah University of Science and Technology with sub-award agreement OSR-2018-CRG7-3706.2.

### Data Availability

All the data used in this study are stored on the Shaheen supercomputer at the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. The data are available upon request.

## References

1. Alkhalaf A, Basset H (2013) Diagnostic study of a severe thunderstorm over Jeddah. *Atmos Clim Sci* 3:150–164. <https://doi.org/10.4236/ACS.2013.31017>
2. Almazroui M (2011) Calibration of TRMM rainfall climatology over Saudi Arabia during 1998–2009. *Atmos Res* 99:400–414. <https://doi.org/10.1016/j.atmosres.2010.11.006>
3. Al Saud M (2010) Assessment of flood hazard of Jeddah area 2009, Saudi Arabia. *J Water Resour Prot* 2:839–847. <https://doi.org/10.4236/jwarp.2010.29099>
4. Atlas (1984) Water atlas of Saudi Arabia. Water Resource Department, Ministry of Agriculture and Water, Riyadh, Saudi Arabia, pp. 2–19
5. Ban N, Schmidli J, Schär C (2014) Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *J Geophys Res Atmos* 119:7889–7907. <https://doi.org/10.1002/2014JD021478>
6. Becker EJ (2017) Prediction of short-term climate extremes with a multimodel ensemble. *Climate Extremes* (eds S.-Y.S. Wang, J.-H. Yoon, C.C. Funk and R.R. Gillies). <https://doi.org/10.1002/9781119068020.ch21>
7. Berrisford P, Dee D, Poli P, Brugge R, Fielding K, Fuentes M, Kallberg P, Kobayashi S, Uppala S, Simmons A (2011) The ERA-Interim archive, version 2.0. ERA report series. 1. Technical Report. ECMWF, pp23
8. Castro CL, Chang H, Dominguez F, Carrillo C, Schemm J, Juang H (2012) Can a Regional Climate Model Improve the Ability to Forecast the North American Monsoon? *J Clim* 25:8212–8237.

<https://doi.org/10.1175/JCLI-D-11-00441.1>

9. Carrillo CM, Castro CL, Chang HI, Luong TM (2017) Multi-year climate variability in the Southwestern United States within a context of a dynamically downscaled twentieth century reanalysis. *Clim Dyn* 49:4217–4236. <https://doi.org/10.1007/s00382-017-3569-1>
10. Christidis N, Stott PA (2015) Changes in the geopotential height at 500 hPa under the influence of external climatic forcings. *Geophys Res Lett* 42:10798–10806. <https://doi.org/10.1002/2015GL066669>
11. Clark P, Roberts N, Lean H, Ballard SP, Charlton-Perez C (2016) Convection-permitting models: a step-change in rainfall forecasting. *Met Apps* 23:165–181. <https://doi.org/10.1002/met.1538>
12. Coelho CAS, Firpo MAF, de Andrade FM (2018) A verification framework for South American sub-seasonal precipitation predictions. *Meteorol Z* 27:503–520. <https://doi.org/10.1127/metz/2018/0898>
13. Coelho CAS, Brown B, Wilson L, Mittermaier M, Casati B (2019) Forecast verification for S2S timescales. Sub-seasonal to seasonal prediction, the gap between weather and climate forecasting. ed: A. W. Robertson and F. Vitart, p 569
14. Collier JC, Zhang GJ (2007) Effects of increased horizontal resolution on simulation of the North American monsoon in the NCAR CAM3: An evaluation based on surface, satellite, and reanalysis data. *J Clim* 20:1843–1861
15. Dasari HP, Attada R, Knio O, Hoteit I (2017) Analysis of a severe weather event over Mecca, Kingdom of Saudi Arabia, using observations and high-resolution modelling. *Met Apps* 24:612–627. <https://doi.org/10.1002/met.1662>
16. Dasari HP, Langodan S, Viswanadhapalli Y, Vadlamudi BR, Papadopoulos VP, Hoteit I (2018) ENSO influence on the interannual variability of the Red Sea convergence zone and associated rainfall. *Int J Climatol* 38:761–775. <https://doi.org/10.1002/joc.5208>
17. Dasari HP, Desamsetti S, Langodan S, Attada R, Kunchala RK, Viswanadhapalli Y, Knio O, Hoteit I (2019) High-resolution assessment of solar energy resources over the Arabian Peninsula. *Appl Energy* 248:354–371. <https://doi.org/10.1016/j.apenergy.2019.04.105>
18. Davis SR, Pratt LJ, Jiang H (2015) The Tokar gap jet: Regional circulation, diurnal variability, and moisture transport based on numerical simulations. *J Clim* 28:5885–5907. <https://doi.org/10.1175/JCLI-D-14-00635.1>
19. De Haan LL, Kanamitsu M, De Sales F, Sun L (2015) An evaluation of the seasonal added value of downscaling over the United States using new verification measures. *Theor Appl Climatol* 122:47–57. <https://doi.org/10.1007/s00704-014-1278-9>
20. de Vries AJ, Tyrlis E, Edry D, Krichak SO, Steil B, Lelieveld J (2013) Extreme precipitation events in the Middle East: Dynamics of the active Red Sea trough. *J Geophys Res Atmos* 118:7087–7108. <https://doi.org/10.1002/jgrd.50569>
21. de Vries AJ, Feldstein SB, Reimer M, Tyrlis E, Sprenger M, Baumgart M, Fnais M, Lelieveld J (2016) Dynamics of tropical-extratropical interactions and extreme precipitation events in Saudi Arabia in

- autumn, winter and spring. *Q J R Meteorol Soc* 142:1862–1880. <https://doi.org/10.1002/qj.2781>
22. de Vries AJ, Ouwersloot HG, Feldstein SB, Riemer M, El Kenawy AM, McCabe MF, Lelieveld J (2018) Identification of tropical-extratropical interactions and extreme precipitation events in the Middle East based on potential vorticity and moisture transport. *J Geophys Res Atmos* 123:861–881. <https://doi.org/10.1002/2017JD027587>
  23. Deng L, McCabe MF, Stenchikov G, Evans JP, Kucera PA (2015) Simulation of flash-flood-producing storm events in Saudi Arabia using the Weather Research and Forecasting model. *J Hydrometeorol* 16:615–630. <https://doi.org/10.1175/JHM-D-14-0126.1>
  24. Feng Z, Dong X, Xi B, McFarlane SA, Kennedy A, Lin B, Minnis P (2012) Life cycle of midlatitude deep convective systems in a Lagrangian framework. *J Geophys Res* 117:D23201. <https://doi.org/10.1029/2012JD018362>
  25. Gao Y, Leung LR, Zhao C, Hagos S (2017) Sensitivity of U.S. summer precipitation to model resolution and convective parameterizations across gray zone resolutions, *J Geophys Res Atmos* 122, 2714–2733. <https://doi.org/10.1002/2016JD025896>
  26. Grell GA, Freitas SR (2013) A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos Chem Phys Discuss* 13:23845–23893. <https://doi.org/10.5194/acpd-13-23845-2013>
  27. Grimm AM, Hakoyama LR, Scheibe LA (2021) Active and break phases of the South American summer monsoon: MJO influence and subseasonal prediction. *Clim Dyn* 56:3603–3624. <https://doi.org/10.1007/s00382-021-05658-3>
  28. Hafez YY, Almazroui M (2016) Study of the relationship between geopotential height anomaly over Europe and extreme abnormal weather over the Eastern Mediterranean and Middle East during December 2013. *Arab J Geosci* 9:409. <https://doi.org/10.1007/s12517-016-2424-8>
  29. Haggag M, El-Badry H (2013) Mesoscale numerical study of quasi-stationary convective system over Jeddah in November 2009. *Atmos Clim Sci* 3:73–86. <https://doi.org/10.4236/acs.2013.31010>
  30. Hewitson BC, Crane RG (2002) Self-organizing maps: application to synoptic climatology. *Clim Res* 22:13–26
  31. Hoteit I, Abualnaja Y, Afzal S, Ait-El-Fquih B, Akylas T, Antony C, Dawson C, Asfahani K, Brewin RJ, Cavaleri L, and Coauthors (2021) Towards an end-to-end analysis and prediction system for weather, climate, and marine applications in the Red Sea. *Bull Am Meteorol Soc* 102:E99–E122. <https://doi.org/10.1175/BAMS-D-19-0005.1>
  32. Houze RA (1993) *Cloud Dynamics*. Academic Press, San Diego, p 573
  33. Hudson D, Alves O, Hendon HH, Marshall AG (2011) Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. *Q J R Meteorol Soc* 137:673–689. <https://doi.org/10.1002/qj.769>
  34. Huffman GJ, and Coauthors (2018) Algorithm theoretical basis document version 5.2. NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG), pp 31, [https://gpm.nasa.gov/sites/default/files/document\\_files/IMERG\\_ATBD\\_V5.2\\_0.pdf](https://gpm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V5.2_0.pdf)

35. Huffman GJ (2019) Caveats for IMERG and the TRMM era. Technical notes.  
[https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/IMERGV06\\_TRMMera-caveats.pdf](https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/IMERGV06_TRMMera-caveats.pdf)
36. Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD (2008) Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J Geophys Res* 113:D13103. <https://doi.org/10.1029/2008JD009944>
37. Jie W, Vitart F, Wu T, Liu X (2017) Simulations of the Asian summer monsoon in the sub-seasonal to seasonal prediction project (S2S) database. *Q J R Meteorol Soc* 143:2282–2295.  
<https://doi.org/10.1002/qj.3085>
38. Joyce RJ, Janowiak JE, Arkin PA, Xie P (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *J Hydrometeorol* 5:487–503. [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2)
39. Kendon EJ, Roberts NM, Senior CA, Roberts MJ (2012) Realism of rainfall in a very high-resolution regional climate model. *J Clim* 25:5791–5806. <https://doi.org/10.1175/JCLI-D-11-00562.1>
40. Kendon EJ, Ban N, Roberts NM, Fowler HJ, Roberts MJ, Chan SC, Evans JP, Fosser G, Wilkinson JM (2017) Do convection-permitting regional climate models improve projections of future precipitation change? *Bull Am Meteorol Soc* 98:79–93. <https://doi.org/10.1175/BAMS-D-15-0004.1>
41. Kim K, Park J, Baik J, Choi M (2017) Evaluation of topographical and seasonal feature using GPM IMERG and TRMM 3B42 over Far-East Asia. *Atmos Res* 187:95–105
42. Köppen W (1936) Das geographische System de Klimate. In: Köppen W, Geiger R (eds) *Handbuch de Klimatologie*, vol 1. GebrBorntreger, Berlin, pp 1–44
43. Kubota T, Shige S, Hashizume H, Aonashi K, Takahashi N, Seto S, Hirose M, Takayabu YN, Nakagawa K, Iwanami K, Ushio T, Kachi M, Okamoto K (2007) Global precipitation map using satellite borne microwave radiometers by the GSMaP Project: Production and validation. *IEEE Trans Geosci Remote Sens* 45:2259–2275. <https://doi.org/10.1109/TGRS.2007.895337>
44. Langodan S, Cavaleri L, Viswanadhapalli Y, Hoteit I (2014) The Red Sea: A natural laboratory for wind and wave modeling. *J Phys Oceano* 44:3139–3159. <https://doi.org/10.1175/JPO-D-13-0242.1>
45. Langodan S, Cavaleri L, Vishwanadhapalli Y, Pomaro A, Bertotti L, Hoteit I (2017) The climatology of the Red Sea – part 1: the wind. *Int J Climatol* 37:4509–4517. <https://doi.org/10.1002/joc.5103>
46. Liang P, Lin H (2018) Sub-seasonal prediction over East Asia during boreal summer using ECCO monthly forecasting system. *Clim Dyn* 50:1007–1022. <https://doi.org/10.1007/s00382-017-3658-1>
47. Livezey RE, Chen WY (1983) Statistical significance and its determination by Monte Carlo techniques. *Mon Weather Rev* 111:46–59. [https://doi.org/10.1175/1520-0493\(1983\)111<0046:SFSaid>2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111<0046:SFSaid>2.0.CO;2)
48. Luong TM, Dasari HP, Hoteit I (2020a) Extreme precipitation events are becoming less frequent but more intense over Jeddah, Saudi Arabia. Are shifting weather regimes the cause? *Atmos Sci Lett* 21:e981. <https://doi.org/10.1002/asl.981>

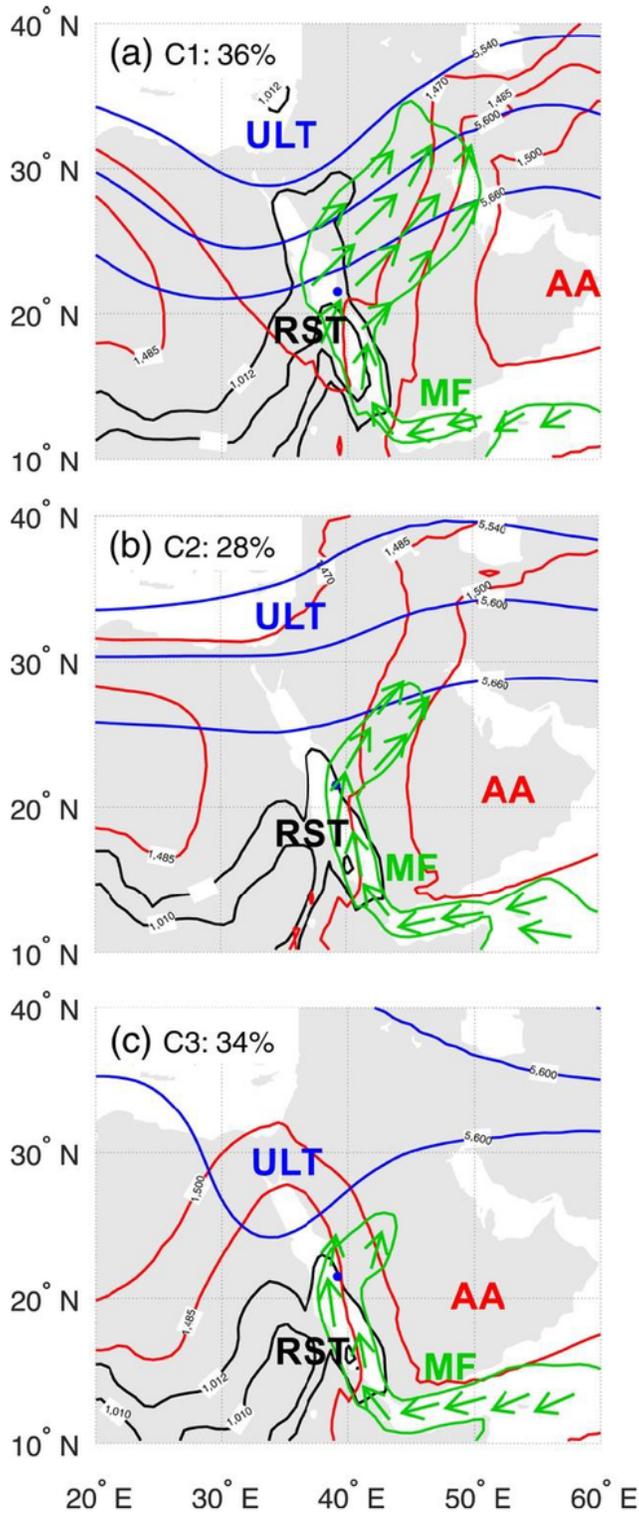
49. Luong TM, Dasari HP, Hoteit I (2020b) Impact of urbanization on the simulation of extreme rainfall in the city of Jeddah, Saudi Arabia. *J Appl Meteor Climatol* 59:953–971. <https://doi.org/10.1175/JAMC-D-19-0257.1>
50. Mariotti A, Baggett C, Barnes EA, Becker E, Butler A, Collins DC, Dirmeyer PA, Ferranti L, Johnson NC, Jones J, Coauthors (2020) Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull Am Meteorol Soc* 101:E608–E625. <https://doi.org/10.1175/BAMS-D-18-0326.1>
51. Mason SJ, Graham NE (2002) Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Q J R Meteorol Soc* 128:2145–2166
52. Merryfield WJ, Baehr J, Batté L, Becker EJ, Butler AH, Coelho CAS, Danabasoglu G, Dirmeyer PA, Doblas-Reyes FJ, Domeisen DIV, Coauthors (2020) Current and emerging developments in subseasonal to decadal prediction. *Bull Am Meteorol Soc* 101:E869–E896. <https://doi.org/10.1175/BAMS-D-19-0037.1>
53. Mitchell DL, Ivanova D, Rabin R, Brown TJ, Redmond K (2002) Gulf of California sea surface temperatures and the North American Monsoon: Mechanistic implications from observations. *J Clim* 15:2261–2281. [https://doi.org/10.1175/1520-0442\(2002\)015<2261:GOCSSST>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2261:GOCSSST>2.0.CO;2)
54. Moker JM Jr, Castro CL, Arellano AF Jr, Serra YL, Adams DK (2018) Convective-Permitting hindcast simulations during the North American Monsoon GPS transect experiment 2013: Establishing baseline model performance without data assimilation. *J Appl Meteor Climatol* 57:1683–1710. <https://doi.org/10.1175/JAMC-D-17-0136.1>
55. Nakanishi M, Niino H (2006) An Improved Mellor-Yamada level 3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound-Layer Meteor* 119:397–407. <http://dx.doi.org/10.1007/s10546-005-9030-8>
56. Orth R, Seneviratne SI (2017) Variability of soil moisture and sea surface temperatures similarly important for warm-season land climate in the Community Earth System Model. *J Clim* 30:2141–2162. <https://doi.org/10.1175/JCLI-D-15-0567.1>
57. Prein AF, Langhans W, Fosser G, Ferrone A, Ban N, Goergen K, Keller M, Tölle M, Gutjahr O, Feser F, Coauthors (2015) A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev Geophys* 53:323–361. <https://doi.org/10.1002/2014RG000475>
58. Pegion K, Kirtman BP, Becker E, Collins DC, LaJoie E, Burgman R, Bell R, DelSole T, Min D, Zhu Y, Li W, Sinsky E, Guan H, Gottschalck J, Metzger EJ, Barton NP, Achuthavarier D, Marshak J, Koster RD, Lin H, Gagnon N, Bell M, Tippett MK, Robertson AW, Sun S, Benjamin SG, Green BW, Bleck R, Kim H (2019) The subseasonal experiment (SubX): A multimodel subseasonal prediction experiment. *Bull Am Meteorol Soc* 100:2043–2060. <https://doi.org/10.1175/BAMS-D-18-0270.1>
59. Powers JG, Klemp JB, Skamarock WC, Davis CA, Dudhia J, Gill DO, Coen JL, Gochis DJ, Ahmadov R, Peckham SE, Coauthors (2017) The weather research and forecasting model: Overview, system efforts, and future directions. *Bull Am Meteorol Soc* 98:1717–1737. <https://doi.org/10.1175/BAMS-D-15-00308.1>

60. Risanto CB, Castro CL, Arellano AF Jr, Moker JM Jr, Adams DK (2021) The impact of assimilating GPS precipitable water vapor in convective-permitting WRF-ARW on North American monsoon precipitation forecasts over Northwest Mexico. *Mon Wea Rev.* <https://doi.org/10.1175/MWR-D-20-0394.1>
61. Robertson AW, Vitart F, Camargo SJ (2020) Subseasonal to seasonal prediction of weather to climate with application to tropical cyclones. *J Geophys Res* 125:e2018JD029375. <https://doi.org/10.1029/2018JD029375>
62. Schumacher RS, Rasmussen KL (2020) The formation, character and changing nature of mesoscale convective systems. *Nat Rev Earth Environ* 1:300–314. <https://doi.org/10.1038/s43017-020-0057-7>
63. Shukla S, Lettenmaier DP (2013) Multi-RCM ensemble downscaling of NCEP CFS winter season forecasts: Implications for seasonal hydrologic forecast skill. *J Geophys Res Atmos* 118:10770–10790. <https://doi.org/10.1002/jgrd.50628>
64. Skamarock WC, Coauthors (2008) A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475 + STR, pp 113, <http://dx.doi.org/10.5065/D68S4MVH>
65. Smith AB, Katz RW (2013) US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Nat Hazards* 67:387–410. <https://doi.org/10.1007/s11069-013-0566-5>
66. Sorooshian S, Hsu KL, Gao X, Gupta HV, Imam B, Braithwaite D (2002) Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull Am Meteorol Soc* 81:2035–2046. [https://doi.org/10.1175/1520-0477\(2000\)081<2035:EOPSSE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2)
67. Tan ML, Duan Z (2017) Assessment of GPM and TRMM precipitation products over Singapore. *Remote Sens* 9:720. <https://doi.org/10.3390/rs9070720>
68. Tewari M, Coauthors (2004) Implementation and verification of the unified Noah land surface model in the WRF Model. 20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction, Seattle, WA, Amer Meteorol Soc 14.2a. [Available online at [https://ams.confex.com/ams/84Annual/techprogram/paper\\_69061.htm](https://ams.confex.com/ams/84Annual/techprogram/paper_69061.htm).]
69. Thompson G, Field PR, Rasmussen RM, Hall WD (2008) Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon Weather Rev* 136:5095–5115. <https://doi.org/10.1175/2008MWR2387.1>
70. Viswanadhapalli Y, Srinivas CV, Langodan S, Hoteit I (2016) Predicting extreme rainfall events over Jeddah, Saudi Arabia: impact of data assimilation with conventional and satellite observations. *Q J R Meteorol Soc* 142:327–348. <https://doi.org/10.1002/qj.2654>
71. Viswanadhapalli Y, Dasari HP, Langodan S, Challa VS, Hoteit I (2017) Climatic features of the Red Sea from a regional assimilative model. *Int J Climatol* 37:2563–2581. <https://doi.org/10.1002/joc.4865>
72. Vitart F, Molteni F (2010) Simulation of the Madden–Julian Oscillation and its teleconnections in the ECMWF forecast system. *Q J R Meteorol Soc* 136:842–855. <https://doi.org/10.1002/qj.623>
73. Vitart F, Robertson AW, S2S Steering Group (2015) Sub-seasonal to seasonal prediction: Linking weather and climate. *Seamless Prediction of the Earth System: From Minutes to Months*, G. Brunet,

Jones S, Ruti PM (Eds), WMO-1156, World Meteorological Organization, 385–401. [Available online at [https://library.wmo.int/doc\\_num.php?explnum\\_id=3546](https://library.wmo.int/doc_num.php?explnum_id=3546).]

74. Vitart F, Ardilouze C, Bonet A, Brookshaw A, Chen M, Codorean C, Déqué M, Ferranti L, Fucile E, Coauthors (2017) The Subseasonal to seasonal (S2S) prediction project database. *Bull Am Meteorol Soc* 98:163–173. <https://doi.org/10.1175/BAMS-D-16-0017.1>
75. Vitart F, Robertson AW (2018) The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Clim Atmos Sci* 1 (3). <https://doi.org/10.1038/s41612-018-0013-0>
76. Vuillaume JF, Dorji S, Komolafe A, Coauthors (2018) Sub-seasonal extreme rainfall prediction in the Kelani River basin of Sri Lanka by using self-organizing map classification. *Nat Hazards* 94:385–404. <https://doi.org/10.1007/s11069-018-3394-9>
77. Wilks DS (2011) *Statistical methods in the Atmospheric Sciences*. Academic Press, pp 676
78. Zhang C, Chen X, Shao H, Chen S, Liu T, Chen C, Ding Q, Du H (2018) Evaluation and intercomparison of high-resolution satellite precipitation estimates—GPM, TRMM, and CMORPH in the Tianshan Mountain Area. *Remote Sens* 10:1543. <https://doi.org/10.3390/rs10101543>
79. Zou KH, O'Malley J, Mauri L (2007) Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 115:654–657. <https://doi.org/10.1161/CIRCULATIONAHA.105.594929>

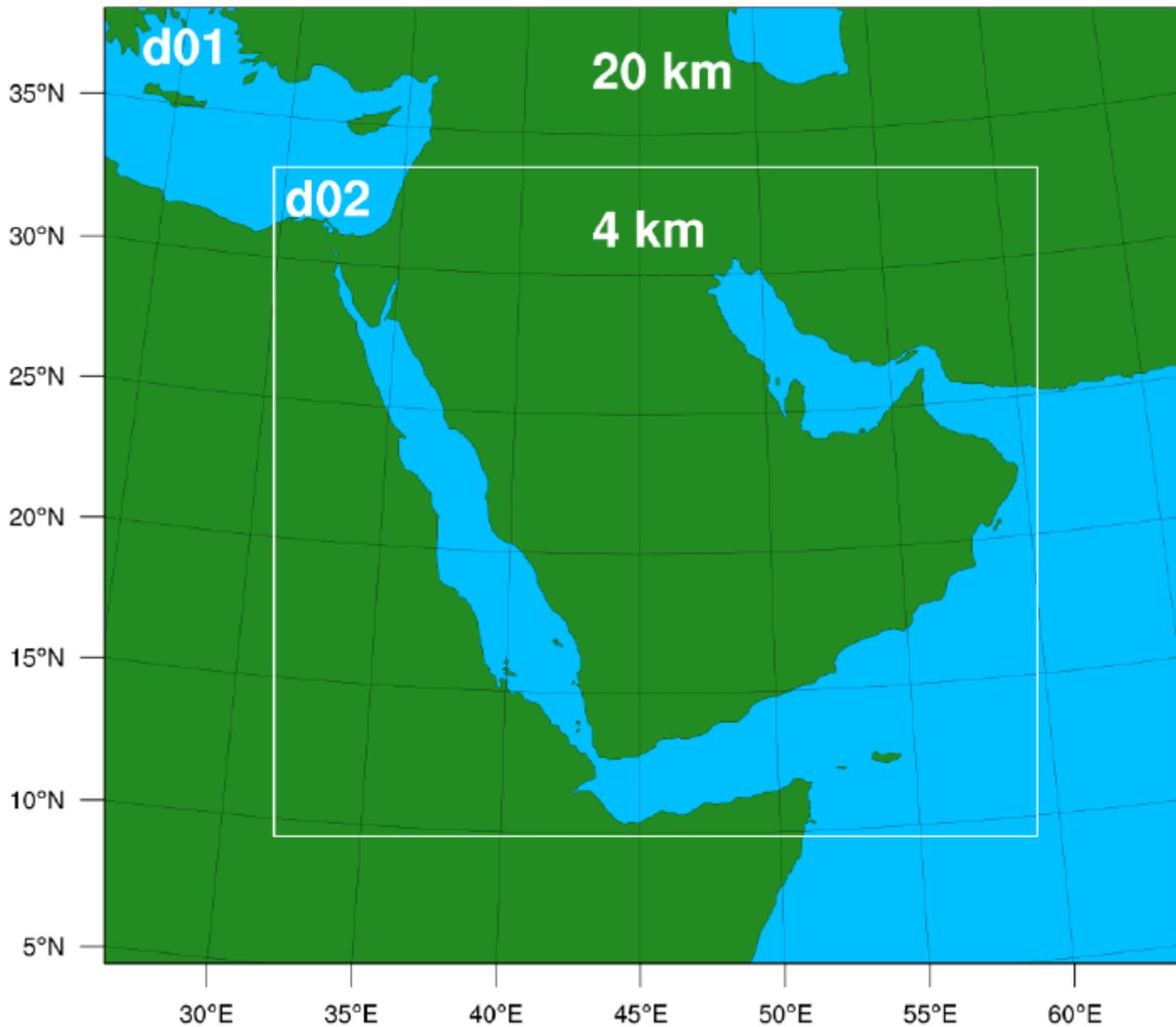
## Figures



**Figure 1**

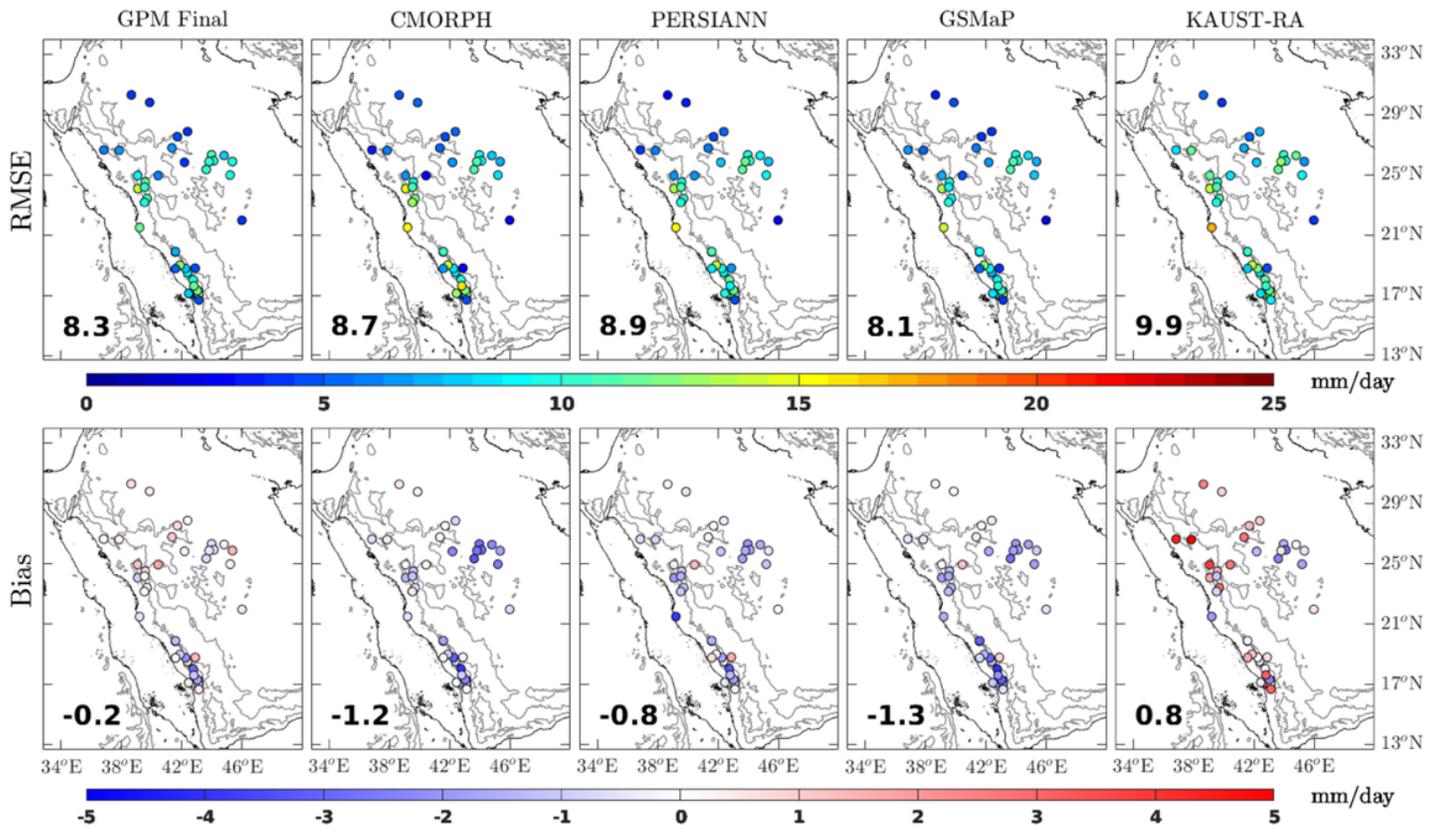
(a–c) Schematics of the large-scale circulation on extreme rainfall days over Jeddah based on the self-organized map (SOM) analysis of clusters 1, 2, and 3 (C1–C3), respectively. Black, red, blue, and green contours represent the mean sea level pressure, geopotential heights at 850 hPa, geopotential heights at 500 hPa, and vertical integral water vapor flux, respectively. The main generators of extreme events in

each cluster are the Red Sea trough (RST), Arabian anticyclone (AA), upper-level trough (ULT), and moisture flux (MF). The blue dot indicates the city of Jeddah. Reprinted from Luong et al. (2020a).



**Figure 2**

Advanced Research version of the Weather Research and Forecasting (WRF-ARW) nested domain configuration. The inner domain (d02) is convective-permitting and allows convective organization at meso- $\gamma$ -scale resolution without a cumulus parametrization scheme.



**Figure 3**

(Upper) Root mean square errors (RMSEs) and (lower) biases of satellite-based precipitation measurements (GPMF, CMORPH, PERSIANN, GSMaP) and a reanalysis product (KAUST-RA) relative to 43 rain gauge precipitation measurements from January of 2000 to December of 2012. The calculation was performed only on days with daily total precipitation exceeding 5 mm (100 days). The value at the bottom left in each panel indicates the average RMSE (upper panels) and bias (lower panels) across the observation sites.

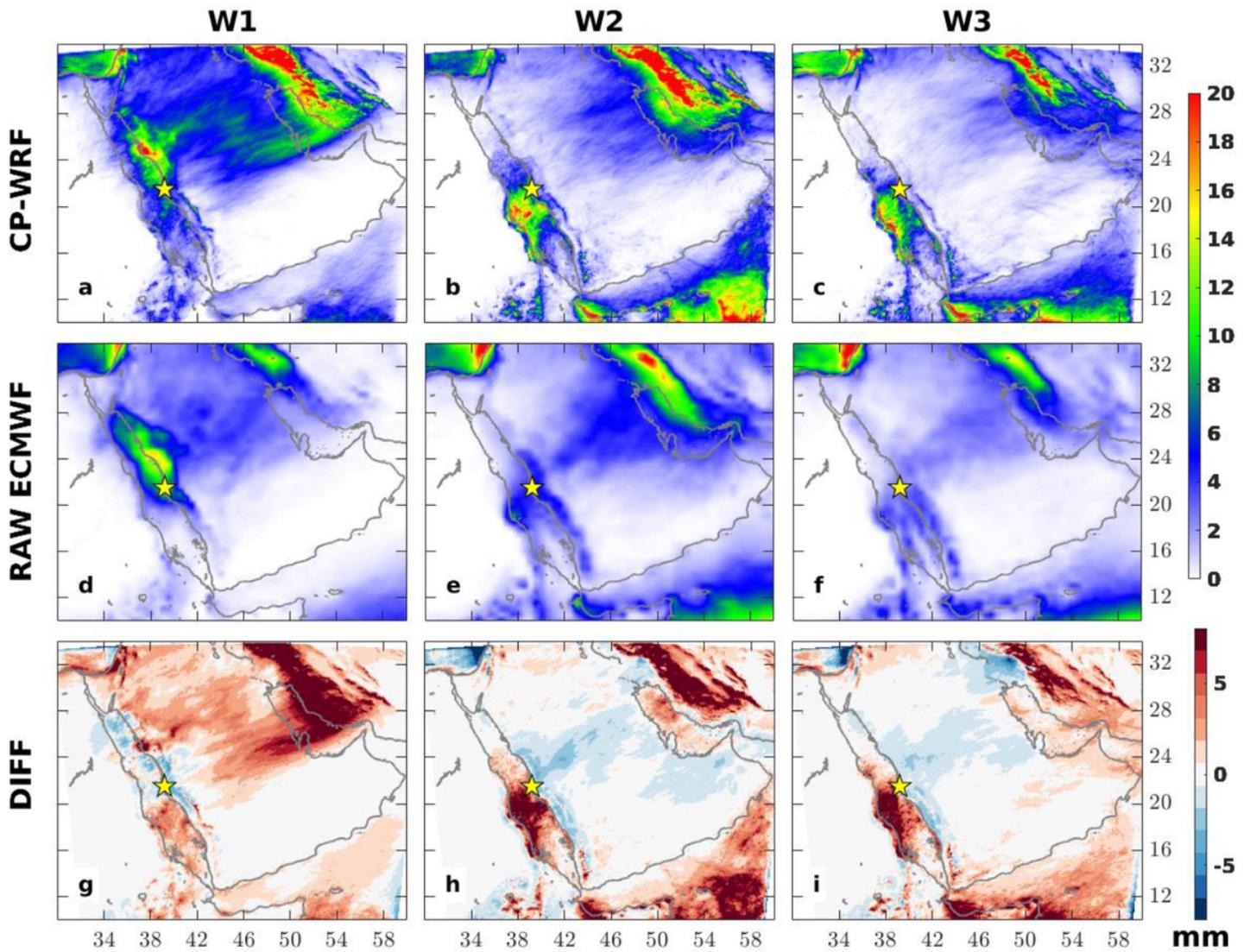


Figure 4

Average total precipitations in the extratropical events obtained by CP-WRF (a–c) and the raw ECMWF (d–f), and their difference (CP-WRF minus raw ECMWF) (g–i). At the 1-week lead time (W1), the precipitation was accumulated for 3 days centered on the day of the event. At the 2- and 3-week lead times (W2 and W3 respectively), the precipitation was accumulated for 7 days centered on the day of the events. The star indicates the location of Jeddah.

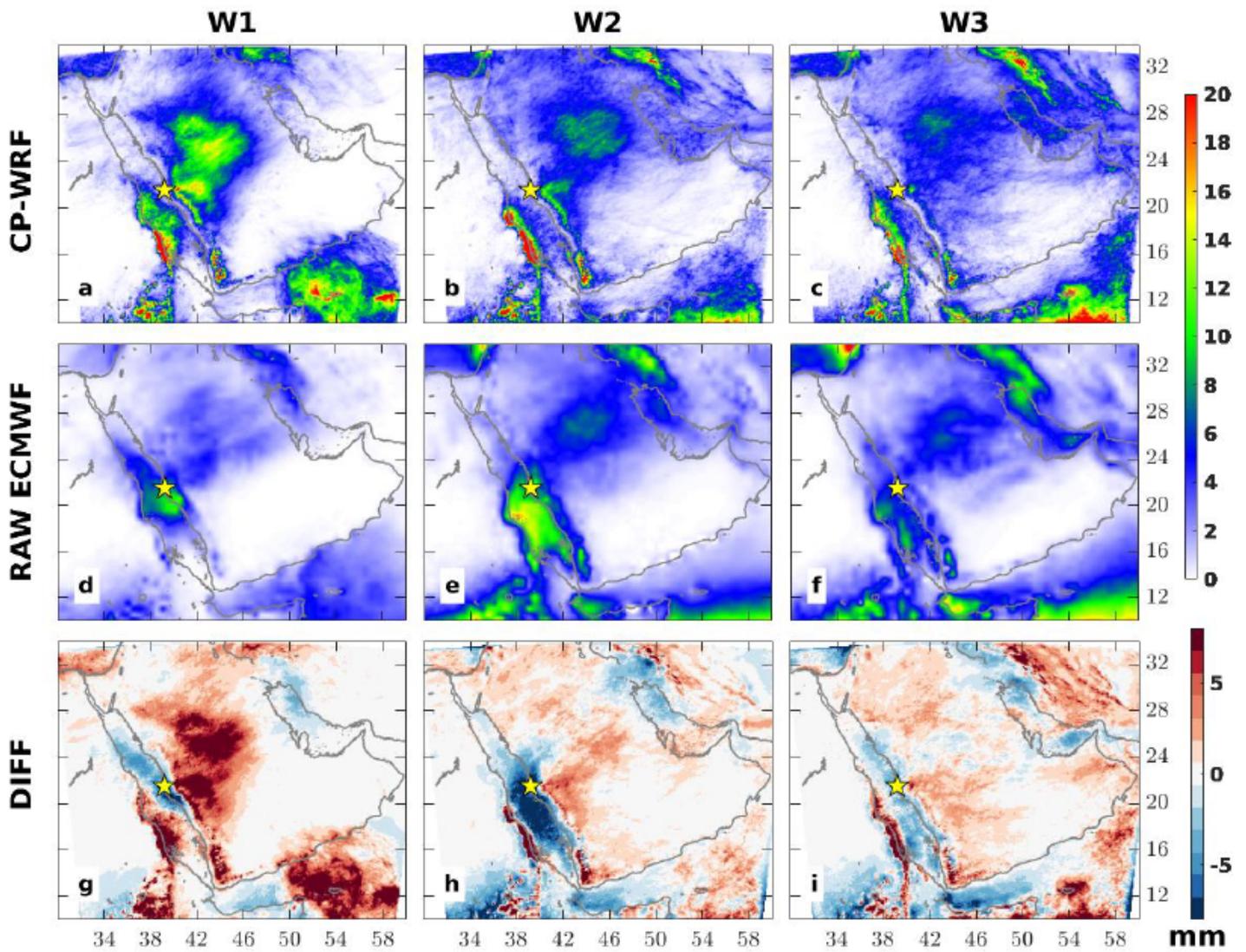


Figure 5

Similar to FIG. 4, but for the tropical events. Note that the CP-WRF does not adequately resolve the precipitation in the central Red Sea at W2 and W3.

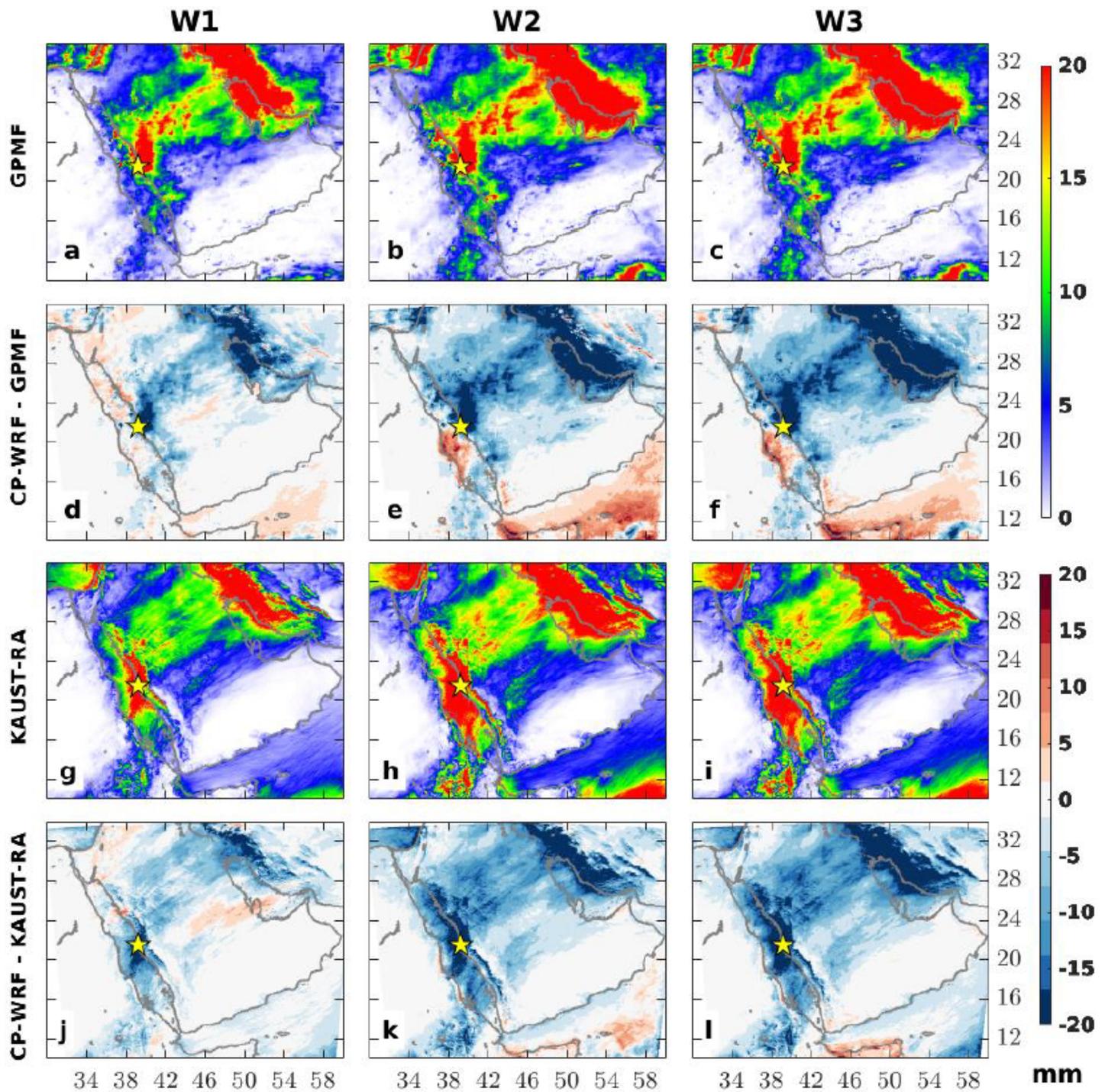


Figure 6

Mean accumulated precipitations of GPMF (a–c), CP-WRF precipitation bias relative to GPMF precipitation (d–f), mean accumulated precipitation total of KAUST-RA (g–i), and CP-WRF precipitation bias relative to KAUST-RA precipitation (j–l) for the extratropical events. At 1-week lead time (W1), the precipitation was accumulated for 3 days centered on the day of the event. At 2- and 3-week lead times (W2 and W3 respectively), the precipitation was accumulated for 7 days centered on the day of the events. The star indicates the location of Jeddah.

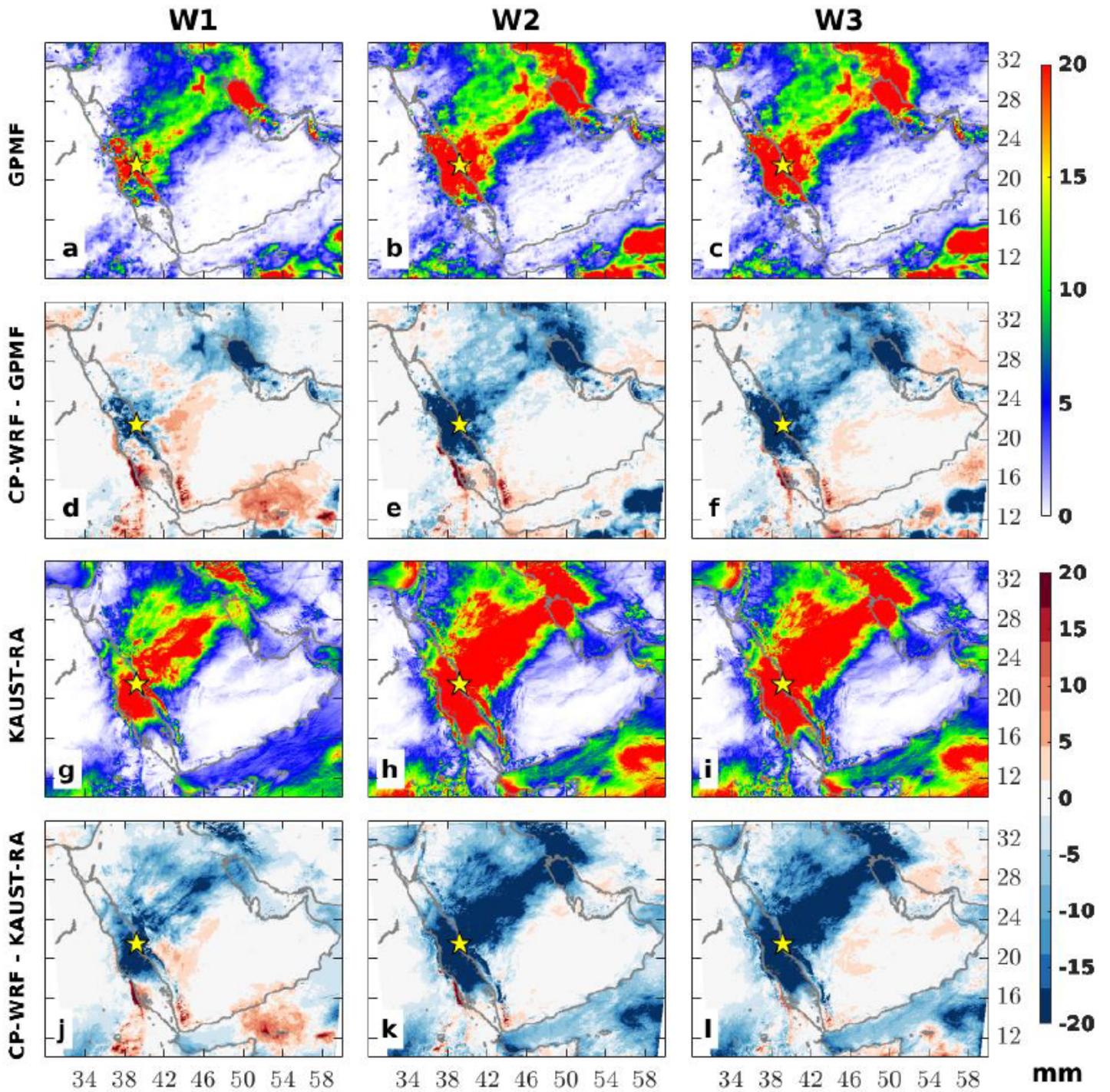


Figure 7

Similar to FIG. 6, but for the tropical events.

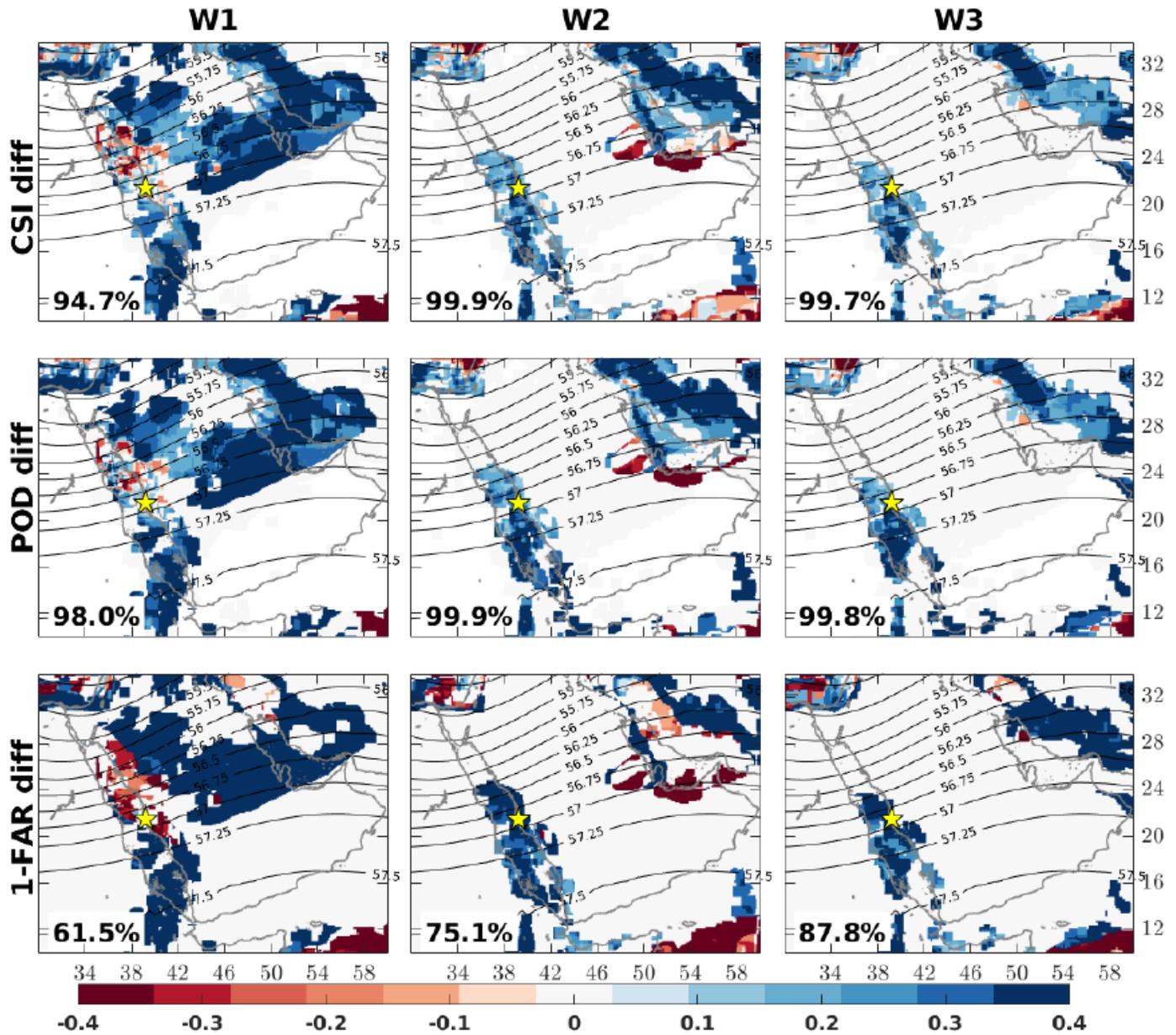


Figure 8

Differences in the CSI, POD, and FAR metrics (relative to the GPMF precipitation) between the CP-WRF and raw ECMWF total precipitations driven by extratropical events at W1, W2, and W3. The total precipitation threshold was set to 20 mm. The blue (red) shading indicates high (low) forecast skill of the CF-WRF. The field-significance values displayed in the lower left of each panel were computed using the Monte Carlo technique with 1000 random permutations. The contour lines are the 500-hPa geopotential heights (in one-tenth of a kilometer) of the extratropical events based on the Luong–SOM analysis. The star indicates the location of Jeddah.

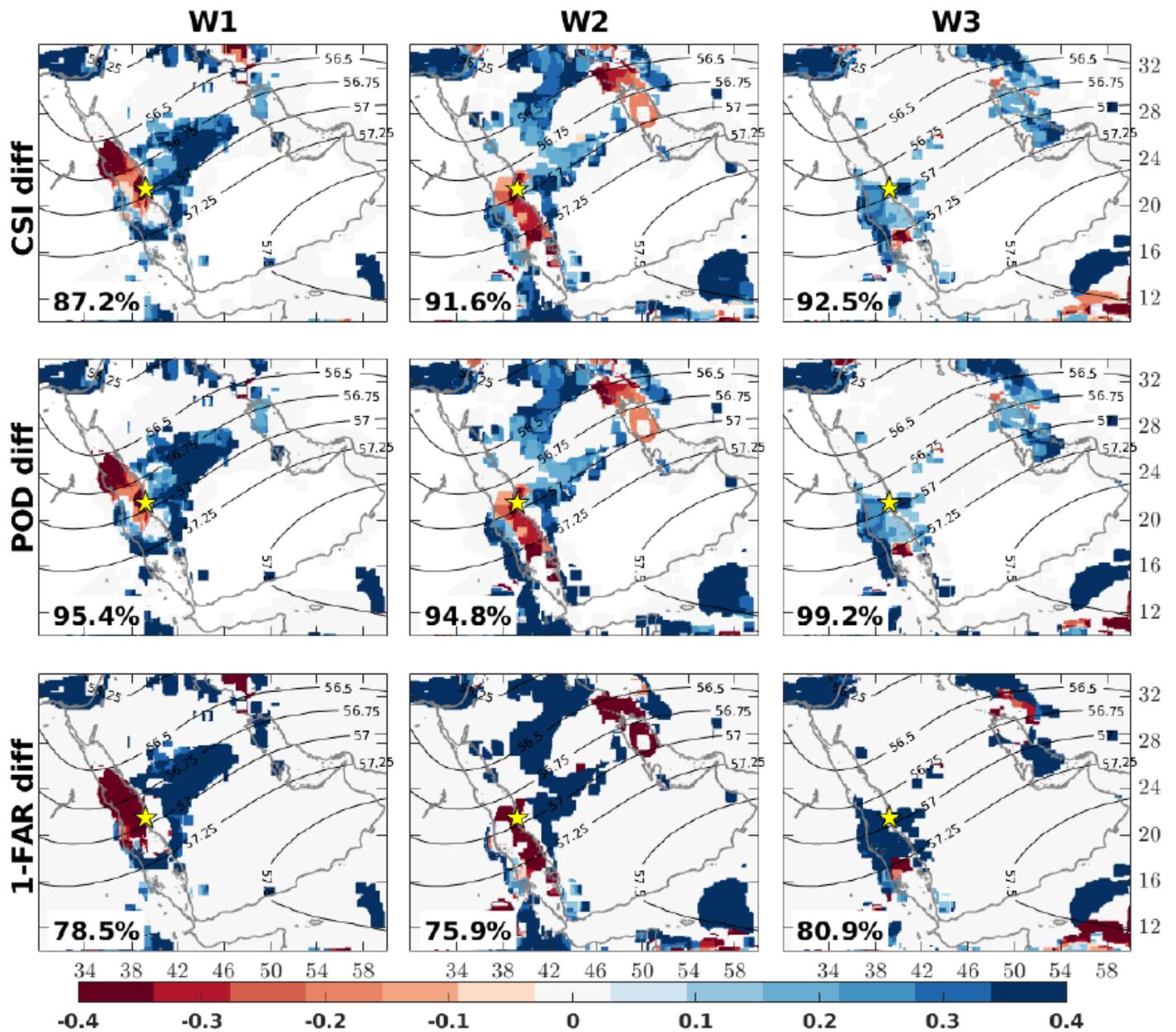


Figure 9

Similar to FIG. 8, but for the tropical events. The contour lines are the 500-hPa geopotential heights (in one-tenth of a kilometer) of the tropical events based on the Luong–SOM analysis. Note that the forecast skill of the CF-WRF was low over the central Red Sea and Jeddah at W1 and W2.

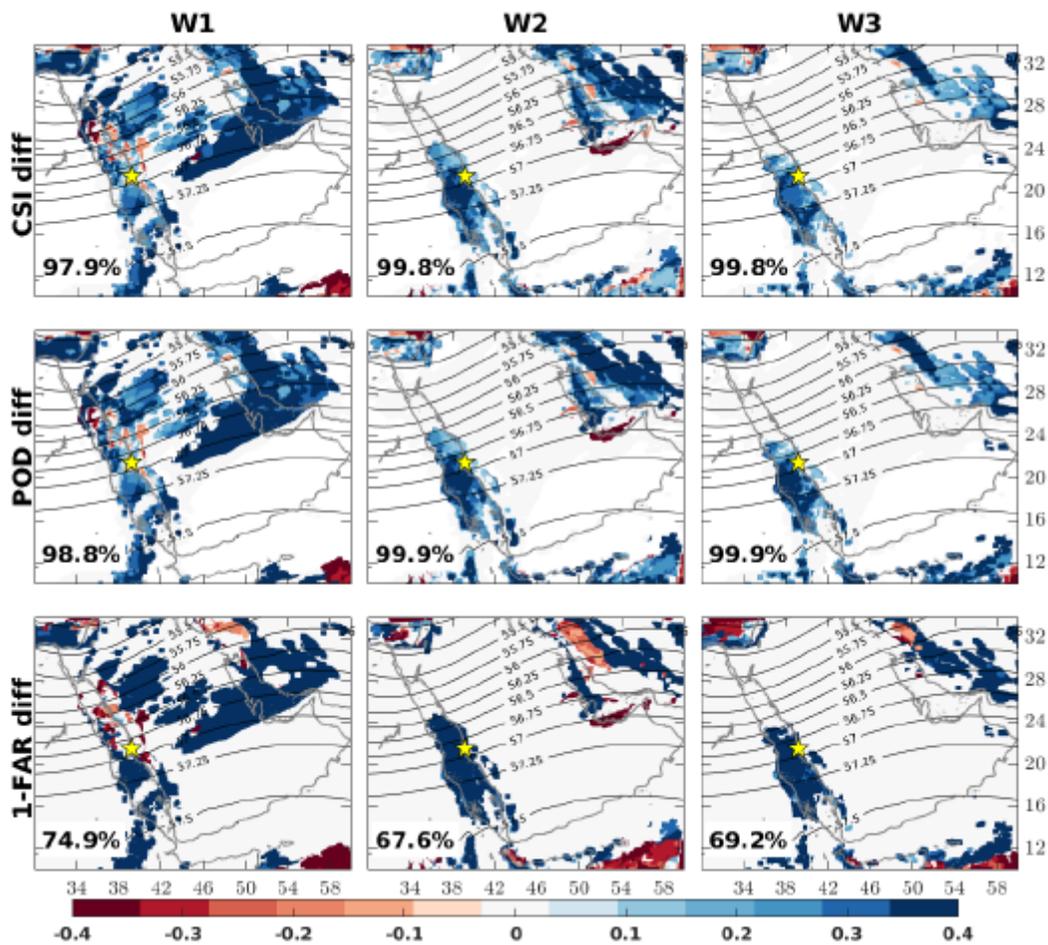


Figure 10

Similar to FIG. 8 but the verification metrics were computed relative to the KAUST-RA precipitation and results are shown for the extratropical events. These results are consistent with those relative to the GPMF precipitation.

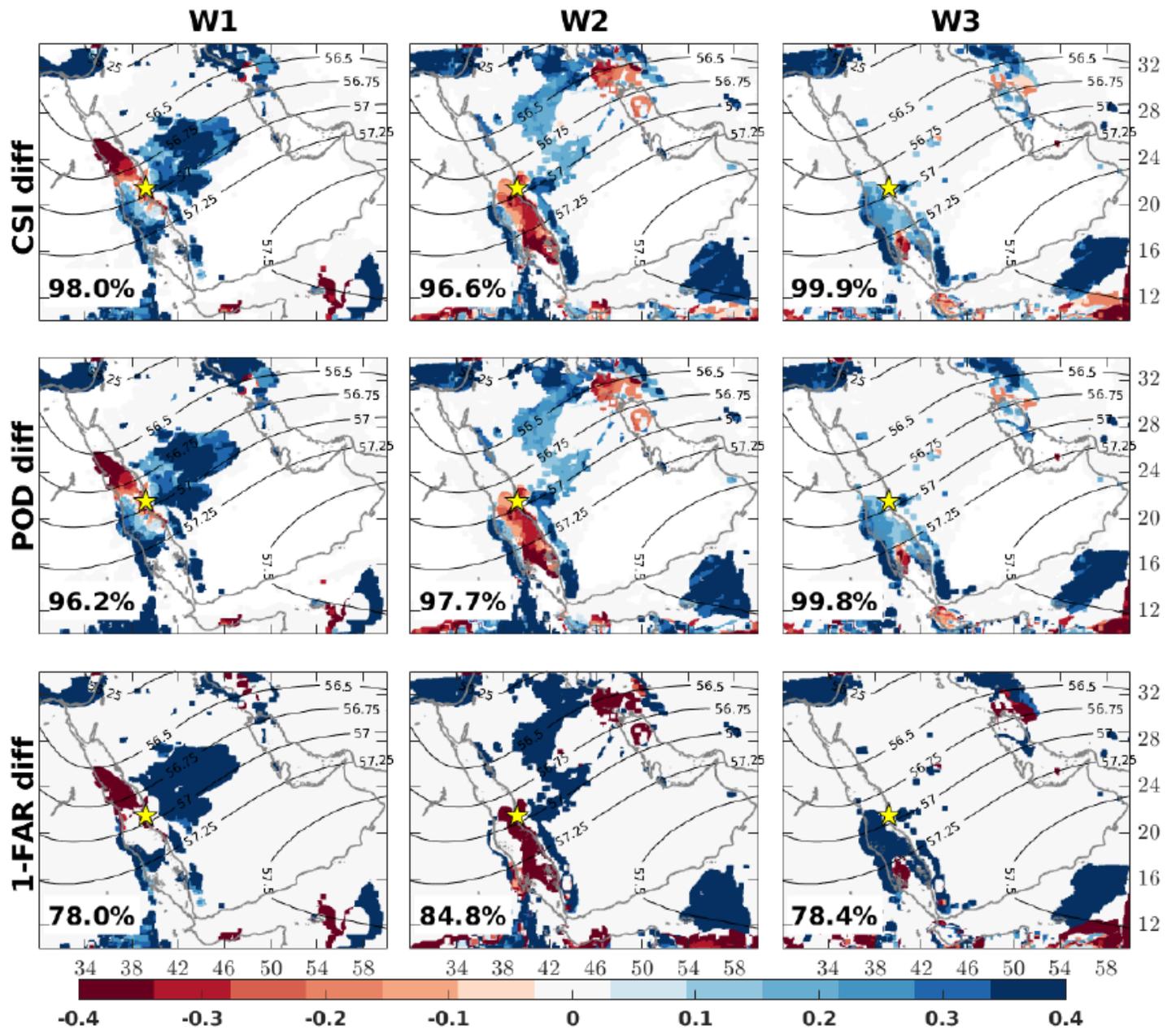


Figure 11

Similar to FIG. 8, but the verification metrics were computed relative to the KAUST-RA precipitation and results are shown for the tropical events. These results are consistent with those relative to the GPMF precipitation.

POD (GPMF) distribution for extratropical – threshold 20 mm

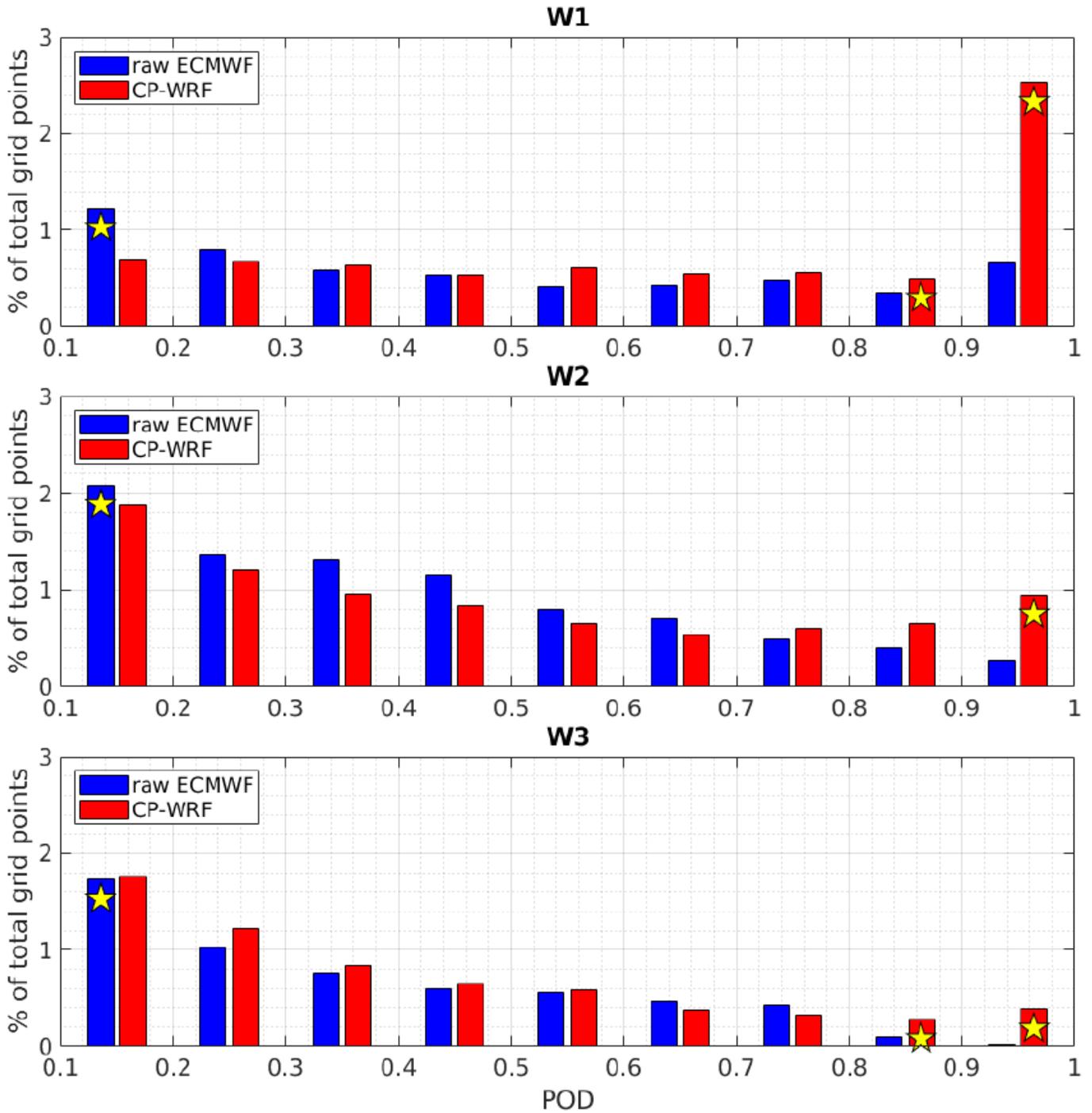


Figure 12

Frequency distributions of POD relative to the GPMF as percentages of total grid points in the domain at W1, W2, and W3 for the extratropical events. Bars marked with stars indicate that the frequency in that CP-WRF bin is significantly different from that of the raw ECMWF (> 90%) based on the Monte Carlo technique with 1000 random permutations. Note that the frequency of PODs > 0.8 is higher in the CP-WRF than in the raw ECMWF at all week lead times.

POD (GPMF) distribution for tropical – threshold 20 mm

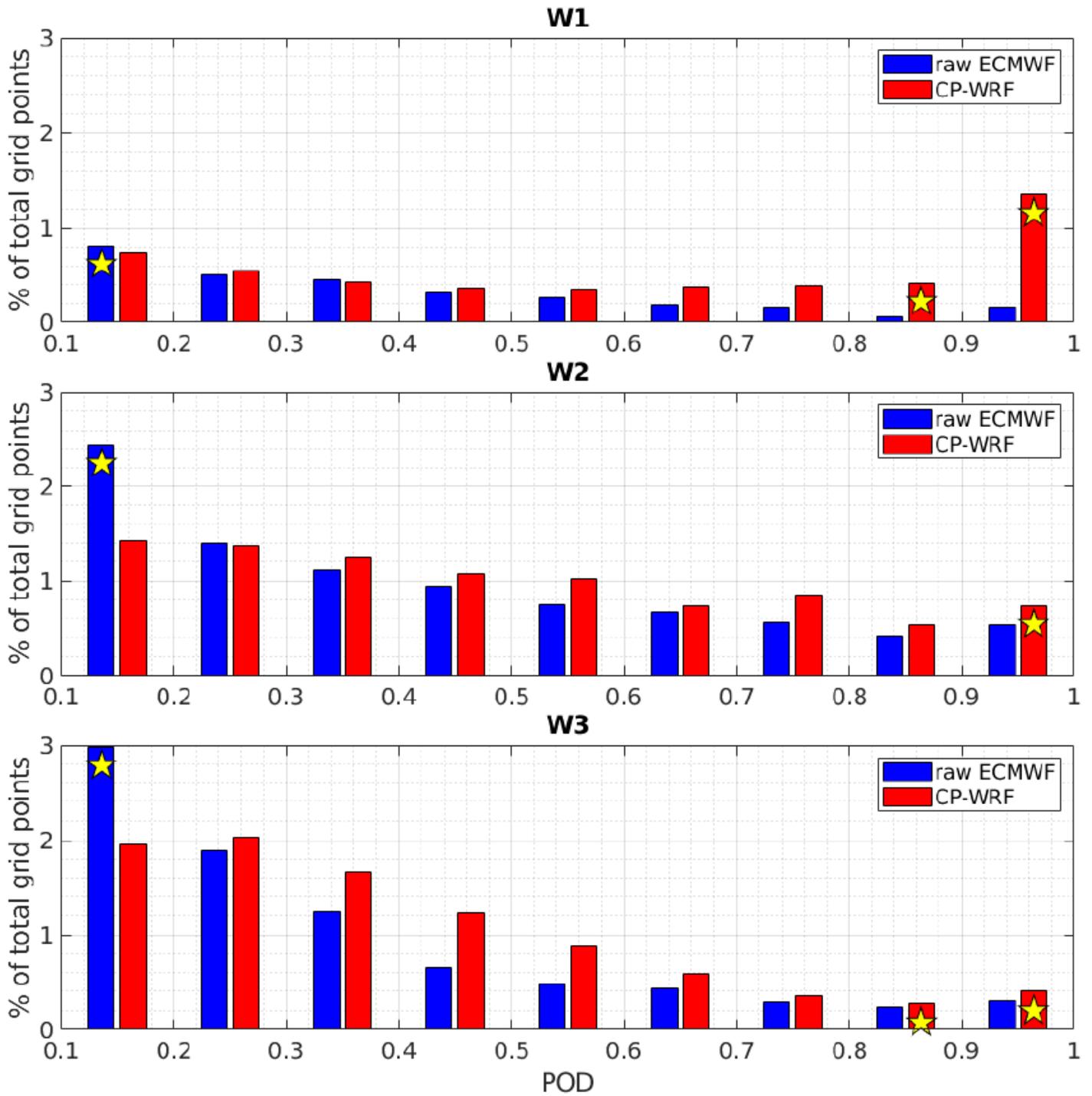
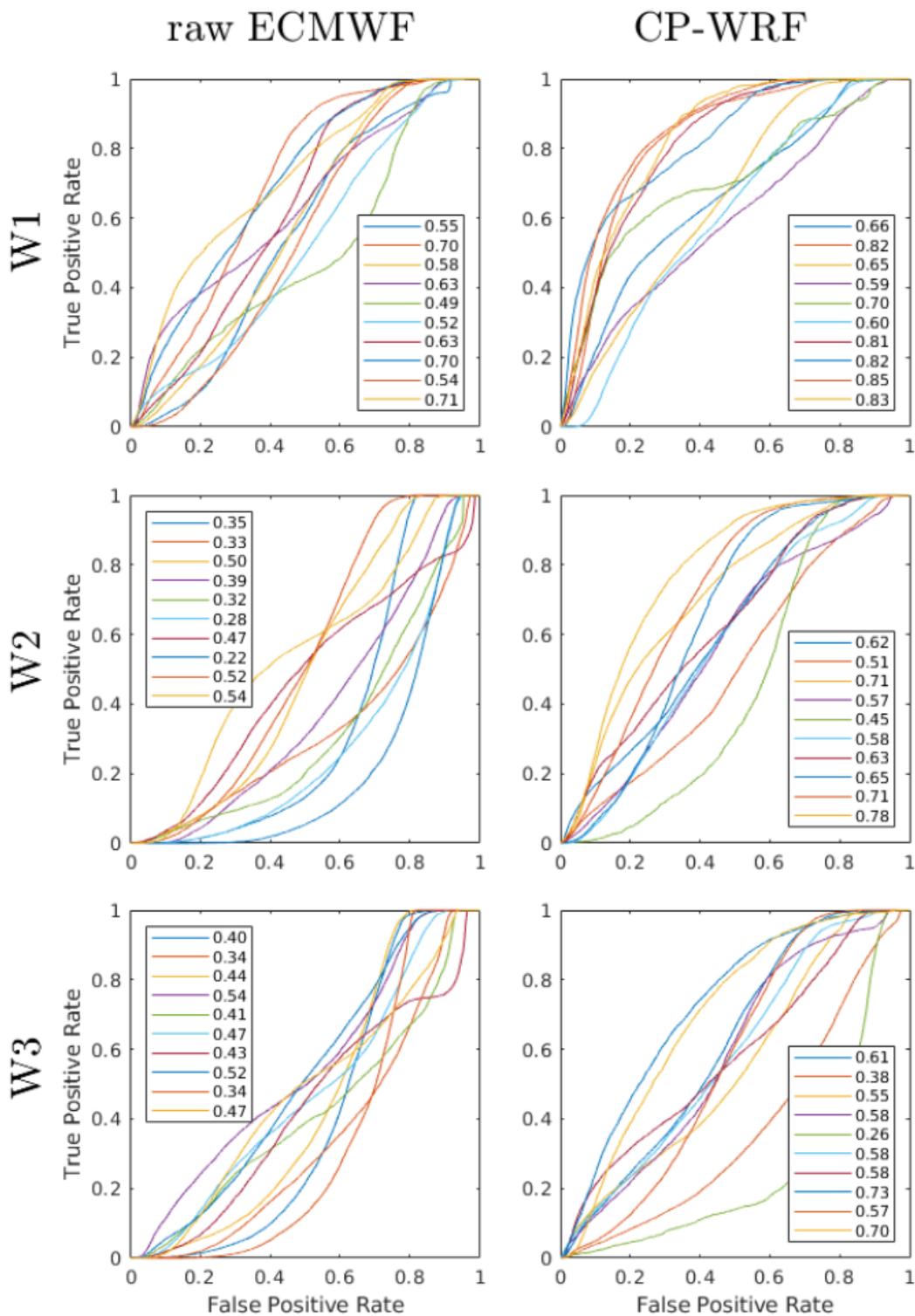


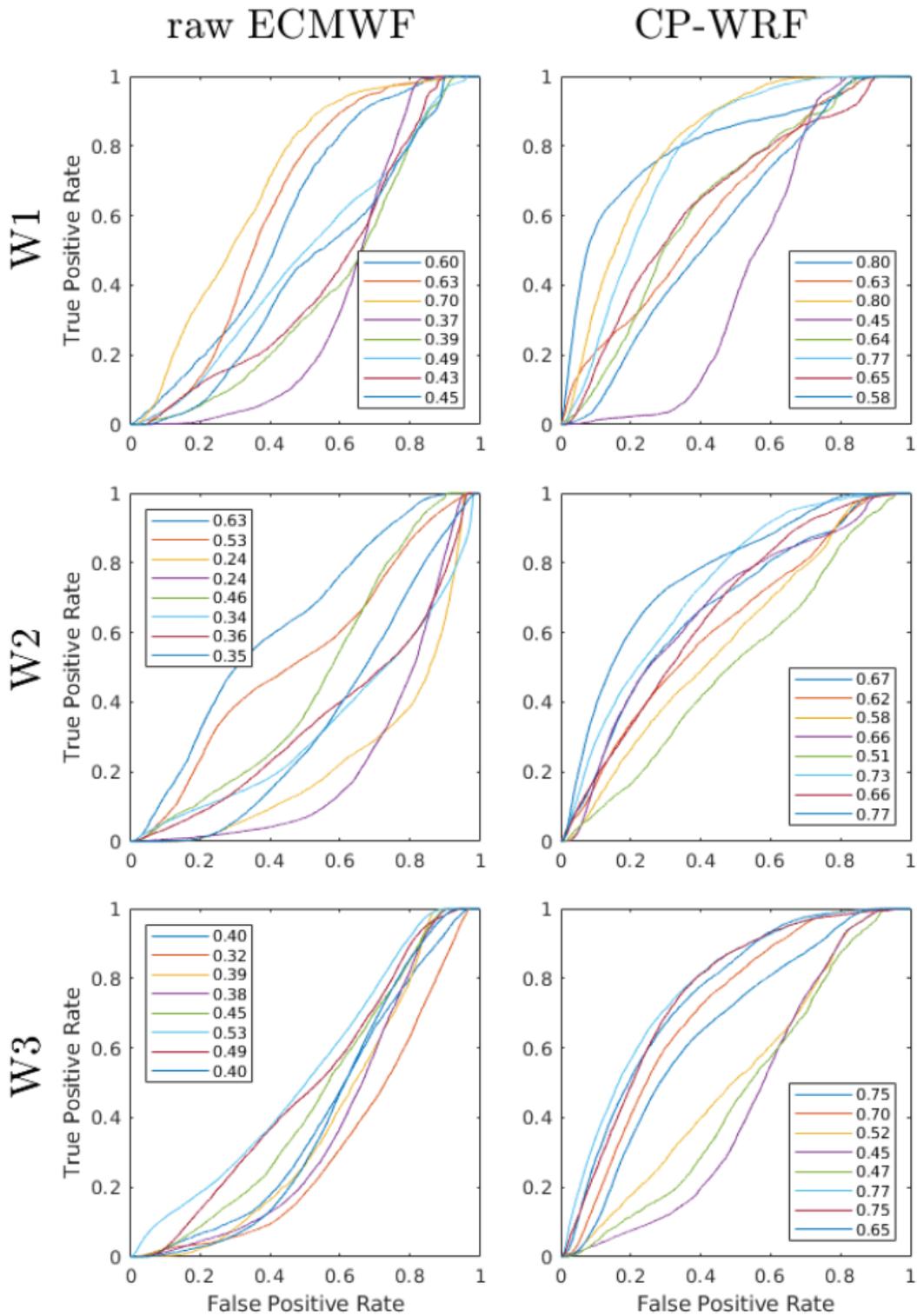
Figure 13

Similar to FIG. 12, but for the tropical events.



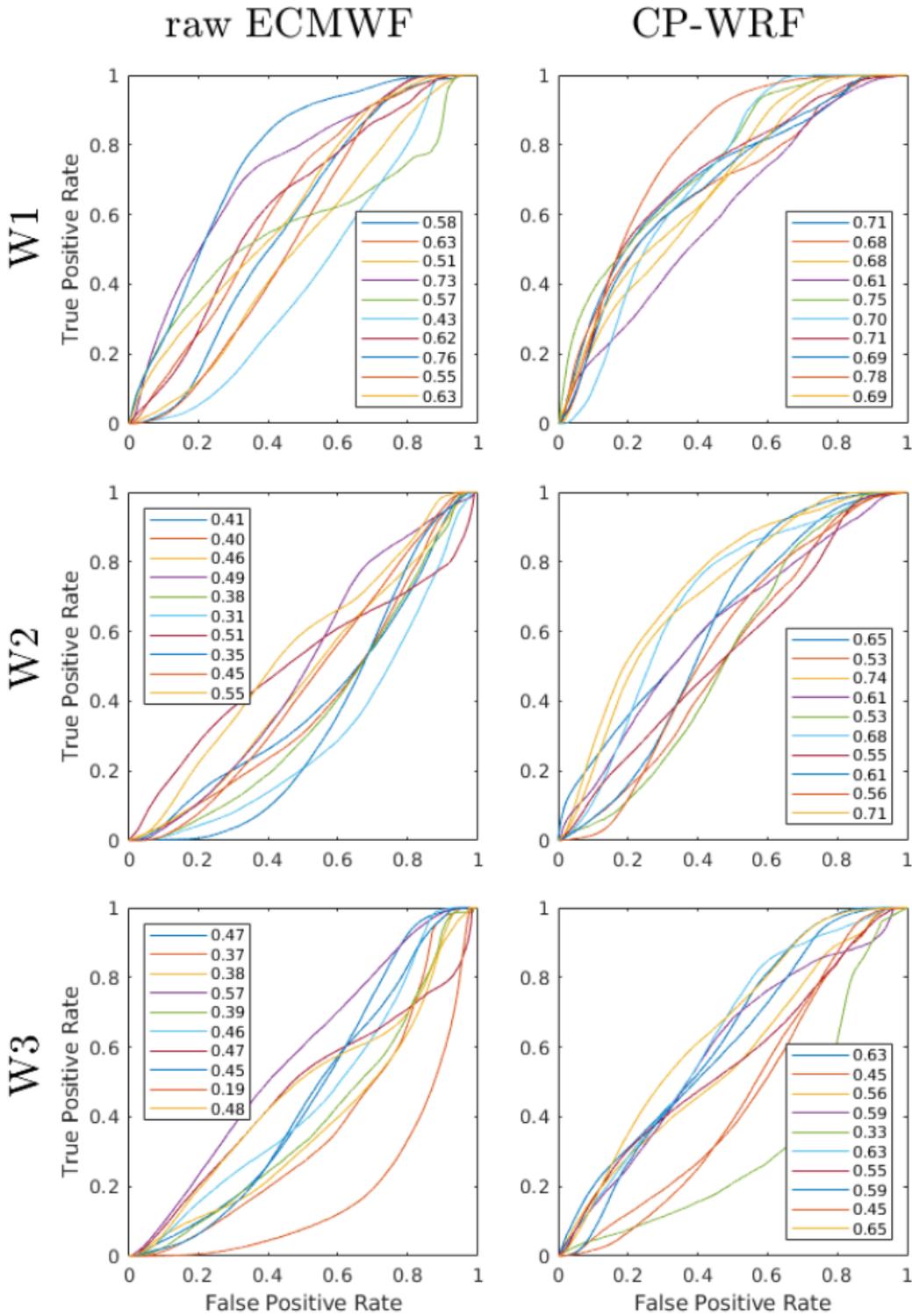
**Figure 14**

Relative Operating Characteristic (ROC) curves and Area under the Curve (AUC) of each extratropical event relative to the GPMF precipitation at W1, W2, and W3. Note the forecast skill improvement in the CP-WRF precipitation (right) compared to the raw ECMWF precipitation (left).



**Figure 15**

Similar to FIG. 14, but for each tropical event.



**Figure 16**

ROC curves and AUC for each extratropical event relative to the KAUST-RA precipitation at W1, W2, and W3. Note the forecast skill improvement in the CP-WRF precipitation (right) compared to the raw ECMWF precipitation (left).

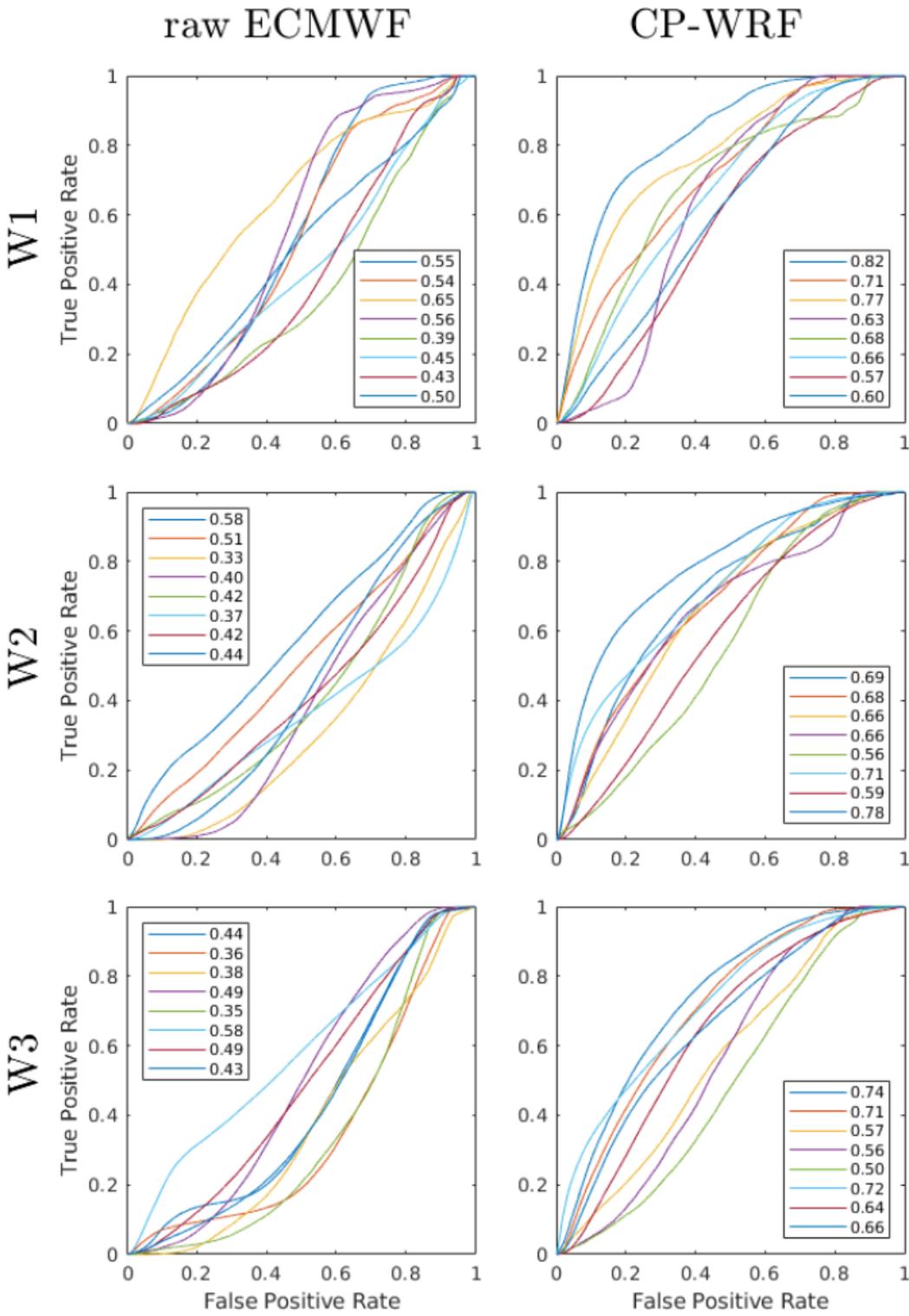


Figure 17

Similar to FIG.16, but for each tropical event.

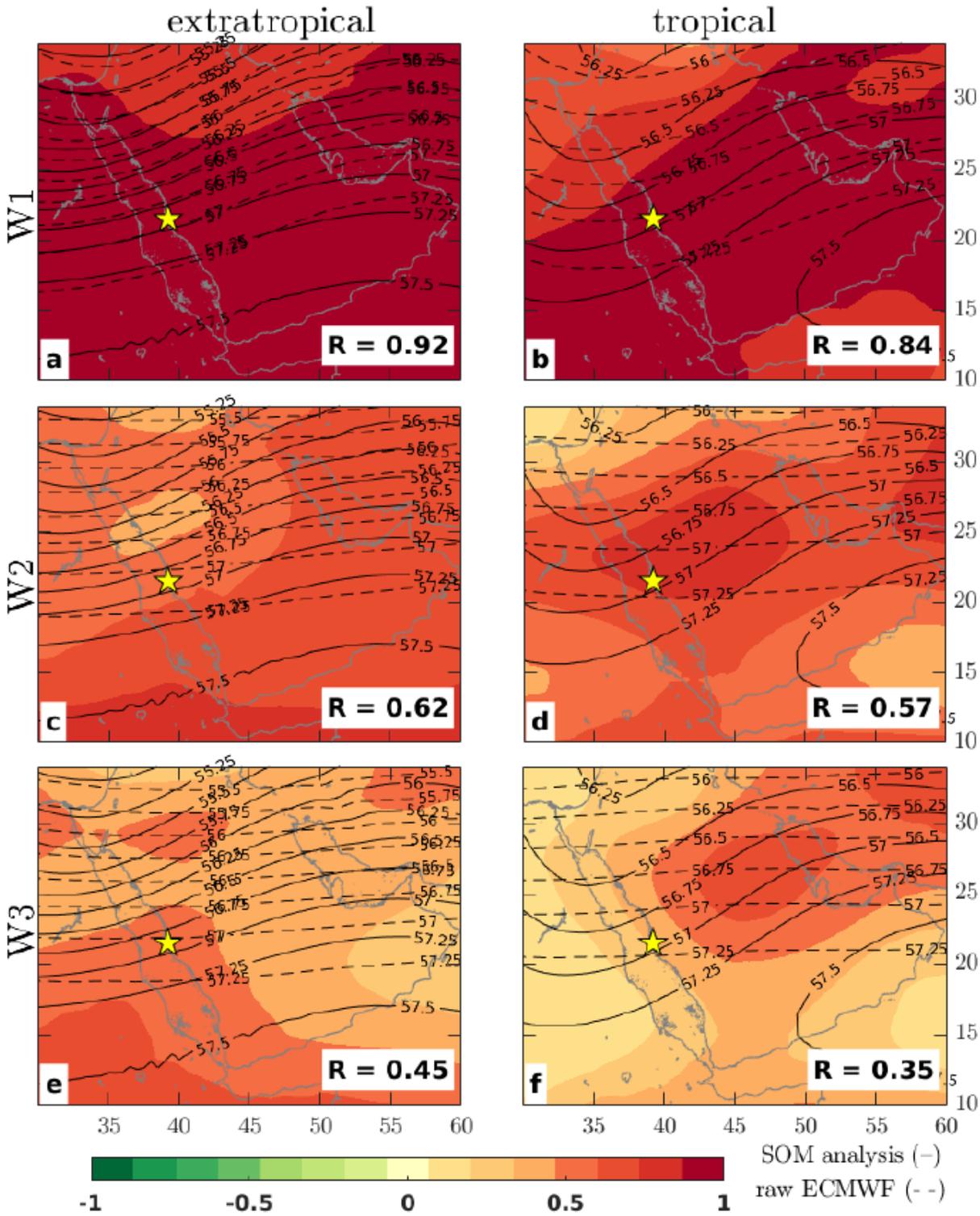


Figure 18

Pearson's correlation coefficient (shading) between the raw ECMWF and the ERA-I 500-hPa geopotential heights. The dashed and solid contours are the average daily means of the raw ECMWF 500-hPa geopotential heights and the 500-hPa geopotential heights of the Luong-SOM analysis, respectively. Both are resolved to one-tenth of a kilometer. The R values are the average correlations across the domain. The star in each panel indicates the location of Jeddah.

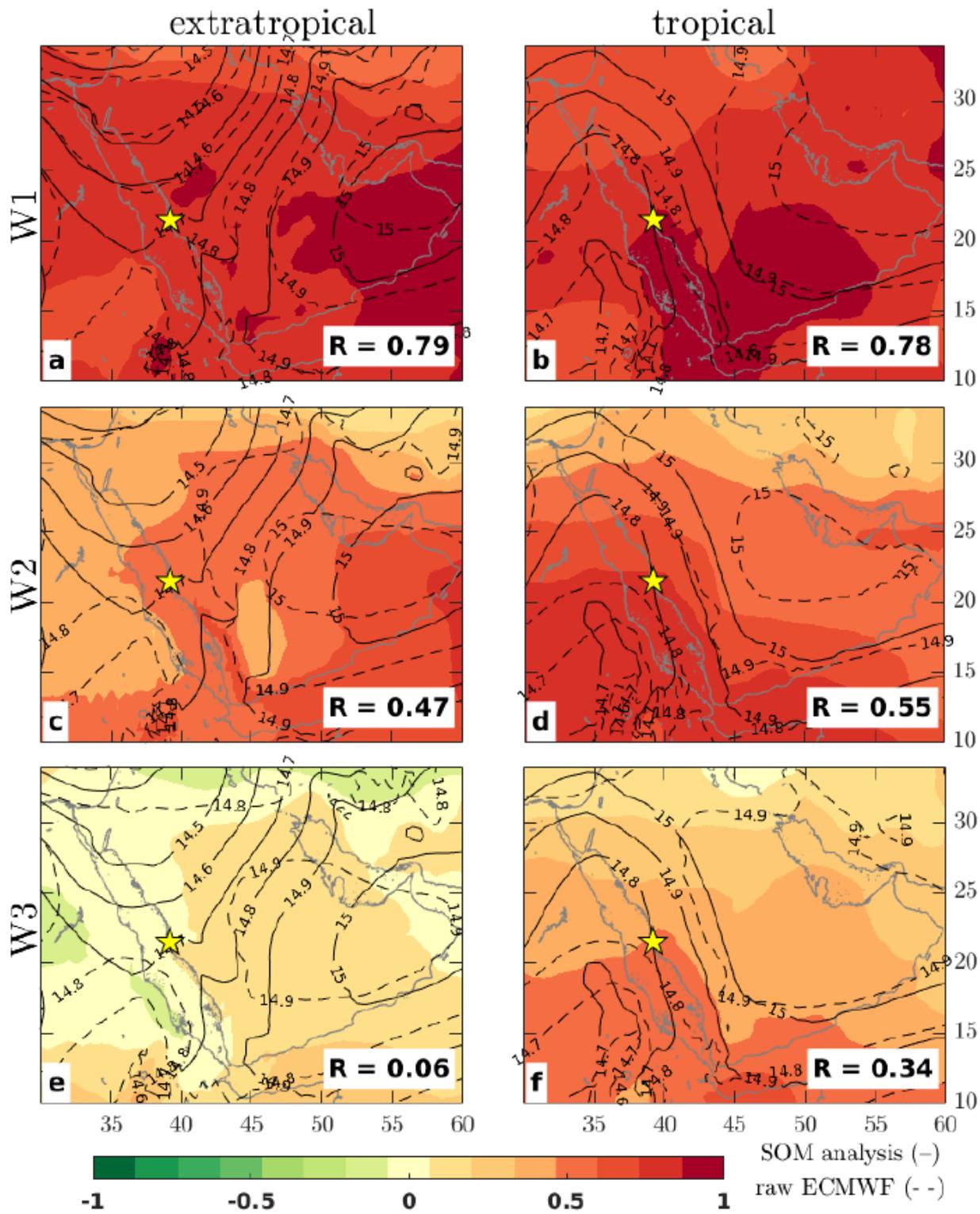


Figure 19

Similar to Fig. 18 but for the 850-hPa geopotential heights.