



**HAL**  
open science

# Improving subseasonal forecast of precipitation in Europe by combining a stochastic weather generator with dynamical models

Meriem Krouma, Damien Specq, Linus Magnusson, Constantin Ardilouze, Lauriane Batté, Pascal Yiou

## ► To cite this version:

Meriem Krouma, Damien Specq, Linus Magnusson, Constantin Ardilouze, Lauriane Batté, et al.. Improving subseasonal forecast of precipitation in Europe by combining a stochastic weather generator with dynamical models. Quarterly Journal of the Royal Meteorological Society, 2024, 10.1002/qj.4733 . hal-04565569

**HAL Id: hal-04565569**

**<https://hal.science/hal-04565569>**

Submitted on 5 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

**RESEARCH ARTICLE**

# Improving subseasonal forecast of precipitation in Europe by combining a stochastic weather generator with dynamical models

Meriem Krouma<sup>1,2</sup>  | Damien Specq<sup>3</sup> | Linus Magnusson<sup>4</sup>  |  
Constantin Ardilouze<sup>3</sup>  | Lauriane Batté<sup>3,5</sup>  | Pascal Yiou<sup>1</sup>

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212

CEA-CNRS-UVSQ, IPSL & Université Paris-Saclay, Gif-sur-Yvette, France

<sup>2</sup>ARIA Technologies, Nanterre, France

<sup>3</sup>Direction de la Climatologie et des Services Climatiques, Météo-France, Toulouse, France

<sup>4</sup>CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

<sup>5</sup>European Centre for Medium-Range Weather Forecast, Reading, UK

**Correspondence**

Meriem Krouma, Laboratoire des Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, IPSL & Université Paris-Saclay, 91191 Gif-sur-Yvette, France.

Email: [meriem.krouma@lscce.ipsl.fr](mailto:meriem.krouma@lscce.ipsl.fr)

**Funding information**

European Union's Horizon 2020 Research and Innovation Programme (XAIDA), Grant/Award Number: 101003469; H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 813844; EU International Training Network (ITN) Climate Advanced Forecasting of Sub-seasonal Extremes (CAFE)

**Abstract**

We propose a forecasting tool for precipitation based on analogues of circulation defined from 5-day hindcasts and a stochastic weather generator that we call “HC–SWG.” In this study, we aim to improve the forecast of European precipitation for subseasonal lead times (from 2 to 4 weeks) using the HC–SWG. We designed the HC–SWG to generate an ensemble precipitation forecast from the European Centre of Medium-range Weather Forecasts (ECMWF) and Centre National de la Recherche Météorologique (CNRM) subseasonal-to-seasonal ensemble reforecasts. We define analogues from 5-day ensemble reforecast of Z500 from the ECMWF (11 members) and CNRM (10 members) models. Then, we generate a 100-member ensemble for precipitation over Europe. We evaluate the skill of the ensemble forecast using probabilistic skill scores such as the continuous ranked probability skill score (CRPSS) and receiver operating characteristic curve. We obtain reasonable forecast skill scores within 35 days for different locations in Europe. The CRPSS shows positive improvement with respect to climatology and persistence at the station level. The HC–SWG shows a capacity to distinguish between events and non-events of precipitation within 15 days at the different stations. We compare the HC–SWG forecast with other precipitation forecasts to further confirm the benefits of our method. We found that the HC–SWG shows improvement against the ECMWF precipitation forecast until 25 days.

**KEYWORDS**

dynamical models, ensemble forecast, European precipitation, subseasonal

## 1 | INTRODUCTION

Making accurate subseasonal forecasts remains a challenge for the scientific community (White *et al.* 2022). The subseasonal time range lies between the medium-term daily weather forecast and the seasonal forecasting (Vitart *et al.*, 2017). Considerable efforts have been made to understand the different processes, interactions, and sources of predictability in order to improve the subseasonal forecast (Domeisen *et al.*, 2022; Robertson & Vitart, 2019; White *et al.*, 2022). The subseasonal predictability is linked to atmospheric, oceanic, and land processes (Robertson & Vitart, 2019). The most important sources of predictability of the subseasonal range are as follows: the Madden–Julian oscillation (Lau & Waliser, 2011; Vitart *et al.*, 2017), owing to its impacts on tropical and extratropical worldwide weather (Cassou, 2008; DeFlorio *et al.*, 2019); the soil moisture (Koster *et al.*, 2010), as this can influence lower atmospheric temperature and local precipitation (Domeisen *et al.*, 2022; Wei & Dirmeyer, 2019); snow cover (Lin & Wu, 2011), in particular for the polar and midlatitude regions (Penny *et al.*, 2019); ocean conditions (Woolnough *et al.*, 2007), which showed capacity to enhance precipitation and temperature forecast over certain areas (Subramanian *et al.*, 2019); and the stratosphere, which has lagged impacts on precipitation and temperature (Butler *et al.*, 2019).

Improving the subseasonal forecast is also related to the improvement of model physics by incorporating coupled processes and many components of the Earth system, such as ocean and sea ice and including the uncertainties in the initial conditions related to the interactions between the different sources of predictability mentioned previously (Merryfield *et al.*, 2020; Newman *et al.*, 2003; Rashid *et al.*, 2011; Vitart, 2014). Subseasonal forecasts have been more accurate with the improvement of numerical weather prediction (NWP) models (Robertson & Vitart, 2019). NWP forecasting has improved over the past decades due to model improvements and the availability of better data and forecast initialization (Magnusson & Källén, 2013). NWP models have shifted from deterministic to probabilistic approaches in the past decades. Indeed, ensemble (probabilistic) forecasts help to catch atmospheric chaos by producing a set of probabilities for the predicted variable (Palmer, 2000). Hence, a probabilistic forecast provides both the most likely scenario and the uncertainties associated with it through a larger ensemble forecast that allows more confident verification of the subseasonal forecast.

As a result of all the aforementioned efforts, the subseasonal ensemble forecasts have showcased their potential to deliver valuable predictions and early alerts for significant climate and weather events (Domeisen *et al.*, 2022; Mariotti *et al.*, 2018; Robertson & Vitart, 2019). These

include many impactful events, such as severe heatwaves, cold spells, and tropical storm occurrences (Domeisen *et al.*, 2022; Merryfield *et al.*, 2020). For instance, ensemble NWP predictions showed a good forecast skill for atmospheric fields such as the geopotential height at 500 hPa for up to 1 month (Robertson & Vitart, 2019). Nevertheless, the subseasonal forecast skill is still insufficient for some weather variables, such as precipitation. There are at least two reasons for this shortcoming. Precipitation results from complex nonlinear and multiscale processes that are not well resolved in NWP models (Stan & Straus, 2019; Zhang *et al.*, 2021). Errors related to the physical parametrization assumptions combined with a lack of resolution explain the poor predictability of precipitation (Karl *et al.*, 1990). In addition, small-scale effects, such as topography or orography, are not well resolved by NWP models. However, those parameters are important for local weather (Wilks, 2011, chap. 6).

Many studies investigated the quality and accuracy of the subseasonal forecast for precipitation and other variables. Domeisen *et al.* (2022) showed that precipitation events, including events in Europe at the subseasonal range time, can be more predictable by associating them with a correct large-scale circulation. Zhang *et al.* (2021) explained that the forecast skill of the subseasonal-to-seasonal (S2S) precipitation forecast depends on the geographical regions and seasons. For extreme precipitation forecast, Olaniyan *et al.* (2018) found that the S2S models could not reliably forecast the extreme precipitation during the West African monsoon. In addition, Rivoire *et al.* (2023) assessed the forecast skill of the S2S extreme precipitation forecast and found high heterogeneity in the forecast skill over regions and seasons in Europe.

To overcome subseasonal forecast issues related to model parametrization or uncertainties, subseasonal forecasts can be obtained by combining numerical or dynamical weather prediction models with machine-learning approaches (Weyn *et al.*, 2021) as outlined by Barnes *et al.* (2023), where they showed the improvement in the forecasting skill of total regional rainfall across Great Britain for up to 1 month using convolutional neural networks, or with statistical techniques that can also contribute to enhance the forecast information on a small scale. Indeed, statistical forecasts based on NWP information tend to correct forecast biases (Klein *et al.*, 1959; Specq & Batté, 2020) as shown in Ben Bouallègue *et al.* (2023), where statistical postprocessing methods have been used to optimize NWP forecast of 2 m temperature and 10 m wind speed.

Stochastic weather generators (SWGs) have been used to generate ensemble weather forecasts for different climate/weather variables (Wilks & Wilby, 1999). SWGs have a good capacity to simulate the behaviour of the

climate variables (Ailliot *et al.*, 2015). They have been used to forecast weather and climate variables such as temperature (Yiou & Déandréis, 2019), precipitation (Krouma *et al.*, 2022), the Madden–Julian oscillation (Krouma *et al.*, 2023), and the North Atlantic oscillation. SWGs have a low computing cost compared with numerical models (Ailliot *et al.*, 2015). Combining SWGs with analogues of atmospheric circulation is a promising approach to simulate the weather. Indeed, the circulation analogue method assumes that the future evolution of the atmosphere will be similar to the flows that followed the historical analogues (Atencia & Zawadzki, 2014; Blanchet *et al.*, 2018; Lorenz, 1969; Yiou *et al.*, 2013).

The aim of this work is to improve the forecast skill of precipitation averages over Europe using an SWG based on analogue circulation for subseasonal lead times ( $\approx 10$  to 35 days). The SWG developed by Yiou (2014) showed the capacity to forecast average precipitation (Krouma *et al.*, 2022) within 5–10 days and temperature within 40 days (Yiou & Déandréis, 2019) with promising probabilistic scores. In this work, we revisit the SWG described by Krouma *et al.* (2022) to optimize the simulation of European precipitation from ensemble dynamical reforecasts of the European Centre of Medium-range Weather Forecasts (ECMWF) and the Centre National de la Recherche Météorologique (CNRM) of Météo France. The idea is to use the gained forecast skill of the Z500 forecast from dynamical models as input to the SWG to have an ensemble forecast of precipitation at different lead times up to 1 month. Then, we evaluate the ensemble forecast skill using skill scores such as the continuous ranked probability score (CRPS) and the receiver operating characteristic (ROC) curve. We also compare our forecasts of precipitation with the ECMWF precipitation forecast using the CRPS and the anomaly correlation coefficient (ACC) as a deterministic score.

The rest of the article is structured as follows: Section 2 details the data used for running our forecast. Section 3 describes the forecast process, including the circulation analogues computation and the SWG and explains the verification metrics used to evaluate the forecast skill. The results of the simulations and the evaluation of the ensemble forecast, as well as the comparison of the SWG forecast with the ECMWF precipitation forecast, are presented and discussed in Section 4. Section 5 contains the main conclusions.

## 2 | DATA

We use daily geopotential at 500 hPa (Z500) data from dynamical reforecasts. Reforecasts (also known as hindcasts) are forecast runs using the same model version

as the real-time forecast for past periods. Two configurations are mainly used to produce reforecasts. The first configuration, the so-called “fixed configuration”, consists of producing reforecasts for all past dates once during the lifetime of a given model version. In this case, a new set of reforecasts is produced with each new model version. The second configuration, known as “on-the-fly configuration”, consists of producing reforecasts at the same time as the real-time forecasts, which means that each reforecast refers to a real-time forecast.

The reforecasts of Z500 were collected from the S2S database for two models: CNRM and ECMWF (Vitart *et al.*, 2017). The ECMWF reforecast is produced “on the fly” and it is composed of an 11-member ensemble covering the past 20 years (Vitart *et al.*, 2019). As initial conditions, the ECMWF reforecast uses the ECMWF Reanalysis v5 (ERA5) (Hersbach *et al.*, 2020) and Ocean Reanalysis System 5 ocean initial conditions (Hersbach *et al.*, 2020). The CNRM reforecasts are produced with the fixed configuration (Ardilouze *et al.*, 2021; Batté & Déqué, 2016). The CNRM reforecast ensemble is composed of 10 members initialized each week over the 1993–2017 period. The CNRM model uses as initial conditions ERA5 for atmosphere and land surface and MERCATOR-OCEAN ocean reanalyses. Both reforecasts feed the S2S database weekly. We considered two models produced with similar properties coupled with ocean and sea-ice models. The main characteristics of the ECMWF and CNRM models are shown in Table 1.

We used two different precipitation databases for verification purposes. Daily observation data (i) at the station scale from the European Climate Assessment & Dataset (ECA&D) (Klein Tank *et al.*, 2002) served as a reference for four different stations in Europe (Berlin, Orly, Toulouse, Madrid), and (ii) data from E-OBS (Cornes *et al.*, 2018) in order to test the forecast skill of our model over Europe. E-OBS data are available in a daily range from 1950 to 2022. We regridded the E-OBS data to a resolution of  $1.5^\circ \times 1.5^\circ$  to comply with the reforecast horizontal resolution. We used the S2S precipitation forecast from the ECMWF model to verify the forecast skill of our model (Vitart *et al.*, 2019). We considered daily data from January to December from 2015 to 2021 at different lead times. Data were extrapolated for each station studied and bias corrected according to the climatology of each lead time.

## 3 | METHODOLOGY

The goal of this study is to simulate forecast ensembles of  $N = 100$  members for European precipitation at the subseasonal lead time (from 2 to 4 weeks). Our methodology enhances the available hindcast ensembles to  $N = 100$

**TABLE 1** Characteristics of the European Centre of Medium-range Weather Forecasts (ECMWF) and Centre National de la Recherche Météorologique (CNRM) subseasonal-to-seasonal ensemble reforecasts.

Model	Model version	Period	Horizontal resolution (km)	Size	Ocean resolution	Sea ice
ECMWF	CY47R2	2001–2021	15–31	11	0.25°, 75 levels	Active
CNRM	CNRM-CM 6.1	1993–2017	50	10	0.25°, 75 levels	Active

members through a random sampling of circulation analogues selected separately from the ECMWF and CNRM reforecasts' ensembles. The following subsections explain how the circulation analogues are computed from reforecasts of the ECMWF and CNRM, and how the random sampling (or SWG) is performed.

### 3.1 | Data processing and analogue dataset

The first step in our forecasting process is to define the analogues of Z500. An analogue is a date where the configuration of the atmospheric circulation is similar to a selected day  $t$  (Krouma *et al.*, 2022; Yiou, 2014). To define an analogue, we apply two rules: the calendar distance between  $t$  and its analogues should not exceed 30 calendar days, and an analogue of  $t$  should be in a different year than  $t$ . Our selection criterion is the Euclidean distance, which we compute between Z500 at a day  $t$  and its analogue day  $t'$ . Then, for each day, we keep the  $K$  analogues with the minimum Euclidean distance. To find analogues of the Z500( $x, t$ ) field, the Euclidean distance is computed as follows:

$$D(t, t') = \left[ \sum_x |Z500(x, t) - Z500(x, t')|^2 \right]^{1/2}, \quad (1)$$

where  $x$  is a spatial index.

To define analogues, some parameters need to be defined, such as the geographical region. We determined analogues over the region with coordinates 30°W–2°E, 40°–60°N defined by Krouma *et al.* (2022) as an optimal region to compute analogues for precipitation forecast in Europe. In this study, we keep  $K = 20$  best analogues as in Krouma *et al.* (2022). We also choose to use analogues only from Z500 to ensure a fair comparison with our previous work (Krouma *et al.*, 2022) and to ensure that the forecast skill improvement is related to the use of the reforecast ensembles and not related to other atmospheric variables.

Before defining the analogues, we start by verifying the spread of the reforecast ensemble of ECMWF and CNRM on the fifth day. The aim is to verify the spread of the ensemble and check whether we can use the whole ensemble to compute analogues. We computed the Euclidean

distance between the members of each ensemble. Then, we compared it separately to the distance between the analogues of each model (Figure 1). We computed analogues from the first and eleventh (or tenth) members, and the ensemble mean (Figure 1a–c). The analogues were computed for the 11 (or 10) members, but for brevity's sake, and as the conclusion is the same, we choose to show the comparison between the analogues of the ensemble mean and the first and last members.

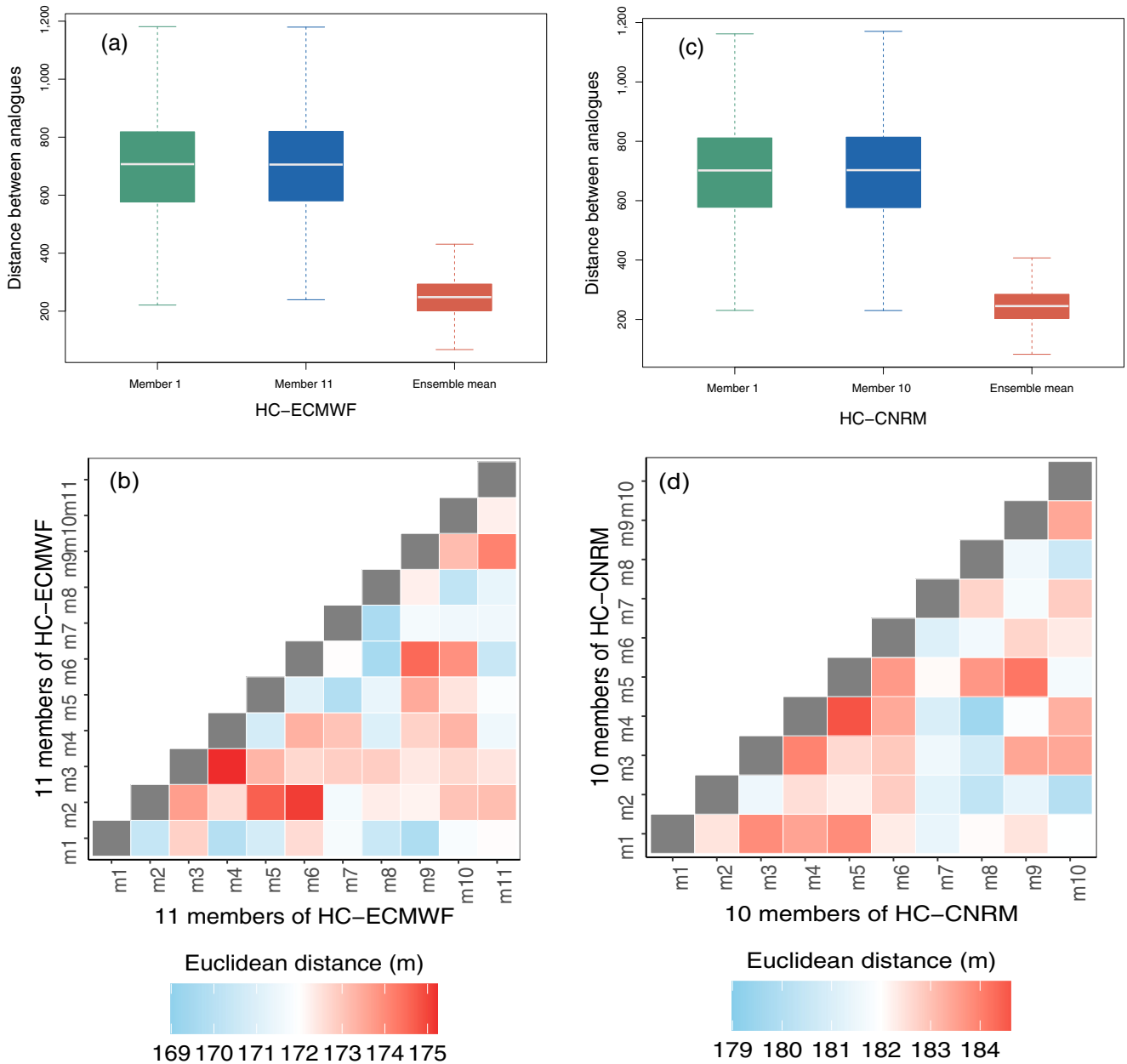
The distance between the 11 members of the ECMWF reforecast of Z500 for 5 days ahead varies between 169 m and 175 m (Figure 1b). The analogues of the ensemble mean are closer than the analogues of the first and last members. For the ensemble reforecast of the CNRM, the maximum distance between the 10 members is 184 m (Figure 1d). The analogues of the ensemble mean show an average distance of 259 m compared with the analogues of the first and tenth members. We conclude that the maximum distance between the ECMWF or CNRM ensemble members is smaller than the average distance between the analogues, as shown in Figure 1.

The ensemble spread is small enough compared with the distance between the analogues, plus the use of one member of each ensemble or of the mean ensemble led to the same analogues with a smaller distance between analogues. Hence, we decided to use the ensemble mean at the fifth day to compute the analogues, as illustrated in Figure 2, instead of using the ensemble members separately.

### 3.2 | SWG configuration

The aim of the analogue SWG is to generate random trajectories based on previously computed analogues (Yiou, 2014). In order to generate a trajectory for a particular day  $t_0$  in year  $y_0$ , we produce series of  $N = 100$  simulations until a time  $t_0 + T$ , with a lead time  $T$  ranging from 10 to 35 days. In this article, we make a forecast for  $t_0 = t + \delta$ . Here, we assume  $\delta$  to be 5 days and refer to ECMWF and CNRM reforecasts 5 days ahead that serve as input for predicting the average precipitation between  $t_0$  and  $t_0 + T$ . This approach is based on the analogues of the atmospheric circulation. For the sake of conciseness, we refer to  $T$  as the lead time owing to its value being





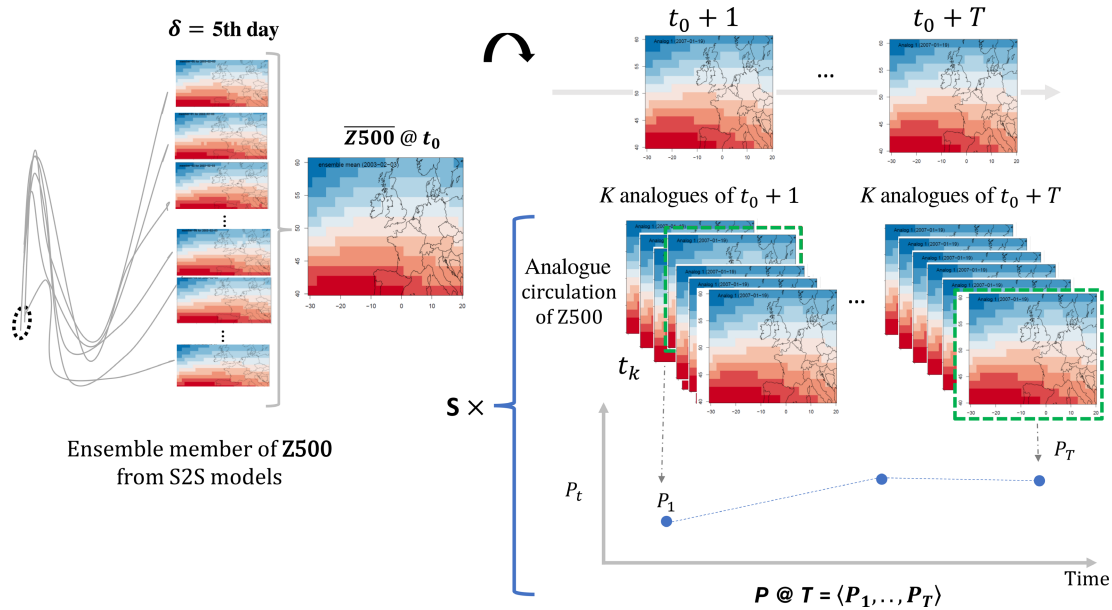
**FIGURE 1** Comparison between the distance between the analogues computed from the first and last members and the ensemble mean of the (a) European Centre of Medium-range Weather Forecasts (ECMWF) and (c) Centre National de la Recherche Météorologique (CNRM) reforecasts and the distance between the members of the ensembles of (b) ECMWF and (d) CNRM ensemble members at 5 days ahead. HC, hindcast. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

greater than zero and our focus on averages beyond  $t_0$ . Nonetheless, it is essential to note that  $T$  and  $\delta$  are distinct quantities.

To produce a single trajectory, we follow the procedure outlined in Krouma *et al.* (2022). Beginning on day  $t_0$ , we randomly select an analogue  $t'_k$  among the  $K = 20$  best analogues for day  $t_0 + 1$ . The random selection of analogues of the day  $t_0 + 1$  is carried out using weights that are proportionate to the calendar difference between  $t_0$  and analogue dates in order to ensure that time progresses (Yiou, 2014).

We also exclude analogue dates with years that are equal to  $y_0$ . This rule is important for the next iterations in order to produce a hindcast simulation. We then replace  $t_0$  with the selected analogue of  $t'_k$  and repeat the operation  $T$  times. Excluding analogues in year  $y_0$  from the selection ensures that we do not use information from the  $T$  days that follow  $t_0$ .

The method involves conducting hindcast simulations by selecting analogues of a specified time period, which are then replaced with the initial time interval. This is repeated



**FIGURE 2** Schematic to illustrate the hindcast–stochastic weather generator (HC–SWG) forecast procedure at a day  $t_0$  for a lead time of  $T$  days. The starting point is the European Centre for Medium-range Weather Forecasts or Centre National de Recherches Météorologiques reforecasts ensemble members of Z500 at  $\delta = 5$  days and the output is the ensemble members of precipitation at a lead time  $T$ . This illustration shows the HC–SWG process to generate one trajectory for a day  $t_0$ . S2S, subseasonal to seasonal. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

$T$  times. To ensure the exclusion of analogues in year  $y_0$  from the selection, their corresponding information from the  $T$ -day period after  $t_0$  is excluded. Subsequently, the simulated precipitation corresponds to the next selected analogue,  $t'_{k+1}$ . The analogues are resampled each day using  $20^n$  potential trajectories, where  $n$  is the length of the simulated sequence. This process produces a hindcast random trajectory between  $t_0$  and  $t_0 + T$ . Then,  $t_0$  is shifted by  $\Delta t \approx T/2$  days and the ensemble simulation is repeated. This provides a set of ensemble forecasts with analogues.

The aforementioned procedure is repeated  $N = 100$  times to simulate  $N = 100$  trajectories from  $t_0$  to  $t_0 + T$ . The daily precipitation of each trajectory is time averaged between  $t_0$  and  $t_0 + T$ . Hence, we obtain an ensemble of  $N = 100$  forecasts of the average precipitation for a day  $t_0$  and lead time  $T$ .

The forecasts of precipitations based on analogues of atmospheric circulation Z500 from the ensemble mean of the CNRM and ECMWF reforecasts are started every  $\Delta t \approx T/2$  days between January 1, 1993, and December 31, 2017, using the CNRM reforecast and between January 1, 2002, and December 31, 2021, using the ECMWF reforecast. This yields a stochastic ensemble hindcast of precipitation and atmospheric circulation (Z500) for the ECMWF and CNRM reforecasts (Table 2).

The added value of this study compared with the previous study (Krouma *et al.*, 2022), where the atmospheric circulation from reanalyses was used, is that we consider

**TABLE 2** Data split for training and verification.

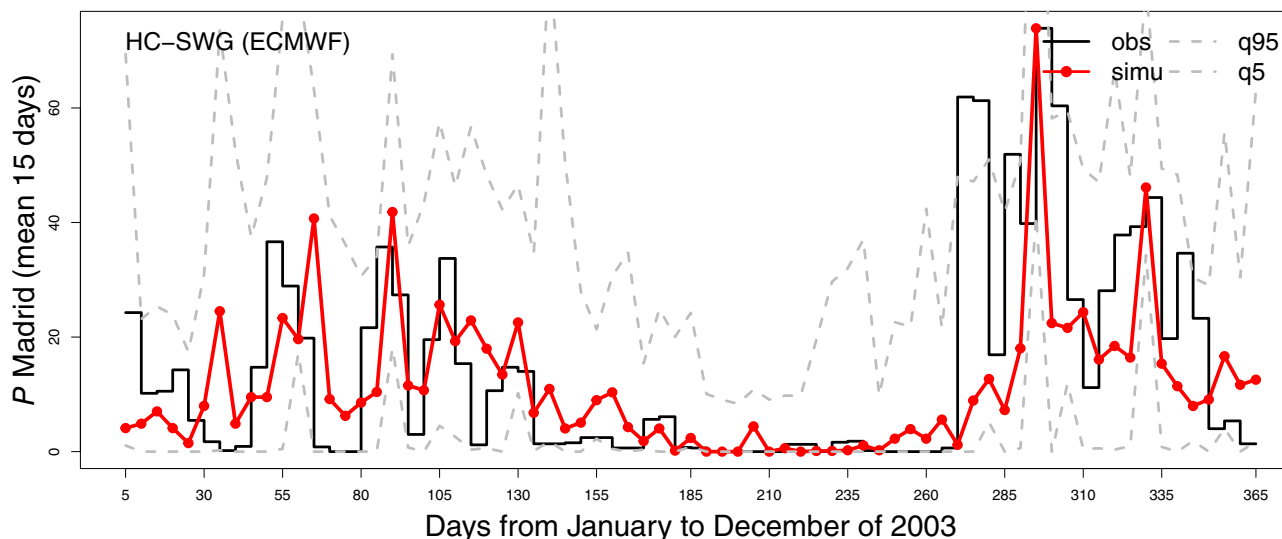
	Training	Test/ verification	Verification vs ECMWF forecast
ECMWF	2002–2014	DJF, JJA	2015–2021
CNRM	1993–2014	DJF, JJA	2015–2018

Abbreviations: CNRM, Centre National de Recherches Météorologiques; DJF, December–February; ECMWF, European Centre for Medium-range Weather Forecasts; JJA, June–August.

Z500 from the reforecasts of S2S models and at  $\delta = 5$  days ahead instead of  $\delta = 0$  days as in Krouma *et al.* (2022), to forecast precipitation beyond 5 days. This means that the circulation analogues calculated in this study are analogues of 5 days ahead (Figure 2). Consequently, the simulations will be called “HC–SWG” forecasts.

### 3.2.1 | Example of the forecast process

To illustrate the procedure, we start from the 11 (or 10) reforecast members of Z500,  $\delta = 5$  days ahead, to get an ensemble forecast of precipitation, as shown in Figure 2. For a given day  $t_0 =$  February 3, 2003, we compute the ensemble average of the Z500 reforecast (either ECMWF or CNRM). Then, we compute analogues from the ensemble average at a day  $t_0 =$  February 3, 2003, and we keep  $K = 20$  best analogues of Z500. For a lead time  $T = 3$  days,



**FIGURE 3** Time series of the simulations (red line), observations (black line), and the 5th and 95th quantiles of simulations (dashed grey lines) for precipitation over Madrid at a lead time  $T = 15$  days for 2003 using the hindcast–stochastic weather generator (HC–SWG) with hindcast of the European Centre for Medium-range Weather Forecasts (ECMWF) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4733)]

we randomly select an analogue from the 20 analogues of  $Z500$  at  $t_0 + 1$ . For instance, the first best analogue for  $t_0 = \text{February 3, 2003}$  is  $t'_k = \text{January 19, 2007}$ . We repeat this operation 100 times. Then, for each analogue, we consider the corresponding amount of precipitation either from E-OBS or ECA&D data. This way, we construct the first trajectory of precipitation, as shown in Figure 2. We generate 100 trajectories of precipitation using the same procedure. Hence, we obtain 100 samples of average precipitation between  $t_0$  and  $t_0 + T$  for the defined day  $t_0$  and at  $T = 3$  days.

As an example, we show the time series of the simulations and observations of precipitation over Madrid within 15 days for 2003 in Figure 3. It shows the mean of the 100 simulations (forecast of precipitation) and the observations for lead times of  $T = 15$  days for the whole year. We notice that the HC–SWG reproduces precipitation fluctuations within 15 days as it gets the high and low values of precipitation for the whole year. We can also see that all the values are covered, as shown by the 5th and 95th quantiles (dashed grey lines in Figure 3).

### 3.3 | Assessment of the forecast quality

In order to evaluate the forecast skill of our model, we employed probabilistic skill metrics, including rank correlation, CRPS, and ROC curve analysis (Wilks, 2011). Temporal rank correlation was computed by assessing the correlation between observed precipitation data and the median derived from 100 simulations. Furthermore, the CRPS is used as a quadratic metric to gauge the

disparities between the forecasted cumulative distribution function and the empirical cumulative distribution function derived from the observed data (Wilks, 2011; Zamo & Naveau, 2018). The CRPS is defined by

$$\text{CRPS}(P, x_a) = \int_{-\infty}^{+\infty} (P(x) - \mathcal{H}(x - x_a))^2 dx, \quad (2)$$

where  $x_a$  is the observed precipitation in  $[t_0, t_0 + T]$ ,  $P$  is the cumulative distribution function of  $x$  of the ensemble forecast, and  $\mathcal{H}$  represents the Heaviside function ( $\mathcal{H}(y) = 1$  if  $y \geq 0$ , and  $\mathcal{H}(y) = 0$  otherwise). A perfect forecast yields  $\text{CRPS} = 0$ .

As the CRPS value depends on the unit of the variable to be predicted, it is useful to normalize it with the CRPS value of a reference forecast, which can be obtained by a persistence or climatology hypothesis. The continuous ranked probability skill score (CRPSS) is defined as a fraction of improvement over such a reference forecast (Hersbach, 2000). We compute the CRPSS with reference values of the CRPS:

$$\text{CRPSS} = 1 - \frac{\overline{\text{CRPS}}}{\text{CRPS}_{\text{ref}}}, \quad (3)$$

where  $\overline{\text{CRPS}}$  is the average of the CRPS of the SWG forecast and  $\text{CRPS}_{\text{ref}}$  is the average of the CRPS obtained by the reference forecast.

As a reference, we used climatology, persistence, and model forecasts. The persistence forecast consists of using the average value between  $t_0 - T$  and  $t_0$  for a specific year. Meanwhile, the climatological forecast is derived



from the long-term average between  $t_0$  and  $t_0 + T$ . To account for natural variability, we introduce randomness into both the persistence and climatological forecasts by introducing small white Gaussian noise. We estimate the standard deviation of this noise through bootstrapping over extended time intervals, denoted as  $T$ . We show an example of the learned values of the standard deviation in Figure A2 (see also Appendix A). Additionally, we incorporate the precipitation forecast from the ECMWF model as a third CRPS reference.

CRPSS values fall within the range of negative infinity to 1. A value of CRPSS  $> 0$  indicates that the forecast demonstrates improvement over the reference. When comparing the HC–SWG forecasts with the ECMWF precipitation forecast, we opted for the “fair CRPSS” outlined in the literature (Ferro, 2007, 2014). The fair CRPSS factors in the difference in ensemble sizes, as we generate 100 members compared with ECMWF’s 51 members in our analysis.

The CRPS can be decomposed into three parts: reliability (Re), resolution (Res), and uncertainty (Unc) (Hersbach, 2000):

$$\overline{\text{CRPS}} = \overline{\text{Re}} - \overline{\text{Res}} + \overline{\text{Unc}}. \quad (4)$$

The reliability Re tests the capacity of the ensemble to generate a cumulative distribution function that has, on average, the desired statistical property (Wilks, 2011). The reliability is sensitive to the average spread of the ensemble and biases in the forecast. The model is perfectly reliable when  $\text{Re} = 0$  (Hersbach, 2000). The uncertainty  $U$  is the CRPS based only on the sample climatology (Wilks, 2011). Resolution Res represents the difference with climatology. A positive resolution indicates that the model’s performance is better than the climatology (Wilks, 2011). In our analysis, we focus more on the interpretation of Re. Indeed, that helps to understand the spread of the ensemble forecast.

The ROC curve is a graphical representation that illustrates the trade-off between false alarms and positive hits in a forecast. This curve serves as a reference point for assessing forecast skill. When the ROC curve lies along the diagonal line it reflects a random classifier with no skill. If the ROC curve falls below the diagonal line it indicates poor forecasting skill; conversely, if it rises above the diagonal line it signifies good skill. In other words, the forecast has the potential to distinguish between successful predictions and false alarms (Fawcett, 2006).

The ROC curve is a valuable tool for quantifying discrimination skill, which measures the model’s capacity to differentiate between binary classes (Fawcett, 2006; Krouma *et al.*, 2023). The area under the curve (AUC)

is a metric used to quantify this discrimination skill. A higher AUC value signifies a greater ability of the forecast model to distinguish between precipitation events and non-events (Fawcett, 2006; Toth *et al.*, 2003). In our study, we assess the discrimination skill of our forecast for two different classifications of precipitation events. We consider a first classification for precipitation events and non-events by considering

- precipitation  $P < 1$  mm to be a non-event, and
- $P \geq 1$  mm to be an event of precipitation.

A 1 mm threshold to define precipitation events is commonly employed in climate studies owing to its established relevance and overall use. Other studies also support our choice: Chen *et al.* (2021) indicate that climate models often show an unrealistic bias towards drizzle; and Frei *et al.* (2006) suggested that there is no difference when lower thresholds are applied, and in our previous study Krouma *et al.* (2022) we used the same threshold to establish the relation between precipitation and Z500.

We also evaluate our model’s capacity to detect extreme precipitation events. We define extreme precipitation as precipitation that exceeds the 90th percentile (Rivoire *et al.*, 2023). We compute the empirical 90th percentile after excluding precipitation values below 1 mm. Our binary classification for extremes is defined as follows:

- $P \geq q_{90}$  is a non-event;
- $P > q_{90}$  is an extreme event of precipitation.

We calculated the ACC to compare the HC–SWG and ECMWF precipitation forecasts. The ACC is a deterministic score commonly used to verify spatial fields (Kam *et al.*, 2021; Peng *et al.*, 2013). It represents the spatial correlation that we computed between the ensemble mean of the HC–SWG forecast anomalies and the ensemble mean of the ECMWF precipitation forecast anomalies. The anomalies are computed with respect to the climatology of each ensemble forecast of precipitation. ACC values of 1 indicate a perfect association of forecast anomalies with the observation anomalies. Values  $< 1$  indicate weaker association (Peng *et al.*, 2013).

$$\text{ACC} = \frac{\sum_{i=1}^n (\text{pred}_i - \overline{\text{pred}})(\text{obs}_i - \overline{\text{obs}})}{\sqrt{\sum_{i=1}^n (\text{pred}_i - \overline{\text{pred}})^2} \sqrt{\sum_{i=1}^n (\text{obs}_i - \overline{\text{obs}})^2}}, \quad (5)$$

where obs is the observed precipitation and pred is the forecasted precipitation. ACC was computed separately for the HC–SWG and the ECMWF precipitation forecast.

## 4 | RESULTS

### 4.1 | Evaluation of the forecast of the precipitation using HC–SWG in Europe

We first evaluate the HC–SWG forecast of the precipitation averages  $t_0$  to  $T$  over Europe using the CRPSS with respect to persistence and climatology at each  $T$ . We make two separate forecasts using the HC–SWG based on analogues of Z500 computed from the ensemble mean of the ECMWF and CNRM S2S reforecast for 5 days ahead for the period between respectively 2002–2021 and 1993–2017. In this article, we choose to show skill scores for December–February (DJF) and June–August (JJA) for the sake of brevity. The computations of the CRPS were made using observations of precipitation from E-OBS (Haylock *et al.*, 2008).

We find that the HC–SWG forecasts show positive improvement against persistence and climatology in different European locations in DJF and JJA (positive values of CRPSS). We notice that the CRPSS for the HC–SWG forecasts using analogues from the ECMWF reforecast has higher scores against persistence and climatology than using analogues from the CNRM reforecast. The CRPSS against persistence is higher in the summer, in particular in July in southwest Europe (Figure 4). However, for the winter, the CRPSS against persistence is higher for northern Europe, mainly in France and Germany. The CRPSS for the HC–SWG forecasts using analogues of the CNRM reforecast still shows a positive improvement for the different locations with smaller values of 0.2 to 0.3 (Figure 4). The higher skill values are also obtained in southern Europe. The difference in the HC–SWG forecast skill using the ECMWF and CNRM reforecast could be mainly related to the difference in the configuration of the two dynamic models and the resolution of each model. Similar results have been found by Ardilouze *et al.* (2021) for 2 m temperature over Europe, where scores highlighted a better performance of the ECMWF over CNRM for every lead time (weeks).

### 4.2 | Evaluation of the forecast at the station level

We evaluate the forecast of precipitation using the HC–SWG at the station level for Orly, Berlin, Toulouse, and Madrid. The motivation behind choosing those particular stations despite the availability of precipitation observations all over Europe is to ensure a comparison with the previous work Krouma *et al.* (2022). The computations of the CRPS were made using observations of precipitation from ECA&D databases.

At a local scale, we find that the CRPSS shows a positive improvement for  $T \leq 35$  days for the different locations studied for both winter (DJF) and summer (JJA) seasons, as shown in Figure 5 for the HC–SWG forecast using analogues of the ECMWF reforecast. We notice that the CRPSS against persistence is higher for Madrid and Toulouse, in particular for JJA, which is consistent with the result in Section 4.1. We notice that the CRPSS against persistence is stable for the four stations in DJF. However, the CRPSS against the climatology decreases with lead time but is still positive within 35 days. This indicates that the HC–SWG performs better than persistence for the different lead times. We argue the increasing of the CRPSS at a lead time of 35 days using the decomposition of the CRPS, as represented in Figure 9. Indeed, we find that the HC–SWG forecast has small reliability values for the different lead times and stations. This confirms the good performance of the HC–SWG. However, we notice that the reliability increases from  $T = 25$  days in Madrid and Berlin compared with the rest of the areas studied, which can explain the highest CRPSS with respect to persistence. This could be related to a larger spread of the ensemble forecast of the HC–SWG from  $T = 25$  days. Similar results were found for the HC–SWG forecast using analogues of the CNRM reforecasts (Figure A1).

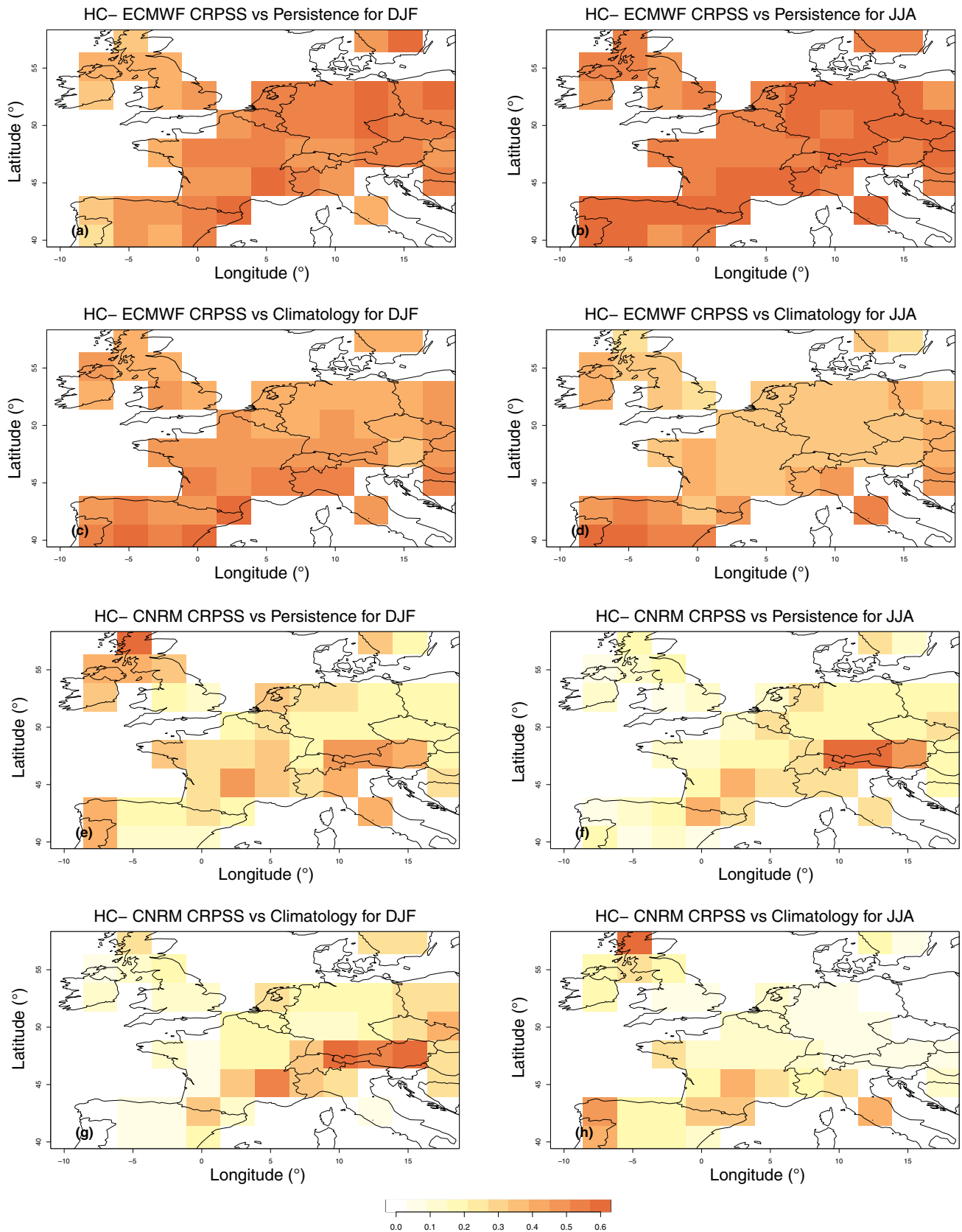
### 4.3 | Comparison of HC–SWG forecasts derived from CNRM or ECMWF

We compare the performance of the HC–SWG using analogues of the ECMWF and CNRM ensemble reforecasts at the station level. We notice that the CRPSS for both sets of HC–SWG forecasts and the four studied areas shows a positive improvement over the climatology (Figure A1).

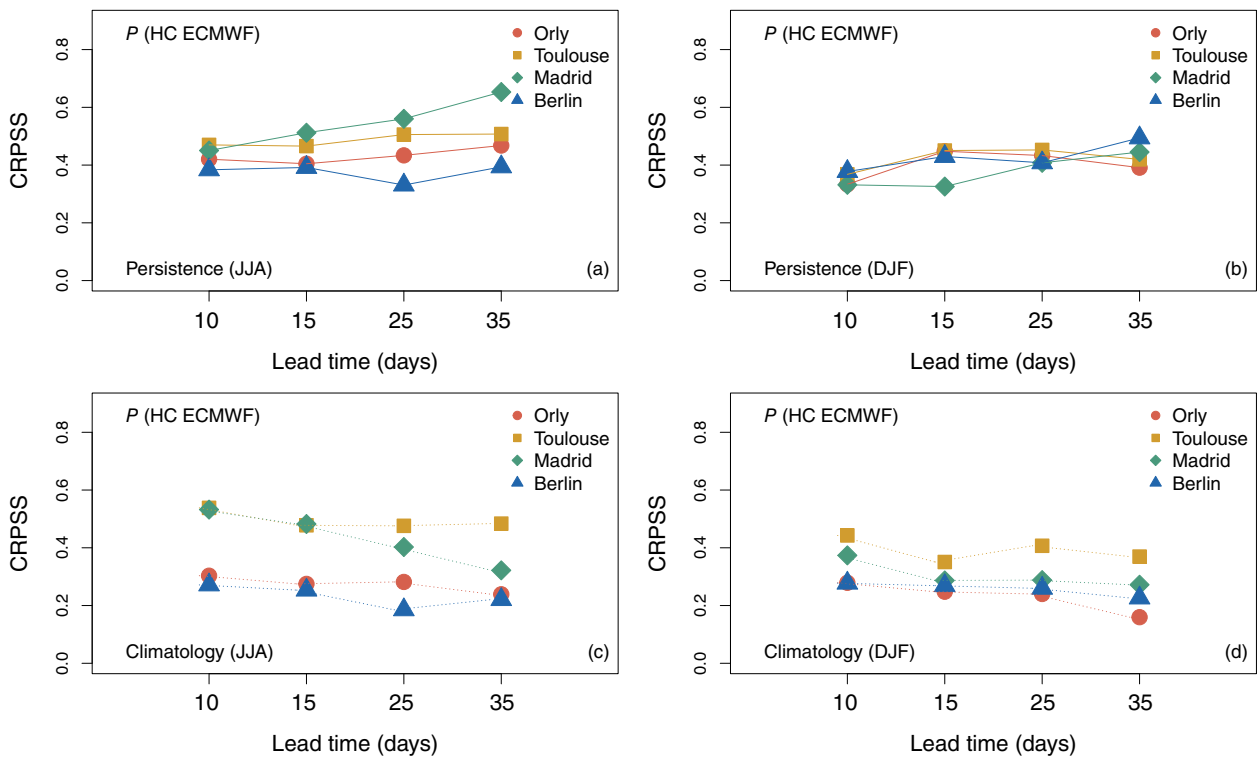
Although the performance of the HC–SWG using analogues computed from ECMWF or CNRM ensemble mean reforecasts of Z500 hPa for  $\delta = 5$  days ahead is different at the European level, we found a comparable performance and skill scores when we compare them at the station level. That may be related to some reasons related to the configuration of the models or even to specific regions where the models perform differently from one region to another, as shown by Hewson and Pilloso (2021).

The HC–SWG shows stable CRPSS values against the climatology (Figure A1). The CRPSS values decreases with  $T$  against climatology. Comparing the correlation between the average of the 100 simulations of the HC–SWG and the observations (ECA&D), we notice a tiny difference in the correlation between the two forecasts, as illustrated in Figure 6.

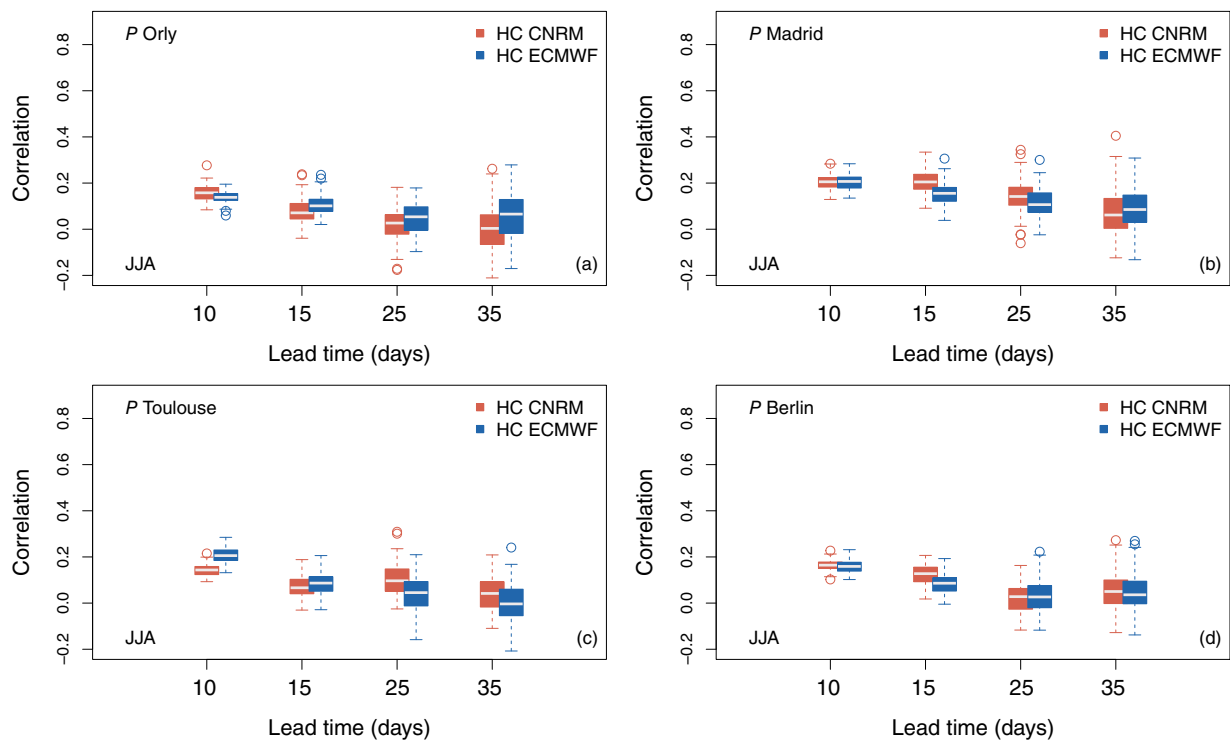
When comparing the forecast skill of the HC–SWG with that of the SWG in Krouma *et al.* (2022), we find



**FIGURE 4** Continuous rank probability skill score (CRPS) with respect to persistence and climatology for the forecast of the European precipitation with the hindcast–stochastic weather generator (HC–SWG) forecast using analogues of European Centre of Medium-range Weather Forecasts (ECMWF) and Centre National de la Recherche Météorologique (CNRM) dynamical models for December–February (DJF) and June–August (JJA) for a lead time of 10 days. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 5** Continuous rank probability skill score (CRPSS) with respect to climatology in Europe for the hindcast–stochastic weather generator (HC–SWG) forecast using European Centre of Medium-range Weather Forecasts (ECMWF) dynamical model for December–February (DJF) and June–August (JJA). [Colour figure can be viewed at wileyonlinelibrary.com]



**FIGURE 6** Comparison of the rank correlation between the hindcast (HC)–stochastic weather generator forecast using analogues of the European Centre of Medium-range Weather Forecasts (ECMWF) and Centre National de la Recherche Météorologique (CNRM) reforecasts for all lead times and for the different studied stations for December–February (DJF) and June–August (JJA). [Colour figure can be viewed at wileyonlinelibrary.com]



**TABLE 3** Comparison between the continuous ranked probability skill score (CRPSS) with respect to the climatology of the hindcast (HC)–stochastic weather generator (SWG) and the SWG in Krouma *et al.* (2022) in June–August for  $T = 10$  and 20 days.

Location	CRPSS			
	$T = 10$ days		$T = 20$ days	
	HC–SWG	SWG	HC–SWG	SWG
Orly	<b>0.38</b>	0.28	<b>0.38</b>	0.2
Madrid	<b>0.58</b>	0.42	<b>0.55</b>	0.38
Berlin	<b>0.35</b>	0.32	<b>0.30</b>	0.22
Toulouse	<b>0.30</b>	0.58	<b>0.50</b>	0.31

Note: The bold was used to indicate the values of the CRPSS from the HC–SWG.

that considering Z500 analogues from the ECMWF and CNRM S2S ensemble mean reforecast for 5 days ahead helped to improve the precipitation forecast for  $T$  up to 35 days in different locations in Europe. The forecast at the station level with the SWG in Krouma *et al.* (2022, e.g., fig. 4) gave a good forecast skill with respect to climatology and persistence for  $T$  up to 10 days; and the CRPSS was higher during the DJF, with values of CRPSS with respect to climatology and persistence ranges of respectively [0.57, 0.48] and [0.6, 0.42] (Krouma *et al.*, 2022). However, the CRPSS decreases considerably during JJA with the SWG, as shown in Table 3. With the HC–SWG, the CRPSS with respect to the climatology and the persistence shows a positive improvement for both seasons DJF and JJA and stays stable for the different lead times (Figure A1 and Table 3). For instance, the HC–SWG forecasts for Berlin are showing good skill (CRPSS against persistence and climatology) for the different lead times using analogues of Z500 from ECMWF and CNRM reforecasts compared with the SWG forecast of precipitation in Berlin based on analogues of Z500 computed from reanalyses (Table 3) (Krouma *et al.*, 2022, fig. 4).

Next, we compared the ROC curves for the HC–SWG forecasts based on analogues of Z500 from ECMWF and CNRM dynamical models for the different areas studied (Berlin, Madrid, Orly, Toulouse) and for lead times going from 10 to 35 days. The ROC curve helps to determine the discrimination skill of the HC–SWG. We considered an event of precipitation when the daily amount of precipitation is above  $P \geq 1$  mm; otherwise, there is a non-event (Figure 7a). We noticed that the HC–SWG is able to distinguish between events and non-events of precipitation for the different areas studied until  $T = 15$  days (Figure 7a). For Madrid, we notice that the HC–SWG is able to distinguish between precipitation events and non-events until 35 days. We found that the HC–SWG forecasts from analogues of ECMWF or CNRM dynamical models show

the same behaviour except for Berlin, where we noticed that the discrimination skill persists until 25 days using analogues of the ECMWF model. The AUC values range between 0.77 and 0.58 at  $T = 15$  days (Figure 7a). The HC–SWG keeps its skill to distinguish between events and non-events of precipitation as the AUC values are over 0.5; that is, above the diagonal of the ROC curve. We notice that the positive rate is 0.85 for 10 days for Madrid, Orly, Toulouse, and Berlin using analogues of ECMWF or CNRM, but it decreases differently with lead time and from one station to another.

In addition, we verified the discrimination skill of the HC–SWG using the analogues of both ECMWF and CNRM models for extreme precipitation (Figure 7b). For  $T = 10$  days, the AUC values range between 0.67 (Madrid) and 0.57 (Berlin), indicating that the HC–SWG conserves its capacity to distinguish between non-events of extreme precipitation ( $P < q_{90}$ ) and events of extreme precipitation ( $P \geq q_{90}$ ).

We assessed the statistical significance of the AUC in both cases using the Mann–Whitney test (Wilks, 2011), which we describe in Appendix B. We found that the AUC of the HC–SWG using the ECMWF reforecast is significant until  $T = 25$  days for Madrid, Orly, and Berlin (Table B1). For extreme precipitation, the discrimination skill is significant until 10 days for all stations and until  $T = 15$  days with the ECMWF reforecast for Madrid and Orly, as highlighted in Table B1.

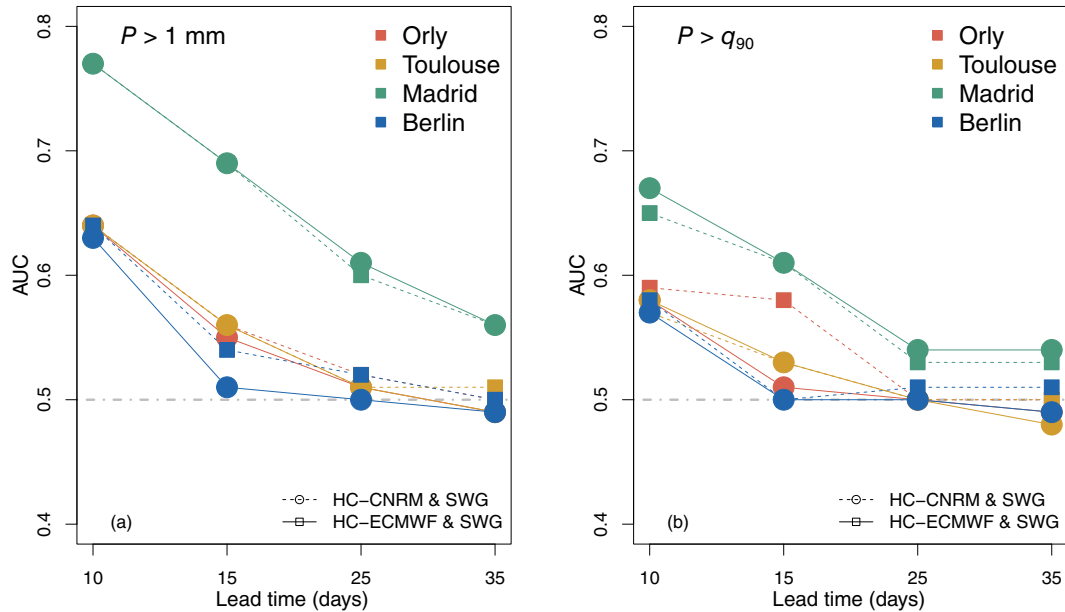
#### 4.4 | Comparison of HC–SWG and ECMWF precipitation forecast

We compared the HC–SWG precipitation forecasts with the ECMWF precipitation forecast by computing the fair CRPSS as shown in Equation (3). We used the ECMWF precipitation forecast as a reference to the HC–SWG forecasts. Positive values indicate an improvement of the HC–SWG forecast over the ECMWF or the CNRM precipitation forecasts. We choose to compare the HC–SWG forecast with the ECMWF precipitation forecast up to 25 days as the AUC values are mainly significant until  $T = 25$  days, as shown earlier (see also Appendix B).

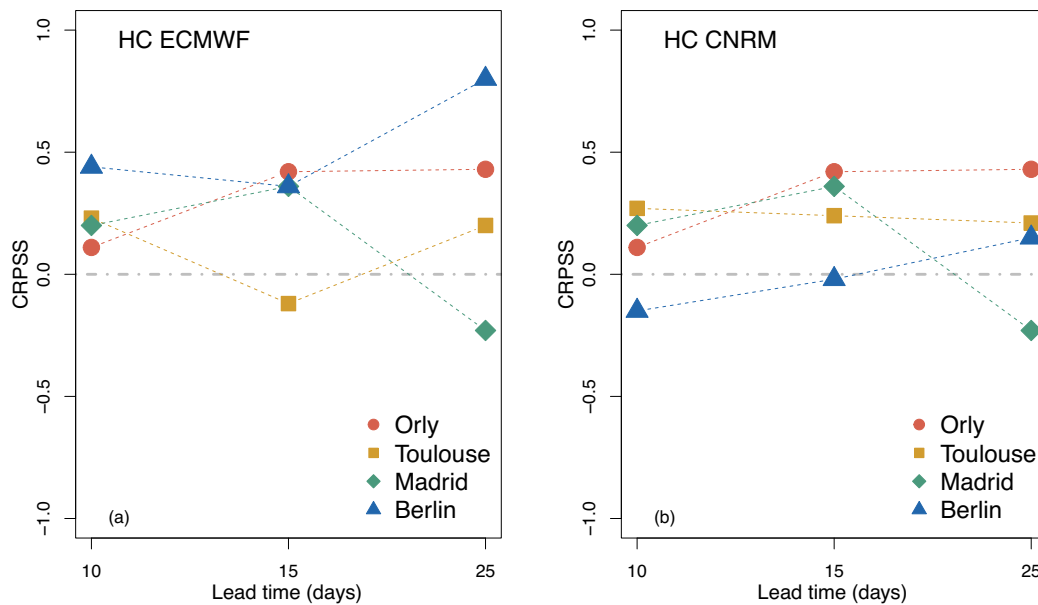
For the different areas studied, we compared the CRPSS of the HC–SWG based on the ECMWF or CNRM reforecast with the CRPSS of the ECMWF precipitation forecast, both against the climatology (see Figure A1). We found that the CRPSS of the HC–SWG using either analogues from ECMWF or CNRM reforecasts is higher than the CRPSS against the climatology of the ECMWF forecast of the precipitation, in particular for Berlin and Orly (Figure A1a,d).

The comparison shows that the HC–SWG improves the precipitation forecast up to 25 days with respect to the





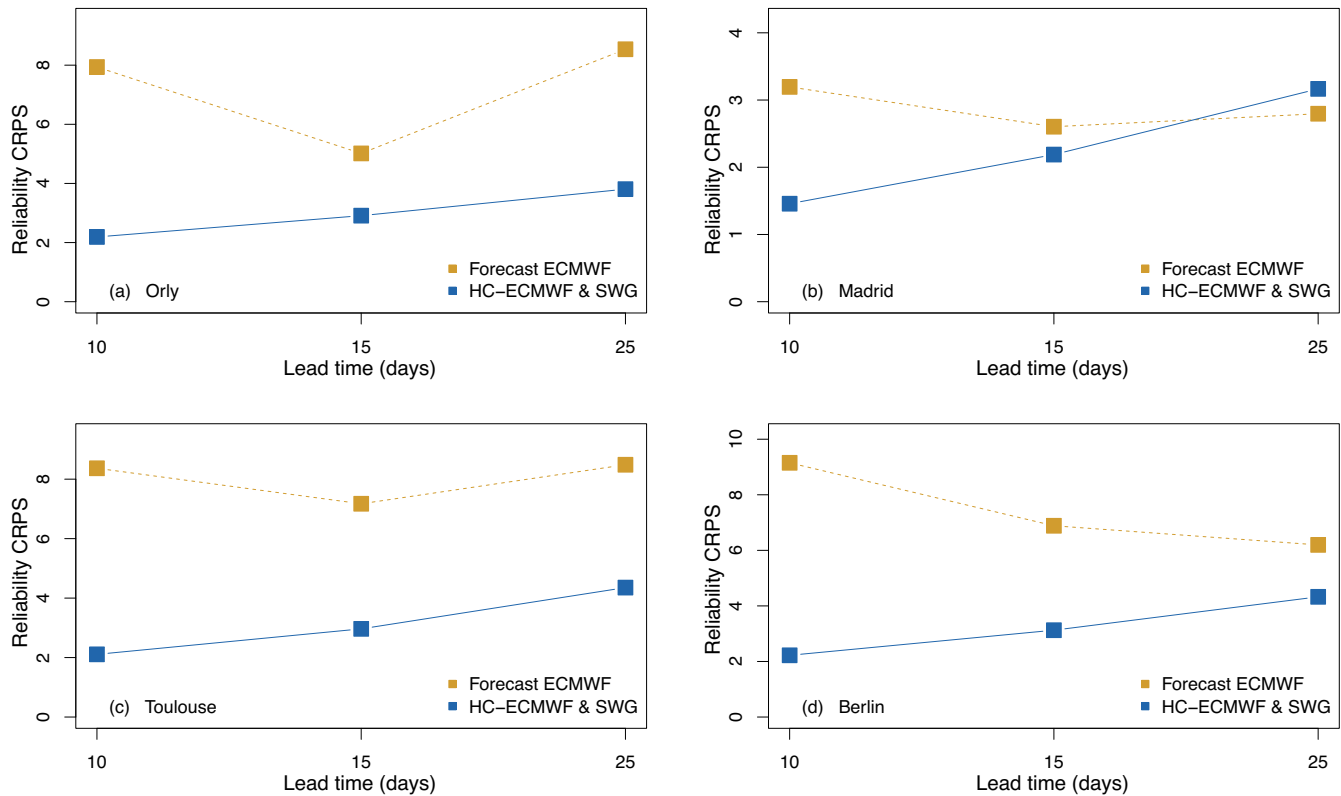
**FIGURE 7** Area under the curve (AUC) for Orly, Madrid, Toulouse, and Berlin for hindcast (HC)-stochastic weather generator (SWG) precipitation forecast based on European Centre of Medium-range Weather Forecasts (ECMWF; solid line with square) and Centre National de la Recherche Météorologique (CNRM; dash line with circle) reforecasts. (a) AUC for precipitation considering 1 mm as a threshold; (b) AUC for extreme precipitation that exceeds the 90th quantile. The dashed grey line represents the diagonal of the ROC curve (AUC = 0.5). [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 8** Continuous rank probability skill score (CRPSS) between the European Centre of Medium-range Weather Forecasts (ECMWF) precipitation forecast and the hindcast-stochastic weather generator (HC-SWG) forecasts based on the ECMWF and Centre National de la Recherche Météorologique (CNRM) reforecasts. The CRPSS is computed between the HC-SWG and a reference, in this case the ECMWF precipitation forecast. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

ECMWF precipitation forecast by using either the CNRM or ECMWF analogues of Z500. As shown in Figure 8, we found that the HC-SWG forecast of precipitation based on analogues of ECMWF (Figure 8a) and CNRM (Figure 8b)

reforecasts of Z500 is more skilful than the ECMWF precipitation forecast for different lead times and for the locations studied, except for Toulouse and Madrid respectively at a lead time of 10 and 25 days. The HC-SWG



**FIGURE 9** Comparison between the continuous ranked probability score (CRPS) reliability of the hindcast (HC)–stochastic weather generator (SWG) forecasts based on European Centre of Medium-range Weather Forecasts (ECMWF) reforecasts and the CRPS reliability of the ECMWF precipitation forecast. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4733)]

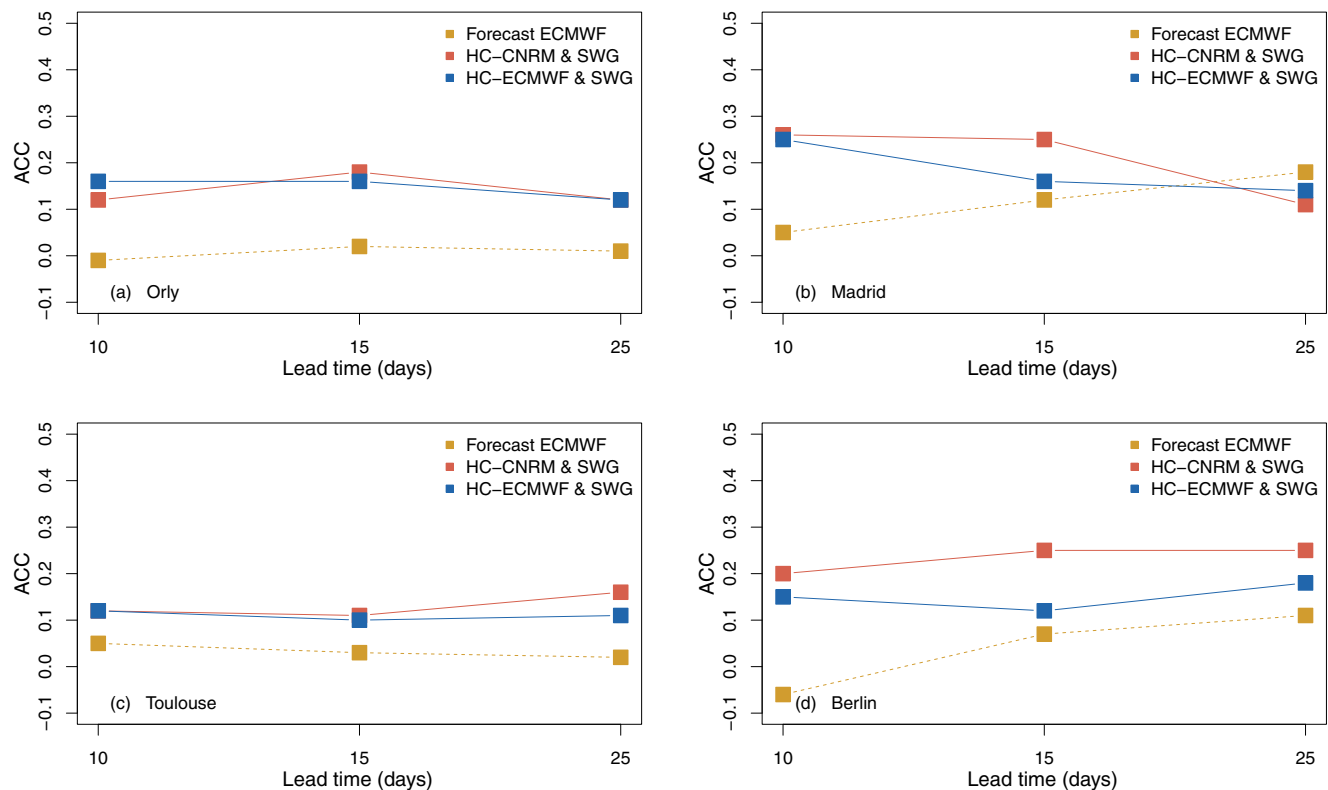
forecasts based on analogues of the CNRM reforecasts of Z500 show no improvement for Berlin for  $T < 15$  days. The ECMWF precipitation forecast shows a better skill for Madrid at 25 days compared with the HC–SWG forecast. Moreover, we notice that the improvement of the HC–SWG using analogues of Z500 of the ECMWF reforecasts is higher than by using CNRM reforecasts.

To better explain these results, we compared the CRPS reliability of the HC–SWG forecasts of precipitation using the ECMWF and CNRM reforecasts with the CRPS reliability of the ECMWF precipitation forecast (Figure 9). We found that the HC–SWG forecast is more reliable for the different lead times and stations compared with the ECMWF precipitation forecast. This confirms the good performance of the HC–SWG compared with the numerical model. However, we notice that the reliability of the ECMWF precipitation forecast is lower than that of the HC–SWG forecast in Madrid at  $T = 25$  days, which can explain the negative CRPS of the HC–SWG at 25 days for Madrid with respect to the ECMWF precipitation forecast. More results about the CRPS decomposition and comparison are detailed in Appendix C.

Both the HC–SWG and the SWG in (Krouma *et al.*, 2022, tab. 6) show comparable improvement in the forecast of precipitation compared with the ECMWF precipitation

forecast. The added value of the HC–SWG is that the forecast is initiated at  $\delta = 5$  days ahead compared with the SWG in Krouma *et al.* (2022), where reanalyses were used to define analogues at  $\delta = 0$ . Also, the HC–SWG can forecast precipitation for further lead times  $T$  and with better skill (Table 3) compared with the SWG in Krouma *et al.* (2022).

To confirm the forecast skill of the HC–SWG precipitation forecast either by using the ECMWF or the CNRM reforecast against the ECMWF precipitation forecast, we computed the ACC for the anomalies of the ensemble mean of the HC–SWG and ECMWF precipitation forecasts (Figure 10). For the four regions studied, we noticed that the ACCs of the HC–SWG using the ECMWF or CNRM reforecast are comparable except for Orly (Figure 10a). The ACC of the HC–SWG for the different areas studied decreases with lead times  $T$ . Comparing it with the ACC of the ECMWF precipitation forecast, we found that the ACCs of the HC–SWG lead times ranging from  $T = 10$  to  $T = 25$  days are higher. Even if the ACC HC–SWG values are lower than 1, which does not indicate a perfect affinity with the observations, they outperform the ACC values of the ECMWF precipitation forecast, which shows negative or close to zero values at some lead times.



**FIGURE 10** Comparison between anomaly correlation coefficient (ACC) of the European Centre of Medium-range Weather Forecasts (ECMWF) precipitation forecast and the hindcast (HC)-stochastic weather generator (SWG) forecasts based on the ECMWF and Centre National de la Recherche Météorologique (CNRM) reforecasts. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

## 5 | CONCLUSION

The use of the S2S ECMWF and CNRM ensemble mean reforecasts of the geopotential height at 500 hPa for 5 days ahead helped to improve statistical features of the precipitation forecast over Europe with an SWG. The HC-SWG confirmed its capacity to forecast precipitation for up to 35 days in Europe and at the station level. However, the forecast skill differs from one region to another and remains higher over southernmost locations for either summer or winter. The SWG based on analogues circulation has been used in previous studies (Krouma *et al.*, 2022, 2023; Yiou & Déandréis, 2019) to forecast temperature, precipitation, and the Madden-Julian oscillation. In this version, the HC-SWG confirmed the capacity of the SWG and analogues circulation to forecast precipitation at a local scale similar to the SWG in Krouma *et al.* (2022), showing good skill scores in different regions despite the variety of the local weather and for a longer lead time of 35 days compared with the 10 days in Krouma *et al.* (2022). The HC-SWG confirmed its capacity to distinguish between events and non-events of precipitation as well as extreme precipitation for a lead time of at least 10 days. In addition, the comparison with the ECMWF precipitation forecast

confirmed the performance of the HC-SWG forecasts for  $T = 25$  days.

Comparing the HC-SWG precipitation forecast with the SWG precipitation forecast in Krouma *et al.* (2022), we also noticed the added value of using the analogues computed from reforecasts of Z500 from dynamical models at  $\delta = 5$  days instead of using reanalyses of the atmospheric circulation with  $\delta = 0$ .

Combining dynamical models with the SWG allowed the improvement of the precipitation forecast to the sub-seasonal lead time. This work highlights the contribution of dynamical models with a correct initialization (Zuo *et al.*, 2016) to get a skilful forecast. These results can be considered as a starting point to implement an operational forecast from the HC-SWG at the subseasonal lead time. This can help to verify the forecast skill of the HC-SWG in real time. The HC-SWG can also be tested to forecast other meteorological variables in different regions of the globe. This can help to improve forecasts and verify the capacity of our forecasting tool.

We used the ensemble mean of the Z500 for 5 days ahead to forecast the precipitation. This approach showed a capacity to improve the European precipitation forecast. However, this approach must be verified and adjusted while using the atmospheric circulation for more  $\delta$  days

ahead, like 10 or even more. The verification would depend on the atmospheric circulation field, its forecast skill at different lead times, and the ensemble member's quality, which would help avoid smoothing data and getting wrong analogues.


## ACKNOWLEDGEMENTS

This work is part of the EU International Training Network (ITN) Climate Advanced Forecasting of subseasonal Extremes (CAFE). The project receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 813844.

## DATA AVAILABILITY STATEMENT

Data openly available from <https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/>.

## ORCID

Meriem Krouma  <https://orcid.org/0000-0003-0617-9956>

Linus Magnusson  <https://orcid.org/0000-0003-4707-2231>

Constantin Ardilouze  <https://orcid.org/0000-0003-3949-4450>

Lauriane Batté  <https://orcid.org/0000-0002-7903-9762>

## REFERENCES

- Ailliot, P., Allard, D., Monbet, V. & Naveau, P. (2015) Stochastic weather generators: an overview of weather type models. *Journal de la Société Française de Statistique*, 156(1), 101–113.
- Ardilouze, C., Specq, D., Batté, L. & Cassou, C. (2021) Flow dependence of wintertime subseasonal prediction skill over Europe. *Weather and Climate Dynamics*, 2(4), 1033–1049 <https://wcd.copernicus.org/articles/2/1033/2021/>
- Atencia, A. & Zawadzki, I. (2014) A comparison of two techniques for generating nowcasting ensembles. Part I: Lagrangian ensemble technique. *Monthly Weather Review*, 142(11), 4036–4052.
- Barnes, A.P., McCullen, N. & Kjeldsen, T.R. (2023) Forecasting seasonal to sub-seasonal rainfall in Great Britain using convolutional neural networks. *Theoretical and Applied Climatology*, 151(1–2), 421–432.
- Batté, L. & Déqué, M. (2016) Randomly correcting model errors in the ARPEGE-climate v6.1 component of CNRM-CM: applications for seasonal forecasts. *Geoscientific Model Development*, 9(6), 2055–2076. Available from: <https://gmd.copernicus.org/articles/9/2055/2016/>
- Ben Bouallègue, Z., Cooper, F., Chantry, M., Düben, P., Bechtold, P. & Sandu, I. (2023) Statistical modeling of 2-m temperature and 10-m wind speed forecast errors. *Monthly Weather Review*, 151(4), 897–911 <https://journals.ametsoc.org/view/journals/mwre/151/4/MWR-D-22-0107.1.xml>
- Blanchet, J., Stalla, S. & Creutin, J.D. (2018) Analogy of multiday sequences of atmospheric circulation favoring large rainfall accumulation over the French Alps. *Atmospheric Science Letters*, 19(3), e809. Available from: <http://doi.wiley.com/10.1002/asl.809>
- Butler, A.H., Charlton-Perez, A., Domeisen, D.I., Simpson, I.R. & Sjöberg, J. (2019) Predictability of northern hemisphere final stratospheric warmings and their surface impacts. *Geophysical Research Letters*, 46(17–18), 10578–10588.
- Cassou, C. (2008) Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature*, 455(7212), 523–527.
- Chen, D., Dai, A. & Hall, A. (2021) The convective-to-total precipitation ratio and the “drizzling” bias in climate models. *Journal of Geophysical Research: Atmospheres*, 126(16), e2020JD034198. Available from: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JD034198>
- Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M. & Jones, P.D. (2018) An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17), 9391–9409 <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JD028200>
- DeFlorio, M.J., Waliser, D.E., Ralph, F.M., Guan, B., Goodman, A., Gibson, P.B. et al. (2019) Experimental subseasonal-to-seasonal (S2S) forecasting of atmospheric rivers over the western United States. *Journal of Geophysical Research: Atmospheres*, 124(21), 11242–11265.
- Domeisen, D.I., White, C.J., Afargan-Gerstman, H., Muñoz, Á.G., Janiga, M.A., Vitart, F. et al. (2022) Advances in the subseasonal prediction of extreme events: relevant case studies across the globe. *Bulletin of the American Meteorological Society*, 103(6), E1473–E1501.
- Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874 <https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X>
- Ferro, C.A.T. (2007) A probability model for verifying deterministic forecasts of extreme events. *Weather and Forecasting*, 22(5), 1089–1100.
- Ferro, C.A.T. (2014) Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683), 1917–1923 <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2270>
- Frei, C., Schöll, R., Fukutome, S., Schmidli, J. & Vidale, P.L. (2006) Future change of precipitation extremes in Europe: inter-comparison of scenarios from regional climate models. *Journal of Geophysical Research: Atmospheres*, 111, D06105. Available from: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD005965>
- Haylock, M.R., Hofstra, N., Tank, A.M.G.K., Klok, E.J., Jones, P.D. & New, M. (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres*, 113, D20119.
- Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15, 559–570.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.
- Hewson, T.D. & Pilloso, F.M. (2021) A low-cost post-processing technique improves weather forecasts around the world. *Communications Earth & Environment*, 2(1), 132.
- Kam, J., Kim, S. & Roundy, J.K. (2021) Did a skillful prediction of near-surface temperatures help or hinder forecasting of the 2012 US drought. *Environmental Research Letters*, 16(3), 034044.



- Karl, T.R., Wang, W.C., Schlesinger, M.E., Knight, R.W. & Portman, D. (1990) A method of relating general circulation model simulated climate to the observed local climate. Part I: seasonal statistics. *Journal of Climate*, 3(10), 1053–1079. Available from: [https://journals.ametsoc.org/view/journals/clim/3/10/1520-0442\\_1990\\_003\\_1053\\_amorgc\\_2\\_0\\_co\\_2.xml](https://journals.ametsoc.org/view/journals/clim/3/10/1520-0442_1990_003_1053_amorgc_2_0_co_2.xml)
- Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A. et al. (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, 22(12), 1441–1453 <http://doi.wiley.com/10.1002/joc.773>
- Klein, W.H., Lewis, B.M. & Enger, I. (1959) Objective prediction of five-day mean temperatures during WINTER. *Journal of Atmospheric Sciences*, 16(6), 672–682 [https://journals.ametsoc.org/view/journals/atsc/16/6/1520-0469\\_1959\\_016\\_0672\\_opofdm\\_2\\_0\\_co\\_2.xml](https://journals.ametsoc.org/view/journals/atsc/16/6/1520-0469_1959_016_0672_opofdm_2_0_co_2.xml)
- Koster, R.D., Mahanama, S., Yamada, T., Balsamo, G., Berg, A., Boiserie, M. et al. (2010) Contribution of land surface initialization to subseasonal forecast skill: first results from a multi-model experiment. *Geophysical Research Letters*, 37(2), L02402.
- Krouma, M., Silini, R. & Yiou, P. (2023) Ensemble forecast of an index of the Madden–Julian oscillation using a stochastic weather generator based on circulation analogs. *Earth System Dynamics*, 14(1), 273–290 <https://esd.copernicus.org/articles/14/273/2023/>
- Krouma, M., Yiou, P., Déandreis, C. & Thao, S. (2022) Assessment of stochastic weather forecast of precipitation near European cities, based on analogs of circulation. *Geoscientific Model Development*, 15(12), 4941–4958 <https://gmd.copernicus.org/articles/15/4941/2022/>
- Lau, W.K.M. & Waliser, D.E. (2011) *Intraseasonal variability in the atmosphere-ocean climate system*. Berlin: Springer Science & Business Media.
- Lin, H. & Wu, Z. (2011) Contribution of the autumn Tibetan plateau snow cover to seasonal prediction of North American winter temperature. *Journal of Climate*, 24(11), 2801–2813.
- Lorenz, E.N. (1969) Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, 26(4), 636–646.
- Magnusson, L. & Källén, E. (2013) Factors influencing skill improvements in the ECMWF forecasting system. *Monthly Weather Review*, 141(9), 3142–3153 <https://journals.ametsoc.org/view/journals/mwre/141/9/mwr-d-12-00318.1.xml>
- Mariotti, A., Ruti, P.M. & Rixen, M. (2018) Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *Npj Climate and Atmospheric Science*, 1(1), 4.
- Merryfield, W.J., Baehr, J., Batté, L., Becker, E.J., Butler, A.H., Coelho, C.A. et al. (2020) Current and emerging developments in subseasonal to decadal prediction. *Bulletin of the American Meteorological Society*, 101(6), E869–E896.
- Newman, M., Sardeshmukh, P.D., Winkler, C.R. & Whitaker, J.S. (2003) A study of subseasonal predictability. *Monthly Weather Review*, 131(8), 1715–1732.
- Olanijan, E., Adefisan, E.A., Oni, F., Afiesimama, E., Balogun, A.A. & Lawal, K.A. (2018) Evaluation of the ECMWF sub-seasonal to seasonal precipitation forecasts during the peak of West Africa monsoon in Nigeria. *Frontiers in Environmental Science*, 6, 4.
- Palmer, T.N. (2000) Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics*, 63(2), 71–116.
- Peng, X., Che, Y. & Chang, J. (2013) A novel approach to improve numerical weather prediction skills by using anomaly integration and historical data. *Journal of Geophysical Research: Atmospheres*, 118(16), 8814–8826.
- Penny, S.G., Akella, S., Balmaseda, M.A., Browne, P., Carton, J.A., Chevallier, M. et al. (2019) Observational needs for improving ocean and coupled reanalysis, S2S prediction, and decadal prediction. *Frontiers in Marine Science*, 6, 391.
- Rashid, H.A., Hendon, H.H., Wheeler, M.C. & Alves, O. (2011) Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system. *Climate Dynamics*, 36(3), 649–661.
- Rivoire, P., Martius, O., Naveau, P. & Tuel, A. (2023) Assessment of subseasonal-to-seasonal (S2S) ensemble extreme precipitation forecast skill over Europe. *Natural Hazards and Earth System Sciences*, 23(8), 2857–2871 <https://nhess.copernicus.org/articles/23/2857/2023/>
- Robertson, A.W. & Vitart, F. (2019) *Sub-seasonal to seasonal prediction: the gap between weather and climate forecasting*, Vol. 569. Amsterdam: Elsevier.
- Specq, D. & Batté, L. (2020) Improving subseasonal precipitation forecasts through a statistical-dynamical approach: application to the southwest tropical Pacific. *Climate Dynamics*, 55, 1913–1927.
- Stan, C. & Straus, D.M. (2019) The impact of cloud representation on the sub-seasonal forecasts of atmospheric teleconnections and preferred circulation regimes in the northern hemisphere. *Atmosphere-Ocean*, 57(3), 233–248.
- Subramanian, A.C., Balmaseda, M.A., Centurioni, L., Chattopadhyay, R., Cornuelle, B.D., DeMott, C. et al. (2019) Ocean observations to improve our understanding, modeling, and forecasting of subseasonal-to-seasonal variability. *Frontiers in Marine Science*, 6, 427.
- Toth, Z., Talagrand, O., Candille, G. & Zhu, Y. (2003) Probability and ensemble forecasts. *Forecast Verification: A practitioner's Guide in Atmospheric Science*, 137, 163.
- Vitart, F. (2014) Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140(683), 1889–1899.
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C. et al. (2017) The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98(1), 163–173 <https://journals.ametsoc.org/doi/10.1175/BAMS-D-16-0017.1>
- Vitart, F., Balsamo, G., Bidlot, J.R., Lang, S., Tsonevsky, I., Richardson, D. et al. (2019) *Use of ERA5 to initialize ensemble re-forecasts*. ECMWF. Technical Memoranda number: 841. <https://www.ecmwf.int/node/18872>
- Wei, J. & Dirmeyer, P.A. (2019) Sensitivity of land precipitation to surface evapotranspiration: a nonlocal perspective based on water vapor transport. *Geophysical Research Letters*, 46(21), 12588–12597.
- Weyn, J.A., Durran, D.R., Caruana, R. & Cresswell-Clay, N. (2021) Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, 13(7), e2021MS002502.
- White, C.J., Domeisen, D.I., Acharya, N., Adefisan, E.A., Anderson, M.L., Aura, S. et al. (2022) Advances in the application and utility of subseasonal-to-seasonal predictions. *Bulletin of the American Meteorological Society*, 103, E1448–E1472.
- Wilks, D.S. (2011) *Statistical methods in the atmospheric sciences*, Vol. 100. Cambridge, MA: Academic Press.



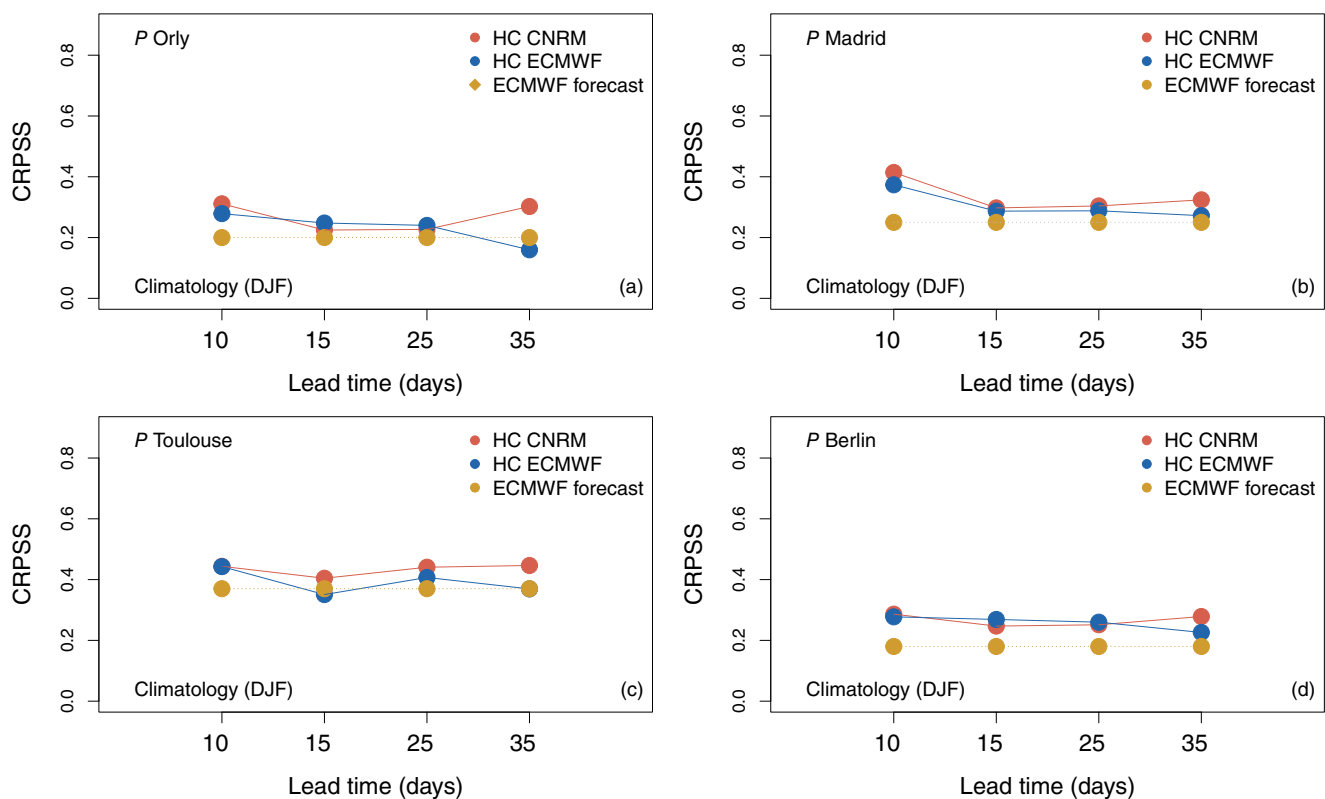
- Wilks, D.S. & Wilby, R.L. (1999) The weather generation game: a review of stochastic weather models. *Progress in Physical Geography: Earth and Environment*, 23(3), 329–357.
- Woolnough, S., Vitart, F. & Balmaseda, M. (2007) The role of the ocean in the Madden–Julian oscillation: implications for MJO prediction. *Quarterly Journal of the Royal Meteorological Society*, 133(622), 117–128.
- Yiou, P. (2014) AnaWEGE: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, 7(2), 531–543 <https://www.geosci-model-dev.net/7/531/2014/>
- Yiou, P. & Déandréis, C. (2019) Stochastic ensemble climate forecast with an analogue model. *Geoscientific Model Development*, 12(2), 723–734 <https://www.geosci-model-dev.net/12/723/2019/>
- Yiou, P., Salameh, T., Drobinski, P., Menut, L., Vautard, R. & Vrac, M. (2013) Ensemble reconstruction of the atmospheric column from surface pressure using analogues. *Climate Dynamics*, 41(5–6), 1333–1344 <http://link.springer.com/10.1007/s00382-012-1626-3>
- Zamo, M. & Naveau, P. (2018) Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Mathematical Geosciences*, 50(2), 209–234.
- Zhang, L., Kim, T., Yang, T., Hong, Y. & Zhu, Q. (2021) Evaluation of subseasonal-to-seasonal (S2S) precipitation forecast from the North American multi-model ensemble phase II (NMME-2) over the contiguous US. *Journal of Hydrology*, 603, 127058.

- Zuo, J., Ren, H.L., Li, W. & Wang, L. (2016) Interdecadal variations in the relationship between the winter North Atlantic Oscillation and temperature in south-central China. *Journal of Climate*, 29(20), 7477–7493.

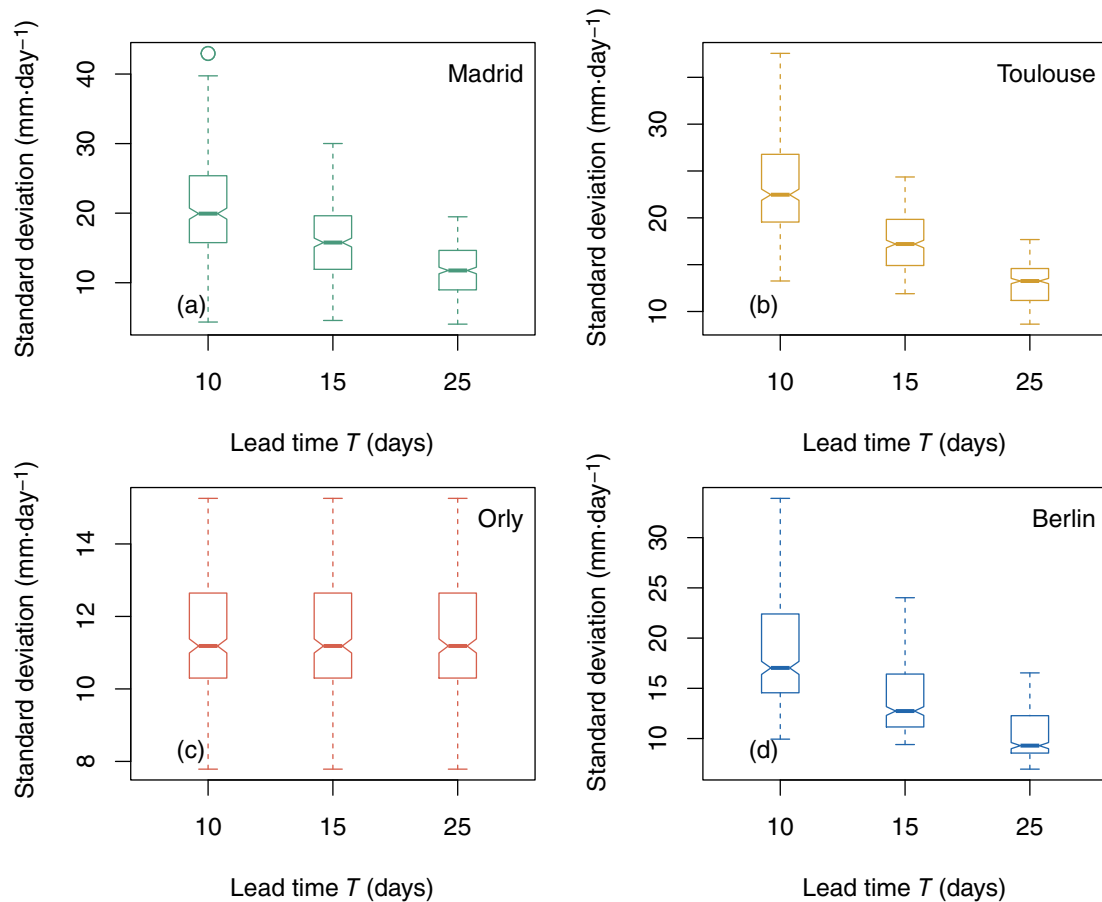
**How to cite this article:** Krouma, M., Specq, D., Magnusson, L., Ardilouze, C., Batté, L. & Yiou, P. (2024) Improving subseasonal forecast of precipitation in Europe by combining a stochastic weather generator with dynamical models. *Quarterly Journal of the Royal Meteorological Society*, 1–21. Available from: <https://doi.org/10.1002/qj.4733>

## APPENDIX A. STANDARD DEVIATION OF LEARNED VALUES OF THE CLIMATOLOGY FORECAST FOR DIFFERENT LEAD TIMES

As explained in Section 3, the added noise to the persistence and climatology forecast is estimated using the



**FIGURE A1** Comparison of the continuous ranked probability skill score (CRPSS) with respect to the climatology of the European Centre of Medium-range Weather Forecasts (ECMWF) forecast (yellow line) and hindcast (HC)–stochastic weather generator forecast using analogues of the ECMWF (blue line) and Centre National de la Recherche Météorologique (CNRM; red line) reforecasts for all lead times and for the different areas studied in winter (December–February, DJF). [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE A2** Boxplots of the standard deviation values of the learned values for the climatology for different lead times  $T = 10, 15,$  and 25 days for the different stations: (a) Madrid, (b) Toulouse, (c) Berlin, and (d) Orly. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

extended time interval equal to  $T$ . The added noise is helpful, as there are truncation errors in observations that make the empirical distribution discrete rather than continuous. It also helps to induce variability in the persistence and the climatology baselines. Figure A2 presents the boxplots of the standard deviation of the learned values for climatology forecast at different lead times for the studied stations Madrid, Toulouse, Berlin and Orly.

The standard deviations of the CRPS values decrease with lead time for Madrid, Toulouse, and Berlin (Figure A2a,b,d) and remain stable for Orly, near relatively small values (Figure A2c). This shows that the forecast sample size decreases with longer lead times  $T$ , which is related to the average forecast provided by the HC-SWG between  $t_0$  and  $[t_0, T]$ . This helps to make the forecast skill of the HC-SWG better over time.

## APPENDIX B. STATISTICAL SIGNIFICANCE OF THE AUC FOR HC-SWG PRECIPITATION FORECAST FOR DIFFERENT LEAD TIMES

We evaluated the discrimination skill of the HC-SWG forecast of European precipitation using the AUC,

as described in Section 3.3. We assessed the statistical significance of the AUC of the HC-SWG forecasts with the ECMWF and CNRM reforecasts against the  $AUC_{obs}$ . The  $AUC_{obs}$  depends on the event and non-event of precipitation in the observations.

To define the  $AUC_{obs}$  is computed as follows (Wilks, 2011, Chap. 7):

$$AUC_{obs} = \frac{1 - U}{n_1 \times n_2}, \quad (B1)$$

where  $n_1$  is the event of precipitation on the observations,  $n_2$  is the non-event of precipitation, and  $U$  is the Mann-Whitney variable defined from  $n_1$  and  $n_2$  (Wilks, 2011, Chap. 7). The interpretation of this significance test is as follows: if  $AUC_{SWG} > AUC_{obs}$ , the AUC value of the SWG is significant, otherwise it is not significant.

We applied this test for the event and non-event of precipitation where (precipitation events are above  $P \geq 1$  mm and non-event with precipitation below  $P < 1$  mm) as shown in Table B1 and extreme precipitation (where extreme precipitation events are above  $P \geq q_{90}$  and non-event with precipitation below  $P < q_{90}$ ) as illustrated in Table B2.

**TABLE B1** Significance test for area under the receiver operating characteristic curve (AUC) of Orly, Madrid, Toulouse, and Berlin for hindcast–stochastic weather generator forecast for non-event of precipitation (<1 mm) and event of precipitation (>1 mm) based on European Centre of Medium-range Weather Forecasts (ECMWF;  $AUC_{ECMWF}$ ) and Centre National de la Recherche Météorologique (CNRM;  $AUC_{CNRM}$ ) reforecasts for different lead times.

Lead time (days)	AUC	Madrid	Orly	Berlin	Toulouse
$T = 10$	$AUC_{ECMWF}$	<b>0.77</b>	<b>0.64</b>	<b>0.64</b>	<b>0.64</b>
	$AUC_{CNRM}$	<b>0.77</b>	<b>0.64</b>	<b>0.63</b>	<b>0.64</b>
	$AUC_{obs}$	0.52	0.51	0.52	0.51
$T = 15$	$AUC_{ECMWF}$	<b>0.69</b>	<b>0.56</b>	<b>0.54</b>	<b>0.56</b>
	$AUC_{CNRM}$	<b>0.69</b>	<b>0.55</b>	0.51	<b>0.56</b>
	$AUC_{obs}$	0.53	0.52	0.52	0.52
$T = 25$	$AUC_{ECMWF}$	<b>0.6</b>	<b>0.53</b>	<b>0.53</b>	0.51
	$AUC_{CNRM}$	<b>0.61</b>	0.51	0.50	0.51
	$AUC_{obs}$	0.54	0.53	0.53	0.53
$T = 35$	$AUC_{ECMWF}$	<b>0.56</b>	0.50	0.50	0.51
	$AUC_{CNRM}$	<b>0.56</b>	0.49	0.49	0.49
	$AUC_{obs}$	0.55	0.54	0.54	0.53

Note: Bold was used to indicate significant values compared to the  $AUC_{obs}$ .

**TABLE B2** Significance test for area under the receiver operating characteristic curve (AUC) of Orly, Madrid, Toulouse, and Berlin for hindcast–stochastic weather generator extreme precipitation forecast based on European Centre of Medium-range Weather Forecasts (ECMWF) and Centre National de la Recherche Météorologique (CNRM) reforecasts for different lead times.

Lead time (days)	AUC	Madrid	Orly	Berlin	Toulouse
$T = 10$	$AUC_{ECMWF}$	<b>0.65</b>	<b>0.59</b>	<b>0.58</b>	<b>0.57</b>
	$AUC_{CNRM}$	<b>0.67</b>	<b>0.58</b>	<b>0.57</b>	<b>0.58</b>
	$AUC_{obs}$	0.52	0.51	0.52	0.52
$T = 15$	$AUC_{ECMWF}$	<b>0.61</b>	<b>0.58</b>	0.50	0.53
	$AUC_{CNRM}$	<b>0.61</b>	0.51	0.50	0.53
	$AUC_{obs}$	0.54	0.54	0.54	0.54
$T = 25$	$AUC_{ECMWF}$	0.53	0.50	0.51	0.50
	$AUC_{CNRM}$	0.54	0.50	0.50	0.50
	$AUC_{obs}$	0.55	0.55	0.55	0.56
$T = 35$	$AUC_{ECMWF}$	0.53	0.50	0.51	0.50
	$AUC_{CNRM}$	0.54	0.49	0.49	0.48
	$AUC_{obs}$	0.57	0.57	0.57	0.57

Note: Bold was used to indicate significant values compared to the  $AUC_{obs}$ .

### APPENDIX C. DECOMPOSITION OF THE CRPS FOR THE HC-SWG PRECIPITATION FORECAST FOR DIFFERENT LEAD TIMES

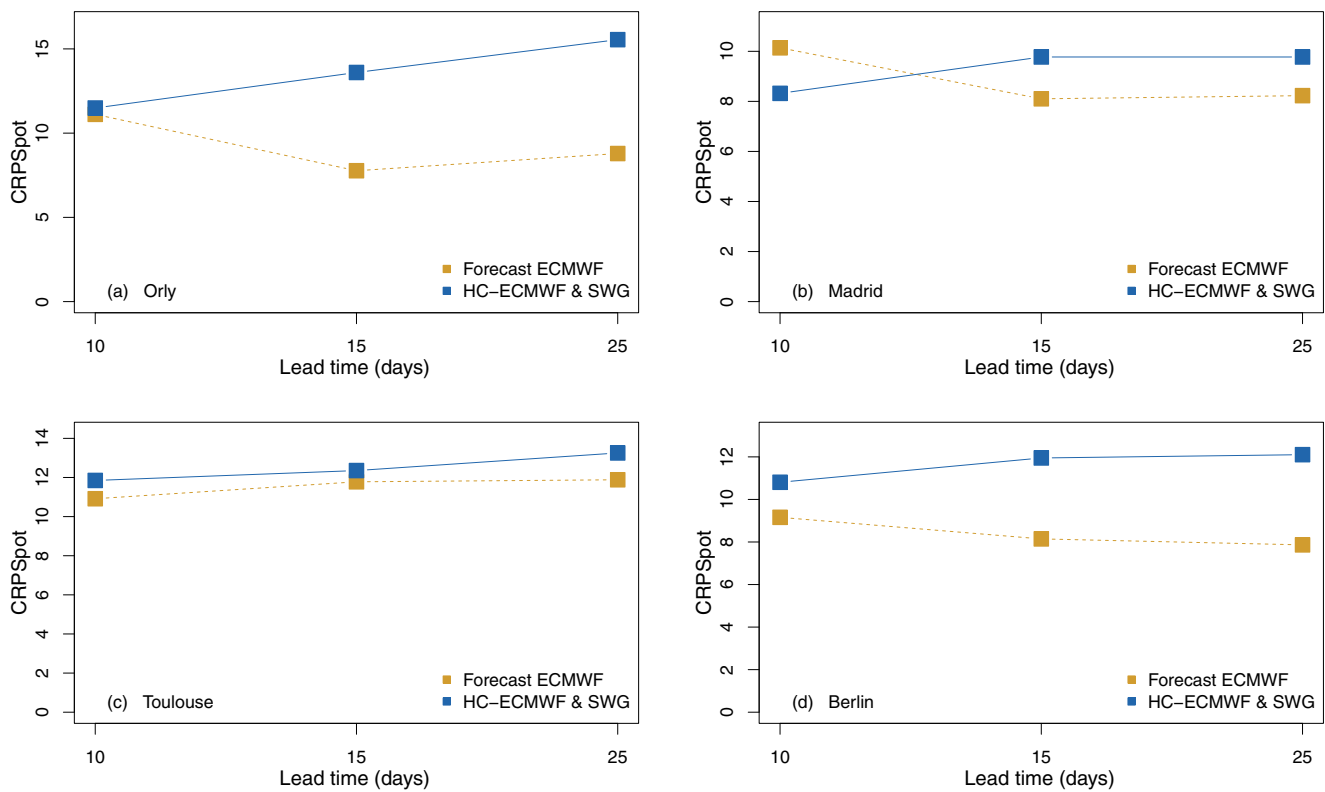
The CRPS can be decomposed into three parts, as explained in Section 3.3 and shown in Equation (4). Table C1 shows the resolution and the reliability values for the HC-SWG using the ECMWF analogues for different lead times. Indeed, the resolution tends to be more independent from the ensemble spread and explains

the predictability better (Hersbach, 2000). The resolution values of the HC-SWG for the different lead times are positive, indicating a better performance than the climatology. However, they increase with lead times, which could be related to the HC-SWG's average precipitation forecast over longer lead times. Figure C1 shows the potential CRPS (CRPS<sub>spot</sub>) (Hersbach, 2000), which we computed as follows:

$$\overline{CRPS_{spot}} = \overline{Unc} - \overline{Res}. \quad (C1)$$

**TABLE C1** Continuous ranked probability score of reliability (Re) and resolution (Res) for the hindcast–stochastic weather generator for different lead times.

Lead time (days)	Re, Res	Madrid	Orly	Berlin	Toulouse
$T = 10$	Re	1.45	2.19	2.22	2.10
	Res	4.60	3.93	6.30	9.51
$T = 15$	Re	2.18	2.91	3.12	2.96
	Res	4.93	5.83	3.37	5.27
$T = 25$	Re	3.16	3.8	4.32	4.35
	Res	3.01	4.16	2.95	6.60



**FIGURE C1** Comparison between the potential continuous ranked probability score (CRPSpot) of the hindcast (HC)–stochastic weather generator (SWG) forecasts based on European Centre of Medium-range Weather Forecasts (ECMWF) reforecasts and the ECMWF precipitation forecast. CRPSpot includes the resolution and the uncertainty of the continuous ranked probability score. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

We decided to look at the CRPSpot to better understand the forecasts' performance. Comparing the CRPSpot of the HC–SWG with the CRPSpot of the ECMWF (Figure C1), we noticed that the CRPSpot of the ECMWF decreases with lead time for the different locations studied. However, the CRPSpot of the HC–SWG tends to increase with lead

times. This can be explained in different ways. It can be related to the average forecast provided by the HC–SWG between  $t_0$  and  $[t_0, T]$ , which leads to filtering out more noise with longer lead times, to the variance in the data due to outliers, or to the relatively small size of the verification periods as indicated.