Seasonal Forecasts of the Saharan Heat Low characteristics: A multi-model assessment

Cedric G. Ngoungue Langue[1,2], Christophe Lavaysse[2,3], Mathieu Vrac[4], Philippe Peyrillé[5], and Cyrille Flamant[1]

[1]Laboratoire Atmosphères, Milieux, Observations Spatiales (LATMOS) - UMR 8190 CNRS/Sorbonne Université/UVSQ, 78280 Guyancourt, France.
[2]Université Grenoble Alpes, CNRS, IRD, G-INP, IGE,38000 Grenoble, France
[3]European Commission, Joint Research Centre (JRC), 21027 Ispra, VA, Italy
[4]Laboratoire des Sciences du Climat et de l'Environnement, CEA Saclay l'Orme des Merisiers, UMR 8212 CEA-CNRS-UVSQ, Université Paris-Saclay & IPSL, 91191 Gif-sur-Yvette, France.
[5]Centre National de Recherches Météorologiques (CNRM) - Université de Toulouse, Météo-France, CNRS, 31057 Toulouse Cedex 1, France.

Correspondence: Ngoungue Langue Cedric Gacial (cedric-gacial.ngoungue-langue@latmos.ipsl.fr)

Abstract.

The Saharan Heat Low (SHL) is a key component of the West African monsoon system at synoptic scale and a driver of summertime precipitation over the Sahel region. Therefore, accurate seasonal precipitation forecasts rely in part on a proper representation of the SHL characteristics in seasonal forecasts models. This is investigated using the last versions of two seasonal forecast systems namely the SEAS5 and MF7 systems respectively from the European Center of Medium range Weather Forecasts (ECMWF) and Meteo-France. The SHL characteristics in the seasonal forecast models is assessed based on a comparison with the fifth ECMWF ReAnalysis (ERA5) for the period 1993-2016. The analysis of the modes of variability shows that the seasonal forecast models have issues with the timing and the intensity of the SHL pulsations when compared to ERA5. SEAS5 and MF7 show a cool bias centered on the Sahara and a warm bias located in the eastern part of the Sahara, respectively. Both models tend to under-estimate the inter-annual variability of the SHL. Large discrepancies are found in the representation of extremes SHL events in the seasonal forecast models. These results are not linked to our choice of ERA5 as a reference, for we show robust coherence and high correlation between ERA5 and the Modern-Era Retrospective analysis for Research and Applications (MERRA). The use of statistical bias correction methods significantly reduces the bias in the seasonal forecast models, improves the yearly distribution of the SHL and the forecast scores. The results highlight the capacity of the models to represent the intraseasonal pulsations (the so-called east-west phases) of the SHL. We notice an overestimation of the occurrence of the SHL east phases in the models (SEAS5/MF7) while the SHL west phases are much better represented in MF7. In spite of an improvement of prediction score, the SHL-related forecast skills of the seasonal forecast models remain weak for specific variations for lead times beyond 1 month requiring some adaptations. Moreover, the models show predictive skills at intraseasonal time scale for shorter lead time.

Keywords : Saharan Heat Low, seasonal forecast, bias correction, wavelet analysis.

1 Introduction

In the Sahel region, food security for populations depends on rain-fed agriculture which is conditioned by seasonal rainfall (Durand, 1977; Bickle et al., 2020), characterized by a strong convective activity in the summer, associated with a large climatic variability (local- and large-scale forcings), generally leading to poor precipitation forecast skills at sub-seasonal and seasonal time scale in tropical north Africa (Vogel et al., 2018). Hence, climate models suffer from biases in the representation of West African Monsoon (WAM) processes and dynamics responsible for rainfall in West Africa (Roehrig et al., 2013; Martin et al., 2017). During the African Monsoon Multidisciplinary Analysis (AMMA) project (Redelsperger et al., 2006), the Saharan Heat Low (SHL) has been used as a key component to assess the variability of the WAM system. In particular, forecasters and researchers have pointed out the need to document the SHL predictability and its link with Sahelian rainfall (Janicot et al., 2008b). Improving precipitation forecasts not only is crucial for agriculture and water supply in the region, but is also of paramount importance for floods and diseases prevention.

The SHL refers to the low surface pressure area that appears above the Sahara region in the boreal summer due to seasonal high temperatures and insolation (e.g., Lavaysse et al., 2009). The SHL is an essential component of the WAM system at synoptic scale (Sultan and Janicot, 2003; Parker et al., 2005; Peyrillé and Lafore, 2007; Lavaysse et al., 2009; Chauvin et al., 2010) and a driver of precipitation over the Sahel region (Lavaysse et al., 2010a; Evan et al., 2015). It plays an important role in the atmospheric circulation over West Africa and brings moisture from the Atlantic Ocean to the region, thereby favoring the installation of the monsoon flow. In the lower atmospheric layers, the cyclonic circulation generated by a strong SHL tends to reinforce the monsoon flow around its eastern flank and the Harmattan flow along the western flank (Lavaysse, 2015). In the mid-layers, the anticyclonic circulation associated with the divergent flow at the top of the SHL contributes to maintain the African Easterly Jet (AEJ) around 700 hPa and modulates its intensity (Thorncroft and Blackburn, 1999). An intensification of the AEJ is observed during strong phases of SHL activity (Lavaysse et al., 2010b). According to Lavaysse et al. (2009), the SHL maximum activity over the Sahara occurs on average from the 20th of June to the 17th of September, and it is located between $7°W - 5°E$ and $20°N - 30°N$ covering much of northern Mauritania, Mali, Niger and southern Algeria [Fig.1]. The maximum of SHL activity happens during the rainfall season in the Sahel region (from June to September, Sultan and Janicot, 2003). The SHL is considered as a reliable proxy of the regional- and large-scale forcings impacting the WAM (Lavaysse et al., 2010b).

Lavaysse et al. (2009) monitored the seasonal evolution of the West African Heat Low (WAHL) using ERA-40 reanalyses and brightness temperature from the Cloud Archive User Service (CLAUS). They found a north-westward migration of the West African Heat Low (WAHL) from a position south of the Darfur mountains in the winter to a location over the Sahara between the Hoggar and the Atlas mountains during the summer. They also estimated the climatological onset of the SHL occuring around the 20th June (from the period 1984-2001) some days before the climatological monsoon onset date. This highlights strong links between the SHL and the monsoon flux. Chauvin et al. (2010) assessed the intraseasonal variability of the SHL and its link with midlatitudes using National Center for Environment Prediction (NCEP-2) reanalysis data. They found a robust mode of variability of the SHL over North Africa and the Mediterranean which can be decomposed into two phases

called : east-west oscillations. The west phase corresponds to a maximum temperature over the coast of Morocco–Mauritania, propagating southwestward, and a minimum temperature between Libya and Sicily, propagating southeastward. The east phase corresponds to the opposite temperature structure which propagates as in the west phase. Roehrig et al. (2011) studied the link between the variability of convection in the Sahel region and the variability of the SHL at intraseasonal time scale using NCEP-2 reanalysis data. They showed that the onset of the monsoon is associated with strong SHL activity when the northerlies coming from the Mediterranean (sometimes called ventilation) are weak. Conversely, they revealed that the formation of a strong cold air surge over Libya and Egypt and its propagation toward the Sahel lead to the decrease of the SHL, which inhibits the WAM onset.

As detailed above, previous work has evidenced the importance and the role of the SHL on the West African climate. These studies are based on a climatological view of the SHL using mostly reanalysis data. One may legitimately wonder how seasonal forecast models represent the SHL evolution?

The seasonal forecast is a long-term forecast which is very useful by allowing an anticipation of seasonal trends. The use of ensemble forecast for seasonal forecasting provides a range of forecasts and gives information about the spread associated with the forecast of a specific variable. Ensemble forecast models lead to an improvement of the predictive skills of some atmospheric variables (Haiden et al., 2015; Lavaysse et al., 2019). The evaluation of the SHL behaviour in seasonal forecast models has not been addressed yet. Roehrig et al. (2013) show that the mean temperature over the Sahara from July to September is well correlated with rainfall position over the Sahel region. Provided that the SHL characteristics (i.e. the east and west pulsations of the Heat low, its intensity and its interannual variability) are well captured in seasonal forecast models simulations, they can be used as predictors for rainfall in the Sahel area.

The goal of this article is: i) to investigate the representation and the forecast skills of the SHL in two seasonal forecast models and ii) to evaluate the added value value of bias correction techniques on raw seasonal forecasts. Bias issues are very frequent in seasonal forecast models; by correcting them with statistical methods, the predictive skills of the models can be improved in order to provide atmospheric variables that better fit the characteristics of the observation.

To reach this aim, we firstly study the SHL variability modes in seasonal forecast models and reanalyses; secondly we estimate the biases between the forecasts and reanalyses. Finally, we assess the recent evolution of the SHL and proceed with an evaluation of forecasts with respect to the reanalyses.

The remainder of this article is organized as follows: in section 2, we present our region of interest and the data used for this work; the description of the methodology adopted is also provided. Section 3 contains the main results of this investigation obtained by following the methodology described in section 2. In section 4, the predictive skills of the seasonal forecast models are discussed; and section 5 provides a conclusion with some perspectives for future studies.

2 Data and Methods

2.1 Saharan Heat Low evaluation metric

The location of the West African Heat low has a strong seasonal variation: North-South owing to the seasonal cycle of insolation and East-West owing to orographic forcing (Lavaysse et al., 2009; Drobinski et al., 2005). It is termed SHL once it reaches its Saharan location generally within 20°- 30°N × 7°W - 5°E during the monsoon season, an area that is bounded by the Atlas mountains to the North, the Hogar mountains to the East, the Atlantic ocean to the West and the northern extent of the WAM to the South (Evan et al., 2015). The SHL has been detected in previous studies using the low level atmospheric thickness (LLAT) computed as a geopotential distance between two pressure levels 700 hPa and 925 hPa (Lavaysse et al., 2009). Because the LLAT is due to a thermal dilatation of the low troposphere and in order to simplify the detection process, the SHL can be monitored by using the 850 hPa temperature field. Lavaysse et al. (2016) using ERA-Interim reanalysis, showed that the 850 hPa temperature (T850) field is well correlated to the LLAT and can be used as a proxy for the monitoring of the SHL (detection and intensity). As ERA5 is an improvement of ERA-Interim, we assume that the correlation between T850 and the LLAT is preserved in ERA5. We suppose this is also true for the forecast models. Consequently in this study, we use T850 to analyse the SHL characteristics. Because fixed boxes are used, the detection of the SHL is not needed, but, strong (weak) phases of the SHL will be associated with high (low) respectively T850.

2.2 Region of interest

The Sahara is located over the [25°W - 40°E] × [20°N - 35°N] and covers large parts of Algeria, Chad, Egypt, Libya, Mali, Mauritania, Morocco, Niger, Western Sahara, Sudan and Tunisia (see topographic map on [Fig1]). The climate is associated with very hot temperatures from May to September around 30°C for mean temperatures and over 40°C for mean maximum temperatures, very low humidity close to the surface (with relative humidities less than 10%) and a critical absence of rainfall . It is also the region with the largest production of dust particles (Prospero et al., 2002). For this study, the North Africa is subdivided in four regions (see [Fig2]) defined as follows:

- the Sahara area between 10°W - 20°E and 20°N - 30°N; it extends from the South of Morocco to Egypt;

- the central SHL here denoted as "$CSHL$", is located between 7°W - 5°E and 20°N - 30°N; it covers most of the North of Mauritania, Mali and the South of Algeria;

- the Western SHL here denoted as "$WSHL$", is located between 10°W - 2°W and 20°N - 30°N; it includes the North of Mauritania, Mali, the South of Morocco and Algeria;

- the Eastern SHL denoted as "$ESHL$", is located between 0°E - 8°E and 20°N - 30°N; it is mostly in the South of Algeria.

The choice of the 4 regions was supported by previous studies: Lavaysse et al. (2009) highlight a maximum activity of the SHL in the CSHL location during summer (JJAS period); Roehrig et al. (2011) show that the SHL tends to migrate from the

West to the East during the season which explain the WSHL and ESHL locations. The Saharan location has been used in some climate studies (Lavaysse, 2015; Taylor et al., 2017).

2.3 Data

In this study, we used two types of data: reanalyses and seasonal forecast model outputs. We used outputs from the 5th Generation European Center for Medium-range Weather Forecasts (ECMWF) Reanalysis (ERA5), (Hersbach et al., 2020). The ERA5 atmospheric variable studied here is daily T850 with a spatial resolution of $0.25°$x $0.25°$downloaded on the climate data store website: https://cds.climate.copernicus.eu/. The Modern-Era Retrospective analysis for Research and Applications (MERRA) dataset was also used. The MERRA data has a spatial resolution of $0.5°$x $0.625°$with 42 vertical levels downloaded on Climserv database. As for ERA5 reanalysis, we used the MERRA T850 to carry out our analyses. To be coherent with the model outputs, we consider only the daily temperature data at 00:00 and 12:00 UTC. We also transformed the spatial resolution of ERA5/MERRA (from $0.25°$x $0.25°$/ $0.5°$x $0.625°$, to $1°$x $1°$) to match the one of the seasonal forecast models. The two forecast models analysed here are the seasonal forecast SEAS5 from ECMWF, and the seasonal forecast system MF7 from Meteo-France. The seasonal forecast model SEAS5 replaces the previous seasonal system S4 (Johnson et al., 2019), it includes upgraded versions of the atmosphere and ocean models at higher resolutions. The SEAS5 model has a horizontal resolution of 36 km over the globe, and contains 91 levels for the vertical resolution. The MF7 seasonal forecast system is based on ARPEGE-IFS global forecast model (Déqué et al., 1994) which was jointly developed by Meteo-France and ECMWF. MF7 uses the climate version of CNRM-CM6 (Voldoire et al., 2019) such that MF7 and SEAS5 only share a common radiation parameterization but the rest of the physical package is different. The horizontal resolution of the MF7 model is around 7.5 km over France and 37 km over the antipodes; it contains 105 vertical levels. Both SEAS5 and MF7 model outputs used in this paper are based on the ensemble retrospective forecast (hindcast) which contains 25 members, meaning that for a given time, we have 25 re-forecasts from each model. The re-forecasts are released on the first day of every month for a period of 6 months for SEAS5. With MF7, one member of the model is initialized on the 1st of the month, and the other members are launched on the 2 last Thursdays of the month. The atmospheric variable investigated in models is also daily temperature at 00:00 and 12:00 UTC with a spatial resolution of $1°$x $1°$. Our dataset covers the period going from the 1st of January 1993 to the 31st of December 2016.

2.4 Strategy for the analysis of forecast

As we analyse the representation of the SHL, we focus on the period going from June to September (denoted by JJAS in the rest of the study) because it corresponds approximately to the period of maximum heat low activity over the Sahara (Lavaysse et al., 2009). Seasonal forecast models usually fail to forecast correctly events a long time in advance for a given target period. Therefore, we are interested in a forecast launched at most two months in advance of the JJAS period. In order to do that, we consider re-forecasts initialised on the 1st of April, May and June, which corresponds respectively to a June lead time of 2, 1, 0 month(s). This technique allows us to quantify the sensitivity of the models in representing the SHL at different lead times.

The re-forecast validation process is made separately for the whole JJAS period and individual months (June, July, August and September) because June and September temperature values are in the same range.

2.5 Methods

This section describes in more details the set of analyses carried out to achieve our goal. The methodology adopted is illustrated below.

2.5.1 Subseasonal modes of variability

A mode of variability represents a spatio-temporal structure highlighting the main characteristics of the evolution of atmospheric variables at a given time scale. There are several statistical methods for assessing the modes of variability that contribute to a raw signal. The one used here is the wavelet analysis of the temperature signal. The wavelet transform consists in applying a time-frequency analysis on a given signal. It is very useful to analyze non-stationary signals in which phenomena occur at different scales. This method provides more information than the Fourier transform about the observed structures in the initial signal (starting and ending time, and the duration of propagation (frequency)). With this type of analysis, we observe the distribution of the signal intensity in time and frequency. A wavelet function is defined by a scale factor and a position factor (Büssow, 2007; Zhao et al., 2004).

Let $\mathbf{f(t)}$ be a real function of real variable, the wavelet transformation of this function denoted as $\mathbf{W(f)(a,b)}$ is given by:

$$\mathbf{W(f)(a,b)} = <\mathbf{f}, \psi_{\mathbf{a,b}}> = \int_{-\infty}^{+\infty} (\mathbf{f(t)} * \psi_{\mathbf{a,b}}\mathbf{(t)})\mathbf{dt} \tag{1}$$

$$\psi_{\mathbf{a,b}}\mathbf{(t)} = \frac{1}{\sqrt{\mathbf{a}}} * \mathbf{\Psi}\left(\frac{\mathbf{t-b}}{\mathbf{a}}\right) \tag{2}$$

The function Ψ is called mother wavelet and must be of square integrable that means $\int_{-\infty}^{+\infty} (\Psi(t))^2 dt$ is finite, and also verify the following property: $\int_{-\infty}^{+\infty} \Psi(t)dt = 0$. The parameter b is the position factor and a is the scaling parameter greater than zero. For a given signal, a represents the frequency and b the time. There exist diverse types of mother wavelets; based on the literature review and its common use, we chose the Morlet wavelet (Tang et al., 2010). The Morlet wave is defined as the product of a complex sine wave and a gaussian window (see "Eq. (3)") (Cohen, 2018). The wavelet analysis has been applied separately on the re-forecasts and the reanalyses for an initialisation of the seasonal forecast models on the 1st of April, May and June for a 6 months period; but we extracted only the signal on the JJAS period to conduct our analyses on variability modes. We focused on signals with a period up to 32 days.

$$\mathbf{\Psi(t)} = \pi^{-1/4}\mathbf{exp}^{-\mathbf{t^2}/\mathbf{2}}\mathbf{cos(w_o t)} \tag{3}$$

(Tang et al., 2010)

2.5.2 Bias Correction

Seasonal forecast models provide a numerical representation of the earth and the interactions between its different components: the atmosphere, the ocean and the continental surfaces. Those interactions are very complex and take place at different spatio-temporal scales. This can lead in certain cases to an over/under-estimation of the evolution of atmospheric variables in the models. The cause of this behavior in the models is often the presence of biases. To overcome this bias issue, we use here two univariate bias correction methods: "Quantile Mapping" (QMAP) and "Cumulative Distribution Function-transform" (CDF-t).

– QMAP

Quantile-mapping aims to adjust climate model simulations with respect to reference data, in determining a transfer function to match the statistical distribution of simulated data to the one of reference values (e.g., Dosio and Paruolo, 2011). When reference data have a resolution similar to climate model simulations, this technique can be considered as a bias adjustment method. On the other hand, when the observations are of higher spatial resolution than climate simulations, quantile-mapping attempts to fill the scale shift and is then considered as a downscaling method (Michelangeli et al., 2009). The QMAP method is based on the assumption that the transfer function calibrated over the past period remains valid in the future. Let $F_{o,h}$ and $F_{m,h}$ be respectively the cumulative distribution functions (CDF) of the observational (reference) data $X_{o,h}$, and modeled data $X_{m,h}$, in a historical period h. The transfer function for bias correction of $X_{m,p}(t)$ which represents a modeled value at time t within a projected period p is given by the following relation (e.g., Cannon et al., 2015; Dosio and Paruolo, 2011):

$$\hat{\mathbf{X}}_{\mathbf{m,p}}(\mathbf{t}) = \mathbf{F}_{\mathbf{o,h}}^{-1}\{\mathbf{F}_{\mathbf{m,h}}[\mathbf{X}_{\mathbf{m,p}}(\mathbf{t})]\} \tag{4}$$

where $F_{o,h}^{-1}$ is the inverse function of the CDF $F_{o,h}$.

– CDF-t

CDF-t is a statistical downscaling method developed by Michelangeli et al. (2009). It can be considered as a generalization of the quantile-mapping correction method. Hence, as QMAP, CDF-t consists in finding a relationship between the CDF of a large-scale climate variable and the CDF of this same variable at the local scale. However, while quantile-mapping method projects the simulated values at large scale on the historical CDF to calculate quantiles, CDF-t takes explicitly into account the change in large-scale CDF between the historical period and the future period. In the CDF-t approach, a mathematical transformation T is applied to the large-scale CDF to define a new CDF as close as possible to the CDF obtained from the station data (e.g., Vrac et al., 2012; Lavaysse et al., 2012).

Let $F_{m,h}$ and $F_{o,h}$ be the CDFs at large and local -scale respectively of the modeled data $X_{m,h}$ and the observational data $X_{o,h}$ over a historical period h, and T the transformation allowing to go from $F_{m,h}$ to $F_{o,h}$. We have the following relation (Vrac et al., 2012):

$$\mathbf{T(F_{m,h}(X_{m,h})) = F_{o,h}(X_{o,h})}. \tag{5}$$

By applying this relation to the CDF $F_{m,f}$ of the modeled data $X_{m,f}$ in a future period f, it provides an estimation of the local CDF $F_{o,f}$ in the future period f:

$$\mathbf{\hat{F}_{o,f} = T(F_{m,f}(X_{m,f}))} \tag{6}$$

A quantile mapping can then be performed between $F_{m,f}$ and $\hat{F}_{o,f}$ to obtain bias corrected values of future simulations. More details about CDF-t can be found in (Vrac et al., 2012). All the computations for the CDF-t method were done with the R package "CDFt".

After applying the bias correction methods on the model outputs, the added value of the bias correction compared to the raw re-forecasts will be assessed by the computation of the Cramer-von Mises (hereafter Cramer) score (Henze and Meintanis, 2005; Michelangeli et al., 2009). The Cramer score measures the similarity between two distribution functions; the closer its value is from 0 the closer are the distributions.

– Application of bias correction

QMAP and CDF-t are usually used for downscaling tasks in a climate projection context. In this study, we adapted the application of these methods for bias correction in a seasonal forecast context. We used a leave-one out approach for the calibration process with CDF-t and QMAP. This method consists in removing the target year (the year we want to apply the correction on) in the historical period before the estimation of the transfer function which allows to pass from the global scale to local scale data. In our case the calibration process has been made using 23 of the 24 years in the historical period 1993-2016 for every year. The correction or projection process is made differently using CDF-t and QMAP. For QMAP, we use as input the target year removed previously during the calibration phase. With CDF-t, we built a new dataset of 24 years which is the concatenation of the dataset used for the calibration and the target year, so that the year at the end of the new dataset represents the target year.

2.5.3 Ensemble forecast verification

Ensemble forecast verification is the process of assessing the quality of a forecast. The forecast is compared against a corresponding observation or a reference, the verification can be qualitative or quantitative. Forecast verification is important to monitor forecast quality, improve forecast quality and compare the quality of different forecast systems. There are many metrics or probability scores developed for ensemble forecast verification depending on the tasks performed. In our preliminary studies (not shown) on the skills of the forecast models, we used different scores (CRPS, Brier Score, Roc Area Curve, Rank Histogram, Reliability diagram) but in the present work, we will only focus on the Continuous Ranked Probability Score (CRPS) (Hudson and Ebert, 2017), which is very similar to the Brier Score. This choice is justified by the simplicity in data

processing when computing the CRPS through some R packages like SpecVerification (Siegert et al., 2017). The CPRS is a quadratic measure of the difference between the forecast CDF and observation CDF. It quantifies the relative error between the model forecasts and the observations; it is a measure of the precision of an ensemble forecast model. The closer the CRPS is from 0, the better is it.

Let $P_F(x)$ and $P_O(x)$ be the cumulative distribution functions respectively for the forecasts and observations, the CRPS is computed as follows:

$$\mathbf{CRPS} = \int_{-\infty}^{+\infty} (\mathbf{P_F(x)} - \mathbf{P_O(x)})^2 \mathbf{dx} \tag{7}$$

The Root Mean Square Error (RMSE) which is a measure of the differences between two samples (model predictions and observations), has also been for the evaluation of the forecasts. Let y_t be the forecast of the model at time t and O_t the corresponding observation at the same time, the RMSE is given by the following relation:

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{t=1}^{N} (\mathbf{y_t} - \mathbf{O_t})^2}{N}} \tag{8}$$

where N is the number of time step.

3 Results

3.1 Climatology of the SHL

The climatology state of the SHL has been assessed from 1993 to 2016 during the JJAS period for ERA5, MERRA, SEAS5 and MF7 (see [FigS2] in supplementary material). The seasonal forecast models tend to develop SHL's climatologies with very similar characteristics as the reanalyses. Strong SHL intensities are located over the CSHL location [FigS2] for all the products (reanalyses and forecast models); this is in agreement with Lavaysse et al. (2009). Another point discussed in this section, is the uncertainty between the reanalyses (ERA5 and MERRA) (see [Fig2]). ERA5 and MERRA exhibit similar behaviours regarding the climatological bias of the SHL with respect to the seasonal forecast models [Fig2-(c-f)]: a cold bias with SEAS5 and a warm bias with MF7. The seasonal evolution of the climatological state of the SHL in ERA5 and MERRA (see [Fig6]), is almost similar over the CSHL location except for the Sahara where a little shift of MERRA to high temperatures is observed but the patterns in the evolution remain very close to ERA5. The distribution of yearly T850 over the JJAs period (see [Fig7-a),b),g),h)]) in ERA5 and MERRA is quite similar, suggesting a good correlation between the two reanalyses. The uncertainties between the reanalyses (see [FigS3] in supplementary material) are very smaller than the biases in the seasonal forecast models with respect to ERA5. We also found large correlation between ERA5 and MERRA (see [FigS1] in supplementary material) around 0.97/0.92 respectively over CSHL/Sahara location during the JJAS period. By considering all these results highlighting high similarities between ERA5 and MERRA, we decided to choose ERA5 as our reference dataset for the rest of the study.

3.2 Variability modes

Through a wavelet transformation, we compared the variability modes in the forecast products (SEAS5, MF7) with respect to ERA5 over central SHL location and Sahara (boxes indicated in Fig2) (see [FigS4] in supplemental material). Especially for the year 2016, three main frequency bands of activity of the SHL have been identified in ERA5. Firstly, the SHL activity within the 4-8-day window with high intensity; secondly an intensification of events with strong intensity (spectral power > 16) is observed for periods of about 8-16 days. Finally, events at very high frequencies are observed and the intensity associated is very higher (spectral power > 64) than the previous ones. This shows the SHL activity becomes stronger at high frequencies. The models tend to reproduce quite differently the pulsations observed in the reanalysis signals; there is an issue regarding the temporality, frequency and intensity of the pulsations in the forecast models.

To assess the climatology of the variability modes [Fig3], we analysed the distribution of days associated with spectral power greater than 1 (normalized value) here defined as significant days during the period 1993 to 2016.This threshold of 1 has been selected arbitrarily after applying a sensitivity test on several threshold values from 0.5 to 10 to focus on predominant events at different periods. We noticed globally a decrease of events occurrence with high threshold values of the spectral power. Note that the sensitivity to the threshold values does not significantly impact our results (not shown). We observe a similar behaviour in ERA5, SEAS5 and MF7 in terms of significant days with an increasing number of days with periods up to 10 days followed by a quite steady activity for longer periods. Over the Sahara area, there is a tendency of both models to reproduce a SHL activity similar to ERA5 at too short period (~ 10 days). ERA5 shows little variations in the number of significant days with periods between $12-26$ days and tends to be constant for high periods (greater than 27 days). SEAS5 overestimates the SHL activity around the 15-day period while MF7 is shifted toward higher frequency and underestimates the longer period. Over the central SHL box, there is a tendency of both MF7/SEAS5 models to generate a significant SHL activity at too short period ($\sim 4/10$ days) compared to ERA5. At longer timescales MF7 tends to overestimate the SHL activity within the 10-23 day period while SEAS5 shows an under-estimation of the SHL activity within the same window. The evolution of significant days over central SHL location and Sahara highlighted three main pulsations based on the period (or frequency). The different pulsations identified are arbitrarily classified as follows: the class $C1 = [0-10 days]$ for low frequency, the class $C2 =]10-22 days]$ for high frequency and the class $C3 =]22-32 days]$ for very high frequency pulsations. In the following, we investigate the inter-annual variability of significant days on those different classes of pulsation [Fig4]. The result for ERA5 shows a high inter-annual variability for pulsations in class C1 both over central SHL box and Sahara [Fig4-a),d)]. This can be caused by the triggering of easterly waves and Kelvin equatorial waves which tend to reinforce the convection activity. Those two types of waves have a periodicity between 1-6 days (Janicot et al., 2008a). The correlation between the seasonal forecast models and ERA5 is very low, less than 0.4. From this analysis, we can see that the seasonal forecast models tend to represent the climatological activity of the SHL at different frequencies even if some discrepancies are observed. However, the representation of the inter-annual variability of the SHL activity remains a big challenge for the seasonal forecast models.

3.3 Seasonal cycle

In this section, we are assessing the spatial representation of the SHL over the Sahara region. In order to do that, we evaluate the bias between the seasonal forecast models (SEAS5, MF7) and ERA5 [Fig5]. The bias is defined here as the difference between the forecasts and the reanalyses; the mathematical expression of the bias is the following:

$$\mathbf{B_t} = \mathbf{F_t} - \mathbf{R_t} \tag{9}$$

where F_t and R_t are respectively the forecasts and reanalyses at time t.

The bias is computed for each month at lead time 0 during the season from January to December for the period 1993-2016. By extending the analysis window over the season, we are able to check if the biases in the seasonal forecast models are constant or specific to the JJAS period. When analyzing the SEAS5 model outputs [Fig5-a)], we notice an overestimation of temperature over the Atlantic Ocean and over the Mediterranean sea. We observe a cold bias between SEAS5 and ERA5 which appears progressively during the first months (January to April) and tends to intensify during the monsoon phase over the Sahel region. This cold bias is centered over the Sahara between the North of Mali, Niger and the South of Algeria; and tends to decrease in intensity during the retreat phase of the monsoon in October. SEAS5 is colder than ERA5 and under-estimates the spatial evolution of the SHL over the Sahara. In fact, biases in the evolution of the coupled ocean-atmosphere system or in the continental sub-surface can play a role in theses biases but their investigation is beyond the scope of this paper. The analysis on MF7 shows a progressive appearance of a warm bias in comparison to ERA5 over the Sahel during January and February [Fig5-b)]. This warm bias tends to develop from March to September and affects the whole Sahara. It is more intense during the monsoon phase and is located over the eastern part of the Sahara. The bias between MF7 and ERA5 tends to decrease in intensity during the retreat phase of the monsoon in October. MF7 is warmer than ERA5 and overestimates the spatial evolution of the SHL over the Sahara. The central SHL area is less affected by this warming in MF7 compared to the rest of the Sahara. This behaviour in the Sahara region, especially in the eastern part of the Sahara, could be related to an under-estimation of air advection coming from the Mediterranean regarding the prevalence of the hot bias to the eastern part of the Sahara [Fig2-b)]. This analysis shows that the two seasonal forecast models have two contrasted representations of the SHL compared to ERA5 with a colder SHL in SEAS5 and a warmer SHL in MF7; which is in agreement with previous global studies (e.g., Dixon et al., 2017; Johnson et al., 2019). The two seasonal forecast models share however a similar seasonal evolution of the bias (increasing bias during the monsoon season) and a large spatial scale of the bias that covers most of the Sahara. Without sensitivity experiments, it is impossible to clearly identify the reasons for these opposite behaviors between the two models. The investigation of the origins of these biases are way beyond the scope of this article. In a more general framework, various european research projects have shown the difficulty of attributing a specific bias to a specific parameterization over West Africa, including the SHL (Martin et al., 2017).

After the evaluation of the spatial evolution of the SHL in the seasonal forecast models, the representation of the temporal drift is assessed [Fig6]. The method used here consists in computing the climatology of daily T850 ensemble mean and

325 ensemble spread for the two models (SEAS5 and MF7); and the daily climatology of T850 for ERA5 from 1993-2016. For the models, we consider only the re-forecasts launched respectively on the 1st of April, May and June for a period of 6 months (see section 2.4 for more details). We can see that the climatology of ERA5 remains contained in the spread described by SEAS5 for all lead times over central SHL location and Sahara; this spread in SEAS5 seems to be constant in time and does not increase with the lead time. We observe for the first forecast days, a large spread with MF7 which is not present in SEAS5,

330 likely associated with different perturbations and initialization techniques that are beyond the scope of this study. For all lead times, an overestimation of temperature is shown with MF7 over the Sahara around mid-June and later over the central SHL location ($\sim$ 10 days after the 1st of July). SEAS5 shows an underestimation of temperature occurring on the 1st of July both over central SHL box and Sahara at different lead times. The maximum intensities of the SHL activity in the two seasonal forecast models are reached during the period of strong activity of the monsoon flux in the Sahel region (July - August). Both

335 models are very consistent at the beginning of the season (April-June) when the Sahara is gradually warming. In the extension of the previous analyses, we decided to check the temporal correlation of the models and ERA5 [FigS6]. We observe a weak correlation between the evolution of the SHL in the seasonal forecast models and ERA5. The scatter plot analysis used for this evaluation, highlights the over/under-estimation of T850 in MF7/SEAS5 with respect to ERA5 as observed in the monthly bias analyses [FigS6-c)].

340 An estimation of bias was carried out for the SEAS5 model data for i) the full available period, running from 1981 to 2016 (denoted $SEAS5_1$), ii) the period common to MF7 and SEAS5, 1993-2016 (denoted $SEAS5_2$). The bias evolution is quite similar over the two periods (see [FigS5]/[Fig5-a)]); but we notice in $SEAS5_1$ a smaller cold bias compared to $SEAS5_2$. This change in bias intensity can be explained by a warming in SEAS5 re-forecasts during the period 1981-1992 which attenuates the cooling effect in the model during the period 1993-2016.

345 ### 3.4 Interannual distribution of the T850

The climatological trend of the distribution of SHL intensities has been analyzed using the seasonal probability distribution function (pdf) [Fig7-a), f)] of the SHL box-averaged T850 (used as the proxy of the SHL intensity) over the JJAS period at June lead time 0 (i.e. the initialization of the model was made on the 1st June). The analysis of seasonal T850 shows a high variability in ERA5 and the presence of a decadal warming trend during the 2005^s both over central SHL location and Sahara

350 [Fig7-a), g)]. The high inter-annual variability of the SHL seen in ERA5 is under-estimated by SEAS5 and MF7. Using raw outputs of the seasonal forecast models, SEAS5 tends to represent much better than MF7 the distribution the SHL intensities over the Sahara [Fig7- a), c), d) and g), i), j)]. Another specificity of MF7 is its slightly larger ensemble spread. SEAS5 seems to underestimate the warming trends present in ERA5 during the 2005^s, an overestimation of this trend is observed with MF7 [FigS10]; this behavior in the seasonal forecast models is present both over central SHL box and Sahara. By using this type of

355 visualization (Heatmap which is a graphical representation of data where values are depicted by color), it is possible to assess the intensity of the climatological trend with respect to the intra-seasonal variability [Fig8]. The inter-annual variability of the SHL anomalies distribution in the seasonal forecast models is too far from ERA5 but some characteristics are captured by the

models (e.g. the increase of the frequency of anomalies in ERA5 during the 2000^s). We observed high frequencies in the SHL anomalies distribution at inter-annual time scale for MF7 and SEAS5 with more intense values over the Sahara [Fig8-b), c),

360 g), h)]. To focus more on the evolution of the tails of the distribution (i.e. the warmest and coldest T850), the anomaly of the pdf of temperature is provided in supplementary materials [FigS7]. An increase of the occurrence of the warmest temperature is observed in SEAS5 and MF7 during the 2010^s. MF7 tends to overestimate the inter-annual variability of the coldest and warmest temperature distribution, and SEAS5 exhibits an overall trend with some features close to ERA5. Despite the fact that seasonal models tend to capture some characteristics of the SHL variability, large differences are observed in comparison with

365 ERA5. These differences can be explained by systematic biases present in models, as well as approximations made during the models implementation (initial and boundary conditions, physical hypotheses, etc...). In order to improve the quality of the forecast, bias correction methods have been applied.

The above analyses revealed the presence of biases in the models, bias correction was applied over the JJAS period for June lead time 2, 1 and 0 (which represent the forecast of the JJAS period initialised respectively in April, May and June). The

370 bias correction techniques used are CDF-t and QMAP (see Section 2.5.2 for more details on their application). The analysis of ensemble forecast models remains very delicate because of the many possible ways to approach the bias correction, i.e. should we use the unperturbed member, mean ensemble member, median ensemble member or the whole ensemble member? In our case, the bias correction is first applied separately on the ensemble members, in order to correct the re-forecasts of each one of the 25 members of the seasonal forecast models (SEAS5 and MF7). A second methodology has been tested by applying a bias

375 correction on the ensemble mean. To evaluate the sensitivity of the Cramer score on the ensemble forecast models, we defined three different approaches as follows:

  – "CORR_NO_MEAN", in this approach the bias correction is applied on the whole ensemble member and the Cramer score is computed using the outputs of the correction;

  – "CORR_MEAN", here we compute first the mean over the outputs of the bias correction on the whole ensemble member;
380 and we use this mean to compute the Cramer score;

  – "MEAN_CORR", the method consists of applying the bias correction on the ensemble mean and the computation of the Cramer is done directly using the outputs of the correction.

The Cramer score was calculated firstly using ERA5 and the raw forecast samples (SEAS5 and MF7), and secondly between ERA5 and the bias corrected forecast samples [Fig9]. We can observe that raw forecasts are not improved with initialisation

385 months (April, May or June) while corrected forecasts show an improvement with decreasing lead times. This can be the result of systematic bias in seasonal forecast models. The "MEAN_CORR" method [Fig9-c), f), i)] is more efficient than the two other approaches "CORR_MEAN" and "CORR_NO_MEAN" based on the Cramer score values. The "CORR_MEAN" approach tends to smooth the corrected forecasts due to the computation of the mean ensemble member after applying the correction. CDF-t and QMAP methods produce very similar results; an illustration of the corrected forecasts using the both

390 methods is provided in supplementary materials (see [FigS9]). MF7 raw forecasts show relatively large correction over the

Sahara [FigS8]; this behavior in MF7 is related to the hot bias occurring over the Eastern part of Sahara during JJAS period as mentioned in Section 3.2 [Fig5-b)]. We can see from these results that bias corrections are efficient and so important to apply to the model outputs. Some illustrations of the corrected forecasts have been made with the CDF-t method. In [Fig7-e), j)], we can notice a significant improvement in the distribution of SHL in MF7 both over central SHL location and Sahara. This improvement is also effective for SEAS5 [Fig7-d), i)]. The corrected forecast distributions are closer to ERA5 than the raw forecasts ones. The investigation of the correlation between the corrected forecasts and ERA5 [FigS6-b), d)], shows clearly that CDF-t corrects the cold/hot bias in SEAS5/MF7 by increasing/decreasing T850 values in order to match with ERA5 T850 values. CDF-t reduces a large part of biases in SEAS5 and MF7 but the interannual correlation with ERA5 is not improved. Indeed, CDF-t is a quantile-based univariate bias adjustment method. As such, it preserves the ranks of the model simulations, and thus preserves as well their rank (Spearman) correlations (e.g., Vrac, 2018; François et al., 2020).

3.5 Evolution of the extreme SHL events

Strong SHL activity contributes to the reinforcement of the monsoon flow over the Sahel along the eastern flank of the SHL. It also modulates the intensity of the AEJ and generates wind shear over the region. The resulting wind shear will generate more instabilities favouring convective activities over the West Africa region. Taylor et al. (2017) showed that strong SHL activity intensifies the convection within the Meso-scale Convective Systems (MCSs). Fitzpatrick et al. (2020) suggest that stronger wind shear may be a key driver of decadal changes in storm intensity in the Sahel. This shows the importance of having a good representation of these SHL characteristics in the models. Therefore, we analysed the variability of the SHL extremes using the raw and corrected forecasts obtained with CDF-t respectively [Fig10]. We distinguished cold and hot extremes which represent events respectively under the quantile 10% and above the quantile 90%. We observe an increase of the SHL hot extremes in the seasonal forecast models during the 2010^s as well as a diminution of the SHL cold extremes which is in agreement with the evolution in ERA5. MF7 raw forecasts tend to overestimate the SHL hot extremes, while they seem to underestimate the SHL cold extremes both over central SHL location and Sahara. SEAS5 raw forecasts underestimate the SHL hot extremes and make an overestimation of the SHL cold extremes over the Sahara. We can see the efficiency of the bias correction (CDF-t) when analyzing the evolution of the SHL extremes from the corrected forecasts. Despite the difference between ERA5 and the corrected forecasts has been reduced compared to the raw forecasts, the observed gap remains significant.

3.6 East-West pulsation modes

The SHL has a typical time scale of 15 days associated with low-level horizontal advections of moist and cold air that modulate the surface temperature on the eastern part of Sahara, and make the maximum surface temperature shift from a more eastern to a more western location of the Sahara (Chou et al., 2001; Roehrig et al., 2011), leading to so-called heat low East (HLE) events and heat low West (HLW) events, respectively (Chauvin et al., 2010). Roehrig et al. (2011) and Lavaysse et al. (2011) highlighted interactions between SHL components and Sahelian rainfall events. In the present work, a simple method is proposed to capture the HLW and HLE oscillations. Our method consists in defining a dipole by computing the mean T850 difference between HLW and HLE boxes here refer to WSHL and ESHL respectively (see section 2.2 for more details):

$$\text{Dipole} = \text{HLW} - \text{HLE} \qquad\qquad (10)$$

425   A positive value of the dipole indicates a HLW occurrence while a negative value corresponds to the HLE event. We evaluate
the method using the LLAT approach and the automatic detection of the SHL barycenter (Lavaysse et al., 2009) used during
the H2020 Dynamics-Aerosol-Chemistry-Cloud Interactions in West Africa (DACCIWA) project campaign (Knippertz et al.,
2017), which aims to evaluate the seasonal location of the SHL with respect to its climatological position. An illustration of our
method for the year 2005 is shown in [FigS11]; and confirms that there is a good agreement between the evolution of the dipole
430   of T850 in ERA5 and the SHL barycenter computed in Knippertz et al. (2017). After the assessment of the detection method,
we evaluate the representation of the SHL components in the seasonal forecasts and ERA5 data [Fig11-a]. As the corrected
forecasts are unbiased compared to the raw forecasts, we use them for this analysis. The results show that ERA5 presents a
bimodal regime; the first one is less accentuated and associated with the HLE events (negative dipole value) while the second
regime is more representative and related to HLW events (positive dipole value). For MF7, we also noticed a bimodal regime
435   and a large range in the distribution of the dipole compared to ERA5 ; the first regime is more frequent and associated with the
HLE events. The second regime is less frequent and related to HLW events. With SEAS5, we also observed a bimodal regime
and a reduced range compared to ERA5. The first regime is less important and associated with the HLE events, while the second
one is more important and related to HLW events. From this analysis, we can notice that MF7(SEAS5) tends to overestimate
the HLE(HLW respectively) phases [Fig11-a]. This behavior in the seasonal models is well highlighted when using the raw
440   forecasts for the computation of the dipole (see [FigS16-a]) in supplementary material). MF7 and SEAS5 here again exhibit
opposite behaviours in terms of frequencies and intensities of the SHL components. The analysis of the correlation between
the models and ERA5 shows that MF7 seems to be slightly better correlated with ERA5 than SEAS5 see [Table1]. We noticed
a little and not significant improvement of the correlation with the corrected signal.

Table 1. Correlation between the dipole values derived from the seasonal models (SEAS5/MF7) and the one derived from ERA5. The "Raw dipole" represents the dipole computed using raw forecasts, and "Corrected dipole" using the bias corrected forecasts).

	SEAS5	MF7
Raw dipole	**0.53**	**0.65**
Corrected dipole	**0.61**	**0.75**

To better understand the reasons for these differences, a separate analysis of the SHL distribution in the two boxes is done
445   [Fig11-b), c)]. First, it is worth noting that the East Sahara is climatologically hotter than the West Sahara in ERA5, SEAS5
and MF7. This is explained by the proximity of West Sahara to the Atlantic ocean and the advection of fresh air masses in that
area (see [Fig2] for the location of the West Sahara). Hence, there is a greater occurrence of the HLE events compared to the
HLW phases in ERA5. The models (SEAS5/MF7) are able to reproduce this partitioning of the SHL phases observed in ERA5
with the same range of frequencies (0.6/0.4) for HLE/HLW respectively. Both models overestimate the occurrence of the HLE
450   events; MF7 tends to develop hot HLE events [Fig11-b)]. The analysis of the HLW phases reveals an under/overestimation of

the intensity/occurrence of these events in SEAS5. We notice, with MF7, a good representation of the intensity of the HLW events with sometimes an overestimation of the frequencies associated with these phases [Fig11-c]. The interactions between the East and West boxes are investigated through a correlation analysis using the outputs of the both seasonal models. The results obtained using raw/corrected forecasts are very similar (not shown), so in the following we present only the results related to the bias corrected forecasts. High correlation would suggest an influence of the large scale processes, whereas low correlation would indicate that smaller scale processes and local impacts to come into play. The correlation between HLE and HLW phases is about 0.45, 0.56 and 0.48 for ERA5, SEAS5 and MF7, respectively (see [Table2]). SEAS5 shows a higher correlation between the two phases compared to ERA5 and MF7. To discriminate the effect of the intra-seasonal and seasonal cycles, the correlations are computed between HLW and HLE by using the daily T850 anomalies relative to daily climatology of T850 in the two boxes. As expected, the seasonal cycle has a strong impact and the correlation reduces from 0.45 to 0.30 for ERA5 compared to a reduction from 0.56 to 0.45 for SEAS5; and from 0.48 to 0.34 for MF7 (see [Table2]). The correlation with SEAS5 remains high compared to MF7 and ERA5; this suggests that T850 over the Sahara in SEAS5 is more affected by large-scale drivers that provide wider temperature field anomalies. MF7 shows a partitioning of SHL variability between the SHL intraseasonal mode and the seasonal cycle that is more in agreement with ERA5. An investigation of the representation of the SHL phases in the models vs ERA5 has been assessed using the corrected forecasts. The models are more correlated with ERA5 for the HLW phases (see [TableS1]). Despite the differences between seasonal models and ERA5, the models are able to capture the seasonal East-West migration of the SHL.

Table 2. Correlation between the HLW and HLE phases : values in bold (brackets) indicate the correlation using the bias corrected temperatures (anomalies of temperature respectively) over the East and West SHL boxes.

	ERA5	SEAS5	MF7
Correlation	**0.45**(0.30)	**0.56**(0.45)	**0.48**(0.34)

4 Discussion

Our results show that the two seasonal forecast models have the capability to capture some of the SHL main characteristics. There is a deficit in reproducing the intensity and the occurrence of events, such as east and west phases of the SHL, as they appear in the reference ERA5. The analysis of the bias in seasonal forecast models evidenced a hot bias in MF7 and a cold bias in SEAS5 which could be explained by large scale processes and forcings occurring at different time scales in the Sahara region. The different behaviors observed in forecast models can be related to their sensitivity to the drivers and the physical processes involved in the SHL evolution.

A preliminary assessment of the quality of the forecasts with respect to the ERA5 reanalysis is discussed. This is done through an evaluation of the skills of the seasonal forecast models to reproduce the interannual variability of the SHL. For this evaluation, we used two metrics : the CRPS score and RMSE (see section 2.5.3 for more details). A first analysis of the forecasts at interannnual time scale was conducted using monthly mean T850 and daily T850, respectively, for the computation

of the CRPS score [Fig12 and FigS12] and the RMSE [FigS14] with initialisation of the seasonal forecast models on April,
May and June. The first period of evaluation is the seasonal time scale that provides a benchmark of the forecast for the rainy
season [Fig12-a), f), k)], (see also [FigS14a) -a),f),k)] in supplementary material). It gives a limited but significant improvement
with respect to the climatology. The scores (CRPS, RMSE) are then decomposed by month [Fig12- b-e), g-j), l-o)] (see also
[FigS14a) - b-e), g-j), l-o)] in supplementary material), to investigate the representation of the intraseasonal variability of the
SHL. We noticed an increase of the CRPS and RMSE values with the lead times over the CSHL location which lead to a loss
of predictability in the seasonal forecast models. SEAS5 shows more predictive skills over the CSHL location for short lead
time forecasting (0 to 1 month) and MF7 is a little better for long lead time (approximately 3 months). MF7 raw forecasts
show very limited skills over the Sahara (see [FigS12] and [FigS13b]) in supplementary material); this behavior in MF7 can
be related to hot biases evidenced in Section 3.2 ([Fig5-a)]). Bias correction improves considerably the predictive skills of
the models. The effect of bias correction on the predictive skills of the seasonal models is more efficient over the Sahara (see
[FigS12] and [FigS13b]) in supplementary material). This can be explained by the fact that climate models usually take into
account large-scale variability. As the Sahara is larger than the Central SHL box, the forecast models will better represent
the variability occurring over the Sahara; and the correction method will adjust the systematic bias present in the models.
Models present lower predictive skills for the daily T850 and an improvement is seen at monthly time scale for short lead time.
However, the correlation found at monthly time scale never exceeds 0.61 (see [FigS15-b)] in supplement material); at daily
time scale, the correlation is less than 0.53 (see [FigS15-a)] in supplement material) for all months at lead time 0. Correlations
at different lead times have also been computed but the resulting coefficients were even lower than those obtained at lead time
0 (not shown). An evaluation of the seasonal forecast models at very short lead time has been done for June, July, August and
September at lead time 0 by computing the CRPS score using daily T850 raw forecasts (see [FigS14] in supplement material),
similar results have been found with the unbiased forecasts (not shown). We observed a progressive increase of the CRPS
score with time before reaching the predictability horizon around 20 days. SEAS5 shows predictive skills at the first time steps
(1-3) days over the CSHL and (1-8) days over the Sahara location; and MF7 is more affect by the spin-up at the beginning
of the prediction. This behaviour in SEAS5 can be explained by the capacity of the model to represent large scale variability
occurring on the Sahara location. From these results, we observed some predictive skills in the seasonal forecast models at
intraseasonal time scale, however they remain weak for long period and for lead time beyond 1 month. These results are in
agreement with previous works which addressed the predictive skills of ECMWF ensemble system (Haiden et al., 2015).

As seen previously, MF7 and SEAS5 present different characteristics in terms of bias, particularly regarding HLW and HLE
events detection. This suggests that a multi-model ensemble approach may be a solution to improve the forecast skills of the
seasonal models. Surprisingly, the multi-model shows a predictive skill comparable to the individual models [Fig12]. This
shows that the predictive skills of an ensemble model do not depend on the number of members in the models.

5 Conclusion

This work assessed the representation of the SHL in two seasonal forecast models (SEAS5 and MF7) using ERA5 reanalyses as reference. The choice of ERA5 as reference for this evaluation was supported by a sensitive study conducted between ERA5 and another reanalysis data set, namely MERRA. Very high correlations have been found between the two reanalyses data, i.e. around 0.97 and 0.92, respectively, over CSHL and the Sahara location; robust similarities are also observed in the yearly distribution of T850. The discrepancies between the reanalyses are much smaller than the biases in the seasonal forecast models with respect to ERA5. Through a set of analyses, we have found opposite biases in the seasonal forecast models compared to ERA5. MF7 has a warm bias and tends to overestimate the intensity of the SHL with respect to ERA5. SEAS5 develops a cold bias and tends to underestimate the intensity of the SHL over the Sahara. The models are able to represent the mean seasonal cycle of the SHL and capture some characteristics of its inter-annual variability like the warming trends observed during the 2010^s. However, the good representation of this inter-annual variability remains challenging for the models. SEAS5 represents more realistically the climatic trend of the SHL than MF7. The bias correction methods CDF-t and QMAP are very efficient at reducing the systematic bias present in the seasonal models. By using bias correction tools, the results highlight the capacity of the models to represent the intraseasonal pulsations (the so-called sast-west phases) of the SHL. We notice an overestimation of the occurrence of the HLE phases in the models (SEAS5/MF7); the HLW phases are much better represented in MF7. This diagnosis is a first validation of the representation of the SHL in seasonal models. In spite of this, the correct timing of these pulsations is still a key challenge and the next step forward. Seasonal forecast models show predictive skills at intraseasonal time scale for short period. Bias correction contributes to improving the ensemble forecast score (CRPS) but the forecast skill remains weak for a lead time beyond 1 month. The issue of the lack of correlation in models cannot be solved through a bias correction approach, only model improvements could provide better correlations between forecasts and observations. In a future study, we will investigate the relationship between the SHL and the extreme rainfall in the Sahel region at intraseasonal time scale.

6 Acknowledgments

This work is supported by the French National Research Agency in the framework of the "Investissement d'avenir" program (ANR-15-IDEX-02) with the project PREDISAHLIM (2019-2021), and under grant ANR-19-CE03-0012 with the project STEWARd (2020-2024).

References

Bickle, M. E., Marsham, J. H., Ross, A. N., Rowell, D. P., Parker, D. J., and Taylor, C. M.: Understanding mechanisms for trends in Sahelian squall lines: Roles of thermodynamics and shear, Quarterly Journal of the Royal Meteorological Society, 2020.

Büssow, R.: An algorithm for the continuous Morlet wavelet transform, Mechanical Systems and Signal Processing, 21, 2970–2979, https://doi.org/10.1016/j.ymssp.2007.06.001, 2007.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes?, Journal of Climate, 28, 6938–6959, 2015.

Chauvin, F., Roehrig, R., and Lafore, J.-P.: Intraseasonal variability of the Saharan heat low and its link with midlatitudes, Journal of Climate, 23, 2544–2561, 2010.

Chou, C., Neelin, J. D., and Su, H.: Ocean-atmosphere-land feedbacks in an idealized monsoon, Quarterly Journal of the Royal Meteorological Society, 127, 1869–1891, 2001.

Cohen, M. X.: A better way to define and describe Morlet wavelets for time-frequency analysis, bioRxiv, p. 397182, https://doi.org/10.1101/397182, publisher: Cold Spring Harbor Laboratory Section: New Results, 2018.

Déqué, M., Dreveton, C., Braun, A., and Cariolle, D.: The ARPEGE/IFS atmosphere model: a contribution to the French community climate modelling, Climate Dynamics, 10, 249–266, 1994.

Dixon, R. D., Daloz, A. S., Vimont, D. J., and Biasutti, M.: Saharan heat low biases in CMIP5 models, Journal of Climate, 30, 2867–2884, 2017.

Dosio, A. and Paruolo, P.: Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate, Journal of Geophysical Research: Atmospheres, 116, 2011.

Drobinski, P., Sultan, B., and Janicot, S.: Role of the Hoggar massif in the West African monsoon onset, Geophysical research letters, 32, 2005.

Durand, J.-H.: A propos de la sécheresse et ses conséquences au Sahel, Les cahiers d'outre-mer, 30, 383–403, publisher: Presses Universitaires de Bordeaux, 1977.

Evan, A. T., Flamant, C., Lavaysse, C., Kocha, C., and Saci, A.: Water vapor–forced greenhouse warming over the Sahara Desert and the recent recovery from the Sahelian drought, Journal of Climate, 28, 108–123, 2015.

Fitzpatrick, R. G., Parker, D. J., Marsham, J. H., Rowell, D. P., Guichard, F. M., Taylor, C. M., Cook, K. H., Vizy, E. K., Jackson, L. S., Finney, D., et al.: What drives the intensification of mesoscale convective systems over the West African Sahel under climate change?, Journal of Climate, 33, 3151–3172, 2020.

François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: which benefits for which losses?, Earth System Dynamics, 11, 537–562, publisher: Copernicus GmbH, 2020.

Haiden, T., Bidlot, J., Ferranti, L., Bauer, P., Dahoui, M., Janousek, M., Prates, F., Vitart, F., and Richardson, D.: Evaluation of ECMWF forecasts, including 2014-2015 upgrades, European Centre for Medium-Range Weather Forecasts, 2015.

Henze, N. and Meintanis, S. G.: Recent and classical tests for exponentiality: a partial review with comparisons, Metrika, 61, 29–45, 2005.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hudson, D. and Ebert, B.: Ensemble Verification Metrics, in: Proceedings of the ECMWF Annual Seminar, Reading, UK, pp. 11–14, 2017.

Janicot, S., Mounier, F., and Diedhiou, A.: Les ondes atmosphériques d'échelle synoptique dans la mousson d'Afrique de l'Ouest et centrale : ondes d'est et ondes de Kelvin, Science et changements planétaires / Sécheresse, 19, 13–22, 2008a.

Janicot, S., Thorncroft, C. D., Ali, A., Asencio, N., Berry, G., Bock, O., Bourlès, B., Caniaux, G., Chauvin, F., Deme, A., et al.: Large-scale overview of the summer monsoon over West Africa during the AMMA field experiment in 2006, in: Annales Geophysicae, vol. 26, pp. 2569–2595, Copernicus GmbH, 2008b.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., et al.: SEAS5: The new ECMWF seasonal forecast system, Geoscientific Model Development, 12, 2019.

Knippertz, P., Fink, A. H., Deroubaix, A., Morris, E., Tocquer, F., Evans, M. J., Flamant, C., Gaetani, M., Lavaysse, C., Mari, C., Marsham, J. H., Meynadier, R., Affo-Dogo, A., Bahaga, T., Brosse, F., Deetz, K., Guebsi, R., Latifou, I., Maranan, M., Rosenberg, P. D., and Schlueter, A.: A meteorological and chemical overview of the DACCIWA field campaign in West Africa in June–July 2016, Atmospheric Chemistry and Physics, 17, 10 893–10 918, https://hal-insu.archives-ouvertes.fr/insu-01518455, 2017.

Lavaysse, C.: Warming trends: Saharan desert warming, Nature Climate Change, 5, 807–808, 2015.

Lavaysse, C., Flamant, C., Janicot, S., Parker, D. J., Lafore, J.-P., Sultan, B., and Pelon, J.: Seasonal evolution of the West African heat low: a climatological perspective, Climate Dynamics, 33, 313–330, 2009.

Lavaysse, C., Flamant, C., and Janicot, S.: Regional-scale convection patterns during strong and weak phases of the Saharan heat low, Atmospheric Science Letters, 11, 255–264, 2010a.

Lavaysse, C., Flamant, C., Janicot, S., and Knippertz, P.: Links between African easterly waves, midlatitude circulation and intraseasonal pulsations of the West African heat low, Quarterly Journal of the Royal Meteorological Society, 136, 141–158, 2010b.

Lavaysse, C., Chaboureau, J.-P., and Flamant, C.: Dust impact on the West African heat low in summertime, Quarterly Journal of the Royal Meteorological Society, 137, 1227–1240, https://doi.org/https://doi.org/10.1002/qj.844, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.844, 2011.

Lavaysse, C., Vrac, M., Drobinski, P., Lengaigne, M., and Vischel, T.: Statistical downscaling of the French Mediterranean climate: assessment for present and projection in an anthropogenic scenario, Natural Hazards and Earth System Sciences, European Geosciences Union, 12, 651–670, 2012.

Lavaysse, C., Flamant, C., Evan, A., Janicot, S., and Gaetani, M.: Recent climatological trend of the Saharan heat low and its impact on the West African climate, Climate Dynamics, 47, 3479–3498, 2016.

Lavaysse, C., Naumann, G., Alfieri, L., Salamon, P., and Vogt, J.: Predictability of the European heat and cold waves, Climate Dynamics, 52, 2481–2495, 2019.

Martin, G. M., Peyrillé, P., Roehrig, R., Rio, C., Caian, M., Bellon, G., Codron, F., Lafore, J.-P., Poan, D. E., and Idelkadi, A.: Understanding the West African Monsoon from the analysis of diabatic heating distributions as simulated by climate models, Journal of Advances in Modeling Earth Systems, 9, 239–270, publisher: Wiley Online Library, 2017.

Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophysical Research Letters, 36, 2009.

Parker, D. J., Burton, R. R., Diongue-Niang, A., Ellis, R. J., Felton, M., Taylor, C. M., Thorncroft, C. D., Bessemoulin, P., and Tompkins, A. M.: The diurnal cycle of the West African monsoon circulation, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131, 2839–2860, 2005.

Peyrillé, P. and Lafore, J.-P.: An idealized two-dimensional framework to study the West African monsoon. Part II: Large-scale advection and the diurnal cycle, Journal of the atmospheric sciences, 64, 2783–2803, 2007.

610 Prospero, J. M., Ginoux, P., Torres, O., Nicholson, S. E., and Gill, T. E.: Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product, Reviews of geophysics, 40, 2–1, 2002.

Redelsperger, J.-L., Thorncroft, C. D., Diedhiou, A., Lebel, T., Parker, D. J., and Polcher, J.: African Monsoon Multidisciplinary Analysis: An international research project and field campaign, Bulletin of the American Meteorological Society, 87, 1739–1746, 2006.

615 Roehrig, R., Chauvin, F., and Lafore, J.-P.: 10–25-day intraseasonal variability of convection over the Sahel: A role of the Saharan heat low and midlatitudes, Journal of climate, 24, 5863–5878, 2011.

Roehrig, R., Bouniol, D., Guichard, F., Hourdin, F., and Redelsperger, J.-L.: The present and future of the West African monsoon: A process-oriented assessment of CMIP5 simulations along the AMMA transect, Journal of Climate, 26, 6471–6505, publisher: American Meteorological Society, 2013.

620 Siegert, S., Bhend, J., Kroener, I., and De Felice, M.: Package 'SpecsVerification', 2017.

Sultan, B. and Janicot, S.: The West African monsoon dynamics. Part II: The "preonset" and "onset" of the summer monsoon, Journal of climate, 16, 3407–3427, 2003.

Tang, B., Liu, W., and Song, T.: Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-Ville distribution, Renewable Energy, 35, 2862–2866, https://doi.org/10.1016/j.renene.2010.05.012, 2010.

625 Taylor, C. M., Belušić, D., Guichard, F., Parker, D. J., Vischel, T., Bock, O., Harris, P. P., Janicot, S., Klein, C., and Panthou, G.: Frequency of extreme Sahelian storms tripled since 1982 in satellite observations, Nature, 544, 475–478, https://doi.org/10.1038/nature22069, 2017.

Thorncroft, C. D. and Blackburn, M.: Maintenance of the African easterly jet, Quarterly Journal of the Royal Meteorological Society, 125, 763–786, 1999.

Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A., and Gneiting, T.: Skill of global raw and postprocessed ensemble predictions of rainfall 630 over northern tropical Africa, Weather and Forecasting, 33, 369–388, 2018.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Mélia, D. S. y., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-635 1, Journal of Advances in Modeling Earth Systems, 11, 2177–2213, https://doi.org/https://doi.org/10.1029/2019MS001683, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001683, 2019.

Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R 2 D 2) bias correction, Hydrology and Earth System Sciences, 22, 3175–3196, publisher: Copernicus GmbH, 2018.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., and Somot, S.: Dynamical and statistical downscaling of the French 640 Mediterranean climate: uncertainty assessment, Natural Hazards and Earth System Sciences, European Geosciences Union, 12, 2769–2784, 2012.

Zhao, G., Jiang, D., Diao, J., and Qian, L.: APPLICATION OF WAVELET TIME-FREQUENCY ANALYSIS, p. 10, 2004.
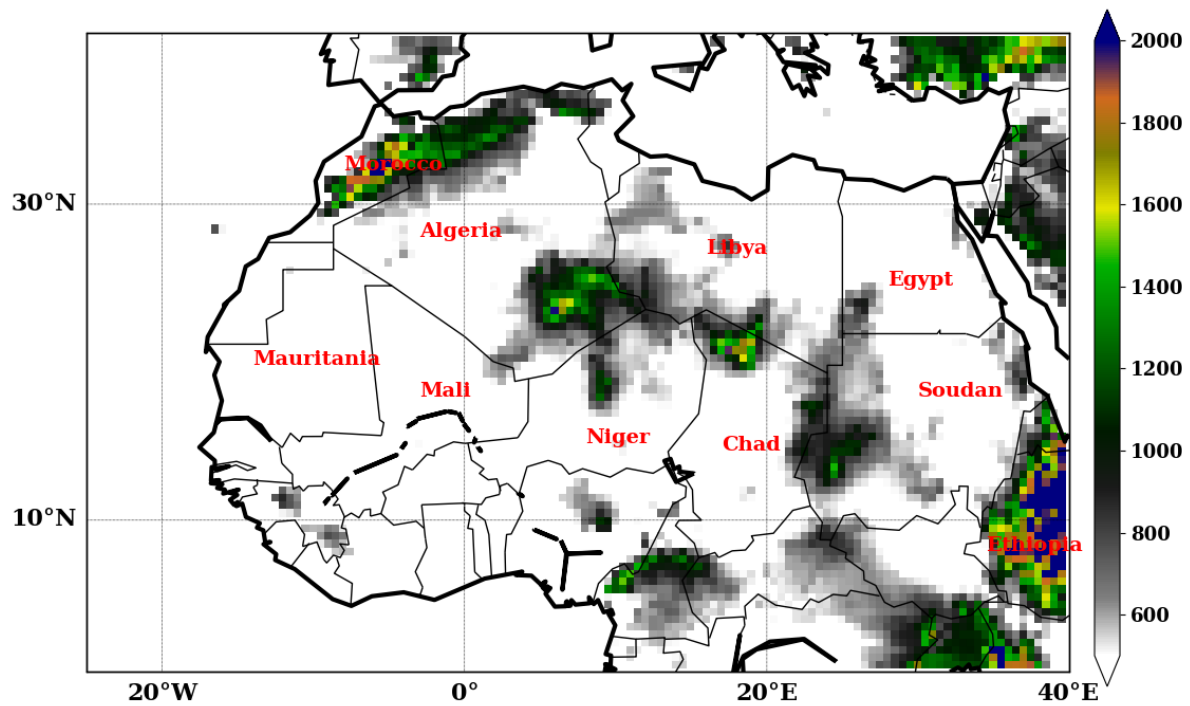
Figure 1. *Topographic map of West Africa using ERA5 elevation data. Y-axis and X-axis represent respectively the latitude and longitude in degree over the domain. The color bar is showing the elevation in meter over the region.*
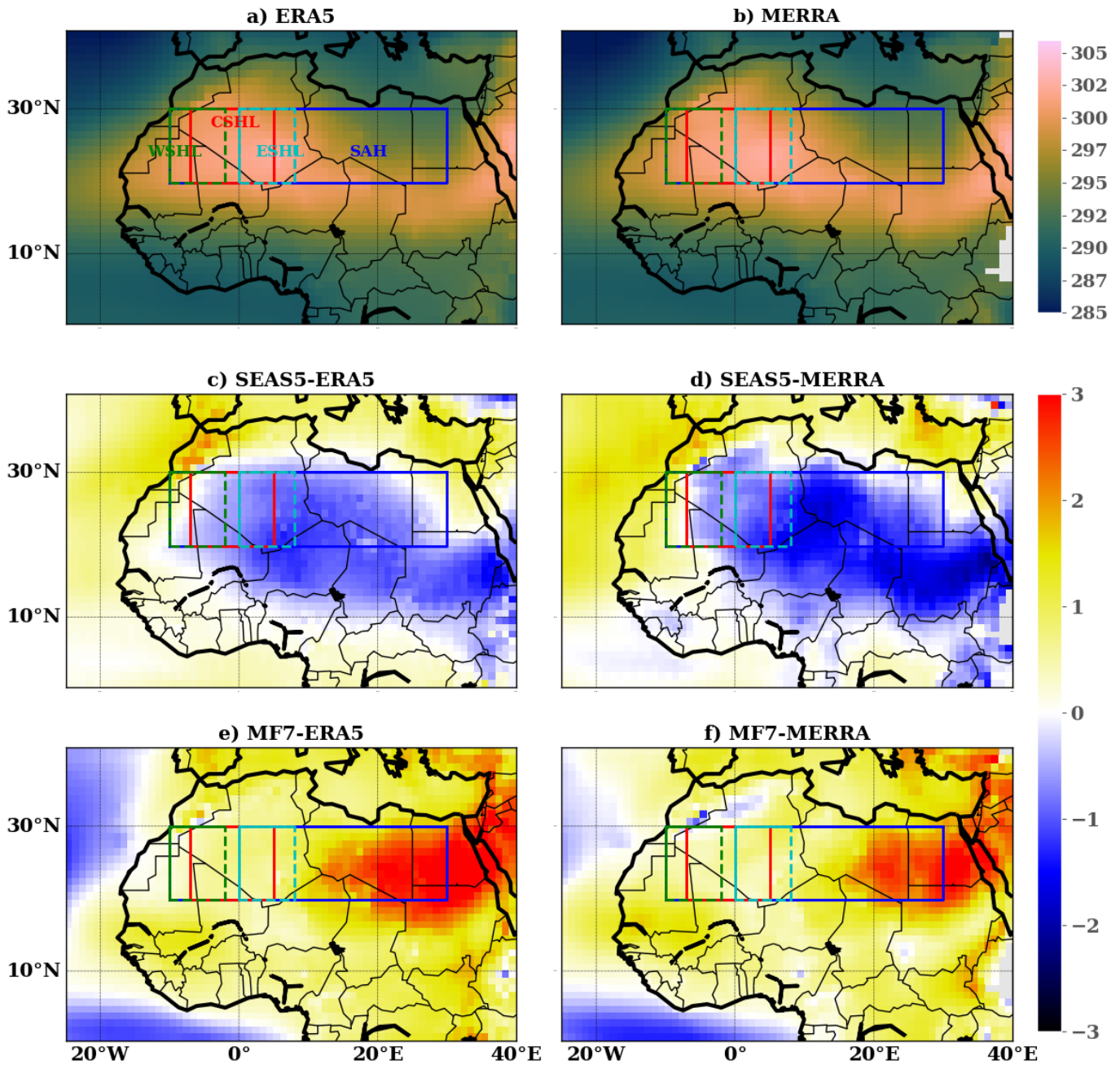
Figure 2. *Climatology of the SHL during the JJAS period over 1993-2016 in the reanalysis data using T850 : a) ERA5 and b) MERRA; and the anomalies of the climatologies of the SHL between: c) SEAS5 and ERA5, e) SEAS5 and MERRA, d) MF7 and ERA5 and f) MF7 and MERRA*. The rectangles indicate the boxes chosen for the computation of the average T850 and their corresponding name. "**WSHL**": West SHL, "**CSHL**": Central SHL, "**ESHL**": East SHL boxes and "**SAH**", the Saharan region. The color bars indicate respectively T850 and the anomalies of T850 in degree Kelvin. The computation was made using the ensemble mean member.
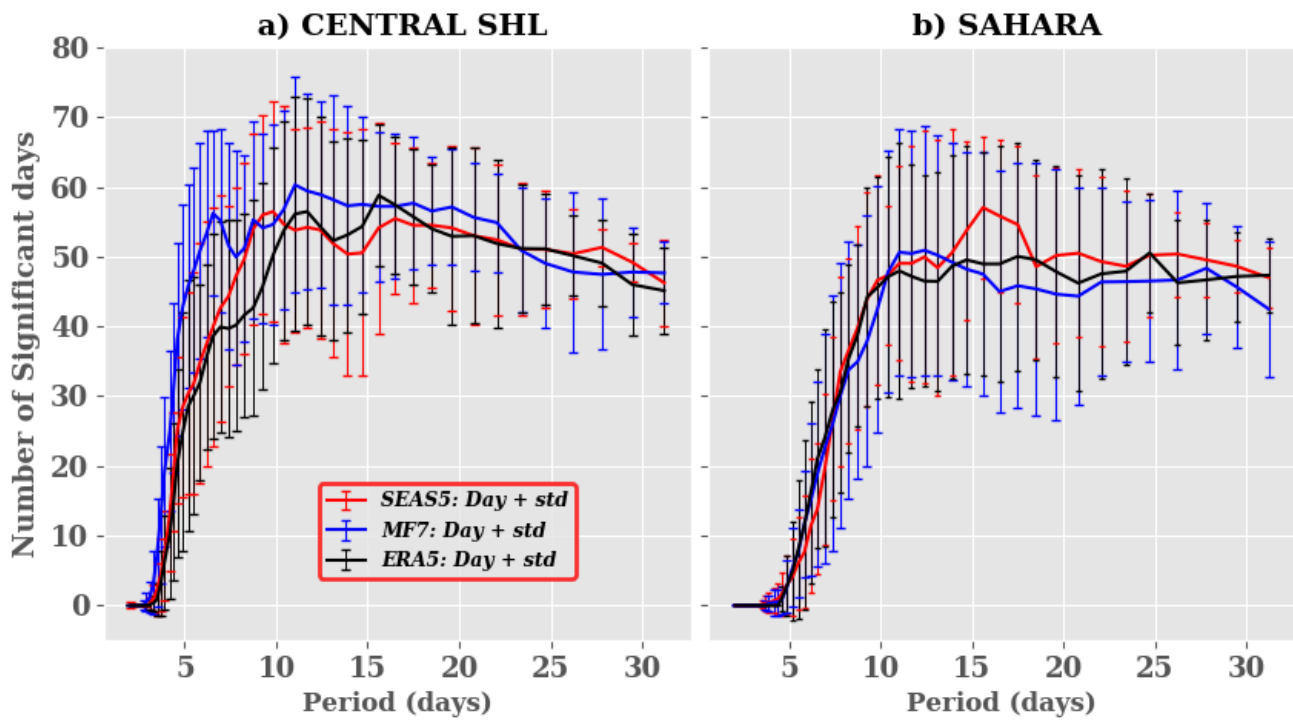
Figure 3. *Climatology of significant days: significant days here refer to days with spectral power signal greater than 1. Red, blue, black curves and bars represent respectively SEAS5, MF7, ERA5 number of days and spread over: **a)** central SHL box and **b)** Sahara during the period 1993-2016. The computation was made just using the unperturbed member of the ensemble forecast models launched from the 1st of June for the JJAS period. Y-axis represents significant days and X-axis the duration of propagation in days.*
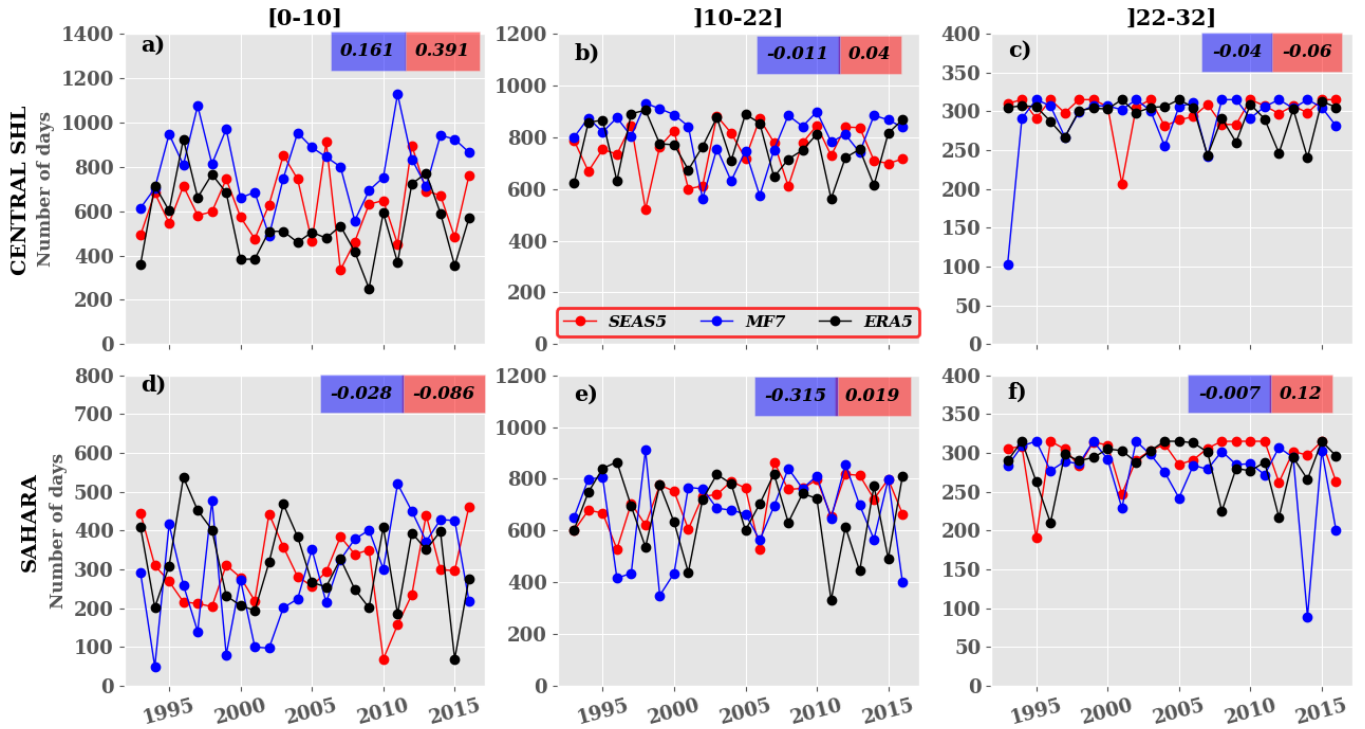
Figure 4. *Inter-annual variability of significant days: significant days here refer to days with spectral power signal greater than 1. Red, blue and black curves represent respectively SEAS5, MF7, ERA5 number of days over: **a) - c)** central SHL and **d) - f)** Sahara. The values on red and blue boxes refer to the correlation respectively between SEAS5 and ERA5, MF7 and ERA5. $[0, 10],]10, 22],]22, 32]$ are the different classes of days identified for the present study.The computation was made just using the unperturbed member of the ensemble forecast models launched from the 1st of June for the JJAS period. Y-axis represents significant days and X-axis the time in year.*
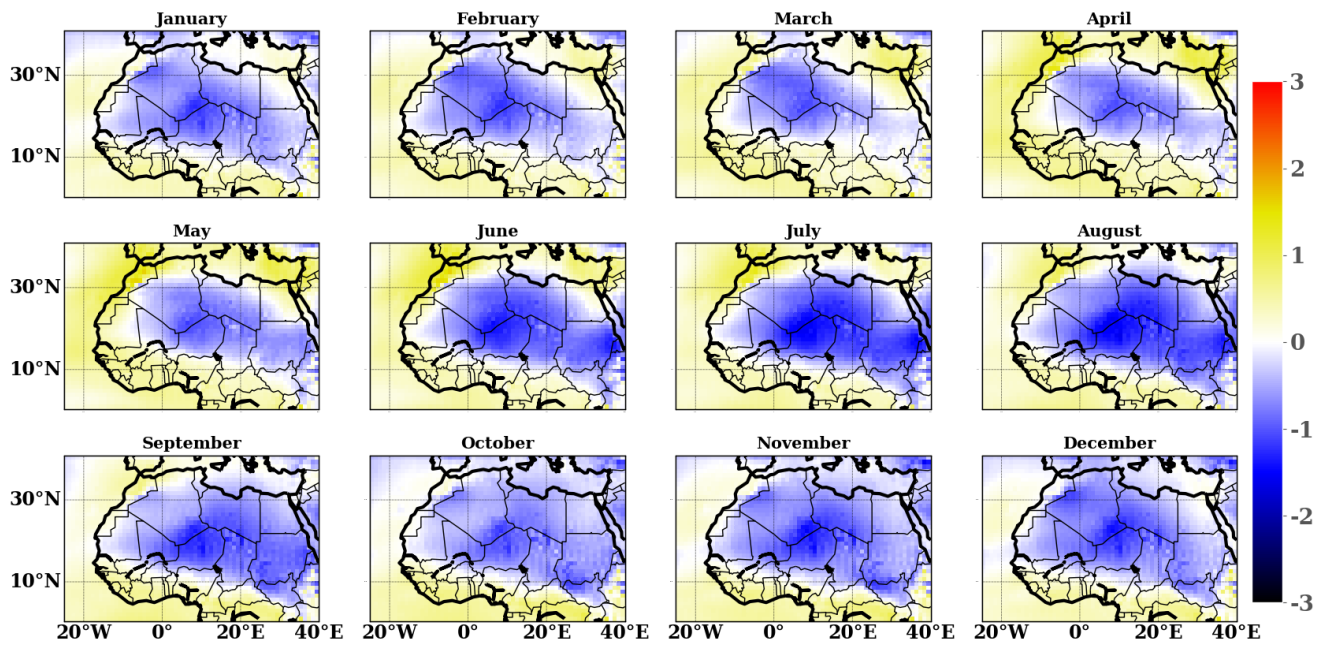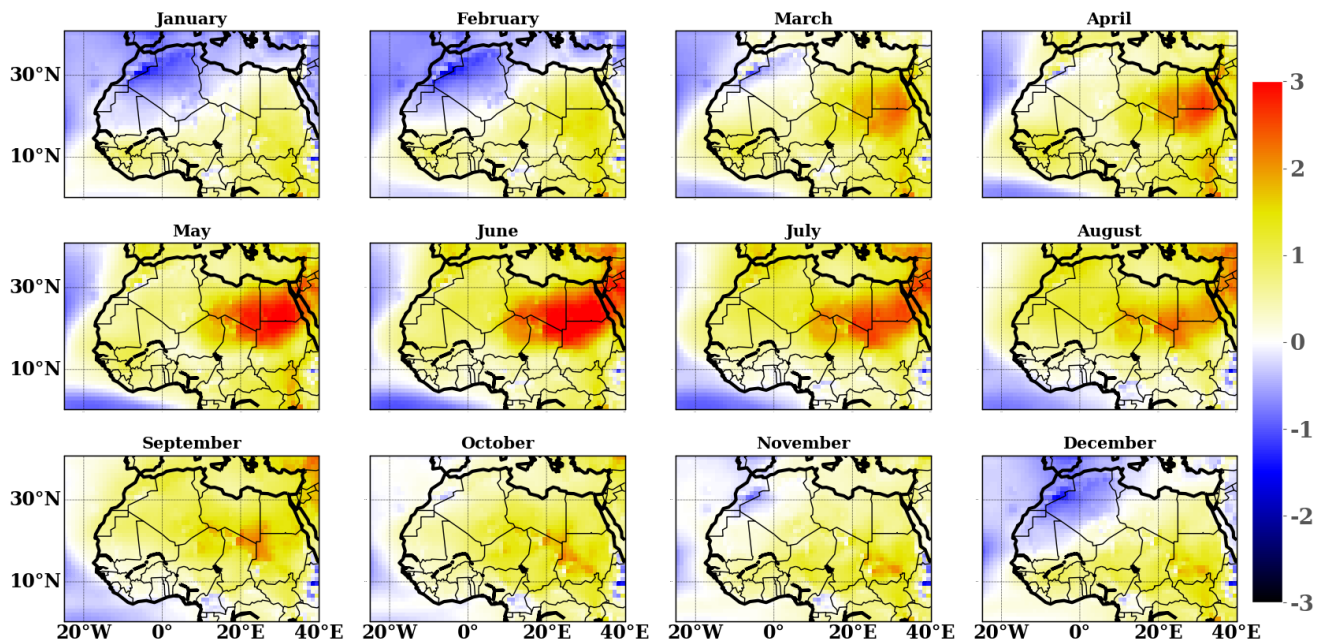
Figure 5. *Climatology of monthly bias temperature over Sahara region during 1993-2016 between: **a**) SEAS5 and ERA5, **b**) MF7 and ERA5. The bias is computed using daily temperature at 00:00 and 12:00 UTC. The computation was made using the ensemble mean for forecast models. The color bar indicates the bias value in Kelvin.Y-axis indicates latitudes and X-axis the longitudes of our domain.*
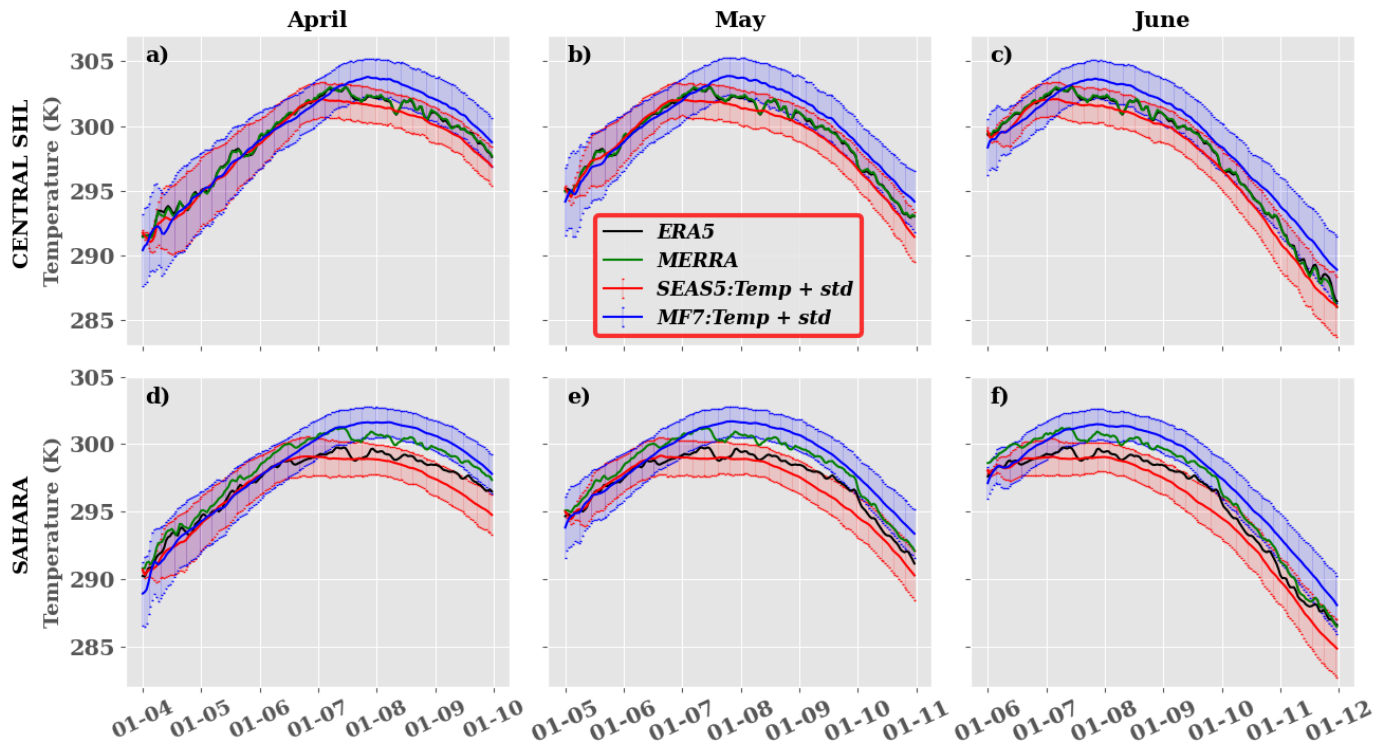
Figure 6. *Climatology and spread of mean daily temperature during 1993-2016 at different initialization months for a 6-month forecast: April, May and June respectively on: **a**) - **c**) central SHL box and **d**) - **f**) Sahara. Black, green, red and blue bold curves refer respectively to ERA5, MERRA, SEAS5 and MF7 mean T850; Red and blue bars represent the inter-member spreads respectively for SEAS5 and MF7. The computation was made using the ensemble mean of forecast models. Y-axis indicates temperature in degree Kelvin and X-axis the time in year.*
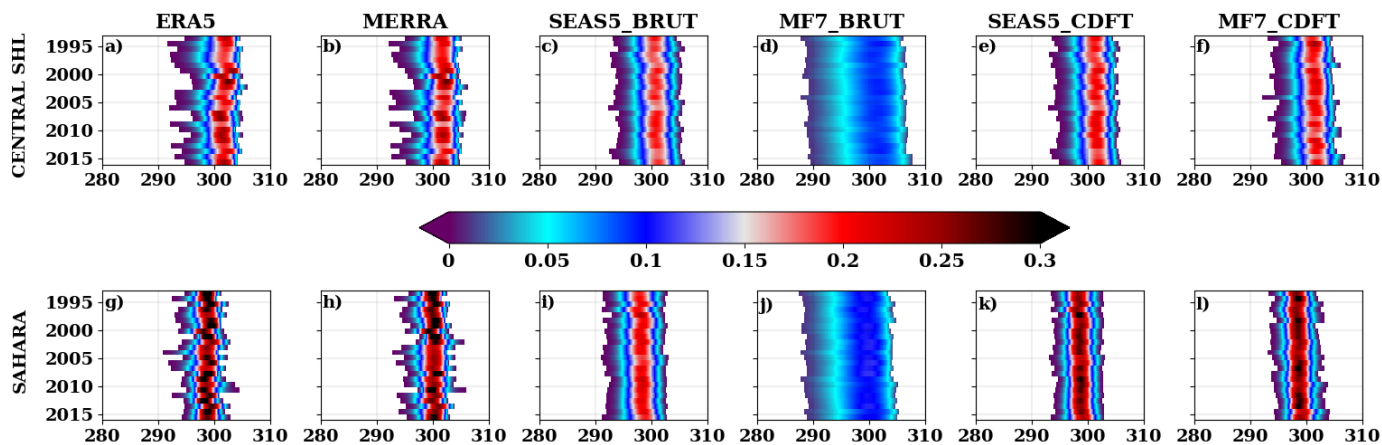
Figure 7. *Distribution of yearly T850 over JJAS period during 1993-2016 respectively over: **a) - f)** central SHL box and **g) - l)** Sahara. "ERA5","MERRA", "SEAS5_BRUT", "MF7_BRUT" here correspond to the intensity of the SHL using ERA5, MERRA for the reanalyses; SEAS5 and MF7 raw forecasts respectively. "SEAS5_CDFT", "MF7_CDFT" refer to the intensity of the SHL using SEAS5 and MF7 seasonal forecasts bias corrected respectively using ERA5. The computation was made using the ensemble members of the forecast models. Y-axis indicates time in year and X-axis T850 in degree Kelvin. The color bar indicates the probability of occurrence.*
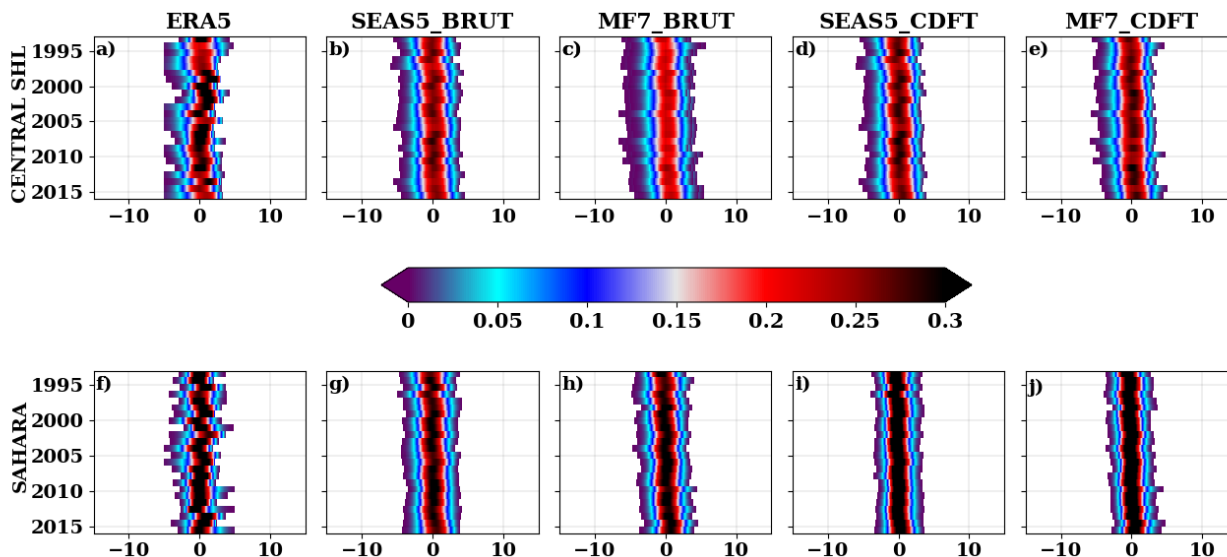
Figure 8. *Same as [Figure7] but for the yearly anomalies of temperature. The anomalies are computed by removing the daily climatology temperature for each year.*
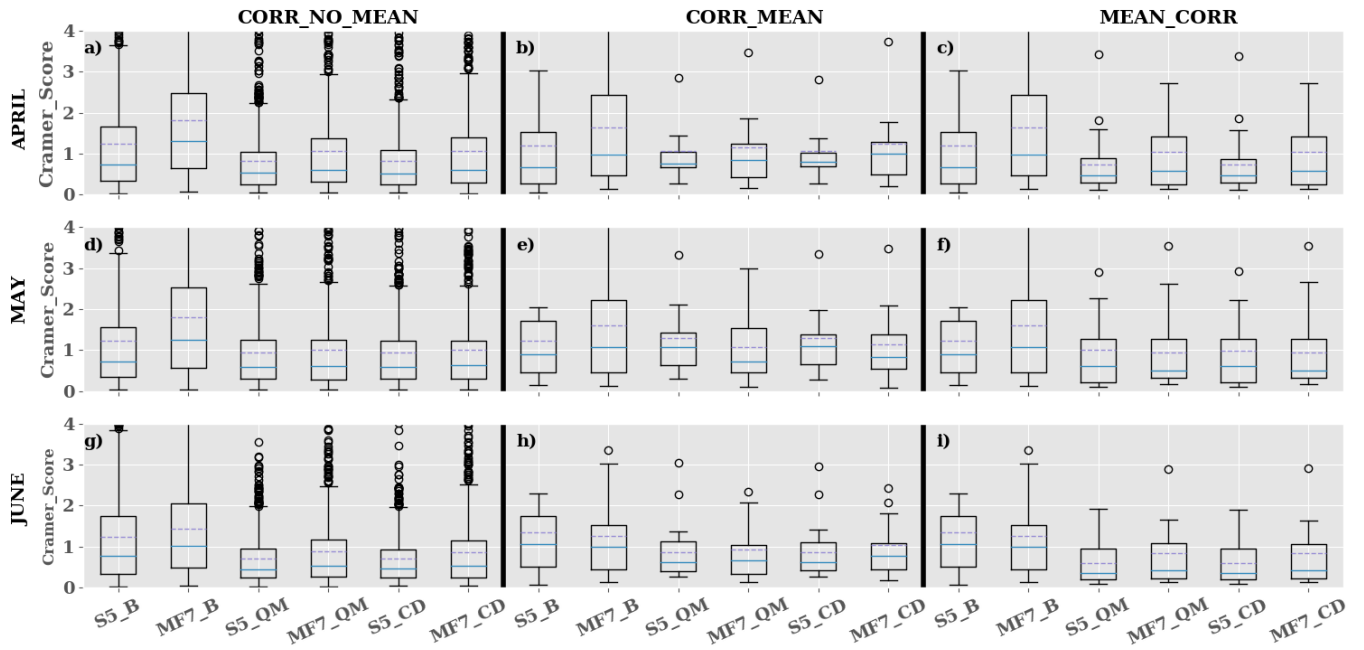
Figure 9. *Bias Correction evaluation using Cramer Von Mises score over JJAS period during 1993-2016 on the central SHL box at different forecast initialization months: **a) - c)** April , **d) - f)** May and **g) - i)** June respectively. **"CORR_NO_MEAN", "CORR_MEAN", "MEAN_CORR"** methods are well described in **section 3.4**. "S5_B","S5_CD","S5_QM" represent the Cramer score computed using respectively SEAS5 raw forecasts, SEAS5 corrected with CDF-t and QMAP methods. Idem for "MF7_B","MF7_CD","MF7_QM" with the MF7 model. Y-axis indicates the Cramer score and X-axis the different products used for the computation of Cramer score.*
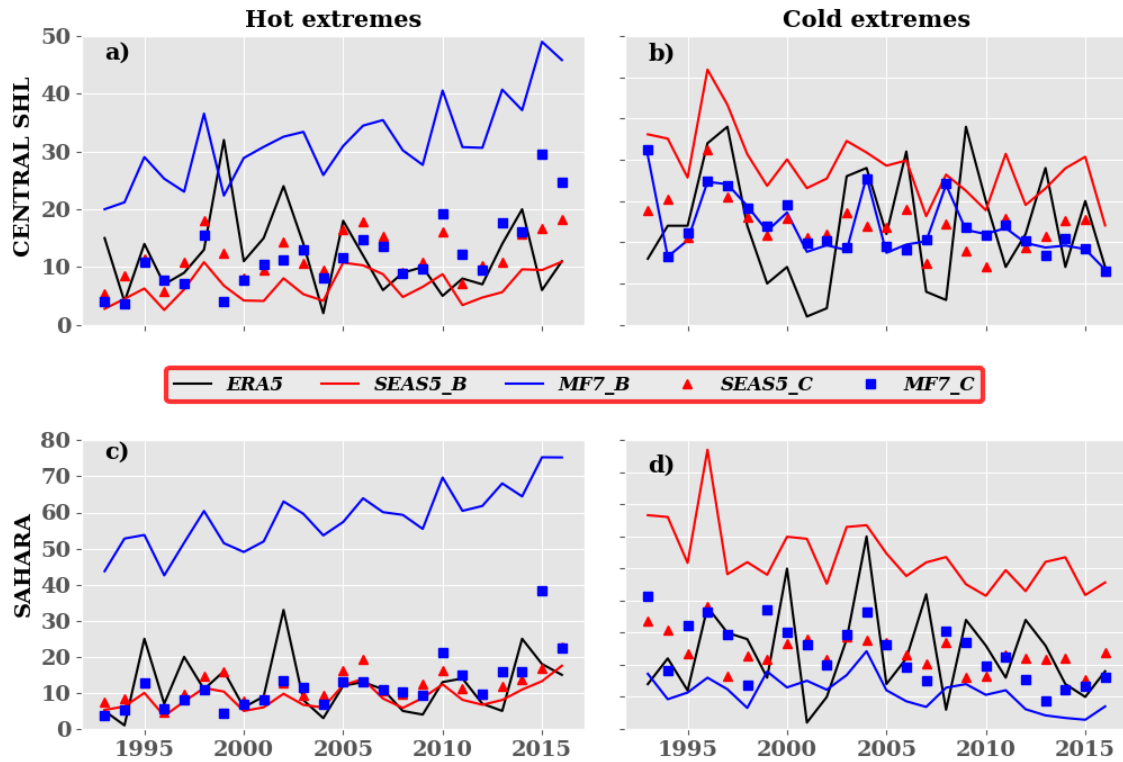
Figure 10. *Inter-annual variability of the SHL extremes over: **a), b)** central SHL and **c), d)** Sahara during the JJAS period from 1993 to 2016. $SEAS5\_C, MF7\_C$ refer respectively to corrected forecast with CDF-t method and $SEAS5\_B, MF7\_B$ represent model raw forecasts. X-axis indicates the time (year) and Y-axis the number of extremes registered for each year. Hot extremes are events occurring above the 90th percentile and Cold extremes associated to events below the 10th percentile.*
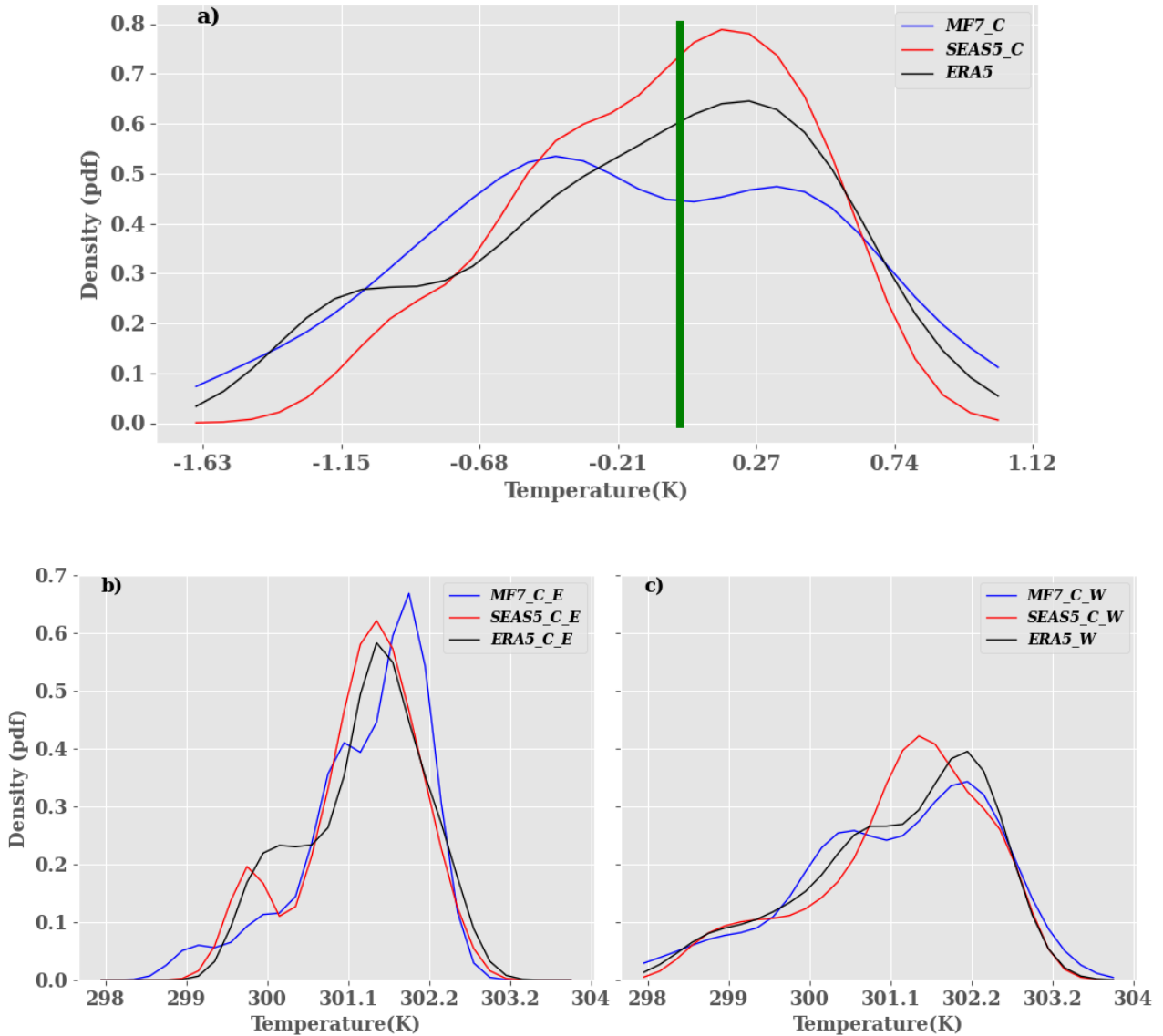
Figure 11. *Distribution of the climatology over the period $20^{th} June - 17^{th} September$ from 1993 to 2016 at June lead time 0 for: a) the dipole which represents the difference between heat low West and heat low East, b) Heat low East and c) Heat low West. $MF7\_C$ and $SEAS5\_C$ refer respectively to the MF7/SEAS5 forecasts corrected with CDF-t method. "$ERA5\_E$", "$MF7\_C\_E$" and "$SEAS5\_C\_E$" refer respectively to the HLE in the reanalyses, MF7/SEAS5 forecasts corrected with CDF-t method. "$ERA5\_W$", "$MF7\_C\_W$" and "$SEAS5\_C\_W$" refer respectively to the HLW in the reanalyses, MF7/SEAS5 forecasts corrected with CDF-t method.Y-axis indicates the probability of occurrence and X-axis the temperature in degree Kelvin.The vertical green bar represents the boundary between the HLE and HLW phases. The analysis was carried out using the unperturbed member.*
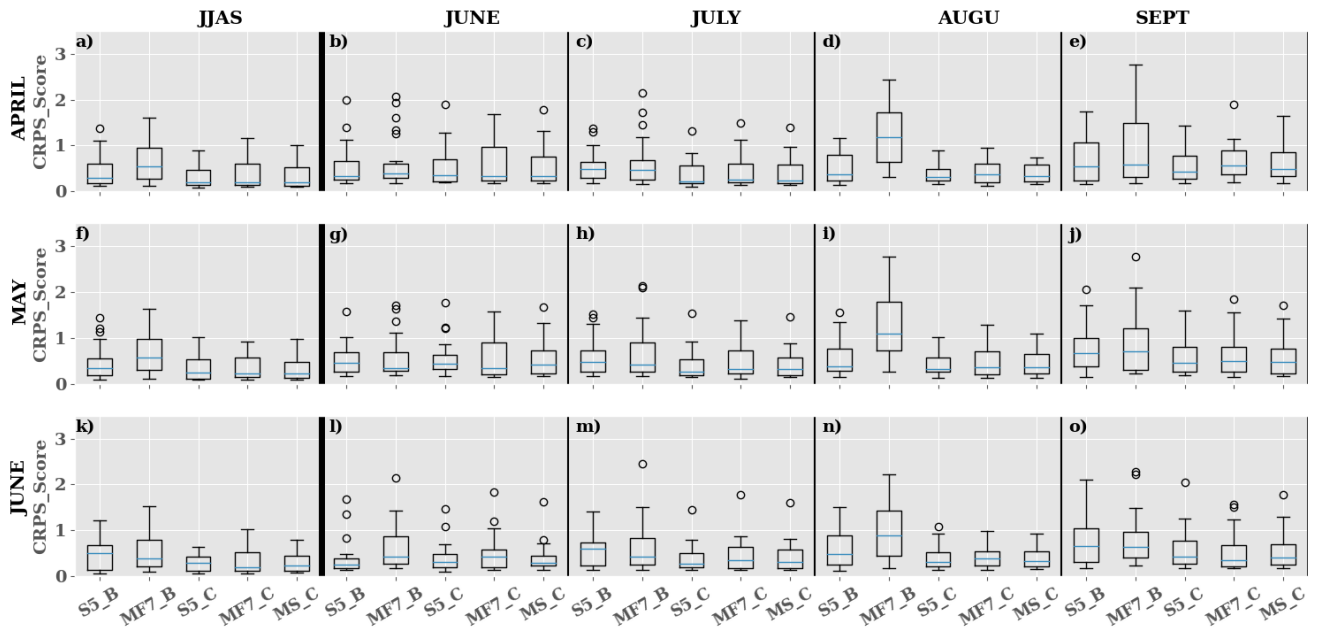
Figure 12. *Evaluation of the interannual variability of the SHL over the JJAS period and separately on June, July, August, September during 1993-2016 using monthly mean T850 over the Central SHL box at different initialization months: **a) - e)** April , **f) - j)** May and **k) - o)** June respectively. "S5_B","S5_C" represent respectively the CRPS score evaluated using respectively SEAS5 raw and corrected forecast with the CDF-t method. Idem for "MF7_B","MF7_C" with the MF7 model. "MS_C" represents the CRPS score evaluated on the multi-model formed by SEAS5 and MF7 corrected forecasts with the CDF-t method. The computation was made using the ensemble member both for corrected and raw forecasts. The Y-axis indicates the CRPS score values and X-axis, the data type used for the computation of the CRPS.*