Forecasting West African heat waves at sub-seasonal and seasonal time scales

Lauriane Batté*, Constantin Ardilouze and Michel Déqué

CNRM UMR 3589 (Météo-France/CNRS), Toulouse, France.

Corresponding author address: CNRM UMR 3589 (Météo-France/CNRS), Toulouse, France.

E-mail: lauriane.batte@meteo.fr

ABSTRACT

Early indication of an increased risk of extremely warm conditions could help alleviate some of the consequences of severe heat waves on human health. This study focuses on boreal spring heat wave events over West Africa and Sahel, and examines the long-range predictability and forecast quality of these events with two coupled forecasting systems designed at Météo-France, and both based on the CNRM-CM coupled global climate model: the operational seasonal forecasting System 5, and the experimental contribution to the WWRP/WCRP S2S project. Evaluation is based on past re-forecasts spanning 22 years from 1993 to 2014, comparing to reference data from reanalyses. On the seasonal time scale, skill in reproducing heat wave duration and number of days interannual anomalies is limited at a grid-point level, but significant for regional averages. Sub-seasonal predictability of daily humidity-corrected apparent temperature drops sharply beyond the deterministic range. In addition to re-forecast skill measures, the analysis of real-time forecasts for 2016 both in terms of anomalies with respect to the re-forecast climatology and using a weather type approach provides additional insight on the systems performance in giving relevant information on the possible occurrence of such events.

1. Introduction

In a warming world, many areas are subject to an increase in the magnitude of heat waves, mainly due to the rise in seasonal mean temperatures, and future projections imply that the number and severity of heat waves should rise during the 21st century (Russo et al. 2014). Sherwood and Huber (2010) argue that consequences on human health could be dramatic for populations around the globe due to an increased exposure to heat stress. Mora et al. (2017) show that the number of days with potentially fatal climatic conditions could increase dramatically across the globe following the IPCC greenhouse gas emission scenarios.

Africa is no exception, although few studies have focused to date on extreme temperature events over the region (Conway 2011), mainly due to limited access to in-situ data. Two studies by Ly et al. (2013) and Moron et al. (2016) found using daily station data a decrease in cold spells and an increase in heat waves in the Sahel and West African regions over the second half of the 20th century. Analysis of daily data in western central Africa (Aguilar et al. 2009) and southern and western Africa (New et al. 2006) led to similar conclusions. Fontaine et al. (2013) found a positive trend in heat wave occurrences over Sahel in gridded surface temperature data. Using different methodologies, Barbier et al. (2017) and Ceccherini et al. (2017) compared heat wave characteristics in in-situ data and reanalysis datasets for the last decades over Sahel and the African continent, respectively. Ceccherini et al. (2017) found an increase in heat wave intensity in the last decades, and Barbier et al. (2017) showed that minimum and maximum temperature trends during heat waves were generally consistent with the background (positive) trends for these variables.

There is no single definition of a heat wave, depending on the time scale, location and impact considered (Perkins 2015), but most heat wave definitions imply that temperature (whether absolute or wet bulb) should reach above a fixed or relative threshold, and the event should last

for several consecutive days. In a recent study, Déqué et al. (2017) evaluated the impact of a 2°C warmer global climate on tropical Africa using CORDEX regional climate model (RCM) simulations. They found a very large increase in the number of heat wave days when using a fixed threshold based on percentiles of present climate, but also (although much more moderate) when adapting the threshold to future climate. Moreover, in Africa the implications for maximum temperature reaching values 2°C higher than in present climate could be catastrophic for human health, since monthly mean maximum temperatures already reach well over 40°C over some areas. Dosio (2017) inferred similar conclusions using RCMs to downscale CMIP5 projections.

The consequences of heat waves on human activities and health make skillful and reliable climate forecasts at sub-seasonal to seasonal time scales all the more needed to anticipate risks and develop adapted responses (Lowe et al. 2016). Unfortunately, seasonal prediction skill for temperature extremes is often quite limited, and directly related to skill in predicting the seasonal mean of the physical variable (Pepler et al. 2015). Most studies on a seasonal time scale focus therefore on monthly or seasonal mean temperature values or quantiles. Lazenby et al. (2014) found evidence of skill in forecasting seasonal mean maximum and minimum temperature over South Africa. Hamilton et al. (2012) studied the seasonal prediction skill of the Met Office GloSea4 system in forecasting counts of extreme temperature days over Northern Hemisphere midlatitudes. In the present study, we include an additional challenge for our seasonal forecasting system by studying its ability to correctly forecast interannual variability of persistent (e.g. beyond several days) extreme heat conditions in a given season.

The focus of studies on sub-seasonal to seasonal predictability over West Africa in general, and Sahel in particular, is often set on precipitation and drought (Conway 2011). Many studies suggest that seasonal forecasting skill over the region remains modest for summer precipitation (Rodríguez-Fonseca et al. 2015; Batté and Déqué 2011; Tompkins and Feudale 2010; Philippon

et al. 2010) and monsoon onset dates (Vellinga et al. 2013). Temperature anomalies over West Africa have received much less attention in seasonal and sub-seasonal predictability studies, although temperature is often more predictable than precipitation at these time scales over most areas of the globe (Doblas-Reyes et al. 2013). For instance, global evaluations of ECMWF System 4 seasonal forecasts reliability by Weisheimer and Palmer (2014) showed that temperature forecasts were more reliable than precipitation forecasts over West Africa. Pepler et al. (2015) showed some gridpoint evaluations of correlation for surface temperature and maximum and minimum temperature over the globe using the ENSEMBLES project multi-model, and some areas with significant skill were found over West Africa. However, both of these studies focus on seasonal means at a global scale for summer and winter seasons. Near-surface temperatures peak over most areas of West Africa and Sahel during the end of the dry season (Fontaine et al. 2013), and heat waves are more likely to occur then. To our knowledge, this study is the first to evaluate heat wave prediction skill at a seasonal time scale with a focus on Sahel during boreal spring.

Depending on their definition and the spatial location, heat waves typically last five days over Sahel (Barbier et al. 2017). This time scale implies that seasonal forecast information can only provide an indication as to an increase in the probability of heat waves occurring, without pinpointing precisely when heat waves are expected to happen during the season. The sub-seasonal scale, ranging beyond deterministic ranges up to several weeks, is relevant for decision-making related to such types of events (White et al. 2017). In the scope of the WMO World Weather Research Programme (WWRP) - World Climate Research Programme (WCRP) Sub-seasonal to Seasonal Prediction Project (S2S, Vitart et al. 2016), several international weather centers routinely provide S2S forecasts for research purposes, to bridge the gap between medium-range and seasonal ensemble forecasts (Brunet et al. 2010). Hudson et al. (2016) illustrate how the Australian Bureau of Meteorology coupled model experimental extended-range forecasts provide information on heat

wave risk for three case studies in Australia. In a recent case study over Europe, Ardilouze et al. (2017) investigated the skill in forecasting extreme surface temperatures with the Météo-France experimental system participating in S2S.

The aim of this present study is to assess the forecast quality of the Météo-France (1) seasonal forecasting System 5 included in the EUROSIP multi-model seasonal predictions (Vitart et al. 2007) and (2) S2S contribution in providing early insight on the occurrence of lasting warm temperature extremes at these time scales over West Africa. For seasonal forecast quality assessments, we focus on the March to June (MAMJ) target season for end of January initializations, corresponding to forecast times of two to five months. The same coupled model is used for the S2S project, and a first assessment of sub-seasonal skill is provided alongside a case study illustration for April 2016.

The paper is organized as follows: section 2 describes in detail the Météo-France seasonal forecasting System 5 as well as the S2S re-forecasts and real-time forecasts, and discusses the heat wave metrics and reference data used for this study. Section 3 shows results in terms of seasonal and S2S forecast skill over a 22-year retrospective period covering 1993–2014. We then focus on the MAMJ 2016 season as a case study illustration for both time scales in section 4, in the context of a strong El Niño event in the preceding winter. Section 5 uses a weather type approach as an additional way of decoding model inconsistencies with the reference data for the case study. Conclusions and perspectives for this work are summarized in section 6.

2. Forecast systems and methods

a. Météo-France System 5 based on CNRM-CM

The Météo-France seasonal forecasting System 5 (henceforth "System 5") is based on the CNRM-CM global coupled model, updated with respect to CNRM-CM5 used for CMIP5 and described in Voldoire et al. (2013). The ARPEGE-Climat atmosphere component is version 6.0, with diagnostic physics as in CNRM-CM5. The 91 vertical levels allow for a high-top explicit representation of the atmosphere up to 0.01 hPa, including stratospheric ozone, non-orographic gravity wave drag and quasi-biennal oscillation parameterizations (Cariolle and Déqué 1986; Lott and Guez 2013). The horizontal grid is a reduced gaussian grid with a linear truncation at t255, corresponding to an approximate resolution of 0.7 degrees in latitude and longitude. The ocean model is NEMO version 3.2 (Madec 2008) as in CNRM-CM5, using the tripolar ORCA1L42 grid with a nominal resolution of 1 degree and a horizontal refinement near the equator. Land surface is modelled with the SURFEX v7.3 scheme (Masson et al. 2013), and the sea ice component is an updated version of the GELATO sea ice model used for CMIP5. The ARPEGE-SURFEX system is coupled daily to NEMO-GELATO using the OASIS3 coupler (Valcke 2006).

Table 1 describes the main characteristics of the seasonal and subseasonal re-forecasts and forecasts based on System 5, commented in the following two paragraphs.

b. Seasonal re-forecasts and forecasts

Retrospective forecasts (referred to as "re-forecasts") for System 5 were run up to 7 months lead for each calendar month over the 1993-2014 period, using a 15-member ensemble, starting from the ERA-Interim reanalysis in the atmosphere (Dee et al. 2011), and initial ocean and sea ice conditions provided by Mercator-Ocean (Ferry et al. 2010). Ensemble members were generated using

the stochastic dynamics technique described in Batté and Déqué (2016). 51-member operational System 5 ensemble forecasts are run each start month since May 2015 based on initial conditions from the second half of the previous month: 25 members starting from the first Wednesday after the 12th of the month, and 26 members the following Wednesday, so as to ensure a timely delivery of the seasonal forecast ensemble at the beginning of the first forecast month. These are initialized from ECMWF IFS 00Z atmosphere and land analyses of the corresponding start date and an upscaling of Mercator-Ocean's GLORYS2V4 operational ocean analysis to the ORCA1 grid.

c. Subseasonal re-forecasts and forecasts

In the case of the subseasonal range, Météo-France does not issue operational forecasts, but takes part on a voluntary basis in the subseasonal-to-seasonal (S2S) project. The participating centers subseasonal re-forecast data and real-time forecasts are made freely available on the project database after several weeks delay (Vitart et al. 2016).

Re-forecasts for the Météo-France contribution to S2S are initialized on the 1st and 15th of each month from January 1993 to December 2014 (Ardilouze et al. 2017). As for the seasonal re-forecasts, the ensemble size is of 15 members, whereas the real-time forecasts form a 51-member ensemble. Note however that unlike the seasonal System 5 runs, the S2S real-time forecasts are generated in "burst" mode (i.e. all members starting on the same date) each Thursday, using stochastic dynamics perturbations to generate ensemble spread. This allows for each member of the ensemble to benefit from the most up-to-date initial conditions both for the ocean and atmospheric counterparts at the time of the forecast start. Initial conditions for the re-forecasts and real-time runs are derived as for the seasonal forecasting system.

d. Heat wave occurrence and verification metrics

Depending on specific applications and the geographical area of interest, there are many ways to define the occurrence of heat waves, and therefore a variety of heat wave indices have been used in past works (see e.g. Perkins (2015) or Grotjahn et al. (2016) for some examples). In the case of the March to June (MAMJ) spring season in West Africa, the use of absolute thresholds commonly defined over mid-latitudes is irrelevant, as temperatures reach well above the values describing extremely warm days or tropical nights on a regular basis. For this study we decided to use the 90th percentiles of each gridpoint daily minimum and maximum screen level temperature (Tmin and Tmax, respectively) climatology over the re-forecast period, after applying a 5-day moving window average to the data. Tmin and Tmax heat waves are defined separately, when Tmin (respectively Tmax) reaches values above its 90th percentile for at least 3 days. This definition therefore takes into account the seasonal cycle of Tmin and Tmax over the region, so that days above this threshold can occur during the whole spring season, even when the amplitude of the seasonal cycle is quite large.

For both Tmin and Tmax heat waves, two indices characterize a given MAMJ season at each grid point:

- HWDI (heat wave duration index) is the number of days per MAMJ season for which a heat wave occurs (once the 3 day limit is applied),
- HWPP (heat waves per period) is the number of separate heat wave events per MAMJ season.

The definition of the daily climatology using a 5-day window may not be the most accurate estimate (with respect to other methods such as polynomial regression fitting, see Mahlstein et al. (2015)). We chose this method for the sake of simplicity, and more elaborate techniques often call for a longer data sample than the re-forecast depth for System 5. Based on results from

Mahlstein et al. (2015) with a perfect model approach, this choice could have a marginal impact on the estimates of re-forecast skill for the percentile-based threshold indices discussed in this paper. Our hypothesis is that this additional source of uncertainty can be neglected (with respect to other sources such as the skill evaluation method, the reference data, and sub-sampling with a short re-forecast and limited ensemble size).

Temperature is not the only physical factor of heat stress for humans. Extreme heat can be made unbearable by high levels of ambient humidity, or conversely be alleviated by a dry environment. We therefore computed a humidity-corrected apparent temperature (AT) as in Fischer and Schär (2010), but for daily mean screen level temperature and relative humidity estimated from the daily mean screen level dewpoint temperature as in Lee and Brenner (2015). Computations were done with daily mean data due to forecast data availability; the AT values computed are likely much lower than the maximum AT values one would get using Tmax and simultaneous relative humidity, since the diurnal cycle of relative humidity during the pre-monsoon season over Sahel is often weak (Gounou et al. 2012). Forecast and re-forecast indices were computed after bias-correcting mean temperature and mean dewpoint temperature fields. Details of the computation are provided in the supplemental material. In the case of AT, we chose to focus on absolute threshold excedences to define an AT35 heat wave when the daily mean apparent temperature exceeds 35 degrees for more than three days. As for Tmin and Tmax, we compute HWDI and HWPP values for each gridpoint and MAMJ season.

The region of study is defined as the land grid points with latitudes ranging from 0 to 25°N and longitudes from 20°W to 35°E, unless stated otherwise.

The re-forecasts are evaluated against reference data to assess the ability of System 5 to represent the mean MAMJ indices (bias) and reproduce interannual variability (gridpoint correlation coefficient, root mean square error). Indices are computed for each ensemble member separately

(before any averaging). Deterministic skill is evaluated by computing ensemble mean indices. The distribution of the indices for the different ensemble members allows for a probabilistic evaluation of the re-forecast and estimation of the ensemble spread. In this study we use ERA-Interim reanalysis data (Dee et al. 2011) as our main reference. The HWDI and HWPP indices are derived separately for re-forecast and reference data so as to account for first-order biases of System 5 in the Tmin and Tmax distributions.

Bhend et al. (2017) showed using a toy model framework that climate indices skill (based on threshold exceedances of daily values) is limited by the skill in forecasting the seasonal mean climate variable used to compute the index. Applied to this study, this suggests that the heat wave index prediction skill is not expected to exceed the seasonal mean Tmin and Tmax skill.

In this study heat wave indices were computed for the seasonal re-forecasts and real-time forecasts only. For S2S runs, AT, Tmax and Tmin anomalies were considered instead of indices to avoid cutoff effects due to shorter forecast ranges and the minimum duration chosen for a heat wave to be defined. The S2S skill assessment is completed by a focus on a specific case study.

e. Reference datasets and observational uncertainty

Skill assessments in climate predictions are subject to uncertainties related to limited sample size (both in terms of re-forecast length and ensemble size), but also to the data used as a reference (Massonnet et al. 2016). Indeed, in order to cover the entire re-forecast period and geographical extension of the phenomena considered, reanalyses datasets, based on observation data assimilation with general circulation models, are often used as benchmarks. These datasets (such as ERA-Interim used in this study) have proven their capacity of reproducing variability of station-based and instrumental observations (see e.g. Dee et al. (2011)). However, they necessarily bear some similarities to the GCMs used in their design. Since ERA-Interim is used to initialize the

atmospheric component of the System 5 re-forecasts, one could also argue that skill evaluations against ERA-Interim do not provide a fair estimate of re-forecast skill.

In this paper, the impact of observational uncertainty on results was accounted for by additionally verifying re-forecasts against Berkeley Earth Surface Temperature (BEST, Rohde et al. (2013)) daily minimum and maximum temperature data. BEST data is derived using the Kriging method from a large amount of station data. It is available up to 2013, so evaluations are restricted to the 1993-2013 period when using BEST. Over West Africa, Barbier et al. (2017) exhibited some differences in the heat wave characteristics derived from BEST Tmin and Tmax data with those of reanalysis datasets, although limited in the case of ERA-Interim.

Figure 1 compares the climatology over MAMJ 1993-2013 of Tmax and Tmin HWDI using ERA-Interim and BEST over the region of study. For Tmax, both datasets yield similar results in terms of spatial distribution of the HWDI, with a zonal gradient between the coasts of the Gulf of Guinea and northernmost latitudes. Over most of the region of study, ERA-Interim Tmax HWDI is slightly higher than BEST. More differences are found with Tmin HWDI, for which the BEST index appears to be much smoother than for ERA-Interim. This could be a consequence of the Kriging interpolation technique used in BEST, whilst ERA-Interim T2m data is derived from optimal interpolation of screen-level temperature from presumably fewer station observations. Further investigation into these aspects goes beyond the scope of this study. However, these differences emphasize the need for caution when evaluating a forecasting system against a unique source of reference data.

3. Re-forecast quality assessment

As previously mentioned, the assessment of re-forecast quality was done separately for the seasonal forecasting system and S2S experiments. This was motivated by the definition of the heat

wave indices based on both threshold exceedance and length of an event. For seasonal re-forecasts, we evaluate the skill of System 5 in predicting the interannual variability of seasonal indices HWDI and HWPP described in section d. In the S2S framework, the sampling of possible past events (when focusing for instance on predictability of a heat wave during week 3) in the re-forecast dataset is limited by the frequency of start dates. We therefore focus first on Tmin and Tmax prediction skill at a weekly scale, and then on a specific case study.

a. Seasonal re-forecasts

Supplemental Fig. S.1 shows the bias for Tmax and Tmin HWDI with respect to ERA-Interim and BEST indices over the re-forecast period. Despite the computation of heat wave indices with respect to the model Tmax and Tmin climatologies, System 5 yields far less heat wave events than in the reference datasets over most of West Africa. Indeed, the model often fails to maintain heat wave conditions for three or more consecutive days. Differences between evaluations against the two reference datasets are (as expected) most striking for Tmin HWDI, but at a regional scale, conclusions are similar for both references.

The Spearman rank correlation coefficient (defined as the Pearson correlation of the ranks of the forecast and reference heat wave indices over the re-forecast period) is computed to evaluate the models' ability to represent interannual variability of the heat wave indices and shown for Tmax and Tmin HWDI in Fig. 2. This skill measure is less sensitive to outliers and appropriate for discrete values such as indices based on counts as used here (Déqué 2012). We estimate a 95% confidence interval for correlation computed for a t-distribution, accounting for dependence between consecutive forecasts as in Zwiers and Von Storch (1995). Dots indicate grid points with correlation values significantly different from zero according to this test.

Results from Fig. 2 can be summarized as follows: System 5 has higher skill in predicting Tmin HWDI anomalies than for Tmax, and skill is highest along the coast of the Gulf of Guinea, where biases are limited. Significant correlation values are found in very few gridpoints of the Sahel region in the case of Tmax HWDI, whereas more areas exhibit skill for Tmin HWDI. Very similar results are found in the case of HWPP indices (not shown).

AT35 HWDI was computed for ERA-Interim and System 5 data for MAMJ 1993-2014. Figure 3 (a) shows the ERA-Interim mean AT35 HWDI for each grid point of the region over the re-forecast period. Accounting for relative humidity shifts the location of the maximum HWDI southwards with respect to Tmin and Tmax indices. The HWDI is highest in Western Sahel around 15°N. Not all areas see such heat waves occurring due to the use of absolute thresholds instead of quantiles. Corresponding AT35 HWDI values were computed for each member of the System 5 ensemble re-forecast, and Spearman rank correlation with ERA-Interim is shown in Fig. 3 (b). Areas where no AT35 heat wave occurred in ERA-Interim during the re-forecast period are shown in gray. Over areas where such events occurred, the model appears to capture the interannual variability of these heat waves better than those defined with exceedences of Tmax quantiles. This is likely linked to the higher skill of the model in representing daily mean temperatures with respect to daily maximum temperatures over the region at a seasonal time scale (not shown).

At a regional scale, results over the Sahel region (10°N-20°N; 10°W-20°E) confirm that skill is higher for Tmin heat wave occurrences than for Tmax. Levels of skill similar to that of Tmin are found when computing AT35 heat wave indices. Figure 4 shows the distribution of ensemble re-forecasts and the corresponding ERA-Interim value in MAMJ 1993-2014 for all six heat wave indices spatially averaged over the Sahel region. Correlation between the ensemble mean and the ERA-Interim indices are shown in the top left corner of each graph. As could be expected by construction, results for HWDI and HWPP indices are quite similar, as these indices are very

highly correlated (over 0.96) in both reference data and model re-forecasts. Correlation reaches over 0.7 for Tmin heat wave indices and over 0.65 for AT35. This implies that although at a grid point level, the model does not manage to capture interannual variability, it does succeed in reproducing part of the variability of regionally-aggregated indices.

Some peaks in interannual variability are reasonably well re-forecast by System 5, such as 1998 and 2010. These events follow a winter El Niño event in the tropical Pacific. Moron et al. (2016) found using reanalysis datasets and station data that MAMJ heat wave indices over the region were highly correlated with indices characterizing the El Niño-Southern Oscillation phenomenon (ENSO) in the preceding winter season. This link could be explained by a general warming of the tropical upper troposphere following a strong El Niño event, which then propagates back to the surface at sub-seasonal to seasonal time scales, and also to the advection of warm mixed-layer anomalies formed over the tropical North Atlantic by low-level winds (Moron et al. 2016; Oueslati et al. 2017). Since ENSO dominates interannual variability and is a prominent source of seasonal forecasting skill, our results are consistent with these hypotheses, and suggest the existence of conditional skill in our seasonal forecasting system following a strong El Niño event. The pronounced 2015/16 El Niño event therefore triggered the provision of real-time forecasts in MAMJ 2016 in the framework of the French Agence Nationale de Recherche (ANR) ACASIS project (`https://acasis.locean-ipsl.upmc.fr`), which focuses on the physical origins, predictability, variability and evolution of Sahelian heat waves and their impact on human health.

b. Subseasonal re-forecasts

As shown in the previous section, seasonal forecasting skill is limited for MAMJ over the West African region, but higher when considering area-averaged heat wave indices over the Sahel. Seasonal forecasts can provide early indications of above-normal heat wave frequency and duration,

but some studies on the uptake of climate predictions by service providers suggest that the sub-seasonal range is also very relevant for action to alleviate consequences of severe heat on human activities and health (White et al. 2017).

Figure 5 shows the Météo-France S2S system correlation with ERA-Interim for apparent temperature at lead times week 1 (days 4-10), week 2 (days 11-17), week 3 (days 18-24) and week 4 (days 25-31) for 1 April and 15 April start dates of the 22-year re-forecast period. Correlation is computed after removing a linear trend. A substantial drop in correlation is found from week 2 onwards, as expected when moving beyond the deterministic range of predictability, although most gridpoints of the West African region still exhibit significant skill at this lead time. Skill is found mainly over lower latitudes for weeks 3 and 4, although some grid points over Sahel still exhibit significant skill. Similar results are found for Tmin and Tmax (not shown). These results suggest System 5 S2S predictions could be used, although with caution, in an early heat wave warning system. Moreover, windows of opportunity for forecasts at these extended time ranges may be provided by planetary waves and teleconnections. This is an area for future work beyond the preliminary assessment presented here, and could be led using a multi-model framework with the S2S database (Vitart et al. 2016).

4. 2016 real-time forecasts

As shown in the previous section with evaluations over 1993-2014 re-forecasts, the Météo-France forecasting systems exhibit modest skill in forecasting Tmin, Tmax or AT anomalies at sub-seasonal and seasonal time scales, and consequently heat wave occurrences over the West African region during the MAMJ season.

However, signal related to ENSO seems to be captured by System 5 at a seasonal time scale. The 2015/16 El Niño was one of the strongest of the last three decades (L'Heureux et al. 2016),

and presented a unique opportunity to assess in real-time the conditional skill in El Niño years of heat wave forecasts with System 5.

As done over the re-forecast period, we computed MAMJ HWDI and HWPP for Tmin and Tmax in each of the 51 ensemble members initialized end of January 2016. Information was provided on the ACASIS website mid-February to the project partners, in the shape of deterministic forecasts (using the ensemble mean anomalies) and probabilistic forecasts.

When ERA-Interim data was available a few months after the MAMJ season, heat wave statistics were computed and verification carried out with this reference data so as to assess the models' success or failure in forecasting the 2016 season.

For the S2S system, we focused *a posteriori* on a particular event as a case study, which occurred in April 2016 over the Matam area in Senegal.

a. Seasonal forecasts for MAMJ

The forecast and ERA-Interim HWDI anomalies in MAMJ 2016 with respect to the 1993-2014 re-forecast period are shown in Fig. 6. Anomalies for HWPP are shown in supplemental Fig. S.2 and give very similar results. As expected from the strong ENSO signal end of January in the tropical Pacific, the model forecasts at a regional scale a strong positive anomaly in heat wave duration and number of heat waves during MAMJ 2016, both in Tmax and Tmin. For Tmax HWDI, some regions towards the east of the domain are forecast to experience near-normal or even slightly cool conditions. When comparing to indices derived from ERA-Interim, it appears that regional contrasts between areas with more heat wave days and less heat wave days are not reproduced by the model. However, the anomaly over the region was on average higher in Tmin than in Tmax, and this characteristic was reasonably well captured in the forecast.

As done previously for the re-forecast, indices are averaged over the Sahel region and results shown in the orange box-and-whisker plots in Fig. 4. The corresponding MAMJ 2016 ERA-Interim value for each index is shown in blue. Over Sahel, unprecedented values are found for Tmin HWDI and HWPP (Fig. 4 (b,e)), whereas Tmax indices are higher than average but lower than recent seasons such as 2010 and 2013 (Fig. 4 (a,d)). It could be tempting to state that the 2016 Tmax indices forecast by System 5 were spot on, but this disregards the fact that, on average, the Tmax indices are underestimated in the re-forecast, and a straightforward bias correction to compute heat wave index anomalies leads to an overestimation of the anomaly observed with ERA-Interim. In the case of AT35, ERA-Interim indices were comparable to values for 2010 (Fig. 4 (c,f)), while the ensemble forecast distribution for HWDI was clearly shifted towards very high values, with the interquartile range predicting 19 to 24 days in AT35 heat wave conditions on average over the area.

We remind the reader that System 5 forecasts consist in 51 ensemble members, which could increase the range with respect to that of a 15-member re-forecast ensemble. It is therefore difficult to directly compare the boxplots of the ensembles.

b. April 2016 case study

In this context of higher than average Tmax heat wave occurrences and record-breaking Tmin heat wave indices, we focus on a particular event as a case study for S2S prediction with the CNRM-CM model. April 2016 was the warmest April month on record since 1910 for the African continent, according to NOAA (NOAA NCEI 2016).

The Matam region in Senegal experienced a prolonged period of very warm temperatures in mid- and late April, reaching locally above 44°C for eleven consecutive days at the Matam SYNOP station from 12 to 22 April (period called CS for ”case study”). As depicted in the ERA-Interim

weekly AT in Fig. 7, the warm anomalies were not restricted to the Matam area delimited by the purple box and defined in this paper between 14 and 17°N and 14 and 11°W (a zoom is provided in supplemental Fig. S.3). These weekly anomalies also show that cold AT anomalies preceded the event during the first week of April. Figure 7 also shows the evolution with the initialization dates of weekly anomalies forecast by the S2S system for three target weeks around the case study period.

Based on skill evaluations shown for AT in Fig. 5, we expect the model to capture anomalies for the first week after the initialization date, but not necessarily beyond. Results for the case study period shown in Fig. 7 show that this is the case for April 2016. Forecasts for the week from 11 to 17 April (second row) are very similar for all start dates before 7 April, and if they suggest the possibility of warm anomalies over Sahel, they overestimate the spatial extent of these anomalies and systematically forecast cold anomalies in the Northwest corner of the domain, possibly due to the model forecasting cold sea-surface temperature anomalies in the Atlantic. The 24 March and 31 March initial conditions were cool in terms of AT over this area (not shown), and both forecasts seem to have persisted these cool anomalies too long before transitioning to warmer-than-average ATs during the week from 18 to 24 April. Results for Tmin and Tmax bear many similarities with AT (see supplemental Fig. S4).

To further assess the model ability to forecast the heat wave experienced in Matam, we average Tmax, Tmin and AT values over the Matam box. Figure 8 shows distributions over 1993-2014 of anomalies with respect to climatology for 11-day averages over the CS dates over the Matam region in ERA-Interim (dashed black line) and the S2S re-forecasts initialized on 15 March and 1 April (dashed blue lines). For Tmax and AT, anomaly distributions differ very little with re-forecast lead time, whereas they are sharper for 1 April initializations in the case of Tmin. For the 2016 case study, the ERA-Interim 11-day average is marked by a black triangle. For all three variables,

the event ranks as the warmest with respect to the reference period. 51-member ensemble forecast distributions are shown in colored full lines for the three start dates that cover the complete CS period. For all three variables, a shift in the distribution with respect to the re-forecast values is found for all start dates, likely due to warm background conditions related to ENSO. The 7 April start, five days before the beginning of the event, correctly anticipates the very high anomalies for CS. However, a shift in the distributions and warm tails beyond the values of the 1993-2014 re-forecasts (which include warm years in the recent period such as 2010) does confirm that warm to very warm events can be partly anticipated at these longer lead times.

The Matam box used in this study is quite small with respect to the horizontal resolution of the coupled system. Results for weekly anomalies on the larger domain shown in Fig. 7 imply that similar conclusions would be drawn by extending the study to a larger area of Sahel concerned by this given heat wave. The sensitivity of results to the choice of the case study period was tested by restricting the case study dates to 12-17 April. Results in terms of Tmin and Tmax anomaly distributions are shown in supplemental Fig. S.4, and are consistent with those of the longer case study period.

Figure 9 shows mean sea-level pressure anomalies over the CS period for ERA-Interim and the three start dates of the 2016 S2S forecasts covering the event. Purple boxes highlight the area used for computing weather types discussed in section 5. During the CS, high pressure anomalies over most of West Africa and the Mediterranean Sea were observed, as well as a low-pressure system centered over the Azores. In the System 5 S2S forecasts, as expected, the ensemble mean sea-level pressure anomalies are close to zero for the longer lead times due to increased noise. Even for the 7 April initialization, the model misses part of the signal over the West African domain, underestimates the high pressure anomalies over the Mediterranean and misplaces the

low pressure anomalies over the Atlantic and Western Europe. This suggests that the system was unable to capture the large scale circulation anomalies at play during this given heat wave event.

Further investigation into links between circulation and temperature anomalies over the region and their representation by System 5 is described in what follows. We consider the use of a weather type approach (Robertson and Ghil 1999; Muñoz et al. 2015) to examine links between the circulation forecast by the model and daily temperature anomalies, and further characterize the 2016 season and case study with respect to both the 1993-2014 reference period, and the ERA-Interim reference data for 2016.

5. Weather type approach

a. Methodology and results over 1993-2014

In the scope of the ACASIS project, weather types (WTs Moron et al. 2008) were computed to characterize the intra-seasonal variability of daily minimum and maximum temperature over West Africa and better understand relationships between near-surface temperature and synoptic weather patterns (Vincent Moron and colleagues, personal communication). The box used to define the WTs covers latitudes from 0 to 30°N and longitudes from 40 W to 40°E and is represented in purple in Fig. 9. The WTs are derived by first computing empirical orthogonal functions of standardized 925 hPa winds and mean sea level pressure from the NCEP reanalysis over the 1967-2014 period, and keeping the first nine EOFs so as to account for 50% of the variance. The iterative k-means clustering approach is used to partition the principal component space into k different WTs, minimizing the sum of within-cluster variance. The number of WTs $k = 7$ is determined by the classifiability index as in Michelangeli et al. (1995). Figure 10 (a) shows the sea-level pressure anomaly composites for each of the seven WTs found, restricted to the MAMJ 1993-2014 period.

WT3 and WT7 SLP anomaly composites present a strong negative pattern correlation. Other WT composites such as WT2 and WT4 exhibit centers of action located mainly outside the domain used to compute the WTs, which are reminiscent of Rossby wave propagation.

We then attributed a WT to each day of the MAMJ 1993-2014 seasons in ERA-Interim (and each member in the System 5 re-forecasts) by projecting the SLP and wind anomalies onto each WT composite, and selecting the closest WT in terms of Euclidian distance. A clear impact on temperature anomalies over West Africa and Sahel is found. Figure 10 (b) and (c) shows the composites of ERA-Interim and System 5 Tmax anomalies for each WT. The SLP-Tmax relationship is clearly non linear, since WT3 and WT7 Tmax composites do not exhibit the same negative pattern correlation as for SLP. Some differences in the Tmax response are found between the model and reference data, most striking in WT1 for which the model extends the cold anomalies too far West, and WT6 and WT7 for which both the amplitude and spatial location of the Tmax anomalies are different from ERA-Interim in the forecasts. Yet, the model does manage to properly reproduce Tmax anomaly composites for each WT. Very similar results are found for Tmin (not shown).

This WT approach is used to further characterize the forecasts for 2016, both the MAMJ 2016 seasonal forecast issued in February, and the S2S forecasts of the CS period.

b. MAMJ 2016

Figure 11 (a) shows the WT frequency histograms for the MAMJ season for ERA-Interim (shades of blue) and System 5 (shades of red). Climatology over 1993-2014 is shown in pale colors, whereas the MAMJ 2016 season is shown in bolder colors. System 5 has reasonable climatological values for WT frequencies, most often differing from ERA-Interim by less than 5 %. Anomalies for 2016 (departures from respective climatologies) show that WT3 was far more frequent than typically observed, while the number of days in WT7 was one third of the climato-

logical average. On the other hand, System 5 strongly favored WT7, reaching a frequency over 30% of the MAMJ season. By referring to the Tmax temperature composites for WT3, this could explain the cooler-than-normal conditions in ERA-Interim for 2016 over the northwestern corner of the region studied as well as around 10°N despite the strong ENSO forcing at play. System 5 failed to capture this, and projected warm conditions almost everywhere over the region, consistent with composite patterns for WT7.

One possible use of WTs would be in a statistical-dynamical forecasting approach, using System 5 forecasted WT frequency anomalies to construct temperature anomaly forecasts instead of using direct model outputs for temperature. However, in terms of predictability of WT frequency at a seasonal time scale, no significant skill is found, save for WT6 (not shown). No additional skill can therefore be expected using such an approach. WTs are therefore used here as a way of interpreting the circulation types preferred by the system for the 2016 season considered.

c. Case study

WTs were also computed for the S2S forecasts of the 12-22 April case study dates. WT frequencies for CS dates over the 1993-2014 reference period and for 2016 are shown in Fig. 11 (b). The corresponding statistics are found in supplemental Table S.1. Note that the ERA-Interim and System 5 climatologies differ from Fig. 11 (a) as we focus here on a 11-day subset of the 122-day MAMJ season. For System 5, small differences between the 15 March and 1 April climatologies result from both sampling errors and forecast lead time. The model underestimates the occurrence of WT6 and overestimates the occurrences of WT1 and WT7 over the re-forecast period, but other WT frequencies are rather well reproduced.

For the 2016 CS anomalies, we use Klee diagrams (Muñoz et al. 2016) to depict the sequences of WTs predicted by the ensembles according to the forecast initialization dates, as well as in

ERA-Interim (Fig. 12). WTs 2, 4 and 7 are not represented in the ERA-Interim values for the 2016 CS (as shown in Fig. 11 (b) and 12 (a)) as the period spans only 11 days, and WT6 was observed for 4 consecutive days from 14 April until 17 April. The S2S ensembles do predict non-zero frequencies of occurrence for each WT over the CS period, even in the case of the latest initialization on 7 April. In contrast to the MAMJ seasonal forecast, the system predicts (correctly) for each start date a lower frequency of WT7 with respect to climatology. However, the WT4 frequency is strikingly overestimated for all start dates except 7 April. WT5 frequencies are off even in the latest initialization, likely due to the fact that this WT occurs mainly at the end of the CS period, beyond the typical deterministic range. The Klee diagram for 7 April initialization (Fig. 12 (b)) provides additional insight into the medium-range predictability of the WTs associated with the case study. In particular, only one ensemble member manages to capture the WT3-WT5 sequence on 12-13 April, whereas most predict WT6 for 2 days or more between 14 and 17 April, suggesting that the anomalies were then more pronounced and persistent (and therefore, closer to an actual weather regime pattern) than conditions on the first two days of the CS.

6. Main conclusions

In this study, we investigated the seasonal and sub-seasonal skill of Météo-France System 5 in forecasting spring heat waves over West Africa and Sahel. For the seasonal range, heat wave duration and number of days per March to June season were computed for minimum and maximum daily near-surface temperature quantile threshold exceedences and compared to statistics obtained with the ERA-Interim and BEST reference datasets. We also studied the modulation by relative humidity by computing a heat wave index based on mean daily apparent temperature threshold exceedence. The coupled system exhibits limited gridpoint correlation over the re-forecast period, consistent with the skill of mean seasonal values for the physical variables considered. However,

the interannual variability of spatially averaged Sahelian heat wave indices are found to be quite well represented at the seasonal time scale. Skill is generally higher for minimum temperature and apparent temperature indices than for maximum temperature. For the S2S range, we focused on apparent temperature anomalies at a weekly time scale, and found evidence of some modest skill beyond the deterministic range. Results found in terms of predictability at the sub-seasonal and seasonal scales are consistent with previous works evaluating the skill of state-of-the-art systems in representing temperature extremes (Pepler et al. 2015; Hudson et al. 2011).

The 2016 season presented a unique opportunity to test the system in a real-time framework, with a strong El Niño forcing in the preceding winter. Consistent with this strong forcing, the forecasting system predicted high probabilities for heat wave indices exceeding the top quintile of the 1993-2014 climatology over most areas of West Africa. This concurs with conclusions from Moron et al. (2016) based on observational datasets, which linked heat wave occurrence over the region with ENSO. As found for the re-forecast evaluations, the model succeeded better in forecasting an area-averaged anomaly than discriminating the geographical locations of the heat waves within the region of interest. A case study evaluation for the Matam area in Senegal for April 2016 illustrated the model capacity in anticipating a particularly warm event.

A weather type approach was used as a means of interpretation of model mismatch with observations in terms of temperature anomalies. Weather types were computed based on sea-level pressure and near-surface wind anomalies over the region. These projections shed light onto the tendencies in the coupled system simulations to favor specific synoptic weather patterns, therefore translating into incorrect location of surface temperature anomalies, although the composite surface temperature response to each weather type in the model is satisfactory.

Results found in this study are based on a re-forecast period of 22 years, using 15-member ensembles, and decadal variability in skill cannot be excluded. We stress that the 2016 season

forecasts should be interpreted with the modest skill of the system in mind, as well as the specific context of a strong ocean forcing in the tropical Pacific. Conversely, 2017 forecasts were issued in the scope of the project but no clear signal was found over the area. To further assess the conditional skill of a forecasting system, much longer re-forecast periods and, ideally, larger ensemble sizes are needed, as well as idealized studies to assess the forecasting system's sensitivity to specific initial conditions. This leaves room for many future improvements in the understanding and forecasting of such events at these longer time scales.

Origins of model limitations over the region are currently being investigated. The coupled system exhibits very poor skill in representing key variables known to affect minimum and maximum surface temperature over the area, especially in pre-monsoon seasons, such as cloud cover and heat fluxes which play a key role in modulating larger-scale variability (Oueslati et al. 2017). At Météo-France, a new seasonal forecasting system with enhanced physical formulation in the atmospheric component of the GCM for clouds, convection and radiation has been developed and will be operational in the framework of Copernicus Climate Change Services (C3S). Future work includes assessing the skill in this new system compared to System 5.

Although many (anticipated) limits in S2S and seasonal skill are found with the current Météo-France System 5, this study examines the possibility of using coupled systems for seasonal and sub-seasonal prediction of extreme heat over the West African region. Future improvements may originate from the use of multi-model combination and an improved calibration of model outputs (beyond straightforward bias removal). This is a goal that could be reached in years to come with international initiatives such as Copernicus C3S and the S2S project database. Then, uptake of reliable forecasts as input for early-warning health risk reduction could enable an improved anticipation of heat-related risks for human health in a context of increased vulnerability related to climate change.

Acknowledgments. The research leading to these results has received funding from the French Agence Nationale de Recherche project ACASIS (Alerte aux Canicules en Afrique et Ses Impacts sur la Santé), grant ANR-13-SENV-0007-01. Matam station SYNOP data was retrieved from the ACASIS project platform. We are indebted to developers of R libraries s2dverification and SpecsVerification used for some of the skill analyses presented in this paper. Weather types cluster centroids were provided courtesy of Vincent Moron (CEREGE, Université Aix-Marseille), whom we wish to thank for sharing his data and ideas. The comments from two reviewers were gratefully received and helped substantially improve the paper.

References

Aguilar, E., and Coauthors, 2009: Changes in temperature and precipitation extremes in western central Africa, Guinea Conakry, and Zimbabwe, 1955–2006. *Journal of Geophysical Research*, **114**, D02 115, doi:doi:10.1029/2008JD011010.

Ardilouze, C., L. Batté, and M. Déqué, 2017: Subseasonal-to-seasonal (S2S) forecasts with CNRM-CM: a case study on the July 2015 West-European heat wave. *Advances in Science and Research*, **14**, 115–121, doi:10.5194/asr-14-115-2017.

Barbier, J., F. Guichard, D. Bouniol, F. Couvreux, and R. Roehrig, 2017: Detection of intraseasonal large-scale heat waves: characteristics and historical trends during the Sahelian Spring. *Journal of Climate, in press*, doi:10.1175/JCLI-D-17-0244.1.

Batté, L., and M. Déqué, 2011: Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: skill of the ENSEMBLES project multimodel ensemble forecasts. *Tellus*, **63A**, 283–299, doi:10.1111/j.1600-0870.2010.00493.x.

595 Batté, L., and M. Déqué, 2016: Randomly correcting model errors in the ARPEGE-Climate v6.1
596 component of CNRM-CM: applications for seasonal forecasts. *Geoscientific Model Develop-*
597 *ment*, **9**, 2055–2076, doi:10.5194/gmd-9-2055-2016.

598 Bhend, J., I. Mahlstein, and M. A. Liniger, 2017: Predictive skill of climate indices compared to
599 mean quantities in seasonal forecasts. *Quarterly Journal of the Royal Meteorological Society*,
600 **143**, 184–194, doi:10.1002/qj.2908.

601 Brunet, G., and Coauthors, 2010: Collaboration of the weather and climate communities to ad-
602 vance subseasonal-to-seasonal prediction. *Bulletin of the American Meteorological Society*, **91**,
603 1397–1406, doi:10.1175/2010BAMS3013.1.

604 Cariolle, D., and M. Déqué, 1986: Southern hemisphere medium-scale waves and total ozone
605 disturbances in a spectral general circulation model. *Journal of Geophysical Research: Atmo-*
606 *spheres*, **90**, 10 825–10 846, doi:10.1029/JD091iD10p10825.

607 Ceccherini, G., S. Russo, I. Ameztoy, A. F. Marchese, and C. Carmona-Moreno, 2017: Heat waves
608 in Africa 1981–2015, observations and reanalysis. *Natural Hazards and Earth System Sciences*,
609 **17**, 115–125, doi:10.5194/nhess-17-115-2017.

610 Conway, D., 2011: Adapting climate research for development in Africa. *Wiley Interdisciplinary*
611 *Reviews Climate Change*, **2**, 428–450, doi:10.1002/wcc.115.

612 Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance
613 of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**,
614 553–597, doi:10.1002/qj.828.

615 Déqué, M., 2012: Deterministic forecasts of continuous variables. *Forecast Verification, A Prac-*
616 *titioner's Guide in Atmospheric Science, Second Edition*, I. Jolliffe, and D. Stephenson, Eds.,
617 John Wiley & Sons Ltd, 77–94.

618 Déqué, M., S. Calmanti, O. B. Christensen, A. Dell Aquila, C. F. Maule, A. Haensler, G. Nikulin,
619 and C. Teichmann, 2017: A multi-model climate response over tropical Africa at $+2^{\circ}$ C. *Cli-*
620 *mate Services*, doi:10.1016/j.cliser.2016.06.002.

621 Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. Pintó Biescas, and L. R. L. Rodrigues, 2013:
622 Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary*
623 *Reviews Climate Change*, **4**, 245–268, doi:10.1002/wcc.217.

624 Dosio, A., 2017: Projection of temperature and heat waves for Africa with an ensem-
625 ble of CORDEX Regional Climate Models. *Climate Dynamics*, **49**, 493–519, doi:10.1007/
626 s00382-016-3355-5.

627 Ferry, N., L. Parent, G. Garric, B. Barnier, N. C. Jourdain, and the Merca-
628 tor Ocean team, 2010: Mercator global eddy permitting ocean reanalysis GLO-
629 RYS1V1: description and results. *Mercator Ocean Quarterly Newsletter*, **36**, 15–
630 27, URL https://www.mercator-ocean.fr/sciences-publications/mercator-ocean-journal/
631 newsletter-36-data-assimilation-and-its-application-to-ocean-reanalyses/.

632 Fischer, E., and C. Schär, 2010: Consistent geographical patterns of changes in high-impact euro-
633 pean heatwaves. *Nature Geoscience*, **3**, 398–403, doi:10.1038/NGEO866.

634 Fontaine, B., S. Janicot, and P.-A. Monerie, 2013: Recent changes in air temperature, heat waves
635 occurrences, and atmospheric circulation in Northern Africa. *Journal of Geophysical Research:*
636 *Atmospheres*, **118**, 8536–8552, doi:10.1002/jgrd.50667.

Gounou, A., F. Guichard, and F. Couvreux, 2012: Observations of diurnal cycles over a West African meridional transect: Pre-monsoon and full-monsoon seasons. *Boundary-Layer Meteorology*, **144**, 329–357, doi:10.1007/s10546-012-9723-8.

Grotjahn, R., and Coauthors, 2016: North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, **46**, 1151–1184, doi:10.1007/s00382-015-2638-6.

Hamilton, E., R. Eade, R. J. Graham, A. A. Scaife, D. M. Smith, A. Maidens, and C. MacLachlan, 2012: Forecasting the number of extreme daily events on seasonal timescales. *Journal of Geophysical Research*, **117**, D03 114, doi:10.1029/2011JD016541.

Hudson, D., O. Alves, H. H. Hendon, and A. G. Marshall, 2011: Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. *Quarterly Journal of the Royal Meteorological Society*, **137**, 673–689, doi:10.1002/qj.769.

Hudson, D., A. G. Marshall, O. Alves, G. Young, D. Jones, and A. Watkins, 2016: Forewarned is forearmed: extended-range forecast guidance of recent extreme heat events in Australia. *Weather and Forecasting*, **31**, 697–711, doi:10.1175/WAF-D-15-0079.1.

Lazenby, M. J., W. A. Landman, R. M. Garland, and D. G. DeWitt, 2014: Seasonal temperature prediction skill over Southern Africa and human health. *Meteorological Applications*, **21**, 963–974, doi:10.1002/met.1449.

Lee, D., and T. Brenner, 2015: Perceived temperature in the course of climate change: an analysis of global heat index from 1979 to 2013. *Earth System Science Data*, **7**, 193–202, doi:10.5194/essd-7-193-2015.

L'Heureux, M., and Coauthors, 2016: Observing and predicting the 2015-16 El Niño. *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-16-0009.1.

Lott, F., and L. Guez, 2013: A stochastic parameterization of the gravity waves due to convection and its impact on the equatorial stratosphere. *Journal of Geophysical Research: Atmosphere*, **118**, 88978909, doi:10.1002/jgrd.50705.

Lowe, R., M. García-Díez, J. Ballester, J. Creswick, J.-M. Robine, F. R. Herrmann, and X. Rodó, 2016: Evaluation of an early-warning system for heat wave-related mortality in Europe: implications for sub-seasonal to seasonal forecasting and climate services. *International Journal of Environmental Research and Public Health*, **13**, 206, doi:10.3390/ijerph13020206.

Ly, M., S. B. Traore, A. Alhassane, and B. Sarr, 2013: Evolution of some observed climate extremes in the West African Sahel. *Weather and Climate Extremes*, **1**, 19–25, doi:10.1016/j.wace.2013.07.005.

Madec, G., 2008: NEMO ocean engine. Note du Pôle de modélisation No 27, Institut Pierre-Simon Laplace (IPSL), France. ISSN No 1288-1619.

Mahlstein, I., C. Spirig, M. A. Liniger, and C. Appenzeller, 2015: Estimating daily climatologies for climate indices derived from climate model data and observations. *Journal of Geophysical Research: Atmospheres*, **120**, 2808–2818, doi:10.1002/2014JD022327.

Masson, V., and Coauthors, 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geoscientific Model Development*, **6**, 929–960, doi:10.5194/gmd-6-929-2013.

Massonnet, F., O. Bellprat, V. Guemas, and F. Doblas-Reyes, 2016: Using climate models to estimate the quality of global observational data sets. *Science*, **354 (6311)**, 452–455, doi:10.1126/science.aaf6369.

Michelangeli, P.-A., R. Vautard, and B. Legras, 1995: Weather Regimes: Recurrence and Quasi Stationarity. *Journal of the Atmospheric Sciences*, **52 (8)**, 1237–1256.

Mora, C., and Coauthors, 2017: Global risk of deadly heat. *Nature Climate Change*, doi:10.1038/NCLIMATE3322.

Moron, V., B. Oueslati, B. Pohl, S. Rome, and S. Janicot, 2016: Trends of mean temperatures and warm extremes in northern tropical Africa (1961-2014) from observed and PPCA-reconstructed time series. *Journal of Geophysical Research: Atmospheres*, **121**, 5298–5319, doi:10.1002/2015JD024303.

Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye, 2008: Weather types and rainfall over Senegal. Part I: observational analysis. *Journal of Climate*, **21**, 266–287, doi:10.1175/2007JCLI1601.1.

Muñoz, A. G., L. Goddard, S. J. Mason, and A. W. Robertson, 2016: Cross-time scale interactions and rainfall extreme events in Southeastern South America for the Austral summer. Part II: predictive skill. *Journal of Climate*, **29**, 5915–5934, doi:10.1175/JCLI-D-15-0699.1.

Muñoz, A. G., L. Goddard, A. W. Robertson, Y. Kushnir, and W. Baethgen, 2015: Cross-time scale interactions and rainfall extreme events in Southeastern South America for the Austral summer. Part I: potential predictors. *Journal of Climate*, **28**, 7894–7913, doi:10.1175/JCLI-D-14-00693.1.

New, M., and Coauthors, 2006: Evidence of trends in daily climate extremes over southern and west Africa. *Journal of Geophysical Research*, **111**, D14 102, doi:10.1029/2005JD006289.

NOAA NCEI, 2016: State of the Climate: Global Climate Report for April 2016. NOAA National Centers for Environmental Information, URL https://www.ncdc.noaa.gov/sotc/global/201604, Published online May 2016, retrieved on June 7, 2017.

Oueslati, B., B. Pohl, V. Moron, S. Rome, and S. Janicot, 2017: Characterization of heat waves in the Sahel and associated physical mechanisms. *Journal of Climate*, **30**, 3095–3115, doi: 10.1175/JCLI-D-16-0432.1.

Pepler, A. S., L. B. Díaz, C. Prodhomme, F. J. Doblas-Reyes, and A. Kumar, 2015: The ability of a multi-model seasonal forecasting ensemble to forecast the frequency of warm, cold and wet extremes. *Weather and Climate Extremes*, **9**, 68–77, doi:10.1016/j.wace.2015.06.005.

Perkins, S. E., 2015: A review of the scientific understanding of heatwaves – their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, **164–165**, 242–267, doi:10.1016/j.atmosres.2015.05.014.

Philippon, N., F. J. Doblas-Reyes, and P. M. Ruti, 2010: Skill, reproducibility and potential predictability of the West African monsoon in coupled GCMs. *Climate Dynamics*, **35**, 53–74, doi:10.1007/s00382-010-0856-5.

Robertson, A. W., and M. Ghil, 1999: Large-scale weather regimes and local climate over the western United States. *Journal of Climate*, **12**, 1796–1813, doi:10.1175/1520-0442(1999)012⟨1796:LSWRAL⟩2.0.CO;2.

719 Rodríguez-Fonseca, B., and Coauthors, 2015: Variability and predictability of West African
720 droughts: a review on the role of sea surface temperature anomalies. *Journal of Climate*, **28**,
721 4034–4060, doi:10.1175/JCLI-D-14-00130.1.

722 Rohde, R., and Coauthors, 2013: A new estimate of the average Earth surface land temperature
723 spanning 1753 to 2011. *Geoinfor. Geostat: An Overview*, **1**, doi:10.4172/2327-4581.1000101.

724 Russo, S., and Coauthors, 2014: Magnitude of extreme heat waves in present climate and their
725 projection in a warming world. *Journal of Geophysical Research: Atmospheres*, **119**, 12 500–
726 12 512, doi:10.1002/2014JD022098.

727 Sherwood, S. C., and M. Huber, 2010: An adaptability limit to climate change due to heat
728 stress. *Proceedings of the National Academy of Science*, **107 (21)**, 9552–9555, doi:10.1073/
729 pnas.0913352107.

730 Tompkins, A. M., and L. Feudale, 2010: Seasonal ensemble predictions of West African Monsoon
731 precipitation in the ECMWF system 3 with a focus on the AMMA Special Observing Period in
732 2006. *Weather and Forecasting*, **25**, 768–788, doi:10.1175/2009WAF2222236.1.

733 Valcke, S., 2006: OASIS3 user guide. Tech. Rep. TR/CMGC/06/73, CERFACS, Toulouse, France.

734 Vellinga, M., A. Arribas, and R. Graham, 2013: Seasonal forecasts for regional onset of the West
735 African monsoon. *Climate Dynamics*, **40**, 3047–3070, doi:10.1007/s00382-012-1520-z.

736 Vitart, F., and Coauthors, 2007: Dynamically-based seasonal forecasts of Atlantic tropical storm
737 activity issued in June by EUROSIP. *Geophysical Research Letters*, **34 (16)**, doi:10.1029/
738 2007GL030740, URL http://dx.doi.org/10.1029/2007GL030740.

739 Vitart, F., and Coauthors, 2016: The Sub-seasonal to Seasonal Prediction (S2S) project database.
740 *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-16-0017.1.

741 Voldoire, A., and Coauthors, 2013: The CNRM-CM5.1 global climate model: Description and
742 basic evaluation. *Climate Dynamics*, doi:10.1007/s00382-011-1259-y.

743 Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *Journal*
744 *of the Royal Society Interface*, **11**, doi:10.1098/rsif.2013.1162.

745 White, C. J., and Coauthors, 2017: Review - Potential applications of subseasonal-to-seasonal
746 (S2S) predictions. *Meteorological Applications*, doi:10.1002/met.1654.

747 Zwiers, F., and H. Von Storch, 1995: Taking serial correlation into account in tests of the mean.
748 *Journal of Climate*, **8**, 336–351, doi:10.1175/1520-0442(1995)008⟨0336:TSCIAI⟩2.0.CO;2.

LIST OF TABLES

Table 1. Characteristics of the seasonal and sub-seasonal forecasts with Météo-France System 5 discussed in this paper. 37

TABLE 1. Characteristics of the seasonal and sub-seasonal forecasts with Météo-France System 5 discussed in this paper.

		Re-forecasts (1993-2014)	**Real time forecasts**
Common features	Ensemble size	15	51
	Initial conditions	Atmosphere/land: ERA-Interim	Atmosphere/land: ECMWF analysis
		Ocean/sea ice: Mercator Ocean PSY2G2V3	Mercator Ocean upscaled op. analysis
Seasonal	Frequency	1st of each month	Two Wednesdays
	Forecast length	7 months	7 months
	Ensemble generation	stochastic dynamics	stochastic dynamics + 1 week lag
S2S	Frequency	1st and 15th of each month	Each Thursday
	Forecast length	60 days	32 days
	Ensemble generation	stochastic dynamics	stochastic dynamics

LIST OF FIGURES

Fig. 1. Climatologies of MAMJ Tmax (a,c) and Tmin (b,d) HWDI computed with ERA-Interim data for 1993-2014 (a,b) and BEST data for 1993-2013 (c,d). The reference period for BEST is restricted to 1993-2013 due to data availability. 40

Fig. 2. Spearman rank correlation of MAMJ System 5 re-forecasts for Tmax (a,c) and Tmin (b,d) HWDI with respect to indices derived from ERA-Interim (a,b) and BEST (c,d) daily data. Evaluation against BEST is for 1993-2013. Dots indicate correlation values significant at a 95% level as estimated by a t-distribution (see text). 41

Fig. 3. (a) Climatology of MAMJ AT35 HWDI (see text) for ERA-Interim over the 1993-2014 period. (b) Spearman rank correlation of MAMJ System 5 re-forecasts for AT35 HWDI with respect to the index derived from ERA-Interim over 1993-2014. Dots indicate correlation values significant at a 95% level as estimated by a t-distribution (see text). Gray areas show regions where no AT35 heat waves are found with ERA-Interim data. 42

Fig. 4. Box and whisker plots of the ensemble re-forecasts of Tmax HWDI (a) Tmin HWDI (b) AT35 HWDI (c) and respective HWPP indices (d-f) for MAMJ 1993 to 2014 (in red). For each plot, boxes show the interquartile range and whiskers the full range of the 15-member System 5 ensemble (up to 1.5σ). Outliers are shown with circles. The blue line shows indices derived from ERA-Interim, and the Pearson time correlation with the ensemble mean re-forecast index over the 1993-2014 period is shown in the top left corner of each plot. The orange box and whiskers is the 51-member MAMJ 2016 forecast, and a blue dot shows the corresponding ERA-Interim index. 43

Fig. 5. (a-d) Weekly mean apparent temperature correlation against ERA-Interim over the re-forecast period according to lead time (weeks 1 to 4, respectively) for 1 April and 15 April 1993-2014 start dates in the Météo-France S2S contribution based on System 5. Dots indicate correlation values significant at a 95% level as estimated by a t-distribution (see text). 44

Fig. 6. Real-time System 5 forecast anomalies for Tmax and Tmin HWDI for MAMJ 2016 initialized end of January, with respect to the re-forecast period (a,c) and verification using ERA-Interim daily data (b,d); probabilistic forecast based on counts of System 5 ensemble members exceeding the highest quintile of the re-forecast for MAMJ 2016 initialized end of January (e,g) and actual observed quintile of ERA-Interim minimum and maximum temperature HWDI (f,h). 45

Fig. 7. Mean daily apparent temperature anomalies over West Africa for target weeks 4-10 April, 11-17 April and 18-24 April 2016 before and during the case study period (from top to bottom along the y-axis) in the S2S forecasts initialized on 17 March, 24 March 31 March and 7 April 2016 (from left to right along the x-axis), and corresponding ERA-Interim anomalies (right column). Anomalies are computed with respect to model re-forecast climatologies for S2S forecasts, and ERA-Interim 1993-2014 climatology for ERA-Interim. 46

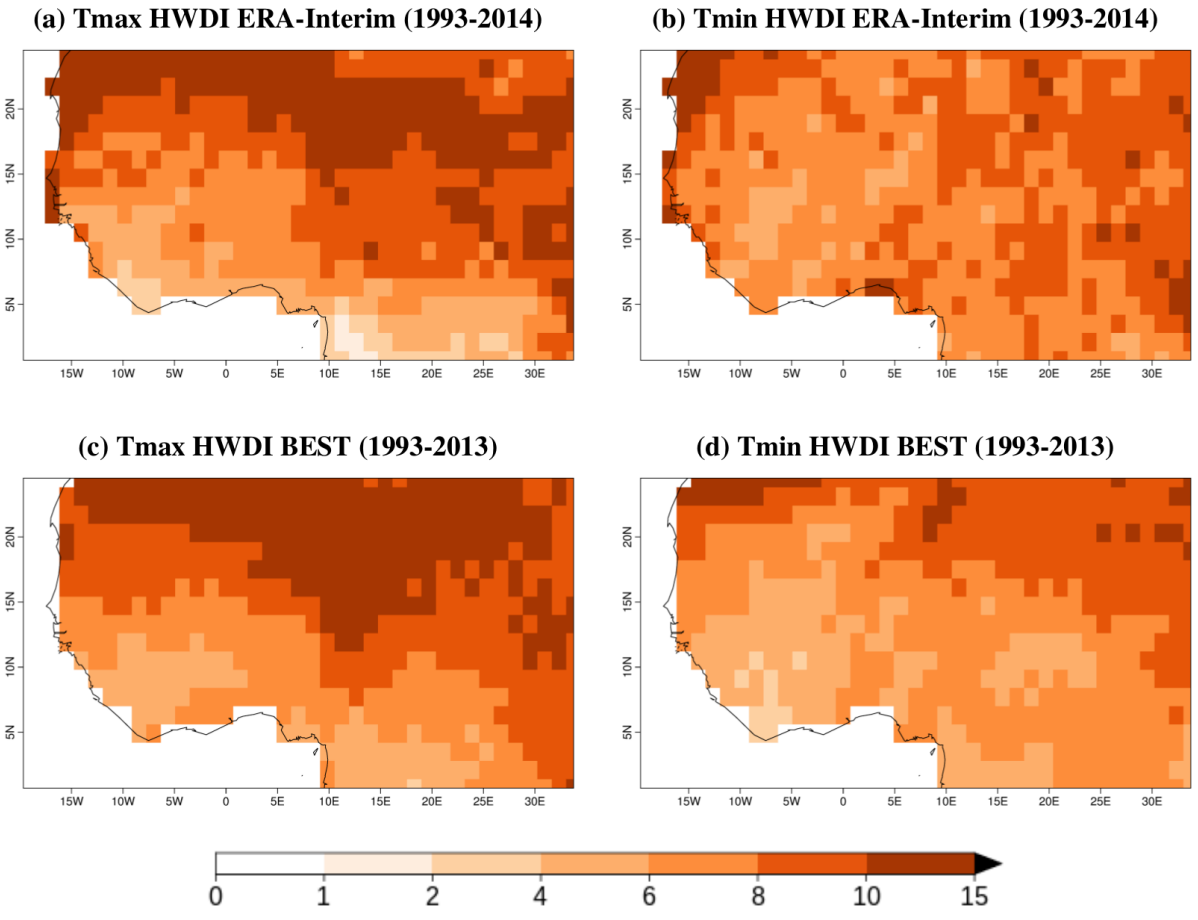
Fig. 8. Tmax (a) Tmin (b) and daily mean AT (c) anomaly distributions over the Matam region for the case study dates (12-22 April) for ERA-Interim 1993-2014, March 15 and April 1 re-forecasts, and three start dates preceding the 2016 case study event. Anomalies are computed for the 2016 S2S forecasts with respect to a weighted linear combination of the climatologies of the two closest re-forecast start dates for the CS dates, depending on the forecast initialization date. 47

Fig. 9. Mean sea level pressure anomalies with respect to 1993-2014 for 12-22 April 2016 for ERA-Interim (a) and different S2S start dates (b-d). 48

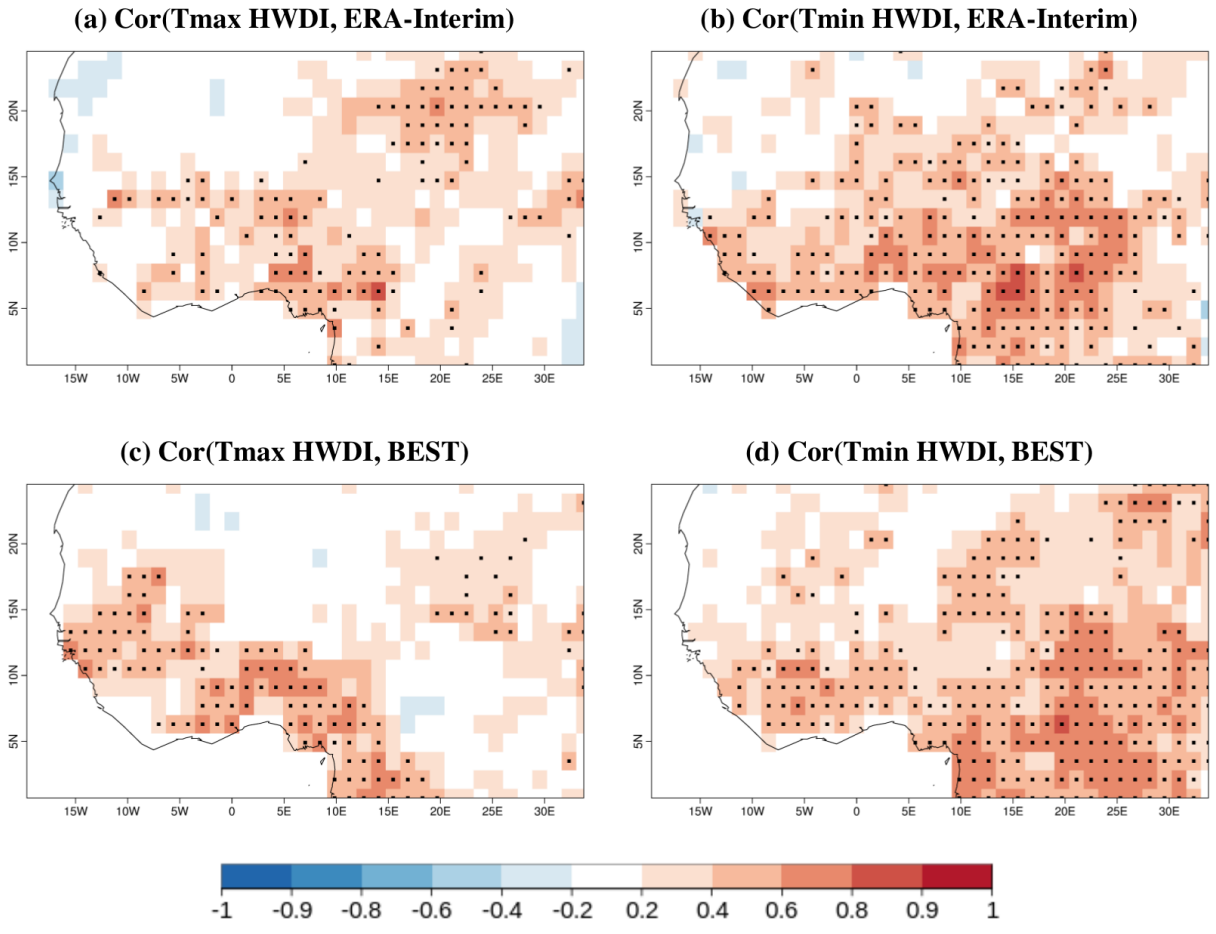
Fig. 10. Column (a): sea-level pressure (SLP) composite in NCEP data for each weather type (WT) 1 to 7 obtained over the West African - Sahel region. Columns (b-c): maximum near-surface temperature composites for each WT for MAMJ 1993-2014 in ERA-Interim (b) and System 5 re-forecasts initialized end of January (c). 49

Fig. 11. WT frequency in re-forecast runs and for 2016 (a) in MAMJ (System 5 re-forecasts initialized end of January) and (b) for the 12-22 April 2016 case study (S2S runs based on System 5). 50

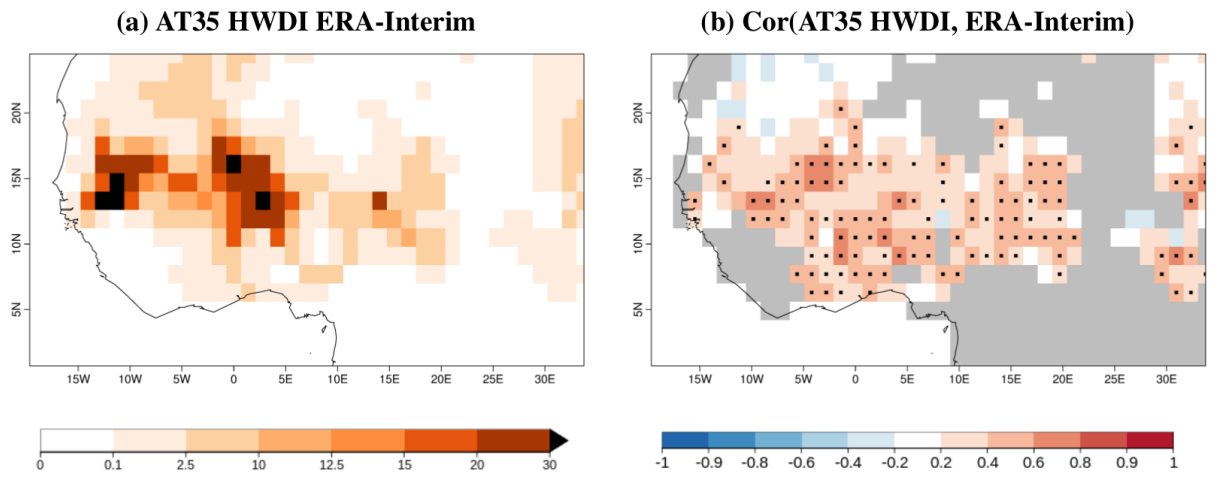
Fig. 12. Klee diagrams of the WTs for the 12-22 April 2016 case study in (a) ERA-Interim and (b-d) S2S ensemble predictions initialized on three consecutive start dates. 51



809 FIG. 1. Climatologies of MAMJ Tmax (a,c) and Tmin (b,d) HWDI computed with ERA-Interim data for
810 1993-2014 (a,b) and BEST data for 1993-2013 (c,d). The reference period for BEST is restricted to 1993-2013
811 due to data availability.



812 FIG. 2. Spearman rank correlation of MAMJ System 5 re-forecasts for Tmax (a,c) and Tmin (b,d) HWDI
813 with respect to indices derived from ERA-Interim (a,b) and BEST (c,d) daily data. Evaluation against BEST is
814 for 1993-2013. Dots indicate correlation values significant at a 95% level as estimated by a t-distribution (see
815 text).



816 FIG. 3. (a) Climatology of MAMJ AT35 HWDI (see text) for ERA-Interim over the 1993-2014 period. (b)
817 Spearman rank correlation of MAMJ System 5 re-forecasts for AT35 HWDI with respect to the index derived
818 from ERA-Interim over 1993-2014. Dots indicate correlation values significant at a 95% level as estimated by a
819 t-distribution (see text). Gray areas show regions where no AT35 heat waves are found with ERA-Interim data.

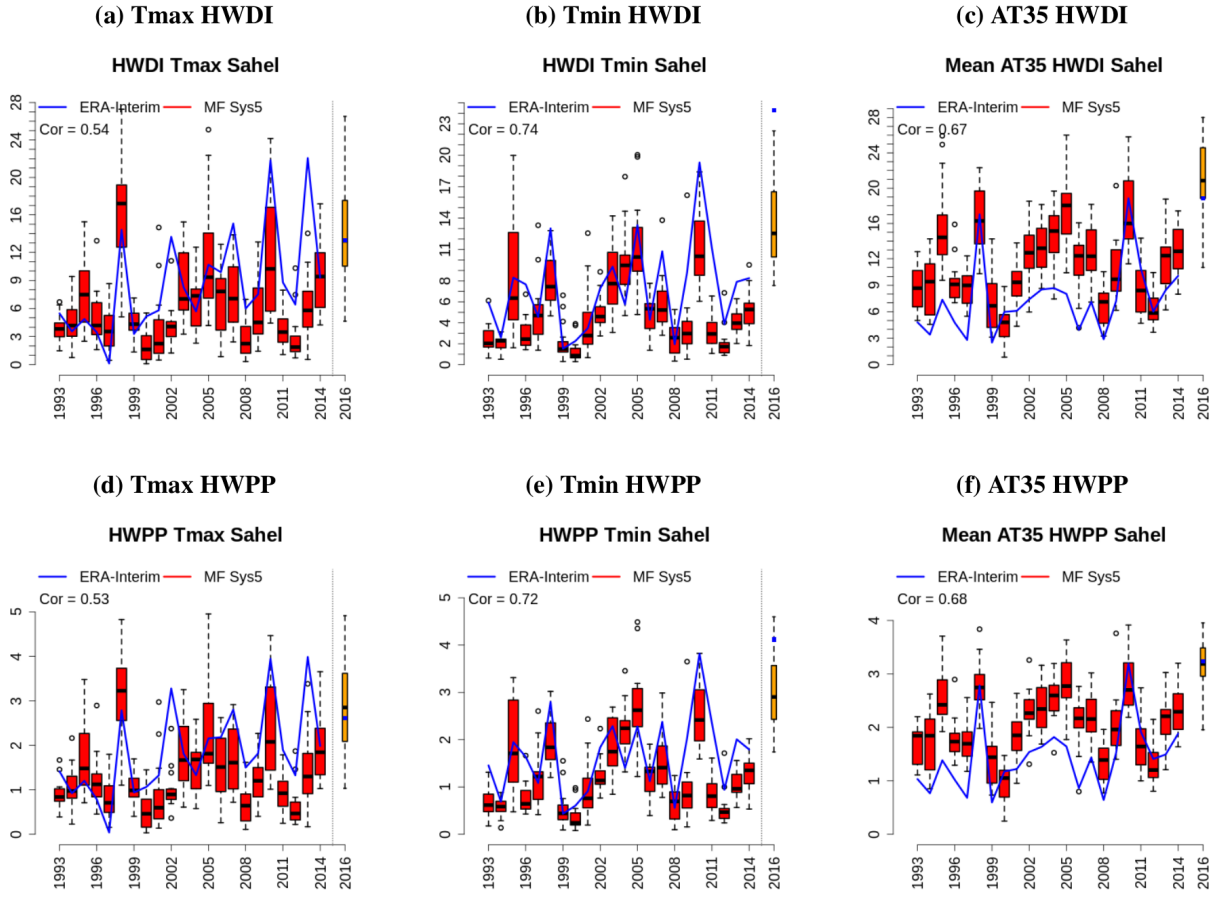
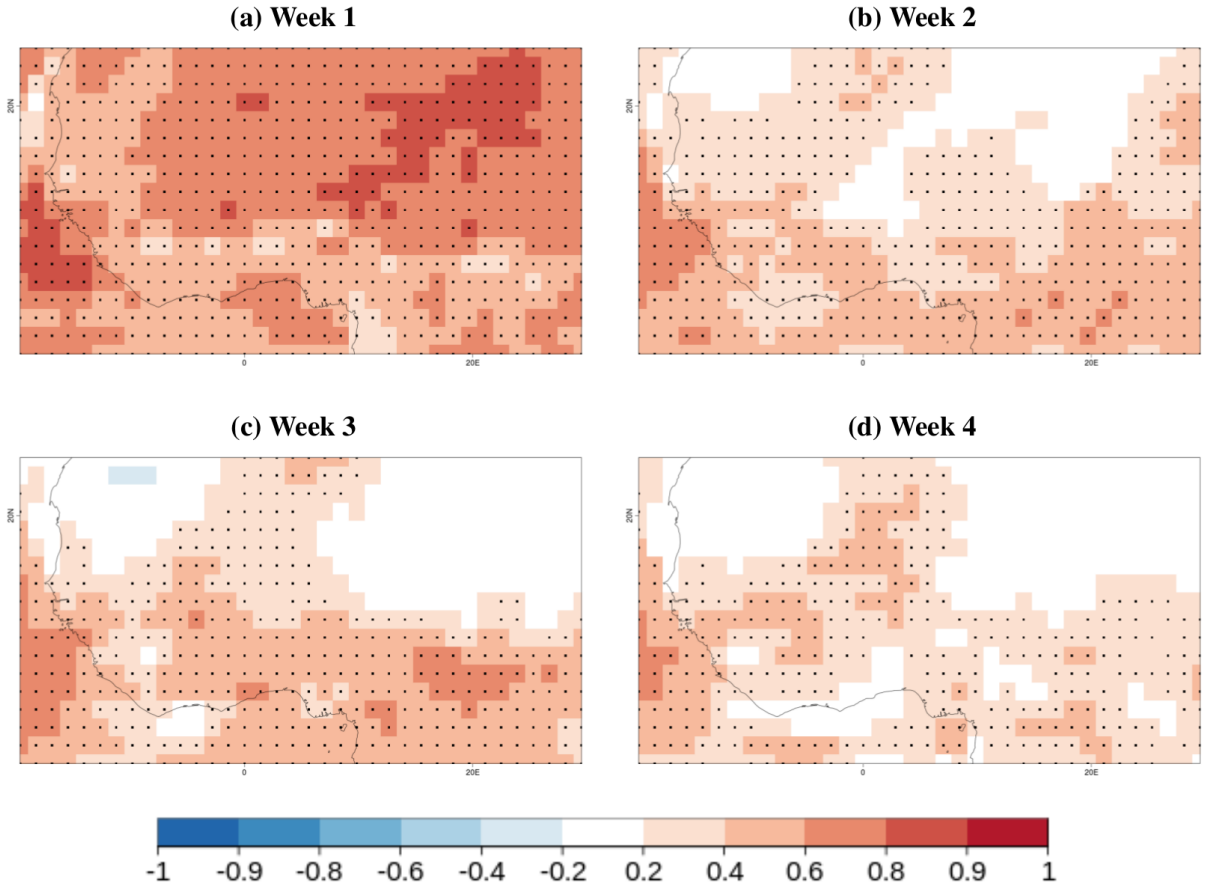


FIG. 4. Box and whisker plots of the ensemble re-forecasts of Tmax HWDI (a) Tmin HWDI (b) AT35 HWDI (c) and respective HWPP indices (d-f) for MAMJ 1993 to 2014 (in red). For each plot, boxes show the interquartile range and whiskers the full range of the 15-member System 5 ensemble (up to 1.5σ). Outliers are shown with circles. The blue line shows indices derived from ERA-Interim, and the Pearson time correlation with the ensemble mean re-forecast index over the 1993-2014 period is shown in the top left corner of each plot. The orange box and whiskers is the 51-member MAMJ 2016 forecast, and a blue dot shows the corresponding ERA-Interim index.



827 FIG. 5. (a-d) Weekly mean apparent temperature correlation against ERA-Interim over the re-forecast period
828 according to lead time (weeks 1 to 4, respectively) for 1 April and 15 April 1993-2014 start dates in the Météo-
829 France S2S contribution based on System 5. Dots indicate correlation values significant at a 95% level as
830 estimated by a t-distribution (see text).

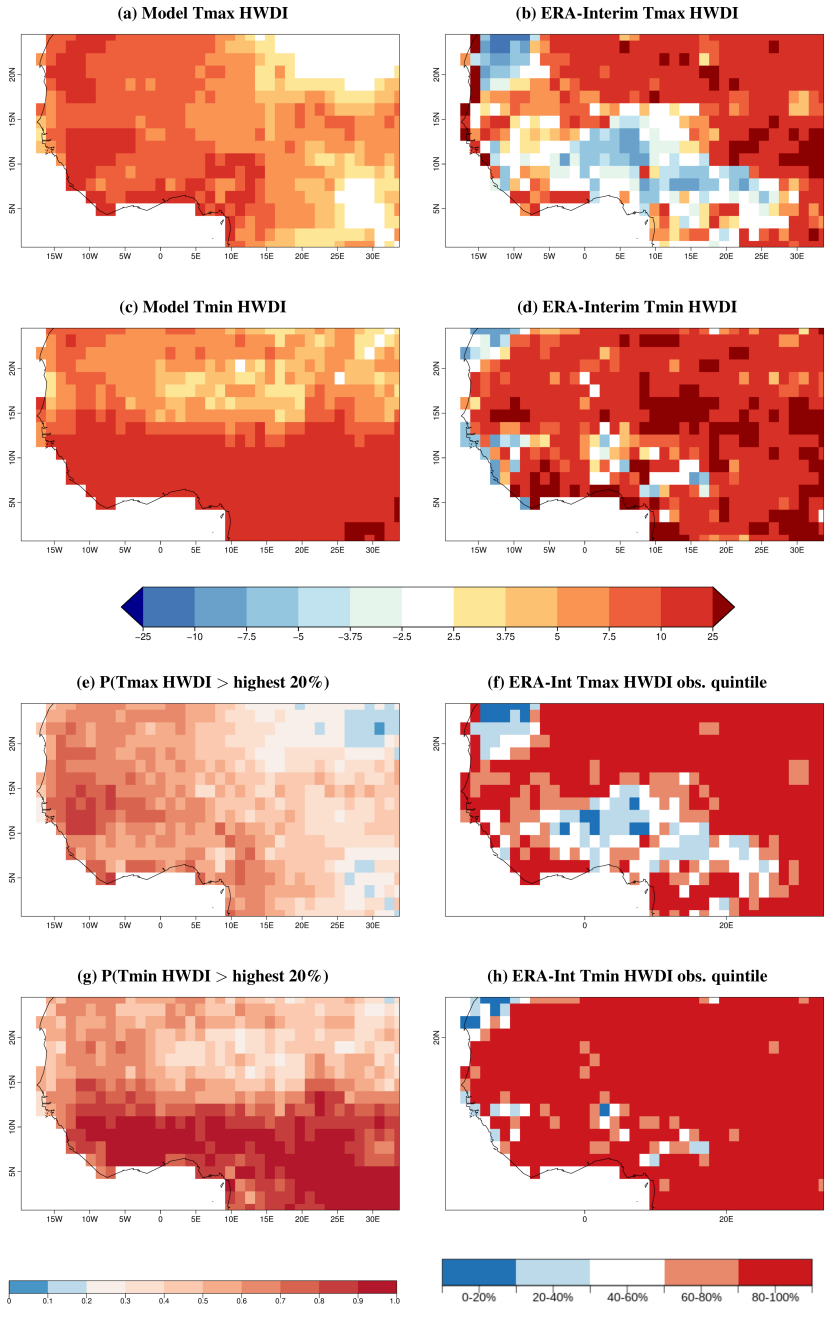


FIG. 6. Real-time System 5 forecast anomalies for Tmax and Tmin HWDI for MAMJ 2016 initialized end of January, with respect to the re-forecast period (a,c) and verification using ERA-Interim daily data (b,d); probabilistic forecast based on counts of System 5 ensemble members exceeding the highest quintile of the re-forecast for MAMJ 2016 initialized end of January (e,g) and actual observed quintile of ERA-Interim minimum and maximum temperature HWDI (f,h).

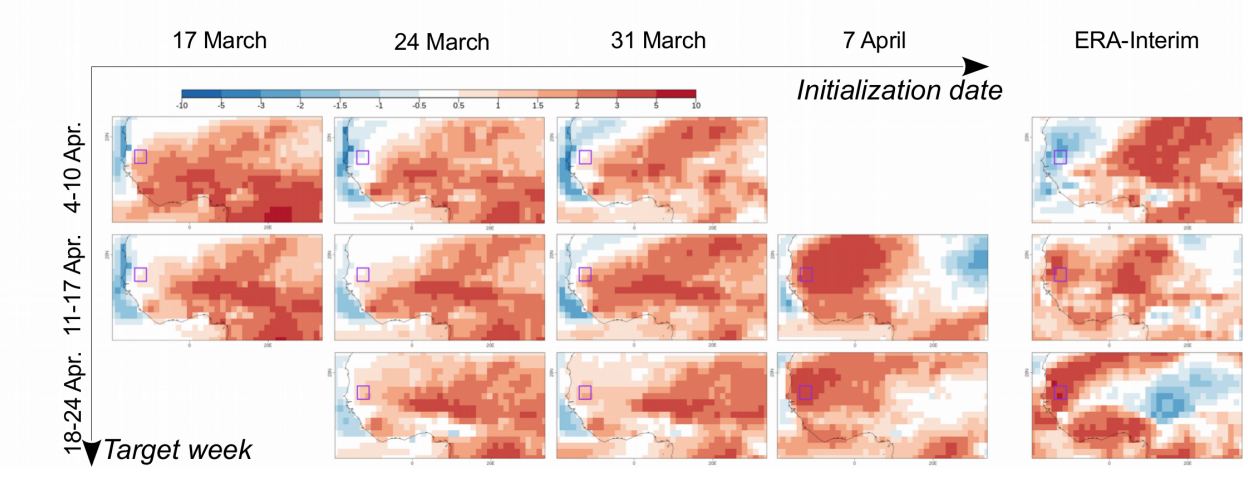


FIG. 7. Mean daily apparent temperature anomalies over West Africa for target weeks 4-10 April, 11-17 April and 18-24 April 2016 before and during the case study period (from top to bottom along the y-axis) in the S2S forecasts initialized on 17 March, 24 March 31 March and 7 April 2016 (from left to right along the x-axis), and corresponding ERA-Interim anomalies (right column). Anomalies are computed with respect to model re-forecast climatologies for S2S forecasts, and ERA-Interim 1993-2014 climatology for ERA-Interim.

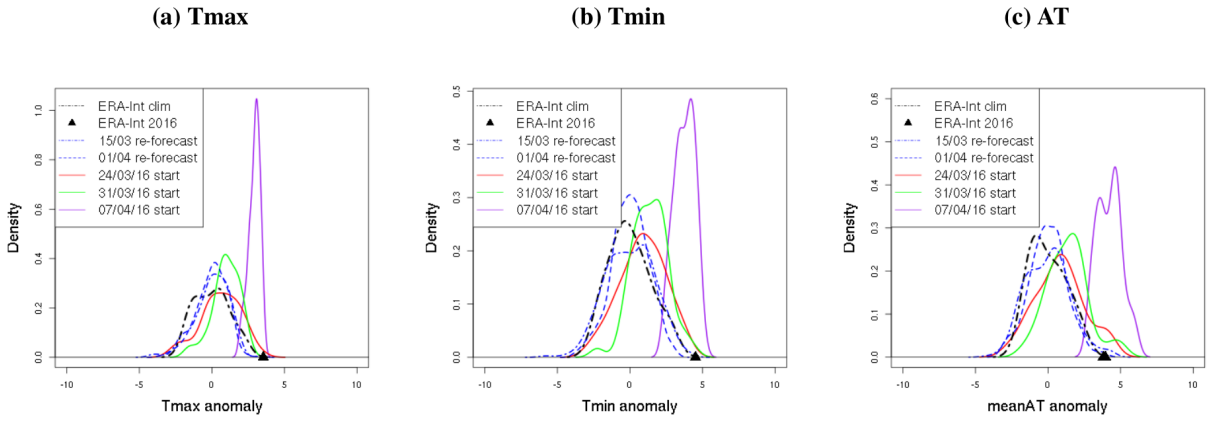


FIG. 8. Tmax (a) Tmin (b) and daily mean AT (c) anomaly distributions over the Matam region for the case study dates (12-22 April) for ERA-Interim 1993-2014, March 15 and April 1 re-forecasts, and three start dates preceding the 2016 case study event. Anomalies are computed for the 2016 S2S forecasts with respect to a weighted linear combination of the climatologies of the two closest re-forecast start dates for the CS dates, depending on the forecast initialization date.

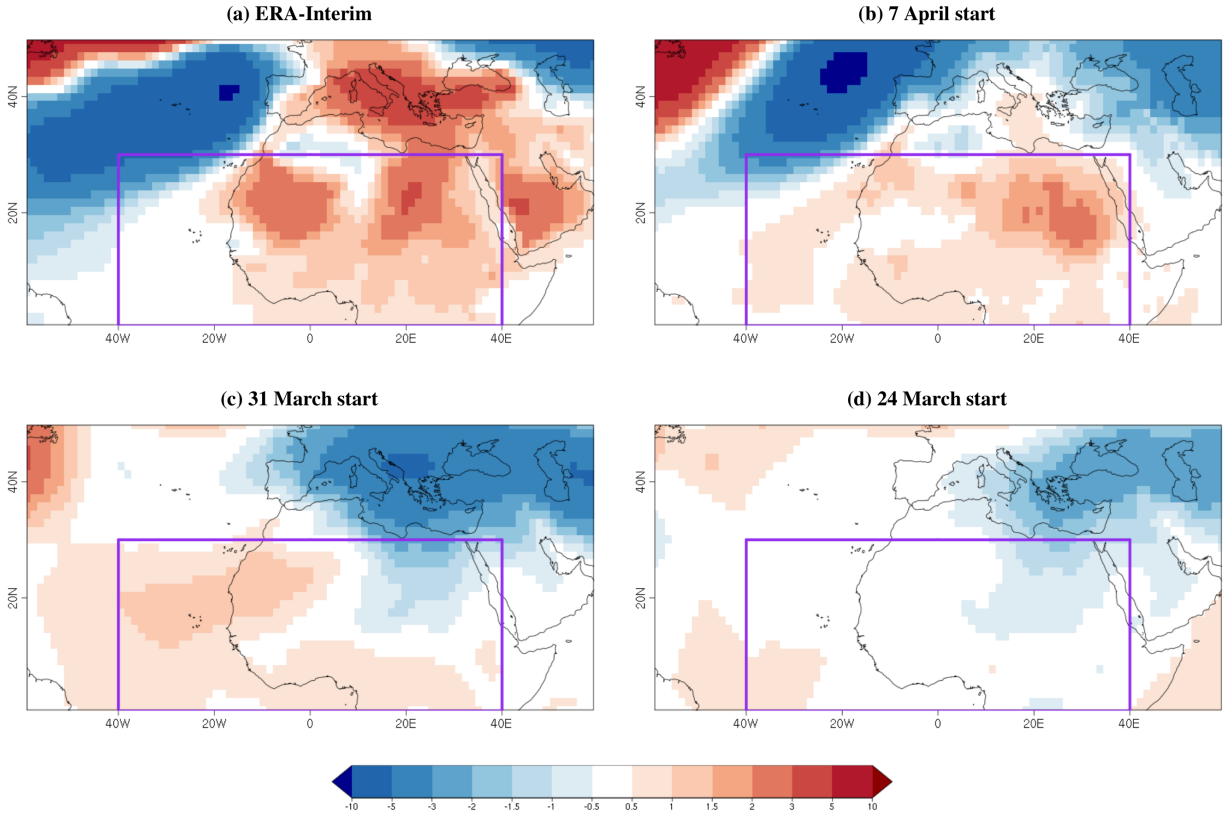
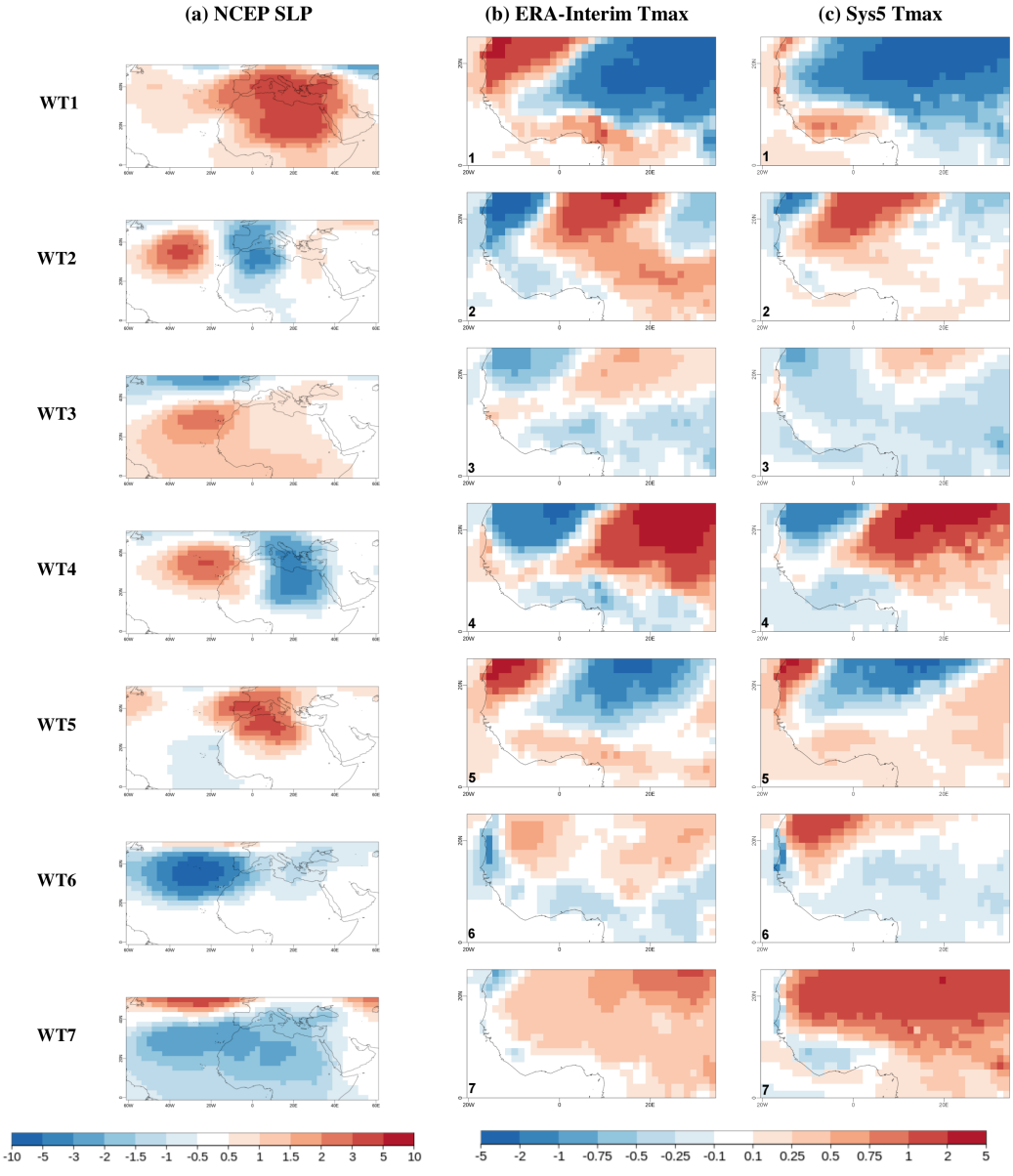
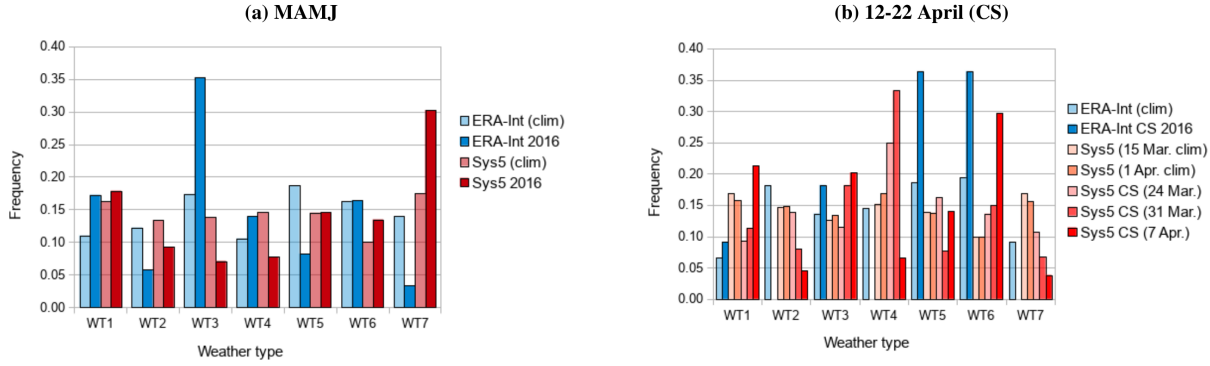


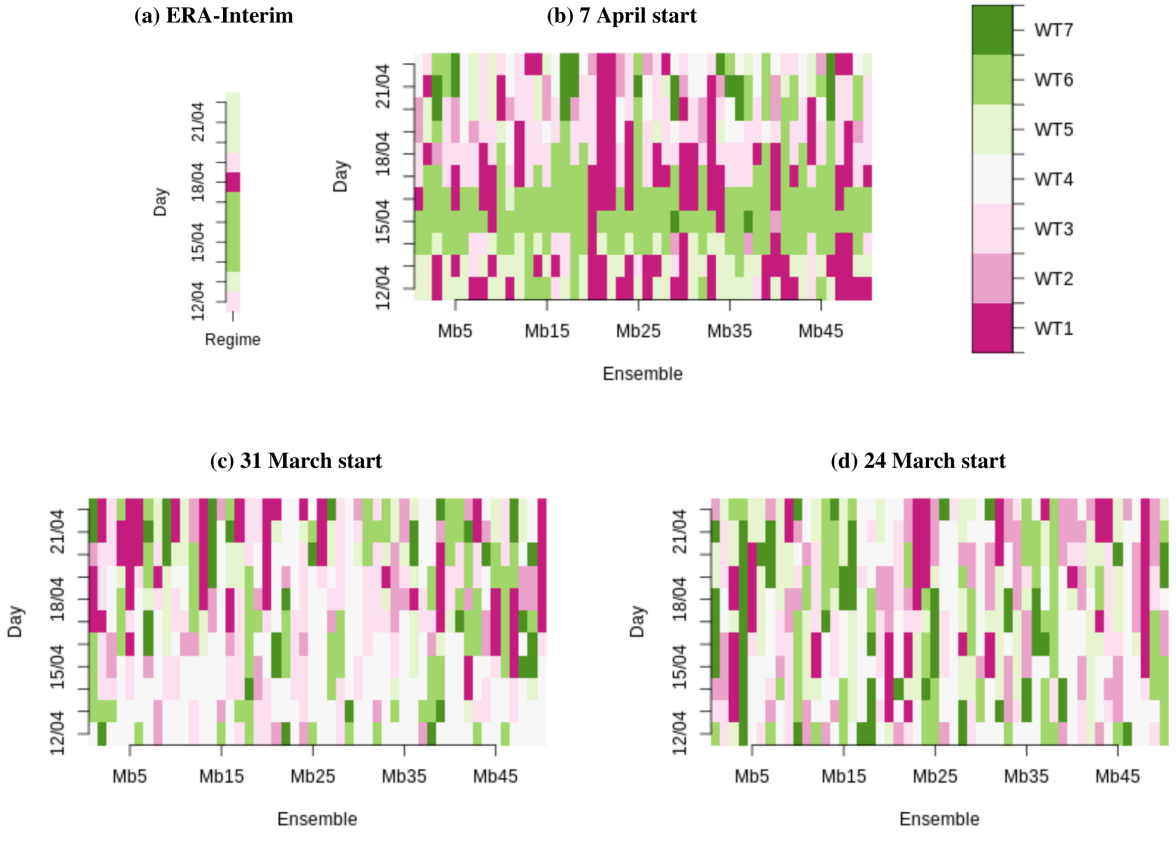
FIG. 9. Mean sea level pressure anomalies with respect to 1993-2014 for 12-22 April 2016 for ERA-Interim (a) and different S2S start dates (b-d).



848 FIG. 10. Column (a): sea-level pressure (SLP) composite in NCEP data for each weather type (WT) 1 to 7
849 obtained over the West African - Sahel region. Columns (b-c): maximum near-surface temperature composites
850 for each WT for MAMJ 1993-2014 in ERA-Interim (b) and System 5 re-forecasts initialized end of January (c).



851 FIG. 11. WT frequency in re-forecast runs and for 2016 (a) in MAMJ (System 5 re-forecasts initialized end
852 of January) and (b) for the 12-22 April 2016 case study (S2S runs based on System 5).



853 FIG. 12. Klee diagrams of the WTs for the 12-22 April 2016 case study in (a) ERA-Interim and (b-d) S2S
854 ensemble predictions initialized on three consecutive start dates.