Subseasonal forecasts of Heat waves in West African cities

Cedric G. Ngoungue Langue[1,2], Christophe Lavaysse[2,3], 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

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

Abstract. Heat waves are one of the most dangerous climatic hazards for human and ecosystem health worldwide. Accurate forecasts of these events are useful for policy makers and climate services to anticipate the consequences of extreme heat. In particular, subseasonal forecasts are of great importance in order to implement actions to mitigate the consequences of extreme heat on human and health. In this perspective, the present study addresses the predictability of heat waves at subseasonal timescales in West African cities over the period 2001-2020. The cities were grouped in three climatic regions based on their climate variability: continental, Atlantic and Guinea. Two types of heat waves were analyzed : dry heat waves using 2-meter temperature and wet heat waves using average wet bulb temperature. Two models that are part of the subseasonal to seasonal forecasting project, namely the European Centre for Medium-Range Weather Forecasts (ECMWF) and the United Kingdom Meteorological Office models, were evaluated using two state-of-the-art reanalysis products, namely the fifth generation ECMWF reanalysis and the Modern-Era Retrospective analysis for Research and Application. The performance of the forecast models in predicting heat waves is assessed through the computation of categorical metrics such as the hit-rate, the Gilbert Skill Score and the False Alarm Ratio. The results suggest that at subseasonal timescales, the forecast models provide a better forecast than climatology, but the hit rate and false alarm rate are sub-optimal and the forecasts may be overestimating the duration of heatwaves, while under- predicting the intensity. Nevertheless, the use of subseasonal forecasts in west African cities can be recommended for prediction of heat waves onset up to two weeks in advance.

1 Introduction

The impact of heat waves on different sectors, in particular the economy and health, makes them one of the most dangerous climate hazards globally (Perkins, 2015). Heat waves pose a significant threat to human health, as they cause discomfort and stress to body temperature regulation. In some cases, heat waves can lead to various cardiovascular and respiratory diseases, increasing the risk of morbidity and mortality [e.g. Anderson and Bell (2009); Gasparrini and Armstrong (2011); Kovats and Hajat (2008); Huynen et al. (2001)]. The combination of heat and humid atmospheric conditions can exacerbate heat stress and lead to deaths, particularly among vulnerable populations such as children and the elderly (Russo et al., 2017). The damages of heat waves on human health are amplified in urban regions due to the urban heat island effect. Over the last decades, West Africa cities were affected by several extreme heat events. In May 2013, the Senegalese town of Matam experienced a severe heat

wave, with temperatures reaching 50°C. This event, which persisted during the day and night, caused the death of 18 elderly in 10 days. At the same period, Mauritania was affected by a devastating heat wave with maximum temperatures exceeding 46°C, causing the death of more than 25 elderly people and children. Recently in April 2024, most of African countries experienced extreme heat conditions. In Mali, for example, temperatures exceeded 48°C, causing the death of more than 100 people.

Climatic projections show an increase in the frequency, intensity and duration of extreme temperatures over the next century and beyond [e.g. Kharin et al. (2007); Fischer and Schär (2010); Perkins et al. (2012)]. Under these warming conditions, the frequency of extreme events such as heat waves will increase. In the latest Intergovernmental Panel on Climate Change report (IPCC report 2023), the authors show that equatorial regions will be more affected by climate change than mid- and high-latitudes. This result from the IPCC is a warning bell, as the predictability of heat waves remains poorly documented in some equatorial regions, such as sub-Saharan Africa.

The impact of heat waves on human activities and health increases the need for skillful and reliable climate forecasts on subseasonal to seasonal time scales in order to anticipate risks and develop appropriate responses (Lowe et al., 2016). Therefore, early warning systems are of crucial importance to provide information on the occurrence of such events. Many early warning systems integrate short and medium-range forecasts of potential weather hazards up to two weeks ahead. The subseasonal forecast range, from 2 to 6 weeks leadtime, is also highly relevant for actions aimed at mitigating the consequences of extreme heat [e.g. White et al. (2017); Moron et al. (2018); Tompkins et al. (2019); Osman et al. (2023)]. Subseasonal forecasts are used to monitor the evolution of specific weather patterns that have been identified in advance with seasonal forecasts. Subseasonal and seasonal forecasts (S2S) are of great importance for humanitarian services in order to build up a "Ready-Set-Go" early warning concept that allows early actions to be taken before a potential disaster [e.g., Bazo et al. (2019); Lala et al. (2022); Domeisen et al. (2022)].

Heat waves are often associated with extreme heat, which can be exacerbated by other factors such as humidity levels. A heat wave is defined as a period of unusual hot temperature over a region persisting at least three consecutive days during the warm period of the year based on local (station-based) climatological conditions, with thermal conditions recorded above given thresholds [e.g. Perkins and Alexander (2013); Déqué et al. (2017); Barbier et al. (2018); Batté et al. (2018); Ngoungue Langue et al. (2023)]. Many factors can affect the definition of a heat wave, including the end-user sectors (human health, infrastructures, transport, agriculture) and also the climatic conditions of the regions (Perkins and Alexander, 2013). Heat waves can be defined from daily meteorological variables such as daily raw temperature [e.g. Batté et al. (2018); Lavaysse et al. (2018); Engdaw et al. (2022); Ngoungue Langue et al. (2023)], minimum, mean, and maximum daily wet bulb temperature [e.g. Yu et al. (2021); Ngoungue Langue et al. (2023)] or heat stress indices [e.g. Robinson (2001); Fischer and Schär (2010); Perkins et al. (2012); Guigma et al. (2020); Ngoungue Langue et al. (2023)] using relative or absolute thresholds. Heat stress indices are used to combine relevant atmospheric variables (such as temperature, humidity, solar and thermal radiation, wind speed) to indicate the impact of the environment on the human body. Examples include apparent temperature, Universal Thermal Climate index, Excess Heat Factor (EHF) and Excess Heat Index (EHI) (McGregor et al., 2015).

A few studies on heat wave forecasting have been carried out in West African regions. Batté et al. (2018) evaluated the predictability of heat waves during spring at subseasonal time scale using the Météo-France model as part of the S2S project. To assess the skills of the models, they used the apparent temperature and T2m anomalies as indicators for heat wave detection. Apparent temperature (AT) represents the temperature actually felt by humans, caused by the combined effects of air temperature, relative humidity and wind speed. The results show that the Météo-France model is able to predict heat waves up to one week in advance. Guigma et al. (2021) assessed the predictability of Sahelian heat waves during spring at subseasonal time scale using the ECMWF extended-range forecasting system (ENS-ext), ERA5 and BEST gridded data for the evaluation. Their approach is based on the prediction of the probability of heat waves occurrence using T2m and heat index (HI) as indicators. They show that ENS-ext is able to forecast Sahelian heat waves with significant skill up to 2 weeks in advance; and with increasing lead time, wet heat waves (those combined with high humidity levels) are more predictable than dry heat waves (those combined with low humidity levels).

Batté et al. (2018) assessed heat waves predictability using T2m and AT anomalies. While this approach provides information about the evolution of T2m and AT for the subsequent days, it is not sufficient to determine heat waves characteristics such as the duration and frequency. In this study, we will adapt the methodology proposed by Lavaysse et al. (2019) when assessing the predictability of heat waves over Europe. This method offers a complete evaluation of heat wave characteristics including not only the evaluation of heat wave intensity, but also of heat wave onset and duration. It involves the computation of evaluation metrics to assess the skills of the models (see Section 2.4.6).

The present study assesses the subseasonal predictability of heat waves in West African cities over the period 2001-2020 using two models part of the S2S project namely, ECMWF and UKMO. To the authors' knowledge, this work is the first of its kind in the region and represents a benchmark for future studies. To achieve our goal, we first analyze the representation of T2m and wet bulb temperature (Tw) in the forecast models with respect to the reanalysis data used as references (see Section 2). Secondly, we evaluate the models on the representation of extreme heat events and the spatial variability of heat wave characteristics. Finally, the skill of the models in predicting heat waves is evaluated.

The remainder of this article is organized as follows: in section 2, we present the region of study and the data used for this work; the description of the methodology is also provided. Section 3 contains the main results of this study following the methodology presented in section 2. Section 4 provides a discussion of the results and section 5 is dedicated to the conclusion.

2 Region, Data and Methods

2.1 Region of interest

This work is carried out for West Africa, covering an area between 5°-20°N and 15°W-10°E, extending from the Atlantic coast to Chad and from the Gulf of Guinea to the southern fringes of the Sahara desert [Fig.1]. The climate of West Africa is mainly influenced by the West African monsoon, which regulates the rainy season and therefore affects agriculture. West

Africa experiences a very hot and dry climate over the Sahel region, and a hot and humid climate over the Guinea coast. The climatic conditions over West Africa and its location over the tropics, make the region exposed to heat waves.

West Africa exhibits a high climate variability at regional- and local-scale. In this study, we focus on a few large, densely populated cities located in the Sahel region and on the Atlantic coast. The cities were grouped in three regions based on their location, climate variability and the evolution of heat waves characteristics in each region (Moron et al., 2016; Ngoungue Langue et al., 2023). The regions are structured as follows :

- Continental region (CO hereafter) englobes the cities of Bamako, Ouagadougou and Niger [Fig.1];
- Coastal atlantic region (AT hereafter) englobes the cities of Dakar, Nouakchott, Monrovia and Conakry [Fig.1];
- Coastal Guinean region (GU hereafter) englobes the cities of Yamoussoukro, Abidjan, Lomé, Abuja, Lagos, Accra, Cotonou and Douala [Fig.1].

2.2 Reanalysis products

African cities suffer from a critical lack of weather observation stations, and the few that exist are not well distributed accross the regions. Therefore, reanalysis data appear as good candidates to overcome this issue. We are aware that reanalysis data have a low resolution for detecting the highest temperatures at a point location, as well as the urban heat island effect. Nevertheless, reanalyses provide a numerical description of the recent climate by combining models with observations and are invaluable to numerous users around the world (Hersbach, 2016).

In this work, we use two state-of-the art reanalysis products namely the fifth-generation European Center for Medium-Range Weather Forecasts reanalysis (ERA5, (Hersbach et al., 2020)) and the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2, (Gelaro et al., 2017)) from the National Oceanic Atmospheric Administration (NOAA). In the following, we will use "MERRA" to refer to MERRA-2 which, with ERA5, are the references for the evaluation of the forecast models. ERA5 and MERRA are part of the most reliable reanalyses used in Africa regions especially on the monitoring of heat waves[e.g. Barbier et al. (2018); Ngoungue Langue et al. (2021); Engdaw et al. (2022)]. Since ERA5 is used to initialize the atmospheric component of the ECMWF extended reforecasts, one could argue that the evaluation of the skills of the forecast models using only ERA5 as reference is circular. For this reason, we have also included MERRA as a second reference. It also allows for estimating the uncertainties of the reanalyses.

2.2.1 ERA5

The ERA5 reanalysis provides hourly estimates of various climate variables for the entire globe using 137 hybrid sigma levels up to 80 km above the surface (Hersbach et al., 2020). The original spatial resolution is 0.28125 degrees, interpolated to a regular 0.25°x 0.25° grid. ERA5 outputs are generated by the CY41r2 model cycle of the Integrated Forecast system (IFS) of ECMWF, which uses a ten-member ensemble of 4D variational data assimilation. The set of atmospheric variables used within

the ERA5 database are the hourly 2-meter temperature (T2m) and hourly 2-meter dew point temperature (d2m) covering the period from 1 January 2001 to 9 February 2021. From these variables, daily T2m (max, min, mean) and wet bulb temperature (Tw) are derived. The dataset is accessed through the Climserv database of the Institut Pierre Simon Laplace (IPSL) server or the Copernicus Climate Data Store (CDS). The land-sea mask used in this study is obtained from the ERA5 reanalysis and is available on the CDS.

2.2.2 MERRA

MERRA, unlike ERA5, has a spatial resolution of 0.625°x0.5° and provides data on 42 standard pressure levels. It uses an upgraded version of the Goddard Earth Observing System Model, Version 5 (GEOS-5) data assimilation system and the Global Statistical Interpolation (GSI) analysis scheme of Wu et al. (2002). To ensure consistency in our analysis, we converted the MERRA data to a spatial resolution of 0.25°x0.25°, similar to ERA5 using a conservative first-order interpolation. In the MERRA database, we use the hourly T2m, hourly 2-meter specific humidity and the pressure field at the surface from 1 January 2001 to 9 February 2021 to calculate daily T2m(max,min,mean) and Tw. The MERRA dataset was also accessed through the Climserv database on the Institut Pierre Simon Laplace (IPSL) server.

2.3 Forecasts products

To bridge the gap between medium range weather forecasts and seasonal forecasts, in 2013 the World Weather Research program (WWRP) and the World Climate Research program (WCRP) jointly launched a 5-year research initiative called the Subseasonal to Seasonal (S2S) project (Vitart et al., 2017) which aims to improve forecast skill and understanding on the subseasonalto seasonal timescale with special emphasis on high-impact weather events. The S2S project focuses on the risk of extreme weather conditions, including tropical cyclones, droughts, floods, heat waves and monsoonal rainfall (see http://www.s2sprediction.net/). In order to obtain more robust statistical results, we conducted our study over a 20-year period, specifically using hindcast data from 2001 to 2020. Among the twelve models involved in the S2S project, we evaluated the ECMWF and UKMO (United Kingdom Met Office) models because they are both available throughout the study period.

2.3.1 ECMWF forecasts

The extended-range ECMWF forecast model used in this work is run on the Integrated Forecast System (IFS) cycle CY47R3 released on October 10th, 2021. The native spatial resolution of the ECMWF model is Tco639 L137 (about 16 km) up to day 15 and Tco319 (about 32 km) after day 15, but the downloaded data are interpolated to a regular 0.25°x 0.25° latitude/longitude grid to match the resolution of ERA5 for evaluation. It contains 137 sigma levels from the surface to 80 km. ECMWF provides two types of outputs for the S2S program: forecasts and hindcasts. Hindcasts are forecasts produced for past dates and allow analysis of how the current system would have performed, alongside a consistent dataset covering a longer time period for evaluation. They are useful for the calibration of the model and post treatment analyses.

The ECMWF extended-range hindcasts are run with 11 members (10 perturbed and 1 unperturbed). ECMWF extended-range hindcasts are produced twice a week, on Monday and Thursday at 00Z. This means that for each week a new set of hindcasts is produced to calibrate the real-time ensemble forecasts for Monday and Thursday of the following week. We have only analyzed the hindcasts produced on Thursdays, as a preliminary investigation into the initialization dates of the hindcasts showed that most models were launched on Thursdays. In the database, some models use fixed dates of the month (1^{st}, 09^{th}, 17^{th} for example) and others, specific days of the week (Monday, Thursday for example), which generate some difficulties to handle. Most of the models do not cover the period under study, except for UKMO which is available but uses fixed initialisation dates of the month (see section below for more details). It is therefore no longer possible to carry out a multi-model evaluation as we had planned and we have limited the evaluation to ECMWF and UKMO. The ECWMF hindcasts produced on Monday cover the same period of the ones produced on Thursday. Thus, using the Monday hindcasts we have no significant changes in the frequency of the events detected. The 11-member ensemble hindcasts start on the same day and month as the forecasts, but covering the last 20 years. In our case, the forecast year is 2021 and we focus on the previous 20 years from that date, and the hindcasts run from 0-46 days. The variables of interest in the ECMWF S2S are T2m(max,min) over the last 6 hours, daily average T2m and d2m from which the daily average Tw was derived. The data are open access and available on the S2S project website (https://apps.ecmwf.int/datasets/data/s2s-realtime-instantaneous-accum-ecmf/levtype=sfc/type=cf/).

2.3.2 UKMO forecasts

The UKMO model runs on the HadGEM3 GC2.0 model which simulates the uncertainties of the initial conditions using a lagged initialisation and the uncertainties of the model using a stochastic scheme. The native spatial resolution of the UKMO model is N216: 0.83°x0.56° (about 60 km at mid-latitudes), but the downloaded data are extrapolated to a regular latitude/ longitude grid of 0.25° x 0.25°. It contains 85 vertical levels from the surface to 85 km and 4 soil levels: level 1 (0 - 0.1 m), level 2 (0.1 - 0.35 m), level 3 (0.35 - 1 m) and level 4 (1- 3 m). Similar to ECMWF, UKMO provides to the S2S program forecasts and hindcasts. The UKMO hindcasts analyzed here are run using the model version released in 2023 which produces 7 members per cycle (6 perturbed and 1 control) for the period 1993-2016 (no UKMO hindcasts available after 2016). They are produced 4 times per month, on the 1^{st}, 9^{th}, 17^{th} and 25^{th}. We are aware that these initialisation dates are not the same as those of ECMWF, but we are interested in this work on the predictability of heat waves in a broad perspective, not on specific events. Our target period is going from January 2001 to February 2021, and as mentioned earlier, the UKMO hindcasts are not available after the year 2016. To solve this problem and get more robust statistical results, we recomposed the products to obtain a new composite that covers the whole target period. The process applied is described in the following expression:

$$\mathbf{T2m\_max} = \mathbf{Hindcasts}(\mathbf{T2m\_max})_{\mathbf{2001-2016}} \bigcup \mathbf{Forecasts}(\mathbf{T2m\_max})_{\mathbf{2017-2021}} \quad (1)$$

The forecasts were extracted for the same days as the hindcasts initialization. In order to apply the concatenation over time between the hindcasts and forecasts, the coordinates dimensions of the two datasets must be the same. As shown early, the number of ensemble members in UKMO hindcasts and forecasts are completely different. Therefore, to meet this requirement,

we reduced the number of ensemble members from 7 to 4 (1-control member and 3-perturbed members)in the hindcasts to match the number of ensemble members in the forecasts. We selected the three first perturbed members in the hindcasts over the 7 available. The UKMO forecasts analyzed in this work are launched for a 6-week duration. The variables extracted from the UKMO database are the same as those in ECMWF. A summary of the main differences between the two models (ECMWF, UKMO) is provided in [Table1].

2.4 Methods

2.4.1 Estimation of temperatures at the city scale from reanalyses

Weather forecasts provide the evolution of atmospheric variables on a global scale, which implies the need to have data from observation stations to access information on a local scale. This is a major problem in areas where there are not enough weather stations to collect data, as is the case in African cities.

To address this issue, downscaling methods can be employed. However, in this study, we study phenomena at the city scale, and the spatial resolution of the reanalyses (ERA5, MERRA) is too coarse for this purpose. Although the reanalysis scale is more representative of the spatial variability of a heat wave occurring in a city than an isolated local station, a validation analysis is needed on test stations in order to determine the best interpolation technique for estimating local temperature from reanalyses. Following the same approach as in Ngoungue Langue et al. (2023), local temperatures over the cities were derived from the reanalysis using the reanalysis grid point closest to the station that satisfies a land-sea mask (lsm) of at least 0.5 ([Table2] shows the lsm values for all the cities considered in this study, the same technique was applied for the forecast models). The lsm indicates the proportion of land contained in a grid point. If the lsm is less than 0.5, this means that the grid point is mainly covered by the ocean, while a lsm greater than 0.5 implies more land coverage.

2.4.2 Heat wave detection

In the present study, two types of heat waves are investigated : dry and wet heat waves. Dry heat waves are associated with high temperatures and low humidity conditions. The detection of dry heat waves is processed using 2-meter temperature (T2m) as an indicator. We distinguished two categories of dry heat waves : those that occur during the daytime and are detected using maximum values of T2m (T2m_max), and those that occur during the night and are detected using minimum values of T2m (T2m_min). Concomitant heat waves, those that occur during daytime and the night, are extremely dangerous because the body does not have the time to recover from the daytime heat waves during the night [e.g. Li et al. (2017); Wang et al. (2020a, b)]. The most lethal heat waves are due not only to high temperatures but also to the effect of humidity [e.g. Steadman (1979a, b); Heo et al. (2019); Yu et al. (2021)]. Humidity is an important driver of wet heat waves. The combination of high heat and humidity can compromise the human body's main cooling mechanism: transpiration. The evaporation of sweat from skin cools our bodies, but higher humidity levels limit evaporative cooling. As a result, we can suffer heat stress and illness, and the consequences can even be fatal. Wet heat waves are detected here using average Tw as an indicator (Yu et al., 2021).

The computation of Tw is given by the following formula:

$$\mathbf{Tw} = \mathbf{T} * \mathbf{atan}\Big[\mathbf{A}(\mathbf{Rh} + \mathbf{B})^{\frac{\mathbf{1}}{\mathbf{2}}}\Big] + \mathbf{atan}\Big[\mathbf{T} + \mathbf{Rh}\Big] - \mathbf{atan}\Big[\mathbf{Rh} - \mathbf{C}\Big] + \mathbf{D} * (\mathbf{Rh})^{\frac{\mathbf{3}}{\mathbf{2}}} * \Big[\mathbf{atan}(\mathbf{E} * \mathbf{Rh})\Big] - \mathbf{F} \tag{2}$$

(Stull, 2011), (Rh is used in percentage, for example 40 for Rh=40%).

The computation of relative humidity (RH) varies depending on the available variables in the products. The first formula is applied for ERA5, and the second for MERRA reanalyses.

$$\mathbf{Rh} = \mathbf{100} * \frac{\exp\Big(\frac{\mathbf{a} * \mathbf{T_d}}{\mathbf{b} + \mathbf{T_d}}\Big)}{\exp\Big(\frac{\mathbf{a} * \mathbf{T}}{\mathbf{b} + \mathbf{T}}\Big)} \tag{3}$$

(August, 1828; Magnus, 1844; Alduchov and Eskridge, 1996)

$$\mathbf{Rh} = \mathbf{0.263} * \mathbf{p} * \mathbf{q} * \Big[\exp\Big(\frac{\mathbf{17.67} * (\mathbf{T} - \mathbf{T_0})}{\mathbf{T} - \mathbf{29.65}}\Big)\Big]^{-1} \tag{4}$$

(Ngoungue Langue et al., 2023)

$\mathbf{a} = 17.625, \mathbf{b} = 243.04, \mathbf{A} = 0.151977, \mathbf{B} = 8.313659, \mathbf{C} = 1.676331, \mathbf{D} = 0.00391838, \mathbf{E} = 0.023101, \mathbf{F} = 4.686035, \mathbf{T_0} = 273.16K$

Where $\mathbf{T}$(°C), $\mathbf{T_d}$(°C), $\mathbf{T_0}$(K), $\mathbf{p}$(hPa) and $\mathbf{q}$ are respectively the ambient temperature, dew-point temperature, reference temperature (water triple point temperature), pressure and specific humidity.

Daily maximum and minimum temperatures are computed respectively from maximum and minimum temperatures in the last 6 hours. This choice of the computation of the extreme daily values is made according to the forecast models outputs. Daily average wet bulb temperature is computed from hourly dew point temperature. This restriction to T2m_min, T2m_max and Tw is related to the atmospheric variables available on S2S outputs preventing us from computing more elaborated indices as in Ngoungue Langue et al. (2023).

We defined a heat wave as a consecutive period of at least 3 days during which the daily temperatures exceed the daily climatological 90^{th} percentile computed over the entire period for T2m_min, T2m_max or Tw respectively [Fig.2]. The 90^{th} percentile threshold is computed independently for the reanalyses and the hindcasts. As mentionned previously, the hindcasts are run at least once every week for a 6-week duration. For each initialization date within a month and for each lead time, we computed the daily climatological 90^{th} percentile over the study period (2001-2020) (see [Fig.S1] in supplement material for an illustration of the computation of the 90^{th} percentile threshold). Heat waves are detected independently for each initialization date within a month using the threshold values computed from this initialization (see [Fig.S2] supplement material). For example, to detect heat waves in ECMWF hindcasts run on 4 January, the daily climatological 90^{th} percentile of each day of

the 4 January run is applied to each corresponding day of the runs on 4 January 2001, 4 January 2002,.. to 4 January 2020. The 90^{th} percentile appears to be a sufficient threshold for monitoring heat waves affecting human health based on previous studies [e.g. Perkins and Alexander (2013); Russo et al. (2014); Barbier et al. (2018); Lavaysse et al. (2019); Ngoungue Langue et al. (2023)]. Nevertheless, it is useful to determine the intensity of the events in order to establish a classification according to their severity (intensity), from "harmless" to "extremely dangerous", for example (see section 2.4.3 for more details on the computation of the intensity of heat waves).

The choice of a relative threshold is more appropriate as it is easily replicable in other regions. When two heat wave events are separated by one day with an indicator value below the daily 90^{th} percentile on this day, they are pooled together to form a single event [Fig.2].

2.4.3 Heat wave characteristics

After the detection of a heat wave, some important characteristics are deduced, namely the duration and the intensity. They are useful to investigate the severity of an event. The predictability of heat waves is assessed for occurrence, duration and intensity. To determine the occurrence and duration of heat waves, we create individual boolean files from the T2m_min, T2m_max and Tw time series at each grid point, which is equal to 1 if it is a hot day and 0 otherwise. This operation is performed on a daily time scale over the study period. Hot days are days belonging to heat waves with the values of one of the 3 indicators (T2m_min, T2m_max or Tw) above their daily 90^{th} percentile. In order to assess the characteristics of heat waves, only hot days belonging to heat wave sequences are considered (Ngoungue Langue et al., 2023). Boolean files are computed separately for reanalyses and forecasts in order to assess the representation of heat wave occurrence and duration.

The intensity of a heat wave was defined as the sum of the daily exceedances of the indicators (T2m_min, T2m_max or Tw) values to a constant threshold for the duration of the event. It should be emphasized that this constant threshold is not used for heat waves detection, but only to compute their intensity. In this study, which is part of the project Agence National de la Recherche STEWARd (STatistical Early WArning systems of weather-related Risks from probabilistic forecasts, over cities in West Africa), we are interested in heat waves, which can be harmful to human health. Therefore, the constant threshold mentioned above is defined as the minimum of the daily climatological 90^{th} percentile over the study period. The choice of a constant threshold for the computation of heat waves intensity is very important because it takes into account the seasonal cycle. This makes it possible to assess the severity of the events using the same reference. In fact, the most dangerous heat waves will have the highest intensity values.

2.5 Evaluation of heat waves forecasting in subseasonal models

Heat waves in the Sahel region occur mainly in spring due to the high temperatures in the region at that time (Barbier et al., 2018; Guigma et al., 2020). In this study, the region of interest was extended to the Guinean region in which heat waves are mainly driven by humidity coming from the Atlantic ocean. This advection of humidity over the Guinean region is active

during the season. Therefore, the detection of heat waves is performed over the whole season and not just in spring, to cover the wet and dry seasons in the region. Similarly, the evaluation of the models is done from January to December for the different initialization dates independently by computing some statistical metrics. The metrics are computed for each week (week 1 to week 6) for each initialization date in a month and the average score for each week is derived.

2.5.1 Evaluation of probabilistic forecasts

The representation of the evolution of T2m and Tw in ensemble forecasts is investigated by computing the bias between the subseasonal models and the reanalyses. To assess the predictive skills of the models in forecasting T2m, Tw and extreme heat events, we used probabilistic scores such as the Continuous Rank Probability Score (CRPS) and Brier Score (BS).

The CRPS is a quadratic measure of the difference between the forecast and reanalysis cumulative distribution functions (CDFs); it quantifies the relative error between forecasts and observations. The values of the CRPS range from [0 to 1], the closer it is from 0, the better the forecast. The CRPS is computed using the following formula:

$$\mathbf{CRPS} = \int_{-\infty}^{+\infty} \Big(\mathbf{P_f(x) - P_o}\Big)^{\mathbf{2}} \mathbf{dx} \tag{5}$$

$\mathbf{P_f}, \mathbf{P_o}$ represent the forecast and reanalysis CDFs respectively.

The BS measures the difference between the probability of a forecast and the outcomes; it is a good metric to quantify the accuracy of a forecast system. The BS is given by :

$$\mathbf{BS} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i=1}}^{\mathbf{N}} \Big(\mathbf{P_i - O_i}\Big)^{\mathbf{2}} \tag{6}$$

$\mathbf{P_i}, \mathbf{O_i}$ are respectively the forecast probability and observed outcome (1 if the event occurs and 0 if not). N is the length of the forecasts or observations time series.

2.5.2 From probabilistic to deterministic forecast

Ensemble forecasting is a tool used for making probabilistic weather forecasts. It is both an alternative and an indispensable complement to deterministic weather forecasts. Ensemble forecasting is of major interest because it provides different scenarios of the evolution of the state of the atmosphere. The particularity of ensemble forecasting is that, unlike deterministic forecasting, many trajectories are simulated in order to take into account uncertainties in the physical component of the model, the chaotic nature of the atmosphere, the observation network and imperfect initial conditions of the forecasts.

Although ensemble forecasting has many advantages over deterministic forecasting, the evaluation of the skills of an ensemble forecasting model remains quite complex due to the amount of information available. Statistical analyses can be performed

by considering the ensemble mean, the median member, the warmest and coldest members, the 1^{st} and 3^{rd} quartiles. Probabilistic scores exist for evaluating ensemble forecasting systems, but when it comes to the evaluation of specific events such as heat waves, the task becomes more complex. Moreover, probabilistic scores are difficult to understand for decision-makers and end-users who are not specialists, hence the need to simplify and transform probabilistic forecasts into deterministic forecasts that are useful for implementing early warning systems. Our approach consists in converting ensemble forecasts into boolean files using threshold values. Threshold values are defined as the percentage of members correctly forecasting heatwave days. Three threshold values were tested to find the optimal deterministic forecasts (see [Table3]). Assuming a threshold of 20% for example (i.e. that at least 20% of ensemble system members are forecasting heatwave days), we determined the probability of the ensemble forecast systems predicting a day as being part of a heatwave. If the probability is greater than the threshold value, the day will be counted as "1" and "0" otherwise. The methodology used here and the choice of the selected thresholds are inspired by Lavaysse et al. (2019).

2.5.3 Evaluation of deterministic forecasts

The predictability of heat waves in the forecast models is evaluated using dichotomous scores based on the coherence between reanalysis and forecasts such as hits, false alarms, misses and correct rejections. Let consider a 2x2 contingency table [Table4] from which 3 metrics have been computed : hit-rate, FAR and GSS. The hit-rate indicates the percentage of observed heat waves that have been correctly forecasted. The False Alarm Ratio (FAR) gives the percentage of forecasted events that did not occur. The Gilbert Skill Score (GSS) measures the fraction of observed events that are correctly predicted, adjusted for hits associated with random chance (close to the climatology). Chance hits (CH) are given as the event frequency multiplied by the number of event forecasts. The GSS takes in account the hits, misses, false alarms and neglects the correct rejections that would artificially improve the score. The values of the GSS range from [-$\frac{1}{3}$ to 1]; a GSS=0 indicates no skill while a GSS=1 perfect skill. The GSS is computed using the following formula:

$$\mathbf{GSS} = \frac{\mathbf{hits} - \mathbf{CH}}{\mathbf{hits} + \mathbf{false\_alarms} + \mathbf{misses} - \mathbf{CH}} \tag{7}$$

Where CH is given by:

$$\mathbf{CH} = \frac{(\mathbf{hits} + \mathbf{false\_alarms})(\mathbf{hits} + \mathbf{misses})}{\mathbf{hits} + \mathbf{false\_alarms} + \mathbf{misses} + \mathbf{correct\_rejections}} \tag{8}$$

2.6 Heat waves forecast strategy

In this study, we adapted the methodology developed by Lavaysse et al. (2019) to evaluate the predictability of heat waves in the forecast models. This assessment is done in two steps:

– The first approach consists to determine the predictability of heat wave days in the forecasts with respect to the reanalyses. To do this, dichotomous scores (hits, false alarms, misses and correct rejection) have been computed for each day at different lead times from which the hit-rate, FAR and GSS were derived. The scores are computed independently for each week and each city. This type of information is useful for early warning systems such as the STEWARd project.

– The second approach focuses on the predictability of the whole heat wave rather than hot days in the heat wave. Therefore, a heat wave occurring in the reanalyses is considered as predicted by the forecast models if at least one hot day during the event is correctly predicted. To facilitate the evaluation, dichotomous scores are now computed for each week, and the hit-rate, FAR, GSS are derived. For example, if a heat wave is detected in the reanalysis during the first week, and a hot day within this heat wave is correctly predicted by the forecast model, then the heat wave is considered as predicted and the hit=1 and miss=0. We repeat the process over all weeks and calculate the hit-rate.

3 Results

3.1 Assessment of skills in probabilistic forecasts

3.1.1 Climatology evolution of some atmospheric variables (T2m and Tw) in the forecasts and reanalyses

In this section, we investigated the representation of the climatological evolution of two key variables for this study, T2m and Tw in the hindcasts with respect to the reanalyses. To do so, the seasonal climatological bias between the hindcasts and the reanalyses is computed over the 20-year period. The bias is computed for each couple of models (ECMWF/UKMO) and reanalysis (ERA5/MERRA) using T2m(min,max) and Tw values [Fig.3] and [Fig.4] respectively).

Firstly, we analysed the bias between the forecast models and ERA5 using T2m_min values. The range of biases associated with T2m spans from -4 K to 4 K. Both models show a negative bias with respect to ERA5 over the Sahel region which is more pronounced with UKMO [Fig.3(i)]. We noticed with UKMO, a progressive shift of this strong negative bias over the northern Sahel during the season from winter to summer [Fig.3 (i)(e-g)]. The negative bias found in ECMWF over the Sahel region is consistent with previous global studies [e.g. Johnson et al. (2019); Haiden et al. (2021)]. The results obtained with T2m_max values are similar to those found with the T2m_min for both models. We notice especially a strong positive bias over the Atlantic coast during the winter and spring [Fig.3 (ii)(a-b)/(e-f)]. The evaluation of the representation of Tw in the hindcasts shows strong negative bias with respect to the reanalyses over the whole Sahel region [Fig.4]. In comparison with the results obtained with T2m, this strong negative bias can be linked mainly to the underestimation of humidity in the models in this region.

Secondly, we analysed the bias between the models and MERRA using T2m and Tw. We noticed significant discrepancies with the results obtained using ERA5 as a reference. Using T2m_min, we observed a positive bias in ECMWF with respect to MERRA over the Sahel and Guinea region during winter and autumn which tends to decrease during spring and summer. These results highlight the uncertainties between the two reanalyses already discussed in Ngoungue Langue et al. (2023). It

can be deduced from these results using T2m_min that ERA5 has a warmer trend than MERRA over the Sahel region, while being cooler than MERRA over the Atlantic Ocean (see [Fig.S3] in supplement material).

The spatial distribution of T2m and Tw in the forecast models shows significant biases with respect to the reanalyses. In order to determine whether these biases vary over time, we analyze the spatio-temporal evolution of T2m and Tw. The daily climatological biases is computed between the hindcasts and reanalyses over the CO, AT and GU regions at different lead times from week1 to week6. We observed a high spatial variability of the biases in the three regions. The models show smaller biases in the AT region in spring, of the order of +/- 0.25 K and larger biases, of the order of +/- 2 K are found in the CO region with T2m_min [Fig.S4]. These large biases observed in the CO region are considerably reduced in summer. The results obtained with T2m_max are quite similar to those obtained with T2m_min (not shown). As observed in the previous results with Tw [Fig.4], the models show lower negative biases in the CO region in winter ([Fig.S5] in supplement). The range of biases associated with Tw spans from -14 K to 0 K. We do not observe any systematic increase in these biases from Week1 to Week6 ([Fig.S4] and [Fig.S5] in supplement). This first assessment of the evolution of T2m and Tw in the models compared to the reanalyses reveals significant biases in the models which may lead to poor predictive skills.

3.1.2 Predictive skills of the models for T2m and Tw

This section addresses a global assessment of the skill of the models on the representation of T2m and Tw at different time scales. To do this, we analyzed the interannual variability of the CRPS between the forecasts and reanalyses over the 3 regions for T2m_min, T2m_max and Tw. The CRPS values are in the same range for the first two weeks (Week1, Week2), the two intermediate weeks (Week3, Week4) and the last two weeks (Week5, Week6) (see [Fig.S6] in supplement). According to these findings, we have chosen to organize our results into medium-range forecasts (Week2) and long-range forecasts (Week5). Week 5 was chosen for the long-range forecasts instead of week 6 because from week 5 onwards, the models generally reach the predictability horizon and are closer to the climatology. This organization of the results applies to the rest of the study.

The results of the CRPS computed using ERA5 as reference are shown in [Fig.5 (i)]. We have noticed that the skill of the models does not necessarily improve as the lead time decreases. This could be related to systematic biases in the forecast models (the representation of atmospheric circulation, local scale processes). The CRPS score shows a high spatial variability in the 3 regions, indicating a high dependance of the skill of the models with the region. The forecast models show better predictive skills in the AT region during all the seasons and lead times with T2m_min except for UKMO during summer. The CRPS associated with T2m_max indicates that ECMWF is more skillful in the GU region (it is also the case with UKMO except during Autumn when the AT region seems to be more predictable). In general, the models present higher skills for T2m_min than T2m_max over the AT region. ECMWF generally shows more skill than UKMO. The evolution of the CRPS computed using MERRA as a reference is similar to that found using ERA5 (not shown).

We observed in the previous analyses [Fig.4] a strong negative bias in the Sahel region with Tw, one could expect high CRPS values in the region. This is indeed the case, with CRPS values ranging from 4 to 13, i.e. 6 times higher than those obtained

with T2_min or T2m_max. The CRPS values are drastically reduced in the CO region in winter, which is consistent with the previous results in section 3.1 (see [Fig.5 (ii)] in supplement). The models show more predictive skill on the representation of T2m than Tw. This is not surprising, given that Tw combines T2m and humidity, which is very difficult to predict.

3.1.3 Predictive skills of the models on extreme temperatures

The representation of extreme heat events referred here as hot days (see section 2.4.3) is evaluated in this section. This is done by the computation of the Brier score between the forecasts and reanalyses (ERA5, MERRA). The Brier score values obtained using ERA5 reanalysis as reference, are very low between 0.05 to 0.175. We could think that the models show skills in forecasting hot days but this is quite difficult to affirm because the brier score is sensitive to the climatological frequency of an event: the more rare an event, the easier it is to get a good BS without having any real skill (https://www.ecmwf.int/sites/default/files/elibrary/2007/15489-verification-probability-forecasts.pdf). This is actually the case with hot days which represent extreme temperatures. Using T2m, we found that in spring the models show less spatial variability than the other seasons [Fig.6 (i)]. ECMWF performs better than UKMO on the detection of hot days except for Tw in the GU region in summer over all lead times, regions and seasons [Fig.6 (ii)]. The results are similar with MERRA reanalysis (see [Fig.S7] in supplement material).

3.2 Spatial variability of heat wave characteristics in deterministic forecasts

After assessing the skills of the models in predicting hot days, we study the spatial variability of heat wave characteristics in the forecast models with respect to the reanalyses. The frequency and characteristics of heat waves were computed using T2m and Tw. The mean duration (resp. mean intensity) of heat waves was computed for each grid point as the sum of heat wave duration (resp. intensity) divided by the number of years affected by a heat wave during the period from 2001 to 2020. A first assessment of the representation of heat waves in the forecast models is done by computing the bias between the models and the reanalyses for heat wave frequency, duration and intensity. For this task, probabilistic forecasts are transformed into deterministic forecasts by computing the ensemble mean.

The spatial variability of heat waves frequency bias between the forecast models (ECMWF, UKMO) and ERA5 shows similar evolution for T2m_min and T2m_max [Fig.7]. The models overestimate the frequency of heat waves in spring and summer over the Sahel for ECMWF, and from the Sahel to the Guinean region for UKMO. This overestimation in heat wave frequency is well marked in UKMO, indicating the inaccurate representation of the daily variability of T2m in the model. The Sahel and Guinean regions exhibit a strong convective activity during spring and summer which is very complex to take into account in the models. The representation of convective processes in the two models is assessed in the discussion at section 4. Some discrepancies are observed in the spatial evolution of heat waves frequency when using Tw over the Guinea coast [Fig.8]; we noticed an underestimation in heat waves frequency over the Sahel and Guinea region in autumn.

In a second step, we assess the evolution of the number of heat wave days in the models versus ERA5. We observed a north-south gradient well established over West Africa and an overestimation of the number of heat wave days over the Guinea region for both T2m_min and T2m_max [Fig.9]. The north-south gradient is well marked in UKMO. This gradient observed in UKMO with T2m is also found in Tw and tends to strengthen from spring to autumn [Fig.10]. The main differences between T2m and Tw on the representaion of heat wave days are found with ECMWF over the Guinea region. In winter, for example, there is an underestimation of the heat waves duration associated with Tw in the ECMWF model while an overestimation is observed with T2m_min and T2m_max.

In a third step, the spatial evolution of heat waves intensity is evaluated. ECMWF generally tends to underestimate the intensity of heat waves over the Sahel region, while UKMO overestimates the intensity of events from the Sahel to the Guinean coast for T2m_min (see [Fig.S8] in supplement). This overestimation of heat wave intensity in UKMO with T2m_min is considerably reduced when using T2m_max. The evolution of heat wave intensity with Tw is similar to that observed with T2m_min, except in summer for the ECMWF (see [Fig.S9] in supplement). We found very similar patterns when computing the bias in intensity using MERRA reanalysis as reference for T2m_min, T2m_max and Tw (not shown).

The bias in heat waves duration using MERRA as reference shows some differences with the results obtained with ERA5 reanalysis for T2m_min and T2m_max. ECMWF shows mostly an underestimation of heat wave days from the Sahel to the Guinean region in spring and summer for T2m_min and T2m_max (see [Fig.S10] in supplemental material). In summer, UKMO shows a large negative bias over Senegal, Guinea, Mali and Cameroon which tends to extend over the east of Sahel in autumn for T2m_min (see [Fig.S10 (i)] in supplement). We do not find significant differences when using MERRA as reference for the evaluation of heat waves duration with Tw (not shown).

3.3 Assessment of the predictability of heat waves in deterministic forecasts

In this section, we investigated the skills of deterministic forecasts in predicting heat waves in the three regions. This is done by computing the following dichotomous scores described in section 2.5.3: the hit-rate, GSS and FAR (see Section 2). The scores are computed for T2m_min, T2m_max and Tw at daily and weekly time scales using the optimized forecasts (see Section 2) and ERA5 reanalysis chosen as reference. The evaluation of the forecast models at daily and weekly time scales provides useful insights for policy makers and climate services. Daily information is relevant for the accurate prediction of heat waves and early warning alert systems.

The results presented below are obtained using a 20% threshold value to optimize the ensemble forecast systems (see Section 2.5.2) [Fig.11]. The hit-rate and GSS values of the optimized forecasts using the thresholds 40% and 60% are lower than those obtained with the 20% threshold. The hit-rate values show a weak spatial variability in the 3 regions, which indicates that the skills of the models are not too sensitive to the topography and climatic characteristics of the three regions. We also noticed a gradual loss of predictability in the models from winter to autumn, with high hit-rate values in winter. The forecast models show skills above the climatology both for medium- and long- range forecasts (Week2, Week5) but the hit-rate values are very low

indicating misses in the forecasts. These results are not surprising, as we are interested in events that are extremely rare. The hit-rate values are slightly better for medium-range forecasts. ECMWF presents higher skills than UKMO for medium-range forecasts in winter. UKMO, for instance, is better for long-range forecasts, mainly for T2m_min [Fig.11 (i)(a-d)].

The models show higher skills in the AT and CO regions in winter and spring for T2m_min values for medium-range forecasts. In order to understand the behavior of the models in these two regions, we investigated the inter-day variability of T2m by computing the standard deviation (std) for each region using ERA5 reanalysis. We found small std values in the AT and GU regions, indicating low variability of daily T2m in these regions [Table5]. Conversely, high variability of daily T2m is observed in the CO region. The low inter-day variability of T2m in the AT region indicates a more stable signal which will lead to favorable conditions for heat wave detection in the models based on a statistical perspective. Statistically, this will contribute to the occurrence of heat waves, as the probability of having consecutive days above the threshold is higher in a stable signal than in one with high daily variability. The skills of the models previously highlighted in the AT region could be partly explained by this low variability of daily T2m (min, max) in the region. The hit-rate values obtained with Tw are very close to those associated with T2m in the different regions and seasons; but we also noticed with Tw, that UKMO is more skillful than ECMWF for medium- and long-range forecasts (see [Fig.S11(a-d)] in supplement material).

The second metric computed for the evaluation of the models is the GSS. The GSS follows the same evolution as the hit-rate with much lower values between 0 and 0.2 for T2m_min, T2m_max and Tw ([Fig.11 (i)(i-l)], [Fig.11 (ii)(i-l)] and [Fig.S11(i-l)] respectively). The highest values of the GSS are observed in winter. During winter, the atmospheric circulation in West Africa regions is mainly governed by the harmattan flow which results in low convective activity in the regions and therefore an improvement of the predictive skills of the models compared to summer. The GSS values are positive, indicating that the forecast models perform better than the climatology mainly for medium-range forecasts (Week2) [Fig.11(i-l)]. Similar results of the GSS are found with Tw (see [Fig.S11(i-l)] in supplement).

An important parameter of a forecast system is its reliability in predicting events: "Do the events predicted by the models always occur in the reanalysis?". This is done by the computation of the false alarm ratio between the forecast models and ERA5 reanalysis for T2m_min [Fig.11 (i)(e-h)], Tw [Fig.S11(e-h)] and T2m_max [Fig.11 (ii)(e-h)]. The FAR values are too high on average, about 0.7/0.85 for the medium/long -range forecasts (week 2/week 5), respectively for all the three indicators. This suggests that the models tend to overestimate the number of heat days with respect to ERA5. The FAR increases with the lead time and ECMWF issues fewer false alarms than UKMO for medium-range forecasts. The FAR is considerably reduced when we increase the threshold values used to optimize the ensemble forecasting systems (see [Fig.S12] in supplement for T2m_min). This is highlighted more clearly in ECMWF than UKMO over all seasons and regions. This result is consistent with the fact that the prediction of an event will be more robust as a large number of model members have predicted the event. However, we found that this is not the case for the hit-rate and GSS for which high values are obtained with lowest threshold (20%) (see [Fig.S13 and Fig.S14] in supplemental material). Thus, according to the context of the study, a compromise must be found in order to obtain a good balance between hit-rate and false alarms.

The skill of the models in detecting heat waves is also assessed on a weekly time scale. As expected, we found an overall increase in the values of hit-rate and GSS for T2m_min, T2m_max and Tw (see [Fig.S15] for T2m_min in supplement).

The predictability of heat wave intensity in the models is also assessed for medium- and long- range forecasts (Week2 and Week5) using ERA5 reanalysis. We found a large spatial variability in heat wave intensity across regions. The strongest heat waves are found in the CO region (as shown in Ngoungue Langue et al. (2023)) and low intensity heat waves in the GU region except for Tw in summer for all the three indicators. Although the models demonstrate consistency in forecasting the spatial distribution of heat wave intensity, they fail in accurately predicting the specific intensity values. The models underestimate the intensity of heat waves (see [Fig.12 (i)] for T2m_min, the results are very similar for T2m_max [Fig.12 (ii)] and Tw (not shown)). The forecast of heat wave intensity in the models remains a difficult task, even though the models show skills in heat wave detection.

4 Discussion

4.1 On the differences between the behaviour of forecasting models in different climatic regions

The behavior of the forecast models varies across different climatic regions. The question that arises is : Why do these differences exist? In order to tackle this question, we analyze some key factors in the models such as the physical parameterizations, the repesentation of the atmosphere-ocean coupling and the spatial resolution of the atmospheric model.

First of all, the physical parameterizations used to simulate atmospheric processes such as convection, turbulence, interactions surface-ocean, surface-radiation and cloud microphysics are different for the two models. For example, for the representation of the convective activity, ECMWF is using the Tiedtke scheme (Tiedtke, 1989) and UKMO, the Met Office convective scheme (Hagelin et al., 2017). The Tiedtke convection scheme is one of the first mass-flow convection schemes, which aims to parameterise the effects of deep convection in numerical weather models. It simulates the vertical transport of heat, moisture and momentum associated with convective updrafts and downdrafts. The system takes into account various factors, including atmospheric instability, moisture content and boundary layer conditions to estimate convective processes. UKMO also uses a mass flux convection scheme, but different from the Tiedtke scheme, which takes into account atmospheric instability and moisture content to determine convective activity. The difference between the two convective schemes could lead to a wrong representation of convective activity in the region, and thus limit the predictive skills of the models mostly for wet heat waves.

Secondly, the models use the same data assimilation methods (4D-Var) for control analyses, but the data and initial conditions are completely different. ECMWF assimilates a wide range of global and regional observational data, including satellite,radar and ground-based measurements. The UKMO focuses on observation data relevant to the United Kingdom and surrounding regions. ECMWF atmospheric model is coupling with the NEMO3.4.1 ocean model, while UKMO uses the NEMO3.6 ocean model. The two systems are not using the same atmospheric and ocean models, which implies different parameterisations.

The differences observed in the representation of T2m_min over the Atlantic ocean [Fig.3] could result from representation of surface-ocean interactions in the models.

Thirdly, the spatial resolution of the atmospheric component of the two models is different: ECMWF has a higher spatial resolution than UKMO (0.32°x0.32° Vs 0.83°x0.56°), which means that it can capture local-scale variability or atmospheric processes and provide more accurate forecasts for specific regions(https://confluence.ecmwf.int/display/S2S/ECMWF+Model). Even if we transform the native resolution of the two models into a regular 0.25°x0.25° grid, some local-scale patterns will be found in the new grid. This may explain why the ECMWF performs better than the UKMO in some regions, for example in the case of extreme temperatures over AT and GU regions [Fig.6]. A more detailed analysis of the influence of each factor on the results is beyond the scope of this paper.

4.2 On the skills of the subseasonal models at medium and long -range forecasts

We noticed that in some regions, the CRPS score is slightly better during week 5 than in week 2; which is rather surprising. This behavior is more apparent with UKMO; it is indeed the case in the Guinea region in January and February (see [Fig.S16] in supplement material). One first hypothesis on this behavior in the UKMO model could be the strong seasonality in the region. Following this hypothesis, we investigate the seasonal evolution of the spread of temperatures in the Guinea region over the lead times. This investigation is very interesting because the variability of the spread can be linked to the skills of a forecasting model. We noticed higher spread values in Winter compared to the rest of the seasons (see [Fig.S17] in supplement material). However, when we looked at week 2 and week 5, we did not find a specific evolution of the spread which can explain the behavior observed in UKMO. The second hypothesis is the presence of a bias in UKMO which decreases over the lead times. Therefore, we analyzed the evolution of the bias of temperatures over lead times. We found an intense cold bias of UKMO with respect to ERA5 over the Guinea region during winter. This intense cold bias could lead to high CRPS values in the region during winter (see [Fig.S18] in supplement material). The evolution of the bias during week 2 and week 5 shows similar patterns in January and February. This hypothesis on the bias evolution is not supported by these findings. One can suggest that atmospheric conditions and physical processes in the region during week 5 are well represented in UKMO compared to week 2. The investigation in more detail of the origins of this behavior in the UKMO model is very complex, and outside the scope of the present study.

5 Conclusions

This study is a first assessment of the predictability of heat waves in West African cities on a subseasonal time scale. Two models that are part of the S2S prediction project, namely: ECMWF and UKMO, were evaluated using two state-of-the-art reanalysis data : ERA5 and MERRA over the period 2001-2020.

To carry out this study, we first analyzed the representation of T2m and Tw in the forecast models with respect to the reanalyses. We found that the models are cooler than ERA5 in the Sahel region, and hotter than MERRA in the Sahel and Guinea region. These uncertainties between the 2 reanalyses have also been highlighted in Ngoungue Langue et al. (2023).

Secondly, we assessed the spatial variability of heat wave characteristics. The models faced some issues in reproducing properly the spatial variability in the frequency, duration and intensity of heat waves. We found an overestimation of the frequency of heat waves in spring and summer over the Sahel for ECMWF, and from the Sahel to the Guinean region for UKMO when using T2m. With Tw, we noticed an underestimation of heat waves frequency over the Sahel and Guinea region in autumn.

Thirdly, we investigated the predictability of heat waves in the models with respect to the reanalyses using dichotomous scores (hit-rate, GSS (Gilbert Skill score) and FAR (False Alarm Ratio)). The hit-rate and GSS values are very weak, while the FAR are higher. On average, only approximately 15% to 30% of the predicted heat wave days are actually observed for Week 5 and Week 2, respectively. This suggests that the models overestimate the duration of the heat waves with respect to ERA5. ECMWF issues fewer false alarms than UKMO for medium-range forecasts. The detection of dry heat waves is slightly better with ECMWF for medium-range forecasts (week 2) during winter and spring, while it is better with UKMO for long-range forecasts (week 5) in the 3 regions. As far as wet heat waves are concerned, UKMO outperforms ECMWF for medium and long- range forecasts. We are aware that the hit-rate and GSS values are very low, but this is not surprising given that heat waves are extremely rare and difficult to predict because the persistence factor comes into play. Furthermore, we also know that the forecasting models underperform in tropical regions due to a poor representation of convective processes in their physical parameterization. Consequently, these scores, which are low but greater than the climatology, are significant for assessing the skill of the models in predicting heat waves in tropical regions which remains a complex task. However the score are significant, the models underestimate the intensity of heat waves with respect to ERA5 at short-, medium- and long- range forecasts.

Regarding these results, we can recommend the use of subseasonal forecasts in African cities to predict the onset and frequency of heat waves, and some days during the heat waves period at least to two weeks in advance, but as far as their intensity is concerned, it is still challenging. Such informations are very useful for the population, hospitals and decision-makers in order to develop some adaptation strategies to reduce the impacts of heat waves in the region.

In future work, we will investigate in more detail the origins of the differences observed in the two forecast models over the different regions. It has been shown recently in some studies that machine learning techniques can be useful to extend the predictability range of weather forecasts [e.g. Salcedo-Sanz et al. (2016); Anjali et al. (2019); Azari et al. (2022); van Straaten et al. (2023)]. Based on these findings, it would be interesting to investigate the potential of machine learning algorithms on heat waves forecasting.

Competing interests. The contact author has declared that none of the authors has any competing interests

References

Alduchov, O. A. and Eskridge, R. E.: Improved Magnus Form Approximation of Saturation Vapor Pressure, 35, 601–609, https://doi.org/10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2, publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, 1996.

Anderson, B. G. and Bell, M. L.: Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States, Epidemiology, 20, 205–213, 2009.

Anjali, T., Chandini, K., Anoop, K., and Lajish, V.: Temperature prediction using machine learning approaches, in: 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), vol. 1, pp. 1264–1268, IEEE, 2019.

August, E. F.: Ueber die Berechnung der Expansivkraft des Wasserdunstes, 89, 122–137, https://doi.org/10.1002/andp.18280890511, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18280890511, 1828.

Azari, B., Hassan, K., Pierce, J., and Ebrahimi, S.: Evaluation of machine learning methods application in temperature prediction, Environ Eng, 8, 1–12, 2022.

Barbier, J., Guichard, F., Bouniol, D., Couvreux, F., and Roehrig, R.: Detection of Intraseasonal Large-Scale Heat Waves: Characteristics and Historical Trends during the Sahelian Spring, 31, 61–80, https://doi.org/10.1175/JCLI-D-17-0244.1, publisher: American Meteorological Society Section: Journal of Climate, 2018.

Batté, L., Ardilouze, C., and Déqué, M.: Forecasting West African Heat Waves at Subseasonal and Seasonal Time Scales, 146, 889–907, https://doi.org/10.1175/MWR-D-17-0211.1, publisher: American Meteorological Society Section: Monthly Weather Review, 2018.

Bazo, J., Singh, R., Destrooper, M., and de Perez, E. C.: Pilot experiences in using seamless forecasts for early action: The "ready-set-go!" approach in the Red Cross, in: Sub-seasonal to seasonal prediction, pp. 387–398, Elsevier, 2019.

Déqué, M., Calmanti, S., Christensen, O. B., Aquila, A. D., Maule, C. F., Haensler, A., Nikulin, G., and Teichmann, C.: A multi-model climate response over tropical Africa at+ 2° C, Climate Services, 7, 87–95, 2017.

Domeisen, D. I., White, C. J., Afargan-Gerstman, H., Muñoz, Á. G., Janiga, M. A., Vitart, F., Wulff, C. O., Antoine, S., Ardilouze, C., Batté, L., et al.: Advances in the subseasonal prediction of extreme events: relevant case studies across the globe, Bulletin of the American Meteorological Society, 103, E1473–E1501, 2022.

Engdaw, M. M., Ballinger, A. P., Hegerl, G. C., and Steiner, A. K.: Changes in temperature and heat waves over Africa using observational and reanalysis data sets, International Journal of Climatology, 42, 1165–1180, 2022.

Fischer, E. M. and Schär, C.: Consistent geographical patterns of changes in high-impact European heatwaves, Nature geoscience, 3, 398–403, 2010.

Gasparrini, A. and Armstrong, B.: The impact of heat waves on mortality, Epidemiology, 22, 68–73, 2011.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A. M. d., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, publisher: American Meteorological Society Section: Journal of Climate, 2017.

Guigma, K. H., Todd, M., and Wang, Y.: Characteristics and thermodynamics of Sahelian heatwaves analysed using various thermal indices, 55, 3151–3175, https://doi.org/10.1007/s00382-020-05438-5, 2020.

Guigma, K. H., MacLeod, D., Todd, M., and Wang, Y.: Prediction skill of Sahelian heatwaves out to subseasonal lead times and importance of atmospheric tropical modes of variability, Climate Dynamics, 57, 537–556, 2021.

Hagelin, S., Son, J., Swinbank, R., McCabe, A., Roberts, N., and Tennant, W.: The Met Office convective-scale ensemble, MOGREPS-UK, Quarterly Journal of the Royal Meteorological Society, 143, 2846–2861, 2017.

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, C., and Richardson, D.: Evaluation of ECMWF forecasts, including the 2020 upgrade, ECMWF, 2021.

Heo, S., Bell, M. L., and Lee, J.-T.: Comparison of health risks by heat wave definition: Applicability of wet-bulb globe temperature for heat wave criteria, Environmental research, 168, 158–170, 2019.

Hersbach, H.: The ERA5 Atmospheric Reanalysis., in: AGU Fall Meeting Abstracts, vol. 2016, pp. NG33D–01, 2016.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, 146, 1999–2049, https://doi.org/10.1002/qj.3803, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3803, 2020.

Huynen, M.-M., Martens, P., Schram, D., Weijenberg, M. P., and Kunst, A. E.: The impact of heat waves and cold spells on mortality rates in the Dutch population., Environmental health perspectives, 109, 463–470, 2001.

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, 1087–1117, 2019.

Kharin, V. V., Zwiers, F. W., Zhang, X., and Hegerl, G. C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, Journal of Climate, 20, 1419–1444, 2007.

Kovats, R. S. and Hajat, S.: Heat stress and public health: a critical review, Annu. Rev. Public Health, 29, 41–55, 2008.

Lala, J., Lee, D., Bazo, J., and Block, P.: Evaluating prospects for subseasonal-to-seasonal forecast-based anticipatory action from a global perspective, Weather and Climate Extremes, 38, 100 510, 2022.

Lavaysse, C., Cammalleri, C., Dosio, A., van der Schrier, G., Toreti, A., and Vogt, J.: Towards a monitoring system of temperature extremes in Europe, 18, 91–104, https://doi.org/10.5194/nhess-18-91-2018, publisher: Copernicus GmbH, 2018.

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.

Li, Y., Ding, Y., and Li, W.: Observed trends in various aspects of compound heat waves across China from 1961 to 2015, Journal of Meteorological Research, 31, 455–467, 2017.

Lowe, R., García-Díez, M., Ballester, J., Creswick, J., Robine, J.-M., Herrmann, F. R., and Rodó, X.: 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, 2016.

Magnus, G.: Versuche über die Spannkräfte des Wasserdampfs, 137, 225–247, https://doi.org/10.1002/andp.18441370202, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18441370202, 1844.

McGregor, G. R., Bessmoulin, P., Ebi, K., and Menne, B.: Heatwaves and health: guidance on warning-system development., WMOP, 2015.

Moron, V., Oueslati, B., Pohl, B., Rome, S., and Janicot, S.: 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, 2016.

Moron, V., Robertson, A. W., and Vitart, F.: Sub-seasonal to seasonal predictability and prediction of monsoon climates, 2018.

Ngoungue Langue, C. G., Lavaysse, C., Vrac, M., Peyrille, P., and Flamant, C.: Seasonal forecasts of the Saharan Heat Low characteristics: A multi-model assessment, Weather and Climate Dynamics, 2, 893–912, 2021.

Ngoungue Langue, C. G., Lavaysse, C., Vrac, M., and Flamant, C.: Heat wave monitoring over West African cities: uncertainties, characterization and recent trends, Natural Hazards and Earth System Sciences, 23, 1313–1333, 2023.

Osman, M., Domeisen, D., Robertson, A. W., and Weisheimer, A.: Sub-seasonal to decadal predictions in support of climate services, Climate Services, 30, 100 397, 2023.

Perkins, S. E.: A review on the scientific understanding of heatwaves—Their measurement, driving mechanisms, and changes at the global scale, 164-165, 242–267, https://doi.org/10.1016/j.atmosres.2015.05.014, 2015.

Perkins, S. E. and Alexander, L. V.: On the Measurement of Heat Waves, 26, 4500–4517, https://doi.org/10.1175/JCLI-D-12-00383.1, publisher: American Meteorological Society Section: Journal of Climate, 2013.

Perkins, S. E., Alexander, L. V., and Nairn, J. R.: Increasing frequency, intensity and duration of observed global heatwaves and warm spells, 39, https://doi.org/10.1029/2012GL053361, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2012GL053361, 2012.

Robinson, P. J.: On the Definition of a Heat Wave, 40, 762–775, https://doi.org/10.1175/1520-0450(2001)040<0762:OTDOAH>2.0.CO;2, publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, 2001.

Russo, S., Dosio, A., Graversen, R. G., Sillmann, J., Carrao, H., Dunbar, M. B., Singleton, A., Montagna, P., Barbola, P., and Vogt, J. V.: Magnitude of extreme heat waves in present climate and their projection in a warming world, Journal of Geophysical Research: Atmospheres, 119, 12–500, 2014.

Russo, S., Sillmann, J., and Sterl, A.: Humid heat waves at different warming levels, Scientific reports, 7, 7477, 2017.

Salcedo-Sanz, S., Deo, R., Carro-Calvo, L., and Saavedra-Moreno, B.: Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms, Theoretical and applied climatology, 125, 13–25, 2016.

Steadman, R. G.: The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science, 18, 861–873, https://doi.org/10.1175/1520-0450(1979)018<0861:TAOSPI>2.0.CO;2, publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, 1979a.

Steadman, R. G.: The Assessment of Sultriness. Part II: Effects of Wind, Extra Radiation and Barometric Pressure on Apparent Temperature, 18, 874–885, https://www.jstor.org/stable/26179217, publisher: American Meteorological Society, 1979b.

Stull, R.: Wet-Bulb Temperature from Relative Humidity and Air Temperature, 50, 2267–2269, https://doi.org/10.1175/JAMC-D-11-0143.1, publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, 2011.

Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, Monthly weather review, 117, 1779–1800, 1989.

Tompkins, A. M., Lowe, R., Nissan, H., Martiny, N., Roucou, P., Thomson, M. C., and Nakazawa, T.: Predicting climate impacts on health at sub-seasonal to seasonal timescales, in: Sub-Seasonal to Seasonal Prediction, pp. 455–477, Elsevier, 2019.

van Straaten, C., Whan, K., Coumou, D., van den Hurk, B., and Schmeits, M.: Correcting sub-seasonal forecast errors with an explainable ANN to understand misrepresented sources of predictability of European summer temperatures, Artificial Intelligence for the Earth Systems, pp. 1–49, 2023.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., et al.: The subseasonal to seasonal (S2S) prediction project database, Bulletin of the American Meteorological Society, 98, 163–173, 2017.

Wang, J., Chen, Y., Tett, S. F., Yan, Z., Zhai, P., Feng, J., and Xia, J.: Anthropogenically-driven increases in the risks of summertime compound hot extremes, Nature communications, 11, 528, 2020a.

Wang, J., Feng, J., Yan, Z., and Chen, Y.: Future risks of unprecedented compound heat waves over three vast urban agglomerations in China, Earth's Future, 8, e2020EF001 716, 2020b.

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J., Lazo, J. K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A. J., Murray, V., et al.: Potential applications of subseasonal-to-seasonal (S2S) predictions, Meteorological applications, 24, 315–325, 2017.

Yu, S., Tett, S. F. B., Freychet, N., and Yan, Z.: Changes in regional wet heatwave in Eurasia during summer (1979–2017), 16, 064 094, https://doi.org/10.1088/1748-9326/ac0745, publisher: IOP Publishing, 2021.

Tables

Table 1. Differences between the two forecasts models

Models	Hindcasts				Forecasts			
	dates	***size***	***range***	***period***	***dates***	***size***	***range***	***Model version***
ECMWF	2/week, on Monday and Thursday	11	0-46 days	past 20 years	2/week, on Monday and Thursday	51	0-46 days	CY47R3
UKMO	4/month on the 1^{st}, 9^{th}, 17^{th}, 25^{th}	3 prior 2016 7 from 25/03/2017	0-60 days	1993-2016	4/month 1^{st}, 9^{th}, 17^{th}, 25^{th}	4	0-60 days	GloSea5-GC2-LI

Table 2. Land sea mask (lsm) of west African towns used in this study

Towns	**latitude**	**longitude**	**lsm**
DAKAR	14.75	-17.25	0.6
ABIDJAN	5.25	-3.75	0.5
NOUAKCHOTT	18	-16	continent
CONAKRY	9.5	-13.5	0.5
MONROVIA	6.25	-10.75	0.6
BAMAKO	12.5	-8	continent
YAMOUSSOUKRO	6.75	-5.25	continent
OUAGADOUGOU	12.25	-1.5	continent
ACCRA	5.5	-0.5	0.8
LOMÉ	6	1	0.5
NIAMEY	13.5	2	continent
COTONOU	6.5	2.5	0.7
LAGOS	6.5	3.5	0.5
ABUJA	9	7.5	continent
DOUALA	4	9.75	0.9

Table 3. Description of the threshold values.

Threshold values	**Description**
20%	20% of the ensemble members are associated to a hot day
40%	40% of the ensemble members are associated to a hot day
60%	60% of the ensemble members are associated to a hot day

Table 4. Contingency table.

2X2 Contingency table		**Event Observed**	
		YES	N0
Event forecast	YES	hits	false alarms
	NO	misses	correct rejections

Table 5. Inter-daily variability of 2-meter temperature over the period 2001-2020 using ERA5 reanalysis during the seasons for T2m_min and T2m_max.

	AT				GU				CO			
	Win	Spri	Sum	Aut	Win	Spri	Sum	Aut	Win	Spri	Sum	Aut
T2m_min	0.5	0.5	0.43	0.77	0.57	0.36	0.44	0.44	1.64	1.74	1.08	1.69
T2m_max	0.38	0.35	0.43	0.58	0.39	0.69	0.76	1	1.71	0.81	2.04	1.66

Figures

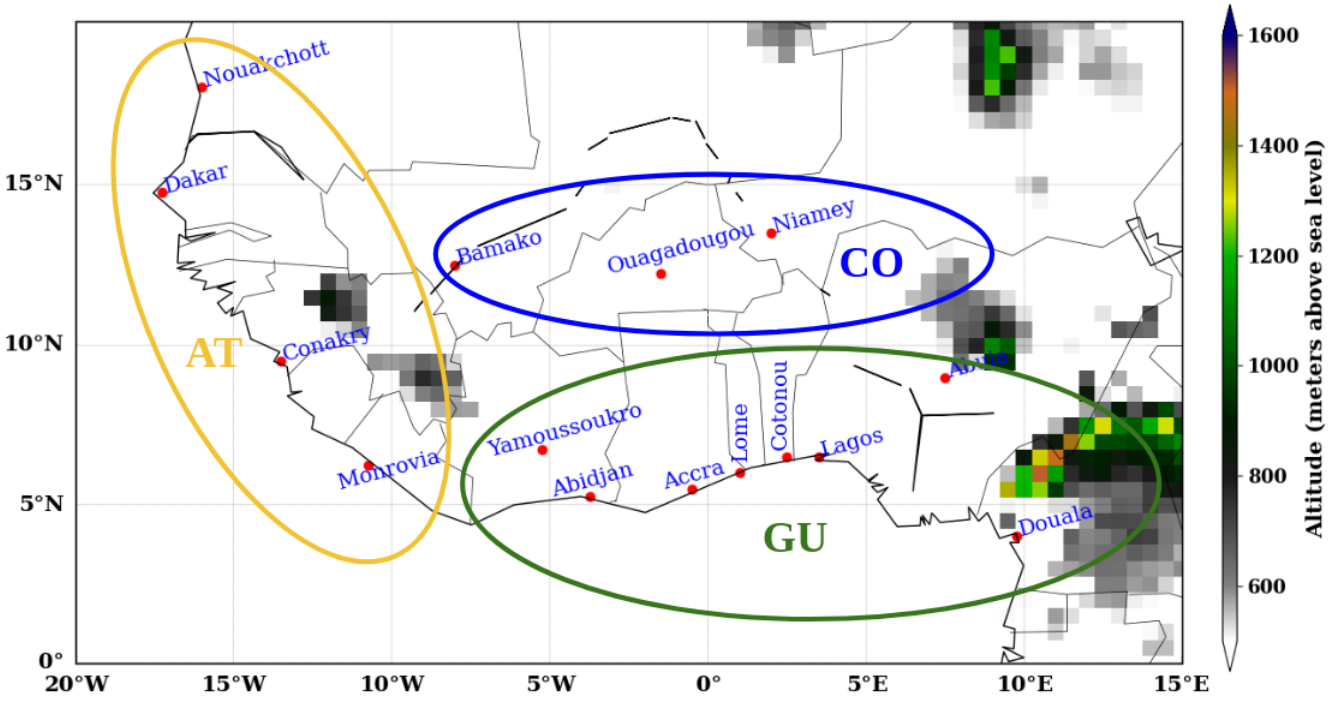


Figure 1. Topographic map of West Africa using ERA5 elevation data. The circles on the map represent the different climatic zones: AT (Coastal atlantic zone), CO (Continental zone) and GU (Coastal Guinean zone). The y and x axes represent the latitude and longitude respectively. The color bar shows the elevation in meters over the region.

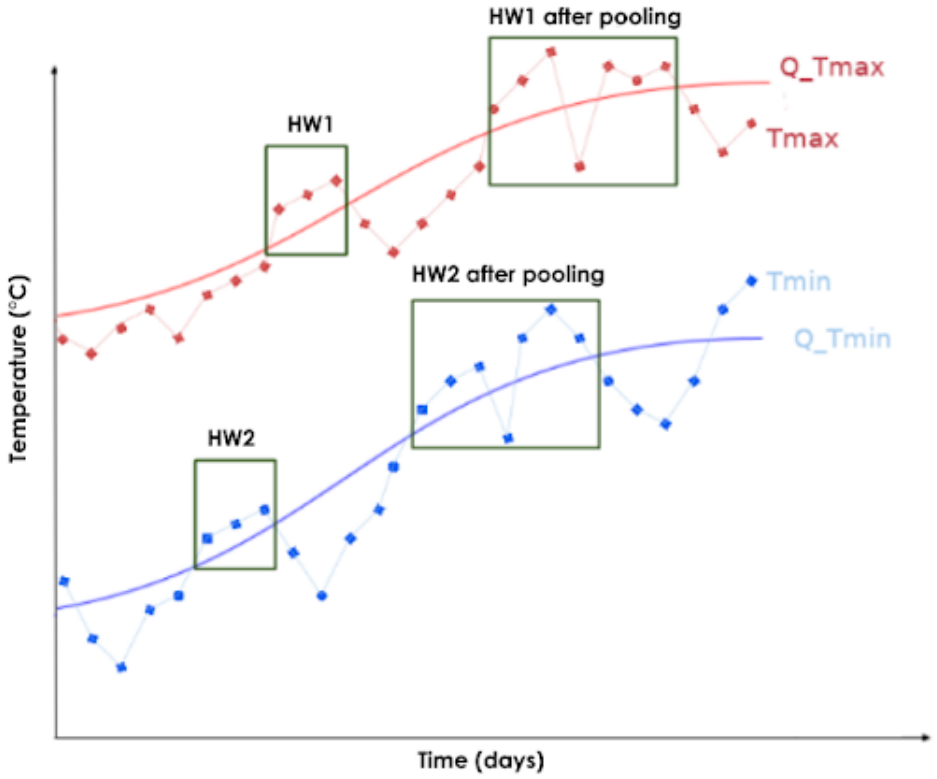


Figure 2. Detection process of heat wave: HW1/HW2 represent events associated respectively to maximum/minimum temperature. The red/blue lines with circles are maximum/minimum daily temperatures. Red/blue solid lines are respectively maximum/mininimum thresholds. X- and Y- axis represent the time in days and the temperature in degrees celsius. The term 'after pooling' refers to the pooling of two (or more) events separated by a day characterized by the value of a given indicator below the daily XX^{th} percentile. This figure is a 'theoretical/schematic' illustration of the different types of heat waves investigated in this work.

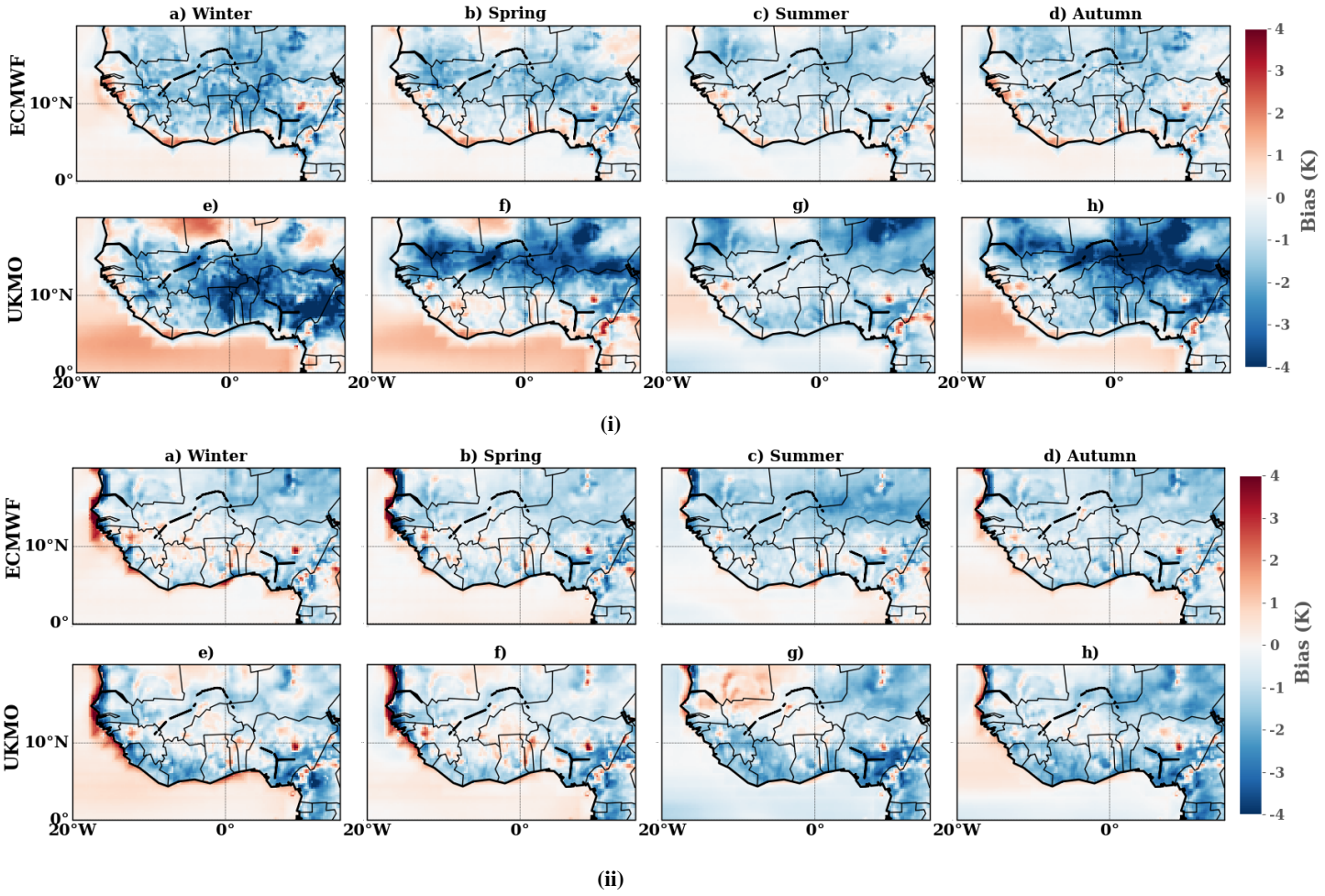


Figure 3. Spatial variability of the climatological bias between the forecast models ensemble mean and ERA5 reanalysis over the period 2001-2020 for : **(i)** T2m_min and **(ii)** T2m_max, during the seasons : (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is computed as the difference between the forecast models and ERA5 considering all the lead times. The color indicates the bias values in degrees Kelvin. The X and Y axes represent the longitude and latitude respectively.

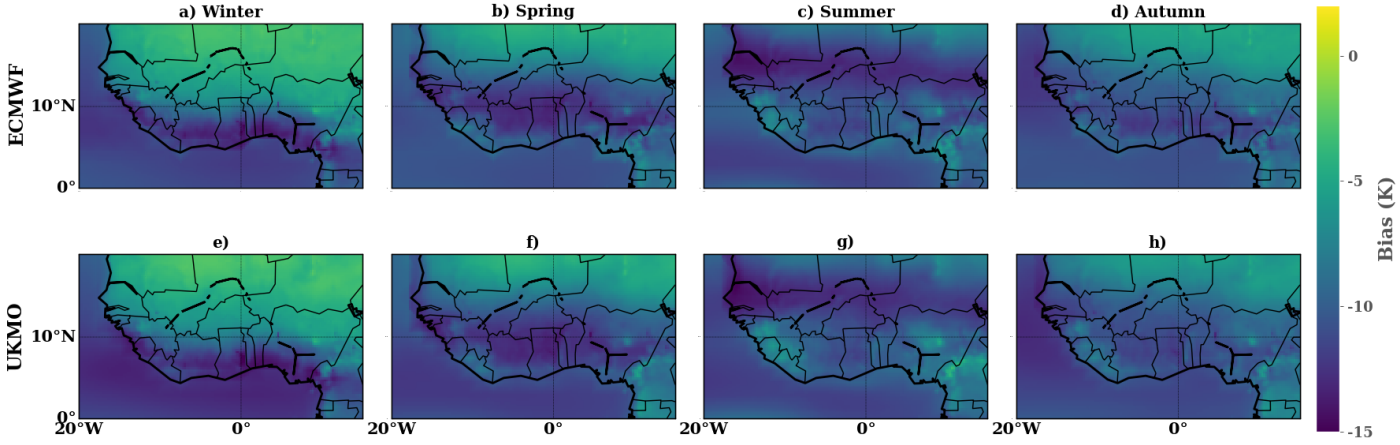
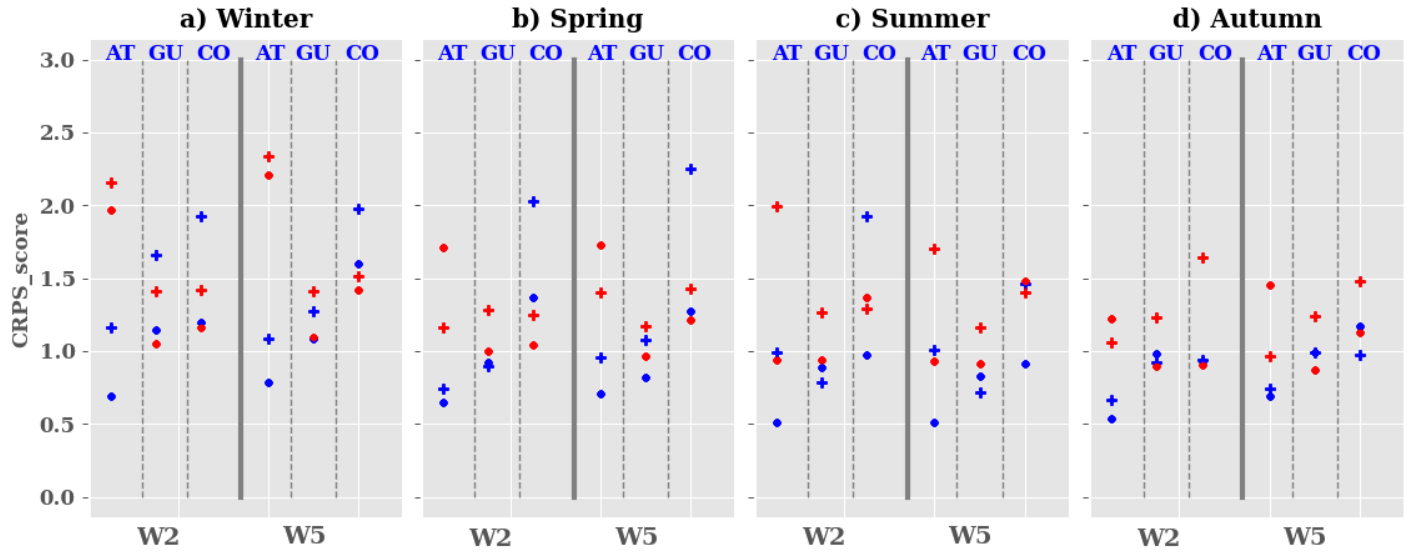
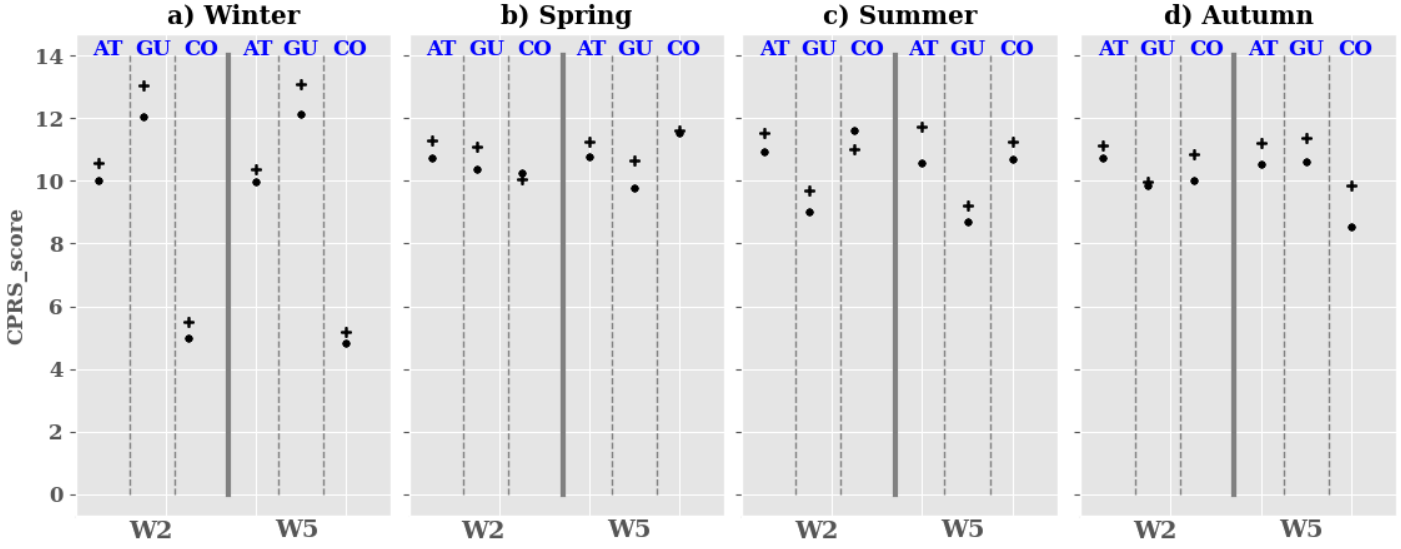


Figure 4. Spatial variability of the climatological bias between the forecast models ensemble mean and ERA5 reanalysis over the period 2001-2020 for Tw during the seasons : (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is computed as the difference between the forecast models and ERA5 considering all the lead times. The color indicates the bias values in degrees Kelvin. The X and Y axes represent the longitude and latitude respectively.

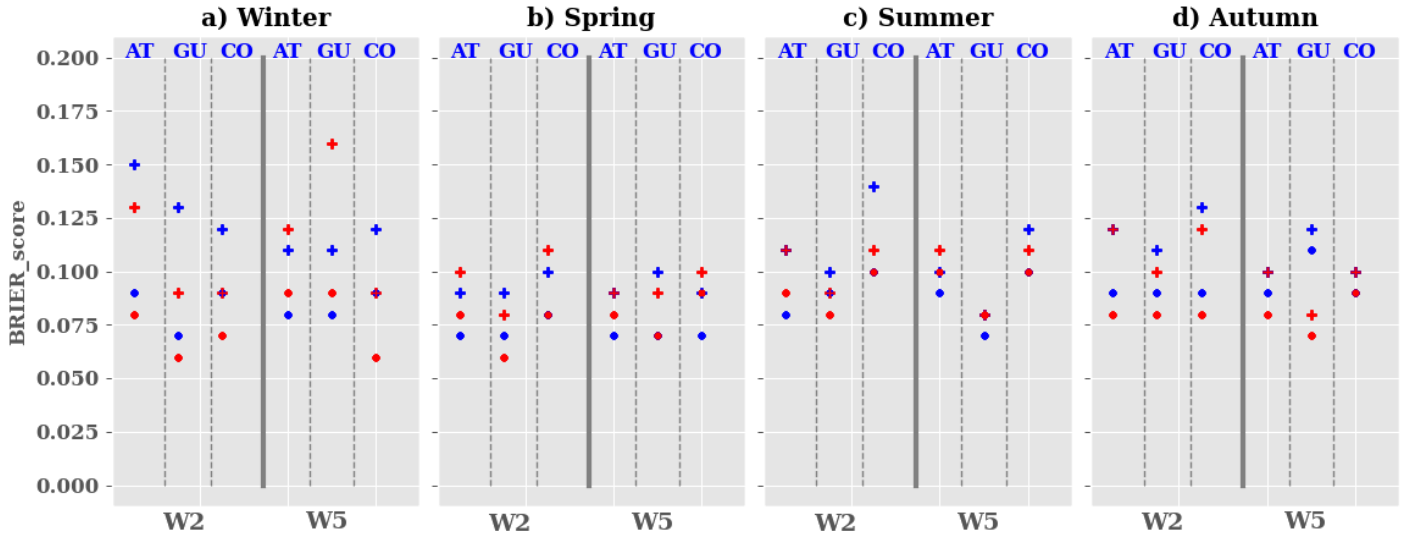


(i)

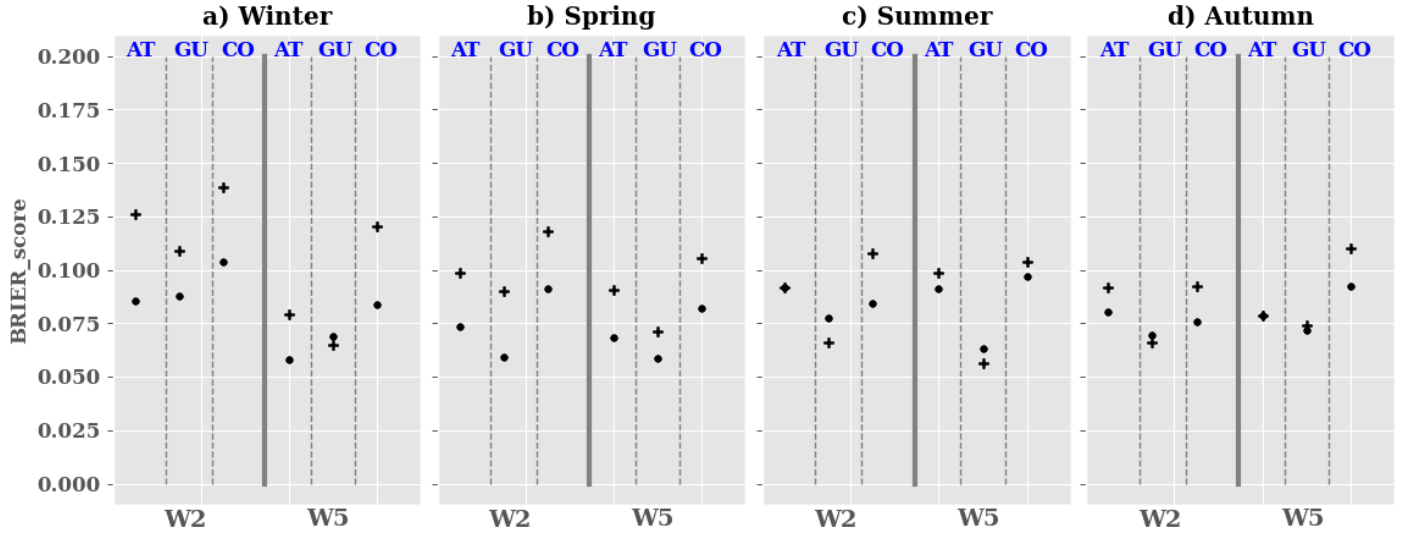


(ii)

Figure 5. Evolution of the CRPS score between the forecast models and ERA5 reanalysis using T2m **(i)** and Tw **(ii)** over the period 2001-2020 during the seasons : (a) winter, (b) spring, (c) summer and (d) autumn. The blue, red and black colors represent the CRPS score computed using T2m_min, T2m_max and Tw mean respectively. The CRPS is computed independently for each initialisation date within a month, and we computed the average CRPS to obtain the CRPS values for the given month. The dot and cross symbols indicate the CRPS score obtained with ECMWF and UKMO respectively. The Y and X axes show the CRPS values and the lead times (W2: week2 and W5: week5) respectively.



(i)



(ii)

Figure 6. Evolution of the Brier score between the models and ERA5 reanalysis using T2m **(i)** and Tw **(ii)** over the period 2001-2020 during the seasons : (a) winter, (b) spring, (c) summer and (d) autumn. The blue, red and black colors represent the Brier score computed using T2m_min, T2m_max and Tw mean respectively. The Brier is computed independently for each initialisation date within a month, and we computed the average Brier to obtain the Brier values for the given month. The dot and cross symbols indicate the Brier score obtained with ECMWF and UKMO respectively. The Y and X axes show the Brier values and the lead times (W2: week2 and W5: week5) respectively.

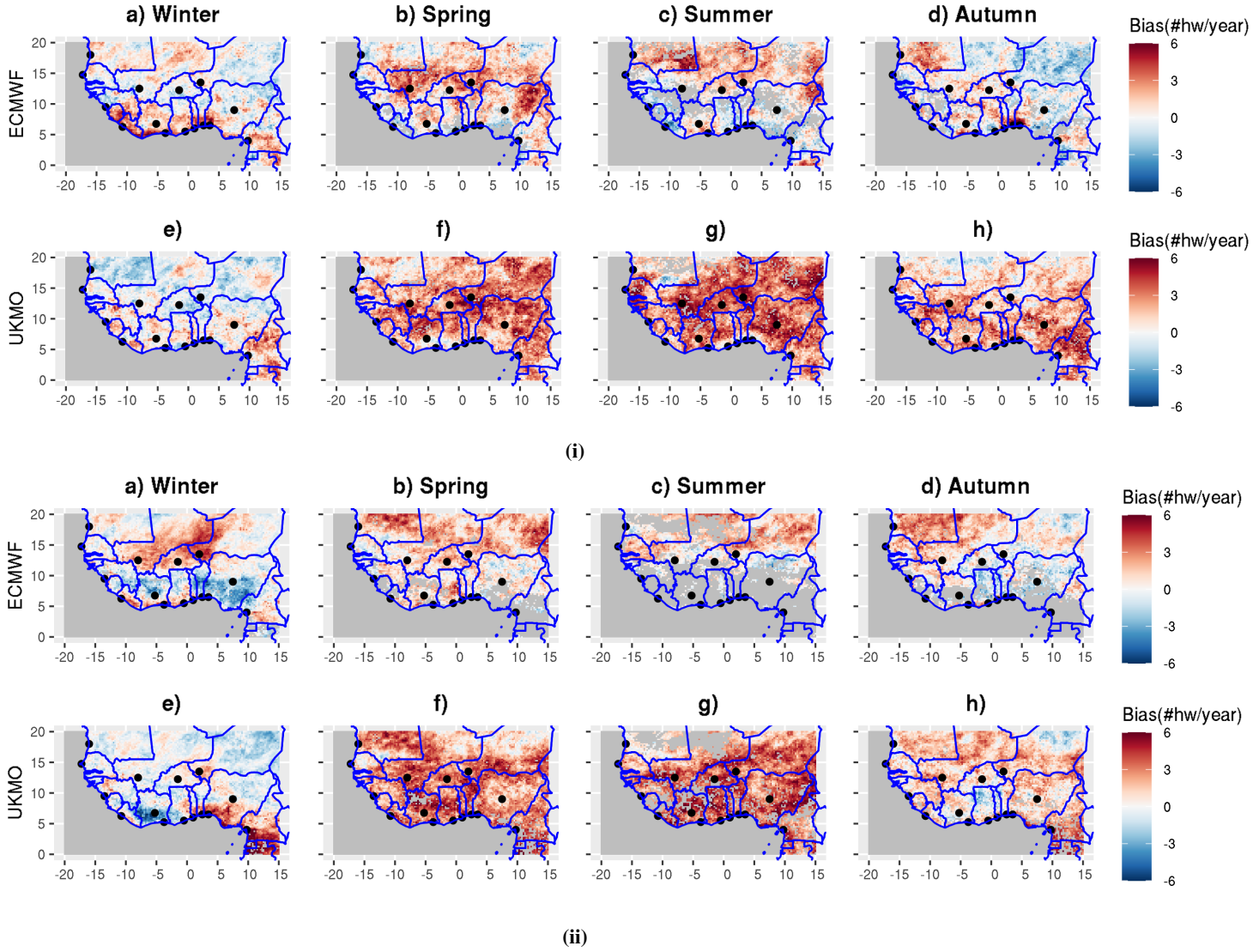


Figure 7. Spatial variability of heat wave frequency bias between forecast models and ERA5 over West Africa from 2001 to 2020 for: **(i)** T2m_min values and **(ii)** T2m_max values, during: (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is computed as the difference in heat wave frequency between the forecast models and ERA5. This analysis is performed using the unperturbed member of the models and the forecasts over all the lead times. The color bar indicates the bias values without units. The grey color represents missing values. The X and Y axes represent longitude and latitude respectively. The solid blue lines indicate the borders between countries; the black dots represent the cities of interest for this study (this applies to the rest of the paper).

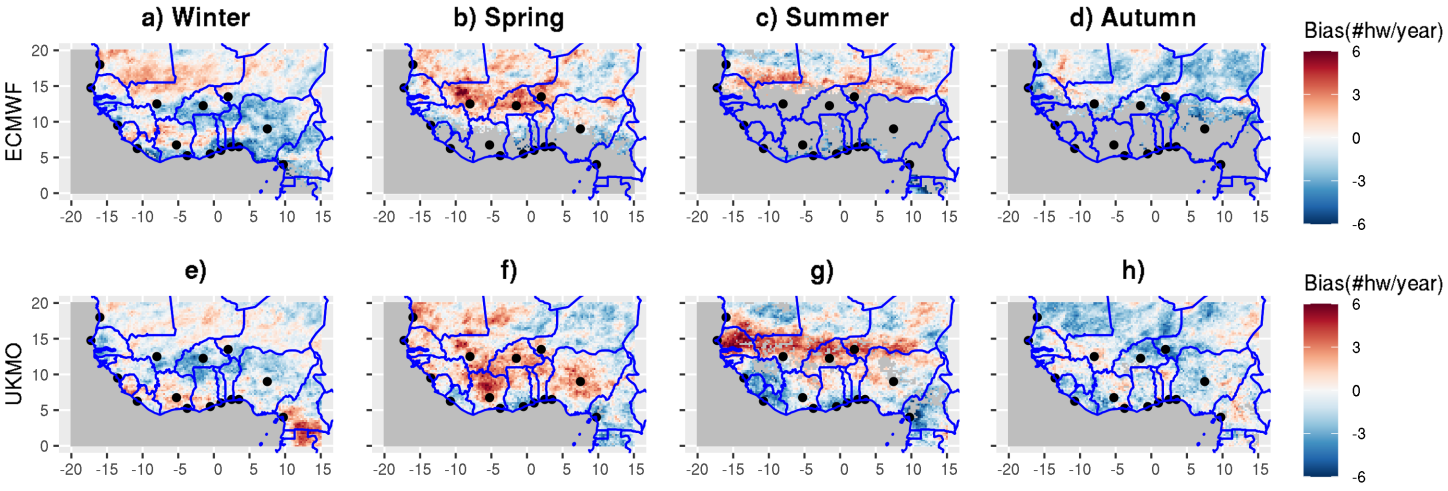


Figure 8. Spatial variability of heat wave frequency bias between forecast models and ERA5 over West Africa from 2001 to 2020 using Tw during: (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is computed as the difference in heat wave frequency between the forecast models and ERA5. This analysis is performed using the unperturbed member of the models and the forecasts over all the lead times. The color bar indicates the bias values without units. The grey color represent missing values. The X and Y axes represent longitude and latitude respectively.

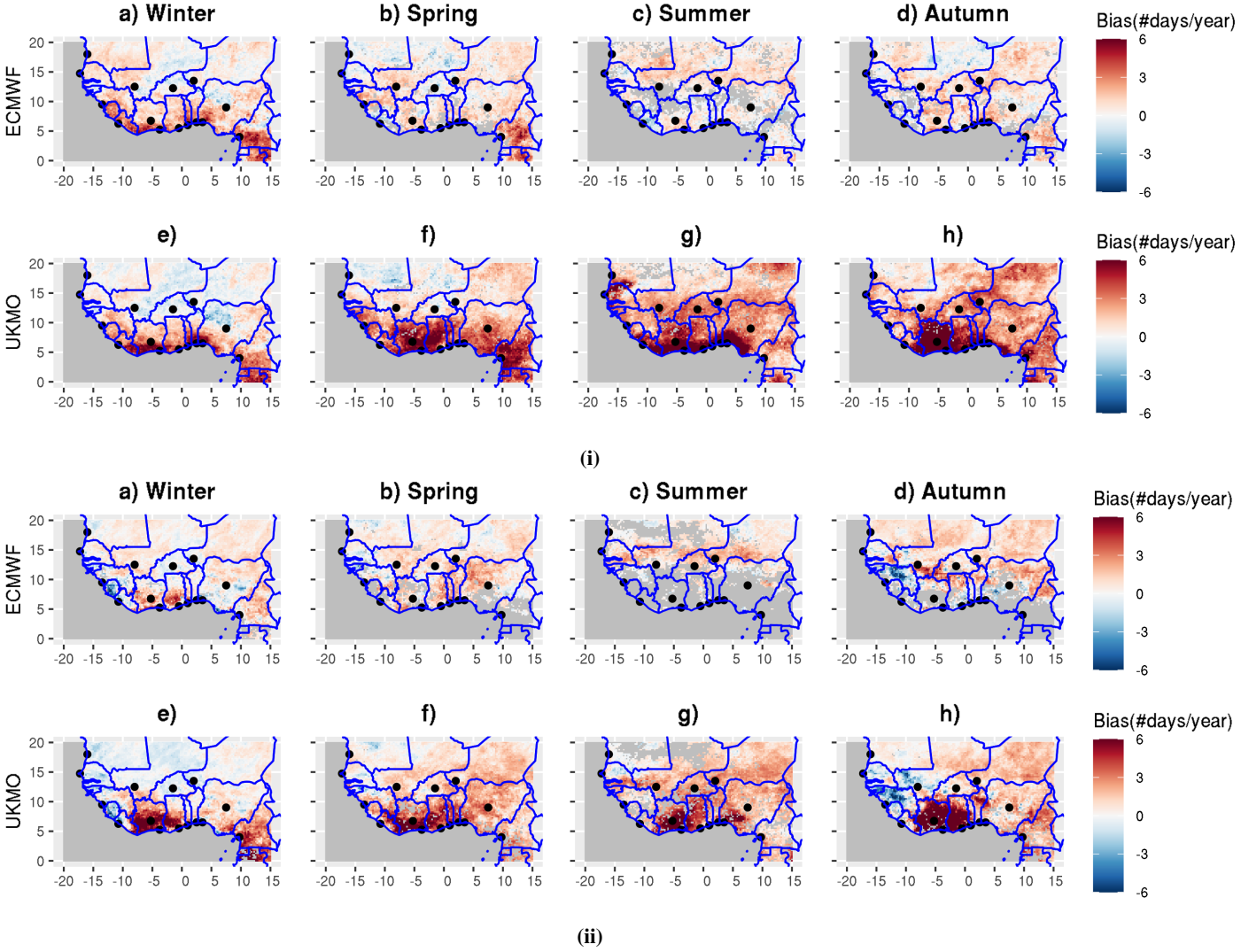


Figure 9. Spatial variability of heat wave duration bias between the forecast models and ERA5 over West Africa from 2001 to 2020 for: **(i)** T2m_min values and **(ii)** T2m_max, during: (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is computed as the difference in heat wave duration between the forecast models and ERA5. This analysis is performed using the unperturbed member of the models and the forecasts over all the lead times during years where heat waves were detected. The color bar indicates the bias values without units. The X and Y axes represent longitude and latitude respectively.

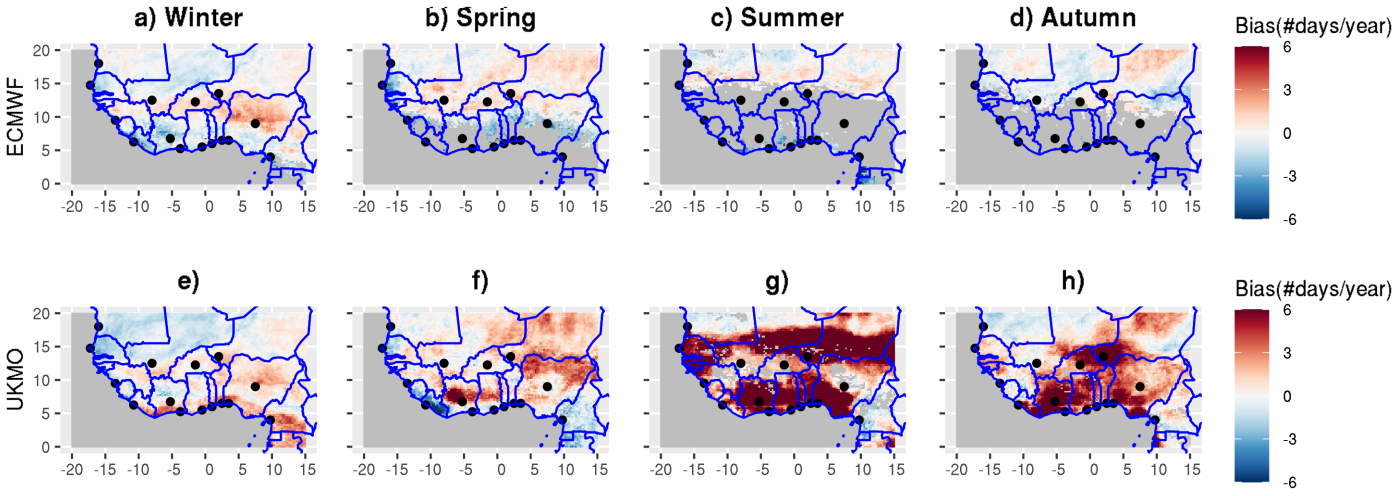
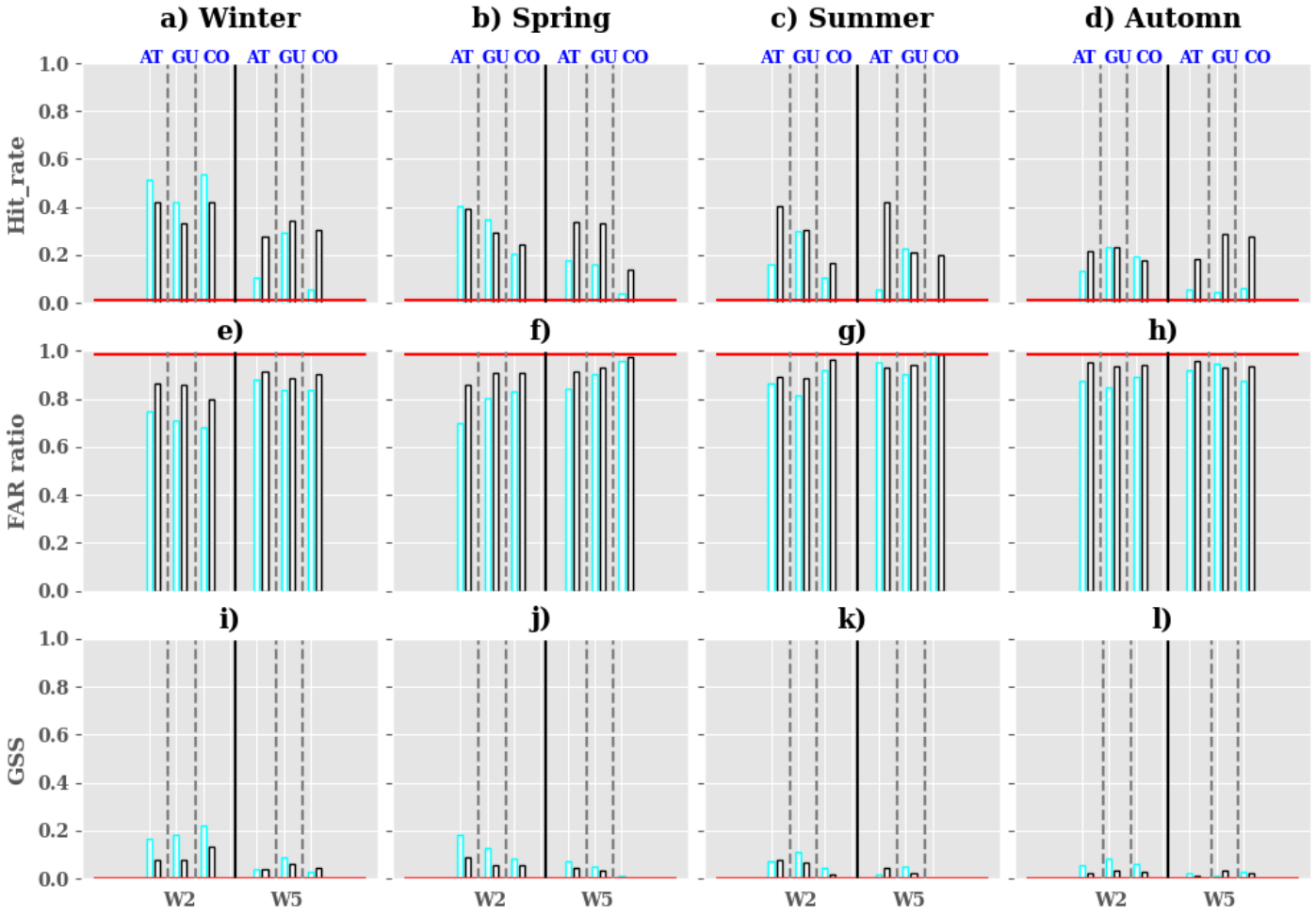


Figure 10. Spatial variability of heat wave duration bias between forecast models and ERA5 over West Africa from 2001 to 2020 using Tw during: (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is computed as the difference in heat wave duration between the forecast models and ERA5. This analysis is performed using the unperturbed member of the models and the forecasts over all the lead times during years where heat waves were detected. The color bar indicates the bias values without units. The X and Y axes represent longitude and latitude respectively.


a) Winter
b) Spring
c) Summer
d) Automn
AT GU CO AT GU CO
Hit_rate
FAR ratio
GSS
e)
f)
g)
h)
i)
j)
k)
l)
1.0
0.8
0.6
0.4
0.2
0.0
W2
W5

(i)

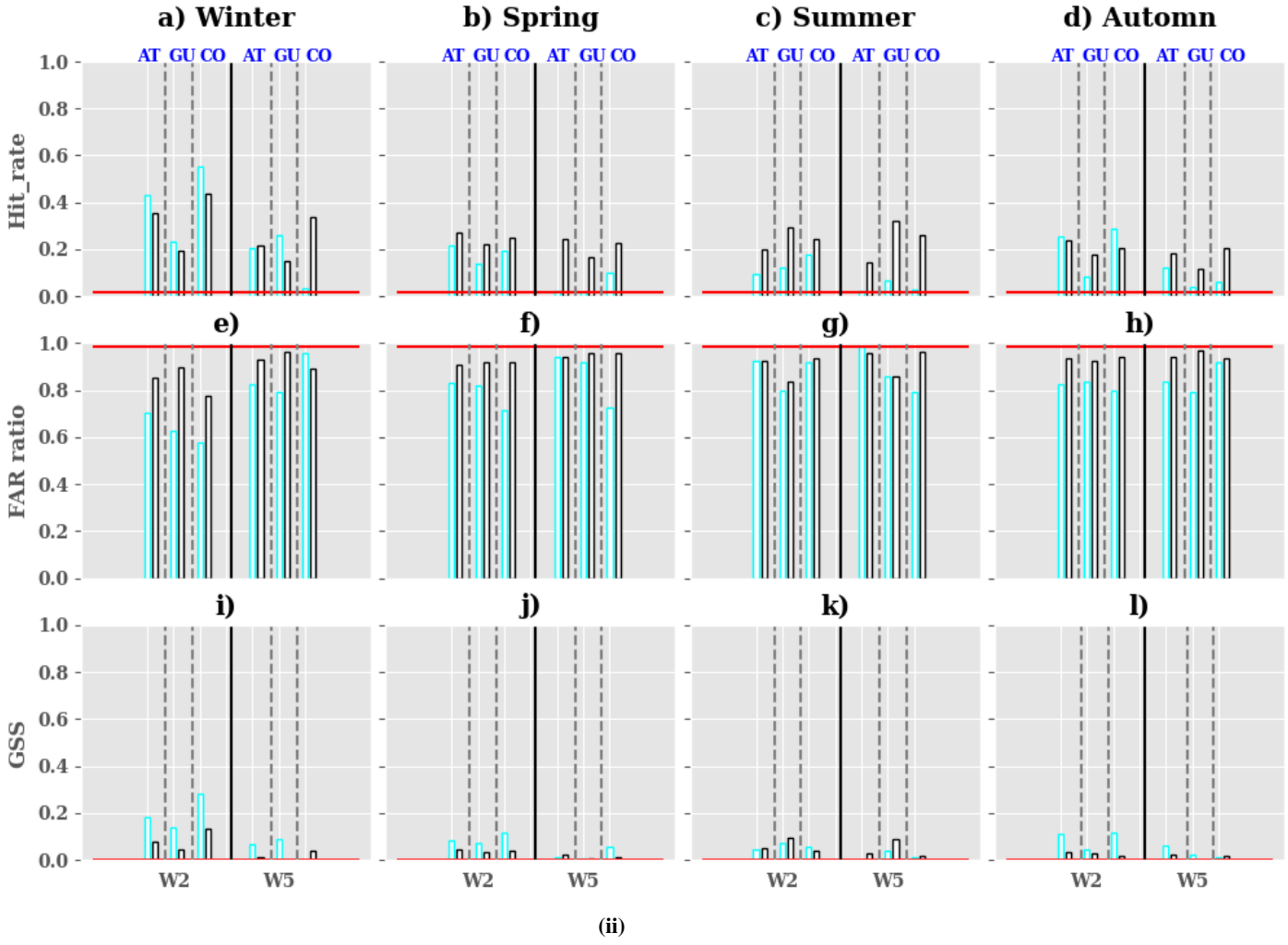
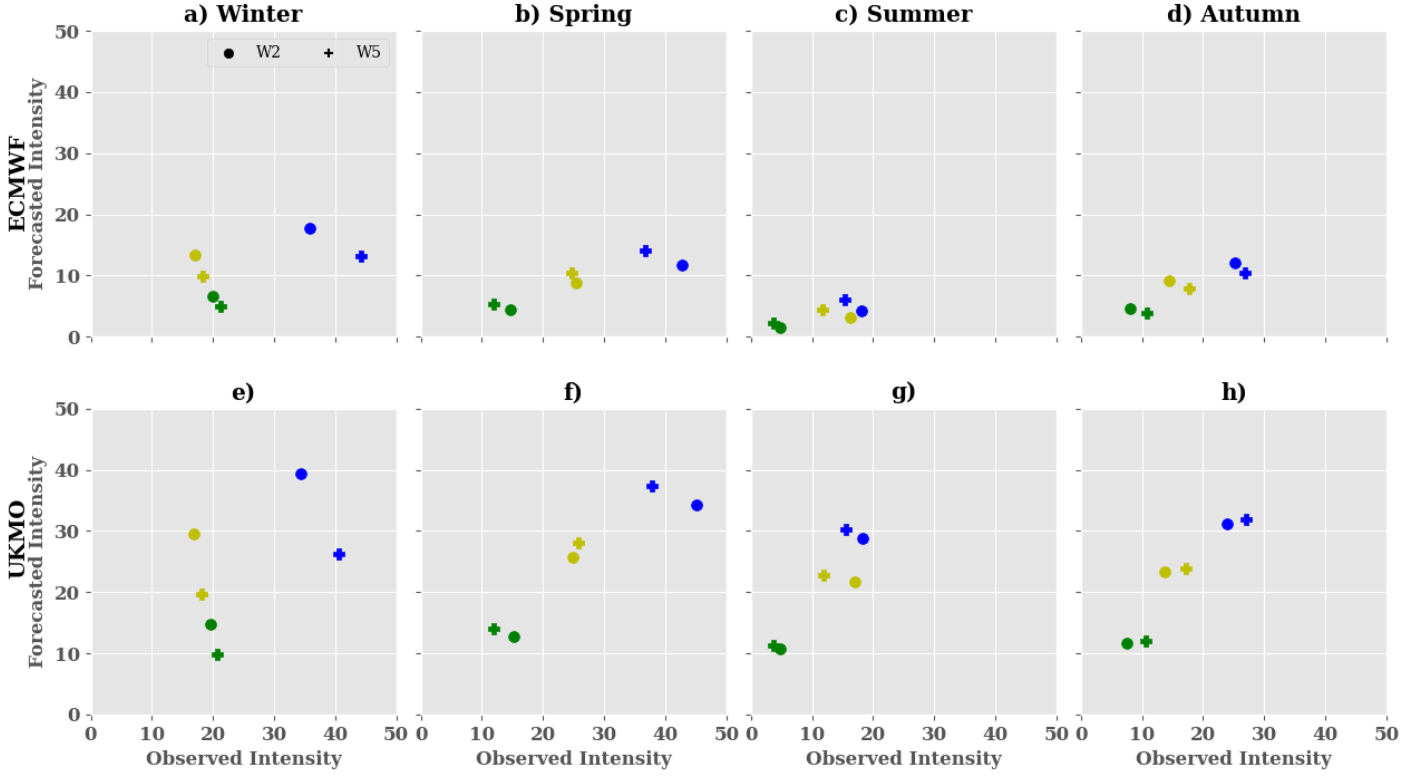


Figure 11. Evaluation of heat waves detection in the forecast models with respect to ERA5 at daily time scale over the period 2001-2020 using T2m_min **(i)** and T2m_max **(ii)** values for : (a-d) hit-rate, (e-h) FAR and (i-l) GSS. The metrics were computed using the optimized forecasts with the 20% threshold (see section Methods for the optimisation of the ensemble forecasts). The metrics are computed independently for each initialisation date within a month, and all initialisations are averaged to obtain the metric values for the given month. The metrics were grouped by seasons : (a,e,i) winter; (b,f,j) spring; (c,g,k) summer and (d,h,l) autumn. The cyan and black borders of bar plots indicate the metrics obtained when using ECMWF and UKMO respectively. The Y and X axes show the metrics values and the lead times (W2: week2 and W5: week5) respectively. The horizontal red line represents the baseline climatology.


a) Winter
b) Spring
c) Summer
d) Autumn
e)
f)
g)
h)
W2
W5
ECMWF
UKMO
Forecasted Intensity
Observed Intensity
0
10
20
30
40
50

(i)

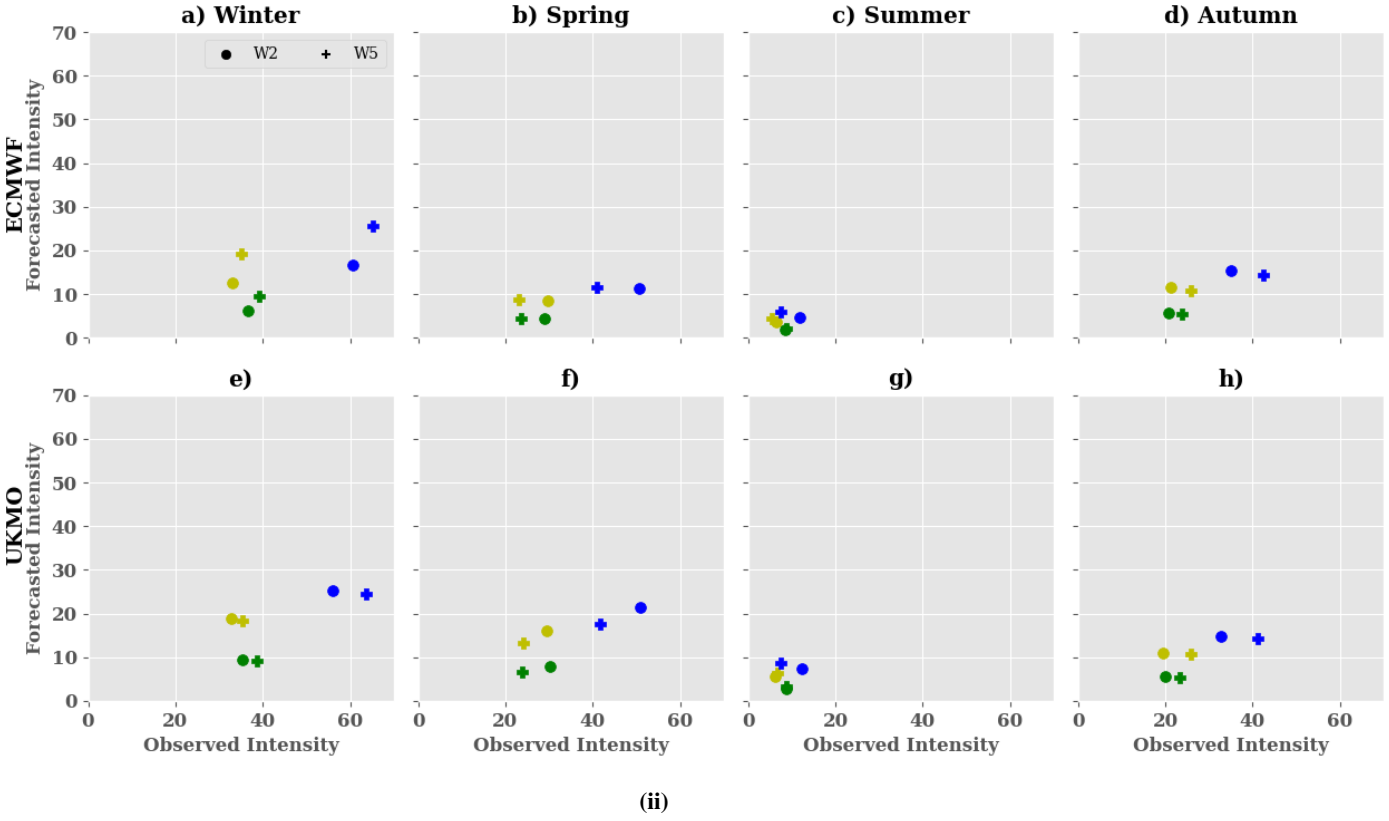


Figure 12. Evaluation of the intensity of heat waves in the forecast models and ERA5 over the period 2001-2020 during the seasons : (a,e) winter; (b,f) spring; (c,g) summer; (d,h) autumn using T2m_min **(i)** and T2m_max **(ii)**. The intensity of heat waves is computed independently for each initialisation date within a month, and all initialisations are averaged to obtain the average intensity of heat waves for the given month. Yellow, green, blue colors represent the values of intensity in the AT, GU, CO regions respectively. The dot and cross symbols represent the intensity of heat waves during week2 (W2) and week5 (W5) respectively. The Y and X axes represent the forecasted and observed intensities in ERA5 reanalysis respectively.