Subseasonal-to-seasonal forecasts of Heat waves in West African cities

Cédric Gacial Ngoungué Langué, Christophe Lavaysse, Cyrille Flamant

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

Subseasonal-to-seasonal 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 dramatic events are useful for policy makers and climate services to anticipate risks and develop appropriate responses. Sub-seasonal to seasonal forecasts are of great importance for actions to mitigate the human and health consequences of extreme heat. In this perspective, the present study addresses the predictability of heat waves at sub-seasonal to seasonal time scale in West African cities over the period 2001-2020. Two types of heat waves were analyzed : dry and wet heat waves using 2-meter temperature (T2m) and wet bulb temperature (Tw) respectively. Two models that are part of the S2S forecasting project, namely the European Centre for Medium-Range Weather Forecasts (ECMWF) and the UK Met Office models, were evaluated using two state-of-the-art reanalysis products, namely the fifth generation ECMWF reanalysis (ERA5) and the Modern-Era Retrospective analysis for Research and Application. The skill of the models to detect hot extreme events is evaluated using the Brier score. The models show significant skills in detecting hot days both for short- and long-term forecasts (2- and 5-week lead times, respectively). The predictability of heat waves in the forecast models is assessed by calculating categorical metrics such as the hit-rate, the Gilbert score and the false alarm ratio (FAR). The forecast models show significant skills in predicting heat wave days compared to a baseline climatology, mainly for short-term forecasts (two weeks lead time) in three climatic regions in West Africa, but the hit-rate values remain below 50% on average. We find that nighttime heat waves are more predictable than daytime heat waves. On average, the False Alarm Ratio (FAR) is excessively high and tends to increase with the lead time. 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. The skill of the models in forecasting dry and hot heat waves are very close. Although the models demonstrate skills on heat wave detection compared to a baseline climatology, they fail in predicting the intensity of heat waves.

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. Increased atmospheric humidity can exacerbate heat stress and lead to deaths, particularly

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

among vulnerable populations such as children and the elderly. 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. In some equatorial regions such as sub-Saharan Africa, the predictability of heat waves remains poorly documented.

Heat waves are often associated with extreme daytime and/or night-time temperatures. Most often, heat waves are defined as a period of at least three consecutive days with daily temperatures above a threshold [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). Therefore, there is no standard definition of a heat wave [e.g. Perkins (2015); Oueslati et al. (2017); Shafiei Shiva et al. (2019)]. Different thresholds, duration and indicators contribute to the divergence in the definition of heat waves (Smith et al., 2013). Heat waves can be defined from daily meteorological variables such as daily raw temperature [e.g. Fontaine et al. (2013); Beniston et al. (2017); Ceccherini et al. (2017); Déqué et al. (2017); Batté et al. (2018); Barbier et al. (2018); Lavaysse et al. (2018); Engdaw et al. (2022); Ngoungue Langue et al. (2023)], min, min or max 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. Some other authors use the daily anomalies of temperature to define heat waves [e.g. Stefanon et al. (2012); Barbier et al. (2018)].

The impact of heat waves on human activities and health increases the need for skillful and reliable climate forecasts on sub-seasonal 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. This is usually done using seasonal weather forecast models. Those forecasts can provide early indications of the frequency and duration of severe heat waves, but some studies on the use of climate forecasts by service providers suggest that the sub-seasonal range is also highly relevant for actions aimed at mitigating the human and health consequences of extreme heat (White et al., 2017). Sub-seasonal to 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 action to be taken before a potential disaster.

Vitart and Robertson (2018), as part of the Subseasonal-to-Seasonal Prediction Project (S2Sproject,https://www.ecmwf.int/en/research/projects/s2s), analyzed the predictability of extreme events across the globe, with a particular focus on the Russian heat wave of 2010. They show that S2S forecasts have the potential to predict the onset, evolution and decay of some large-scale extreme events several weeks in advance. The time that elapses between the initialization of the forecast and the occurrence of the target event is called the lead time. To assess the predictability of the 2010 Russian heat wave, their approach was based on the predictability of 2-meter temperature (T2m) anomalies a few weeks before the onset of the event. They found that the S2S models were able to predict extreme T2m anomalies over Russia up to three weeks lead time. Katsafados et al. (2014) also assessed the predictability of the 2010 Russian heat wave at seasonal time scale using the National Center for

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences
Discussions

Atmospheric Research Community Atmosphere Model version 3 (NCAR CAM3) and the latest ECMWF operational forecasts for validation. Their forecasting approach is based on the predictability of atmospheric blocking using temperature at 850 hPa and the geopotential height at 500 hPa. They found that only a few members reproduced the main features of the blocking system three months in advance. Most of the members predicted the blocking system in the wrong location and underestimated its duration. In West Africa, few studies have been conducted on heat wave predictability. Batté et al. (2018) evaluated heat wave forecasting at sub-seasonal and seasonal time scales using the Météo-France model in the S2S project and the Météo-France seasonal forecasting System 5 (MF5) respectively. To assess the skills of the models at seasonal time scale, they defined two indices from the Expert Team on Climate Change Detection and Indices (ETCCDI) database,namely, the heat wave duration index (HWDI) and the number of heat waves per season (HWPP) using the apparent temperature and 2-meter temperature. At the sub-seasonal time scale, they used the apparent temperature and T2m anomalies instead of the HWDI and HWPP. They found that at the seasonal time scale, the skills of MF5 to reproduce inter-annual anomalies of heat wave duration is limited at the grid point level because of the high spatial variability in the region, but it is significant when taking the average over the Sahel region [10°N-20°N, 10°W-20°E]. At sub-seasonal time, the skills of the model decrease beyond one week. Guigma et al. (2020) assessed the predictability of Sahelian heat waves at sub-seasonal time scale using the ECMWF extended long-range forecast system (ENS-ext), ERA5 and BEST gridded data for the evaluation. Their approach is based on forecasting the probability of occurrence of heat waves. They show that ENS-ext is able to forecast Sahelian heat waves with significant skill up to 2 weeks ahead; and with increasing lead time, nighttime heat waves are more predictable than daytime events.

Most previous studies assess heat wave predictability using T2m anomalies and large-scale predictors such as temperature at 850 hPa and geopotential height at 500 hPa [e.g. Vitart and Robertson (2018); Katsafados et al. (2014)]. While this approach is adequate for obtaining information about the weather situation for the future days, it cannot provide useful information about the onset and duration of heat waves. In this study, we will adapt the method proposed by Lavaysse et al. (2019) when assessing the predictability of heat waves over Europe. This method is more robust and takes into account the occurrence and duration of events.

The present study assesses the predictability of heat wave frequency and characteristics in West African cities over the period 1993 to 2020 using two sub-seasonal to seasonal (S2S) models namely, ECMWF and UKMO. To the author's knowledge, this work is the first of its kind in the region and represents a first assessment of heat wave predictability at subseasonal to seasonal time scale in West Africa. To carry out this study, 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. Finally, the predictive skill of the models in forecasting heat waves is assessed.

The remainder of this article is organized as follows: in section2, we present the region of study and the data used for this work; the description of the methodology is also provided. Section3 contains the main results of this study following the methodology presented in section 2.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences Discussions

2 Region, Data and Methods

2.1 Region of interest

This work is carried out in 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. The West African climate is characterized by a short wet season followed by a long dry season. West Africa exhibits high climate variability at regional- and local- scale. The present study focuses on major cities in the coastal and continental parts of West Africa. 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 available are not well distributed in the regions. Therefore, reanalysis data appear as good candidates to overcome this issue. 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) as our 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 model which is one of the forecast models used here, one could argue that the evaluation of the skills of the forecast models using only ERA5 as reference is circular. Therefore, we 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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

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.

The assessment of dry and wet heat waves is done using T2m and Tw respectively. The computation of Tw is given by the following formula:

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

(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.

$$\mathbf{Rh} = \mathbf{100} * \frac{\exp\Big(\frac{\mathbf{a} * \mathbf{T_d}}{\mathbf{b} + \mathbf{T_d}}\Big)}{\exp\Big(\frac{\mathbf{a} * \mathbf{T}}{\mathbf{b} + \mathbf{T}}\Big)} \quad (2)$$

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

$$\mathbf{Rh} = \mathbf{0.263} * \mathbf{p} * \mathbf{q} * \Big[\exp\Big(\frac{\mathbf{17.67} * (\mathbf{T} - \mathbf{T_0})}{\mathbf{T} - \mathbf{29.65}}\Big)\Big]^{-1} \quad (3)$$

https://earthscience.stackexchange.com/questions/2360/how-do-i-convert-specific-humidity-to-relative-humidity

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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

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, pressure and specific humidity.

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 sub-seasonal to 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 ECMWF ENS model runs on the Integrated Forecast System (IFS) cycle CY41r2. The native spatial resolution of the ECMWF model is Tco639 (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. It contains 91 sigma levels from the surface to 80 km. ECMWF provides two types of outputs for the S2S program: real-time forecasts and reforecasts called "hindcasts". Real-time forecasts are forecasts for the coming days, while hindcasts refer to forecasts over historical periods that are used to calibrate a new version of a forecast model. ECMWF real-time forecasts are run with 51 ensemble members (50 perturbed and 1 unperturbed), while hindcasts are run with 11 members. In this study, we focus on hindcasts only. ECMWF ENS hindcasts are produced twice a week, on Monday and Thursday at 00Z. We only analyzed the hindcasts for Thursday, as a first investigation revealed that some models had been initialized on the same date, but we realized after that those models were not available for the whole period. This means that each week a new set of hindcasts is produced to calibrate the real-time ensemble forecasts for Monday and Thursday of the following week using the latest version of the IFS. The 11-member ensemble hindcasts start on the same day and month as the real-time forecast (Monday and Thursday), 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 for 6 weeks. 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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

model is N216: 0.83°x0.56° (about 60 km at mid-latitudes). 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 real-time forecasts and hindcasts. The UKMO real-time forecast consists of a set of 4 members run daily for a period of 60 days (3 perturbed members and 1 control member). The UKMO hindcasts are produced 4 times per month, on the 1st, 9th, 17th and 25th, and cover a 24-year period from 1993 to 2016. We are aware that these init dates are not the same as those of ECMWF, but we are interested in this work on the predictability of heat waves in a global perspective, not on specific events. The ensemble hindcasts are composed of 7 members per cycle (from the 25 March 2017 hindcasts, prior to that 3 members per cycle). 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 recompose the products to obtain a new composite that covers the whole target period. The process applied is described in the following expression:

$$\mathbf{UKMO_{2001-2021}} = \mathbf{Concatenation}\left[\mathbf{UKMO_{hindcast2001-2016}}, \mathbf{UKMO_{realtimeforecast2017-2021}}\right] \quad (4)$$

The concatenation operator is applied to the time dimension. The real time forecasts were extracted for the same days as the hindcasts initialization. The number of ensemble members in the hindcasts is set to 4 (1-control member and 3-perturbed members) to be consistent with the number of ensemble members in the real time 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.

2.4 Metrics

2.4.1 Estimation of temperatures at the city scale

Weather forecasts provide the evolution of atmospheric variables on a global scale, which implies the need to have data from local stations to access information on a local scale. This is a major problem in areas where there is a lack of 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 developed 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 ([Table1]shows the lsm values of all cities considered in this study). The same technique was used for the forecasts.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences
Discussions

2.4.2 Heat wave detection

The detection of heat waves is carried out using 2-meter daily minimum temperature (T2m_min), 2-meter daily maximum temperature (T2m_max) and the average wet bulb temperature (Tw) as indicators. T2m_min and T2m_max are used to investigate the evolution of nighttime and daytime heat waves respectively. The most lethal heat waves are due not only to high temperatures but also to the effect of humidity (Steadman, 1979a, b). We therefore used Tw to assess the evolution of wet heat waves. 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 exceeds the calendar 90th percentile threshold computed over the entire period for T2m_min, T2m_max or Tw respectively [Fig.2]. The 90th percentile is calculated for each calendar day of the year. 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 90th percentile, 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 boolean files from T2m_min, T2m_max and Tw time series at each grid point following the steps below :

- proceed to heat wave detection at each grid point;
- for each grid point, we define a zero vector with the length of T2m_min or T2m_max time series containing only zeros;
- for days in T2m_min, T2m_max or Tw time series, if days are hot days, we replace in our zero vector the values corresponding to those days by 1. Hot days are days with T2m_min, T2m_max or Tw above the 90th percentile daily thresholds. In order to assess the characteristics of heat waves, only hot days belonging to heat wave sequences are kept. This is applied for all grid points and we obtain boolean files containing 0 or 1 (Ngoungue et al. 2023). These boolean files will be processed both for the reanalyses and the forecasts to assess the representation of heat waves occurrence and duration.

The intensity of heat waves was defined as the cumulative sum of the daily exceedances of daily values of indicators to the climatological daily threshold in a sequence of hot days (see Ngoungue Langue et al. (2023) for more details). This study is in the framework 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) project which focuses on the human impacts of climate extremes. Therefore, the climatological daily threshold is chosen to be constant over the whole period; and it is defined as the minimum of the daily climatology thresholds over the study period. This approach allows us to properly assess the severity of a heat wave and its potential human impacts.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

2.4.4 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. Although ensemble forecasting has many advantages over deterministic forecasting, the evaluation of the performance of an ensemble forecasting model remains quite complex due to the amount of information available. Statistical analyses can be performed by considering only the mean of the members, the median member, the warmest and coldest members, the 1st and 3rd quartiles. In order to extract deterministic information from a probabilistic system for the calculation of prediction scores, we defined an approach based on member thresholds inspired by Lavaysse et al. (2019). Three threshold values were tested (see [Table2]) to optimize the ensemble forecast system based on results found in Lavaysse et al. (2019).

2.4.5 Skill scores

The skill of the probabilistic models in forecasting the evolution of T2m, Tw and extreme events are assessed using several adapted scores. The forecast of T2m and Tw in the models is done using the Continuous Rank Probability Score (CRPS). 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 calculated using the following formula:

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

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

The representation of hot days in the forecasts is addressed using the Brier score. The Brier Score (BS) is similar to the CRPS with the only difference that it focuses on specific events. It 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 :

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

$\mathbf{P_i}, \mathbf{O_i}$ 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.

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 [Table3] from which 3 metrics have been computed : hit-rate, FAR ratio and GSS. The hit-rate indicates the percentage of

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

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 calculated 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:

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

2.4.6 Heat waves forecast strategy

The assessment of heat wave predictability in the forecast models is a very difficult task. To address this concern, previous studies [e.g. Vitart and Robertson (2018); Batté et al. (2018)] have used T2m anomalies and large-scale atmospheric variables such as the geopotential at 500hPa for example, to predict their occurrence a few days in advance. In this study, we adapted the methodology developed by Lavaysse et al. (2019) to evaluate the predictability of heat waves in the forecasts. This assessment is done in two steps:

- The first approach is 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 calculated for each day at different lead times from which the hit-rate, FAR and GSS have been derived. The scores are calculated 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 of the event is correctly predicted. To facilitate the evaluation, dichotomous scores are now calculated 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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences Discussions

3 Results

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

The first assessment focuses on the representation of the climatological evolution of two key variables for this study, T2m and Tw in the forecasts compared to the reanalyses. To do so, the seasonal climatological bias between the forecasts 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). First, we analysed the bias between the forecast models and the ERA5 reanalysis using T2m_min values. Both models show a cold bias over the Sahel region which is more pronounced with UKMO [Fig.3_a]. We observed in UKMO, a progressive shift of this strong cold bias over the northern Sahel during the season from winter to summer [Fig.3_a (e-g)]. UKMO exhibits a hot bias with respect to ERA5 over the Atlantic Ocean that decreases considerably in summer. The cold 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 hot bias over the Atlantic coast during the winter and spring [Fig.3_b (a-b),(e-f)]. The evaluation of the representation of Tw in the forecasts shows strong cold bias with respect to the reanalyses over the whole Sahel region [Fig.4]. Compared to the results found with T2m, this strong cold bias can be related mainly to the underestimation of humidity in the models in this region. The assessment of the bias between the forecast models and MERRA on the evolution of T2m highlights significant discrepancies with the results obtained using ERA5 as a reference. Using T2m_min, we observed a warm 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.S1] in supplement material). The spatial distribution of T2m in the forecast models shows significant biases with respect to the reanalyses. In order to determine whether these biases change over time, we analyze the spatio-temporal evolution of T2m and Tw. To do this, we calculate the daily climatological biases between forecasts 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.S2]. These significant 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 cold biases in the CO region in winter ([Fig.S3] 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.S2 and [Fig.S3] 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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences Discussions

3.2 Assessment of the predictive skills of T2m and Tw forecasts

In this section, we carry out a global assessment of the predictive skills 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 calculated between forecasts and reanalyses over the 3 regions for T2m_min, T2m_max and Tw. The forecasts used in this work are launched for a 6-week duration; for each week, the median CRPS was calculated from 2001-2020 using ERA5 and MERRA. The CRPS results are similar for the first two weeks (Week1,Week2), the two intermediate weeks (Week3,Week4) and the last two weeks (Week5,Week6). 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-term 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. Firstly, the results of the CRPS obtained using ERA5 as reference [Fig.5] are analyzed.We have noticed that the skill of the models does not improve necessarily with decreasing lead time. This could be related to systematic biases (the representation of atmospheric circulation, local scale processes) in the forecast models. 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 more skills for T2m_min than T2m_max over the AT region. ECMWF generally shows more skill than UKMO. The evolution of the CRPS calculated using MERRA as a reference is similar to that found using ERA5 (not shown). As we previously observed a significant cold bias in the Sahel region with Tw, we expect high values of CRPS in the different regions. 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.S4] 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.

3.3 Extreme temperature values and heat waves in the models and reanalyses

3.3.1 Extreme temperatures

After the evaluation of the forecast models on the representation of T2m and Tw, we focus now on the extreme heat events referred here as hot days (see section 2.4.3). This is done through the calculation of the Brier score between the forecasts and reanalyses (ERA5, MERRA). In order to compute the Brier score, the raw forecasts and reanalyses are converted to boolean data, where a value of 1 represents a hot day and a value of 0 indicates no hot days. This approach based on a relative threshold (see section 2.4.3) will contribute to partially correct the biases previously found in the models. We firstly evaluated the Brier score between the forecasts and ERA5 for T2m and Tw. The range of values for the Brier score is between 0.05 to 0.175, indicating that the models are able to detect hot days in agreement with the reanalyses over the three regions for short and long term forecasts (Week2,Week5) ([Fig.6] and [Fig.S5] in supplement material). We found with T2m that in spring the models are

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences Discussions

more skillful and present less spatial variability compared to other seasons. In winter and autumn, for the short term lead time (Week2), the forecast models perform better for T2m_max than T2m_min. Over all lead times, regions and seasons, ECMWF performs better than UKMO on the detection of hot days except for Tw in the GU region in summer. Using MERRA reanalysis as reference, we also found that ECMWF has more skills than UKMO, except over the CO region during winter (see [Fig.S6] in supplement material) and the Brier score is in the same range of values as the Brier score computed using ERA5 for T2m and Tw. The results show a good performance of the models in predicting hot days for short and long term forecasts.

3.3.2 Spatial variability of heat wave characteristics

After assessing the skill of the models in predicting hot days, we study the spatial variability of heat waves in the forecast models versus the reanalyses. Heat waves are defined as a period of at least three consecutive hots days. The frequency and characteristics of heat waves were calculated using T2m and Tw. The mean duration (resp. mean intensity) of heat waves was calculated 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 made by calculating the bias between the models and the reanalyses for heat wave frequency, duration and intensity. The spatial distribution of the heat wave frequency bias between the forecast models (ECMWF, UKMO) and ERA5 reanalysis 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. 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 reanalysis. 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 north-south gradient observed in UKMO with T2m is also found in Tw and tends to strengthen from spring to autumn ([Fig.S7] in supplement material). The main differences between T2m and Tw on the representaion of heat wave days are found with ECMWF. 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 calculating the bias of intensity using MERRA reanalysis as reference for T2m_min, T2m_max and Tw (not shown). The bias of heat wave 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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences Discussions

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_a)] in supplement). We do not find significant differences when using MERRA as reference for the evaluation of heat waves duration with Tw (not shown).

3.4 Predictability of the heat waves in extended forecasts

In the previous section, we analyzed the spatial variability of heat wave characteristics in the forecast models versus the reanalyses. We are now assessing the skills of the models to detect heat waves in the three regions. This is done by calculating the following probabilistic scores described in Section 2: the hit-rate, GSS and FAR ratio (see Section 2). The scores are calculated 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 calculated metrics are then compared to a baseline climatology defined as the probability of having a heat wave in the ERA5 reanalysis over the period 2001-2020. The results presented below are obtained using a 20% threshold value to optimize the ensemble forecast system (see Section methods) [Fig.10]. The hit-rate and GSS values of the optimized forecasts using 40% and 60% percentile threshold values 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 geographical 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 reference both for short- and long- term forecasts (Week2, Week5) but the hit-rate values are below 0.5 indicating misses in the forecasts. The hit-rate values are slightly better for short-term forecasts. ECMWF presents more skills than UKMO for short-term forecasts in winter. UKMO, for instance, is better for long-range forecasts, mainly for T2m_min [Fig.10(a-d)]. We also noticed that the predictability of heat waves is slightly better for heat waves associated with T2m_min than for T2m_max. We can infer from this result that nighttime heat waves are more predictable than daytime heat waves. The models show more skills in the AT and CO regions in winter and spring for T2m_min values for short-term forecasts. The analysis of the inter-day variability of T2m is assessed by calculating 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 [Table4]. 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. The skills of the models previously highlighted in the AT region could be partly explained by this low variability of daily T2m. 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 short- and long-range forecasts (see [Fig.S11(a-d)] in supplement material).

The second metric calculated 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 both for T2m_min, T2m_max and Tw ([Fig.10(i-l)] and [Fig.S11(i-l)] in supplement). This suggests that the skills of the models to predict heat waves are considerably reduced when random hits

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences
Discussions

(chance hits) are removed from the total hits. We found that the GSS values are significantly decreasing with lead time and season; the highest values are observed in winter. In 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 greater than 0, indicating that the forecast models perform better than a random forecast mainly for short-term forecasts (Week2) [Fig.10(i-l)]. Similar results of the GSS are found with Tw (see [Fig.S11(i-l)] in supplement). An important aspect of the evaluation of a forecasting system is to assess its ability to predict events that do not occur in the observations: the so-called "false alarms". This is done by the calculation of the FAR ratio between the forecast models and ERA5 reanalysis for T2m_min [Fig.10(e-h)], Tw [Fig.S11(e-h)] and T2m_max (not shown). The FAR values are too high on average, about 0.7 and 0.85 for the short and long term forecasts, 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 ratio increases with the lead time; ECMWF makes fewer false alarms than UKMO for short-term forecasts. The FAR ratio 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 forecast models show skills at weekly time scale compared to the baseline climatology. To be consistent with the previous analyses, the predictability of heat wave intensity in the models is also assessed for short- 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 heatwave intensity, they fail in accurately predicting the specific intensity values. The models underestimate the intensity of heat waves (see [Fig.11] for T2m_min, the results are very similar for T2m_max 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

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 data assimilation technique and the spatial resolution of the atmospheric component. First of all, the physical parameterizations used to simulate atmospheric processes such as convection, turbulence, interactions surface-ocean, surface-radiation and cloud

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

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 models used the same data assimilation methods (4D-Var) for control analyses but the data and initial conditions are completely different. These differences may influence the representation of surface-ocean interactions in the models. The differences observed in the representation of T2m_min over the Atlantic ocean [Fig.3] can result from these types configurations. Another important factor is the spatial resolution of the atmospheric component of the two models : ECMWF has a higher spatial resolution than UKMO (0.16°x0.16° 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. This may explain why the ECMWF performs better than the UKMO in certain regions, for example in the case of extreme temperatures over AT and GU [Fig.6]. A more detailed analysis of these differences between the models is beyond the scope of this study.

5 Conclusions

This study is a first assessment of the predictability of heat waves in West African cities on a sub-seasonal to seasonal 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 investigate the representation of hot days and heat wave characteristics in the models with respect to the reanalyses. The Brier score exhibits small values for T2m and Tw, indicating that the models are able to detect extreme heat events up to 5 weeks lead time. Over all lead times, seasons and regions, ECMWF performs better than UKMO in detecting hot days when compared to both ERA5 and MERRA. The models face some issues in reproducing properly the spatial evolution of heat wave characteristics. 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 assessed predictive skills of the models using the hit-rate, GSS and FAR. The models perform well against a baseline climatology, but the hit-rate values remain below 50%, indicating missed events in the forecasts. The hit-rate values are slightly better for short-term forecasts (up to two weeks lead time). We found that nighttime heat waves are more predictable than daytime heat waves. This result is consistent with Guigma et al. (2021). On average, the False Alarm Ratio (FAR) is excessively high and tends to increase with the lead time. 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 short-term forecasts. The prediction of dry heat waves is slightly better with ECMWF for short-term forecasts, while it is better with UKMO for long-term forecasts. For wet heat waves, UKMO outperforms ECMWF for both short- and long-term forecasts. Although the models demonstrate

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023


Natural Hazards and Earth System Sciences
Discussions

skills on heat wave detection compared to a baseline climatology, they fail in predicting the intensity of heat waves; the accurate forecast of heat waves intensity remains a challenging task for the models.

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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

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.

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.

Beniston, M., Stoffel, M., and Guillet, S.: Comparing observed and hypothetical climates as a means of communicating to the public and policymakers: The case of European heatwaves, 67, 27–34, https://doi.org/10.1016/j.envsci.2016.11.008, 2017.

Ceccherini, G., Russo, S., Ameztoy, I., Marchese, A. F., and Carmona-Moreno, C.: Heat waves in Africa 1981–2015, observations and reanalysis, 17, 115–125, https://doi.org/10.5194/nhess-17-115-2017, 2017.

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.

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.

Fontaine, B., Janicot, S., and Monerie, P.-A.: Recent changes in air temperature, heat waves occurrences, and atmospheric circulation in Northern Africa, 118, 8536–8552, https://doi.org/10.1002/jgrd.50667, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50667, 2013.

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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences Discussions

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.

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.

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.

Katsafados, P., Papadopoulos, A., Varlas, G., Papadopoulou, E., and Mavromatidis, E.: Seasonal predictability of the 2010 Russian heat wave, Natural Hazards and Earth System Sciences, 14, 1531–1542, 2014.

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.

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.

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.

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.

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.

Oueslati, B., Pohl, B., Moron, V., Rome, S., and Janicot, S.: Characterization of Heat Waves in the Sahel and Associated Physical Mechanisms, 30, 3095–3115, https://doi.org/10.1175/JCLI-D-16-0432.1, publisher: American Meteorological Society Section: Journal of Climate, 2017.

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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

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.

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.

Shafiei Shiva, J., Chandler, D. G., and Kunkel, K. E.: Localized Changes in Heat Wave Properties Across the United States, 7, 300–319, https://doi.org/10.1029/2018EF001085, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018EF001085, 2019.

Smith, T. T., Zaitchik, B. F., and Gohlke, J. M.: Heat waves in the United States: definitions, patterns and trends, 118, 811–825, https://doi.org/10.1007/s10584-012-0659-2, 2013.

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.

Stefanon, M., D'Andrea, F., and Drobinski, P.: Heatwave classification over Europe and the Mediterranean region, 7, 014 023, https://doi.org/10.1088/1748-9326/7/1/014023, publisher: IOP Publishing, 2012.

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.

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. and Robertson, A. W.: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events, Npj Climate and Atmospheric Science, 1, 3, 2018.

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.

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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

Tables

Table 1. 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 2. 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 3. Contingency table.

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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

Table 4. 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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

Figures

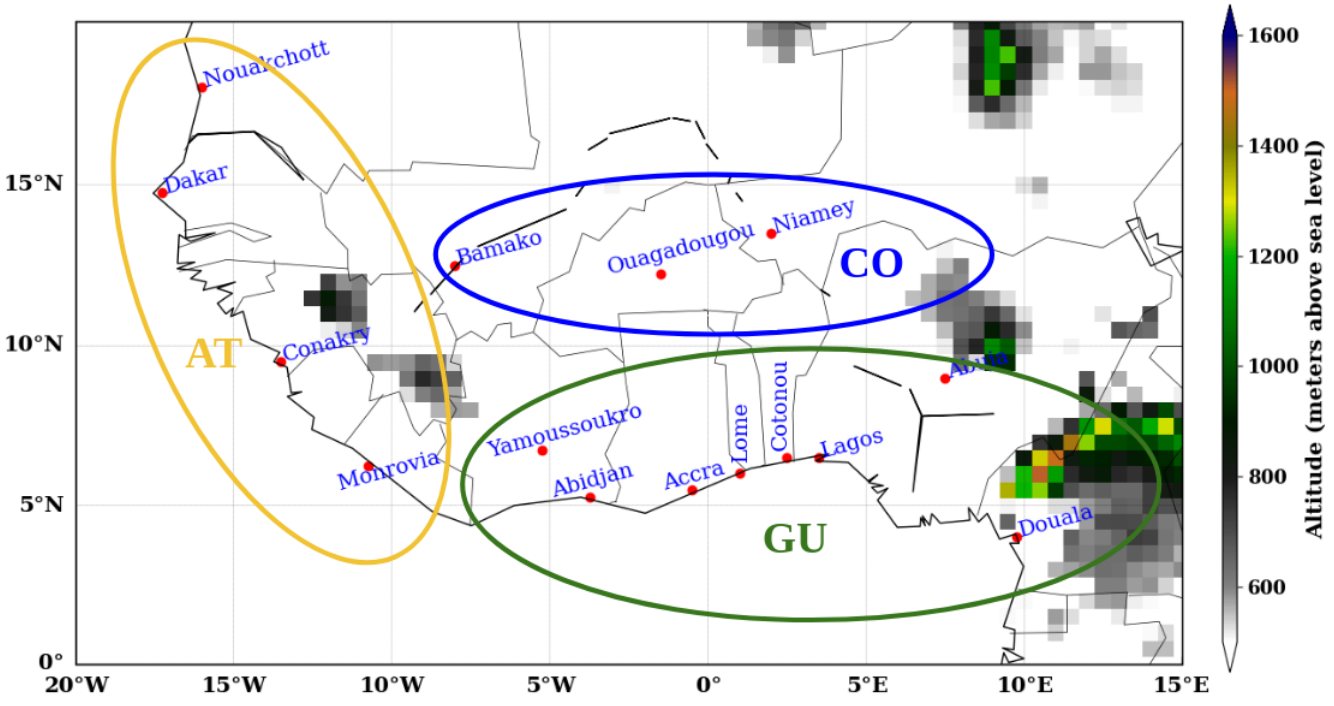
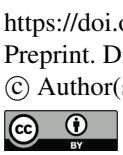


Figure 1. Topographic map of West Africa using ERA5 elevation data. The circles on the map represent the different climatic zones: AT (Coastal atlantic zone), CO (Continental zone) and GU (Coastal Guinean zone). The y and x axes represent the latitude and longitude respectively. The color bar shows the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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023



Natural Hazards and Earth System Sciences
Discussions

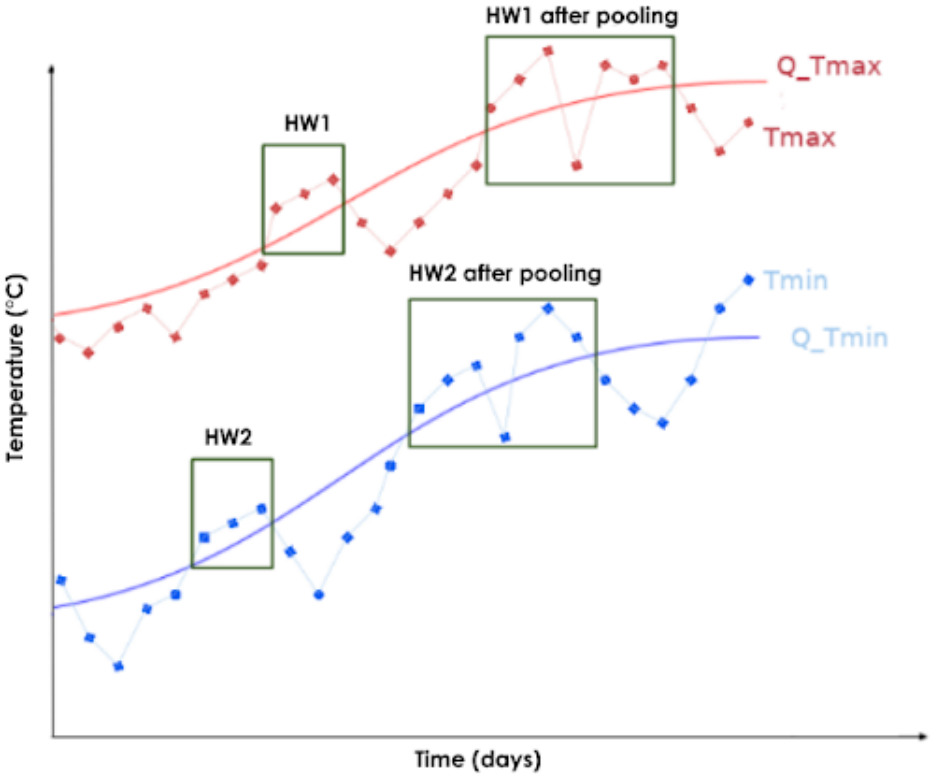


Figure 2. Detection process of heat wave: HW1/HW2 represent events associated respectively to maxima/minima temperature. The red/blue lines with circles are max/min daily temperatures. Red/blue solid lines are respectively max/min thresholds. X- and Y- axis represent the time in days and the temperature in degrees celsius. 'With pool' 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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

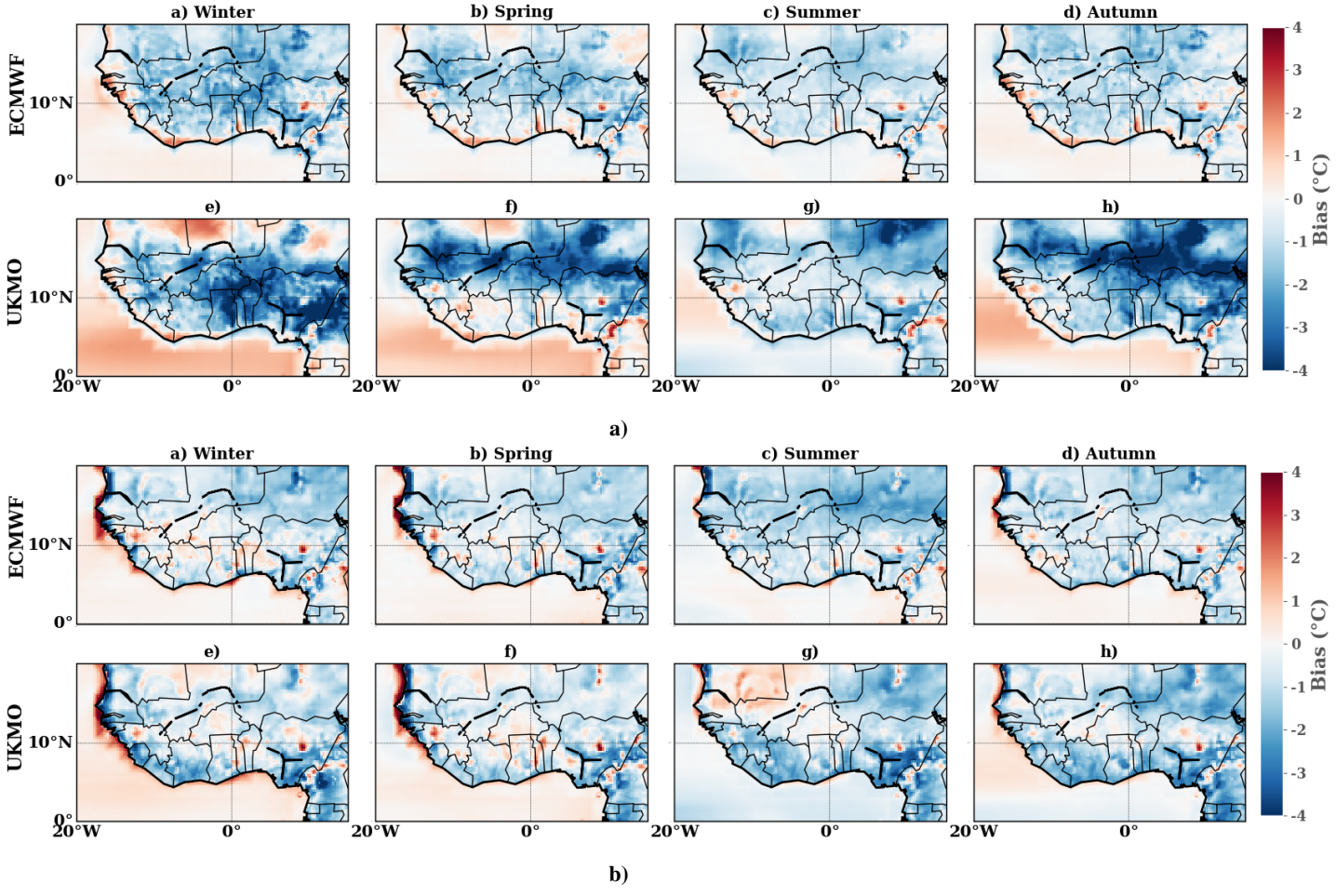
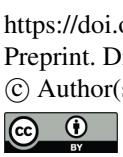


Figure 3. Spatial variability of the climatological bias between the forecast models ensemble mean and ERA5 reanalysis over the period 2001-2020 for : **(a)** T2m_min and **(b)** T2m_max, during the seasons : (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is computed as the difference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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023



Natural Hazards and Earth System Sciences
Discussions

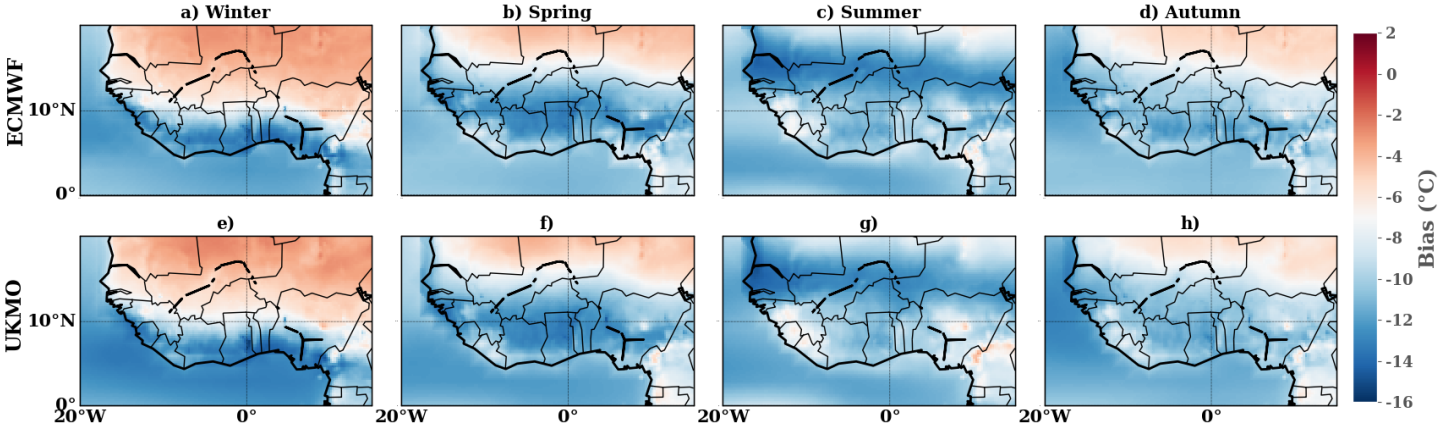


Figure 4. Spatial variability of the climatological bias between the forecast models ensemble mean and ERA5 reanalysis over the period 2001-2020 for Tw during the seasons : (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is computed as the difference between the forecast models and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

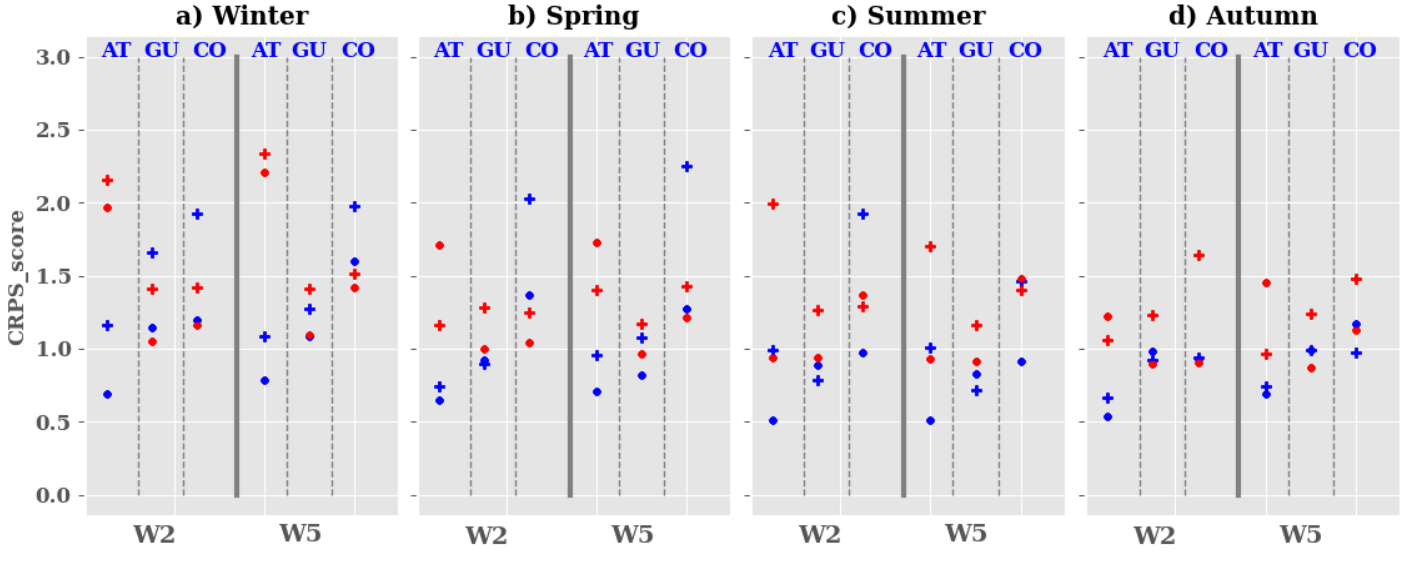


Figure 5. Evolution of the CRPS score between the forecast models and ERA5 reanalysis over the period 2001-2020 during the seasons : (a) winter, (b) spring, (c) summer and (d) autumn. The blue and red colors represent the CRPS score calculated using T2m_min and T2m_max respectively. The dot and cross 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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

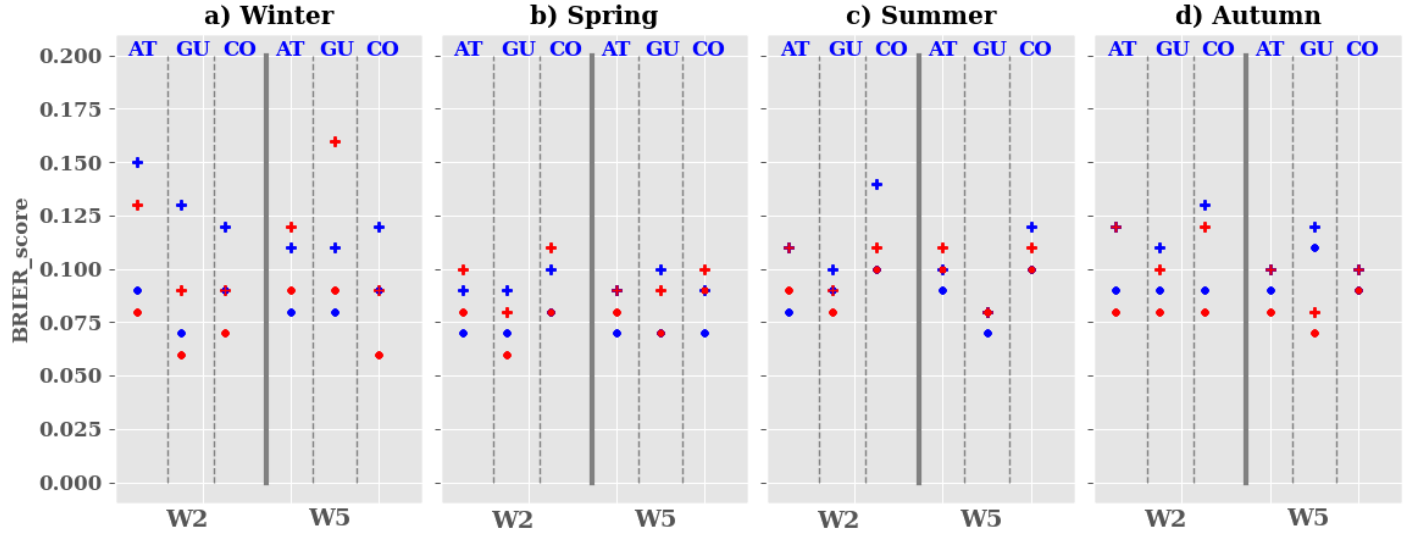


Figure 6. Evolution of the Brier score between the models and ERA5 reanalysis over the period 2001-2020 during the seasons : (a) winter, (b) spring, (c) summer and (d) autumn. The blue and red colors represent the Brier score calculated using T2m_min and T2m_max values respectively. The dot and cross symbols indicate the Brier 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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

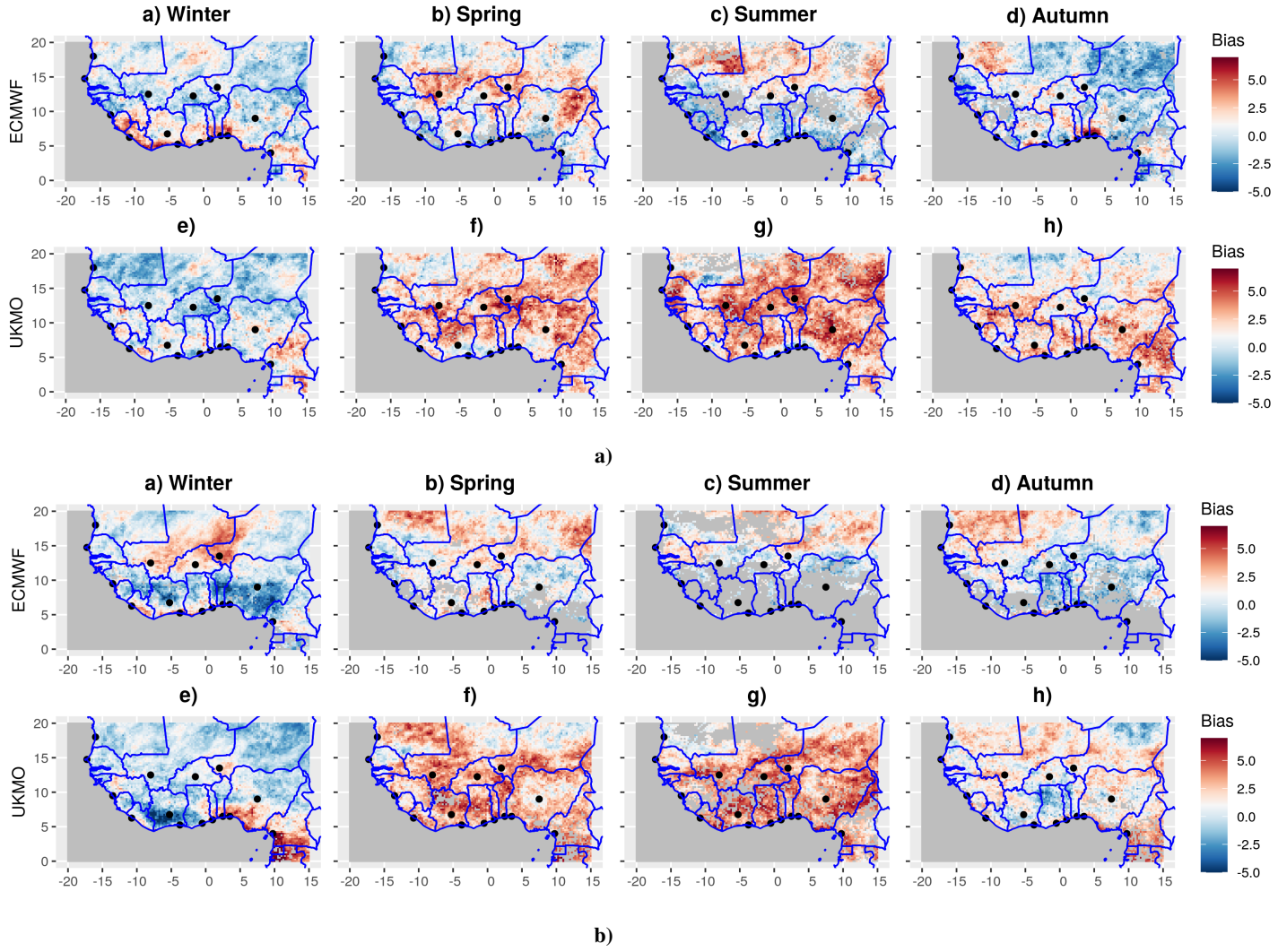


Figure 7. Spatial variability of heat wave frequency bias between forecast models and ERA5 over West Africa from 2001 to 2020 for:**(a)** T2m_min values and **(b)** T2m_max values, during: (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is calculated as the difference in heat wave frequency between the forecast models and ERA5. This analysis is performed using the unperturbed member of the models. The color bar indicates the bias values without units. 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).

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

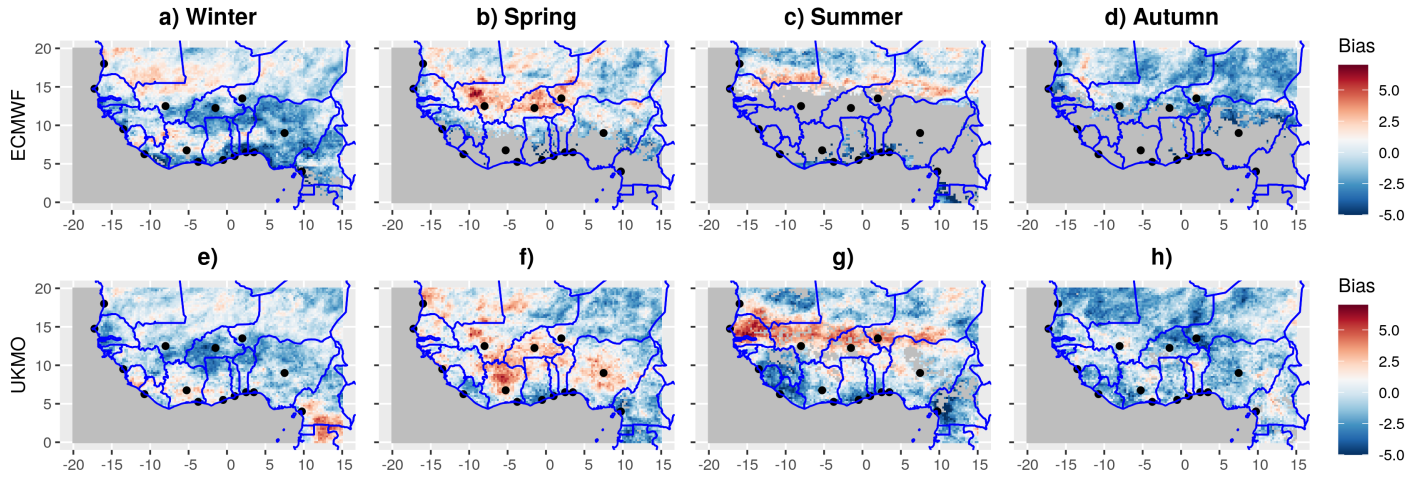


Figure 8. Spatial variability of heat wave frequency bias between forecast models and ERA5 over West Africa from 2001 to 2020 using Tw during: (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is calculated as the difference in heat wave frequency between the forecast models and ERA5.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

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

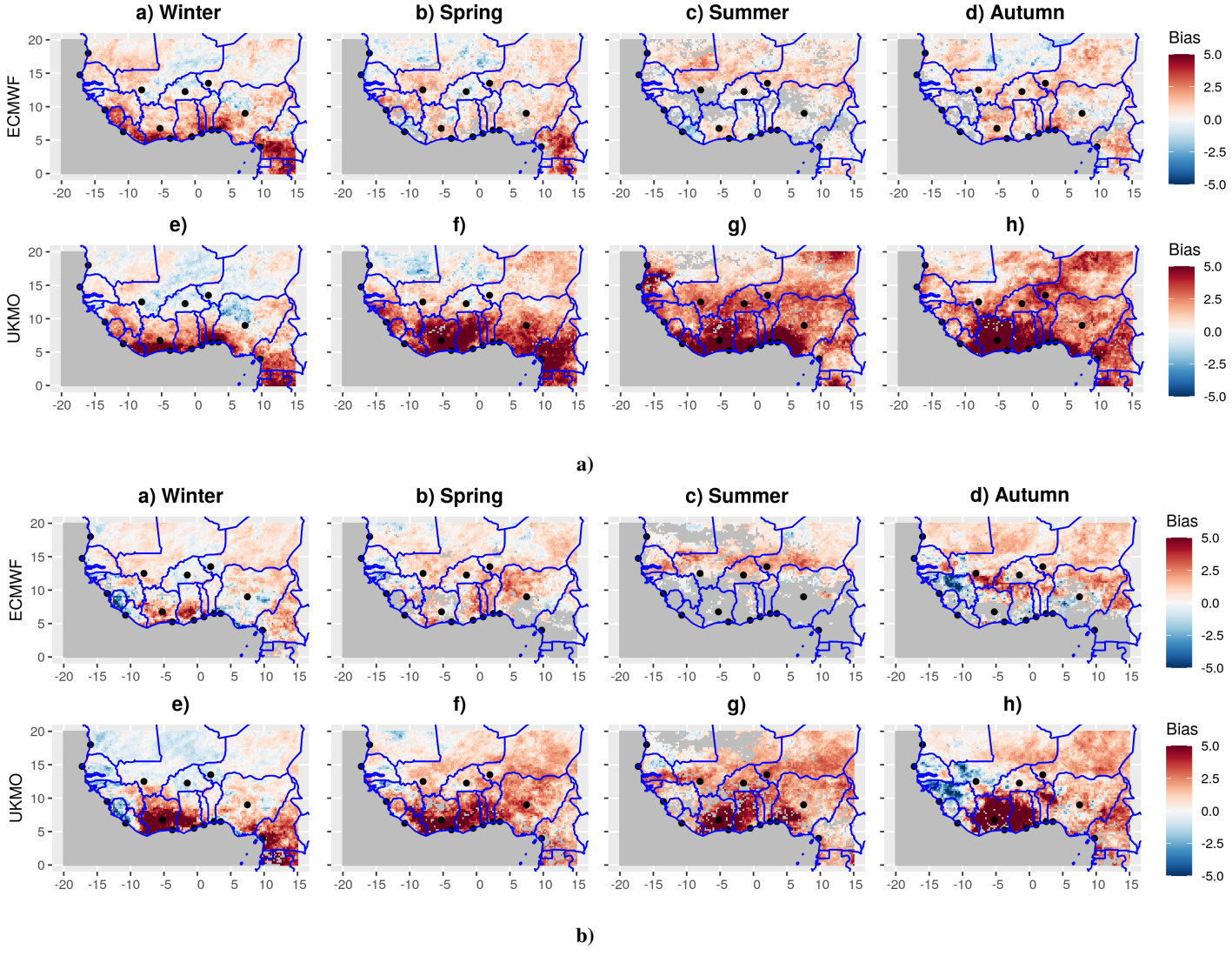


Figure 9. Spatial variability of heat wave duration bias between the forecast models and ERA5 over West Africa from 2001 to 2020 for: **(a)** T2m_min values and **(b)**, during: (a,e) winter; (b,f) spring; (c,g) summer and (d,h) autumn. The bias is calculated as the difference in heat wave duration between the forecast models and ERA5. This analysis is performed using the unperturbed member of the models over 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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

Natural Hazards and Earth System Sciences
Discussions

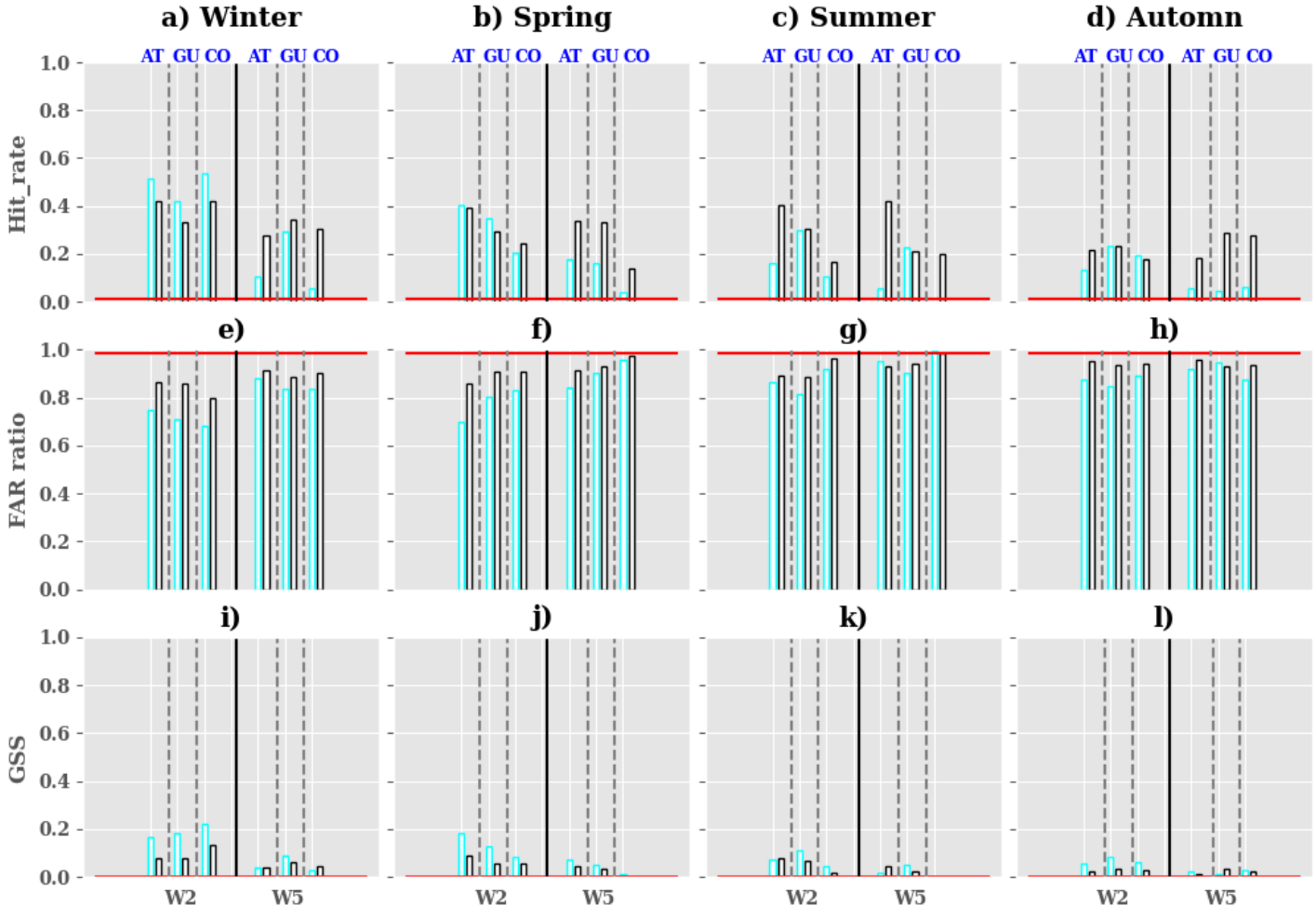


Figure 10. Evaluation of heat waves detection in the forecast models with respect to ERA5 at daily time scale over the period 2001-2020 using T2m_min values for : (a-d) hit-rate, (e-h) FAR ratio and (i-l) GSS. The metrics were computed using the optimized forecasts (see section Methods for the optimisation of the ensemble forecasts). The metrics were calculated during the 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.

https://doi.org/10.5194/nhess-2023-144
Preprint. Discussion started: 10 August 2023

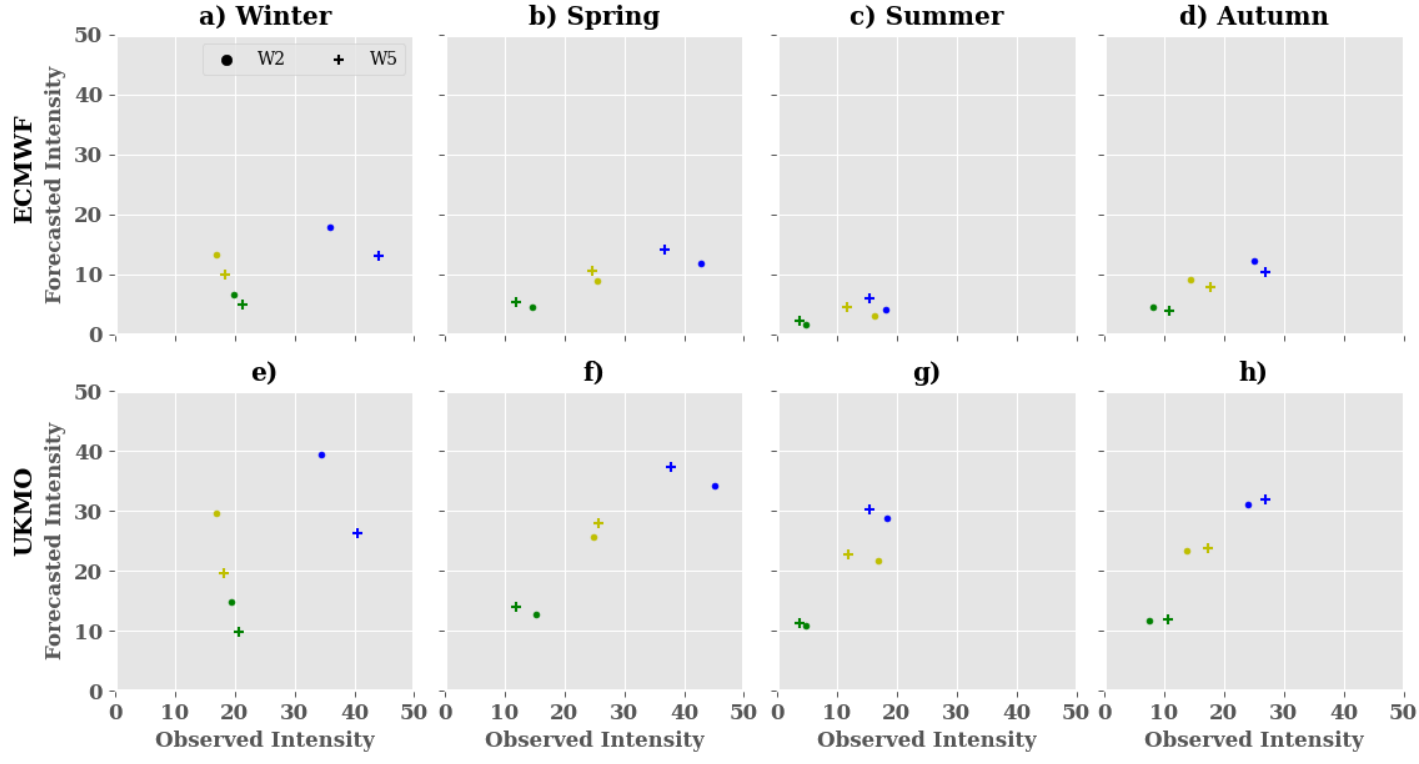


Figure 11. Evaluation of the intensity of heat waves in the forecast models and ERA5 over the period 2001-2020 during the seasons : (a,e) winter; (b,f) spring; (c,g) summer; (d,h) autumn using T2m_min values. Yellow, green, blue colors represent the values of intensity in the AT, GU, CO regions respectively. The dot and cross symbols represent the intensity of heat waves during week 2 (W2) and week5 (W5) respectively. The Y and X axes represent the forecasted and observed intensities in ERA5 reanalysis respectively.