Comment on nhess-2022-192
Anonymous Referee #1

Referee comment on "Heat waves monitoring over West African cities: uncertainties,characterization and recent trends" by Cedric Gacial Ngoungue Langue et al., Nat. Hazards

Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-192-RC1, 2022
Heat waves monitoring over West African cities: uncertainties, characterization and recent trends

General comments

This manuscript assesses potential uncertainties encountered in the process of heat wave monitoring and analyse their recent trend in West African cities. This is investigated using downscaled ERA5 and MERRA variables for the period 1993-2020. Three types of uncertainties are discussed. The first type is related to the reanalyses themselves; the second, to the sensitivity of heat wave frequency to the threshold values used to define them; and, finally, the third is related to the choice of indicators and the methodology used to define heat waves.

We thank the reviewer for his/her availability and interest to examine this work; and also for his/her insightful suggestions to improve the quality of the paper.

Specific comments

The abstract is rather long. It contains a surplus of information that could better fit in other sections of the paper such as the introduction. I might suggest keeping it more concise, only stating the main problems & objectives, how they have been dealt with and the main results obtained. The abstract does not mention that local stations have been used nor the downscaling methodology applied. As it is now it seems that it only uses reanalysis data, and this gives a sense of contradiction with the title (which emphasizes the application to cities).
Thanks for this constructive remark, we rearranged the abstract as suggested in the comment.

We changed :

“Heat waves can be one of the most dangerous climatic hazards affecting the planet; having dramatic impacts on the health of humans and natural ecosystems as well as on anthropogenic activities, infrastructures and economy. Based on climatic conditions in West Africa, the urban centers of the region appear to be vulnerable to heat waves. In this study, we assess the potential uncertainties encountered in the process of heat waves monitoring and analyse their recent trend in West Africa cities. This is investigated using two state-of-the-art reanalysis products namely ERA5 and MERRA for the period 1993-2020. Three types of uncertainties are discussed. The first type of uncertainty is related to the reanalyses themselves, with MERRA showing a cold bias with respect to ERA5 over the Sahel and Guinean regions except over some countries (Guinea Bissau, Sierra Leone,Liberia). Furthermore, large discrepancies are found in the representation of extreme values in the reanalyses over the southern Sahel and the Guinea coast. The second type of uncertainty is related to the sensitivity of heat waves frequency to the threshold values used to monitor them. Heat waves detected using the lowest threshold value are very persistent and last for several days; while the duration of heat waves related to high threshold values is shorter. The choice of indicators and the methodology used to define heat waves constitutes the third type of uncertainty. Three sorts of heat waves have been analysed, namely those occurring during daytime, nighttime and both daytime and nighttime concomitantly. Four indicators have been used to analyse heat waves based on 2-m temperature, humidity, 10-m wind or a combination of these. Nighttime and daytime heat waves are in the same range of occurrence while concomitant day- and nighttime events are extremely rare because they are more restrictive. The climatological state of heat wave occurrence shows large differences between the indicators. We found that humidity plays an important role in nighttime events; concomitant events associated with wet-bulb temperature are more frequent and located over the north of Sahel. Most of the events detected in the regions (75%) have a duration around 3-6 days. The most dangerous events with a duration of at least 10 days contributed up to 12% of the total number of events.For all indicators, the interannual variability of heat waves in the west Africa region evidences 4 years with a significantly higher frequency of events (1998, 2010, 2016 and 2019) possibly due to higher sea surface temperatures in the Equatorial Atlantic corresponding to El Nino events. All indicators also highlight that the cities in the Gulf of Guinea region experienced more heat waves than those lying along the Atlantic coastline and those located in continental Sahel during the last decade. The heat wave events occurring in the Guinean region show short duration and weak intensity, while in the coastal and continental regions, events are persistent with strong intensity. We find a significant increase in the frequency, duration and intensity of heat waves in cities during the last decade (2012-2020) compared to the previous two decades. This is thought to be a consequence of climate change acting on extreme events."

to :

" Heat waves can be one of the most dangerous climatic hazards affecting the planet; having dramatic impacts on the health of humans and natural ecosystems as well as on anthropogenic activities, infrastructures and economy. Based on climatic conditions in West Africa, the urban centers of the region appear to be vulnerable to heat waves. The goals of this work is firstly to assess the potential uncertainties encountered in heatwaves detection; and secondly analyse their recent trend in West Africa cities during the period 1993-2020. This is done using two state-of-the-art reanalysis products, namely ERA5 and MERRA, as well as two local station datasets, namely Dakar in Senegal and Abidjan in Ivory Coast. An estimate of station data from reanalyses is proceed using an interpolation technique : the nearest neighbor to the station with a land sea mask >=0.5. Three types of uncertainties are discussed: the first type of uncertainty is related to the reanalyses themselves, the second is related to the sensitivity of heat waves frequency to the threshold values used to monitor them; and the last one is linked to the choice of indicators and the methodology used to define heat waves. Three sorts of heat waves have been analysed, namely those occurring during daytime, nighttime and both daytime and nighttime concomitantly. Four indicators have been used to analyse heat waves based on 2-m temperature, humidity, 10-m wind or a combination of these. We found that humidity plays an important role in nighttime events; concomitant events detected with wet-bulb temperature are more frequent and located over the north Sahel. For all indicators, we identified 4 years with a significantly higher frequency of events (1998, 2010, 2016 and 2019) possibly due to higher sea surface temperatures in the Equatorial Atlantic corresponding to El Nino events. A significant increase in the frequency, duration and intensity of heat waves in the cities has been observed during the last decade(2012-2020); this is thought to be a consequence of climate change acting on extreme events.
"

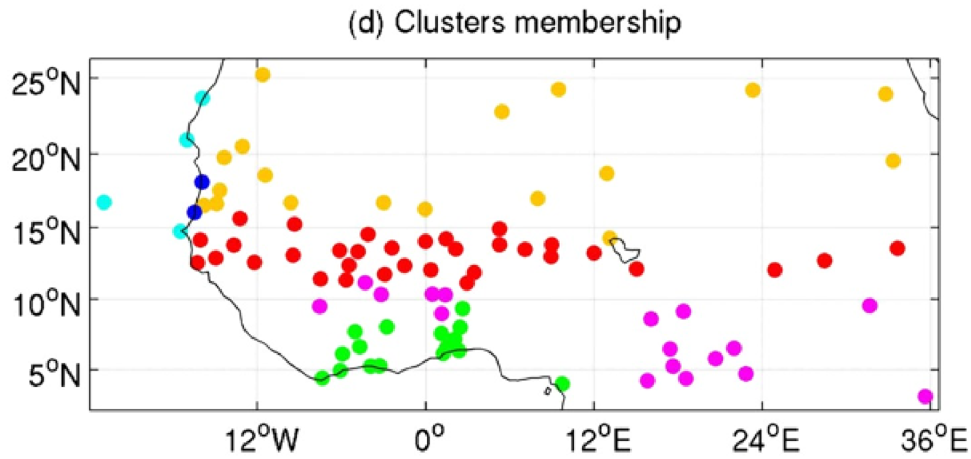
The analysis conducted to define the three areas is not provided (this is highlighted by the authors). Since this division is a core aspect of the paper, I think it is important to provide, at least, a description of the method followed to obtain these three areas (it can be included as Supplementary Material if the authors consider that it is too dense for the main document).

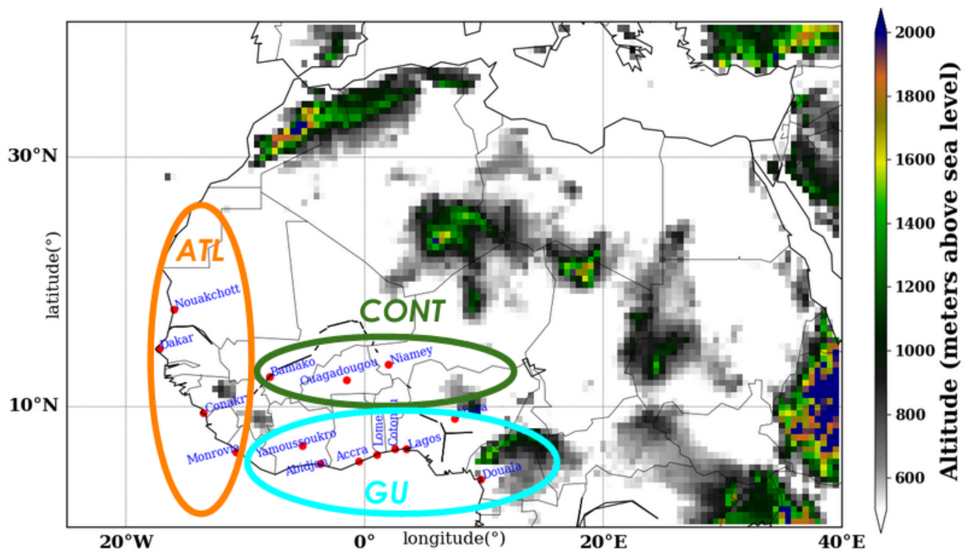
Thanks for this remark.
We added some description of the method in the manuscript.

We changed:
"The choice of these regions has been validated by conducting some analyses over the cities belonging to each region (not shown). The repartition of the different climatic regions is given as follows :
– Continental zone (CONT hereafter) including the cities of Bamako, Ouagadougou and Niamey [Fig1];
– Coastal atlantic zone (AT hereafter) including the cities of Dakar, Nouakchott, Monronvia and Conakry [Fig1];
– Coastal Guinean zone (GU hereafter) including the cities of Yamoussoukro, Abidjan, Lomé, Abuja, Lagos, Accra, Cotonou and Douala [Fig1]."

To:
" The choice of these regions is coherent with Moron et al. (2016) who used a hierarchical clustering approach to define some blocs of cities over West Africa. The fifteen cities investigated here have been classified in three regions as follows:
– Continental zone (CONT hereafter) including the cities of Bamako, Ouagadougou and Niamey [Fig1];
– Coastal atlantic zone (ATL hereafter) including the cities of Dakar, Nouakchott, Monrovia and Conakry [Fig1];
– Coastal Guinean zone (GU hereafter) including the cities of Yamoussoukro, Abidjan, Lomé, Abuja, Lagos, Accra,
Cotonou and Douala [Fig1].

The CONT and GU regions are very similar to the clusters found by Moron et al. (2016) (see figure below under the title 'Clusters membership'). The ATL region is a specific case because all the cities belonging to the region are not present in the clusters defined by Moron et al. (2016). Therefore, we have investigated the spatial variability of heatwave characteristics for each city in the ATL region. As result, we found coherent evolution between the cities (see [FigS1] in supplement material for maximum values of T2m using the 90th percentile as threshold); and we put them together to form the ATL bloc."





Besides, the authors note that their interest is in the coastal zone of West Africa region (lines 109). In that case, I am not sure why the analysis of a 'Continental zone' is needed. I would suggest to either rephrase the 'focus' on the coastal area or take out the continental zone analysis.

We rephrased the sentence according to the reviewer's suggestion.

We changed :

“In this study, we are interested in the coastal zone of West Africa, therefore, we identified three regions based on their location and their climate variability on which we conducted our analyses.”

to

“In this study, we are interested in the coastal and continental parts of West Africa, therefore, we identified three regions based on their location and their climate variability on which we conducted our analyses.”

In section two: Region of interest, Data and Methods, I would suggest the authors rearranging the contents to have only three subsections: 2.1 Region of interest; 2.2 Data and 2.3 Methods (with the corresponding sub-subsections). In 2.2 Data, for instance, I would suggest including the general information on the different reanalysis used (resolution, time-period, climate variables, etc.) as well as the information on the local station data (location, source, time-period, climate variables, percentage of missing values, quality of the series, etc.). Now it is not easy nor clear to find which local information has been used and its characteristics.

We reorganized section 2 according to the reviewer's remarks.

We changed :

“2. Region of interest, Data and Methods
2.1 Region of interest
2.2 Heat wave monitoring: Data and indicators
2.3 Methods ”

to:

“2. Region of interest, Data and Methods
2.1 Region of interest
2.2 Data
2.3 Methods ”

We added some clarifications in section 2.2 on the data.

We changed :

" In this work, to access information with a regular spatial grid and a large horizontal coverage, we used two state-of-the-art reanalysis products: the fifth-generation European Center for Medium-Range Weather Forecasts (ECMWF) 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). The spatial resolution of the products is 0.25°x0.25° and 05.°x 0.625° for ERA5 and MERRA-2, respectively. To be consistent in our analyses, we transformed the spatial resolution of MERRA-2 from 0.5°x0.625° to 0.25°x0.25° to match the one of ERA5. This is done using a first order conservative interpolation. We use hourly data covering the period going from 1 January 1993 to 31 December 2020 for all the reanalyses. We focus on atmospheric variables at the surface such as 2-meter temperature (T2m), 2-meter relative humidity (Rh), 2-meter dew-point temperature, 2-meter specific humidity, 10-meter wind components and water vapor pressure (e) from which wet bulb temperature (Tw) and Apparent Temperature (AT; (McGregor et al., 2015)) have been computed. Daily minima and maxima values were computed for T2m, Tw, AT and the Universal thermal Comfort Index (UTCI; (Di Napoli et al., 2021)). "

to:

" Reanalysis products are often taken as an alternative solution to observational weather and climate data due to availability and accessibility problems, particularly in data-sparse regions such as Africa (Gleixner et al. (2020)). In this work, to access information with a regular spatial grid and a large horizontal coverage, we used two state-of-the-art reanalysis products: the fifth-generation European Center for Medium-Range Weather Forecasts (ECMWF) 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). The ERA5 reanalysis has a native spatial resolution of 0.28125 degree (~31 km) with 137 hybrid sigma/pressure levels from the surface up to 80 km, yet downloaded data are interpolated to a regular latitude/longitude grid of 0.25°x 0.25°. ERA5 is produced using 4D-Var data assimilation and model forecasts in CY41R2 of the ECMWF Integrated Forecast System (IFS). The MERRA-2 reanalysis has a spatial resolution of 0.625°x0.5° with 42 standard pressure levels. MERRA-2 is using 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). These two reanalyses dataset are assessed through the Climserv database from the Institut Pierre Simon Laplace (IPSL) server. To be consistent in our analyses, we transformed the spatial resolution of MERRA-2 from 0.625°x0.5° to 0.25°x0.25° to match the one of ERA5; this is done using a first order conservative interpolation. We use hourly data covering the period going from 1 January 1993 to 31 December 2020 both for ERA5 and MERRA. Our choice of ERA5 and MERRA-2 to conduct this study is supported by some previous work showing that these two reanalyses are part of the most relevant used in Africa regions (e.g. , Barbier et al., 2018; Ngoungue Langue et al., 2021; Engdaw et al., 2022). As the main objective here is to process heat waves detection, we focus on atmospheric variables at the surface such as 2-meter temperature (T2m), 2-meter relative humidity (Rh), 2-meter dew-point temperature, 2-meter specific humidity, 10-meter wind components and water vapor pressure (e) from which wet bulb temperature (T w) and Apparent Temperature (AT ; McGregor et al., 2015) have been computed. These atmospheric variables have a significant impact on human thermal comfort (McGregor et al., 2015). Daily minima and maxima values were computed for T2m, T w, AT and the Universal thermal Comfort Index (UTCI; Di Napoli et al., 2021).

The land sea mask dataset used in this work has been derived from ERA5 reanalysis; it can be accessed on the Copernicus Data Store (CDS). T2m daily maximum and minimum observations at Dakar-Yoff station in Senegal and Aéroport Félix Houphouët Boigny (FHB) station in Ivory Coast have been used to evaluate our interpolation method. This is because we do not have access to other station datasets in the regions. The data from Dakar-Yoff extend from 1 January 1973 to 31 December 2018 containing almost 16% of missing values; and the data from Aéroport FHB are from 1 January 2005 to 31 December 2017 with 0.35% missing values. These data have been provided by some colleagues at Agence Nationale de l'Aviation Civile et de la Météorologie (ANACIM) for the Dakar-Yoff station, and Institut des Géosciences de l'Environnement (IGE) for the Aéroport FHB station."

Although the authors have clearly stated which are the uncertainties that have been studied in the paper, there is a general lack of justification why each number of choices is enough to characterize each uncertainty. Why using only ERA5 and MERRA, for example? There are other reanalyses. Perhaps is it ok to stay with these two, though. In any case, there is a need to better justify whether two are enough to 'characterize' (which implies some sort of specific quantification from a statistical point of view) or, conversely, if they can only be used to 'illustrate' the magnitude of the uncertainties can be important enough to affect the conclusions.

Thanks for this valuable comment.
First of all, we agree with the reviewer that there are others reanalysis products than ERA5 and MERRA; our choice on ERA5 and MERRA-2 to carry this study is supported by some previous work which show that these two reanalyses are part of the most relevant reanalyses used in Africa regions (e.g. Engdaw et al 2022, Ngoungue et al 2021, Barbier et al 2018). We added this information to the main document.

We changed :
"We use hourly data covering the period going from 1 January 1993 to 31 December 2020 for all the reanalyses."

to:
"We use hourly data covering the period going from 1 January 1993 to 31 December 2020 both for ERA5 and MERRA. Our choice on ERA5 and MERRA to carry this study is supported by some previous work which show that these two reanalyses are part of the most relevant reanalyses used in Africa regions (e.g. , Barbier et al., 2018; Ngoungue Langue et al., 2021; Engdaw et al., 2022)."

The same applies to the uncertainties linked to thresholds and the different ways to define a heatwave. There is a need to better justify and discuss why the authors think their choice of methods and thresholds is enough to map the uncertainties and to which degree this could be achieved.

The choice of the thresholds used to evaluate the uncertainties on heat wave detection is based on previous work. We will justify and discuss our choice of threshold values to conduct the sensitivity analysis in the main document.

We changed :
" As we have seen previously in the section "heat wave detection", the threshold value used for heat waves monitoring has a significant impact on the characteristics of the heat waves. The threshold value is related to the application we want to achieve. In this part of the work, we investigate the sensitivity of heat waves occurrence on different thresholds. To achieve this goal, we define 4 relative

threshold values computed over the entire period: the 75 th , 80 th , 85 th and 90 th daily percentiles "

to :

" As we have seen previously in the section "heat wave detection", the threshold value used for heat waves monitoring has a significant impact on the characteristics of the heat waves. The threshold value is generally tailored to the application one wants to achieve. In this part of the work, we investigate the sensitivity of heat waves occurrence on different thresholds. To achieve this, we define 4 relative threshold values computed over the entire period: i.e. the 75th, 80th, 85th and 90th daily percentiles. The choice of these thresholds to evaluate the changes on heat wave detection is based on previous work. Many studies are using the 90th percentile to define a heat wave (e.g., Fischer and Schär, 2010; Perkins et al., 2012a, b; Déqué et al., 2017; Lavaysse et al., 2018; Barbier et al., 2018); other studies are using the 75th percentile (Guigma et al., 2020). Based on these studies, we decided to test the sensitivity of threshold values from the third quartile (75th) to 90th percentile by steps of 5% to quantify significant changes in the occurrence of heat wave events. As we are studying extreme events, it is not relevant to go below the third quartile; knowing also that this study focuses on human impacts of heat waves, the 90th percentile is enough as a maximum threshold.``

The choice of the methods used to evaluate the sensitivity on heat wave detection is based on previous work. We justified and discussed our choice of the methods to conduct the sensitivity analysis in the main document.

We changed :

"Heat waves are usually defined as consecutive days of extremely hot temperatures above a threshold value of temperature (e.g., Tan et al., 2010; Gasparrini and Armstrong, 2011; Perkins and Alexander, 2013; Wang et al., 2019). 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 universal and standard definition of a heat wave (Perkins, 2015; Oueslati et al., 2017; Shafiei Shiva et al., 2019). Different thresholds, duration and indicators contribute to divergence in defining heat waves (Smith et al., 2013). Heat waves can be defined from daily meteorological variables such as daily raw temperature (T min , T mean and T max) (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), mean daily wet bulb temperature (Yu et al., 2021) or heat stress indices (e.g., Robinson, 2001; Fischer and Schär, 2010; Perkins et al., 2012; Guigma et al., 2020) 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). In our case, we use the daily min and max values of: T2m (T 2m min , T 2m max), T w (T w min , T w max), AT (AT min , AT max) and UTCI (U T CI min , U T CI max) as indicators for the detection of heat wave events. Three types of heat wave were detected (namely those occurring during daytime, nighttime and both daytime and nighttime concomitantly) using the following methods (see [Fig2]):"

To:

"Heat waves are usually defined as consecutive days of extremely hot temperatures above a threshold value of temperature (e.g.,Tan et al., 2010; Gasparrini and Armstrong, 2011; Perkins and Alexander, 2013; Wang et al., 2019). 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 universal and standard definition of a heat wave (Perkins, 2015; Oueslati et al., 2017; Shafiei Shiva et al., 2019). Different thresholds, duration and indicators contribute to divergence in defining heat waves (Smith et al., 2013). Heat waves can be defined from daily meteorological variables such as daily raw temperature (Tmin, Tmean and Tmax) (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), mean daily wet bulb temperature (Yu et al., 2021) or heat stress indices (e.g., Robinson, 2001; Fischer and Schär, 2010; Perkins et al., 2012a; Guigma et al., 2020) using relative or absolute thresholds. The use of absolute thresholds is very interesting to detect heat waves during the year in regions where the seasonal cycle is well marked. In mid-latitudes for example, the difference of T2m between the summer and winter is very important, approximately +30°C. Using this approach in tropical regions is not suitable, because the seasonal cycle is not so well marked; therefore a relative threshold for heat wave detection is more adapted. Some authors use the daily anomalies of temperature to define heat waves (e.g. Stefanon et al., 2012; Barbier et al., 2018). Most of the previous studies are focused on daytime or nighttime heat waves, ignoring events which occur during the day and night concomitantly. These type of heat waves are very dangerous for human health because the body suffers from heat stress during the day and night (Lavaysse et al., 2018). In our case, we defined 3 methods to detect specific type of heat waves (namely those occurring during daytime, nighttime and both daytime and nighttime concomitantly) using the daily min and max values of: T2m (T2min, T2max), Tw (Twmin, Twmax), AT (ATmin, ATmax) and UTCI (UTCImin, UTCImax) as indicators. The selected atmospheric variables have been used for heat wave detection in previous studies; they take in account some key variables (air temperature, wind, humidity, radiant temperature) to assess the body heat stress and they are easy to compute. The methods applied are defined below : ").

There is a need to further explain the downscaling method applied as well as the need for it. The method is not clear, nor how is it applied, as well as which stations were used and why. If I'm not mistaken, the method is applied because the reanalysis products are not enough to go to the city level, and there are not enough stations in the cities of interest to just use point station data. If this is the case, this has to be better explained in the Methods subsection (and, possibly, in the introduction and conclusions sections, too). Hence, I would suggest the authors to expand this section or provide a more detailed description of the methodology as Supplementary material.

We explained in more details the interpolation technique used in this study, which we completed as well in the paper.

We changed :
"Climate models used for weather studies are generally run at global scale, therefore information at local scale is missing in many regions; this is a critical issue. To overcome this problem, downscaling methods can be used. In this work, we studied phenomena at the scale of the city while our products have much coarser spatial resolution. In this context, we need a downscaling approach to attribute variables of interest from global to local scale. Another problem we faced is that most of the cities are located along the coast and influenced by the ocean flow (see [Fig1]). The evaluation of the spatial variability of the correlation between the local scale variable (station) and reanalyses (ERA5), showed high correlation values over the continent [FigS8]. To estimate the proportion of land on a grid point, we used the land sea mask whose values range from 0 to 1. A land sea mask (lsm) of 0 means no land (a point located in the ocean), and a lsm of 1 means that the model cell is fully covered by land. Hence, to estimate the temperature over the city using reanalyses, we use the nearest grid point of reanalyses to the station which satisfies a lsm equal or greater than 0.5 (see [Table3] for lsm values of all the cities considered in this study)."

To:

“ Reanalysis dataset used for weather studies are generally run at global scale, therefore information at local scale is missing in many regions; this is a critical issue in regions where there is a lack of observation stations as is the case of African cities. To overcome this problem, sometimes downscaling methods can be used. In this work, we study phenomena at the scale of the cities and reanalyses (ERA5 and MERRA) have too coarse a spatial resolution. The scales of the reanalyses are more representative of the spatial variability of a heat wave occuring in a city than an isolated local stations. Nevertheless, a certain validation must be conducted of testing stations, especially to find the best interpolation technique to estimate local temperature from the reanalyse. This is especially important over the coastal regions. Indeed, most of the cities used in this study are located along the coast and influenced by the ocean air masses (see [Fig1]). The evaluation of the spatial variability of the correlation between the local scale variable (station) and reanalyses (ERA5) for T2m for example, showed high correlation values over the continent [FigS2] (Dakar, Abidjan). This suggests that the station data are well correlated with ERA5 grid points which are located on the continent; so there is a need to know whether a ERA5 grid point is over the continent or not before applying an interpolation technique. To estimate the proportion of land on a grid point, we used the land sea mask (lsm) whose values range from 0 to 1. The land sea mask is a measure of the land occupation on a grid point. A lsm of 0 means no land (a grid point located in the ocean), and a lsm of 1 means that the model cell is fully covered by land. Therefore, to estimate the climate variables over the cities from reanalyses, we use the nearest grid point of reanalyses to the station which satisfies a lsm equal or greater than 0.5 (see [Table1] for lsm values of all the cities considered in this study). This approach was chosen after evaluating different methods such as (see [FigS3 a)] for more details) :

– a bilinear interpolation using the four nearest grid points of reanalyses around the station [FigS3 (a,d)];

– a linear gradient approach which considers that the gradient of temperature is constant between two grid points based on a linear interpolation with a condition on the lsm value (>0.5) [FigS3a (c,f)];

– the selection of the nearest grid point of reanalyses from the station with different values of lsm (>=0.5, 0.75 and 1; we only show for lsm>=0.5) [FigS3a (b,e)] “

- a dynamical downscaling approach taking into account the effect of winds (not shown).

The use of relative thresholds to establish heatwave duration for all the year, though it is systematic, implies that for some regions and periods, the 'heat waves' have different impacts. It could happen that for some regions and periods, though formally there could be a heat wave, in practice, there would not be any impact at all from it. This needs to be highlighted and discussed, to justify that, in any case, the analysis for all the year it is still useful.

We agree with the reviewer point of view about the definition of relative thresholds to process heat waves detection over the year; this is especially true for mid-latitude regions. First of all, our region of interest is West Africa, and in this region the seasonal cycle is not as well marked as in the mid-latitudes. The seasonal thermal amplitude is about 6 °C. Secondly, this study is part of the STEWARd (STatistical Early WArning systems of weather-related Risks from probabilistic forecasts, over cities in West Africa) project which focuses on climate extremes human impacts. Therefore, an estimation of the intensity of heat waves is added by using a fixed yearly threshold (the minimum of the daily climatology percentiles over the period) from which we computed the daily exceedance of hot days. Using this approach, we can clearly evaluate the severity of a heat wave and its potential human impacts which will be higher when occurring during the hottest period of the year . This has been clarified in the manuscript as follows.

We changed :
"Heat waves can be defined from daily meteorological variables such as daily raw temperature (Tmin , Tmean and Tmax) (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), mean daily wet bulb temperature (Yu et al., 2021) or heat stress indices (e.g., Robinson, 2001; Fischer and Schär, 2010; Perkins et al., 2012; Guigma et al., 2020) using relative or absolute thresholds."

To:
" Heat waves can be defined from daily meteorological variables such as daily raw temperature (Tmin , Tmean and Tmax) (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), mean daily wet bulb temperature

(Yu et al., 2021) or heat stress indices (e.g., Robinson, 2001; Fischer and Schär, 2010; Perkins et al., 2012; Guigma et al., 2020) using relative or absolute thresholds. The use of absolute thresholds is well suited to detect heat waves during the year in regions where the seasonal cycle is well marked. In mid-latitudes for example, the seasonal thermal amplitude of T2m is large, approximately 20°C. In tropical regions this method is not suitable since the seasonal thermal amplitude is strongly reduced (6°C). Therefore a relative threshold for heat waves detection is adopted in our study as our region of interest is West Africa."

We change :

" The mean intensity of a heat wave has been defined as the sum of the daily exceedance of daily values of indicators over the daily threshold in a sequence of hot days divided by the total number of heat waves. In the scope of this study, we are interested in human impacts of heat waves, therefore we defined a constant threshold value over the whole period to compute the intensity. "

To :

"The mean intensity of a heat wave has been defined as the sum of the daily exceedance of daily values of indicators to the climatological daily threshold in a sequence of hot days divided by the total number of heat waves. This study 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) project which focuses on climate extremes human impacts. 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. From this approach, we can properly evaluate the severity of a heat wave and its potential impacts on humans."

When performing the comparison through statistical metrics, besides clearly stating that the data is downscaled, I would also suggest comparing the 'downscaled values' with the station ones (whenever possible).

This is done in the validation process of our method through the anomaly of correlation between the different interpolation methods and the station datasets. We added this result to supplement material (see [FigS3a]).

It would also be needed to specify why choosing ERA5 as the reference (instead of MERRA or any other station network).

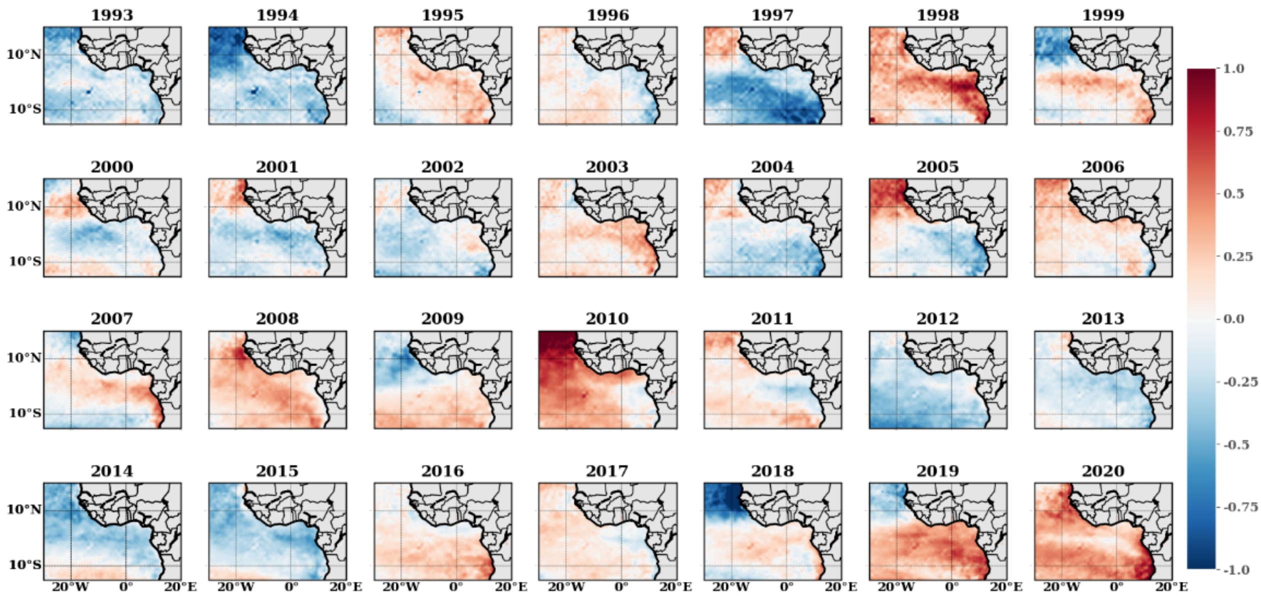
Thanks to the reviewer for this remark.

The choice of ERA5 as reference for this analysis is based on previous work. For instance, Olauson, 2018 and Ramon et al., 2019 found that ERA5 provides a good representation of various near surface meteorological variables including near surface humidity and wind speed in comparison to others reanalyses including MERRA. However in order to clarify this point, we added in the manuscript some analyses on heatwaves evolution with MERRA reanalysis. Reanalyses are more representative of the spatial variability of the city than a local station.

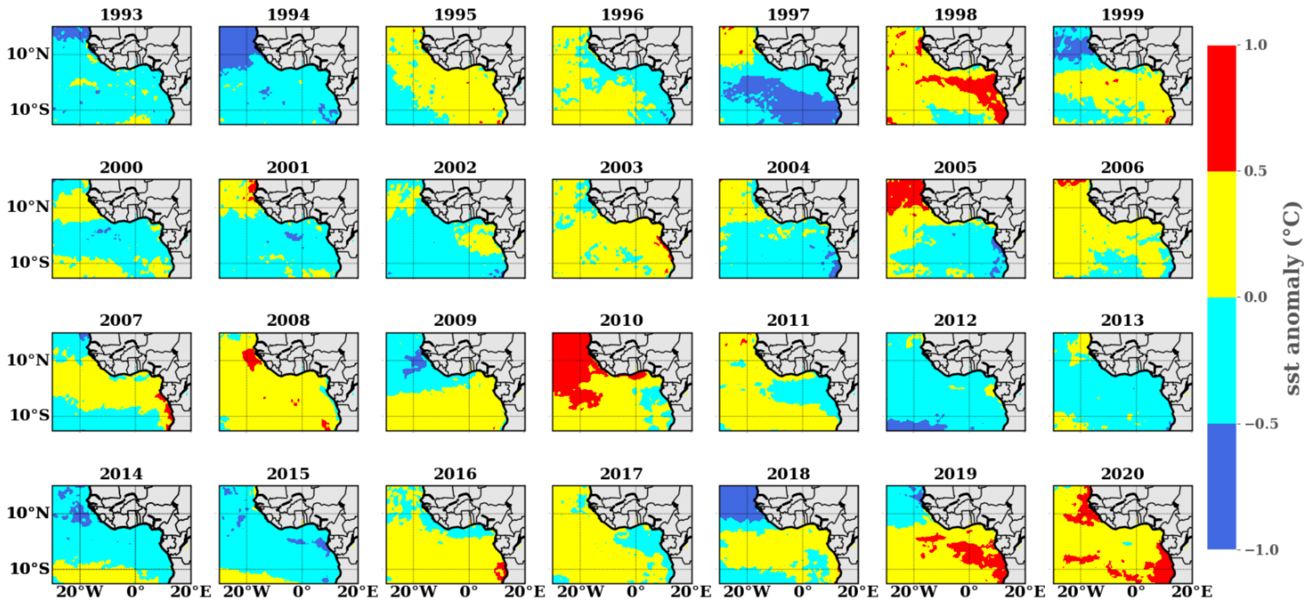
In the maps at the end, I would suggest using discrete colour bars (continuous ones are not suitable for assigning values). I would also include the cities of interest in all the maps (since this is the focus of the paper).

Thanks to the reviewer for this valuable suggestion. We plot the maps accordingly.

We changed:



To:



We changed :

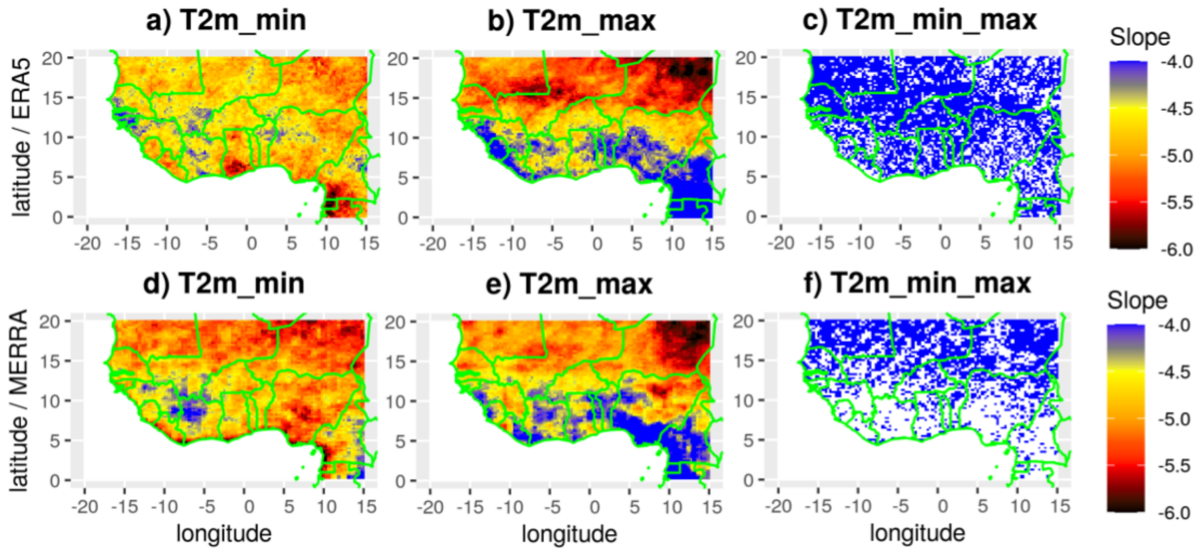


Figure 5. *Evolution of the heat wave duration with respect to the threshold values using T2m as indicator respectively for : a-c) ERA5 and d-f) MERRA. The slope is computed using the* $75^{th}, 80^{th}, 85^{th}$ *and* 90^{th} *percentiles. X and Y-axis respectively represent the longitude and latitude in degrees. The color bar shows the values of the linear evolution of the duration of heat wave per percentile. The white blanks indicate non significant changes in the duration of heat waves per percentile.*

To :

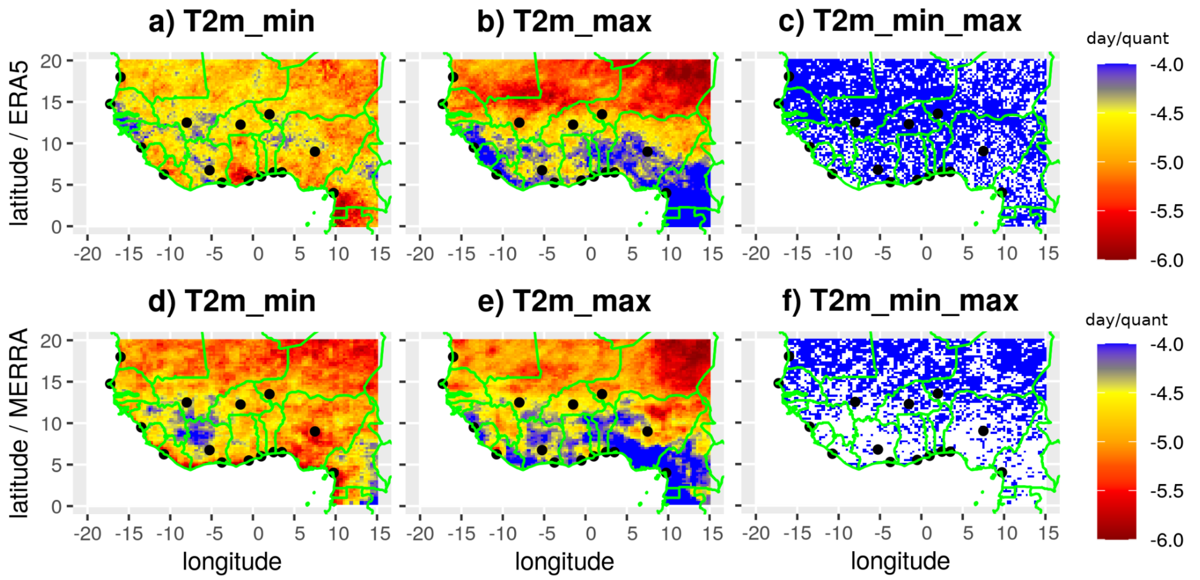


Figure 5. *Evolution of the heat wave duration with respect to the threshold values using T2m as indicator respectively for : a-c) ERA5 and d-f) MERRA. The slope of the regression line in day per percentile is computed by fitting a linear regression between the threshold values (75, 80, 85, 90) and their corresponding heat waves's duration (D75, D80, D85, D90) . X- and Y- axis respectively represent the longitude and latitude in degrees. The color bar shows the values of the slope. The white blanks indicate non-significant changes in the duration of heat waves per percentile.The significance of the slope of the regression line has been computed using a two-sided Chi-square test.*

Figure 4 depicts the slope of the linear regression in heatwaves (also figure 5). The caption says that the slope is computed using the 75th, 80th, 85th and 90th percentiles, but there is only one map per variable and reanalysis. Does this mean that the trend is computed for all the four percentiles simultaneously? This is not very clear when reading the methods section, and a rephrasing and/or extension of the description would be advisable. Besides, there is also a need to state more clearly (both in the methods and in the captions) which method has been applied to compute the significance of the slope. That said, if the slope is computed with all the four thresholds simultaneously, it wouldn't be a conventional approach and, consequently, a more thorough justification about its correctness and its utility would be needed (compared to performing the analysis independently for each threshold).

Thanks to the reviewer for this comment.

The assessment of changes in heat waves occurrence or duration with respect to the threshold (75th, 80th, 85th, 90th percentiles) is processed independently for the 4 thresholds; this is done by the computation of the linear evolution coefficient over each grid point. The linear evolution coefficient is defined as the slope of the linear regression line fitted between the threshold values (Q75, Q80, Q85, Q90) and the number of events associated to each threshold (N75, N80, N85, N90) or their corresponding duration (D75, D80, D85, D90). In fact, for each grid point, a regression line is firstly fitted between the threshold values and their corresponding occurrence or duration; and secondly the slope is computed. We make it clear in the manuscript.

We changed:

" To quantify the changes of heat waves occurrence with respect to the threshold values, we analyse for each grid point the linear evolution of events detected and their duration. The linear evolution is computed by fitting a linear regression between the threshold values (75, 80, 85, 90) and the number of events associated to each threshold (N1, N2, N3, N4) or their corresponding duration (D1, D2, D3, D4). We are aware that this regression based on 4 points is not very robust, nevertheless it makes it possible to obtain information on the evolution of the heat wave characteristics with respect to the thresholds. Therefore, we evaluated the significance of the slope values according to the thresholds using a confidence level of 95%. "

To:

" The sensitivity in heat waves occurrence or duration with respect to the threshold (75th, 80th, 85th, 90th percentiles) is processed simultaneously for the 4 thresholds; this is done by the computation of the linear evolution coefficient over each grid point. The linear evolution coefficient is defined as the slope of the linear regression line fitted between the threshold values (Q75, Q80, Q85, Q90) and the number of events associated to each threshold (NQ75, NQ80, NQ85, NQ90) or their corresponding duration (DQ75, DQ80, DQ85, DQ90). The computation of the linear evolution coefficient is done by the following steps:

- After processing to heat waves detection at each grid point for the 4 thresholds separately, we compute for each of them the heat waves frequency and duration;

- then fitted a regression line between the threshold values (Q75, Q80, Q85, Q90) and their corresponding occurrence or duration. This is done for each grid point;

- Finally, the changes in heat waves occurrence/duration from the 75th to 90th percentiles at each grid point, is evaluated by the computation of the slope of the regression line fitted at step 2 between the threshold values and their corresponding heat waves occurrence/duration.

We are aware that this regression based on 4 points is not very robust, nevertheless it makes it possible to obtain information on the evolution of the heat wave characteristics with respect to the thresholds. Therefore, we evaluated the significance of the slope values according to the thresholds using a confidence level of 95%. The significance of the slope has been evaluated using a two sided Chi-square statistics test (Pandis, 2016)."

Technical corrections

Figure 6 lacks titles in the top row

This is correct, we add titles in the top row.

Figures should include units when necessary. For example, in figure 1, 'meters above sea level; in figure 6 it would be 'number of days' or 'Number of events / occurrences'); in figure 3 and figure 4, number of events or days / year; figure 7; figure 11...

We add units in figures when necessary.

It is not clear how figure 2 has been obtained. Is it built with data from all the stations?Cities? Grid-points? It is just an illustration for a single grid-point? This has to be included in the caption (as well as in the main text).

In fact, figure2 represents a schematic illustration of the different types of heat waves analysed in this paper. It is not obtained from a specific station nor grid point.

We changed:

"Figure 2. Detection process of heat wave: HW1/HW2 represent events associated respectively to maxima/minima temperature, HW3 are events detected at same time in maxima and minima temperatures. 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."

To :

“Figure 2. Detection process of heat wave: HW1/HW2 represent events associated respectively to maxima/minima temperature, HW3 are events detected at the same time in maxima and minima temperatures. 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 XXth percentile. This figure is a 'schematic' illustration of the different types of heat waves investigated in this work”

Sometimes x- axis and y- axis is written is capital letters and sometimes it is not.
The acronyms for variables should be the same in figures (titles, for instance), captions as well as in the main text.
We corrected the manuscript accordingly. We changed x- axis and y- axis to capital letters everywhere in the document.

Column titles in figure 8 and 9 are difficult to understand. Besides, the idea to display different parameters in the same format it is confusing (apparently, from the caption, 2nd, 3rd and 4th columns display percentages instead of duration of heatwaves). I would suggest to only maintain the same format when displaying the same elements.

Thanks to the reviewer for this suggestion, we reorganized the figure 8 and 9 accordingly.

We changed:

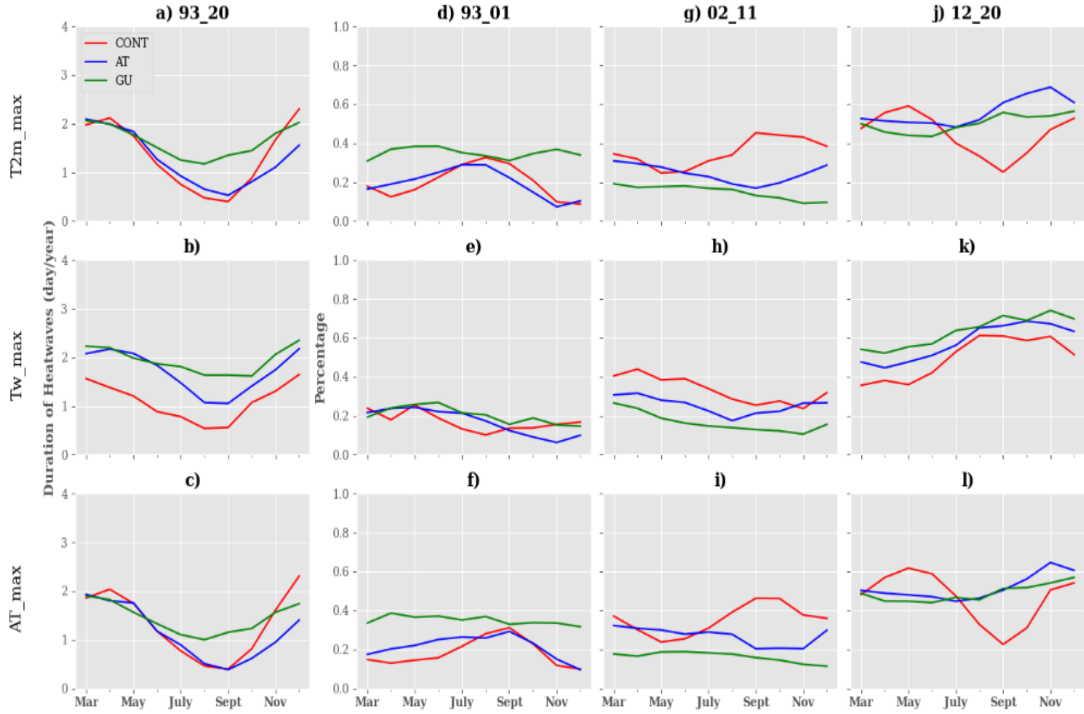


Figure 8. *Seasonal variability of heat wave yearly duration using maximum values of indicators: a) $T2m$, b) TW and c) AT. The first column shows the evolution of heat wave duration per year over the whole period 1993-2020. The $2^{nd}(d,e,f)$, $3^{rd}(g,h,i)$ and $4^{th}(j,k,l)$ columns represent respectively the contribution in percentage of the sub-periods 1993-2001, 2002-2011 and 2012-2020 to heat wave duration over the whole period. We compute a a 3-month running mean to smooth the seasonal cycle. The X-axis represents the time in days; and the Y-axis respectively the duration of heat waves, and the contribution of each decade. Red/blue/green lines represent the evolution of heat wave duration over CONT/AT/GU regions (see region of interest section for more details).*

To :

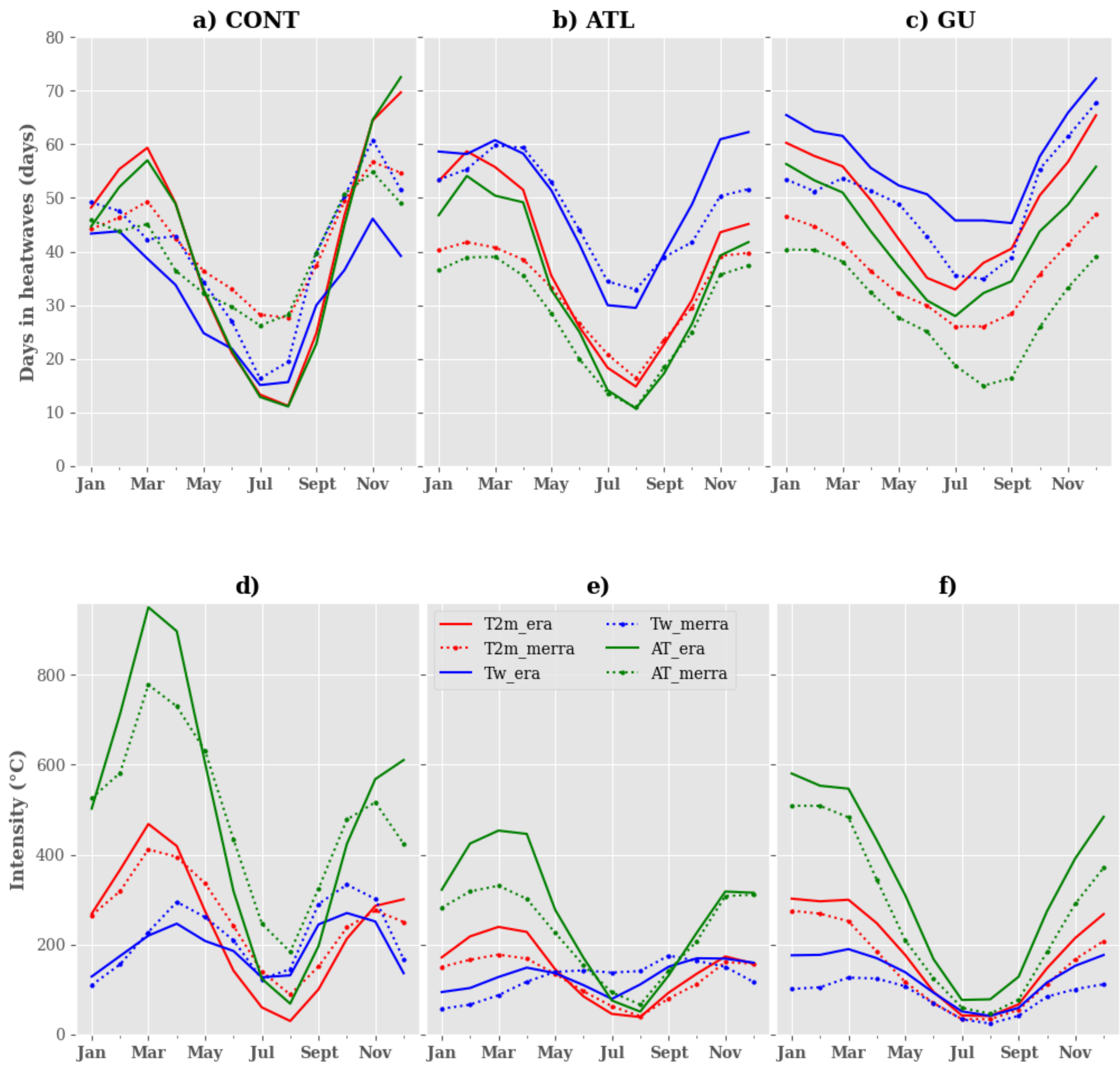


Figure 8. Seasonal variability of heat waves characteristics using maximum values of T 2m, T w, AT : a-c) duration and d-f) intensity. We compute a 3-month running mean to smooth the seasonal cycle. The detection of heatwaves is done using the 90th percentile as threshold over : CONT (a − d), ATL (b − e), GU (c − f) regions. Red/blue/green strong and dashed lines represent respectively the results using T2m, Tw, AT from ERA5 and MERRA. The Y- and X- axis represent the duration and intensity of heat waves and the time in month respectively.

We did the same for Figure 9 (not shown here)

In figure 10 it is not clear what those percentages refer to. Are percentages from the total of days? From the total of heat wave days? Do they have to sum 1 in total? The phrase 'using maximum values of indicators based on the duration' is not very clear, either. What does this refer to? The thresholds? The methods? The variables? This also extends to the other figures applying the same approach

Thanks to the reviewer for this comment; we clarify all these points in the paper.

Figure10 represents the classification in terms of duration of heat waves detected with the 90 th percentile as threshold using maximum values of indicators (T2m,Tw and AT) over the period 1993-2020. Firstly, we detect heat waves and compute their duration; after we construct clusters of heat waves based on their duration (3d, 4d-6d, 7d-9d, 10d-12d, +13d) and finally, we quantify the proportion of each class of heat waves to the total number of events detected.

We changed :

"Figure 10. Classification of the heat waves detected using maximum values of indicators based on the duration: a) T 2m, b) T W and c) AT. The X and Y-axis represent respectively the percentage of the heat wave per class and the duration in day. Red/blue/green bars represent the percentage of heat waves detected over CONT/AT/GU regions (see region of interest section for more details)."

To:

"Figure 10. Classification of the heat waves detected with the 90th percentile as threshold using maximum values of indicators based on their persistence over the period 1993-2020 : a) T2m, b) Tw and c) AT. Firstly, we detect heat waves and compute their duration; after we construct clusters of heat waves based on their duration (3d, 4d-6d, 7d-9d, 10d-12d, +13d) and finally, we quantify the proportion of each class of heat waves to the total number of events detected. The Y- and X- axis represent respectively the percentage of the heat waves per class and the duration in day. Red/blue/green bars represent the percentage of heat waves detected over CONT/ATL/GU regions (see region of interest section for more details). The sum of the contribution of heat waves in different clusters is equal to 1 for each region."