Note the following errata in the first submission that will be addressed in the revised publication:

In the whole text, the potential temperature indicator is said to be taken as the mean potential temperature between 925 and 750 hPa. We actually used the potential temperature at 700 hPa for the calculations as specified in the literature, the “750” in the text is a typing error. The authors want to apologize in advance.

Review of: “Saharan warm air intrusions in the Western Mediterranean: identification, impacts on temperature extremes and large- scale mechanisms” by Cos et al.

Summary and General Assessment

Warm air from the Sahara can be advected to the Mediterranean and possibly cause extreme temperatures in the Mediterranean and beyond. Cos et al. use an identification method based on the geopotential thickness between 1000 and 500 hPa, and the mean potential temperature between 925 and 750 hPa to investigate the so-called Saharan warm air intrusions into the Western Mediterranean. Using this method, they calculate a catalogue of past intrusions. The authors perform a clustering of intrusion days based on the distribution of geopotential thickness and mean potential temperature, and find three to five distinct clusters per season, that mainly differ in their geographical location. The transition probabilities between the clusters, the correlations with the teleconnection indices, and the composites of meteorological fields are computed per cluster per season. Extreme temperatures are often observed during intrusion days. Based on the conditional probabilities the authors argue that intrusions have a great impact on heat events, but not such a big contribution, i.e. intrusions often cause heat, but heat is not always caused by intrusions. The composites of the meteorological fields show that the different locations of the clusters are due to different locations of the driving pressure patterns.

The study is motivated by increasing temperatures in the western Mediterranean and the mention of such intrusions in the literature. While the connection between the intrusions and heat is not too surprising, the concept of Saharan warm-air intrusions into the western Mediterranean gives a new perspective on extreme temperature events in the region.

The overall structure of the study is well-organised, however some parts need improvement as they are not so easy to follow. The figures support the presented results mostly well, but their readability could be improved. Below I have major, minor, and figure-specific concerns, comments, and questions that should be addressed.

We want to thank the reviewer for taking the time to read the text and provide feedback that we find to greatly improve the quality of the manuscript.

Major comments

Section 3

In this section, you describe the algorithm used to identify the intrusions, based on the two indicator variables

$\Delta GH_{500-1000}$ and $\theta_{750-925}$. You refer to the work of Sousa et al. (2019), however, I still lack a motiva- tion/explanation for why these levels are useful to determine the air mass. Especially considering that daily and seasonal variations may change the height of the boundary layer height drastically, for example. Have you looked at how sensitive the results are to your choice of indicators?

Thanks for this very pertinent comment. We agree that it might not be enough to follow the methodology from Sousa et al. 2019. Their choices come from the climatological characteristics of the Sahara region described in the book “An Introduction to the Meteorology and Climate of the Tropics” by J.F.P. Galvin from 2016.

To verify that the definition can work for our case, we looked at the monthly climatologies of the vertical profiles of the troposphere over the Saharan and Mediterranean. Figure R1 shows the climatological vertical profiles in the Mediterranean region (blue) and in the Sahara (red) for every month of the year.

Summer months have a most distinct lower troposphere potential temperature in the two regions, especially above 950/925 hPa. This is an argument in favour of using 925hPa as the lower bound. From April to October, the potential temperature vertical gradient is larger between 1000hPa and around ~925hPa, then becomes lower from 925hPa to around 700/600hPa. This is the main characteristic of Saharan air masses in the warm months, which are warm and homogeneous in the vertical due to surface longwave radiation heating a thick layer of the lower troposphere over the Sahara.

In colder months, the Saharan and Mediterranean potential temperature vertical gradients are similar. Between November and March, the difference in the vertical profiles lies only in the magnitudes of the potential temperature in all vertical levels, but not in their vertical gradients. Even if the vertical temperature profile The difference in potential temperatures is still relevant to distinguish between Saharan and Mediterranean air masses (as seen in the inspection of the events), although the Saharan air mass is not as homogeneous in the vertical as in the warm months. An intrusion of such an air mass in the cold months should still be distinguishable and have an impact on the Mediterranean region.

In terms of geopotential height, Figure R1 gives little insight as the vertical gradient is much larger than the latitudinal gradient which separates the two regions. We plot Figure R2, which shows the monthly climatological geopotential thickness between different levels. The results suggest that the geopotential thickness between 500 and 1000 hPa has sufficient separation between its Saharan and Mediterranean climatologies. Nonetheless, we perform some sensitivity studies.

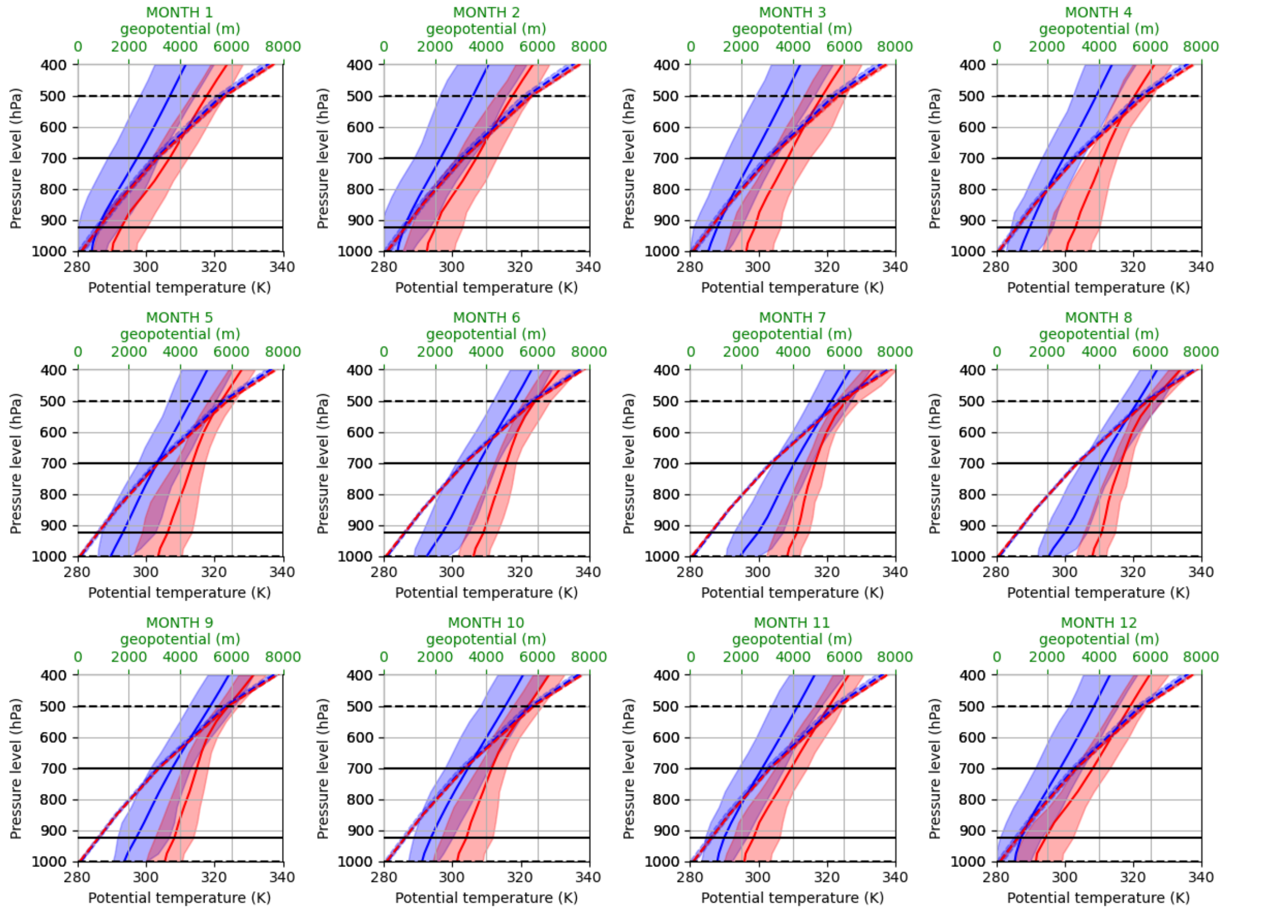


Figure R1: Monthly climatologies of vertical potential temperature (solid) and geopotential (dashed) profiles in the Mediterranean region (blue) and in the Sahara (red). The climatology of different months can be seen in the different panels. The shaded area is the interannual variability along the period 1981-2010.

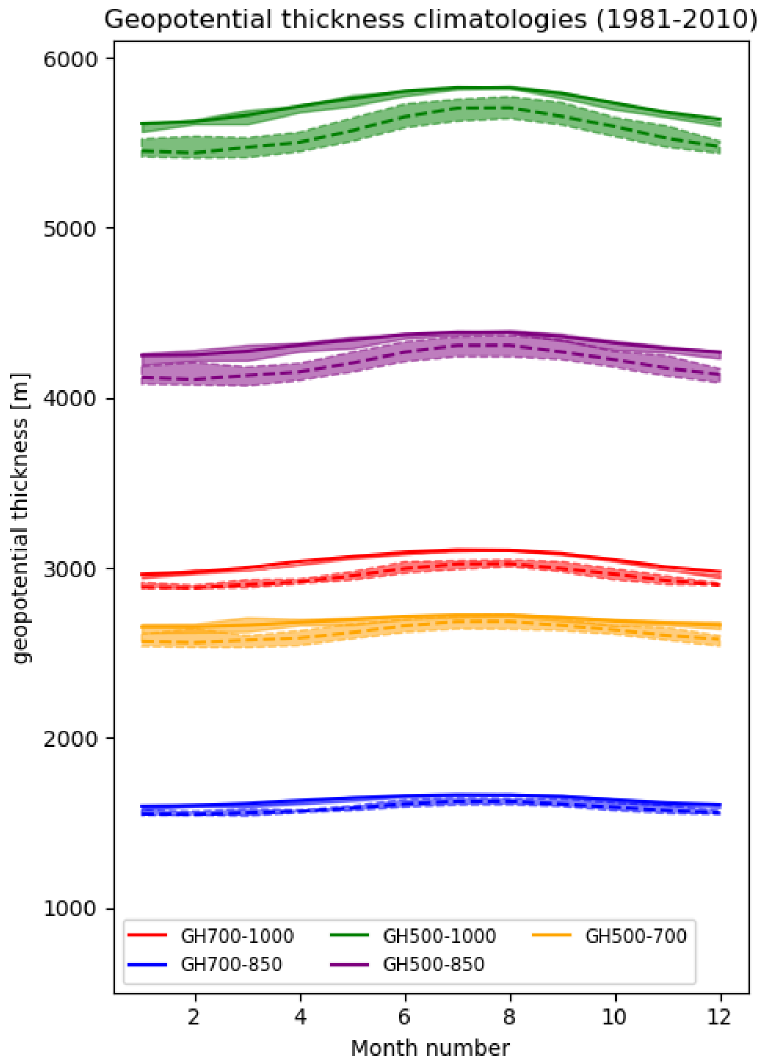


Figure R2: Climogram with monthly climatologies of geopotential thickness between different levels in the WMed (dashed) and in the Saharan (solid) regions. The shaded area is the interannual variability along the period 1981-2010.

A sensitivity study of the choice of the levels in the two metrics has been performed using different levels of geopotential thickness ($\Delta GH500-1000$ and $\Delta GH500-850$) and mean potential temperature ($\theta700-925$ and $\theta700-850$). The captured intrusions remain similar.

In summer, similar results are obtained for the combination of indicators θ700-925/ΔGH500-1000, θ700-850/ΔGH500-1000 and θ700-925/ΔGH500-1000 (The last one does not capture Sousa et al. 2019 events). θ700-850/g500-850 identifies longer events and exceeds the intrusion day classification from Sousa et al. 2019. In winter the choice of metric is not sensitive to the levels used. In spring and autumn, changes in the levels used have some small influence in the length of the events.

In general, using a higher lower potential temperature bound leads to more and lengthier events, and using a lower geopotential bound leads to a decrease in the amount and length of the events.

We will expand the sensitivity of the study on the levels in the revised version hoping to obtain a more robust manuscript.

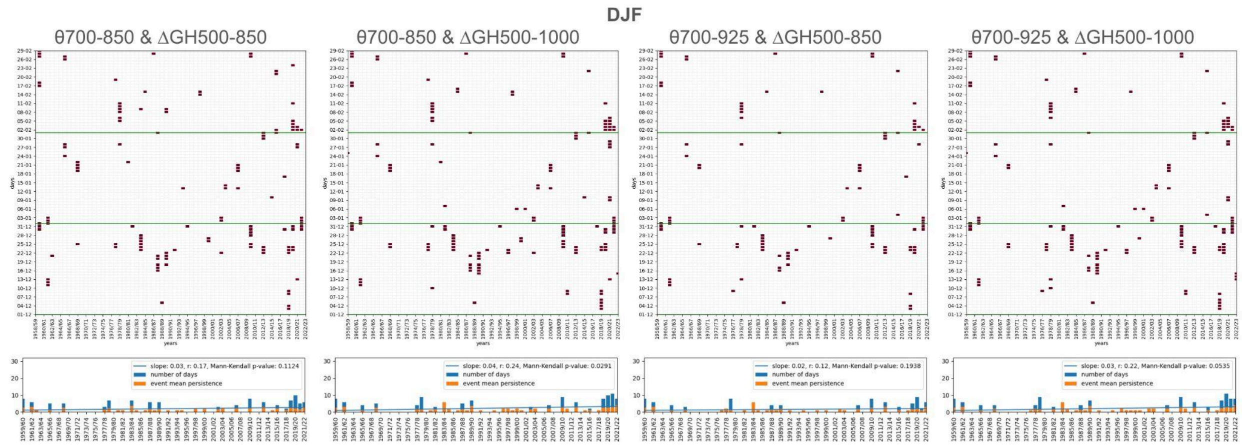


Figure R3: Catalogues of intrusion days in winter (DJF) from 1959/1960 to 2021/2022 using 4 different indicator combinations: i) θ700-850/ΔGH500-850, ii) θ700-850/ΔGH500-1000, iii) θ700-925/ΔGH500-850 and iv) θ700-925/ΔGH500-1000 (from left to right). The days with intrusion are painted in dark red, and the panels below the catalogues show the aggregated number of intrusion days and mean event duration during DJF for all the period. A least squares linear regression fit to the seasonal number of intrusion days is shown in blue.

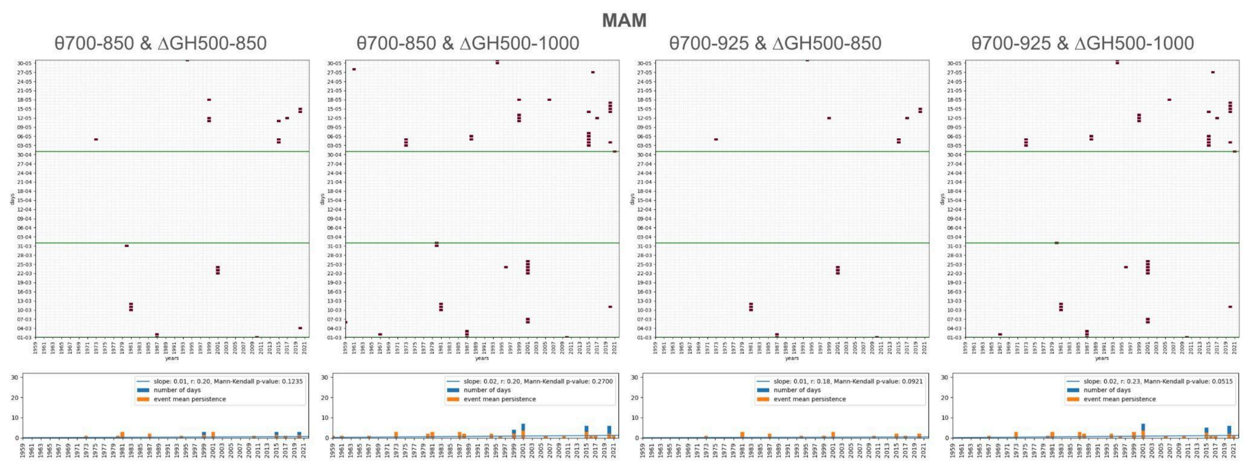


Figure R4: Same as R3 but for MAM in the period 1959 to 2022.

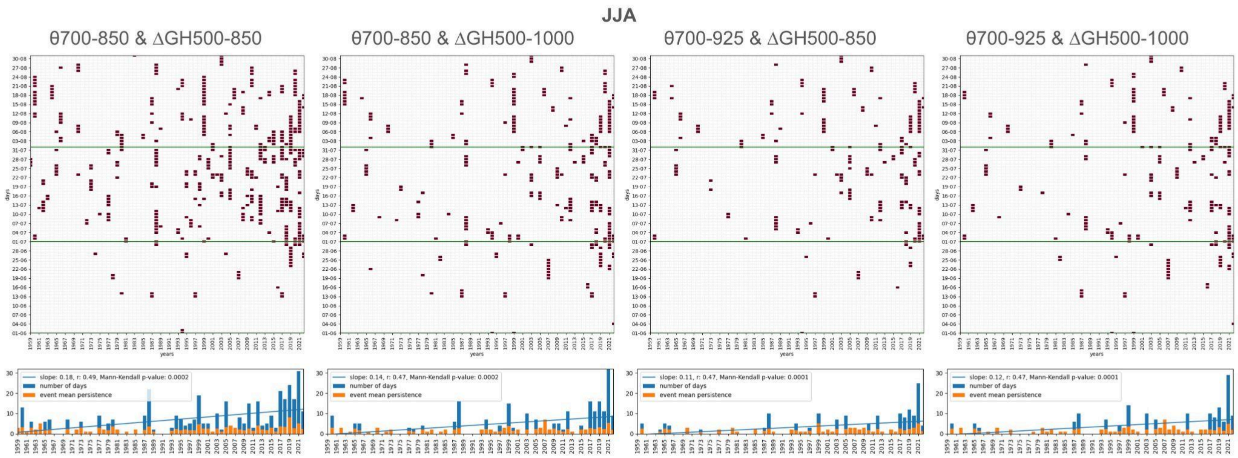


Figure R5: Same as R3 but for JJA in the period 1959 to 2022.

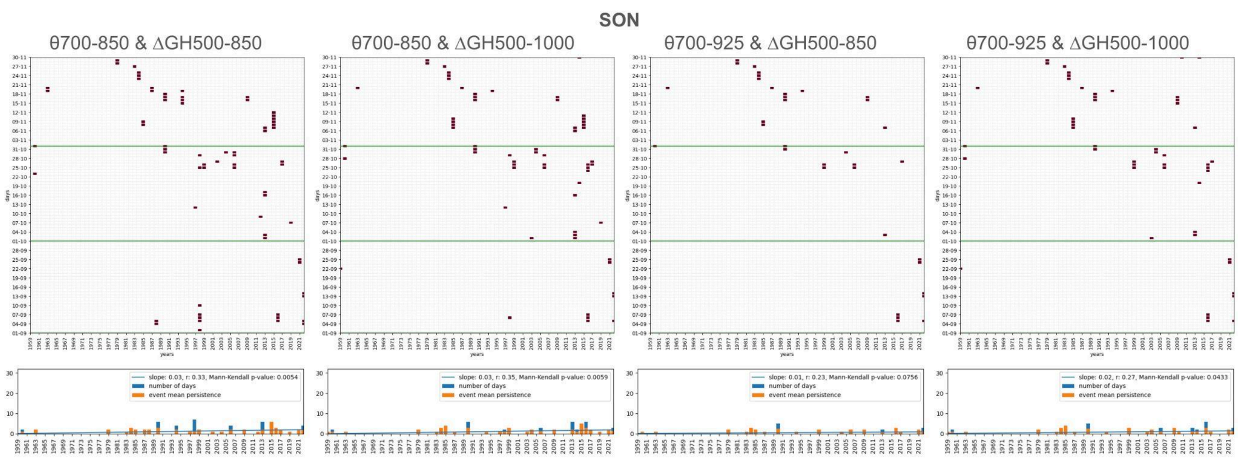


Figure R6: Same as R3 but for SON in the period 1959 to 2022.

l 91-105

This paragraph is very difficult to follow. It can be improved by rewriting/rephrasing, but since it is a crucial paragraph to the paper, I consider this a major comment. Let me summarise what I understood:

Monthly climatologies of ΔGH and θ show that in summer values are maximised over the Sahara, but during the rest of the year they decrease slowly and constantly between 15 and 35◦N.

Then, you choose a region as representative for the Sahara. You average over this region, which was chosen so that you meet the values described in literature. I am missing the exact extent of this region in the text.

Then, you calculate 31-day rolling means, averaged over the region, to define thresholds. Why do you chose 31d in particular and how sensitive is your result to this choice?

The air mass is then defined as all grid cells that exceed both threshold values. Does this always yield a coherent air mass as depicted in Fig 1? I would imagine there might be single cells within the air mass that are not identified, or single cells outside of it that are?

Please improve this paragraph by explaining step by step what you did and why.
In lines 109-115 you mention again that in the colder seasons, the values in the Sahara are not as distinct, since there is a constant gradient present. First, this is a little bit repetitive to what was mentioned before (l 92 following). Second, you do not really explain what the consequences of this are. I would imagine that the sensitivity to the threshold values varies a lot across seasons because of this? Does this mean your method works better in summer than in the rest of the year?

We see how this paragraph might be confusing and understand that it raises questions about sensitivity to parameter choice. Therefore, sensitivity studies have been conducted (Figures R1-R6) in order to better illustrate the implications of changing the pressure levels where metrics are computed. In the revised text we try to explain better the Sahara climatological properties, taking advantage of the figures that were produced thanks to Reviewer #1's previous question.

Answers to other questions:

Why using a 31-day rolling mean?
The original work by Sousa et al. 2019 uses the climatology of the summer months as thresholds for the whole summer. We identified that using the climatology of three months for the thresholds makes the identification criteria vary for events along the 3-month period. For instance, using the 3-month JJA climatology might make early June and late august events harder to identify than events in the middle of summer (closer to the climatological maximum temperature). Therefore we decided to make a rolling climatology that would take into account the seasonal cycle of the metrics. Different periods were explored, from 15-day to 3-month rolling means. The choice of 15 days was very restrictive, while three months was too permissive. After exploring the results we went for a 31-day rolling mean, which although arbitrary, is an educated guess.

Spatial continuity of the intrusions
We see, from the visual inspection of all the events, that there are no spurious masses generated outside the Sahara. From this inspection we could ascertain that during Saharan warm air intrusions, air masses are linked to air mass movements from the Sahara to the Mediterranean. Nevertheless, there are intrusion days when the Saharan air mass seem to split up, most likely due to orographic effects.

Consequences of the winter less distinct Saharan properties
The consequences of having less distinct values in the Sahara during the cold months is that air masses that are considered to be within the threshold of a Saharan air mass might not come directly from the Sahara, but maybe from the Atlantic ocean at Saharan latitudes. The properties of these air masses are still uncommon in the Mediterranean region in the winter season. That is why some intrusions might be more precisely referred to as subtropical ridges in this season.

We have modified the text so that the differences between seasons are better understood and the thresholds used fully justified.

l 107

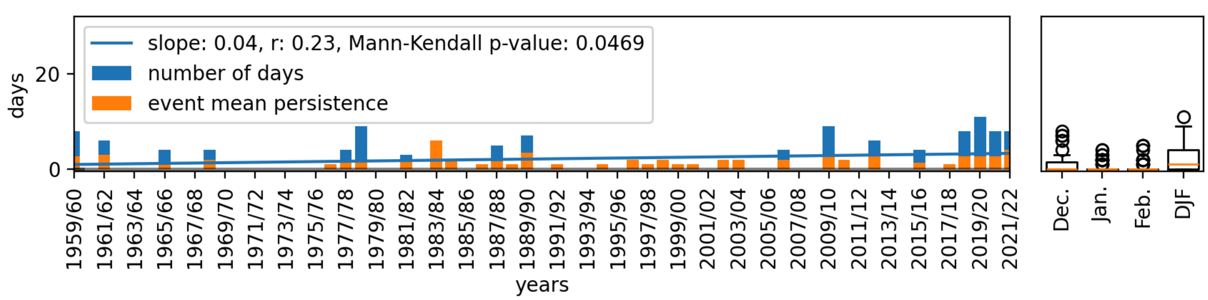
“air masses that are displaced from the Sahara”. Does that mean the air mass has to cover parts of the Sahara at all times to ensure the origin?
Is the visual inspection necessary for your detection algorithm in general, or was this just a proof of concept in this case? Because if there is doubt about the origin of the air masses, and this step is always necessary, this is a big issue in the applicability of the method to other cases.

The inspection was a proof of concept to make sure that the parameters chosen make physical sense. Also, as a method based on thresholds with a certain degree of arbitrariness, some parameter tuning is needed to make sure that the method is actually capturing real events and avoid spurious warm air masses generated away from the Sahara.
The only issue that can’t be detected from the visual inspection is the difficulty to detect events that are not captured by the method. This is a consequence of the novelty of this work, which makes that an alternative method to detect missed events is not available yet. This will be discussed in the revised version of the paper.

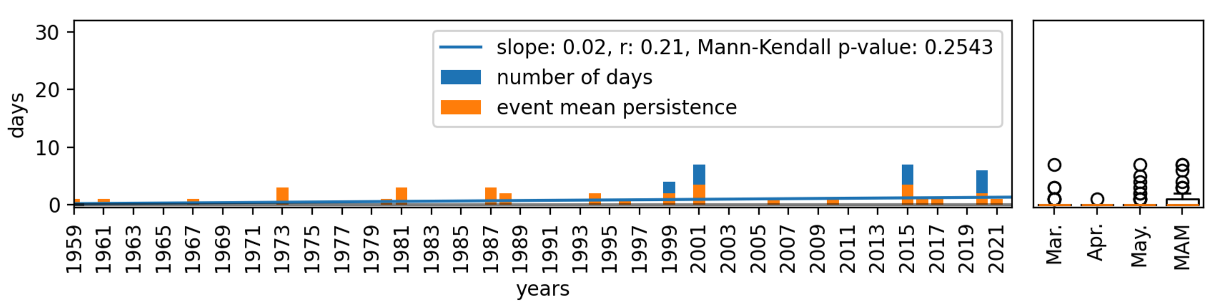
l 116-118

As far as I understand, the development of the catalogue and analysis of seasonal and inter-annual variations is a crucial part of your study. I was therefore disappointed to be referred to the Supplement for such a central information. Why don’t you show an average number of events per month and their spread across years (e.g. as a boxplot). Then your description of the seasonal variability can refer to this plot.

DJF



MAM



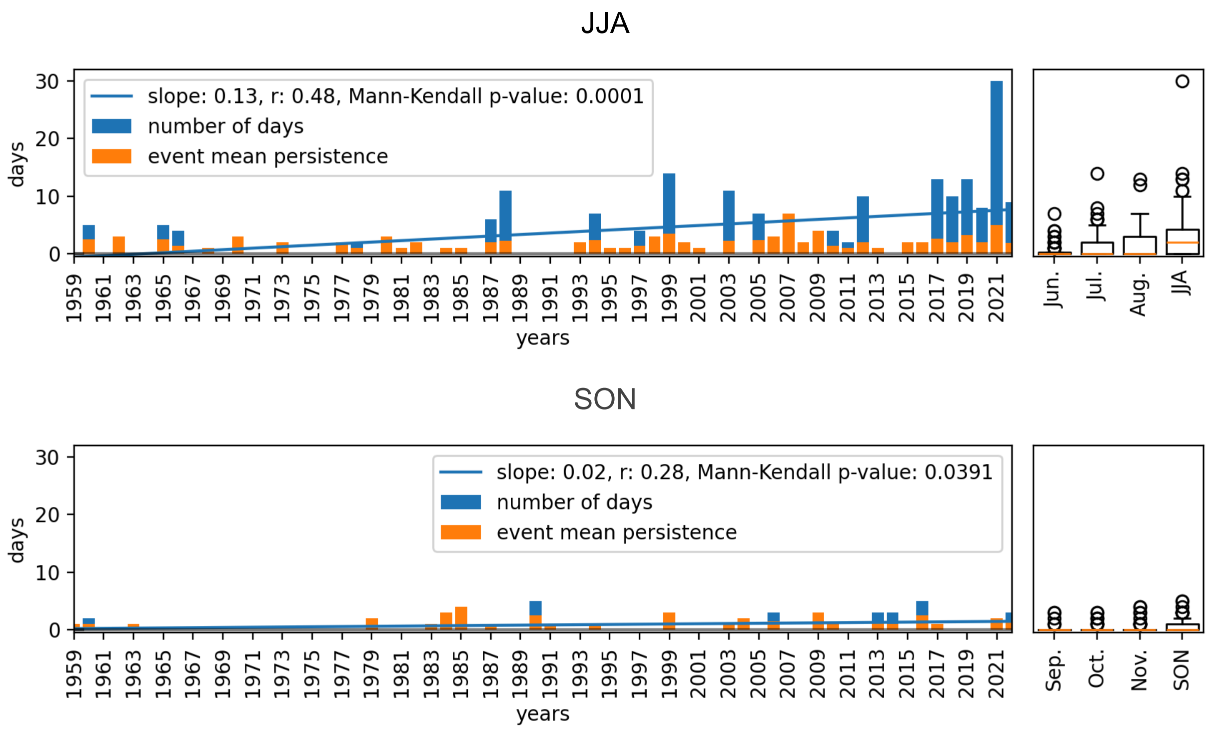


Figure R7: Number of identified Saharan warm air intrusions in a) DJF, b) MAM, c) JJA and d) SON per year between 1959-2022 (blue bars). The mean duration of each intrusion event, from the first to the last continuous day, is shown with orange bars. Note that when a single event is recorded for a specific year the blue and orange bars are the same size, therefore, only the orange is shown. A least squares linear regression fit to the seasonal number of intrusion days is shown in blue.The panels on the right show the distribution of intrusion days for each month within the season and the whole season.

We want to thank the reviewer for pointing this out. We agree that leaving the catalogue in the supplementary material goes against the interests of this publication and we will produce the suggested plot displaying the seasonality of events in the revised version of the paper. We will move Figure 6 earlier in the text and add visual data that informs about the average intrusion days per month and its interannual variability. An example is presented in Figure R7

Fig 3
From the cross-sections it becomes obvious that the air mass tilts up towards the edges, especially in the North and West. This will not be captured by your methodology, which focuses on the lower levels only. I think, however, this warm, dry air mass might still have considerable effects in those regions. This goes in the same direction as the question about how sensitive your results are to the choice of the indicator variables and their respective levels. Could you discuss this a little bit further?

The effects of a warm air mass in the higher parts of the troposphere are not direct but rather a consequence of the air mass intrusion on the atmospheric dynamics. This indirect effect escapes the scope of the current work. Further studies with the catalogue will focus on the extratropical dynamical effects and the impacts on cloud cover and surface downward shortwave radiation.

l 153 and following and Fig 4
I was wondering how similar the clusters between the seasons are. The different colour scales make it difficult to compare by eye, but there seem to be great similarities (referring to the figures in the appendix). I would be interested whether one kind of IT appears in all season, or another exists only in summer, or only in MAM and SON, or something like that. This then raises the question, if similar/the same clusters appear across seasons, is it necessary to split the seasons for the clustering?

The clusters look similar, but we prefer to separate the clusters per season because we think that it is interesting to illustrate the large-scale characteristics across seasons. The large-scale structure shows differences, therefore we decided to keep the results of the clustering separated. We will make this motivation more explicit in the text

l 168 and following The paragraph starting with “Kendall-tau correlations...” seems to be quite interesting but lacks depth. First, I think there should be a new paragraph, i.e. linebreak. Second, the topic is not introduced at all. It would be helpful to mention the teleconnection indices and then explain that you calculated correlations. Also, you mention you use “some” of the indices but do not motivate which ones. Third, the acronyms used in the equations are not introduced, and the dashes as bullet points can be mistaken for minus signs. Most importantly: I am lacking a discussion of what your resulting numbers mean. I.e. does the first equation imply that IT2 in summer occur more often in a positive AMO phase? Is this reasonable/expected or not? Please elaborate on these results more.

We agree that the results and discussion around the teleconnection indices is shallow. Therefore, we will expand on this and justify the different indices used with the literature of Mediterranean drivers. Also, we will discuss the implications of obtaining significant correlations with different teleconnection indices.

Sect 5
The intrusions are identified based on temperature thresholds, the conditional probability P (heat|intrusion) can therefore be expected to be extremely high, at least in the area that is covered by the affected air mass.
Your approach adds to this as it shows any area that is impacted. However, since the information shown in Fig 7 is a combination of many intrusion days, it cannot be distinguished between areas that are hot “by definition” just because the air mass is present, and areas that are hot on an intrusion day, but not covered by the intrusion itself. Could you please elaborate on this. One idea could be to show Fig 7 for cell-wise calculation of the impact: P (heat in this cell|intrusion in this cell) in order to see how big the difference is between this direct vs your “also -remote” impact.

Another thing about this section is the way you look at the contribution. I spent a long time thinking about this. My immediate question was why don't you show P (heat|no intrusion), as this would be the intuitive measure to compare P (heat|intrusion) to.
I see how your version of the contribution answers the question "In how many percent of heat days was an intrusion present?" and the intuitive conclusion would be "in this many percent of the heat days, an intrusion was present and therefore played a role, i.e. contributed". However, since the heat cannot be the cause of the intrusion, it is difficult to interpret this in terms of causality.
I would definitely ask you to also show P (heat|no intrusion). This shows you how often it was hot, despite no intrusion being present. Contrasting P (heat|intrusion) and P (heat|no intrusion) by looking at a difference, relative risk, odds ratio or log-odds ratio will give insight into how the presence of an intrusion increases or decreases the chance of heat. I.e. when the two probabilities are similar, intrusions do not influence heat very much, while when they are different, you can see the effect of the presence of an intrusion.

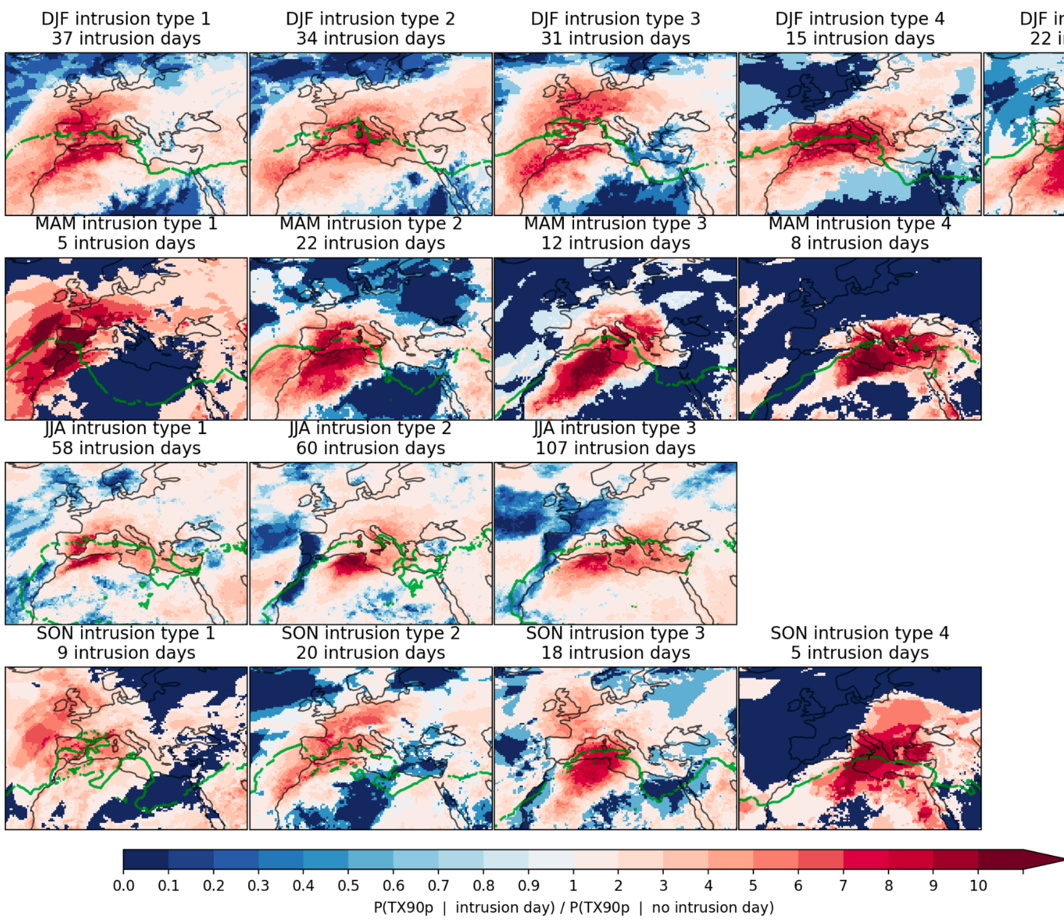


Figure R8: Odds ratio of the Impact of the Saharan warm air intrusions on extreme temperatures in the EM region for the different seasons (rows) and ITs (columns). The odds ratio is measured as the fraction of the percentage of intrusion days that coincide with an extreme temperature day (TX90p) divided by the percentage of extreme temperature days that do not coincide with an intrusion day. The area of impact is displayed in green contours, and represents the amount *of Saharan air masses recorded in the historical period for a specific IT. The values of the contours go from one Saharan air mass event to the maximum number of days recorded in a gridpoint. Note that the outer contour always represents one recorded day in presence of a Saharan air mass.*

We thank very much this thorough comment on the role of the intrusions to extreme temperatures. We will add contour lines showing which is the accumulated extent of intrusion days for each IT (effectively showing the area of impact, which will be redefined to the boundary where intrusions have been recorded).
Regarding the probabilities to show, it was a tricky part to convey and could definitely be improved by using the P(heat|no intrusion). We computed the odds ratio and displayed it in Figure R8. The regions where the probability of having an extreme is higher when intrusions happen are shaded in red. These results will be added to the revised manuscript and discussed.

l 259 Here you conclude, that Saharan intrusions in the cold months could be similar to subtropical intrusions. To a reader who is not totally familiar with the literature, it is not obvious why and how Saharan intrusions differ from subtropical intrusions. Please make this more clear.

Thanks for pointing this out. We will expand on this so that the difference between a Saharan intrusion and a subtropical intrusion is better understood. We believe that introducing this concept when we talk about the Sahara climatologies in the cold months might be helpful.

l 273 Here you argue that the increasing number of intrusions can be explained by thermodynamic changes as the warming of the troposphere. I cannot follow that argument. An increase in intrusions according to your detection algorithm could either be due to an increase in the synoptic situations causing the advection, or due to a general warming of the atmosphere in the WMed region. The former is not necessarily connected to thermodynamic changes, but rather to dynamic changes in a warming climate. The latter is closely tied to thermodynamics and a warming troposphere, but it might be a false trend due to the detection method. If your WMed is generally warmer and the latitudinal gradient weaker, your indicator variable thresholds might be exceeded without having advection from the Sahara present. Please elaborate on this in more detail.

We agree with the reviewer, the increase could also be due to a change in the dynamics induced by climate change. We will include the discussion of how climate change might also affect changes in the dynamics and in turn have an effect on Saharan warm air intrusions.
Regarding the increasingly WMed region and the indicator thresholds being more easily exceeded, we agree that there might be intrusions that correspond to an advection of air masses from the north of the Sahara region. This could be the case in the most recent years due to warming. Nonetheless, the characteristics of the air mass will be climatologically Saharan. There is literature indicating projected desertification of Northern Africa, Southern Iberian Peninsula and other regions of the Southern Mediterranean (lite lite lite). If this is the case, we shouldn't be surprised about Saharan-like air masses being formed north of the

Sahara in a warming climate. The framework we present here, aims to define a Saharan air mass in the recent historical period to be used as a baseline (benchmark).

We thank the reviewer for the thorough minor comments that follow, they will greatly improve the text. We will apply the suggestions to the text. See specific comments that require answers below:

Minor Comments

Throughout the manuscript:

Please be consistent in your spelling and the usage of the terms for the air mass (Saharan warm air intrusion, warm-air intrusion, warm intrusion, intrusion, air mass, etc.). It will help the readability if you use one clear term throughout the manuscript. Examples, e.g. 50,53,57 or the title of Section 5.

Similarly: Use cross-section (or a precise term of your choosing), as “cut” can be misunderstood (you use cut, vertical section,...).

l 14 You are referring to changes over what time period? Climatological short(er) term changes? We refer to both historical and projected. We will specify in the text.

l 38 this should be “forest fires in Portugal”

l 45 gain insight into

l 46 applicable

l 50 I don’t understand this sentence. These are the aims of the study, no? What do you mean by “objectives to characterise intrusions into the WMed”? Agreed, sentence is changed to “This study has four objectives with regard to Saharan warm air intrusions (referred to in the text as SWAIs) into the Western Mediterranean …:”

l 65-66 This was already mentioned very similarly in l. 43 and following. The last sentence is removed.

l51/69 Please define the “historical period” in terms of years somewhere.

l 69 Do you mean daily resolution?

l 73 Daily maximum temperature at what level?

l 74 Please define the WMed and Euro-Mediterranean region in terms of lon/lat

l 77 For better readability I suggest:

“We adapt the definition of Sousa et al. (2019), which takes into account...”

l 81-82 Please reference equations 1 and 2

l 90 the gas constant of dry air

l 106 “the detected events”

l 120 Why are you showing August 2006 here and May 2015 in Fig 1? Are they specifically representative? But why chose 2 different ones? There is no specific reason for choosing one or the other. Both correspond to intrusion days. We think it is good to show different events.

l 121 Sardinia is not at 17◦E? Do you mean the air mass penetrates east of Sardinia? also: “intense trough in the North Atlantic”. That is right, wrong synaptic connection from my end. I meant Sicily, it will be corrected.

l 132 “different expressions” is not a precise term, maybe rather use “properties”?

l 137 I am not sure whether I understand this clustering correctly. You use ∆GH and θ in the WMed region on days that were identified as intrusion days? And then you want to cluster those days into different types. Your results seem as if the position of the anomalously high indicator values defines the clusters. But would the clustering also pick up on two clusters that affect the same region but with different magnitude? From the obtained results we gave more weight to the interpretation of the positioning of the intrusion, but it is true that the clustering seems to also capture differences in the magnitude (IT2 of Figure 4 clearly has larger magnitude as well as other ITs in the rest of seasons). Thanks for the comment, we will include these findings in the results and discussion.

l 151 Looking at the plot for JJA, I cannot really see how you conclude on 3 clusters. The elbow seems to be higher and the high value in the Silhouette score looks a bit like an irregularity?

l 151 The last sentence in this paragraph "Finally,..." is really inherent to the clustering method and not necessary here.

l 155 You are referring to Fig S6, not 5 here.

l 162 I think this should be DJF and SON, you are referring to panels a,c. It is DJF, JJA and SON which are panels a,c,d. It will be fixed

Fig 6 How do you handle the duration of events, where a long period of intrusion is interrupted by a very short period, where the 5% criterion is just not met for a short time? If you handle these as separate events, this would reduce the mean duration very much compared to when you use a filling or filter. We are considering them as separate events.

l 167 from Fig S8, to me it looks more like IT1 and IT2 often transition to IT3, not IT1 to IT2 and IT3? Or are you interpreting 3 in more than 20 as often? That is right, IT1 and IT2 often transition to IT3. It will be addressed in the text-

l 178 This is not the first time you use the term Euro-Mediterranean. Please introduce the acronym and the exact definition when you mention it first (in the data section?)

l 180 the 90th percentile of what variable (there is simply a word missing I think).

l 197 What do you mean with "largest contributions come from JJA"? The largest contribution of extreme temperatures that come from intrusion days (P(intrusion day|TX90)).

l 212 subtracting them from the intrusion days' composite

l 217 "for" instead of "from" in the parentheses

l 223 What do you mean with "the [...] circulation and SLP anomalies are coherent"? That the wind directions and SLP fields in the composites make sense, which shouldn't be surprising. It will be made clearer in the text.

l 224 What is the tilde over the 3? It is a typographic error, thanks for noticing.

l 225 always

l 228 negative sea level pressure anomaly

l 243 Your description earlier implied you are detecting intrusions that originate in the entire Sahara, why are you mentioning only the western Sahara here now? This is a typing error it should say only Sahara

l 251 You do not really show that the threshold you choose is robust, or how sensitive result are to it's choice. We will avoid saying robust.

l 256 What is a "climatological air mass"? I understand what you are referring to, but this can be misleading. Also "the latitudinal band" is not precise. This will be re-worked in line with the second major comment

l 262 Please leave out the "backward", as the studies you cite use forward and backward trajectories.

What do you mean by "historical contribution"?

l 292 remove parenthesis, “which” instead of “and” afterwards

We want to thank the suggestions done in the Figures. We will apply them in the revised version as we agree with all the suggestions

Comments about Figures

While many of your figures are nice in general, all have very small font sizes on the axis labels, the axis ticks, etc. Please change this for better readability. Often the use of shared axes or removal of figure titles that are redundant with the caption can create space.

MAJOR: Fig S1 Caption is wrong, the red box is not the Saharan region. Also here the figure title is redundant. Axis ticks are not necessary in all rows and columns but only at the bottom and left, which makes space for bigger labels ad better readability. Please mention what contours and shading represent in the caption.

Fig 1: The green and cyan lines are not so easily distinguishable, please use additional different linestyle. The black dashed box is supposed to be an example for how big 5% are, this needs to be stated very clearly, otherwise it is not obvious that this is just exemplary.

Fig 2 I do not see the additional value of this figure. When it is mentioned in line 96 it is rather confus- ing. As far as I understand the helpful information here is that the averages over the Saharan region are in the high tails of the distribution of the WMed region, I think this can be mentioned without showing the figure. One other question this figure raised for me is the spread in the blue dots in February.

We will consider moving this figure to the supplement and using information from Figures R1 and R2 to better convey our point.

Fig 3 Add dates to the caption when the exact time of the event is, and what you refer to as before, dur- ing and after.

Fig 4 Caption could be clearer. These are composites of the intrusion days of the respective type in the JJA season? The labels in the contour lines are difficult to read and interpret.

Fig 5 As IT 1-3 are categorical values, a non-sequential colourmap would make the perception easier. Espe- cially in connection with the figures in the Supplement, where more ITs exist. Then the question is also, relating to an earlier comment of mine, whether all IT1 across seasons are similar to each other or not - using the same colour for all IT1 implies that they are more similar to each other than to others.

Fig 6 Do I understand it correctly, that the orange bar can never be larger than the blue? Consequently, when I only see orange, this means it is a single event? Exactly, when only orange appears means that only one event exists for that season and year. I’ll mention it in the caption.

Fig 7/8/9 Can you add the area of influence as a contour. You discuss the relation between the impact and the area of influence (e.g. l 192), but it is difficult to see that by comparing diffrent plots on different pages. This is a very good idea, we will add the contours of the area of impact. (See Figure R8)

Fig 9 be constistent in use of SLP and PSL, I think it should be wind speed anomalies in the caption here. And what arrows are you referring to in panels (a) that are not the quivers? This caption is not easy to follow. It might also help to add numbers to the climatological contours for reference. And the quivers are way too small to see their direction properly.

Is Fig 9 panel (a) the same as Fig (4) panel (b) (except for the significance masking)??? Fig 9 (a) and Fig 4 (b) are not the same. Any similarities are due to compositing the same days, although the variables are different