We would like to thank the reviewer for the very thorough review. We hope all small technical errors are now solved and we believe that the questions and suggestions provided in this round of revisions make the study more robust.

I thank the authors for answering to my comments in depth and improving the manuscript. The manuscript has
improved a lot with the current version, however, I still have some minor concerns that need to be addressed
before publication, most of which are technical, but some more substantial.

Minor comments

line 72 I think if you explain all your sections like this, you should indeed mention all of them. You do not mention Section 7 here.

Thanks for the suggestion, section 7 has been added in the manuscript structure.

line 81 What do you mean by roughly? You must somehow have a specific region that you use, i.e. to compute the TX90 values? But thank you for clarifying.

TX90 values are calculated at specific grid points (see "...we compute extreme temperature days for the historical period at each grid cell…" in the manuscript), so there is no issue in taking a larger or smaller domain and results will remain the same. I agree, though, that the "roughly" is a bit ambiguous, so I'll remove it as we are using that exact domain.

line 89 double parentheses could be avoided

Thanks for noticing it has been addressed.

line 110, Fig S3 In the caption you refer to a red box, which in fact is black

Thanks for noticing this error, the caption has been fixed.

line 124, Fig S4 y-axis and caption differ. Is it θ700−925 or θ750−925?

It is 700-925 as in the caption, the figure has been fixed. Thanks.

line 133,135 be consistent with red vs orange shading

Thanks for noticing the discrepancy, I changed it to orange.

line 132, Fig 1 be consistent with purple/fuchsia

It has been changed to fuchsia.

line 135 ”...a representation of how much area 5% of the WMed IS, ...”

Thanks for pointing this out, it has been fixed.

line 139-141 Thank you for adding the sensitivity study. I thought this result was really interesting. My immediate reaction to this sentence was ’why?’. Maybe you could add a thought about what mechanism leads to these changes in the length of the detected events depending on the lower level bounds.

We agree that this is an interesting point, thanks for bringing it up. We think we can hypothesise about that only for the geopotential height, where changes are slightly more noticeable. Therefore, we suggest that obtaining longer and more intrusion events when using an upper lower-bound of the potential temperature metric this is due to the fact that separating the indicator’s lower bound from the ground diminishes diabatic cooling with the surface and therefore the Saharan air masses will change slower when taking potential temperature at 850 hPa. We modify the text about sensitivities to:

“In general, using a higher lower-level potential temperature bound leads to more and lengthier events, and a higher lower-level geopotential bound leads to a very slight decrease (increase in JJA) in the amount and length of the events. The increase in events due to taking 850 hPa potential temperature might be related to the diminished diabatic cooling of the air mass as the lower bound of the indicator is further away from the surface. Changes due to GH lower bound are less obvious and we can’t provide a robust hypothesis for their small sensitivity.”

line 158 in panels a-d you show seasonal results, not monthly as you state in this line, or did I misunderstand?

I agree, it has been changed to seasonal. Thanks for pointing this out.

line 159-160 And Fig2: ”Results show that the months with most of the intrusions are July and August, as well as December. There is also some notable activity during January-February and May-June.” Thank you for adding the panels e-h) in Fig2. However, I still have some reservations about this.
First: panel indicators are missing in the figure.
Indicators have been added, thanks for pointing it out.
Second: panel a has a different y-scale, which makes it very difficult to compare the panels and follow your reasoning.
The y-scale was actually the same but the 10 and 30 ticks were missing. It has been fixed.
Third: The boxplots are hard to read as they are small, the circles big, and the y-axis too tall, so that it
is squished in the bottom. Also, the median is marked in orange, which can be confusing as the data you
show in the box plots refers to the blue bars not the orange ones in panels a-d.

The boxplots have been changed to violin plots spanning the whole distribution. We think it is now easier to compare the different months.

Apart from these technical/visual things, I cannot follow the sentence I pasted above. To me it does not
become obvious that e.g. December is has many intrusions, only that the spread seems higher? Also you
say Jan and Feb have some activity, but from the plot they don't look too different from March or the
Autumn months. This definitely needs to be addressed and made more clear before publication.

We understand the concern. After producing the violin plots it is now easier to see that the distributions of January, February, March, May and June look quite similar and some intrusion activity is present, although saying "notable activity" might be a bit much. Therefore we changed the sentence to:
"There is also some activity during January-March and May-June."

lines 213-216 These 2 sentences do not belong in the discussion about the ITs, but rather right after line 160.

I agree, thanks. They have been moved.

line 229 Are you using the same acronym for WMed region and WMed Oscillation? This is confusing.

Wrong acronym, thanks for noticing. The acronym should be WeMO and it has been changed in the text.

line 233 double parentheses. Also 'is' and 'fluctuations' - plural or singular?

The double parenthesis has been fixed and changes "fluctuations" to "fluctuation".

line 246 I still think that the itemised list of correlations should not be itemised using dashes, as they can be read as minus signs

True, the dashes have been removed. Thanks.

line 255 double parenthesis should be avoided

It has been addressed in the text.

line 257 There is no table S2, it should be S1, and then in the supplementary you should call it that as well, it is
'Table 1' at the moment

Thanks, it has all been homogenised to Table S1.

”We aim to quantify the probability of getting...”

Thanks for the suggestion, it has been modified in the text.

lines 270 ff ’impact on’

It has been addressed, thank you.

lines 265-267 This finding is really interesting. However, you do not really explain why you think this seasonal difference exists.
I was wondering here: S12 looks similar for all seasons. But you motivated it in order to explain what is happening in JJA. I am missing your reasoning then, if the behaviour in JJA is explained by a slower dynamics, or what else it could be

True, what we can say from S12 is that the reason why the impact beyond the area of influence is very low in JJA is not because the circulation is slower. The impact in the day after an event does not increase. We add this to the text:
“Therefore, Figure S12 disproves that JJA intrusions might move slower and have an impact outside the area of influence the days after an intrusion event. This result suggests that if JJA has no impact outside the area of influence it is because Saharan warm air intrusions, and their impacts on extreme temperatures, are simply more confined to the area of influence.”.

line 280 Thank you for following my suggestion on how to look at the conditional probabilities. However, the equation you show here is not the odds ratio, but the relative probability or relative risk. While it has a similar meaning, please be careful about the wording and the exact conclusions you draw. Decide for either of them, but then check the meaning and be consistent.
For reference, the odds ratio is defined as:
O1 = P 1 / 1 − P 1 (1)
O2 = P 2 / 1 − P 2 (2)
OR = O1 / O2 (3)
where in your case: P 1 = P (T X90p|IT nday), P 2 = P (T X90P |noIT nday)

We want to thank the reviewer for pointing this out, we agree that the metric we showed in the revised text is not the odds ratio but rather the risk ratio. See Figure R1 for the results of the Odds ratio.

The two metrics do not have exactly the same meaning, but the conclusions we arrive at are the same. For simplicity we will keep the risk ratio metric, so we adapt the text to:

“To quantify how the presence of an intrusion increases or decreases the risk of having TX90p days we compute the risk ratio (Equation 4) between the probability of having TX90p

conditioned on having an ITn intrusion day versus that of having TX90p conditioned on being on any other day than the ITn intrusion day.
(…)
The risk ratio is the ratio of how much more risk there is to have an extreme temperature event if the day is an ITn intrusion day than if it is not. Risk ratio will be closer to one when the two probabilities are similar, and therefore the effect of the intrusion is not affecting the probability of having a TX90p day. When the risk ratio has values above 1 it means that having an extreme temperature is more probable when an intrusion is present. Risk ratio under one means that having an extreme temperature day is more probable on non-intrusion days. Values above 10 suggest a strong difference in the occurrence of TX90p between having or not an ITn intrusion day (Ellison 1996).
Figure 7 displays the risk ratio and we see that …"

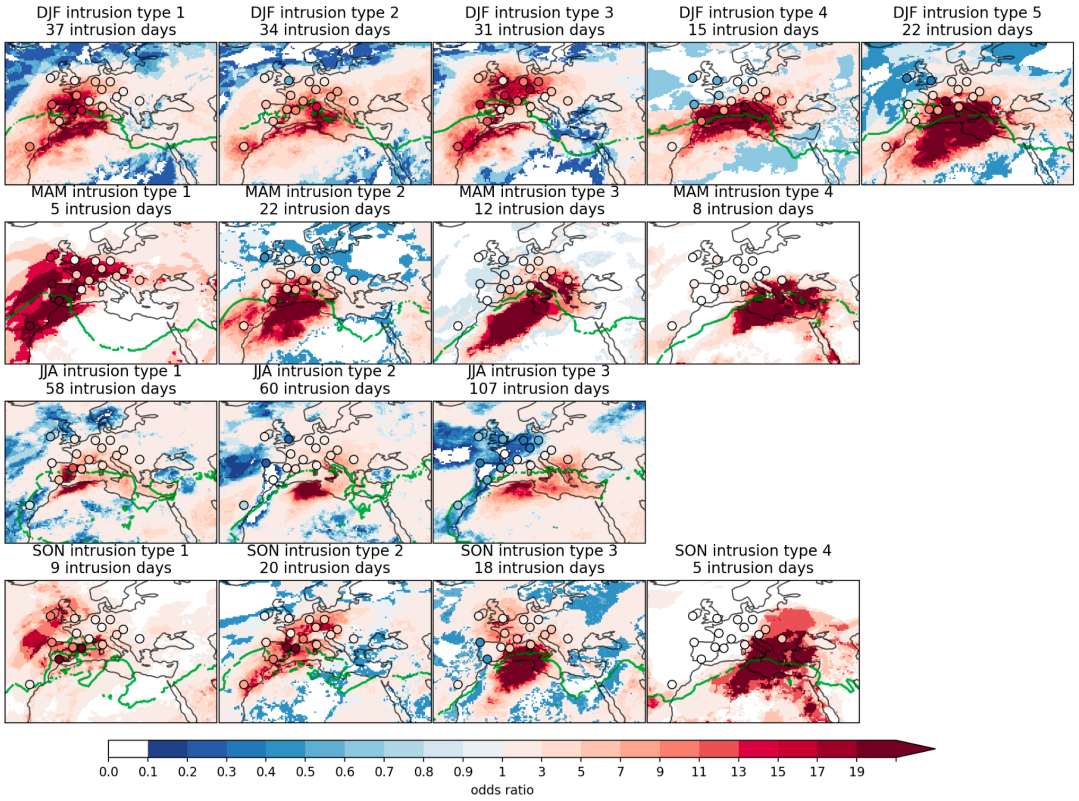

Figure R1: Odds ratio of days with and without Saharan warm air intrusions to the extreme temperature days (TX90p) of each season (rows) and IT (columns). It is computed as the fraction of the impact divided by the probability of having an extreme temperature day when no intrusion is recorded. Colored dots are results from ECA&D station data (Table S1). The area of influence is displayed with a green (dotted) contour, and represents the limit where Saharan air masses have been recorded in the historical period for a specific IT.

line 281 '...when the day is not in ITn' A day cannot be IN an intrusion type? please rephrase.

The text has been changed to: "when the day is not an ITn day".

line 283 ”isn’t” should be is not

The text has been addressed, thanks.

line 289 to have an influence ON

Thanks for pointing this out, it has been fixed.

line 292 the word coincide indicates a joint probability, not a conditional one

The text has been addressed: “the percentage of TX90p days that are also an ITn intrusion day”.

Fig 7 Thank you very much for showing this, it improves the manuscript greatly. I had a follow-up question now. Did you look at the same thing for all ITs combined? I.e. P (T X90p|intrusionday)/P (T X90p|nointrusionday). Since the intrusion types are still somewhat similar, comparing to ”no ITn intrusion day” will still incorporate all other ITs (which have an effect on similar regions). I assume you could see even more clearly that intrusions in general have an effect on temperature if you showed it for intrusion days in general.

Thanks for the suggestion. We performed a study of the impact on extreme temperatures of all the intrusion days in each season at the beginning of this work. We saw that the impacts were smoothed over the EM region due to the spatial heterogeneity of the Sharan warm air intrusions. That is why we use ITs to help us discriminate between the impacts in the different parts of the WMed.
It is true that in the risk ratio there are intrusion days being included in the no ITn days. Nonetheless, we think the effect of those days is minimal and actually is useful to include them in the no intrusion days: For example, the impact of IT1s in the east of WMed is generally very small (and the risk ratio is sometimes below 1). Therefore, counting that IT1 intrusion day decreases the influence of intrusions to temperature extremes in the east. This is why we want to keep the ITs separated.
Below we show Figure R2 where we compute the relative risk or risk ratio for all season’s intrusion days and it can be seen how, even if the risk ratio is positive in most of the EM region (and especially strong in DJF and MAM), the temperature effects decrease in magnitude for the reasons explained above (see how the impacts in eastern Mediterranean are very reduced).

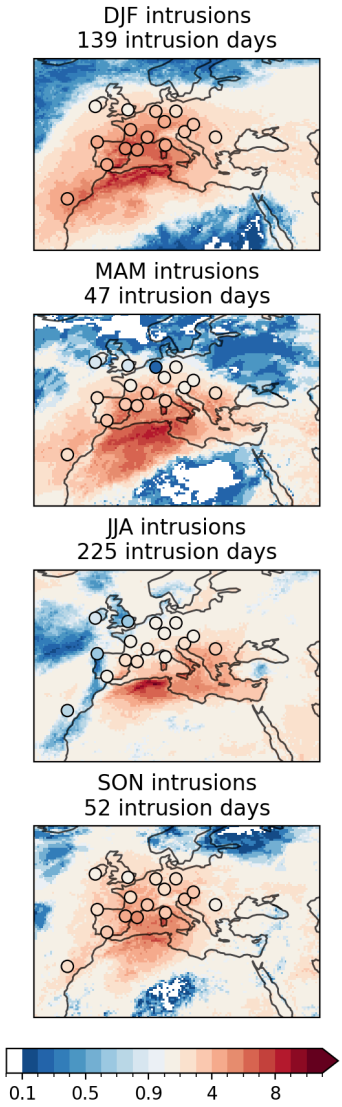

Figure R2: Risk ratio of having TX90p days with and without Saharan warm air intrusions for each season (rows). It is computed as the fraction of the impact divided by the probability of having an extreme temperature day when no intrusion is recorded

line 306 you are referring to Fig4 not 5 here?

True, thanks for pointing this out.

line 328 change the passive to an active citation

I have applied this change in the text. Thanks for noticing.

Fig 8/9 Your description is a little bit hard to follow, since the colour scales are different, which makes it difficult to estimate if a signal is stronger in Fig 8 or 9.

I agree, thanks for the suggestion. We changed the color bars and quivers to the same scale for all seasons. The differences between seasons should now be clearer.

line 351 reference Fig 7 here

It has been added in the text.

line 363 S4 is the wrong figure to reference here (since this is not the only instance of a wrong figure being referenced, please check carefully if all references are correct) general be consistent in your spelling of upper-troposphere/tropospheric, i.e. with or without hyphen

It has been addressed. Thanks for pointing out the issues with the figures, we have gone over all references to make sure that they are correct.