The authors revised the manuscript according to several of my suggestions. The content of the manuscript is generally interesting, and I think it is worth to be published. Yet, I still have some concerns about ambiguities of several statements due to unclear language. Additionally, I would like to know why the authors chose to replace the original trend estimation by the robust linear model, as the replacement seems to change parts of the result in a non-negligible way.

The authors thank the reviewer for the time spent on a thorough review and have provided answers to the comments below.

General & minor comments:

• The results displayed in Figure 2 changed considerably compared to the first manuscript version. I guess this is due to the usage of the robust linear model instead of Mann-Kendall's test. Yet, I did not find any explanation in the author's response why this new method has been chosen (apologies in case I missed it) and why it yields quite different trend estimates. This makes me wonder a bit about the robustness of the estimated trends. Additionally, as mentioned also further below, I think it would be necessary to shortly explain in the manuscript what the robust linear model is and why it has been chosen.

Thank you for this comment.

The robust linear model (RLM) test was chosen in place of the Mann-Kendall (MK) test because it is a reliable methodology (also with regards to outliers) and it allows us to be quantitative with regards to the trends (i.e. know by how much the decrease, increase is, e.g. Huber, 2011). Moreover, we found the following other papers which also adopted the same method for heatwaves trend analysis (e.g. Kishore et al., 2022)

The trends that we found do not differ significantly because of the change in the trend test methodology (i.e. RLM or MK). The differences are due to the use of the false detection rate (FDR). We implemented FDR in the revised version of the paper following the suggestion of the reviewers, in contrast with the single fixed threshold as done in the original version of the paper.

To illustrate this, a comparison of the MK test and the RLM with respect to the same significance level assessment based on the FDR has been performed. For each pixel and for the different return levels threshold, we computed and reported in Table 1 the percentage of the study area in which the two methods provide: i) positive statistically-significant agreement; ii) negative statistically-significant agreement; iii) positive statistically non-significant agreement; iv) negative statistically non-significant agreement; v) disagreement in statistically-significant cells; vi) disagreement in statistically non-significant cells.

The highest percentage of the study area in disagreement for any of the return levels is less than 5%, for CWMId>CW5Y. The highest percentage of the study area in disagreement for statistically significant regions is less than 4%, for HWMId>HW5Y. The NA represents the cases in which both methods had no value for the corresponding row of the table. Results show that the two methods (MK and RLM) provide consistent outcomes when both are assessed using the FDR methodology.

Table 1: the percentage of the total study area in which the two methods (MK and RLM) agree or disagree.

	HWMId > 0	HWMId > HW5Y	HWMId > HW10Y	CWMId > 0	CWMId > CW5Y	CWMId > CW10Y
i) Agreement significant positive	98.86	86.87	NA	NA	NA	NA
ii) Agreement significant negative	NA	NA	NA	NA	NA	NA
iii) Agreement non-significant positive	0.76	9.42	91.45	59.34	75.73	47.46
iv) Agreement non-significant negative	NA	NA	7.7	39.03	19.89	52.51
v) Disagreement significant perc	0.38	3.71	NA	NA	NA	NA
vi) Disagreement non-significant perc	NA	NA	0.85	1.63	4.38	0.03

1

An explanation of the robust linear model is provided here and in the relevant minor comment below. This new sentence has been added to the paper.

Old sentence:

The trends are analyzed using the robust regression technique (Huber, 2011). This method is often used throughout the literature for assessing trends in natural hazards (Formetta and Feyen, 2019 for multiple hazards and Kishore et al., 2022 specifically for HWs).

New sentence:

The trends are analyzed using the robust regression technique (Huber, 2011) which is often used to assess trends in natural hazards (Formetta and Feyen, 2019 for multiple hazards and Kishore et al., 2022 specifically for HWs). Robust regression seeks to overcome part of the limitations of traditional regression analysis.

For example, the linear regression least squares method is optimal when the regression's assumptions (normal distribution, independence, equal variance) are valid (Filzmoser and Nordhausen, 2021; Khan et al., 2021). This method can be sensitive to outliers or if normality is dissatisfied (Khan et al., 2021; Brossart et al., 2011). The robust regression method is designed to limit the effect that invalid assumptions have on the regression estimates (see Filzmoser and Nordhausen, 2021 and Alma, 2011 for more details).

References:

Hipel, K.W. and McLeod, A.I., (2005). Time Series Modelling of Water Resources and Environmental Systems. Electronic reprint of our book orginally published in 1994

Huber, P. J.: Robust Statistics, in: International Encyclopedia of Statistical Science, edited by: Lovric, M., Springer, Berlin, Heidelberg, 1248–1251, https://doi.org/10.1007/978-3-642-04898-2_594, 2011

• The format of references to figures is broken in several instances – please check! I encourage the authors to check the manuscript for formatting issues before (re-submission).

Thank you for this comment, this is a Mircrosoft Word issue when going from track changes to untracked changes, the authors apologies for this occurring and have resolved this in the submission for this revision.

• The language should be checked again. Particularly, the introduction was partly hard to read and understand. Sometimes, the language obscures the meaning of statements (e.g., what does “would change their risks to society” in line 34 mean? I think it would need to be “change the risks they pose to society”) I anticipate that some of this will be tackled by the journal proof reading, but nevertheless it should be guaranteed that ambiguities stemming from unclear language are minimized.

Thank you for this comment.

For the sentence line 34, we agree with the comment. The authors changed this specific part of the sentence, to simplify it from:

“ which would change their risks to society”

To:

“changing the risks they pose to society”

The authors have also been through the text again and have attempted to minimize the number of ambiguities. This can be viewed from the tracked changed version of the manuscript.

• Results: The results section begins with the description of several Supplementary figures, which I do not find so ideal. I think it is best practice to start the results by describing figures that are shown in the main manuscript. I am not familiar with the NHESS manuscript requirements, but I think it would be good if one or two figures could be moved from the supplementary to the main manuscript, if possible.

Thank you for this comment. We agree for the first figure that is therefore moved from the supplementary material to the main text. The presentation and the comment of this figure was already present in the main text. However, the second figure (the new S2) that we mentioned in the result section has been kept in the in supplementary material. This figure only shows that that the KS test indicates no rejection for the entire region and is therefore not interesting to show in the main text.

• Lines 447-463: I think that Figure 6 is really an interesting figure, but it could be highlighted and explained better. The current manuscript just goes through the single figure panels one by one, without making connections between them. I think that restructuring the text by focusing on the main features of the figure and on what the message of this figure should be, would make the description much more interesting.

Thank you for this comment, the authors have restructured the text by focusing on the main features of this figure and its message. The text was changed from to

From:

Figure 6-a (Figure 6-b) show increasing trends in risk (due to change in vulnerability only) in the main cities and nearby areas. Decreasing trends are found for most of the remaining region.

Figure 6-c (Figure 6-d) show increasing trends in risk (due to change in exposure only) in/near urban areas and decreasing trends in zones at high elevations and far from the urban centers.

Figure 6-d show the hazard is the main driver of risk for HWs, with statistically significant increasing trends, more evident in and around highly populated areas. Finally, Figure 6-e show no significant trends in CWs risk (due to change in hazards only).

To:

Figure 7a and Figure 7b show trends in risk due to changes in vulnerability only, effectively indicating the locations of the increases/decreases in risk due the changes of vulnerability indicators, that are equally weighted (seen in Figure 5). These trends are found to be increasing in the main cities and nearby areas and are found to be decreasing for the rest of the region.

Figure 7c and Figure 7d show trends in risk due to change in exposure only, indicating the locations of changing risk due to the changes in population (exposed) only. The HW and CW risks are found to be increasing in/near urban areas and decreasing in zones at high elevations and far from the urban centers.

Figure 7e shows the trends in HWs risk due to hazard only, with statistically significant increasing trends being more evident in and around highly populated areas. The figure shows that hazard is the main driver of risk for HWs, with the significant increasing hazard trends cancelling (as can be seen in Figure 6a) most of the significant decreasing trends of the other two elements (exposure and vulnerability) seen in the Figure 7a and 7c.

Finally, Figure 7f shows no significant trends in CWs risk due to change in hazards only. The figure indicates that the combination of three elements of the risk equation (Equation 9) is the main driver of its risk (Figure 6b) rather than the CWs hazard only.

Specific comments:

• Line 17-18: I am not sure that the sentence about the Tweedie distribution is helpful in the abstract if the purpose of the application of the Tweedie distribution is not clearly specified. I would suggest to either add that information or remove this sentence.

Thank you for this comment, the reason for using the Tweedie distribution has been specified. The authors find it is important to mention the Tweedie distribution in the abstract given it is one of the main novelties of this paper.

The paragraph has therefore been adjusted from:

To obtain HWs and CWs risk maps we combined: i) occurrence probability maps of the hazard, ii) normalized population density maps, and iii) normalized vulnerability maps based on eight socioeconomic indicators. The occurrence probability of the hazard is obtained using the Tweedie zero-inflated distribution.

To:

To obtain HWs and CWs risk maps we combined: i) occurrence probability maps of the hazard obtained using the zero-inflated Tweedie distribution (accounting directly for the absence of events for certain years) ii) normalized population density maps, and iii) normalized vulnerability maps based on eight socioeconomic indicators.

• Line 19: I think rather "contributions" than "effects"

Thank you, this has been adjusted. The sentence went from:

The methodology allowed us to disentangle the effects of each component of the risk to its total change.

To:

The methodology allowed us to disentangle the contributions of each component of the risk to its total change.

• Line 48: Which increase in mortality does this refer to? The previous sentences do not mention any increase in mortality due to CWs.

Thank you for this comment, you are right this sentence is not needed, it was meant to be interpreted with regards to the previous sentence. This has been adjusted from:

With regards to CWs in Europe, recent winters have claimed 790 deaths in 2006 and 549 deaths in 2012 (Kron et al., 2019). The increase in mortality and among elder people is also found in Italy for CWs. For example, de'Donato et al., (2013) reported a notable increase in mortality (47%) for the timeframe of the 2012 CW in the city of Bolzano.

to:

With regards to CWs in Europe, recent winters have claimed lives with 790 deaths in 2006 and 549 deaths in 2012 (Kron et al., 2019).In Italy, de'Donato et al., (2013) report an increase in mortality (47%) for the timeframe of the 2012 CW in the city of Bolzano compared to the 4 previous winters (2008-2011).

- Lines 56-59: Might be worth to also mention what hazard refers to.

Thank you for this comment. This new sentence has been added to the main text:

Hazard is defined as a process, phenomenon or human activity that may cause loss of life, injury or other health impacts, property damage, social and economic disruption or environmental degradation and hazards being characterized by location, intensity or magnitude, frequency, and probability.

- Line 68: What does "qualitative" mean in this context? Does it mean arbitrary thresholds?

Thank you for this comment. We mean that the limits were subjective. The authors of HWs and CWs studies often time use a threshold for hazard, ie (0<HWMId<3, 3<HWMId<6, 6<HWMId<9), these thresholds (3, 6, 9) are subjective.

This has been specified in the sentence, changing it from:

Most studies on HWs and CWs have used qualitative numerical thresholds on the indicator to define severity and exposure to the hazards.

to:

Studies on HWs and CWs typically have used subjective numerical thresholds, on the indicator to define severity and exposure to the hazards (e.g. 0<HWMId<3, 3<HWMId<6, 6<HWMId<9).

• Lines 77-90: The description of the Tweedie distribution is rather technical. That is fine, but I think it would help the readers (at least, it would help me) if one or two sentences could be added explaining in a few simple words, why the Tweedie distribution is the right choice here (e.g., what are its main advantages for the application here?). Also, it would be nice if the implications of the limitations of the Tweedie distribution for this study (lines 88-90) could be briefly explained.

Thank you for this comment. The authors have made clearer why the Tweedie distribution is the right choice, as well as specified the implications of the limitations.

The paragraph has changed from:

Instead, for the first time we use the zero-inflated distribution of Tweedie families (Jorgensen, 1987; Tweedie, 1984) to estimate HWs and CWs frequency of occurrence, which enabled us to directly account for the possible zero values. The Tweedie distribution has been used mostly for the purpose of insurance claims analysis, but has seldom been applied in the field of natural hazards, such as HWs mortality (Kim et al., 2017), droughts (Tijdeman et al., 2020), and rainfall analysis (Dunn, 2004; Hasan and Dunn, 2011). The main advantage of the Tweedie distribution is the possibility of considering many distributions for the continuous and semi-continuous domain such as: normal, Gamma, Poisson, Compound Gamma-Poisson, and Inverse Gaussian (Bonat and Kokonendji, 2017; Rahma and Kokonendji, 2021; Shono, 2008; Temple, 2018). Moreover, for some of these distributions (i.e. Poisson mixtures of gamma distributions) it explicitly enables the fitting of zero-inflated data. Tweedie distribution main limitation is the complex distribution's fitting methodology and the difficulties to compare it to other models via information criteria such as the Akaike's information criterion (Shono, 2008).

To:

Instead, for the first time we use a distribution allowing for the direct fitting of zero-values (years with no event): the zero-inflated distribution of Tweedie families (Jorgensen, 1987; Tweedie, 1984). This distribution is also used to estimate HWs and CWs frequency of occurrence. The Tweedie distribution has been used mostly for the purpose of insurance claims analysis, but has seldom been applied in the field of natural hazards, such as HWs mortality (Kim et al., 2017), droughts (Tijdeman et al., 2020), and rainfall analysis (Dunn, 2004; Hasan and Dunn, 2011). The main advantage of the Tweedie distribution is the possibility of considering many distributions for the continuous and semi-continuous domain such as: normal, Gamma, Poisson, Compound Gamma-Poisson, and Inverse Gaussian (Bonat and Kokonendji, 2017; Rahma and Kokonendji, 2021; Shono, 2008; Temple, 2018). Moreover, for some of these distributions (i.e. Poisson mixtures of gamma distributions) it explicitly enables the fitting of zero-inflated data. The distribution's main limitation is the complex distribution's fitting methodology and difficulties in obtaining its relevant information criteria such as the Akaike's information criterion (Shono, 2008) The implication of these limitations are that the fitting of the Tweedie distribution is computationally intensive and that it is difficult to compare its goodness of fit to other distribution via the information criteria.

• Line 98: What does "temperature vulnerability" mean? Can you specify this in the manuscript?

Thank you for this suggestion and yes indeed this is now specified with the text going from:

Temperature vulnerability has also been appraised at city scale for HWs mortality (Ellena et al., 2020) and at regional scale (López-Bueno et al., 2021) for CWs mortality. Karanja & Kiage (2021) and Cheng et al. (2021) provide an overview of the different types of indicators used in the literature to quantify vulnerability

to:

Vulnerability indicators, in combination with the temperature-mortality relationship, have also been appraised at city scale for HWs (Ellena et al., 2020) and at regional scale (López-Bueno et al., 2021) for CWs. Karanja & Kiage (2021).

• Lines 105-106: It might be worth to name some examples of vulnerability factors used by Frigerio & De Amicis that are also important in the context of your study

Thank you for this comment, indeed and these have been added. The sentence being changed from:

Studies on social vulnerability to natural hazards in Italy used a diversity of indicators derived from the census records (Frigerio and De Amicis, 2016).

To:

A study on social vulnerability to natural hazards in Italy (Frigerio and De Amicis, 2016) used 7 indicators (i.e. family structure, education, socioeconomic status, employment, age, race and ethnicity and population growth) derived from the freely-available census records.

• Lines 125-126: Something seems to be missing here

Thank you for this comment, indeed, you are right:

The sentence has been changed from to:

The latter compared the hospital admissions due to HWs in summer months of three years (2003, 2006, and 2009) possible heat health issues among elder women.

To:

The latter compared the hospital admissions due to HWs in summer months of three years (2003, 2006, and 2009) and found heat health related issues among elderly women.

• Line 135: Might be worth to name the indicators again

Thank you for this suggestion, this has been done by changing the sentence from:

Quantify HWs and CWs hazards and their return level at a very high spatial resolution (250m) by combining for the first time i) the indicators proposed by Russo et al., (2015) and Smid et al., (2019), together with ii) the Tweedie distribution

to:

Quantify HWs and CWs hazards and their return level at a very high spatial resolution (250m) by combining for the first time i) the indicators (HWMId, CWMId) proposed by Russo et al., (2015) and Smid et al., (2019), together with ii) the Tweedie distribution

• Lines 169-174: To me, this seems to fit better to the discussion or conclusions, as it is an assessment (and opinion) rather than a description of methods.

Thank you for this comment, the authors agree with this and the paragraph in question has been moved and is now the first paragraph of the discussion.

• Line 211: “always <T 0” - maybe a typo?

Thank you for noticing this, the T was indeed a typo and has been removed.

• Lines 341ff: Same text appears twice.

Thank you, the duplicate has been removed.

• Line 344: I think that it should be shortly explained what “robust regression technique” is, and how it differs, e.g., from least-squares regression (see also general comment above)

This has been done. The relevant section has been changed from:

The trends are analyzed using the robust regression technique (Huber, 2011). This method is often used throughout the literature for assessing trends in natural hazards (Formetta and Feyen, 2019 for multiple hazards and Kishore et al., 2022 specifically for HWs).

To:

The trends are analyzed using the robust regression technique (Huber, 2011) which is often used to assess trends in natural hazards (Formetta and Feyen, 2019 for multiple hazards and Kishore et al., 2022 specifically for HWs). Robust regression seeks to overcome part of the limitations of traditional regression analysis.

For example, the linear regression least squares method is optimal when the regression's assumptions (normal distribution, independence, equal variance) are valid (Filzmoser and Nordhausen, 2021; Khan et al., 2021). This method can be sensitive to outliers or if normality is dissatisfied (Khan et al., 2021; Brossart et al., 2011). The robust regression method is designed to limit the effect that invalid assumptions have on the regression estimates (see Filzmoser and Nordhausen, 2021 and Alma, 2011 for more details).

• Lines 366-368: I do not understand this sentence. Why does the test reveal a significance level? Isn't the significance level chosen by you? And what does the false discovery rate have to do with this? The latter is rather a method to correct for potential overestimations of the number of grid cells classified as statistically significant.

Thank you for this comment, this is indeed not clear. The authors have changed the sentence from:

A KS test (Figure S3 in the supplementary material) shows that the Tweedie distribution provides a good fit for both CWMId and HWMId, with power parameter values between [1,2] for the entire region. The KS goodness of fit test reveals a significance level of αsig=5% as well as the false discovery rate for the significance level 2αsig for any pixel in the region.

To:

The KS tests p-values (Figure S2 in the supplementary material), indicate that the fitting of the Tweedie distribution with power parameter values between [1,2] cannot be rejected for both HWMId and CWMId.

• Lines 386-395: I am not sure how I should interpret the population numbers (especially the ones above 1 million) given that the analyzed region has a total population of about 1 million. How can the affected population be 5 million? Is it something like person-years, i.e, the sum over all exposed people in all years? This should be specified in the paper.

Thank you for this comment, it is indeed the sum of the people affected over this time frame. This is now specified in the paper and the sentences are changed from:

In total, between 1980 and 2000, in the study region, about 900 000 people were exposed to a 5-year HW event, 250 000 to 10-year HW event, 3million to 5-year CW event and 1.9 million to 10-year CW event. Between 2000 and 2018, the values increased to over 5millions for 5-year HW event and to about 2.5million for 10-year HW event but decreased to 2.4 million for 5-year CW event and to 500 000 for 10-year CW event.

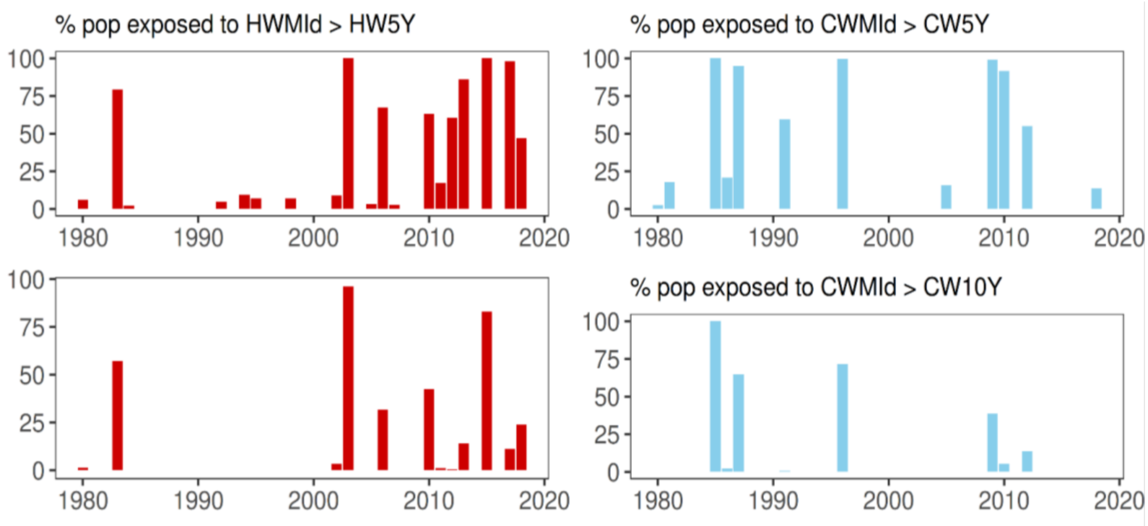
To:

Summing the overall number of people exposed over intervals (i.e. one person can be exposed each year and therefore counted multiple times over the interval), between 1980 and 2000 about 900 000 people were exposed to a 5-year HW event, 250 000 to 10-year HW event, 3million to 5-year CW event and 1.9 million to 10-year CW event. More recently, between 2000 and 2018, the population exposure values increased significantly to over 5 million for 5-year HW event and to about 2.5 million for 10-year HW event but the numbers decreased for CW events, to 2.4 million for 5-year CW event and to 500 000 for 10-year CW event.

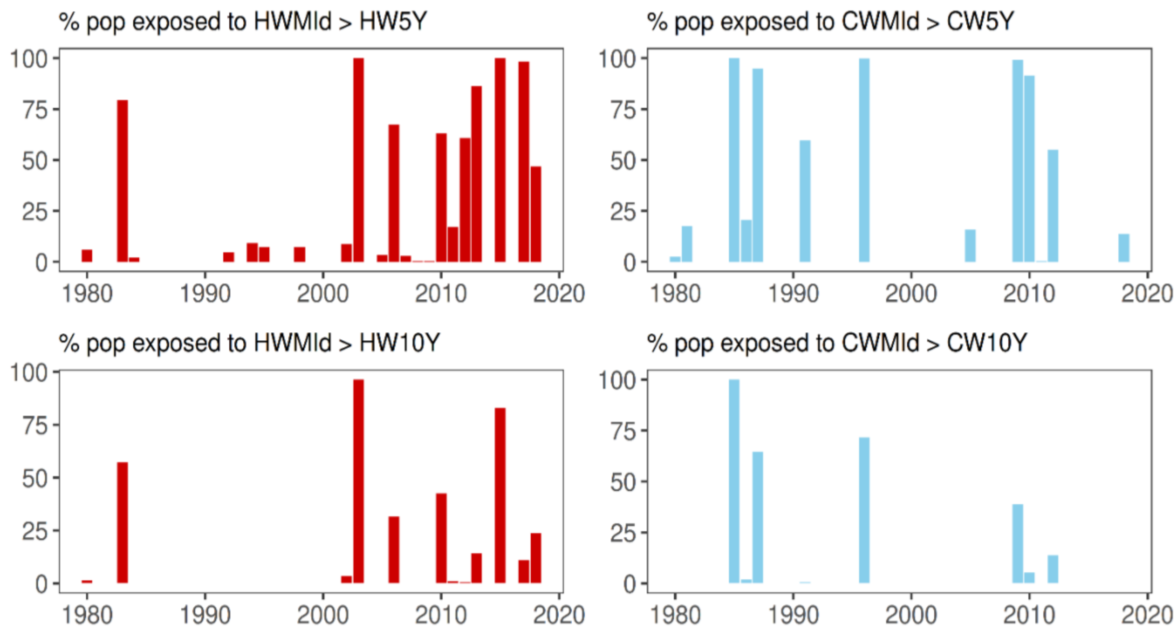
• Figure 3: The title of one subplot is missing

Thank you for noticing this, this has been adjusted.

The previous figure:



The new figure:



• Lines 483-486: I am not totally convinced by this argument given that the highest mountains in that region are not where the strongest HWs occur. I would be more cautious with this statement.

Thank you for this comment.

We agree with the reviewer, and we rewrote the sentence in order to be more cautious in the concept expressed:

The text was nonetheless adjusted to justify this statement from:

The location of our highest increasing trends in HWs events are concordant to those of the higher increase in temperatures found at higher elevations by Acquaotta et al., (2015) in north-west Italy

To:

Figure 3a indicates that a strong increase in heatwave trends is observed in the northwest and the north of our study area. Both areas are at a high elevation (between ~1000m and ~3900m) and one includes the highest mountain in the analyzed area. These results are consistent to those presented by Acquaotta et al., (2015), which found higher increases in temperatures at higher elevations in north-west Italy.

• Lines 492-493: I do not understand this sentence. What is it meant to say?

Thank you for asking, this has been clarified, the sentence has been changed from:

The two driving factors behind the increase in vulnerability (elderly population and isolation) have also been found as some of the main factors for vulnerabilities in other regions of Europe (López-Bueno et al., 2021; Poumadère et al., 2005)

To:

Consistently with previous studies (e.g. López-Bueno et al., 2021; Poumadère et al., 2005), we found that the elderly population and isolation were the indicators most affecting the increase in extreme temperature vulnerability.

• Lines 517ff: I would say that in certain areas also the change in population plays a role according to Figure 6.

Thank you for this comment, this is what the authors meant by demographic changes (ie. an increasing and aging population). This has been specified further in the article, the sentence is changed from:

The changes in CWs risk is mainly explained by the demographic and vulnerability changes, which are increasing in/around urban areas and decreasing elsewhere.

To:

The changes in CWs risk are explained by the demographic (i.e. an increasing and aging population) and vulnerability changes, which are increasing in/around urban areas and decreasing elsewhere.