We thank the Anonymous Referee #2 for his/her positive appreciation of our work and useful comments. Here we provide some replies to his/her comments and suggestions:

I/ Certain pieces of the results suggest that the data had not been checked for outliers before homogenisation and/or that QC was not very effective. The study does not contain anything about general QC. —- The suspicious pieces of the results: a) The monthly variation of spatial correlations (Table 4) is rather hectic, b) seasonal changes of corrections terms have strange structures, which hardly can be explained with any physical phenomenon apart from a low signal to noise ratio whose likely origin is random large errors in the data.

The quality control is mostly manual. Some definite outliers appearing in the data sets, most probably, due to typing or/and OCR procedure were corrected after first visual inspection in tabular and graphical forms. This information is now added in the text. As to the data from Table 4, we see here no serious problems. First of all, the series from Porto is very well correlated with series from Coimbra whereas correlations with Lisbon series are smaller. This is in agreement both with the locations of the stations (Coimbra is much closer to Porto than Lisbon) and with local climatic zones – the region around Porto (north part of the country) is under strong influence of the Atlantic cyclones. Coimbra region, despite its relative proximity to Porto, is more or less protected from this influence by the inland location and surrounding mountains. The region around Lisbon is located quite far from Porto and the local climate depends less on circulation patterns passing through the north of the country, leading to overall lower correlation between temperature series of Lisbon and Porto.
This zonality is particularly visible in correlations of Tmax in autumn-winter seasons (from November to January) – see Table 4 in the reviewed manuscript as well as prepared on its base Figure R1 in this Reply – the period of strong cyclonic influence in the North (Figure R1, right panel)

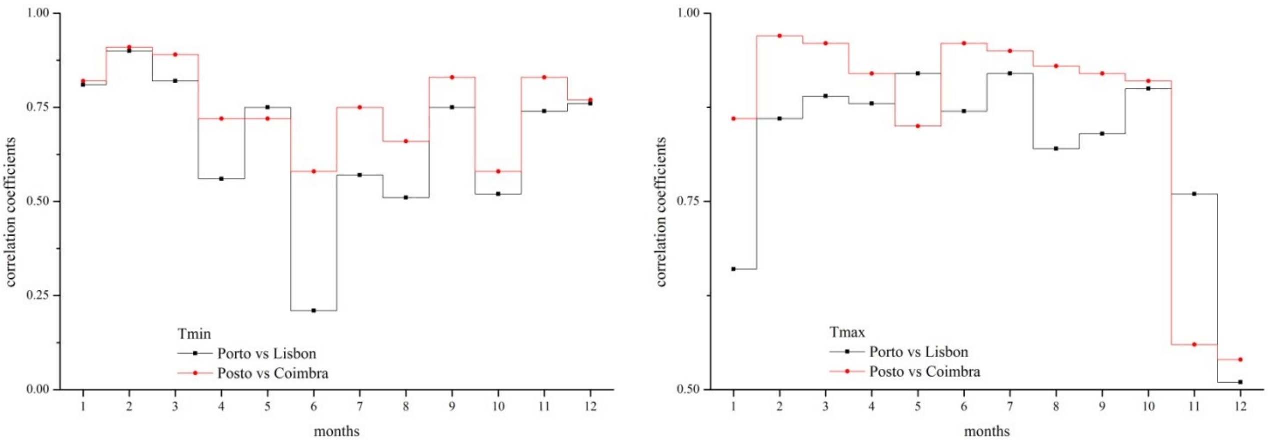

Figure R1. Correlation coefficients between Tmin (left) and Tmax (right) of Porto vs temperature in Coimbra and Lisbon (data are from Table 4 in the reviewed manuscript). Beware that the correlations axes have different scales in each figure.

As to the shape of the correction curves, we suppose that this depends on the sensitivity of the temperature parameter (Tmin or Tmax) to the particular changes in the microclimate due to thermometer relocation for a particular season. However, we have not paid specific attention to the particular changes in the thermometer environment and their possible effect on Tmin/Tmax for different seasons.

II/ Justified scientific facts are mixed with hypotheses. I think that the inclusion of hypotheses is allowed, but it should be clear from the text that they are hypotheses (e.g. by using "we think", "we suppose", "likely", etc. If sg. is only a hypothesis, then it is better not to repeat that too frequently. Only hypotheses and not justified facts are:

i) Volcanic effects on inhomogeneity-biases: Volcanic effects usu. last 1-4 yr, with sharply lowering intensity from the second year. In contrast, the study deals with inhomogeneities whose biases last for one or more decades. If Authors suppose that short-term effects sometimes could provoke long-lasting changes in the climate systems and consequently in the observed data, first they should write clearly their hypothesis. This hypothesis has little evidence, thus I suggest reducing the references to volcanic effects throughout the paper.

The homogeneity tests (HT) we used in the study do show homogeneity breaks that coincide with strong volcanic eruptions. All this coincidences are seen in the Figs. 9, 17, 25 that show HT results for corrected data (and sometimes in Figs. 5, 13, 21 that show HT results for original data). Of course, not all remaining inhomogeneities coincide with volcanic eruptions, but some of them do coincide. We have no evidence (and did not intend in this paper to prove) that these coincidences have cause-and-effect relationships. However, we think that it is a correct scientific approach to pay attention to these coincidences and to suggest an explanation. Also, we suppose that even short-term temperature variations after volcanic eruptions could be picked up by the HT if these variations are above the ordinary deviation level of the data series (as it could be for homogenized series).
On the other hand, we agree with the Referee that the references to the volcanic effect have to be reduced, at least, to avoid repetition and improve the style.

ii) DTR is a better indicator than Tmax or Tmin. It has more evidence than long-term volcanic effects have, but is still a hypothesis only. I suggest changing the wording to: "non-climatic jumps [of DTR] often can be seen more clearly" (bottom of page 532)

On the whole, we agree with the Referee and corresponding changes are inserted in the text.

iii) "The use of tests of different type (parametric, non-parametric, ... helps to obtain more significant results" —- It is often believed, but it is in fact a hypothesis only.

In our opinion, when some results (dates of homogeneity breaks, in our case) could be supported by the different methods (different kinds of HT, in our case), they are more robust, than results obtained by single methods. Of course, we did no math test to prove this wide believe, so we could change the text "… helps to obtain…" to "… could help to obtain…".

III/ Some statements are imprecise, and the meaning becomes misleading: i) In the definition of Tn and Tx: I am sure that they are not monthly minimum and monthly maximum as it is written in the study, but they are the monthly averages of daily maxima and minima.

Appropriate changes are inserted in the text.

ii) Pieces of the results that no non-climatic breaks are found with 95% significance transformed to the text (in several places of the study) to: there is no non-climatic break with 95% confidence. However, from the former statement does not follow the later. —- Such statements of confidence should be deleted, they are very misleading, in fact the corrected series likely still have inhomogeneities but the signal to noise ratio does not allow their detection and correction with adequate confidence.

The HTs of the final corrected series show no breaks around dates of thermometer relocation that go beyond the significance level of 95%. In fact, for many cases the significances of still existing breaks are much lower than 95%, but on the whole we chose the 95%-level that satisfies all cases.
In our mind, this comment from the Referee shows that the sentences in our paper are not clear and require re-phrasing. We changed "Thus, we consider the data sets of *Tmin* and *Tmax* corrected by the

procedure described in the paper as free of non-climatic changes with a significance of at least 95%." to "Thus, the data sets of corrected *Tmin* and *Tmax* have no non-climatic breaks that could be detected by the described homogeneity tests with more than 95% significance."

iii) Measurement error 0.1^{o}C—mentioned several times in the paper. However, it refers to the expected measurement errors, which sometimes might be strongly exceeded (e.g. by personal error). Please write first that the standard measurement error is +/- 0.2^{o}C, and there is no need to repeat that later.

Done.

IV/ There are too much repetitions throughout the paper, e.g. the correction method is written well in Sect. 2.2, but its principles are repeated again and again when the corrections for a selected station data are discussed.

The repetitions have been removed.

V/ "...four of the most popular homogeneity tests have been used" (p. 525). Popularity is not a scientific argument when statistical methods are selected. As in this study the role of statistical tests is secondary relative to the use of metadata, I suggest explaining it in the introduction. Then it can be written that the inclusion of some simple and widely used statistical tests are appropriate to the examinations of the study.

Done. Now it states that "Despite the fact, that the main role in the detection of the breaks was assigned in this study to the metadata, four simple and widely used statistical homogeneity tests have been applied to the data:..."

Minor comments:
i) "homogeneity breaks" — it is not a usual term (breaks or change-point are more usu.), I suggest writing first "homogeneity breaks (hereafter: breaks)" then using "break".

Done.

ii) Metadata lists are often incomplete and it would be nice to mention that in the paper. Note: Metadata list can be incomplete even with high level data-management, since it is hardly possible to realise and document all the possible changes in the influencing factors (changes of environment, of the treatment of natural vegetation, in the cleanliness and reflectivity of the screen, habitual personal errors, etc.)

Done. The following sentence was added to Section 2.2: "It is possible that the metadata do not list all changes in the stations' environments occurred during the measurements periods, however, in this study we found no significant (as estimated by the statistical tests) breaks that could not be associated to metadata records or other sources."

iii) At the bottom of p. 534 the composition is strange, because from November to December and from July to October of the same year is simply from July to December, so maybe that something is incorrect here.

We grateful for finding this mistyping. Here should be "November to December **1920**" *iv) "unbiased homogeneity tests" (p. 541): I suggest writing statistical tests or statistical homogeneity tests instead.*

Done.