We thank Referee#1 and #2 for the insightful comments, which will surely improve the quality of the manuscript. We have revised our manuscript in accordance with the comments and suggestions received from two reviewers.

We believe we have properly addressed all concerns and added the necessary material to the text and figures to strengthen our manuscript.

Please find below the details responses to the referees' comments (*in italic*) with associated changes (**in bold**) in the revised version. We thank the reviewers for his/her positive evaluation and appreciation of our work, as well as for the constructive comments and valuable suggestions to further improve the manuscript.

D. Lee (Referee#1)
Received and published: 23 December 2015
The results are presented clearly and concisely in the text, but the figures and supplement should be completely reworked before publication.
In general, the paper's conclusions are limited to the scope of point observations made at stations which are unevenly spatially distributed. This is acknowledged by the authors and for the most part treated sensibly.

1) *However, certain statements in the paper are misleading due to the nature of the data analyzed - e.g. on lines 12, 13, pg. 738 in the abstract. Here the authors state that "up to 75% more events [occured] in the last 10 years." Strictly speaking, up to 75% more events were observed in the data. Coverage in wide areas, e.g. in Brazil and Argentina, is sparse, and especially dense in other areas, e.g. in Ecuador. As heat/cold waves do not occur homogeneously throughout all of South America, the locations of the observing stations surely leads to a difference between the increase in occurrences in all of South America vs. observations made at these specific stations. I agree with the authors' choice to use the station data, especially since the metrics they use are based on daily minima and maxima, which are more accurately captured in the spatially sparse but temporally high resolution stations than in reanalysis data. However, it is important to acknowledge the limitations of the data.*

We do recognize that the statement "up to 75% more events" could be misleading and it has been removed thereof. We were uncertain whether to use more 'quantitative' facts of results obtained. The expression "*up to 75% more events in the last 10 years*" has been rephrased such as "**especially in the last 10 years**". This expression has been modified throughout the entire manuscript.
In addition, a sentence has been added in the Discussion section to accentuate the spatial limitation of the dataset:

"In addition, as heat/cold waves do not occur homogeneously throughout all of South America, the location of the observing stations surely influences the results"

2) *Also, the supplement is not really helpful - it is composed only of plots which have the same general problems as the plots included in the paper (see the next section for details). It would be very interesting if the actual data - i.e. the occurrences of heat / cold wave events*

and their magnitude - were included in the supplement, e.g. as CSV. As it is, the supplement could just as well be left off.

The actual data (both heat and cold wave annual occurrences for each GSOD station) has been included in the supplement as CSV. Moreover, the figures has been modified according to the reviewer's suggestions (see replies 13 to 21 reviewer#1).

3) *Lines 3-6, pg. 7383: I'm not convinced that this paragraph is necessary, and in its present form it is definitely not meaningful. First of all, the proportion of available station data per country is not relevant for the rest of the study. Additionally, comparing the absolute number of available stations without normalizing for the area of their containing units - countries - is misleading. It should be clear that Brazil has many more stations than French Guyana, since it's so much larger.*

We do agree with the reviewer, and for the sake of clarity this paragraph and the relative table (i.e. Tab. 1) have been removed.

4) *Line 10-11, pg. 7385: The authors state that "... The intermittent nature of heat waves prevents from carrying out trend analysis..." Without having seen the data, I would assume that a Poisson regression would be well-suited to analyzing its trends. If this type of regression is unsuitable, there are surely other types of trend analysis that can be used on sparse observations that occur with low frequency. The trends such an analysis would project would be much more interesting for the paper; heat and cold events are simply not expressed in the yearly averages and the strength of this paper is that it examines these events using a robust, generalized metric suitable that captures these events well.*

Following the suggestion of the reviewer we have tried to fit heat and cold waves distributions using Generalized Linear Models (GLMs). Specifically, we counted the number of heat or cold waves for each 5-year period greater than a fixed level of magnitude (e.g. 2) and we applied GLMs with

1) formula = HWgt2 ~ Time
2) family=poisson or binomial.

Where HWgt2 is the number of heat (or cold) waves greater than 2 for each 5-year period, and Time is the temporal variable, i.e. a sequence ranging from 1 to 7 (step 1) which represents the 5-year periods from 1980 to 2014.
By applying GLMs we got a number of stations across South America which display a statistically significant regression always lower than 10 (out of 254 stations!).
Maybe, the size of the temperature sample is very small in order to be robustly modelled with a non-stationary peaks-over-threshold (POT) process. We have modified the text accordingly.
"**Note that** the intermittent nature of heat and cold waves **and the abrupt change in the last 10 years (see section 3.1)** prevent from carrying out trend analysis for HWMI and CWMI **using non-stationary peak over threshold models or other trend analysis (for further information, annual values of both HWMI and CWMI are available in the Data Annex)**.

5) *Lines 1-3, pg. 7387: This is a good observation, but the authors are comparing a period of 9 years with a period of 24 years. It makes the conclusion they draw from it all the more*

interesting, but it should either be normed (e.g., "between 2005 and 2014 the frequency of extreme heat waves had increased to 40 observations per year, as compared to 8.5 per year in the period from 1980 to 2004") or accentuated in the text.

Thanks for this fair comment. According to the reviewer's suggestion, we have corrected the text in the revised manuscript.

6) *I find the figures problematic (more about that in the following section), but the information they contain is very interesting. It would be nice if the hot spots of increases in HMWI frequency and magnitude would be discussed in the text. As it is, this gem disappears unless the reader discovers it on his own.*

Thanks for this fair comment. According to the reviewer's suggestion, we have corrected the text in the revised manuscript (see reply 19 reviewer#1).

7) *Lines 9-16, pg. 7387: Here the authors describe an example year. Is this year significant for some reason, or was it chosen at random? Both would be fine, but it should be stated in the paper.*

This year (i.e. 2013) has been used as an example because it is characterized by a high number of heat wave events (as 1982 or 1997 among others - Fig. S3 of the Annex). "**By way of example**" has been added to introduce the Figure.

8) *Section 3.4., specifically pg. 7390: I'm not convinced that the indicators used are suited. First of all, the second indicator is Mean Absolute Error, not Absolute Error (AE would be for a single observation). Secondly, both RMSE and MAE are used for measuring errors, e.g. in predictive models, and this is not the case here. The authors are examining deviance of a subpopulation from its superpopulation. In this case, deviance indicators, not error indicators, should be used.*

We agree with the above reviewer comment and for the comment 8 from reviewer #2. A Kolmogorov-Smirnov test has been used instead. Text has been modified accordingly:

"The two-sample Kolmogorov-Smirnov test has been employed to test whether daily average in the 2010-2014 and daily average for 1980-2014 come from the same distribution, for both maximum and minimum temperature.

The p-value has been calculated, getting the result of 1.14e-09 and 1.82e-05for maximum and minimum temperature, respectively. With a so low p-value, it is reasonable to assume that they do not come from the same distribution. The null hypothesis is rejected and we can infer that different statistical distributions are indeed present, thereby confirming our graphical inference."

9) *Lines 16-18, pg. 7380: This sentence can be inferred from the previous ones. I suggest leaving it out.*

According to the reviewer’s suggestion, we have corrected the text in the revised manuscript.

10) *The English in the paper is good, but the study would benefit nonetheless from proofreading.*

The revised manuscript has been checked by an English mother colleague.

11) *Line 14, pg. 7381: I would write "sensing network" rather than "gauge network".*

Done.

12) *Line 16, pg. 7382: It would be nice if section 2.2. were referenced here so that the reader knows where to look to find criteria for selecting stations.*

According to the reviewer’s suggestion, we have corrected the text in the revised manuscript.

3.1 Figures
In my opinion, the figures should all be completely reworked.

13) *First of all, figures should be interpretable without having to know where they are described in the text. The text can, of course, refer to them and provide context, but units, variables, etc. should be in captions next to the figures rather than in the text. Too often relevant information is written in the text and not in the caption, which both makes the figures difficult to interpret and interrupts the flow of the text (e.g. lines 15-18, pg.7386; lines 16-17, pg. 7388; lines 12-15, pg. 7389; other paragraphs might be affected as well).*

Figures have been completely reworked and the relative text has been edited (and often removed) in the most part.

14) *Also, in many cases the figures are scaled so that the reader is required to analyze and interpret a lot in order to understand it in a way that is not misleading. Readers shouldn’t have to do math in their head - figures should display the data in a way that the data can be interpreted as intuitively as possible.*

According to the reviewer’s suggestion, we have corrected the figures in the revised manuscript. Scales have been harmonized throughout the manuscript.

15) *Concerning the maps - South America is such a big area that the latitude/longitude axes are not necessary. Especially in the small maps they’re distracting.*

According to the reviewer’s suggestion, we have corrected the figures in the revised manuscript. Axes have been removed.

16) *Figure 2: The axis labels are too small. Also, the combination of bargraphs and maps reduces readability. Consider breaking this up into two plots. Also, the magnitude classes are a bit confusing - does 16-32 occur at all? If it does not (I was unable to find an occurrence) it should be removed from the legend.*

Axis labels have been removed; and bargraphs and maps have been split in two parts.
Since the magnitude class 16-32 occurs only three times (see HWMI_TX.csv in the annex), it has been merged with the class 8-16 for the sake of clarity.

17) *The increase in heat wave frequency could be better demonstrated by using the same scale in all bargraphs - otherwise the reader sees only that frequency decays with intensity if he does not mentally scale the bars. The same applies to figure 5.*

As suggested by the reviewer, we modified the figures using non-log scales y-axis.

18) *Figure 3: The authors use a logarithmic y-axis. It would be helpful to state that on the scale. Also, the legend is interpretable, but it's clearly from the plotting program and written differently than if it were being prepared for readers. I would suggest completely redoing this figure. First of all, the variable names need to be in the caption. The plot could also be interpreted more intuitively if different colors were reserved for magnitues and time were plotted on the x-axis, rather than magnitude on the x-axis and colors signifying time periods. The authors might also consider adding trend lines. The same applies to figure 6.*

According to the reviewer's suggestion, we have corrected the figures in the revised manuscript. However, we do prefer not to add trend lines to not weight down the figure. Text has been modified accordingly.

19) *Figure 4: This plot shows an interesting spatial distribution. It would be nice if this were discussed in the text. Also, as in figure 2, the class 16-32 does not occur!*

According to the reviewers' suggestions we modified the text such as:
"The spatial distribution of the HWMI also shows the hot spot of increase in HMWI frequency and magnitude from 2009 onwards across Paraguay and southeastern Brazil, as shown in Fig. S3 of the Annex. However, spatial patterns of heat waves are heavily influenced by the distribution of GSOD stations throughout the continent."

For the magnitude class 16-32, as shown in point 16 reviewer#1, class 16-32 has been removed and merged with class 8-16.

20) *Figure 7: Please use uniform descriptions in the caption for the panels within the plot, e.g. "...mean annual maximum temperature (TX, left panel)...". I would suggest using uniform maxima / minima for all 3 subplots. However, MTR should use a different color scale. It's not describing an increase or decrease of temperature, but an increase in the the spread between minimum and maximum. A grayscale or some other scale that indicates intensity would be better suited, otherwise readers might interpret a bright red value of 1.0 to mean that warmer temperatures were observed, whereas this value could theoretically occur at*

a station where the minimum gets colder faster than the maximum temperature, even though both means decrease over time.

As suggested by the reviewer, we modified figures and captions. However, uniform minima for all 3 subplots hinder the major changes in temperature. We used a colorscale ranging from green to purple for the MTR to avoid misinterpretation.

21) *Figure 8: The axes are too small to be readable with many printers, please enlarge them. It looks like the authors are using R - in this case the plot could be painted in a smaller viewport and the font size would be larger in proportion to the rest of the plot automatically.. The values are very jittery, and this makes interpretation of the figure unnecessarily difficult - the long-year average temperature on a specific day in the year is not meaningful. I would smooth the values using e.g. a moving window of 5-10 days. This would also remove the outliers, which don't provide meaningful information, and simultaneously reduce the range on the y-axis, which would increase the information density shown. The same applies to figure 9.*

Thanks for this fair comment. We increased the axis fonts and smoothed the values using a moving average of 10 days for both Fig. 8 and 9. Formulas have been removed, and text has been rephrased accordingly such as:

'Figure 8 and 9 show the **10-day moving average** of daily maximum and minimum temperatures, respectively. The average values (i.e., mean ± 2 standard deviation) of the 1980-2014 time series are in **dark brown**, the highest and lowest values in **light brown**.

The black lines in Fig. 8 and 9 represent the **the 10-day moving average** of daily average for 1980-2014 of maximum and minimum temperature, respectively.

Similarly, the **orange** lines represent the **10-day moving average** of daily average in the most recent 5-year period (i.e., 2010-2014).**"**

Anonymous Referee #2

The paper by Ceccherini et al. describes the extreme temperature regime of heat waves and cold waves across South America over recent years (1980–2014). To do so, the authors use the Global Surface Summary of the Day (GSOD), a climatological dataset produced by the National Climatic Data Center. The topic of this study is of interest to be published. However, the manuscript is unclear on several occasions and have several methodological issues that should be dealt with. Hence, the paper is returned for a revision. The list of comments below should be taken into account before the article is ready for acceptance.

1) Please remove the references from the abstract;
According to the reviewer's suggestion, we have corrected the text in the revised manuscript

2) L21- Which areas?

Thanks for this comment. We added "**of the world**".

3) Section 2.1 and 2.2. More information regarding the weather stations is needed.

Unfortunately, we used all the information available in literature about GSOD dataset.

Can the authors comment the percentage of stations at different altitudes?

The vast majority of GSOD stations lies below 1000 m asl, as shown in Fig.1. However, as stated in section Methods,
"The HWMI is a normalized index, therefore it automatically removes the effect of the different elevations of GSOD stations"

In addition, I find the 30% gaps threshold a bit higher. Why the use of 30% and not 25% or even 20%? Can the authors give some information regarding the percentage of stations with different missing values ?

Thanks for this fair comment. 30% of gaps does not hinder HWMI and CWMI retrieval and represents the trade-off between accuracy of the time series and representability of GSOD stations. We also noticed that the number of stations was incorrect. There are **912** GSOD station across South America (instead of 851), and only **254** (instead of 705) pass the quality check (i.e. more than 30 years of data and less than 30% of gaps). The text and these figures throughout the paper have been corrected. Also Fig. 1 has been modified. We added these sentences:

"Note that 30% of gaps in a time series does not hinder the HWMI and CWMI retrieval, and using that threshold we already exclude ~72% of the GSOD stations (i.e., 658 out of 912). A further decrease in that threshold will reduce excessively the number of available temperature stations."

4) Section 2.2.1. I did not understand all the details regarding the computation of the heat and cold wave magnitude indices. The authors stated :"In this work this index and the corresponding Cold Wave Magnitude Index (CWMI), defined below, are used to detect South American heat and cold waves in the present climate." So, how many indices were computed by the authors ? In the present form is seems that the use of CWMI was also used to compute the heat wave indice. Therefore section 2.2.1 must be completely re-organized with both definitions computation (HWMI and CWMI) to be clear as possibly. With this it is impossible to fully understand the main results.

We have rephrased this part of the revised manuscript as in the following:

"Recently, Russo et al. (2014), have introduced the HWMI able to overcome the limitation above by merging a few climate measures, as duration and temperature anomaly, into a single numerical index (Hoag, 2014). Basically, the magnitude index sums the probability scores associated to consecutive daily temperatures above a threshold (**for further details see Russo et al., 2014). The HWMI is a normalized index, therefore it automatically removes the effect of the different elevations of GSOD stations**. The HWMI computations requires a 30-year long time series of daily temperature records; the latter normally refers to the 1981-2010 timespan, taken as reference period. **Because of the lack of agreement on Cold Wave definition, Forzieri et al. (2015) have recently introduced the Cold Wave Magnitude Index (CWMI). The CWMI is computed in a similar way to the HWMI, merging thus the duration and the intensity of the extreme events into a single numerical index. In this work, both HWMI and CWMI indices are computed to detect South American heat and cold waves in the present climate. For further details on HWMI computation, see Russo et al. (2014). Regarding the CWMI, it has been defined for each GSOD station by the following steps:"**

> *5) P7384L22. The computation of the indices end at 2014 ? The date "30 June 2015" seems a contradiction to what was said before.*

We rephrased the text and we added a sentence to explain this 'shift':

"Since heat waves generally occur between December and January **in the southern hemisphere (i.e. throughout the vast majority of South America**), the calendar year used for HWMI starts in July and ends in June. In such manner we avoid splitting in two heat waves that are likely to happen at the end of December and the beginning of January. Therefore, the HWMI computation starts on July 1, 1980 and ends on June 30, 2015. **Because of this 6-month "shift" we will refer - for heat waves - to 2014 as the year starting on July 2014 and ending on June 2015, and so on**."

> *6) Why the particular analysts of the 2013 heatwave? It was the most intense during the 30 years analysis? Why not doing a similar example to the cold wave?*

2013 is presented as an example because of the many heat waves that occurred. The text has been edited accordingly (see reply 7 reviewer#1). Maps of coldwave are shown in the annex.

> *7) P7389L3 The stations with statistical significant trend do they appear clustered in some areas? or they are scatter throughout the domain?*

Only stations with statistically significant trend have been shown. Having said that, stations in Fig.7 are clustered across southeastern Brazil and Paraguay, but it is heavily influenced by the spatial distribution of the GSOD stations, as discussed in the Discussion section.

> *8) Section 3.4 I believe that they are to many equation in this section. Most of the computed indices are rather straightforward, therefore a simple explained is enough. In addition, i do not believe that the use of equation 5 and 6 are the most appropriate in this context. Probably a two-sample Kolmogorov–Smirnov test would be more appropriate.*

Thanks for this fruitful comment. We do agree that the two sample Kolmogorov-Smirnov test is more appropriate. In additions, since the equations are rather straightforward, they have been removed (see reply 21 reviewer#1).

9) Figure Captions need to be improved. Figure 1. You can improve this Figure by added a color representing a range of altitudes to each dot..

Thanks for this comment. We modified Fig.1 according to the reviewer's suggestions showing the elevation of each GSOD station.

10) Figure 2 and 5. What does the bar plots represent?

Fig.2 and 5 have been modified. Specifically, maps and bars have been split in two figures (see point 16 reviewer #1). Bars represented the occurrence of heat and cold waves for each 5-year period.

11) Figure 3 and 6. The X - axis must be improved in order to represent as much classes a s the ones in Table 2.

According to reply 18 reviewer#1 we decided to re arrange Fig. 3 and 6. Magnitude categories are represented with different colors, while the x-axis represents the time. Since the class 16-32 occurs only three times, we decided to merge it with class 8-16 (see reply 16 reviewer#1). However, we decided to provide the reader with the annual values of HWMI and CWMI with two .csv files in the data annex.

12) Figure 7 Which trends are significant?

All trends are significant, as explained in the text: "**Note that only stations with a statistically significant trend (i.e., p-value smaller than 0.05, or 5%) are shown**,"

And

"The number of GSOD stations that display statistical significant trends for maximum temperature, minimum temperature, and MTR **(i.e. all the stations shown in Fig. 7**) **is equal to 75**."