Response to RC1

We thank reviewer 1 for the constructive comments. Our response is written in blue and located behind each reviewer comment.

Review of "Wilcke, R. A. I., Kjellström, E., Lin, C., Matei, D., and Moberg, A.: The extremely warm summer 2018 in Sweden – set in a historical context"

This manuscript uses long-term observations and 5 climate model ensembles to quantify how extreme, or how unusual the extended summer season of 2018 was in Sweden. The main result is that although all climate model ensembles include summers of similar warmth, they are quite to very rare (depending on which metric is considered), and in particular May 2018 as an individual month was exceptionally warm. Furthermore, the authors show the likelyhood of a summer as warm as 2018 is higher in recent decades compared to in pre-industrial times or in the mid 20th century. The manuscript is within scope of ESD and is based on sound methods. The conclusions which are drawn are supported by clear evidence. Overall, the English language is of a good standard. There are a few odd turns of phrase / constructions but I expect copy editing will be able to address these. Some aspects of the manuscript could be clearer and as such I suggest some minor revisions which I list below.

1. The single largest issue with this manuscript is that there are a large number of time periods which are quite confusing to follow at times. Related to this, it is unclear why the reference period (1981- 2010) and a recent past (1989–2018) are both required. Furthermore, it is hard to remember for which time periods each observational data set and each ensemble is available. I strongly encourage the authors to include a table with all of this information in one place. A table could also indicate which data sets / climate models are available for each of the 4 time periods presented at the start of section 2. - We have removed one reference period (1961-1990) and further added a table about observational data sets to the manuscript. The data availability is now clear from Table 1 (observational data) and Table 2 (model data). The 30-year periods used for analysis are stated clearly in L80-81. We use one reference period, which is the new reference period defined by WMO, 1981-2010. The recent past or present day climate period is relevant as we are studying extreme temperature events in the recent past. Excluding the years beyond 2010 would give an out-dated picture of the available climate knowledge about extreme temperatures. We therefore argue, to keep both periods.

2. Line 33. There is no reference item for Räisänen (2018) in the reference list at the end. - Now included: Räisänen J. (2018). Energetics of interannual temperature variability. Clim. Dyn. 52: 3139–3156. DOI: 10.1007/s00382-018-4306-0

3. Line 36. Suggest adding "temperature" after "average". - We agree and add "temperature" to the sentence.

4. Line 36. Using the reference period 1961- 1990 adds to the confusion discussed in point 1 above. Could the 1981 – 2010 reference period be used instead ? - We

understand the possibility of confusion here and therefore will use only one reference period, 1981-2010.

5. Lines 62 – 63. How does this study differ to those of Leach et al (2019) and Yiou et al (2019)? Please add some details to the introduction. - Both studies are pure attribution studies. So we refer to their results as they underpin our conclusions drawn from simpler statistics without a factual-counterfactual approach. The main differences are the metrics which have been analysed, meaning different aspects of the heat event. Additional differences are how the heat event is defined, for which region the analysis was undertaken and what ensembles have been used.
    - Yiou et al (2019) analyse a 19-day window in July 2018 over a larger Scandinavian domain. Using regional climate model ensembles as well as a CMIP5 ensemble they do an unconditional and a conditional attribution approach to demonstrate the thermodynamic contribution of human-induced climate change to describe the probability and intensity of the summer 2018 event in Scandinavia.
    - Leach et al. (2019) analysed the return values of annual maximum of the 1-, 10-, and 90- day running mean of daily mean 2-m temperature for the factual and counterfactual simulations of HadGEM3-A global atmospheric model.

   We add to the introduction text in L63: "<u>Determining the probability of an event like 2018 being observed, plus the probability of occurrence conditioned on large scale patterns</u>, Yiou et al. <u>(2019) concluded</u> that the probability of an event such as the one observed in the second half of July 2018 has increased as a result of human-induced climate change. "
   Further in L69 we change the sentence to "The aim of the study is to <u>expand findings from previous studies by evaluating a broad range of</u> temperature conditions in Sweden during summer 2018 in relation to the historical climate."

6. Lines 73 – 76. Add details here to make it clear to a reader that only historical climate model simulations are considered – no long term future projections are considered here. - Changed to: "In a next step we investigate the likelihood of such a summer to <u>have occurred in the past century</u> using five large global climate model ensembles, some of which are covering a period since 1860, and others starting in the second half of the 20th century<u>, up to 2018</u>."

7. Lines 90 – 94. This information about the observations could be added to table 1. Please add details if daily or monthly values from E-OBS were used here. - We use daily and monthly data from E-OBS, depending on the kind of analysis step we were at. This has been clarified now in the sentence in L90: " i) the gridded daily and monthly climatology E-OBS version 19.0e …" We also added a table to give an overview over the observational data used.

8. Line 98. typo. "complement of correct" → complement or correct. - Corrected

9. Line 134. Again this relates to point 1 above. Here it is stated "we analyse 294 simulations adding up to 8820 summers over a 30 year period." Which 30-year period? Or multiple 30- year periods? Please clarify the text here. - The 8820 summers are for each 30 year periods (294 * 30 = 8820).
   This has been written more clearly now, "In total, we analyse 294 simulations expanding the sample size for each 30 year period to 8820 (30 summers times 294 simulations)."

10. Lines 138-139. From which observational data sets and climate models are daily maximum and daily averages taken? This information is partially given in Table 1 but it is not clear which data was used from the CMIP5 multi-model ensemble. - We added a table (Table S1) to the supplement which lists all CMIP5 simulations used here. In the comparison with climate models we used E-OBS only as observational data.
    Furthermore, are the results comparable if for some datasets / ensembles daily values are used whereas for others monthly values are used? - The monthly and daily analysis are meant to complement each other, rather than being strictly compared. The monthly mean is a good indicator for a larger heat event as a single heat day doesn't affect the monthly average as much. Another reason was that we wanted to use the MPIGE which, with its 100 members, is the largest ensemble, so far. The caveat is that only monthly data had been made available. For each metric we calculated and analysed we used the same temporal resolution for each ensemble. That means, analysis based on daily data does not include MPIGE.

    In L138 we write: "All analysis is based on daily and monthly temperature data." It continues further with "To assess the average temperature for the summer season we use monthly mean temperatures for four individual summer months (May, June, July, and August), separately and for two summer seasons (JJA, MJJA; averaged)." Here we use only the monthly data as stated.
    "Furthermore, three warm day indices, based on daily values, awere used to assess the temperature variability during summer:"
    At the end of the section we add: "The analysis is carried out for all model ensembles where respective data is available. I.e. the pdf analysis based on monthly data includes all five ensembles (c.f. Table 2 and Table S1), whereas the warm day indices are based on daily temperature values which are not available from MPI-GE."

11. Line 144 – 146. The definition of a warm day here is not completely clear. Specifically, is the 98th percentile threshold for each grid point the same for all days i.e. no time dependence between May – August? Please clarify this. - For each grid point we calculate the 95th percentile from all days within the 4 month period of May to August. Those 95th percentiles are then used as thresholds to define the warm days at each grid point. That means, the relative threshold is the same for all grid points (95th percentile). Sweden is a rather long country covering a big range of latitudes. A fixed threshold value for all grid points would lead to no heat days in the Northern part, though locally experienced and affecting the environment there had been heatwaves, relative to the average temperatures there.

We changed the sentence to: “For simplicity reasons we define a warm day as a day i with T_(max,i) greater than a relative threshold. The threshold is the 95th percentile calculated of all T_max in May to August (MJJA) during the reference period (see Eq. (1)).”

12. Section 2.3. It is quite hard to understand all of these diagnostics. Please explain what the differenes are between the WSDI and the warm days as defined by equation 1. Please also revise the explanation of the HWMI in lines 162 – 167 as currently I cannot understand this. - We restructured the section partly which reads now:
“All analysis is based on daily and monthly temperature data. We look at both daily average and daily maximum temperatures, as well as monthly means of daily average temperature. To assess the average temperature for the summer season we use monthly mean temperatures for four individual summer months (May, June, July, and August) separately and for two summer seasons (JJA, MJJA; seasonal averaged). Temperature anomalies are calculated against the reference period 1981–2010.

Furthermore, three warm day indices, based on daily values, are used to assess the temperature variability during summer:. The indices based on equation 1 are i) the “total number of warm days per year” (totWarmD), ii) the “maximum number of consecutive warm days per year” (max_conWarmD), iii) the “number of heat events” (tot_event), where an event is defined as minimum 3 consecutive days of T_max>threshold.
For simplicity reasons we define a warm day as a day i with T_(max,i) greater than a relative threshold. T, where the threshold is the 95th percentile calculated of all of MJJA T_max in May to August (MJJA) during the reference period (see Eq. (1)).
T_(max,i)>p95(T_(max,(MJJA 1981-2010)))
(1)

This simple definition based on a relative measure is chosen to make it possible to compare conditions in different parts of Sweden to each other. For example, a perceived heat wave in the colder North of Sweden, may not even appear in an analysis involving an absolute threshold representative of conditions in southern parts of the country like, e.g., summer days defined as days with T_max >25° C. Other examples of benefits with a relative measure involve comparison of coastal and inland stations or between low and high altitude stations for similar reasons.

Additionally we calculated two heat wave indices commonly used. The Warm Spell Duration Index (WSDI, Orlowsky and Seneviratne, 2011) that can be compared to max_conWarmD, that differ in their definition of the warm day. The WSDI is calculated based on an individual threshold for each day of the year (doy). A warm day is defined as a day with T_(max,i) larger than the 90th percentile of A_i, as defined in Eq. 2 (from Eq. 1 in Russo et al. 2014) .

$$A_i = \bigcup_{y=1981}^{2010} \bigcup_{d=i-15}^{i+15} T_{max,y,d} \qquad (2)$$

Where ∪ denotes the union of sets and T_(max,y,d) is the daily maximum temperature of the day d in the year y. The WSDI calculates then the maximum number of consecutive warm days (larger than 3), i.e. for a given year (or season), the WSDI is the longest duration of any such heat-wave event.
The second heat-wave index is the Heat Wave Magnitude Index (HWMI, Russo et al. 2014), which uses the same warm day definition as WSDI. Whereas the WSDI takes into account only the duration, the HWMI also considers the magnitude of the heat-wave. The HWMI takes into account multiple sub-maximum temperatures of an event by summing them up and mapping them to a probability (called magnitude) related to annual maximum magnitudes of the reference period. By that it weights the duration more than the absolute maximum temperature of an event. This relates to the heat stress which builds up with many warm days in a row rather than a couple of very warm days in a row. A more detailed description of how to calculate the HWMI can be found in Russo et al. (2014)."

13. Line 173. Add the years after "reference period" to remind a reader. - Added.

14. Line 178. Add "temperature" before "data". - Added.

15. Line 194. Add that the 95th percentile referred to here is calculated from the 1756 – 2005 Stockholm temperature data. This is another example of yet another different averaging period (see point 1 above). - We understand that we use many different periods, which we tried to give a better overview in Table 1 and 2 as well as in the text: "The historical context is given by comparing observed conditions in 2018 to observed and simulated climate conditions for: i) a pre-industrial period 1861–1890, ii) a mid 20th century period 1951–1980 and iii) our reference period 1981–2010. For some analysis we also look at the most recent past 30-year period ending 2018 (1989–2018) partly overlapping the reference period. For the longest possible time perspective, we also consider the period 1756–2018 using the Stockholm temperature series." However, the advantage and reason why we include the Stockholm temperature series, is its exceptional length. With comparing 2018 to only 1981-2010 we would miss out on the opportunity to compare an extreme summer to the past 250 years.
We edited the sentence: "For the full four-month period MJJA we note that more than 35 % of the days were above the long-term (1756–2005) climatological 95th percentile calculated for each day (upper grey line in the figure)."

16. Line 194. It would be interesting to know which other year exceeded the 95th percentile. Could this information be added to the text? - The only other year that has more than 20% days above the 95th percentile is 2002. We add that information to the text: "Only one additional year, the year 2002, exceeds the 95th percentile with 20 % of the days in a full May−August season."

17. Line 214. This paragraph needs to include what data the numbers / results discussed here are based on which I think is E-OBS. - Correct, the observations used here were the E-OBS data set. Edited to "Probability distributions (pdfs) of the four single

summer months (May, June, July, and August) from E-OBS temperature data in Fig. 2 (a–d) further illustrate the rarity of the weather situation in 2018 in Southern Sweden. "

18. Line 231, Add the relevant years after "For the recent past period (…).." - Added.

19. Line 249. Given that there are only 5 figures in this manuscript, Figure S3 could be moved from the supplement to the main article. - We agree with this suggestion and move Figure S3 to the manuscript, changing it to Figure 3 and changing former Figures 3-4 accordingly.

20. Line 268. Please clarify what is meant by "… the chosen fit". - The "chosen fit" refers to the way the distribution is calculated and presented. For E-OBS as "one-member ensemble", we chose to show a histogram which best represents how the temperature anomalies are distributed. For each ensemble we pooled together all ensemble members and calculated the kernel density estimate with a gaussian fit to gain a smooth curve. It has a strong visual aspect here, as nine histograms wouldn't be distinguishable in one figure. However, the density function curves go asymptotically to zero probability in their tails. When calculating the 100th percentile (indicated by 1 on top of the bars in Fig. 2) this value can be larger than the empirical value, meaning that it can be outside of the actual value range provided by the ensemble members.
We edit Line 268 and add: "This also indicates that CMIP5 temperatures like in May 2018 fall only into the distribution because of the chosen fit of the Gaussian pdf estimate. (The pdf curve continues smoothly to zero even if there is no value at the end of the tail.)"

21. Figure 2. Please expand the explanation in the caption of the colorbars at the top of each panel. - The caption reads now "Probability distributions for monthly average temperature anomalies, calculated for 1861-1890 (dashed) and 1989-2018 (solid) against 1981-2010 for the Southern half of Sweden, for the months of May, June, July, and August, and seasonally pooled for JJA and MJJA. The bars in the upper part of each panel are a guide to compare the positions of percentiles for each ensemble and period. The opacity of the bars indicates in steps the .9, .95, .99, and 1 percentile ranges marked on the uppermost bar. Ensemble distributions are a kernel density fit, whereas the histogram for EOBS is based on actual data. The observed year 2018 for respective months/season is indicated by the vertical dotted black line."