Investigating the observed sensitivities of air-quality extremes to meteorological drivers via quantile regression

Porter et al.

Referee #1 Comments

(authors' responses in italics)

This seems to me a very nice paper. It furthers an understanding of the meteorological drivers of air pollutant extremes across the U.S., both how they differ regionally and seasonally, and the differences between ozone and aerosol pollution. It also helps clarify the meteorological drivers of extreme events. I think this will be a valuable contribution. I would recommend publication after the rather minor comments below have been addressed.

Major Comment: My main concern is that the methodology is not always clear. (i) A central theme of the paper concerns quantile regression. I would guess this procedure is not widely known in the meteorological/chemical community. I would recommend adding a short section to the paper within the methodology section explaining in more detail what quantile regression is. (ii) The analysis procedure and variable selection were not clear to me. I read the relevant section several times and still did not come away with a precise understanding of the procedure. The authors need to take the time to fully explain their procedure. Maybe a schematic diagram would help (also see minor comments below).

Thank you for the positive feedback and general comment. We've tried to address each of the concerns you've brought up here through our responses to the individual points you raise below.

1. P14077 L13-14: It is my understanding that measures in Beijing were not taken because of a "particularly extreme events" but because of normal high pollution levels. Paris in the last few years might be a better example of extraordinary measures taken during high pollution events.

   Good point – the example has been modified following this suggestion:

   "In addition, particularly extreme events may hinder day-to-day activities, and require the implementation of drastic tactical air pollution control measures (e.g. the temporary banning of vehicles with even-numbered license plates from driving in Paris during the Spring of 2015)."

2. P14075 L5: "fans out". I think I know what you mean, but it would be better to explain more explicitly instead of using a term in quotes.

   This phrase has been removed, with the explicit definition and a relocated reference to the example in Figure 1a taking its place.

3. P14079 : As the paper is nominally about quantile regression more background on the methodology would be appropriate as it may not be generally known. The paper goes over this in a few sentences in the introduction and provides a nice example (Fig. 1) but it would make sense to educate the community in somewhat more detail.

   The introduction of quantile regression has been expanded and reorganized:

   "This situation is one common example of a distribution that might be better characterized through the use of more advanced statistical tools, such as quantile regression (Koenker and Bassett Jr, 1978). A semi-parametric estimator, quantile regression (QR) seeks to minimize the sum of a linear (rather than quadratic) cost function, making it less sensitive to outliers than OLS regression. Unweighted, this simple change produces a conditional median (or 50^{th} quantile regression), rather than the conditional mean of OLS regression. Applying appropriately chosen weights to the positive and negative residuals of this cost function then targets specific percentiles of the response, allowing for the quantification of sensitivity across nearly the entire response distribution. An example of this regression performed across a broad range of percentiles is shown in Figure 1b, including the 5^{th} quantile in black, the 50^{th} quantile in yellow, and the 95^{th} quantile in red."

4. P14080, L21: "all" is a strong word. I suggest you delete it.

   Done.

5. P14080, L8: "averages" – this seems to imply all variables are 3-hour averages. This does not seem consistent with some of the variable descriptions.

   Prior to the calculation of daily values (maximum, minimum, etc…) all of the original 3-hourly fields were scaled to hourly values using cubic splines, allowing for time zone normalization. This has been clarified in-text:

   "We use the 3-hourly NARR output to reconstruct hourly resolution diurnal cycles for each meteorological variable at each station through time series cubic splines and bilinear interpolation of the gridded fields to station latitudes and longitudes."

6. P14081: I believe the RPI as defined previously is actually equal to the ratio of vector/scale sums. This would make a low RPI (close to 0) indicative stagnant air masses.

   Following Levy et al 2009 we define RPI as this ratio of vector/scale sums subtracted from 1, so that higher values indicate more recirculation. This has been more clearly explained in-text:

   "To measure this effect we calculate a daily Recirculation Potential Index (RPI) from surface wind speeds based on the ratio between the vector sum magnitude (L) and scalar sum (S) of wind speeds over the previous 24 hours (Levy et al., 2009):

   $RPI = 1 - \left(\frac{L}{S}\right)$*."*

7. P14080: Variable generation. Some variables the authors averaged regionally (e.g., tke), some they do not. It would be appropriate to provide some rationalization for which variables are averaged regionally.

   Variables showing too many identical values (usually zero) were averaged regionally to introduce increased variability. This has been explained in-text:

   "In some cases regional means were included, primarily due to insufficient variability in individual cell values for that variable at some sites."

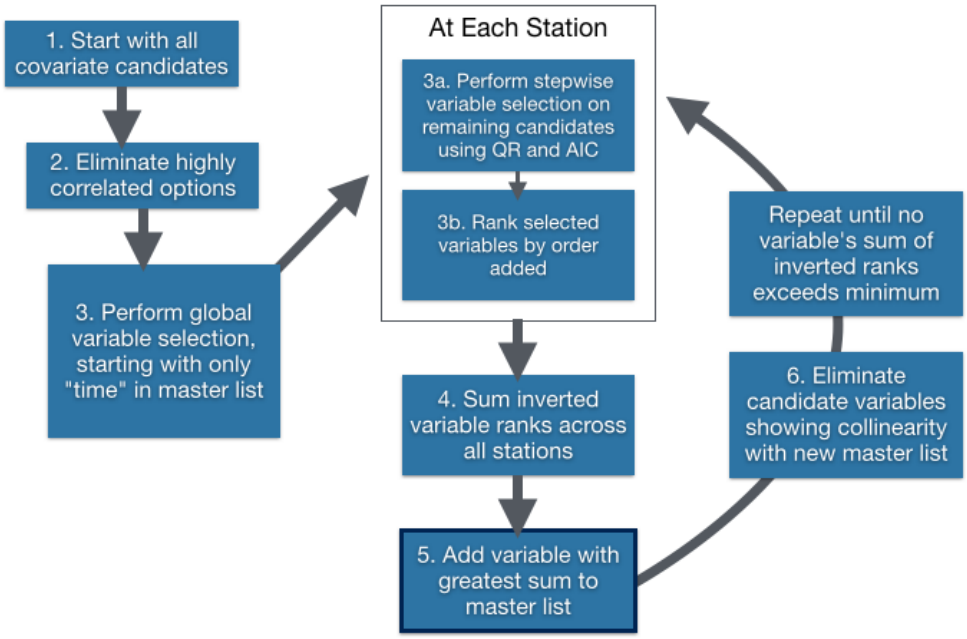
8. P14083, L17: "pollutant levels". I think you said this previously – but it might be worthwhile reminding the reader here which metrics you use for ozone and pm2.5 (e.g., daily average?)

   We use daily mean values for $PM_{2.5}$ and peak 8-hour average for O_3. Clarification text has been added to the section in question:

   "We use O_3 and $PM_{2.5}$ measurements from the US Environmental Protection Agency's (EPA) Air Quality System (AQS) network, including daily peak 8-hour average measurements of O_3 and daily mean $PM_{2.5}$ levels."

9. P14084: The procedure to select variables here is not altogether clear to me. I have read this section a number of times and am still unclear on the exact procedure. The authors should make sure it is clearly explained. Maybe a diagram would be helpful here?

   This section has been revised, including an accompanying flowchart (now Figure 3), to help clarify the procedure.



10. P14085: “summed inverse rank threshold of at least 2”. This is not clear to me. Perhaps when the procedure above is explained in more detail this will also become clear.

    We have attempted to make this “summed inverse rank” value clearer throughout section 2.4, and hope that this (along with the flowchart described above) helps to clarify the calculation and purpose of this metric.

    “We then rank the final set of included variables by order of selection, invert those ranks, and sum these inverted ranks over all 100 test stations (Figure 3, step 4). This sum represents an overall importance metric, and will be large for variables that either appear somewhat valuable at many stations, or that appear to be exceptionally valuable at just a few stations.”

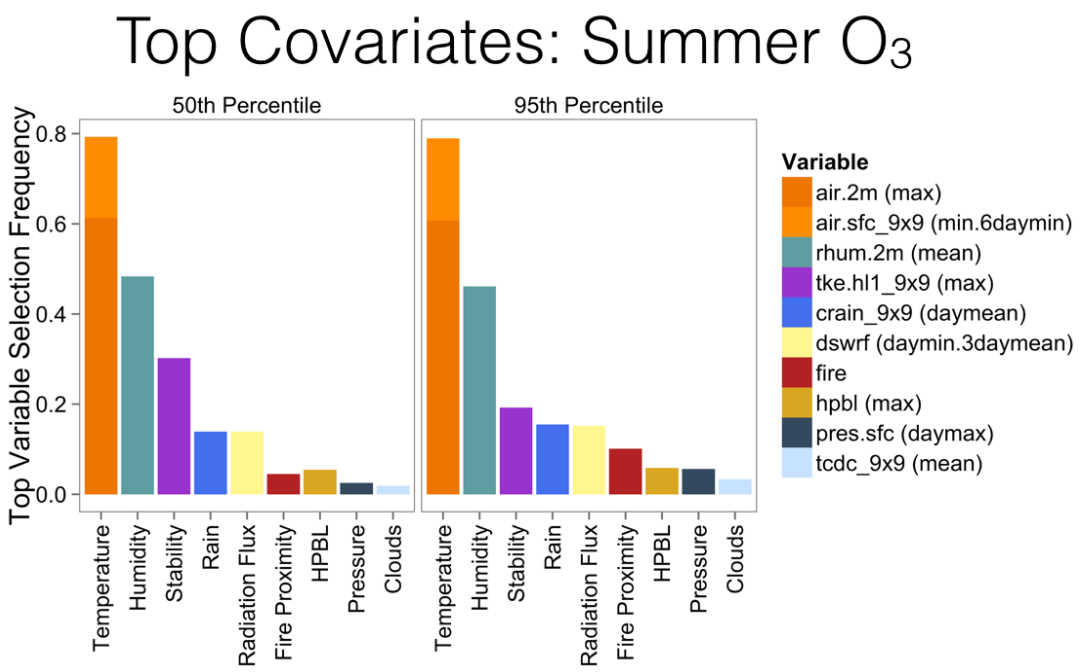
11. P14085, L22-23: “multivariate quantile regression”. Here I assume it is linear regression?

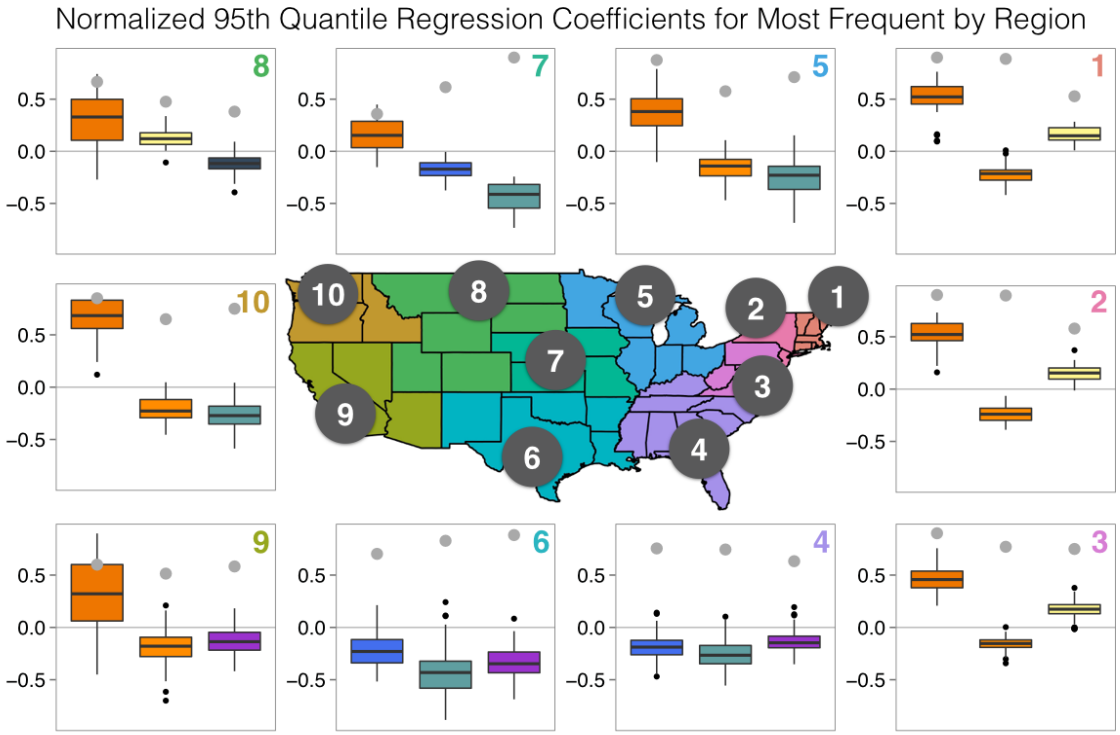
    Correct. All regressions performed here are linear. Clarifying text has been added:

    “Using these selected meteorological variables, we next perform linear multivariate quantile regression to identify sensitivities for percentiles from 2% to 98% at each station in the full set of AQS sites.”

12. P14086, L7: “frequency of appearance”. Actually the frequency of appearance is not shown, but the number of stations is shown. I would suggest showing the actual frequency would be a better metric.

These figures have been modified to show frequency rather than raw counts:





13. P14087, L7: “inverse correlation”. I assume by inverse the authors mean a negative correlation.

    Yes.

14. Fig 1. Please state the seasonality of the measurements and how many years are used in the figure caption.

    These data are from JJA measurements from 2004 through 2012, and this information has been added to the figure caption.

15. Table 1. Some of these variables names are not obvious and could be explained better with a footnote. What is categorical rain, best lifted index, the difference between apcp and prate, projected cloud cover? I suspect turbulent kinetic energy was generated in the boundary layer scheme – please clarify?

    There were several typographical errors in this table, leading to some confusion. The table has been corrected, which should hopefully clarify some of these questions. For more detailed information on specific NARR variables themselves, we would like to refer readers to Mesinger et al. 2006 and the websites linked therein.

    The revised variable descriptions include:

    apcp = accumulated total precipitation
    crain = binary precipitation flag
    tcdc = total column cloud cover

16. Fig. 3. I had to blow this figure up to make anything out of it. I would suggest making the panel sizes bigger and possibly separating into separate figures. I always find it hard to match colors precisely. In the lower panel in each figure the individual sensitivities should be specified (the names for these variables could probably be easily shortened). In addition either in the text or the figure caption it should be specified to what extent the correlations are significant.

    To a certain extent the legibility of this figure has been reduced by the ACPD formatting scheme, a problem which should be somewhat resolved in the larger ACP layout. However, we have also taken steps to increase the size of the inset figures to make them easier to interpret.

    Since these inset boxplots show the results of many regressions, rather than just one, there is no straightforward methodology for determining overall significance. However, we have added a grey dot to each box, representing the fraction of stations at which each variable showed significance at the 95% level. We hope that this, in combination with the normalized distributions themselves, are sufficient for assessing covariate significance.

    See final figure example above, in response to comment 12.

17. Fig 7. At what level are these slopes significant? I was struck by how similar the results were across the different quantiles. Even a 10% change seems rather small. Is this really significant? In general throughout the paper a number of correlations and regressions are made. The authors should really comment on the significance of these quantities.

The y-axis of this figure (now Figure 8) was mislabeled as a percentage, rather than a simple decimal value. Under the correct labeling, it should be clear that many of these coefficient ranges across quantiles are in fact of magnitude similar to the averages themselves, leading to drastically different sensitivities between the lowest and highest response percentiles. We hope that the relabeled y-axis (along with additional explanation in text) helps to indicate the importance of these differences:

"Quantifying the extent to which these differences in quantile sensitivities might impact the response distributions themselves is beyond the scope of this work, but the magnitudes of sensitivity differences relative to the mean sensitivities themselves suggest large differences between mean and extreme behavior. For example, the sensitivity change of summer O_3 to maximum air temperature is shown to be roughly equivalent to the mean sensitivity itself. Thus, a location showing a mean increase of 1 ppb O_3 per ºC could be expected to exhibit an increase of only 0.5 ppb O_3 per ºC at the 5^{th} percentile, but a much larger increase of 1.5 ppb O_3 per ºC at the 95^{th} percentile. This could clearly have important consequences for the resulting O_3 distribution, given increasing temperatures."