Response to Reviewers

We would like to thank the three reviewers for their positive and extremely constructive comments. We have implemented almost all of their suggestions and, as a result, we believe the readability of the paper has improved. We have included responses to individual comments below.

Review A

General comments.

1. Page 14, 1st full paragraph. An estimate of sampling error is computed by reconstruction of subsampled NCEP/NCAR reanalysis from 1990-2000. Subsampling is done to simulate 19th century sampling. However, the modes used (if I read this correctly) are computed using the 1951-2002 data. Because the modes span the reconstruction period, this reconstruction only partially simulates the skill that could be expected in the 19th century. Using dependent modes will inflate the skill even with subsampling. A much better estimate could be obtained using modes that are independent of the reconstructed period, say 1951-1989. I believe that this experiment should be repeated using independent modes for reconstruction.

We believe Reviewer A has raised and important point and so have repeated the experiment exactly as suggested. The text has been amended to reflect this change. As noted by the reviewer, the skill was slightly inflated by using dependent modes.

2. Page 24, top paragraph. Part of EMSLP is run without J86 fields, with J86 incorporated when it is available. Was a version of EMSLP run for the full period without any J86? It would be interesting to see how much difference adding J86 makes. Does it contain any independent information?

The J86 fields originate from synoptic charts and contain many thousand station observations over land per day; they include the equivalent of the surface input MSLP data that ERA-40 and NCEP-NCAR have been given since the 1940s. Post 1880 we only have 41 terrestrial stations with data (see Table 1) and so the J86 fields contain a large amount of independent information and, accordingly, are a critical part of the product. To omit these fields would have a serious impact.

The Reviewers comments indicate that the important of these fields has not been fully explained. We have revised the section describing the J86 fields (section 2.3.1) to better illustrate what information the J86 fields contain and how important they are to the EMSLP product.

Editorial suggestions.

1. Page 6, line 8, change to "... back to 1850 and forwards to 2003."

Changed, as suggested.

2. Page 13, 4th line from bottom, change to "... are prevented from producing noisy or spurious fields ..."

Changed, as suggested.

3. Page 18, 5th line from bottom, change to "... the ERA-40 product compared to EMSLP."

Changed, as suggested.

Also, does this imply that the ERA-40 is a better product?

Does it have more observations that the J86 fields? …

4. Page 24, end of top paragraph, change last sentence to "Much of central Europe is dominated by high pressure associated with these events."

Changed, as suggested.

Review B

General comments

My only suggestion: in Section 4, more attention should be paid to the lack of independence between ADVICE and EMULATE data sets when making comparisons; particularly in regards to the correlation maps in Figures 5 and 7.

We have changed the text to better reflect this issue.

Another suggestion: It is often difficult to distinguish between positive and negative contours in many of the black and white figures. Contour labels are also difficult to read. The authors should make an effort to improve these features.

We have re-plotted most of the figures, following this suggestion, particularly Figures 11 and 12 where the positive and negative contours were indeed difficult to read.

Section 4.5: Figures 11 and 12 show the presence of consistent local SLP anomalies corresponding to extreme temperature and rainfall events in the UK, and I don't doubt

the ability of the EMSLP data to represent historical circulation anomalies. But the evidence presented that these local anomalies were part of consistent, larger spatial scale SLP anomaly patterns is weak.

In this section we have aimed to demonstrate that EMSLP can be used to examine sub-monthly extreme events; most diagnostics presented thus far used monthly fields. We have focused on recent events in the context on historical events, drawing from the work of Burt (2004) and using the Chronology of British Hydrological Events and Daily Weather Records held in the Met Office archives. We believe it is important to demonstrate that EMSLP can resolve such events, as it is currently be used to characterize circulation patterns, a secondary component of the EMULATE project. This work will be submitted shortly. EMSLP is also being used to examine the circulation patterns associated with changes in North Atlantic storms, extending the work of Alexander et al (2005, GRL).

We would like the main focus of this paper to be the development of the dataset. We believe it is beyond the scope to provide an in depth analysis of these events, particularly as this work is shortly to be published. However, we do believe that reviewer B does make some pertinent comments and so we have tried to address these points in the text, to present a more coherent and tighter argument.

Figure 11 supposedly shows that "Anomalously high pressure centered over or near southern Scandinavia is a common feature in all 6 events", but the July 1868 case, at least, stretches this conjecture a bit. The fact that July 1868 was constructed from a 9 day average, while the 1881 case represents two days, suggests that an alternate criterion could be devised to present better examples of extreme large scale events. The authors' case is strengthened if the stronger point "Much of central Europe is dominated by high pressure" was emphasized as the reader was referred back to Fig. 10.

These dates were selected with reference to the historical heat wave events described in Burt (2004). Some of events were long, and we accept that it would be better to present events of similar length rather than 2 day and then 9 day averages. We have been able to refine our choice of dates, while still being consistent with Burt's analysis. For example the heat wave during 14-22nd July 1868 was comprised of two main episodes (15-16th and 21st-22nd); we have now plotted the first episode. Similarly, we have plotted subsets of the 1911, 1923 and 2003 events.

The same could be said of Figure 12. Data are stratified by wettest Octobers, but a single event is shown in each month. There is no mention in the text if the days selected represent significant flood events. I'm skeptical that any significance can be attached to the two anomaly patterns in the SLP fields, as presented, without additional statistical corroboration.

Although this was not explicitly stated in the text, each event is associated with a significant storm and flooding event. We have now altered the text to reflect this.

Our method was been to initially search for extreme Octobers using historical monthly precipitation records (no historical daily precipitation dataset is currently available pre 1931). However, as we wished to examine EMSLP's suitability to resolve extreme sub-monthly events, we then searched daily weather records, held at the UK Met Office, and the Chronology of British Hydrological Events to select the days, if any, within these extreme months which corresponded to flooding events.

Minor comments

1. pg. 6, 26: EMULATE web site has wrong address.

The correct address is now given

2. pg. 14: "highlighting regions were the technique"; were = where

Changed, as suggested.

3. Last line in Section 3.2 should be moved to 3.2.1

Changed, as suggested.

4. p. 16: how is "maximum permitted distance" defined?

This has now been defined in the text.

5. p. 16: More precise accounting of "majority of flagged grid points"

This has now been quantified in the text.

6. p. 18, Fig. 3: Jul, Aug, Sep undecipherable blobs over Middle East, not mentioned in text. Also, eliminating shading for $|slp'| < 0.5$ might help the reader interpret the figure.

The figure has now been replotted, as suggested by the reviewer. We have also commented upon the feature present in the Middle East.

7. Fig 7 is described, but the significance of its results are not commented on.

We have now added in text, expanding the discussion of this result, including the implications for the user.

8. CCs given on Figure 9, but not mentioned in text or in Figure caption.

These results are now mentioned in the text and in the figure caption.

9. In Conclusions, last point of second paragraph appears for the first time. It should also be included in Section 4.3

Changed, as suggested.

10. Figure 12: Is the label for 1882 (24-24 Oct) correct?

Changed (see General comments above).

Review C

General Comments

1. Abstract: (6th sentence) "…captures generally 80-90% of (add "daily" here) variability…"

Changed, as suggested.

2. In the introduction: It would help to briefly describe the nature of the problem introduced by lack of observations, especially in early years and at some locations with few ships or fixed stations. It seems to me that you are trying to obtain unbiased monthly data, even where the daily grids still may not be well defined for all locations.

The reviewer raises a pertinent point. It is very important to consider the impact of limited observations. We have now included a section in the introduction raising this issue.

3. Introduction: Please define "Central Europe" in the text. This term often includes countries like Poland and Hungary, but I do not think that you include them here.

We have included the main countries we consider to be in Central Europe in the introduction. We have also removed the word 'Central' from the description of the flooding in 'Central' Europe in 2002 and replaced it with the countries that were mainly affected, namely Germany, Austria, Romania and The Czech Republic.

4. Please define these "daily" analyses. Most SLP analyses that the reader sees are an analysis of pressure at a given GMT time each day. Often we may have 2 grids per day (or 4/day in the case of most reanalysis products). These "daily" analyses are different. First, they start with a daily average pressure for each observing station, and then there may be a higher than normal space averaging (true?) All ship observations in a 24-hour period are probably used (add text). Thus, changes within a day are removed, and the added smoothing removes some more variability.

The reviewer raises a very important point. The EMSLP fields are different to 'normal' synoptic charts, which represent a 'snap shot' at a particular time in the day. Our fields, which represent the mean pressure over a 24 hour period are quite different and are smoother. We have tried to make this clearer in the text, raising this point in section 2.0.

5. Page 7.7 (in 2.1.1): Convert station pressure to SLP. For stations at high elevation, people in the field (at least in US) usually used a more complicated conversion from stn P to SLP. It included elevation, the station, time of year. This probably helps to compensate for surface inversions, etc., that can be rather specific (and perhaps persistent) for the location of the station. If the station does report both stn P and SLP, do you recalculate or use their SLP?

If SLP is provided, we always use this rather than re-calculating the reduction to MSLP using the station pressure. This is because the original reductions were almost always done with the daily temperature value, which we did not always have access to. We hope this is now clearer in the text.

6. (2.2.1) First paragraph: After the first gridding step, I assume that most of the 1^o boxes might still be empty because no observation is close by. In the 2nd step that goes out about 3.5^o, it appears that an empty box gets filled in if there are any observations in the 7^o area. Otherwise it stays empty. Is this true? Also, if there is only one observation in a region, then all of the 1^o boxes within 3.5^o of the obs will be filled with the value of this one observation (true?) Please add a little to the text to clarify this.

This procedure is a little confusing and we acknowledge more information is required in the text to clarify it. Within the 7 degree area, there are 42 grid boxes (target included). If the target box value is missing, but one of the surrounding 41 1^o x1^o boxes in the 7 degree area contain an observation, then the target value is replaced with this value. If all grid box values in the 7 degree area contain data, including the target, then the target value will be replaced with the median of all 42 grid box values.

Additional text has been added to hopefully clarify this procedure.

7. (2.4) page 12, 2nd paragraph: "…jumps in adjustments…binomial filter with 7 terms was applied…" Please add a comment like this if true: This process gives a smooth daily adjustment series, but almost the fully daily variability of the station data is still preserved in the analyses.

The smoothed adjustments applied serve to correct the series for any potential heterogeneities; however the daily variability of the station data is almost fully preserved, as the reviewer indicates. We have changed the text to better describe this, as suggested by the reviewer. SHOULD WE INCLUDE A PLOT ALSO?

8. (3.2.1) Procedure. Last sentence, add "daily": "…to yield the smoothed (daily) climatology."

Changed, as suggested.

9. (3.2.1) The binomial filter (21 terms). Please give a rough feeling for the response of this filter (does it remove noise under about 2 weeks, or 4 weeks, etc.)

This filter removes noise under 15 days (about 2 weeks). Text has been added to clarify the response of the filter.

10. The figures 2.9 that show monthly values are encouraging. I am still having some trouble in visualizing what is happening with the daily grids. If one of your good observing stations is within 2^o or 3^o of a 5^o grid point, then will RMS daily data for the grid point be rather close to the RMS for the station? Have you made this check for a few stations and a few sample months? If a station is nearly at a grid point, I wonder if the day by day observed values are fairly well preserved in the output.

The reviewer raises an interesting point. We have performed grid box comparisons with the original station and similar results to that seen in Figure 8 are evident. The gridded data is smoother than the original daily series. We believe this is probably a result of the application of RSOI, which tends to produce dampened fields, but may also be a result of the smoothing applied in the marine gridding procedure for coastal grid points. We have included this result in section 4.3 where Figure 8 is discussed. We acknowledge that this feature is one of the greatest limitations of the dataset. We hope that by highlighting this issue to the reader/potential they are better informed as to what analyses EMSLP is appropriate for.

11. Fig 10: In the caption, please call it "daily variability."

Changed, as suggested.