Reply to Mathias Raschke comments 29th January 2016

I thank Mathias Raschke for spending time reading the manuscript and providing comments which have improved the content. More specific replies to all comments are now given.

1. *It would be helpful if the sampling error (called standard error or estimation error in statistics) and the observational error would be briefly defined/explained in section 3 (Analysis method).*

Done – Section 3 has been expanded to include description of analysis of sampling and observational uncertainties.

2. *The estimation method for the parameters (cluster coefficient and/or expectation and variance) should be described in section 3. A corresponding statistical reference would be also helpful.*

Done – the estimation method has been added to text in Section 3.

3. *The analysis results are compared with the findings of recent publications. However, the dispersion statistics is applied in many publications and not the cluster coefficient methodology. That is why any comparison would need more details and explanations on the differences of the applied methods.*

An explanation is added to the manuscript on how comparisons can be made to previously published results. In brief, if storm rate information is published in earlier studies, then CC values can be derived from dispersion values using Equation (3).

4. *The result of an increasing cluster coefficient by an increase in the threshold/return period is in contrast to the results of Raschke (2015) for winter storms in Germany wherein the cluster coefficient does not depend on the threshold/return period. This could be discussed in the current paper.*

The result in Raschke (2015) using Karremann et al. (2014) climate model data, is added to the discussion in the revised manuscript.

5. *The modelling of Raschke (2015) is based on winter storm data, which have been extracted from climate model simulations (for Germany, details see Karremann et al. (2014), including supplemental). These data indicate a constant cluster coefficient for different thresholds/return periods (difference between the point estimations are not statistically significant). If in reality a cluster coefficient exists, which increases with an increase in the threshold/return period of e.g. a winter storm index (magnitude or the like), than the climate model approximations are incorrect, which need to be discussed in detail.*

There are many potential causes of different clustering behaviour in climate models, and analysis of climate model errors goes beyond the scope of observed clustering in this study. However, I agree this is an important point, and a statement has been added that the potential benefits of longer GCM integrations – providing smaller sampling errors – may not be available currently, due to differences in clustering behaviour between coupled climate models and observed.

6. *The samples with a large history are not complete. Figure 3 illustrates this fact; older time periods include much less storm events than younger periods. The problem of completeness is well known in earthquake magnitude statistics (e.g., Hakimhashemi and Gruenthal, 2012; Kijko, 2012). The incompleteness mainly concerns the smaller storms and could cause a bias. The filtering by incomplete documentation and inhomogeneous perception might be very complex. The documentation of three storms with short time delays might be more likely than the documentation of one storm event.*

Incompleteness is a general problem with longer historical datasets, and the datasets and methods were adapted to alleviate this problem. The length of some datasets was shortened to more recent periods (UK-Lamb, FR-Garnier and CZ-Brazdil) and the analysis focussed on the more severe damaging storms which are well represented in EU documentary records. The CZ dataset shown in Figure 3 was the most problematic and is the subject of more analysis than any other dataset: a discussion of its incompleteness is given in Section 2, and impacts of this problem upon CC values are discussed in Section 5, and evaluated in Tables 2 and 3 of the revised manuscript.

7. *How is the homogeneity of the event definition with respect to the event duration considered? For example, while one definition might identify two events in short succession, another definition might detect only one event.*

There is evidence that EU documentary records resolve independent meteorological events for the past few centuries. For example, the French dataset resolves the three major storms hitting France in a 5-day period, between 14^{th} and 18^{th} January 1739, and the CZ dataset resolves two storms one week apart in December 1514, and four days apart in March 1537. Europe is rich in documentary evidence of weather and other events over the past few centuries.

8. *If the cluster coefficient really increases with an increase in the threshold/return period than a stochastic model/process should be formulated for this phenomenon in the same way as e.g. done by Raschke (2015); for the case of constant cluster coefficient (key word: thinning process). I do not expect that the author formulates such a stochastic model but the gap in the theory should be mentioned.*

This gap in theoretical modelling is mentioned in the revised text.

9. *The samples could get a name or number in section 2 (data) which could be also used in section 4. This way the reading would be much easier.*

A new Table 1 has been added with a summary description and a brief name for each storm dataset. These brief names are used throughout the new version of manuscript.

10. *It would be also helpful if the analysed data are published in a supplemental to ensure simple reproducibility.*

Most datasets are referenced in Section 2. Some datasets, such as Garnier's French storm list, are available for a fee and cannot be provided by the author.