Reply to Joaquim Pinto review   3rd February 2016

I thank Joaquim Pinto for spending time reviewing this manuscript, and providing comments that improve its content. More specific replies to all comments are now given.

1. *Abstract, Page 7458, lines 14+15 and 19-21, and elsewhere: The statement may be true based on the results analysed here but the author actually only looks in detail at RPs up to 10 years – which is fine by itself, given the data. Still, care should be taken not to generalise this result for "clustering increases with intensity for all return periods". I would suggest adding more precise information and staying close to the actual results, e.g., "stronger clustering is found up to a certain return period (10 years)" or ". . .. within the range 1-10 years return period". It is unclear how clustering for very long return periods actually looks like (large uncertainties), and results by Karremann et al. (2014a) based on GCM data do show that clustering may in some cases actually reach a stabilisation level or even decrease for long return periods. It is unclear why this may happen, it might be simply to do with the length of the datasets – the rarer the events, the more "random" their occurrence may be given the limited sample – or it might have physical reasons - the jet cannot intensify infinitely or remain quasistationary for months in a row - but we simply do not know.*

More precise information on analysed RPs has been added to the Abstract and Summary.

2. *Introduction, Page 7459: The introduction about clustering is generally fine but very short. In my opinion, it would be helpful to shortly discuss also these three recent papers*

a) Blender et al. (2015) – which provide a different view of clustering (based on the Fractional Poisson processes)
b) Pinto et al. (2014) – who provided a "modern" synoptic and dynamic view of the phenomena, thus explaining the physical reasons for the clustering of windstorm losses
c) Hunter et al. (2015) – closer look at frequency intensity dependence, role of teleconnections.

The Introduction serves to focus attention onto observed clustering, which is the subject of the manuscript. The Introduction has been revised to include insights into observed clustering in Pinto et al. (2014) and Hunter et al. (2015). The replication of observed clustering using models such as the FPP in Blender et al. (2015) is very important, but not the focus of the manuscript. Overall, the revised Introduction has 30% more words than the original version.

3. *Results and discussion, page 7466, lines 3-5. There is a bit of confusion here regarding cyclone based results vs (potential) loss based results. The given statement is true for cyclone data as analysed in Mailier et al. (2006), Vitolo et al., (2009) or Pinto et al. (2013). However, this is not necessarily true for potential losses and longer return periods– see Karremann et al. (2014a, 2014b) and comment #1 above. Note also that the latter papers analysed much longer return periods than the former papers, and thus the slightly different conclusion is not necessarily a contradiction. As mentioned in #1, clustering for long return periods is uncertain and may actually decrease. I would suggest writing here two sentences, one focussing on cyclones and one focussing on losses, and shortly discuss the differences. See also page 7467 lines 20-22 and elsewhere.*

Some extra notes are relevant: (i) The last data point in Figure 6 of Vitolo et al (2009) represents 10 most extreme storms in 53 winters, which is equivalent to the RL5 used in Karremann et al. (2014a); (ii) the potential loss metric used in Karremann et al. (2014a) and Raschke (2015) is similar to the damage metrics used in the new study, and the reduced clustering strength for longer return periods in the ECHAM5 climate model is different from behaviour in a wide variety of historical storm datasets.

I agree with the main point that different measures of severity confound comparison of clustering-severity relations between studies. The discussion of earlier results is expanded in the revised manuscript to describe the uncertainty from this issue.

4. *General: I believe that the description of "southern countries (off the main storm track)" (e.g., 7458, line 15, and elsewhere) is quite misleading, because the author is actually talking about countries in Central Europe and not the Mediterranean area. I suggest changing the denomination to "Central Europe" or similar. (for "Northern countries" it is clear).*

The text in Abstract is changed to "regions off the main storm track in central Europe and France", and similar changes are made in the Summary.

5. *Results, page 7469, lines 28-29. Well, this is not unexpected, as similar results were obtained in Karremann et al (2014a) for the longer 505y PRE simulation – clustering can apparently change over time (in longer time scales), hence the "weaker clustering" in general if compared to a period of comparatively high clustering*.

Lines 28-29 of page 7469 discuss the stronger clustering of more severe storms in the CZ-Brazdil-500 dataset, whereas the multi-decadal variability of the PRE simulation is referring to a different feature (Karremann et al. (2014a) use multi-decadal variability as an explanation of why the PRE simulation has a longer RP for three storms of RP1 severity compared to other simulations).

6. *Summary, page 7471, lines 25ff: the sentence is very long and hard to explain, I would suggest writing two sentences*.

It is split into two sentences in revised text.

7. *Summary, page 7472, lines 13-16: another possibility to reduce uncertainties is to use GCM data as performed by Karremann et al (2014a, 2014b), this should be clearly stated as a valid alternative.*

A new sentence is added at the end of the Summary on how climate models with validated clustering behaviour would be beneficial.