Reviewer 1:

I think that the authors have addressed only part of my previous concerns, but the core of the problems that I find have not been solved. Specially, I am not convinced by the application of the clustering technique to the indices generated by the authors. The indices are semi quantitative, continuous and nonlinear variables. I think that, under these conditions, the use of the cluster techniques is questionable. In the new version I do not find convincing arguments justifying this application. The inconsistency of the results reinforces this. Now, the authors emphasize the limitations in the DI Mountain cluster, which is Ok. However, I still think that it is a purely statistical artifact. I have provided different reasons I my previous reviews. On top of them, I provide an additional one. Figure 7 shows the spatial patterns of extreme drought years. 1685 and 1701 (bottom right panels of the figure) show that in these 'extreme' years of the mountain cluster, Teruel shows an index value of 0. I think this is not acceptable and is another proof of the lack of physical foundation of this cluster analysis in an area of high precipitation and temperature variability, as acknowledged by the authors in lines 647-658.

The new version of the discussion is focused in justifying that rogation records are good proxies for droughts. This is has been proven in the literature and I am not questioning at all their value as local indicators of droughts. My point is on how the cluster is applied and interpreted in that specific case, not on the validity of rogations.

Panel a) of figure 8 is from my viewpoint another proof of lack of consistency. How can the authors explain that they find in the Mediterranean cluster a significant signal before the volcanic eruptions? (year -1) Do they have an explanation for this?. I think that the only robust result is the impact of the Tambora eruption in panel b, a result previously reported in papers co-authored by some of the authors.

So, I do not think the paper is acceptable for publication.

Additional comments

- The authors claim that they have computed the SPEI index since 1787. This index requires instrumental data to compute the AED. Which data have they used? This should be explained in the text, since, as they claim, the only long instrumental series in the region is the Barcelona temperature series. Is this a combination of instrumental and proxy records?. If so, this must be carefully described in the text.
- The language needs further revision. I asked in my previous review but it does not seem to have been taken into account too seriously.
- Some references are missing or wrong. Examples:

line 168 AEMET 2012 missing

line 744-745 I have not been able to find Dominguez-Castro and Garcia Herrera GRL 2016

The reference to Garcia-Ruiz 2001 about climate change impacts in the Mediterranean should be updated. See for example

The climate of the Mediterranean region: from the past to the future

2012 , Elsevier Insights, 592pp, ISBN: 978-0-12-416042-2 , Ed. Lionello P.

Mediterranean Climate Variability

2006, Elsevier, Amsterdam, ISBN: 0-444-52170-4, 438 pp, Eds: Lionello P., P. Malanotte-Rizzoli and R. Boscolo

- DIMED appears in l 272, while its meaning is explained in line 324

Reviewer 2 (new):

The manuscript is an interesting work on historical droughts in Northeast Iberian Peninsula, but in my opinion it needs some revisions before publishing:
1) In the literature you can find different definitions of the drought concept (atmospheric, meteorological, hydrological, agricultural), depending on the physical variable studied (relative humidity, rainfall, other elements of the hydrological cycle), and the duration of the event (days, months, seasons). I understand that here the authors are studying agricultural droughts, due to the origin of the data (rogations linked to agricultural production). In any case, it would be important to precise this point.

2) The nature of rogation ceremonies must be explained with more detail. For instance, is it possible to find a 'preventive rogation', that is, a ceremony organized before the event occur? In this sense, the date of the rogation is an important information. It may be the case that a dry winter provoked the rogation, but timely spring rainfalls yielded a good harvest. In that case, can we speak on 'drought'? In relation to previous comment, perhaps here we could speak on dry winter (meteorological drought), but not on 'agricultural drought', and, in consequence, this event is not comparable with other characterized by the water deficit during an entire year. The 'annual' index (from December to August) may mask important intra-annual fluctuations, in my opinion it is preferable to divide the information into seasonal indices, following the different phases of the plant growing, from seed (autumn) to harvest (summer). In addition, all rogations are linked to cereal production? Other plants (fruits, olive trees) have different climatic limitations, and it would be possible that a single meteorological event (for instance, dry spring) was harmful for a specific plant, but not for another (for instance, the barley is more tolerant to drought than wheat).

3) I have doubts on the classification of the rogations (lines 205-207, Table 2). Were the ceremonies the same in all the cities and during the whole time period, from 1650 to 1899? Severity indices are based on the type of ceremony, but is it a reliable criteria? In the discussion (lines 404-411), authors say that 'an index of level 2 does not necessarily imply that a drought was twice as intense as a drought classified as level 1, nor that the change in the intensity of droughts from level 1 to level 2 or from level 2 to level 3 has to be necessarily equivalent'. In that case, how must we interpret these indices? In my opinion, these indices only specify the nature of the ceremonies organized as response to natural hazards, but do not inform on the severity of the climatic event. In consequence, what is their utility from a climatic point of view? In my opinion, the binomial distribution (occurrence or not) is the more appropriate statistical approach to the treatment of this information.

4) Clustering is an appropriate tool to classify and group local series into regional series. There are very different clustering algorithms, hierarchical and not hierarchical. Why have you used Ward method with Euclidean distance, and not, for example, the non-hierarchical k-means, or other methods as the principal component analysis? Results of clustering must yield groups more or less homogeneous, but the chosen number of clusters is normally arbitrary. Why do you distinguish between Mediterranean and Ebro Valley group (dendrogram, Figure 3), if, as you say (lines 450-451) 'the high correlation between DIEV and DIMED is suggesting similar climatic characteristics'?

5) Validation of the regional drought indices is made using the overlapping period 1786-1899 between documentary and instrumental data. But, as you say in the discussion (lines 390-392) 'the apparent low frequency of rogations in the 19th century could be explained by a combination of political instability, and the loss of religiosity and historical documents'. I would add changes in the socioeconomic structures, organization of the cereal production, agricultural techniques, etc. In consequence, this period is not valid to calibrate and/or validate the rogation series in previous centuries. The cultural background, economic organization and technology of the 19th century was not the same that in previous centuries, and calibrations established for 19th century are not applicable to 17th century! In fact, you do not use this calibration (regression in Figure 5) to interpret previous data, only to validate the index during the 19th century. Besides, this analysis is only made for DIMED

(Barcelona), and not for the other points in the studied area. I suggest to remove this analysis (and the Figure 5).

6) Superposed epoch analysis (SEA). Although you give a reference, a brief explanation on the basis and procedures of this method would be important.

7) Minor questions:

Line 301:'The cluster analysis (CA, see methods) using the DI of the 13 locations for the period of 1650-1899 AD revealed three significantly coherent areas…' Erratum, I suppose, clustering is made using the period 1650-1770, common to all the stations, although the classification obtained is after applied to the complete period until 1899.
Lines 373-375: 'However, two years after the Tambora eruption in April 1815, there was a significant ($p<0.05$) increase in the three drought indices…' However? The time life of volcanic aerosols in the atmosphere is around one to two years. The Tambora aerosols caused a radiative forcing of the global climate system of about 5.6W/m2 for one to two years following the eruption (Brönnimann and Krämer, 2016). In consequence, this increase in the drought indices may be caused not by the volcanic eruption, but by the return to 'normal' conditions (or not forced climate variability).
Lines 417-418: 'the local series are compared with the regional reference series as a basic element of quality control'. But, if the regional series is obtained from the average of local series, here we have a circularity problem.
Line 432: 'the local series are separated by tens or hundreds of kilometers'. If you speak on meteorological droughts, this is not a problem, because the dynamical conditions provoking dry conditions are associated to the predominance of anticyclonic conditions, and the spatial extension of an anticyclone may be much greater. Again, we are speaking on the drought concept. Meteorological, hydrological, agricultural?
Figure 5. Significance level must be added in the figures. D), E), F), correlation is the Pearson correlation coefficient?
Figure 7.The legend is arbitrary, why do you distinguish between 2.1-2.6 and 2.7-3 DI values?
Reference
Brönnimann S, Krämer D. 2016. Tambora and the 'Year Without a Summer' of 1816. A perspective on Earth and Human Systems Science. Geographica Bernensia G90, 48 pp, doi:10.4480/GB2016.690.01

Reviewer 3 (new)

This study compiles and quantitatively analyses drought indices derived from documentary data on rogation ceremonies in northeastern Spain to further insights into historical droughts but also to understand the role of volcanic forcing on past event.

The most importantly contribution of the paper is that it provides a very interesting insight into both the strengths and weaknesses of documentary sources, especially approaches that seek to derive quantitative estimates from qualitative data. It is my view that the paper should be accepted following minor revision. I suggest some points below that that the authors need to address. Most of my comments seek clarity and explanation. I also request the authors to give the paper a thorough edit. There are instances of misspelling or improper English scattered throughout the paper. This deserves considerable attention. There are also instances of long tracts of text that make it hard for the reader to track key points. Please use paragraphs more effectively to deal with this. There are also very long sentences at times (e.g. end of introduction) that need to be broken up.

I would like to know more about the documentary sources consulted. We don't get to know much about this aspect despite so much interpretation later depending on these sources. In line with comments from previous reviewers the original sources and their consultation/analysis needs to be given greater attention in the paper.

The key methodological steps in the paper are as follows:
1) Development of a semi-quantitative series from qualitative data derived from documentary sources on rogation ceremonies. This is done using an established technique. I have no concerns to note.
2) Clustering of series to develop regional drought indices
This again seems to follow best practice. Importantly the analysis is not entirely statistically based and physical reasoning around the derived clusters is given. This is important as such techniques are somewhat subjective and the authors are transparent in their choices. The limitations of the approach are clearly articulated in the discussion.
3) Validation of the resultant series against instrumental records.
4) The performance of an epoch analysis to detect volcanic influence on historical droughts. This section is given the least attention in methods and most prominence in the abstract. I think that the authors need to explain this approach in more and sufficient detail to allow reader fully understand what they are doing here. A short paragraph should suffice. Why this method and what are the assumptions/strengths/weaknesses around the approach and desired attribution statements.

I do not have local knowledge of the region but I find the results interesting, especially for the mountainous region. It seems the other two regions show similar results that are coherent with findings from previous studies. Indeed in discussion this aspect of the coherence of results needs to be moved further up. This is important information to have before getting into the limitations.

I find the weaker results from the mountain region interesting. Some effort is expended on trying to explain why this difference and at times the authors get into attributing different processes. First, is this something seen in recent times when we have measurements? Second, given that the performance of a rogation was done based in part on the wishes of agricultural institutions, is there a risk that mountainous regions would have weaker political power in influencing rogations. Therefore, the lack of intense droughts in mountainous region or the disagreement with other regions, may be due to its weaker economic importance rather than anything to do with drought directly? I think the authors need to mention this possibility.

I also think that the authors could make more of the issues they run into for the mountainous region as a case of the challenges of using documentary sources. This needs to be mentioned in the abstract as its lessons are important for other studies.

The authors rightly state that the drought index for the mountainous region should be treated with extreme caution.

Need to be careful of tense used in abstract.