Dear referee #1,

Many thanks for the positive evaluation of our work. We highly appreciate your explicit comments and suggestions which have helped us improve the original manuscript. Please find our response to your questions and suggestions below (we first repeat the referee's comment in *italic* and then provide our answer)

General comment

"However, a brief assessment on the main uncertainties would be beneficial to understand the reliability of main results."

Thank you for raising the issue of uncertainties in the database. In the discussion section, all results are explicitly discussed in relation to the uncertainties in EDII data. We have amended the reference to Stahl et al 2016 (doi:10.5194/nhess-16-801-2016) in the data section as this publication gives a detailed account of the uncertainties and added more text on the potential uncertainty of different information sources and the need to consider uncertainty in the discussion at the end of Sect 2.1.

Particular comments

1. Regarding the indicators and factors used, while you state which indicators you selected, I don't fully understand how you did the regional and temporal aggregation of indicators. A brief quantitative assessment in this respect would be beneficial.

We revised the manuscript by distinguishing more clearly between:

a) the derivation of the underlying database of monthly drought hazard indices such as SPI and SPEI (at various accumulation periods) as well as the vulnerability factors for each NUTS-combo region. Aggregation procedures are now explained better in Sect. 2.2 and 2.3.

b) the subsequent use of these variables as predictor variables in the models. This selection involved the drawing of a sample (pooling) of NUTS regions within the macro region for which the model is fitted (no more averaging). This is now explained better in the beginning of the method Sect. 3.x

2. Please briefly describe the drought indicator aggregation procedure for severity calculation.

The definition of drought severity for SPI, SPEI, ΔpF, ΔfAPAR is inspired by the definition of McKee(1993) who assigned standard deviations from normal to hazard severity levels for SPI, with a threshold of '1' corresponding to a return period of 6.3 years, classified as moderate drought, and '-2' as extreme drought conditions. We added this explanation to the last paragraph of the methods section, i.e. where the severity is introduced for use in the hazard map scenarios.

2. Baselines used to calculate indicators

The requested information was added to the belonging indices descriptions in chapter 2.2

4. ***Fully named impact categories***

Impact category names have been changed accordingly throughout the text.

5. Page 12519 lines 16-17: It's not clear what the authors mean with multiple hazard predictor? What hazards?

Changed to "multiple drought indices".

6. Objectives are stated in the paragraph starting at line 17 of page 12519 and somehow repeated in the last paragraph of page 12520. To avoid recurrences, I would recommend to condense these two paragraphs in one, preferably at the end of the introduction.

Many thanks for this suggestion, we merged the starting line with the formulation at the end of paragraph 1 as follows:

"In this study we expand the method of Blauhut et al. (2015a) into a 'hybrid' approach, which implies the consideration of vulnerability factors into the probabilistic impact prediction. The approach builds on earlier work developed for the agricultural sector (Zhang et al. 2011; Ahmed and Elagib 2014; Han et al. 2015; Yin et al. 2014) and an European assessment by De Stefano et al. (2015), who considered several physical and socio-economic factors to calculate sensitivity and adaptive capacity, and used impact information collected in the EDII to estimate exposure. More specifically, the hybrid approach aims to: "

Page 12522 Lines 25-29: "Drought impact reports stem from various sources and are assigned with a certain level of reliability, decreasing by its enumeration-rank: academic work, governmental reports and documents, reports, media and webpages and other sources." How this reliability was accounted for? Different weights were assumed for each rank?

No, unfortunately we had no reliable basis to decide on such a weighting. This is something future work will have to address. At the end of chapter 2.1 we added the following sentence for further explanations. "Nevertheless, uncertainties due to the nature of the impact data need to be discussed and considered in the interpretation of any study that are based on this or similar sources of data."

Page 12523 Section 2.2: Different levels of severity are shown as representatives of drought severity. However there is little explain on how drought severity was defined, how it was aggregated in time and space, etc. As far as I understood there is one annual value that defines drought, usually there are single months as stated in Figure 3. These months are related to some drought characteristic, like the peak of the drought, or are just the values of SPEI regionally aggregated? Were other metrics like total area, absolute minimum, etc. tested? A better description on this key issues would be beneficial to the understanding on how the functions were built.

Grid cell values were averaged for all NUTS regions of the NUTS-Combo level. Different possibilities (e.g. the regional Minimum of SPEI instead of the average) were tested in an early stage of this work, but made little difference and hence the mean was chosen. For the modeling, the NUTS region values make up the pooled sample. (Also see comments above on revisions of these explanations. In time, indices for different months and different accumulation periods (SPI-n's) were tested in the stepwise model building. Indeed that means that possibly a sub-annual characteristic predicts the likelihood of impact of which we don't know when in the year it happened. These necessary simplifications due to the lack of more accurate pan-European time stamping of the impacts are discussed.

Page 12523 line 16: "indictors" should read as "indicators"

Changed accordingly.

Page 12523 lines 22-25: In this form this affirmation is not completely true. Neither in this work (you stated in page 12524 that a transformation using Gamma distribution was used) nor in McKee 1993, was the SPI computed as the difference from the mean divided by the SD. This is true for normally distributed variables, but usually this is not the case for precipitation.

Thanks for pointing out this mistake. Changed it to:

"the SPI is the transformation of the precipitation time series into a standardised normal distribution" (Lloyd-Hughes and Saunders 2002), and is commonly used to estimate wet or dry conditions based on long term records of precipitation."

Page 12524, line 11: EOBS-9 was used for a specific reason? Several updates since this version were made to this dataset that now integrates more data and with better quality of information. Do you think that using a different or updated dataset results could change largely?

The choice for the EOBS-9 dataset was made at the beginning of the project and was used throughout by different partners in different project studies. A comparison to the older EOBS-8 dataset at the time showed marginal differences in results. As both the SPI and SPEI are relative measures, they are likely robust to systematic changes in the dataset that do not change the order of the values. As the baseline period would not change with an update we therefore do not expect substantial changes to more recent releases.

Page 12526 Section 2.3: I'm not sure if Table 2 is completely necessary. I found a bit confusing that at the beginning 16 vulnerability factors are presented but then in Table 3, 69 factors were selected. Also the header of the Table 2 is not clear, do exposure should be after the line as it is a component? Are drought characteristics the definition of exposure? If you decide to keep Table 2 factors in Table 3 should be associated to any category in Table 2.

Following your advice we removed Table 2

Page 12527 lines 5-18: It's not clear how vulnerability factors with different time steps were used? There are different maps of vulnerability or the last available data was used?

We added the following text at the end of section 2.3

"Vulnerability data for which multiple timesteps were not available, the most recent information for the entire period of investigation was applied.. Vulnerability data with multiple timesteps was assigned to the corresponding year, and preceding years up to the next time step available (e.g. available timesteps 1976, 1990, 2003, → 1970-1976: 1976; 1977-1990:1990; 1991-2012: 2003)."

Section 3: Is there any other function that could be suitable for fitting drought LIO?

We did not explicitly consider or test other link functions given our priority to assess the role of the variables in this paper. However, other studies have considered different approaches to link textual impacts with indices. Thus we added the following line to the introduction:

" Stagge et al. (2015b) considered variations of the logistic regression and expanded the approach to include multiple hazard predictors. Bachmair et al. (2015a) applied regression tree and correlation approaches to link impact number and occurrence with a range of indices. Both studies relied on a rather high temporal resolution of reported impact occurrence, and hence considered only a few regions with particularly good data coverage."

Page 12529 lines 10-14: Is unclear how SPEI data was included in the model. I guess that monthly data for each aggregation period was used. Please be more specific.

Clarified this in the revised version under section 2.2.

Page 12532 line 3-6: did the authors checked D(fapar) for the growing season instead of the annual average? This could improve the suitability of the fapar as a predictor.

We agree. We did not analyse the growing season particularly, but found in accompanying studies that monthly aggregations actually do show a better suitability for the timeframe of growing season. However, since the growing season length varies with region across Europe, a seasonal investigating would also require smaller-scale regional models, which at this point was beyond the scope of the study. Thus, we added the following to section 5.1:

"Concurring e.g. with Shakun et al. (2014), fAPAR proved its usage as drought index for vegetation-process-related impact categories, for the growing season particularly. Thus, of the use of a fAPAR based seasonal index in further studies appears promising.."

Page 12532 lines 9-11: SPEI performs better because it has more data availability than the CDI or because it better represent the conditions?

Thank you for pointing out that this should be made clearer. A detailed test of SPEI for only the short period of 2001-12 (as EDO products) (see supplement Table S5) revealed that the overall better performance is not only due to the longer time series. Even though it cannot be neglected that a longer time series has a strong positive impact on model performance. Accordingly we added the following lines to

4.2.: "To estimate the influence of longer time series for model input, Table S5 shows model performance for SPEI applied for the shorter time period 2001-2012. Resultant model performance follow similar performance pattern, but less strong, as for longer time series."

And to 5.1.:" Generally, the tests showed that the hazard-impact-linkage will benefit from longer time series and thus a wider range of drought conditions. Furthermore, it was found that the overall better performance of SPI and SPEI to JRC hazard indices was not due to the differences in time series length."

Page 12537 lines 24-26: "For an application like this one, this can be interpreted in two ways: prior standardisation, composition and weighting appears unnecessary or a composite of factors may well replace the many individual ones." This sentence seems a bit contradictory in the present form, please consider rephrasing.

We agree and rephrased to: “For an application like in our study, this can be interpreted as prior standardisation, composition and weighting of vulnerability factors appears unnecessary.”

Page 12539 lines 2-7: “Hence, the most relevant SPEI may differ in month selected. This corresponds to different aggregation times, e.g. detected by Lei et al. (2011) for Northern China and Potopováa et al. (2015) for Czech Republic for 5 maize. Furthermore, some combinations of selected hazard indicators may have been affected by the criterion of variable independence employed.” This two sentences are not completely clear to me. Please consider rephrasing them.

We agree and rephrased this part to:

“Hence, the most relevant SPEI for each region may differ in accumulation time and month selected. This corresponds e.g. to Lei et al. (2011) and Potopováa et al. (2015) who detected different optimal accumulation times of SPEI for maize productivity for Northern China and Czech Republic. A reason for the selection of more unexpected combination of SPEI (e.g. SPEI-6 of August was selected together with SPEI-1 in December for ‘Agriculture and Livestock Farming’ in Southeastern Europe) might be due to the criterion of variable independence employed.

Figure 1. Please review the caption of this figure. It could be shortened as: “Number of annual NUTS-combo scale impacts reported and archived in the European Drought Impact report Inventory (EDII) by European macro region (left panel) and by NUTS-combo region (right panel).”

We changed the caption accordingly.

Referee #2

Many thanks for the positive evaluation of our work. We highly appreciate your comments and suggestions which have helped us improve the original manuscript. Please find our response to your questions and suggestions below (we first repeat the referee’s comment in *italic* and then provide our answer)

1. *Indicators vs. index*

Thanks a lot for highlighting this issue. Indeed, the terminology of indicator, indices and combined indicators in this paper had not been used appropriately, even though it had been defined. The mixture of terminology is, unfortunately, taking place throughout literature. Reconsidering Hayes 2000, Zargar et al.2011, Vincente-Serrano 2012 this paper was revised to use the terminology as follows:

Drought indicator: are variables that directly measure climatological parameters, such as precipitation, temperature, etc.

Indices: are quantitative measures that characterise drought levels by assimilating data from one or several drought indicators

Combined drought indices: a combination of several drought indicators and or indices that are categorised to relative drought hazard severity levels

We changed the terminology throughout the text according to these definitions and added the following to section 2.2:

“Variables which describe drought hazard are numerous, and can be categorised into two main groups: indicators and indices (Heim Jr 2002; Zargar et al. 2011) Drought indicators directly measure a certain facet of the drought hazard, e.g. climatological conditions, vegetation health, or soil moisture, by a quantitative measure. Drought indices, such as the Standardised Precipitation Index (SPI) or Soil Moisture Anomaly (ΔpF), are quantitative measures characterising drought levels by assimilating data from one or multiple drought indicators to a single numerical value (Zargar et al. 2011). Unlike these, combined drought indices, e.g. Drought Intensity of the US Drought Monitor (Svoboda et al., 2002) or the ‘Combined Drought Indicator’ of the European Drought Observatory (Sepulcre Canto et al., 2012) blend drought indicators and indices to a categorical hazard-severity index. For the purpose of this study, focus is on drought indices that are commonly recommended (Stahl et al. 2015), readily available, monitored, and used operationally in Europe for drought monitoring (Table 1).”

Still, the ‘Combined Drought Indicator’ of the EDO is a proper name, and thus we had to keep this, but mainly used the acronym after defining it.

2. Page 17/Line 13: Any ideas as to “why” all impact categories have reported impacts post-2000? Is this simply due to more contemporary collection methods for incorporation into EDII by the team that built the database?

We added the following text to 4.1: “The observed increase in the occurrence of reported impacts from 2000 onwards may have several reasons. One of the most important one being an increased reporting behaviour (governmental and news) due to an increased awareness of natural hazard impacts and the possibility of easy and fast communicated information (internet). Nevertheless, we cannot exclude the fact that Europe is warming and that this warming may lead to an increase in reported drought impacts.”

3. Page 25/Lines 8-15: Good to see the “fire” issue included as it is very hard to discern regular fire season activity from drought exacerbated fire. Temperatures also play a key role, particularly winter temps. Fuel loads and such are often tied to much longer time frames leading up to the fires themselves with droughts providing the trigger in many cases after forest stands are vulnerable to pests and disease, and thus mortality.

Thank you.

I would like to see the Figures, 2-6 in particular, be larger in order to be more readable.

Indeed, larger images are desirable, but the amount of information provided is very high. Since we decided to focus on the differences between e.g. impact categories or hazard severities, it was found to be important to have all information comparable on one site. Thus, we prefer to keep this information level. We will work with the HESS production department to maximize the size of the figures in the final HESS layout (which is considerably larger than the Discussion paper format)

I do like the format for Figures 4-7.

Thank you.