This is the Authors' reply to comments from Referee #1. We will use blue color for our reply and black color for Referee #1 comments.

First of all, the Authors want to thank the Referee for the work and comments which without doubt will help to improve the paper.

It should be noted that the comments of the Referee #1 make reference to pages and lines of the production paper. We will use the same criteria in this reply.

Interactive comment on "Searching for the optimal drought index and time scale combination to detect drought: a case study from the lower Jinsha River Basin, China" by Javier Fluixá-Sanmartín et al.

Anonymous Referee #1

Received and published: 25 May 2017

Summary: This manuscript presents an analysis of meteorological drought metrics over the lower Jinsha River Basin in China. They use precipitation data from 29 meteorological stations and calculate various formulations of the Standardized Precipitation Index (SPI), the Rainfall Anomaly Index (RAI), the Percent of Normal precipitation (PM) and Deciles (DEC). These indices are then evaluated spatially and in the context of their intensity using what the authors call the Overall Drought Extension (ODE) and the Overall Drought Intensity (ODI). Characterizations of these metrics are then compared to historical documentation of droughts over the lower Jinsha Basin from 1960-2014 to assess their efficacy in characterizing historical drought events. The authors make various conclusions about which of the indices and their spatial characterizations best represent the historical data.

General Remarks: This is generally a well written paper (it nevertheless could benefit from some English writing improvements and attention to typos throughout) that seeks to evaluate how best to characterize meteorological droughts over the lower Jinsha Basin. It should be published after some major revisions regarding clarity and content, which I outline as general comments below.

We want to thank Referee #1 for the General Remarks. English writing improvements will be carried out before submission of the revised manuscript. Aspects regarding clarity and content are treated below.

2. The authors make clear that their assessment is specific to meteorological drought and therefore focus exclusively on precipitation. This is fine as far as it goes, but they make several statements about temperature and river discharge assessments in the context of droughts that are too critical and not entirely accurate.

The authors wanted to stress that, in this particular case, precipitation data is the most reliable source of information. The questionable statements about temperature and discharge will be reviewed and adjusted.

Moreover, they point out that temperature/ET plays an important role in droughts within their study region (e.g. Pg. 3, Lns. 27-29). While they note as a caveat in their conclusions that ET has not been considered and may explain some of the deficiencies in their assessments, it is too little and too late in my opinion. The authors need to take on this obvious criticism of their study more directly and provide more guidance on how it might impact their results, if not try to quantify the impact of ET in an assessment metric. They also should not be so dismissive of the vast amount of work that has shown integrated drought metrics like modeled soil moisture, PDSI, SPEI, etc. to work as a suitable measure of drought (they mentioned the US Drought Monitor, but fail to note it is based on PDSI!).

For instance, their paragraph starting on Pg. 2, Ln. 31 is far too dismissive of integrated metrics and reads like a poor justification for why they focus only on precipitation. If they only have reliable precipitation data over their study region that is fine, but a focus on precipitation alone in this case should not be falsely justified by an attempt to dismiss integrated metrics. This aspect of the paper needs to be modified throughout.

We have only used precipitation-based indices for several reasons:

- The availability of measured meteorological data was limited; precipitation was found as the single most reliable type of information.
- It is true that the use of integrated drought metrics such as PDSI or SPEI could improve the scope and quality of the study and enrich the procedure. However, potential evapotranspiration (PET) data is required to compute these indices, and no reliable PET data was available for the study region. PET calculation depends on solar and longwave radiation, temperature, wind speed, and humidity. Although approximations may be used to estimate this variable, for example by only using temperature data, some studies (Jeevananda Reddy, 1995; Shaw and Riha, 2011; Staage et al., 2014) showed a high sensitivity of the PET to the chosen approximation method. A deeper analysis that helps selecting and applying such methods is needed.
- While this study is specific for the lower Jinsha River Basin, the procedure proposed is intended to serve as a basis for further studies in other regions where only precipitation data is available. This study should be seen as a test for other cases to validate whether precipitation-based indices can be used to predict droughts at a basin scale.

The authors will present these considerations more clearly in the manuscript to justify the only use of precipitation-based indices.

Same answer applies to the corresponding comment of Referee #2.

2. I am not convinced that the metrics proposed by the authors are new. They claim that the ODE and ODI are newly developed metrics and tout their development at multiple points within the manuscript. The ODE is just a form of drought area index and is no more than a measure of the total area of their study region in drought. A similar criticism can be made of the ODI. I therefore have no criticism of the application of these methods, just that they should not be touted as newly developed metrics or metrics of particular novelty that somehow add to the importance of their study.

As indicated by Referee #1 (and also Referee #2), some works (Bhalme and Mooley, 1980; Fleig et al., 2011; Mitchell et al., 1979) have already developed and used drought area indices, although without specifically using the SPI, RAI, PN and DEC indices for their definition. Consequently, the authors will mention these references in the manuscript and avoid presenting the ODE and ODI indices as newly developed. Instead, they will indicate that these indices (ODE and ODI) are an adaptation of existing ones.

3. The authors present quantitative metrics for comparing drought conditions based on their metrics and the historical records of droughts in the region. What is not clear, however, is how they actually translate the historical data into quantitative measures that can be compared to the drought metrics. In other words, they define skill scores in terms of hits, misses, etc., but what actually constitutes a hit or a miss? Is it just timing? Are magnitudes considered?

For the original paper, the authors had considered the temporal coincidence (timing) of drought events as defined by the ODE indices surpassing a predefined threshold (magnitude). The following definitions were used:

- A hit: when one (or more) drought detected according to the ODE values happened during the same year of a historical drought.
- A miss: when, during a year where a drought has been recorded, no event has been detected.
- A false alarm: when one (or more) drought detected according to the ODE values happened during a year when no event has been detected.

- A correct rejection: when, during a year where no droughts have been recorded, no drought has been detected according to the ODE values.

The maximum number of hits (or misses) was limited to the number of years of the study period, i.e. 55, which impacted on the confidence interval of the PSS and thus the precision of the PSS-based results.

The comments of Referee #1 have entailed a discussion among the authors of potentials for improving this approach. We now propose using a discretization by months (instead of by years) for the matching between detected and recorded droughts. That means that:

- We create 2 **monthly** series of events, each month being either “drought” or “non-drought”: one series for the historical events; and one series for the detected events.
- For each month, we check if a drought in the detected series corresponds to a drought in the historical series, thus defining the hits, misses, etc.
- In this way, we increase the number of possible hits, misses… thereby increasing the sample size and reducing the confidence interval of the PSS.

Moreover, instead of calibrating different ODE thresholds for the different index-timescale combinations, we will now use the same threshold across all index/timescale combinations for the analysis, investigating the sensitivities of our results for a range of thresholds. This decision is based on the fact that ODE captures the coverage of droughts at the basin scale, which should not depend on the type of index used. We think this new approach is less arbitrary and more consistent than the original one. It is worth mentioning that new results have been obtained, which differ in some way from those presented in the original manuscript. In particular:

- The confidence intervals corresponding to each PSS value have been reduced, which implies a greater statistical confidence on the new results.
- In general the 3- and 6-month timescales offer better results than the 12- and 24-month timescales, practically for all the thresholds, while in the original work best results were found for the 6- and 12-month timescales. Figure 1 shows an example of results for 2 different ODE thresholds (0.45 and 0.8).
- Based on the PSS values and taking into account their confidence interval, there are no statistically significant differences of results across the different indices for the 3- and 6-month timescales. This indicates that indices perform similarly well, consistent with the fact that they all rely on the same type of data (precipitation).

A complete description of this new approach and the results obtained will be included in the final version of the manuscript.

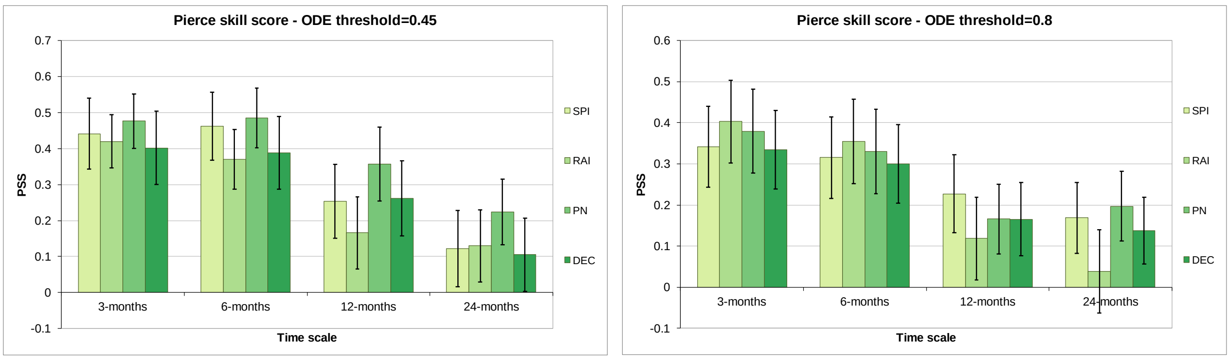

Figure 1. New PSS results. Comparison between ODE thresholds of 0.45 and 0.8.

Note that drought magnitudes are not directly considered for the comparison between historical records and our metrics.

It is also not clear why the authors consider the historical accounts a reliable benchmark, relative to the more quantitative measure of droughts that they develop in their study. I do not think enough emphasis is placed on skill scores that are impacted by inaccuracies in the historical records (in terms of how well they characterize the timing, severity and spatial extent of droughts) relative to what the authors construct from the network of precipitation records.

The catalogue of historical droughts is mainly used to find which combination of index and timescale best fit these catalogued events for different thresholds of the ODE indicator. That means we considered that the information collected during the compilation process is the basis to which quantitatively detected droughts must be compared, simply because no other benchmark exists in this region. While the historical accounts may not be entirely reliable, in this study they are used as a reference for lack of better information.

However, as mentioned in the manuscript, historical accounts should be addressed carefully, in particular regarding the reliability of the data sources and their ambiguity. The compilation process of information relative to historic drought events will be described in more detail in the manuscript. Particular focus will be set on the type and form of information which was available and used. And the authors will discuss the availability and accuracy of the information on the drought characteristics (such as date, duration, area, etc.).

The authors will discuss the expected sensitivity of the results to the historical records.

Specific Comments:

Pg. 10, Ln. 17: While not essential, the authors might consider using two consecutive positive or negative years to end or start a drought. There are definitely periods in Figure 2 that identify droughts as separated by a single year of positive SPI values or very short droughts that represent just single-year excursions. If more persistent and widespread droughts are the interest, a 2-yr criterion for beginning and ending droughts might help.

Indeed, Figure 2 shows very short drought events (1-2 months). This is more noticeable for low timescales and is due to the identification criteria based on the index values (e.g., the criteria defined by McKee et al. (1993) for the SPI). The authors agree with Referee #1 and consider that it would be convenient to set a minimum duration of droughts. According to the range of drought durations of the historical events recorded (≥3months and ≤13months), a **2-year** criterion for beginning and ending droughts seems too restrictive. The authors will adapt the applied identification criteria to avoid an overestimation of events by defining a minimum duration of dry (and wet) periods of 3 months.

Pg. 15, Ln. 6: The authors optimize the characteristics of their drought metrics based on skill assessments over the full historical interval. This is akin to calibrating the forecast model and then performing in-sample skill assessments. A more rigorous assessment would be to optimize over a specific period and then assess the skill in an out-of-sample period. This could be done using block hold out periods or leave half out assessments. As it stands, however, the authors optimize over the same period that they assess the skill of their metrics. This is particularly relevant when considering the authors' methods for future drought assessments. Their in-sample skill assessment is very likely to exaggerate the efficacy of their metrics for future droughts.

The main objective of our work is to identify which combination of index and timescale offers good correlation with the observed events. The fact that only 13 events have been documented is an important limitation, and the authors realize that an optimization of the ODE threshold based on this limited sample is not robust. Any meaningful cross-validation of this optimization would require a larger sample. Instead, the authors no longer search for an optimal threshold but explore the effect of varying the threshold in a reasonable range. Since the general findings turn out to be independent of the specific threshold, we consider them robust.
Same answer applies to the corresponding comment of Referee #2.

REFERENCES

Bhalme, H. N. and Mooley, D. A.: Large-Scale Droughts/Floods and Monsoon Circulation, Monthly Weather Review, 108(8), 1197–1211, doi:10.1175/1520-0493(1980)108<1197:LSDAMC>2.0.CO;2, 1980.

Fleig, A. K., Tallaksen, L. M., Hisdal, H. and Hannah, D. M.: Regional hydrological drought in north-western Europe: linking a new Regional Drought Area Index with weather types, Hydrological Processes, 25(7), 1163–1179, doi:10.1002/hyp.7644, 2011.

Jeevananda Reddy, S.: Sensitivity of some potential evapotranspiration estimation methods to climate change, Agricultural and Forest Meteorology, 77(1–2), 121–125, doi:10.1016/0168-1923(95)02239-T, 1995.

McKee, T. B., Doesken, N. J. and Kliest, J.: The relationship of drought frequency and duration to time scales, pp. 179–184, American Meterological Society, Anaheim, CA., 1993.

Mitchell, J. M., Stockton, C. W. and Meko, D. M.: Evidence of a 22-year Rhythm of Drought in the Western United States Related to the Hale Solar Cycle since the 17th Century, in Solar-Terrestrial Influences on Weather and Climate, edited by B. M. McCormac and T. A. Seliga, pp. 125–143, Springer Netherlands, Dordrecht., 1979.

Shaw, S. B. and Riha, S. J.: Assessing temperature-based PET equations under a changing climate in temperate, deciduous forests, Hydrological Processes, 25(9), 1466–1478, doi:10.1002/hyp.7913, 2011.

Staage, J. H., Tallaksen, L. M., Xu, C.-Y. and Van Lanen, H. A. J.: Standardized precipitation-evapotranspiration index (SPEI): Sensitivity to potential evapotranspiration model and parameters, in Hydrology in a changing world: environmental and human dimensions ; proceedings of FRIEND-Water 2014, Montpellier, France., 2014.