Reviewer #2 (Anonymous Referee)

Burgdorf and colleagues, motivated to better understand drought forcing in North America, use the LBDA and EKF400 datasets to relate multi-year droughts in North America (via LBDA) to their synoptic circulation drivers (via EKF400) over a sufficiently long record to make robust claims.
The authors rely on clustering analysis of multiyear drought events (5-yr running mean on the standardized PDSI values) to identify their prevailing spatial patterns. They find two dominant modes of soil moisture anomalies (consistent with previous findings), and building on work, are then positioned to link those patterns to their atmospheric drivers via the EKF400 data assimilation product. They find (generally consistent with the previous literature) that particular configurations of ocean-atmosphere variability select for different drought types.
Overall the paper appears to be in a position to make a nice contribution. I have a few larger comments and some minor ones the authors might find helpful in a revision.

We thank the reviewer for the positive feedback and constructive comments and suggestions. In the following, we will take a stance on them.

Major comments

1. How does the spatial domain presented (page 5, line 4) influence the clustering of the drought events and thus the spatial patterns presented? Presumably the clustering is quite sensitive to the domain selection and its odd that Figs. 1 and 5a use a constrained North American domain to show the drought patterns, while the rest of the analysis puts North America more fully in perspective. The LBDA v1 on which the central analysis is based, encompasses all of North America, so I wonder why the authors chose to constrain their analysis in such a way, particularly given their emphasis on both pattern identification (which is likely domain-sensitive) and synoptic scale circulation on such drought events. I recognize the authors' point at page 3, line 31 about version 2 being more limited spatially. But since the 20th C. drought is dropped from the central analysis, why not expand the domain to encompass all of NA, or at the least, all of CONUS (as in Fye et al.)?

That is a legitimate point we focused on for a considerable amount of time. We first looked at the first version of the NADA (Cook et al., 2004, 2008) and compared the droughts resulting from our detection method with the droughts found by Fye et al. (2003) who used an earlier version of the PDSI (Cook et al., 1996, 1999). In these first examinations, we looked at the entire domain (all of North America) and ended up with a very similar set of droughts. We then looked at the droughts in the even more advanced version the LBDAv1 (Cook et al., 2010) which again showed an almost identical drought catalog, affirming the robustness of our approach. Since all multi-annual droughts we found mainly affect the Great Plains and the Southwest, we tested the idea of subsetting our domain to even better capture the spatial signal of droughts in this particular drought-prone area. In doing so, our detection method revealed in addition to the pervious drought catalog, further multi-annual drought periods that are very prominent in the relevant domain but remain disguised when looking at entire North America.
So while we argue that the development and subsistence of droughts in North America are driven by large scale circulation, we recognized that their spatial signature is more limited to particular regions in central and western North America which calls for a more regional vantage point in terms of the detection and classification of droughts. We ended up choosing a domain that includes prominent drought regions in the Mississippi Valley, Northern Mexico, and the southern Canadian Great Plains but excludes the East Coast and the tropical South as well as most of Canada.

*2. I wonder about the comparison of EKF400 anomalies to the LBDA anomalies: do they share the same standardization intervals? I *believe* the LBDA is standardized relative to the instrumental period of 1931-90 (could be wrong here), and then the authors here take the 5-year running mean to define a drought event (plus a spatial scale threshold). The GPH analysis is just relative to +/- 5 years centered on the drought. I wonder if the atmospheric fields should first be centered to the same interval as the LBDA and then those EKF400 anomalies can be composited with the +/- 5 year approach. This may make some more consistent GPH, T2M, and SLP patterns emerge.*

In our opinion this is negligible for the following two reasons: Firstly, we would like to point out that since we are looking at anomalies alone, the absolute values and how they are standardized are not relevant. Secondly, we do not directly compare LBDA and EKF400 anomalies. We use the LBDA solely to detect multi-annual droughts (± 5 year running mean plus a spatial scale threshold) based on a standardized PDSI relative to the instrumental period. The EKF400 which does not have the same standardization interval is analyzed strictly independently of the LBDA, we just apply the composite analysis over the detected drought years.

3. Are the EKF400 T2M data representative of ocean skin temperatures? Certainly SSTs and T2M should share the same variations over climate timescales, but some kind of validation of that, or just using SSTs over ocean basins would be more sound for making claims about oceanic forcing.

We are currently looking into this comparison and will provide some statement/validation in the revised version addressing this matter. Furthermore, we plan to include a new supplementary figure in the revised manuscript showing the comparison of SSTs and T2m.

4. It strikes me as a pretty large missed opportunity to not also leverage the PHYDA in this work as a check on EKF400 results, given the uncertainties in the latter that the authors concede (e.g., page 6, line 33) and the authors' search for robustness. As I understand it, EKF400 simulations are forced with SST reconstructions that use a number of the same proxies that are also then used in the data assimilation process itself, which seems potentially problematic. The authors' ability to make robust claims about wave trains, jet positions, and SSTs would be greatly enhanced if there is consistency across more than one atmospheric reconstruction, which is now publicly available.

We discussed before submitting our paper whether to include some comparison with PHYDA or not. Since Steiger et al. (2018) show that PHYDA is highly consistent with the LBDA we decided to stick with the LBDA which has a higher spatial resolution. Nevertheless, we will perform our T2m composite analysis with the PHYDA dataset to compare with our results and add some comparison to the revised manuscript. Unfortunately, T2m is the only variable we can compare with our EKF400 analyses, PHYDA does not include SLP and GPH fields.

5. Updating the Fye et al. paper seems to be a central motivation in this work and there are places where contrasts are drawn between the findings here and those in Fye, which is interesting, but it would be great to have the reasons for those differences explained or hypothesized about a bit more.

A possible hypothesis for the slightly different drought catalogue in Fye at al., 2003 and our study is the fact that they 1. use a different, less sophisticated version of PDSI reconstruction, 2. use a different drought detection metric and 3. use the entire domain over the contiguous U.S. Given these methodological differences, it is rather remarkable how similar the results are, pointing to their robustness.

6. Finally, the outlier pattern associated with the most recent drought is really quite compelling as the authors suggest this one is anomalous based on their pattern clustering. Are there any droughts in the original two clusters (Dust Bowl and 1950s) that look somewhat like the modern drought? Some more validation of that finding would be really great. Could it be a methodological artifact due to its being in

version 2 and not 1, and the need to put v2 (PMDI) on equal footing with v1 (PDSI)? It might be easier to drop this from the paper and do a more rigorous treatment of it in a separate analysis.

There is one drought (1652-1656) in our catalog of 16 droughts that somewhat resembles the turn-of-the-century drought. It also features the unique (among the other droughts in our catalog) spatial pattern with a diagonal divide between dry anomalies along the entire Westcoast and Southwest and wet anomalies (!) in the north-eastern Great Plain stretching down south along the Mississippi Valley. We will add plots of all the individual multi-annual droughts as a new supplementary figure so droughts can be looked at individually. We also add here the correlations plot showing how different the turn-of-the-century drought is compared to the two drought types. We don't think that the fact that the drought has a very distinct spatial signature is an artifact caused by the calibration of the LBDAv2 to fit the LBDAv1. The differences are only minor and would not explain the very different spatial pattern featuring the 21th-century drought.

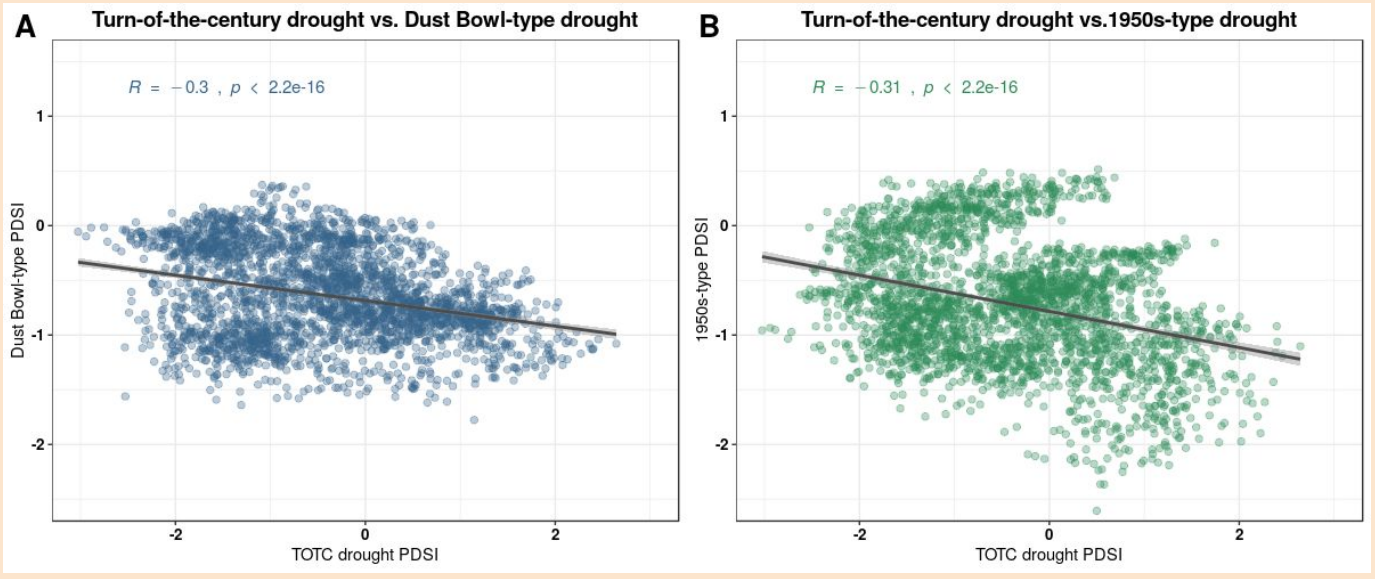

Figure1 | Correlations (pearson) between the turn-of-the-century drought and the "Dust Bowl"-type drought **A** and between the turn-of-the-century drought and the 1950s-type drought respectively **B.** Each point depicts a gridbox.

Minor comments

- *P2, L27: You cite internal variability here; a recent Cook et al. 2018 paper ("Revisiting the Leading Drivers of Pacific Coastal Drought Variability in the Contiguous United States," Journal of Climate) shows that there are numerous ocean-atmosphere configurations that can give rise to the same drought pattern in the West Coast of North America.*
  Thank you, we will add that reference to the revised version and point out the combined influence of atmospheric variability and forced ocean low-frequency variability.
- *P5, L5: is the PDSI < -1 consistent with Fye?*
  Yes. However, Fye at all used a different approach for drought detection (they are identifying decadal moisture regimes).
- *P5, L10: point to the Supplemental Figures here?*
  Thank you, that is a good idea, we will point to the supplement here.
- *P6, L9: You discuss two drought types, but in this and the subsequent sentences you reference three.*
  That is a mistake and will be corrected in the revised version.

- *Online it's fine, but in a print out, Fig. 2's color bar is difficult to discern.*
  Thank you for pointing that out, we will adjust the color bar in the revised version.
- *P6, L29: Are these statistical tests on the patterns of droughts or the jet positions? As written it's not clear. (Seems like it should be on the jet positions.)*
  They are on the jet position. We will formulate this more clearly in the revised version.
- *the quotes around "Dust Bowl" and such are upside-down(?); usage of e.g. requires a parenthetical, etc.; please just check the manuscript for the typos throughout.*
  Thank you for pointing out these typos, we will correct them and check for further typos.