Reviewer 1

We thank the reviewer for the careful review of this paper, and for the comments and suggestions provided. Below, we address the comments point by point. We are confident that all of the changes can be implemented, contributing to a stronger overall message and clearer communication of our key points.

1) Delineation of "drought" and "non-drought" period is based on USDM data seems a bit random. 01-01-2008 to 31-12-2011 defined as "non-drought" period but it contains periods of "Extreme" and "Severe" droughts. Similarly, 01-01-2012 to 01-01-2019 "drought" period contains drought-free days along with period of "Extreme" and "Severe" droughts. Since this classification is a basis of the analysis that follows, a more robust classification, perhaps based on drought categories, is needed.

We acknowledge that our initial delineation of drought categories may seem arbitrary. In order to strengthen our analysis, we took the reviewer's suggestion of dividing the drought periods based on 3 categories that emerge directly from the Drought Monitor. Specifically, we have updated this by dividing our study period into three different categories. 1. no drought, 2. moderate drought (everything in categories D0 and D1) and 3. extreme drought (everything in categories D2–D4). We believe this delineation better highlights the changes in onset and propagation of drought and provides a more robust characterization of the drought responses. We have added description of this step in the methods and updated all Figures where we show data in drought categories (i.e. Fig. 4, Fig. 5a, b, Fig 6a, b, c, Fig. 7). .

2) Analyzing and comparing PET and P between drought and non-drought periods, based on NMDC data seems like going in circles since NMDC drought categories are derived from the very dataset.

We would like to note that the data used in this study is from local weather stations at our two sites. We do not use the same data the NDMC uses. Furthermore, we would like to point out that the U.S. Drought Monitor bases its drought intensity categories on a wide range of indicators, including, but not exclusive to P and PET. We only use the Drought Monitor maps in this study to delineate periods falling into particular drought categories for our study area.

3) NDVI derived from Landsat-5, Landsat-7, and Landsat-7 are not comparable and must be homogenized and filtered from clouds and other types of data noise (Goulden and Bales, 2019). I was unable to figure out if homogenization and cloud correction was performed or not. Also, considering the short growing season, a median NDVI value may not be appropriate as it may end up representing the NDVI at the beginning or end of the month. See Roche et al. 2018 for centering technique.

We really appreciate the thoroughness of the reviewer here. First, we only used cloudless images in our analysis. Second, based on the approach described in Goulden and Bales, 2019, we have homogenized our NDVI data and replotted relevant figures. We have also used the suggested centering method presented in Roche et al. (2018). These points have been updated in the methods.

4) NDVI exhibits saturation beyond some threshold precipitation or available water, it can be seen in Figure 7a. You don't expect the NDVI to continue to increase with increasing water availability. Some vegetation expansion is possible when ample

water supply is available and other resources (energy, nutrient etc.) are not limited but eventually max out. Fitting an exponential model ignores this fact.

Yes, it is true the NDVI saturates for any particular vegetation type. Therefore, we have introduced a threshold value that represents maximum greenness based on all historic values over the study period. These site-specific thresholds were then applied to our exponential model to prevent NDVI values from saturating. We have redone our analysis, updated this point in the methods, and replotted relevant figures. These changes did not significantly affect the results.

5) The definition of polygons with homogeneous vegetation and soil textural properties requires further explanation. Considering the fact that you have a mixture of vegetation at both sites, how did you define "homogenous?

We based our delineation of these polygons on observations, both from the field and from remote sensing. The polygons were drawn in a way that were restricted to a common vegetation (grass) cover from NDVI, excluding other types of vegetation (i.e. trees). The assumption of homogenous soil properties is also based on field data, which was obtained via soil plots when the sites were set up by UCSB as part of IDEAS project. These points have been clarified in the methods.

6) The scenarios can be better described in the methods, I could not understand Scenario A and B until looking at figure 9. What is the meaning of the truncated rainy season and how annual P from the truncated months are redistributed? Also, these scenarios represent intense future drought as posted in the research question (iii) but the presentation of results and discussion comes out as typical climate change scenarios.

Our rationale for a truncated season was to explore the effects of an even earlier onset of the dry season then what has already occurred. In this experiment, we are trying to simulate specific scenarios of climate change that would lead to more intense drought conditions (as discussed in the text). The truncated rainy season scenario was designed to explore the effects of an earlier shift in the onset of the dry season than what occurred in the recent drought. Spring rains are important to soil moisture stores and seed germination in grassland ecosystems. If the onset of the dry season were to shift towards early spring, soil moisture stores would be exhausted earlier, senescence and browning of vegetation would start earlier, and lead to conditions more conducive to wildfires for more extended periods of time.

In Scenario A we shortened the rainy season to occur between November – March and rain in other months was lost. In Scenario B the same applies but the rainfall recorded after March was redistributed between Nov-Mar, effectively increasing the intensity of the rain events, but keeping seasonal totals the same. We have updated the description of these scenarios in the text to make things more clear.

7) Figure 9 is interesting but can be conceptually predicted without running a model. Perhaps these results can be analyzed to better understand the onset and longevities of the drought. Something similar to 5a but for different scenarios.

We agree that increased soil moisture drying and drought responses could be predicted from a conceptual model. However, quantifying these changes for a real environment is not possible without running a model. It was our intent to apply our model to illustrate and quantify the potential changes of earlier drought onset and in response to different scenarios that may not yet have occurred. To further emphasize this, we have added to Fig 9 and the relevant discussion the relative % changes of time below the browning threshold to highlight the impacts of the different scenarios.

<u>Minor points:</u>

1) Suggesting removing the unnecessary background information from the methods, i.e. do we need introductory sentences like these "Soil moisture is essential for plant growth and -health and accordingly, there are strong seasonal responses of vegetation to temperature and precipitation (Coates et al., 2015; Roberts et al., 2010)" to describe the study sites?

We disagree here. The work cited was specifically looking at drought and soil moisture deficits from a remote sensing perspective, and it is relevant to the region. It serves as a key background for our study. We have modified the sentence to link our study area and remote sensing to this reference.

2) Precipitation values reported on top of the page 7 don't match the 20% difference reported on top of page 17

We acknowledge this mistake in our calculations. The actual average difference of precipitation per water year between the two sites is about 10%. We have corrected this in the manuscript.

3) You mentioned inland site is not used for grazing, how about the coastal site?

The coastal site is also not used for grazing. We have added clarification to the relevant passage at L139.

4) Provide mean temperature for the two sites.

Mean temperatures for both sites have been added to the relevant passage in the text in L143.

5) Table S1, note the data formatting issue

We assume the noted issue refers to the number formatting of the silt content in Table S1. The issue has been noted and was corrected.

6) Shortwave and longwave radiation measurements: are these net radiations?

Yes, these are net radiations. The meteorological stations at the coastal site includes a four-channel net-radiometer measuring upwelling and downwelling longwave and shortwave radiation, while at the inland site is equipped with a one-channel net radiometer. We have clarified this point in the text.

7) L155: "For each site, we extracted daily maximum daytime temperatures, humidity and precipitation totals and calculated monthly averages to define the meteorology of the drought"- not clear. Which variables are daily maximum and which ones are totals? What do you mean by the monthly average of precipitation total?

Due to the high resolution of the data set (15min), temperature and relative humidity were summarized to diurnal maximum daytime values. Precipitation was summarized to daily totals. The passage in the text at L155 has been clarified.

8) PET calculation using the Penman-Monteith model need more information on how other inputs were derived i.e. conductance, ground heat flux etc.

Due to the comprehensive nature of the dataset, a wide range of variables, such as net radiation, soil temperatures and windspeed, was available. This allowed us to estimate inputs such as conductance and soil heat flux and use them in our Penman

Monteith calculations. We acknowledge the lack of information provided on this approach, so we have modified the text to add additional information in the relevant section at L157.

9) Stevens hydro probe, provide manufacturer and model

Soil moisture content was measured using in-situ probes (Stevens Hydro Probe II, Stevens Water Monitoring Systems Inc., Portland). We have provided additional information on the manufacturer and model of the in situ probes in L163.

10) L166: here you argue for using the degree of saturation but then end up comparing VMC in Figure 9. Relative saturation may have been more appropriate as it accounts for differences in residual WC between the two sites.

The historical soil moisture data is presented as % saturation to account for the difference in soil textural properties between the two sites. We have also now presented our model results as saturation to maintain consistency.

11) Fig S1 SMD can be equal to RAW as stated in the text

SMD can indeed be equal to RAW. If the reviewer refers to the line RAW<SMD< TAW in the text box, we acknowledge the mistake in the formulation and have changed it to RAW ≤ SMD ≤ TAW.

12) Showing Fc, Wp, RAW, and TAW in figure 3 is misleading. The picture depicts a soil profile and not a unit volume. In its current form, it looks like the Wp is always at the bottom of the root zone.

The drawing is based on the FAO conceptual model, which can be found in Allen et al., (1998). Wilting point is indicated there in a similar way towards the bottom of the volume. However, we acknowledge that the indication of parameters and processes in 3-D may lead to confusion as pointed out by the reviewer. We have therefore changed the figure into 2-D to represent the bucket approach in a simplistic way.

13) Equation 1, I don't quite understand what minimum and maximum Kc & median minimum and maximum NDVI means. Are not you regressing the monthly Kc values against monthly NDVI values with the index I being the month 1 through 12?

After some deliberation, we decided to take a different approach on estimating kc via remote sensing to make it more robust. Specifically, we adopted in the approach of Glenn et al. (2011), in which the crop coefficient can be replaced by vegetation indices (such as NDVI) that reflect the actual growth stage of the plant at the time of measurement. No reference values of kc are needed in this case but a direct relationship between kc and NDVI can be used to estimate ET. The text has been updated to reflect this change. This new approach enables a more dynamic kc than the previous approach, making the subsequent analysis more realistic.

14) P-PET is not really a net precipitation, it is closer to aridity P/PET?

Net precipitation is generally defined as the difference between precipitation and evapotranspiration. We recognize the confusion here as we use the acronym of PET instead of AET. We had used netP=P-PET as a leading indicator of NDVI. In order to avoid confusion, we havel changed the terminology to 'available P (aP) for infiltration' to avoid confusion. We have updated the methods to clarify this point.

15) L320 2012-2019 drought is only relevant for southern California. Statewide, the drought ended in 2016.

We have changed the text in L320 to specify Southern California, rather than the whole state.

16) 16) Fig. 8: At what depth these soil moisture measurements were made? Is the simulated VMC are for the same depth or integrated over the entire root zone?

The measurements were taken at several depths (15,20,50cm), however we are only interested in the balance of shallow soil moisture as it captures the dynamics of precipitation and evapotranspiration we are interested in. We use the shallow soil moisture observations to calibrate our model so that we are able to capture the main processes. The simulated moisture content represents an integrated bucket over the root zone and is not a direct reproduction of observed values. This point has now been clarified in the text.