Review of **"Using statistical models to depict the response of multi-time scales drought to forest cover change across climate zones" by** *Yan Li, et al., 2023*

The authors use linear models to explore the influence of forest cover, temperature, and precipitation on the drought indices in various climate zones. The study's motivation and goal are exciting for the community. The exploratory data analysis used in this study is robust and could be interpreted very well. However, I have a few major comments, which shall be clarified/discussed further:

- How do you isolate the local effects of forest cover and drought from the global drivers and large-scale atmospheric patterns? For example, increase/decrease in precipitation, anthropogenic global warming, jet-stream shift, ITCZ, etc.? The tree growth dependency on T and P depends on the biomes (Boreal forest, Temperate seasonal forest, etc.). Each tree has its characteristics.

- It should be described why the authors used linear models for their analysis.

- One suggestion which might be considered to add value to the results:
Using the linear regression model is an excellent approach to analysing the interactions between the variables and features. However, as mentioned by the authors, the interplay among precipitation, temperature, soil, land cover and drought might be complex and non-linear. Authors could add some complexity to the model by using decision-tree-based models already implemented in R and comparing the results with the linear model. On the other hand, simple/shallow decision tree models are also interpretable.

Other comments:

Lines 1-5: What do you mean by forest cover change? Do you mean human-made changes or natural changes?

Line 6: Hard to understand: ""to explore the changes in forest fraction and drought from 1992–2018."". Do you mean to find a kind of relationship between those two? Or exploring them separately? And why those 27 years?

Line 7: which various factors? Please clarify! Are they natural factors or management factors, etc. ..

Lines 8-9: Is precipitation the dominant one among the two variables? Please mention!

Lines 9-10: It needs to be clarified: You mention precipitation and temperature (which describe the climate state), then forest cover and finally, short and long-term drought. The reader needs to catch up

on the clear goal. Please clarify which relationships or driving effects you will explore in this manuscript.
Some chains like: T, P => forest cover => drought?

Line 30: "–500 million hectares up to +1000 million hectares" what do -500 million hectares mean? And all the SSPs show the same trend, or do they differ from each other?

Lines 35-45: maybe you could also mention that extensive forests like the Amazon are the sink of CO2 and are predicted to become a source of CO2 under the recent trend of climate change we are following:

Boulton, C.A., Lenton, T.M. & Boers, N. ***Pronounced loss of Amazon rainforest resilience since the early 2000s.*** *Nat. Clim. Chang.* **12**, 271–278 (2022). https://doi.org/10.1038/s41558-022-01287-8

Line 57-58: "in this region," which region mention again.

Line 59-60: Is drought a condition or a phenomenon? Clarify? There are also many definitions for drought, like meteorological, agricultural, etc.… Please clarify how you define the drought. Which index do you use? Is it based on temperature and precipitation, or other variables, like soil moisture, evaporation, etc. are, involved?

Line 62-63: What do you mean by "forest structure and carbon content"? Please clarify.

Line 65: How much increase in the frequency and intensity? Is it significant? With respect to which period? Please describe in more detail!

Lines 79-80: The word ""change"" is used frequently.

Lines 87-88: Why didn't you use the newer version of the data with a higher resolution or cite this study:

Beck, H., Zimmermann, N., McVicar, T. *et al*. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Sci Data* **5**, 180214 (2018). https://doi.org/10.1038/sdata.2018.214

Figure 1: Please insert the number of grid points belonging to each main climate classification.

Line 116: Does the potential evapotranspiration data have a reference?

Please include a table with the characteristics of the data used in this study to have a better overview. For example, it is boring to know when and where you downloaded each dataset. A table would be enough, which describes all the datasets. And please include the citation of each dataset in the table.

Figure.2. Given that the scPDSI values between -1 and 1 are considered normal, how significant are the annual trends shown in the drought indices in Fig.2?

Line 172: Why not consider the precipitation sum (yearsum) instead of mean (yearmean)?

Line 175: What do you mean by complex? Clarify!

Lines 200-215: How about the problem of collinearity? There might be correlations between the forest cover change and P or T. How do you consider this? A correlation matrix might show the collinearity between the predictor variables, or the Variance inflation factors (VIF) method might help. The other concern is how many grid points you achieve for each climate zone. How big is the training dataset for each climate zone? I assume you have a more extensive training dataset for the temperate than the equatorial zone. How about the seasonality? You have an arid zone in both the North and South hemisphere. Averaging over all those grid points might mix seasons. Could one include the latitude as an extra feature in the lm model?

Equation 6: Is "i" indicating the observation over different grid points and times? Or do you average the gridpoints of each climate zone at each time, and "i" is just the time? Do you train for each grid point a separate linear model? Do you train one linear model for each climate zone? Please clarify in more detail....

Line 261: Must be moved to data and methods.

Line 262: You mentioned the regions before.

Lines 283-284: How does the time deviation of forest cover look like in equatorial regions? There may be some temporal changes in tropical forest cover. This is because the trees receive enough energy (T) and moisture (P) throughout the year. Have you removed the seasonal cycle from the "lm" features, i.e., T, P and forest?

Lines 295-297: It is a strong conclusion based on a single linear statistical model. I would be cautious about concluding solid results on this.

Figure 6: I see green and yellow colours and not blue and red lines. Using symbols instead of colours could help readers with colour blindness.

Lines 364-366: Given that the trees' species might change in the snow and equatorial regions, how do you isolate those impacts?

368-369: I am unsure if this is the correct English: "The colour change … should be vertical". Please re-frame.

Line 382-383: Please mention that your conclusion is valid only under the assumptions you use here. There might be other models more accurate than your linear model.

Line 414: You have to spell out CMIP.