Reviewer 1

Review of Grant et al., Historical Trends of Seasonal Droughts in Australia

This manuscript analyzes historical drought trends in Australia using multiple drought indicators, including meteorological, agricultural, hydrological, and impact-based metrics. It employs explainable machine learning to assess hydrometeorological drivers and the influence of variability versus mean changes. The study is well structured, the methodology is generally robust, and offers a valuable contribution to the understanding of historical drought trends in Australia. However, I have a several recommendations for improvement to enhance clarity, statistical transparency, and methodological rigor. This review evaluates the manuscript against HESS criteria, with recommendations to enhance its impact and alignment.

We would like to sincerely thank the reviewer for taking the time to review the manuscript and for their positive assessment.

Scientific Significance

1. Does the paper address relevant scientific questions within the scope of HESS?

- Yes, the manuscript examines historical drought trends and their drivers, contributing to the understanding of water availability and hydrological variability. The study's integration of multiple drought indicators and machine learning techniques is relevant to advancing hydrological monitoring and data analysis. Additionally, its focus on drought impacts aligns with HESS's interest in hydrology's interaction with climate and society.
- **Suggested Revision:** The discussion could further highlight how the findings might support sustainable water resource management and decision-making in response to drought variability.

Thanks for this suggestion. We will expand the discussion section to highlight the implications on water resource management and decision-making.

2. Does the paper present novel concepts, ideas, tools, or data?

- Yes, ML integration with traditional drought assessment is emerging in Australia (line 135, citing Devanand et al., 2024; Hobeichi et al., 2022). While advancements exist, the field is evolving. This manuscript contributes by combining multiple drought indicators with an impact-based ML model, offering a novel approach.
- **Suggested Revision:** Provide a more explicit comparison with prior ML-based drought assessments to clarify how this approach improves upon past methods (e.g., better accuracy, broader applicability, or deeper insights into drought impacts).

Our study uses machine learning in two ways – first, we use machine learning to develop an impact-based drought metric that captures droughts using drought impact reports from local and federal government authorities. This approach builds upon previous studies that have used machine learning in combination with drought impact reports to assess the drivers and predictability of drought (Devanand et al., 2024; Hobeichi et al., 2022). Our approach goes beyond previous research as we use the impact-based drought metric to quantify historical drought trends, complementing our drought trend analysis using more traditional drought metrics. In doing so, we show that traditional drought metrics are able to capture the historical trends in drought seen in the impact-based drought metric.

Secondly, we use machine learning to understand drivers of drought trends (i.e. which hydrometeorological variables contributed most to past changes in drought in each season and region). The application of machine learning in this context is novel and provides important insights into the most important drivers of drought trends.

We will edit the manuscript to detail the novelty and advancement over past approaches and studies, as suggested.

3. Are substantial conclusions reached?

- Yes, the manuscript draws key conclusions about the historical variability of droughts and suggests that recent trends are not unprecedented. However, the phrase "within historical variability" appears multiple times without clarifying whether it implies a lack of anthropogenic influence or if natural variability obscures a long-term trend.
- **Suggested Revision:** Add a clarifying statement in the discussion that "within variability" does not necessarily imply no human influence, as variability can mask emerging signals. Also, consider discussing whether these trends are expected to continue or if they are part of cyclical natural variability.

We agree with the reviewer that a drought trend being within the bounds of historical variability does not imply that there is no human influence. We were careful with the wording in our manuscript to ensure we stated "within observed variability" (as opposed to natural variability) as the historical changes can be due to both natural variability and anthropogenic influences.

We cannot quantify the contribution of the trends from human or natural influences separately with the data used in this study. An attribution study of this nature would typically rely on climate models, however these are not yet reliable for this purpose (Lane et al., 2023). We will revise the manuscript to explain that observed variability could include human and natural variability.

Scientific Quality

4. Are the scientific methods and assumptions valid and clearly outlined?

The methodology is well-described, but the manuscript could be improved with more explicit discussion of the following points:

- **Threshold Selection:** The authors justify the choice of the 15th percentile threshold, but should briefly discuss how this selection might impact results.

Past studies have found drought trends to be largely insensitive to the choice of threshold (e.g., Kirono et al., 2020; Ukkola et al., 2018). We expect that a lower percentile would likely lower the significance of trends due to a smaller sample size but would be unlikely to affect the sign of changes. We will revise the manuscript to clarify the choice of threshold and outline any potential impacts different threshold choices would have on the results.

- **Groundwater Use:** The limitations mention that AWRA-L does not account for groundwater use, but there is no discussion of how reliance on groundwater varies regionally and how this affects hydrological drought impacts.

We thank the reviewer for this suggestion and agree that this is an important point. Groundwater sources have large regional variation across Australia, and this could have influence over the impacts of drought. For example, areas with large groundwater reserves may have a delayed response to hydrological drought onset during extreme dry periods. On the other hand, groundwater can be linked to prolonged post-drought recovery, with catchments taking years to fully recover after long periods of drought (Fowler et al., 2020, 2022).

These responses to drought in regions with differing groundwater availability could have influence over the subsequent impacts of hydrological droughts. We will revise the manuscript to highlight the role of groundwater for hydrological drought impacts.

- **Drought Impact Data Bias:** The methodology explains that the RF model was trained on government drought reports, but it does not specify how those reports define or measure drought impacts. Furthermore, potential biases in training data due to regional population/economic factors should be acknowledged.

We agree it would be helpful to provide more details on how droughts were defined in the impact reports. We will include examples of how a drought impact was defined and acknowledge possible biases due to population and economic factors.

- **Uncertainty Consideration:** While variability in trends is discussed, the manuscript does not explicitly assess whether uncertainty bounds overlap with observed trends.

We are not sure we understand the reviewer's question, but we think it relates to whether observed trends are within the bounds of observed variability, i.e. if the trends are statistically significant in the context of observed variability.

Our study includes a comprehensive assessment of the emergence of drought trends from observed variability (see Section 2.4 and S1.2 for details of the methods). This includes a comparison of historical trends against uncertainty bounds of decadal means and the application of two trend emergence tests; the signal-to-noise (S/N) ratio and Kolmogorov-Smirnoff (KS) test (see the description of these results in Sections 3.1 and 3.2). We found that while some areas showed signs of trend emergence, the trends largely remain within the observed variability.

In summary, we believe that this comment is already addressed in the current version of the manuscript.

5. Are the results sufficient to support the interpretations and conclusions?
While the results are well presented, statistical transparency needs improvement. Specific concerns include:

- **Random Forest Model Overfitting:** The results discuss Mean Decrease in Impurity (MDI) for feature importance but do not report whether individual decision trees reached pure leaf outcomes (fully deterministic splits).
- **Regional Data Sparsity:** The manuscript does not explicitly address how the RF model handles regional data sparsity and whether feature importance rankings shift across different drought types and regions.
- **Suggested Revision**: Report maximum and mean tree depth and include a diagnostic plot of impurity reduction versus feature correlation. Additionally, summarize MDI variations between data-rich and datapoor regions and provide out-of-sample error distributions to assess model consistency.

We appreciate the reviewer's attention to the details of our Random Forest implementation. Please find our responses to each part of the suggested revisions below.

Random Forest Model Overfitting
We do not think that information on the purity of individual decision tree leaves contributes meaningfully to interpreting feature importance via Mean Decrease in Impurity (MDI). Due to the nature of bootstrap sampling and inherent randomness in feature selection in Random Forest models, it is likely that some trees may reach pure leaves simply by chance while others may not.

In our analysis, we employed the default settings of the `RandomForestClassifier` in the scikit-learn Python package, which allows trees to grow until no further impurity decrease is possible. This approach follows standard practice and is consistent with typical use of MDI.

Regional Data Sparsity
We believe there may have been a misunderstanding regarding the data used to train the Random Forest models for investigating drivers of drought trends. The training data are uniformly gridded datasets (gridded climate and hydrological data and gridded drought trends) and so have equal density across the regions. It is true that the underlying observations used to develop the gridded data have more density in some regions of Australia than in others. However, quantifying the effects of this on drought trends or feature importances is not within the scope of our study as it would require re-generating the climate data using a subset of stations, which is not a trivial undertaking. However, our manuscript already reports how the performance of the Random Forest and the feature importances vary between regions (see Table S1 for regionally and seasonally varying RF performance, and Figure 7 for regionally and seasonally varying feature importances).

Out-of-sample error distributions
All Random Forest performance metrics shown in our manuscript are based on out-of-sample predictions using a cross-validation approach (Section 2.2.2, lines 143-149, and Section 2.6, lines 234-240).

6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

- The methodology is detailed, including data sources, statistical methods, and ML model parameters.

- **Suggested Revision:** However, some calculations (e.g., Mann-Kendall test and signal-to-noise ratio calculations) lack explicit mathematical definitions. Adding these formulae to the supplement would improve reproducibility.

We will modify the manuscript to include the signal-to-noise formula in the supplementary information. Python packages were used to calculate Mann-Kendall and Kolmogorov-Smirnoff tests. We will add citations of these libraries to the methods section.

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution?

- The manuscript references key studies on Australian drought trends and ML applications (e.g., at line 135, Devanand et al., 2024; Hobeichi et al., 2022).
- **Suggested Revision:** Strengthen the discussion on how this ML-based approach differs from previous studies. Explicitly highlight whether it provides higher accuracy, broader applicability, or novel insights into drought impacts.

As stated above, we will revise the manuscript to include this more explicit comparison as suggested (see our response to comment 2).

<u>Presentation Quality</u>

8. Does the title clearly reflect the contents of the paper?

- The title is relevant but could better emphasize the study's significance or approach.

9. Does the abstract provide a concise and complete summary?

- Yes, the abstract effectively outlines the research gap, methodology, key findings, and conclusions.

10. Is the overall presentation well-structured and clear?

Generally clear, but some linking between results and discussion needs improvement.

- Strengthen the explanation of seasonal drought trends by linking them explicitly to known climate drivers (ENSO, IOD, SAM).

We appreciate the reviewer's suggestion on strengthening the link between results and discussion and explaining seasonal drought trends by linking them to known climate drivers. Previous research has shown that seasonal climate drivers (such as ENSO, IOD, SAM) only explain a small proportion of Australian droughts (Hobeichi et al., 2024) and droughts are influenced by complex synoptic patterns and the frequency of heavy precipitation events (Holgate et al., 2025). As such we have decided not to explicitly link these climate drivers to seasonal drought trends in our analysis. Holgate et al. (2025) provide a review of meteorological droughts in Australia, and explain the connection with large-scale climate drivers in a more comprehensive manner than is possible within the scope of our study. In addition, Wasko et al. (2021) comprehensively discuss historical drivers of trends in hydrological variables and extremes.

- Clarify why evapotranspiration (ET) is a dominant drought driver in some regions but not others.

Differences in the most important drivers of drought trends reflect the differences between climate zones and hydrological regimes across Australia (e.g. the Monsoonal North with a distinct wet and dry season vs a temperate to dry climate in the South-East). We will expand the discussion of these results to explain why the importance rankings of ET are as expected.

11. Is the language fluent and precise?

- Yes, the manuscript is well-written with clear and precise language.

12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?

- Yes, but some statistical methods (e.g., Mann-Kendall test, signal-to-noise ratio calculations) lack explicit mathematical definitions.
- **Suggested Revision:** Include these formulae in supplementary materials for clarity, ensuring that readers unfamiliar with these methods can fully understand their application.

As stated above, we will include the necessary formulae and python package citations (see our response to comment 6).

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?

- Table S1 presents R^2 values for RF models, but it is not easy to compare performance across seasons/regions.
- **Suggested Revision:** Add a column with mean R^2 scores across all seasons/regions.

We agree that it would be useful to be able to compare the skill of the models across different regions and seasons. However, we believe taking a mean of the R^2 scores across the different random forest models would not be a meaningful statistic. Instead, we will add a column indicating the range of the R^2 scores for the different models.

14. Are the number and quality of references appropriate?

- Yes, but some references (e.g., Ukkola et al., 2024) appear incomplete.
- **Suggested Revision:** Ensure consistent formatting and completeness.

We will edit the references to ensure consistent formatting and completeness.

15. Is the amount and quality of supplementary material appropriate?

- The supplementary material is valuable, but some critical methodological details (e.g., preprocessing steps for government drought reports) should be included.

We will include these details as suggested by the reviewer in the above comments.

References

Devanand, A., Falster, G. M., Gillett, Z. E., Hobeichi, S., Holgate, C. M., Jin, C., Mu, M., Parker, T., Rifai, S. W., Rome, K. S., Stojanovic, M., Vogel, E., Abram, N. J., Abramowitz, G., Coats, S., Evans, J. P., Gallant, A. J. E., Pitman, A. J., Power, S. B., Rauniyar, S. P., Taschetto, A. S., and Ukkola, A. M.: Australia's Tinderbox Drought: An extreme natural event likely worsened by human-caused climate change, Sci. Adv., 10, eadj3460, https://doi.org/10.1126/sciadv.adj3460, 2024.

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K.-W., and Western, A.: Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections, Water Resour. Res., 56, e2019WR025286, https://doi.org/10.1029/2019WR025286, 2020.

Fowler, K., Peel, M., Saft, M., Peterson, T. J., Western, A., Band, L., Petheram, C., Dharmadi, S., Tan, K. S., Zhang, L., Lane, P., Kiem, A., Marshall, L., Griebel, A., Medlyn, B. E., Ryu, D., Bonotto, G., Wasko, C., Ukkola, A., Stephens, C., Frost, A., Gardiya Weligamage, H., Saco, P., Zheng, H., Chiew, F., Daly, E., Walker, G., Vervoort, R. W., Hughes, J., Trotter, L., Neal, B., Cartwright, I., and Nathan, R.: Explaining changes in rainfall–runoff relationships during and after Australia's Millennium Drought: a community perspective, Hydrol. Earth Syst. Sci., 26, 6073–6120, https://doi.org/10.5194/hess-26-6073-2022, 2022.

Hobeichi, S., Abramowitz, G., Evans, J. P., and Ukkola, A.: Toward a Robust, Impact-Based, Predictive Drought Metric, Water Resour. Res., 58, e2021WR031829, https://doi.org/10.1029/2021WR031829, 2022.

Hobeichi, S., Abramowitz, G., Sen Gupta, A., Taschetto, A. S., Richardson, D., Rampal, N., Ayat, H., Alexander, L. V., and Pitman, A. J.: How well do climate modes explain precipitation variability?, Npj Clim. Atmospheric Sci., 7, 1–9, https://doi.org/10.1038/s41612-024-00853-5, 2024.

Holgate, C. M., Falster, G. M., Gillett, Z. E., Goswami, P., Grant, M. O., Hobeichi, S., Hoffmann, D., Jiang, X., Jin, C., Lu, X., Mu, M., Page, J. C., Parker, T. J., Vogel, E., Abram, N. J., Evans, J. P., Gallant, A. J. E., Henley, B. J., Kala, J., King, A. D., Maher, N., Nguyen, H., Pitman, A. J., Power, S. B., Rauniyar, S. P., Taschetto, A. S., and Ukkola, A. M.: Physical mechanisms of meteorological drought development, intensification and termination: an Australian review, Commun. Earth Environ., 6, 1–14, https://doi.org/10.1038/s43247-025-02179-3, 2025.

Kirono, D. G. C., Round, V., Heady, C., Chiew, F. H. S., and Osbrough, S.: Drought projections for Australia: Updated results and analysis of model simulations, Weather Clim. Extrem., 30, 100280, https://doi.org/10.1016/j.wace.2020.100280, 2020.

Lane, T. P., King, A. D., Perkins-Kirkpatrick, S. E., Pitman, A. J., Alexander, L. V., Arblaster, J. M., Bindoff, N. L., Bishop, C. H., Black, M. T., Bradstock, R. A., Clarke, H. G., Gallant, A. J. E., Grose, M. R., Holbrook, N. J., Holland, G. J., Hope, P. K., Karoly, D. J., Raupach, T. H., and Ukkola, A. M.: Attribution of extreme events to climate change in the Australian region – A review, Weather Clim. Extrem., 42, 100622, https://doi.org/10.1016/j.wace.2023.100622, 2023.

Ukkola, A. M., Pitman, A. J., Donat, M. G., De Kauwe, M. G., and Angélil, O.: Evaluating the Contribution of Land-Atmosphere Coupling to Heat Extremes in CMIP5 Models, Geophys. Res. Lett., 45, 9003–9012, https://doi.org/10.1029/2018GL079102, 2018.

Wasko, C., Shao, Y., Vogel, E., Wilson, L., Wang, Q. J., Frost, A., and Donnelly, C.: Understanding trends in hydrologic extremes across Australia, J. Hydrol., 593, 125877, https://doi.org/10.1016/j.jhydrol.2020.125877, 2021.