Reviewer #1

This work analyzes and predicts streamflow drought in the CONUS via machine learning and statistics. Random forests are trained to predict droughts from meteorological forcings. Gini impurity values are computed to investigate variable importance at the regional scale. Principal component analysis (PCA) on the variable importances allows the investigation of large-scale drivers of drought across the CONUS. Linear regression is used to predict principal components (PCs) from basin characteristics in ungaged basins. Similarity between PCs is used to select donor gages from which forcings are translated to the ungaged basin and converted by the random forest into drought prediction.

The study is well executed and very well documented. The motivation and objectives are clear from the beginning. The methodology is explained in detail and likely allows for reproduction. The dicussion nicely picks up on the research objectives, offers explanations for surprising findings, and draws elaborate connections to existing literature.

Thank you very much for taking the time to review our submission! Your comments are very helpful and appreciated.

Specific comments (most of them are suggestions and therefore optional):

- l 23-24: Sentence about dividing CONUS into nine regions could be removed

Removed this sentence.

- l 61-62: Indeed, ML requires large amounts of data processing during training. Once a model is trained, though, costs are often amortized

This makes sense from a financial perspective for commercial ML applications, though in the paper, the focus is on acquiring the large amount of data and storage needed upfront for some ML workflows, rather than a cost-benefit analysis of ML.

- l 62-63: Sentence about interpretability feels out of place here, though interpretability is important, so I would put it somewhere else

Agreed - moved this sentence above and incorporated it into the sentence about uncovering patterns since interpretability is a caveat to that (ML can uncover complex patterns but it still must be interpretable enough to describe those patterns)

Old: "ML has a distinct advantage over more traditional approaches in its ability to utilize a wide variety of disparate datasets, potentially uncovering complex patterns in space and time (Jiang et al., 2022; Nunes Carvalho et al., 2022; Zhu et al., 2021)."

New: "ML has a distinct advantage over more traditional approaches in its ability to utilize a wide variety of disparate datasets and potentially uncover complex patterns in space and time (Jiang et al., 2022; Nunes Carvalho et al., 2022; Zhu et al., 2021), though interpretability of AI and ML methods for high-stakes decision making remains a crucial objective (Rudin et al., 2022)."

- l 66-68: Should conflicting training signals from different basins be called noise? Also, isn't the main advantage of the donor-based approach that it allows PUB?

This first point is a good one. I have replaced "noise" with "conflicting signals".

To the second point, I have reorganized the paragraph to highlight the primary advantage of being able to predict in ungaged basins.

Old: "In hydrologic modeling, one approach to reducing the volume of training data required while minimizing the amount of information loss is to use dynamic regionalization to select a subset of models trained at "donor" gages, similar in character but not necessarily proximal to the ungaged location of interest. This subset is then used for prediction at the ungaged location of interest (McIntyre et al., 2005). This method has the advantage of reducing the volume of training data and potentially avoiding the introduction of conflicting signals from including training data from dissimilar basins (Pagliero et al., 2019)."

New: "In hydrologic modeling, dynamic regionalization enables prediction at ungauged locations by selecting a subset of models trained at "donor" gauges that are similar in character, though not necessarily proximal, to the ungauged location of interest (McIntyre et al., 2005). Beyond enabling prediction in ungauged basins, this approach has the additional advantage of reducing the volume of training data required and potentially avoiding conflicting signals introduced by training on dissimilar basins (Pagliero et al., 2019)."

- l 81: Optionally replace "faced across the nation" with "present across the country" or similar

Replaced as suggested.

- l 88: Did you mean non-uniform?

I think we will retain non-random because we are referring to the deliberate, objective-driven placement of USGS streamgages, which is the underlying cause of the documented biases.

- ll 87-91: I feel a little lost with this information and I don't think it is picked up again. Are you trying to say that the results should be taken with a grain of salt?

This is a fair point – we have added a sentence connecting this information to the donor-based approach in Sec 2.4, clarifying these biases as a motivation for the similarity-based selection approach.

New: "These biases further motivate the similarity-based donor selection approach described in Section 2.4."

- ll 105-107: Once you understand that sentence, it is clear, but it took me a while. Maybe the formula could be extended a bit to make it easier

Modified the explanation so that it is a little more understandable:

Old: "To remove this seasonality, streamflow percentiles were based on the Weibull plotting positions computed from data for each day of the year using the values for a 30-day window centered on the given day from all years:"

New: "To remove this seasonality, streamflow percentiles were computed for each day of the year using the Weibull plotting position. For a given day, all streamflow values within a 30-day window centered on that day across all years of record were ranked, and the percentile was calculated as:"

Also,

Old: "where r is rank and n is the number of data points."

New: "where r is the rank of the daily streamflow value among all values in the 30-day window across all years, and n is the total number of values in that window."

- l 131-132: Comment: Also, if you would include antecedent streamflow, the task would be much easier since droughts are defined based on streamflow, and the model likely wouldn't learn the connections you're after

Agreed. Appended the additional information to the end of this sentence: "...and would likely dominate model predictions given the autocorrelation between streamflow and the drought response variable, obscuring the meteorological connections of interest."

- Table 1: I would add horizontal lines to delimit different sources and references. Also, it would be nice if the table wasn't split across two pages

Added horizontal lines in Table 1 and added a page break to Table 1 to avoid the splitting.

- ll 150-152: What exactly is meant with robustness of the method, especially considering that robustness to redundancies and correlations are separately mentioned right after?

Clarified this sentence – the first mention of "robustness" refers to the method's well-established predictive performance and resistance to overfitting through ensemble averaging, which has been made explicit:

Old: "We chose to implement random forest classification models (Breiman, 2001) for the modeling framework because of the robustness of the method (Biau and Scornet, 2016; Tyralis et al., 2019), robustness to redundancies and correlations among predictor variables (post-hoc dimension reduction, described in 2.4.2, is utilized over pre-processing for this reason), and the ability to quantify the relative importance of individual predictor variables (Archer and Kimes, 2008)."

New: "We chose to implement random forest classification models (Breiman, 2001) for the modeling framework because of the method's well-established predictive performance and resistance to overfitting..."

- ll 156-162: This could be shortened a bit

Agreed, we have condensed the explanation of the class-weighting to the following:

"Because of the class imbalance (drought events comprise 20% of observations), training data were weighted inversely proportional to class frequency (0.8 for drought, 0.2 for non-drought), effectively treating each drought instance as equivalent to four non-drought instances in the node-splitting calculations."

- l 169: Did you play around with the node size parameter and found this value empirically or is this a common choice?

The minimum node size of 10 was selected as a common default, consistent with recommendations in literature (e.g., Probst et al., 2019). We have clarified this in the text:

Old: "Minimum node size was set to 10 to avoid overfitting."

New: "Minimum node size was set to the package default of 10 for classification. Given the scale of the analysis (3,198 individual models), individual hyperparameter optimization was not practical; models with poor performance were excluded through the filtering described in Section 2.4.1."

- l 180 and l 176 are the same formula

This comment refers to the formulas for observed agreement, p0, and overall model accuracy. I have removed overall model accuracy.

- ll 191-192: But didn't you explicitly reweight the data so that the classes are balanced? Most likely I'm missing something here, so maybe you could elaborate a bit

The class reweighting described in Section 2.3.1 addresses imbalance during model training, but the test data retain the natural 80/20 class ratio. Kappa is used for evaluation because it corrects for expected agreement by chance on these imbalanced test data. We have clarified this distinction in the revised text.

Old: "By adjusting for agreement that might happen by random chance, it is a helpful metric for evaluating models trained on imbalanced data where a naïve model simply predicting the majority class might result in misleading metrics."

New: “By adjusting for agreement that might happen by random chance, it is a helpful metric for evaluating predictions on imbalanced test data (80% non-drought, 20% drought), where a naïve model predicting only the majority class would achieve high accuracy but a Kappa near zero. While class weighting during training (Section 2.3.1) addresses imbalance in the learning process, Kappa provides a fair evaluation of the resulting predictions on the naturally imbalanced test set.”

- ll 198-201: It was not directly obvious to me how the enumeration relates to the figure. I think adding the paragraph numbers to the enumeration sentence would already help a lot

Added section numbers to each step in the text to help with the connection to Figure 2.

- Figure 2: This figure contains more than the "development and evaluation of dynamic regionalization method", which is the name of this subsection. It contains all the study's research questions! So I would suggest to reference it not only in that subsection but more generally. Also, the figure is almost completely linear, so I think simply plotting it from top to bottom would be more intuitive than the current layout, but that might be a personal preference

We have added a reference to Figure 2 in Section 2.3 where the methodology described in the figure starts. Regarding the layout of the figure, we considered the suggestion but decided to keep the figure as-is to accommodate for the branching associated with the three research questions.

- ll 248-250: These sentences seem redundant to me

Agreed, we have revised to reduce the redundancy:

Old: “To determine a set of suitable donors, a weighted sum of absolute differences between the estimated PC scores for the LOI and the PC scores for each candidate donor was computed. That is, the absolute value of the difference between each estimated PC for the LOI and each corresponding PC generated for each candidate donor by the random forest models was computed. Individual PC scores were weighted by the percent of total variance explained by the PC as described in 2.4.2. For example, if PC1 explains 25% of the total variance in the original dataset of variable importances, the calculated values for the absolute difference between PC scores for PC1 were weighted by 0.25. This ensured that *high variance and therefore more determinative PCs for donor selection were given more weight for the donor selection process. The weighted differences for each PC score between the LOI estimate and each donor were then summed for each candidate donor."*

New: "To determine suitable donors, the absolute difference between each estimated PC score for the LOI and the corresponding PC score for each candidate donor was computed and weighted by the percent of total variance explained by that PC (Section 2.4.2), giving greater influence to more determinative components. For example, if PC1 explains 25% of the total variance in the original dataset of variable importances, the absolute difference in PC1 scores was weighted by 0.25. The weighted differences were then summed across all PCs for each candidate donor."

- l 256: Similarly to node size, I'm wondering whether the 1% were found empirically

Yes, the 1% threshold was selected pragmatically to provide a sufficient ensemble of donors for averaging predictions while maintaining reasonable similarity to the location of interest. We clarified this in the text and acknowledge that sensitivity to this parameter was not formally tested:

Added: "The 1% threshold was selected as a pragmatic choice to provide a sufficient ensemble size for averaging while maintaining similarity to the LOI; sensitivity to this parameter was not formally tested."

- l 259: I understand that you found 0.35 empirically here, but wouldn't one expect it to be 0.2 given you drought definition? On the other hand, you reweighed your training data, so one might even expect a threshold of 0.5. Interestingly, 0.35 is right in between. Is there something to it, or did I miss something?

Yes, the reviewer's intuition is correct – the 0.35 threshold reflects the interaction between the natural 20% drought base rate and the class weighting (0.8 and 0.2) applied during training. The weighting inflates predicted drought probabilities above the natural base rate (0.2) but not to 0.5 because the model's learned feature-space boundaries still reflect the underlying imbalance. The threshold of 0.35 was determined empirically through Kappa maximization, but its position between 0.2 and 0.5 is consistent with this expected behavior.

- Figure 4: First (and only) occurrence of "Julian Day" is in this caption, maybe introduce it earlier. Also, how does it relate to "Decimal Date"?

Thank you - we have introduced the term "Julian Day" earlier in the text where the "weibull_jd" naming convention is first described (Section 2.2.2). We have also clarified the distinction between "Julian Day" (the day-of-year used for the variable-threshold Weibull transformation) and "Decimal Date" (a continuous numeric time predictor included to capture non-seasonal temporal trends). These are unrelated concepts that happen to both involve time.

- ll 313-314: You explain what loadings are here, but you already used the term in l 229

Moved the definition of loadings to its first occurrence in Section 2.4.2 and removed the redundant parenthetical from Section 3.2.

- Figure 7: If I understand correctly, sometimes, prediction works better with the donor approach than with an at-site models. I think this should be discussed, what could be the reason for this?

Agreed, we have added this to the Discussion section at the end of paragraph 2 of Section 4.3: "Notably, the donor-based predictions outperformed the at-site models at some locations, particularly where at-site model performance was low. This is likely attributable to the averaging of 19 donor predictions, which reduces variance from individual model overfitting to local noise. In effect, the donor ensemble provides regularization, where predictions from characteristically similar basins may generalize better than a single model trained on potentially noisy local data."

- ll 546-550: The conclusion is mostly on a rather high level (which is good), so I wouldn't explicitly mention PC1 etc. here, because those terms are quite technical. Maybe there is a way to phrase it more generally

Agreed – we rephrased this section to describe the PCA results in terms of the physical driver groupings and their regional associations in lieu of referencing PC numbers.

Old: “From PCA, we found that, within PC1, teleconnections, temperature, and PET are highly interrelated in the context of streamflow drought. Similarly, we found from analysis of PC2 that the importance of precipitation, SPEI, and soil moisture to streamflow drought predictability are also highly interrelated at consistent smoothing window lengths. These PCs, along with the SWE- and soil-moisture-driven PC3, exhibit the strongest regional differences across the CONUS.”

New: “From PCA, we found that streamflow drought drivers cluster into distinct, regionally coherent groupings: teleconnections, temperature, and evaporative demand covary in the western CONUS; precipitation, SPEI, and soil moisture covary in the eastern CONUS; and snow water equivalent and soil moisture distinguish snow-dominated from baseflow-driven systems. These groupings exhibit the strongest regional differences across the CONUS.”

Technical corrections:

- l 293: "Decimal Date" -> "Decimal date"

Edited.

- l 350: Remove closing parenthesis

Removed.

- l 409: "Date" -> "date"

Edited.

- l 596: doi should be https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2

Edited.

- l 603: doi should be https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2

Edited.

- l 635: doi should be https://doi.org/10.1175/1520-0477(1998)079<2715:IMOET>2.0.CO;2

Edited.

All in all, this is a very good study in my opinion. After some minor revisions, which mostly concern intelligibility, it is ready for publication in HESS.

Thank you very much for your review!

Reviewer #2

Predicting streamflow drought in the conterminous United States using machine learning and a donor-gage approach, 1982-2020

General

The paper covers the prediction of streamflow drought with random forest approaches and a new dynamic regionalization method for ungauged catchments. The study site covers most of United States and over 3000 sites with different topography and hydrogeological conditions. The goals are to identify the most common drivers of streamflow drought, find the defining basin-characteristics of the most impacted by the drivers, and investigate the use of the new method for ungagged basins.

The paper is generally well-structured and well written. And the subject and approach are very relevant for the readers of HESS and the scientific field. Figures are well presented and easy to decode. The method section is detailed and well-described. There are, however, some points in the comments below that should be addressed prior to publication. Most important are:

- Lack of background meteorological and hydrological information in the study description
- The impact on the very broad definition of droughts applied

Thank you very much for taking the time to review our submission! Your comments are very helpful and appreciated.

Comments

L28: if space allows, the abstract could benefit from a concluding remark on potential of dynamic regionalization in other studies

Added a concluding sentence to the abstract: "This method may be applicable to other hydrologic prediction problems requiring transferability to ungaged locations."

L38: consider moving "have all resulted in economic losses totalling billions of dollars" to L35 after "modern droughts"

Moved the economic impacts statement earlier in the sentence as suggested.

L41: could you add a reference for this statement

Added two references for this statement: Seneviratne (2012) and Mishra & Singh (2010).

L55: "traditional approaches" such as?

Added examples: "...traditional approaches (e.g., process-based hydrologic models and statistical regression methods)..."

L61: remove "potentially"

Removed as suggested.

L78: the study site description is very short, and main talks about the data from GAGES-II. Even though you cover an immense area, it would be good to add a little more information on precipitation, streamflow and temperature patterns in general terms.

Added a brief description at the beginning of 2.1. of the general precipitation, temperature, and streamflow patterns across the CONUS to provide more hydroclimatic context for the study area:

“The CONUS encompasses a wide range of hydroclimatic conditions. Mean annual precipitation ranges from less than 250 mm in the arid Southwest to over 2,500 mm in the Pacific Northwest and Southeast, with a general east-west gradient reflecting moisture availability (Gorski et al., 2025). Temperature regimes vary from subtropical in the southern CONUS to continental in the northern interior, influencing the seasonality of streamflow through snowmelt timing and evaporative demand. Streamflow regimes are correspondingly diverse, ranging from snowmelt-dominated systems in the Rocky Mountains and northern latitudes to rainfall-driven and baseflow-dominated systems in the eastern CONUS (Rice et al., 2015).”

Figure1. it seems a little odd to divide the area by administrative boundaries and not hydrological/geological or climatological zones. Is there a specific reason for this approach?

The nine regions are used solely as a geographic reference for describing spatial patterns in the results and are not used as analytical units in the modeling framework. We chose recognizable geographic regions over hydrologic or climatological boundaries for ease of communication, as the regions are not intended to delineate hydrologically distinct zones. We have added a clarifying sentence noting this in the Figure 1 caption:

Old: “Panel b) shows regional naming conventions for reference.”

New: “For ease of reference in describing results, we divided the CONUS into nine geographic regions (Panel b), which shows regional naming conventions.”

L119: this means that a dry wet-season, would be defined as a drought. In your opinion is this a relevant measure for streamflow drought? Would it not be more relevant to select a threshold based on the dry season? And what about the different predictors of droughts would you not expect these to be different depending on the season they occur. E.g., droughts during low flow season may be very impacted by lower-than normal groundwater baseflow or high PET, while spring droughts could be more related to too little meltwater from mountain ranges or precipitation deficits. The impact of the choice of defining droughts in the way done here in relation to the identified predictors should at least be part of the discussion.

We appreciate this thoughtful comment. Our variable-threshold (day-of-year) drought definition identifies anomalously low flows relative to what is expected for each day of the year, removing the influence of the seasonal cycle. This means a "dry wet-season" is *identified as drought because flows are below the 20th percentile of what is normally observed during that season - not because flows are absolutely low. We believe this is appropriate for our objectives because it captures meaningful departures from expected conditions regardless of season and avoids conflating normal seasonal low flows with true drought anomalies.*

The reviewer raises a valid point about seasonal differences in drought predictors. We acknowledge that the drivers of anomalously low flows likely differ between seasons (e.g., snowmelt deficits in spring vs. high PET in summer). Our random forest models implicitly capture these seasonal differences through the inclusion of date information as a predictor, which allows the model to learn season-dependent relationships. However, we have added a paragraph in the discussion acknowledging that our single-model approach does not explicitly partition seasonal drivers, and that future work could explore season-specific models to better isolate these mechanisms.

Added to end of paragraph 3, Section 4.3: "Our variable-threshold drought definition identifies departures from expected seasonal flows, meaning that anomalously low flows during any season are classified as drought. While this approach appropriately removes the seasonal cycle and captures meaningful departures year-round, it does not distinguish between the potentially different mechanisms driving drought in different seasons. For example, spring drought may be more closely linked to snowmelt deficits or precipitation shortfalls, while summer and fall drought may be driven primarily by elevated evaporative demand or depleted soil moisture and baseflow. Our random forest models partially account for these seasonal differences through the inclusion of date as a predictor variable, but a season-specific modeling approach could more explicitly isolate the distinct drivers of drought in different parts of the hydrologic year."

L120: is there trends in dataset? There must be other studies on this

Added a citation to documented CONUS streamflow trends (Rice et al., 2015) to acknowledge that trends do exist in the record. Our stationary drought definition means that long-term trends would shift the temporal distribution of identified droughts toward the drier end of the record at locations experiencing declining flows.

L126: " Teleconnections data " would be good to add examples of these in () in the text

Added the specific teleconnections indices (AMO, PDO, ENSO, PNA) where teleconnections are first mentioned.

L144: what about groundwater or connectivity to groundwater, are there any predictors covering this aspect?

We did not explicitly represent groundwater as a time-varying predictor for this study due to the lack of a nationally consistent, daily groundwater level dataset at the scale of this analysis. However, groundwater connectivity is indirectly represented through several other variables: NLDAS soil moisture (shallow subsurface storage), 2) baseflow index (BFI) and water table depth (WTDEPAVE) in the PC regression for donor selection (Section 2.4.3), and 3) subsurface flow contact time (CONTACT) is also included in the donor selection variables. We add the following explanation to the manuscript:

Added, line 132: "Additionally, no explicit groundwater level data were included due to the lack of a nationally consistent daily dataset at this scale. Groundwater connectivity is instead indirectly represented through soil moisture and through static basin characteristics (e.g., baseflow index, water table depth) used in the donor selection process (Section 2.4.3)."

Figure 2: agree with reviewer 1. good figure illustrating the workflow, but the location of the figure and the reference to it in L197, seem very strange and as an afterthought. More introduction to the figure should be added.

We agree with both reviewers and have added a reference to Figure 2 at the beginning of Section 2.3 to introduce it as a guide for the full study workflow, rather than only referencing it in Section 2.4. We also added section numbers to the enumeration in Section 2.4 to help the reader connect the text to the figure.

L325: extra space should be removed after "range"

Removed the extra space.

L375: very good and informative to include specific examples

Thank you!

L402: “This is likely indicative of hydrologic memory effects that influence the persistence of drought through sustained dry periods, acting as a critical buffer to transient meteorological anomalies.” I don’t understand this. Should it be the case for low aridity areas more than for high?

We agree that this original phrasing is unclear. The intended point was that in low-aridity regions, when deep soil moisture is depleted, it can sustain drought by resisting a quick recovery even when the short-term meteorological conditions return to normal. This effect is more apparent in low-aridity regions because these basins have greater subsurface storage capacity and longer hydrologic memory. We have revised the sentence to clarify this:

Old: “This is likely indicative of hydrologic memory effects that influence the persistence of drought through sustained dry periods, acting as a critical buffer to transient meteorological anomalies.”

New: “This is likely indicative of hydrologic memory effects in these wetter regions, where deep soil moisture integrates antecedent conditions over long periods and can sustain low streamflow even after short-term meteorological conditions improve. In other words, deep soil moisture deficits in low-aridity regions may prolong drought by resisting rapid recovery in response to transient precipitation events.”

L417: please add numbers to guide the reader: “1. longer duration, less frequent, and less intense droughts; 2. moderate duration, moderately frequent, and moderately intense droughts; and 3. shorter duration, more frequent, and more intense droughts”

Added numbers as suggested.

L509: this sentence is very long, consider rephrasing

Split this sentence into three shorter sentences for clarity:

Old: “The primary motivator lies in the idea that the complexity of streamflow drought, in terms of both the interconnections and overlapping of meteorological drivers, as well as the mitigating and exacerbating factors of the land surface as the signal propagates to the basin outlet, imply a need for a model workflow capable of independent and flexibly-derived predictions.”

New: “The primary motivator lies in the complexity of streamflow drought itself. Meteorological drivers are interconnected and overlapping, and the land surface both *mitigates and exacerbates the drought signal as it propagates to the basin outlet. This complexity implies a need for a model workflow capable of independent and flexibly-derived predictions."*

L513: I don't understand this sentence

Rephrased the sentence to clarify that "independent" refers to the method's sensitivity to heterogeneous, sub-regional drought drivers that regional models might miss.

Old: "Here, the term 'independent' is defined in contrast to regional models that may be insensitive to sub-regional scale, highly heterogeneous, and influential drought drivers affecting individual basins within the region."

New: "Here, the term 'independent' is defined in contrast to regional models that may be insensitive to drought drivers that are highly heterogeneous at sub-regional scales and influential at individual basins within the region."

L530: maybe add here that it makes impossible for the user to understand and evaluate the model performance

Added as suggested.

Old: "That is, the internal functions of a machine learning model are not easily interpretable. This can be an issue in situations where the model may be drawing misleading or inaccurate conclusions from the input data."

New: "That is, the internal functions of a machine learning model are not easily interpretable, making it difficult for users to understand and evaluate model performance. This can be an issue in situations where the model may be drawing misleading or inaccurate conclusions from the input data."

L542: a donor-based?

Rephrased for clarity.

Old: "Finally, we found potential in the dynamic regionalization of a donor-based method for predicting streamflow drought at pseudo-ungaged locations."

New: "Finally, we found potential in a donor-based dynamic regionalization method for predicting streamflow drought at pseudo-ungaged locations."

L543: new line shift before “We conclude”

Added.

Discussion:

Reflections on the impacts of drought definition and the results obtained in this study should be added.

Groundwater is not mentioned a single time in the paper, apart from the mention of baseflow-driven basins and the baseflow-index. How information on groundwater or the geological setting of the basins is included in this analysis is not clear to me. Could you add some reflections on this.

Would be interesting to add some reflections on the future implications of your findings in relation to predicted climate change in the region. What is expected in the region and would we expect the drivers of drought to shift, and how?

Thank you again for your thoughtful review. We have added a paragraph in Section 4.3 (in response to an earlier comment) discussing the implications of our variable-threshold drought definition, acknowledging that it does not distinguish between potentially different mechanisms driving drought in different seasons and that future work could explore season-specific modeling approaches.

We have also added clarification in Section 2.2.2 noting that no explicit groundwater level data were included as time-varying predictors due to the lack of a nationally consistent daily dataset at this scale.

To the point on climate change, we have added a brief discussion of potential future shifts in drought drivers under climate change, noting that the driver typology developed in this study provides a baseline for assessing how drought mechanisms may evolve under warming conditions.

Added at the end of paragraph 3 in Section 4.3: “Under projected climate change, rising temperatures and shifting precipitation patterns are expected to alter the relative importance of drought drivers across the CONUS. For example, increasing evaporative demand and declining snowpack may amplify the role of temperature and PET in the West and Northern Rockies, while changes in precipitation seasonality could shift the dominant drivers in currently precipitation-driven regions of the eastern CONUS. The driver typology *developed here provides a baseline against which future shifts in drought mechanisms could be assessed."*