Comments by Referee #1

RC: I find the toolchain and approach presented in this study to be highly practical for large-scale investigations of drought and vegetation resilience. However, the current methodology is not well-suited for a localized study. The authors have overlooked significant opportunities provided by the specific site, such as ground-truth monitoring, which could have enhanced the robustness of their findings. Given these limitations, I recommend rejection but encourage resubmission following a substantial revision that incorporates site-specific data and methodologies

AC: Thank you for your constructive feedback regarding the methodological scope of our study. We appreciate your recognition of the practicality of our approach and toolchain for large-scale drought analysis, and we acknowledge your concerns about its suitability for a localized context.

Our study was designed to examine drought propagation dynamics across multiple hydrological compartments at three distinct locations along a single stream. With this our goal was to bridge the gap between large-scale drought assessment tools in the form of standardized indices and their application at the local scale. This approach was chosen to ensure comparability across sites while enabling integration with broader regional studies.

However, we fully recognize that our methodology has inherent limitations for a highly localized study, particularly regarding the lack of ground-truth monitoring of vegetation and soil moisture. As stated in the manuscript, the use of MODIS NDVI at 250 m resolution may not capture fine-scale vegetation responses, and the lack of field validation constrains the ecological interpretation of our findings.

The analysis is based on data from three monitoring stations along the Tegeler Fließ, with the focus on local variability and drought propagation within this specific catchment. The time frame (restrained by data availability, compare later comment) and retrospective nature of our study did not allow for in-situ measurements or ground-truth monitoring. Our study's findings highlight significant differences between sites, underlining the importance of local context and limiting direct extrapolation to larger regions without additional data.

We agree that the lack of ground-truthing represents a key limitation, and we discuss this in the manuscript while emphasizing the need for future research to incorporate higher resolution and field-based data to improve the accuracy of drought assessment.

In response to your recommendation, we plan to revise the manuscript to:

- *incorporate higher resolution remote sensing data (e.g., Landsat) where applicable,*
- *include additional land cover information,*
- *expand the discussion on methodological limitations and more explicitly justify the applicability of our approach in a localized setting.*

While our methodology is scalable and well-suited to regional analysis, its current application is indeed localized through the use of site-specific hydrological and hydrogeological data. We will clarify this conceptual distinction in the revised version to avoid confusion and better contextualize the strengths and constraints of our approach.

We thank you again for your thoughtful suggestions and the encouragement to resubmission.

RC: My primary concern is that the authors did not take advantage of the opportunity to visit the field site and use field data to better understand and support the observed differences between the chosen indices and locations. Instead, the manuscript offers speculative explanations without providing the depth of analysis that could have been achieved through direct field observations and data.

AC:

One of the main motivations behind the paper was a series of visits to the field site, when we visited multiple locations in the catchment (as part of a walk along the whole stream). These visits were the inception of the general methodology framework, after experiencing the stark differences between the different landscapes within the catchment.

More visits to the field site probably could have provided opportunities for additional measurements and observations, however we would argue whether they could significantly improve our understanding of the catchment.

The presented methodology is based on the analysis of a multi-year time series. Such a data analysis approach cannot directly improve just by additional field visits, as it would just extend the time series by a single data-point (also out of the investigated timeframe).

Field measurements could have helped to reveal some of the hydrological characteristics of the stream, most notably by measuring the river profile to estimate the stage-discharge relations. In the case of the Tegeler Fliess, this however was not possible at 2 of the 3 existing stations (please see our comments later regarding this issue specifically). We will acknowledge this limitation in the revised version of our manuscript.

To increase our understanding of droughts in the soil, vegetation and groundwater, the installation of additional observation infrastructure would have been necessary. This however was beyond the scope of our study. Additionally, the operation of such instruments over a longer time period would have been necessary to incorporate their results to our workflow.

RC: l375: "In particular, Tegel groundwater experienced a severe 17-month drought from 2019 to 2020, while Schildow showed more resilience in groundwater levels but showed stress in surface water systems". How are the aquifer properties in the different areas which may lead to this resilience? For instance, If the porosity is larger, water level declines are less for the same amount of water extracted.

AC: Thank you for this comment! This was indeed an aspect which we overlooked in our interpretation.

The borehole profiles from the three different areas indeed show a very diverse geological picture, especially the profile in Luebars. The monitoring station in Luebars is filtering a local isolated groundwater body, which is formed by local geological processes. This aquifer is located under more than 10 meter thick shale, suggesting a much weaker coupling with recharge processes as the other two sites.

In Schildow, a similar thick sand layer with some silt is present, that could have a similar, although much lesser isolation effect - which could lead to better resilience. The aquifer at Tegel has no confining, or semi-confining layer above, it is just a sequence of different sand layers. This should indicate a more direct connection to atmospheric processes - which can be a reason behind the weaker resilience.

The most likely explanation for the extended groundwater drought in Tegel after 2019, is that after the initial phase of the drought, any further incoming precipitation was already used up at the soil level. This peri-urban site has a lot of sealed surfaces that obstruct any rain percolation, and also the vegetation in this area is well developed, which could have caused high evapotranspiration.

Unfortunately no porosity logs were available for the three investigated locations, hence we cannot surely say whether it impacts the droughts index calculations, however the main aquifer material is a similar middle sand at each location, which suggests similar porosities. We can still note that the aquifer at Tegel has the most sand, which could have higher storage than at the other sites. This could explain the smaller observed yearly oscillation in the groundwater values.

In the revised manuscript, we will include a detailed hydrogeological description of the three sites, highlighting the differences in their geologies. We will also revise our discussions regarding the groundwater drought results (section 4.1), including the geological diversities as potential explanations for the observed differences.

RC: “In-situ measures of vegetation health could provide more precise insights and supplement satellite-derived indices in these analyses.” While this is a valid suggestion, the study would benefit from a more comprehensive inclusion of land cover distribution for each MODIS pixel analyzed. The variation in NDVI changes is highly dependent on vegetation type. For instance, if the NDVI signal predominantly reflects trees with deep root systems, a summer drought may have little impact on leaf greenness. In contrast, grasslands are likely to show more noticeable NDVI changes under similar drought conditions. Additionally, although the manuscript briefly mentions the influence of agriculture, it lacks specific and quantitative details. It would be beneficial to clarify the proportion of agricultural land within the MODIS pixels and to specify which types of crops are present in these areas. For a study focused on local conditions, this information should be readily obtainable and incorporated into the analysis. For the reader, at least one picture of each study site would be helpful

AC: We thank the reviewer for his thoughtful and constructive suggestion. We completely agree that the interpretation of NDVI signals—particularly in terms of drought impacts—is heavily influenced by vegetation type and land cover structure. Your feedback helped us improve our spatial interpretation of NDVI responses across our research locations.

To address this, we have made two significant revisions to the updated manuscript:

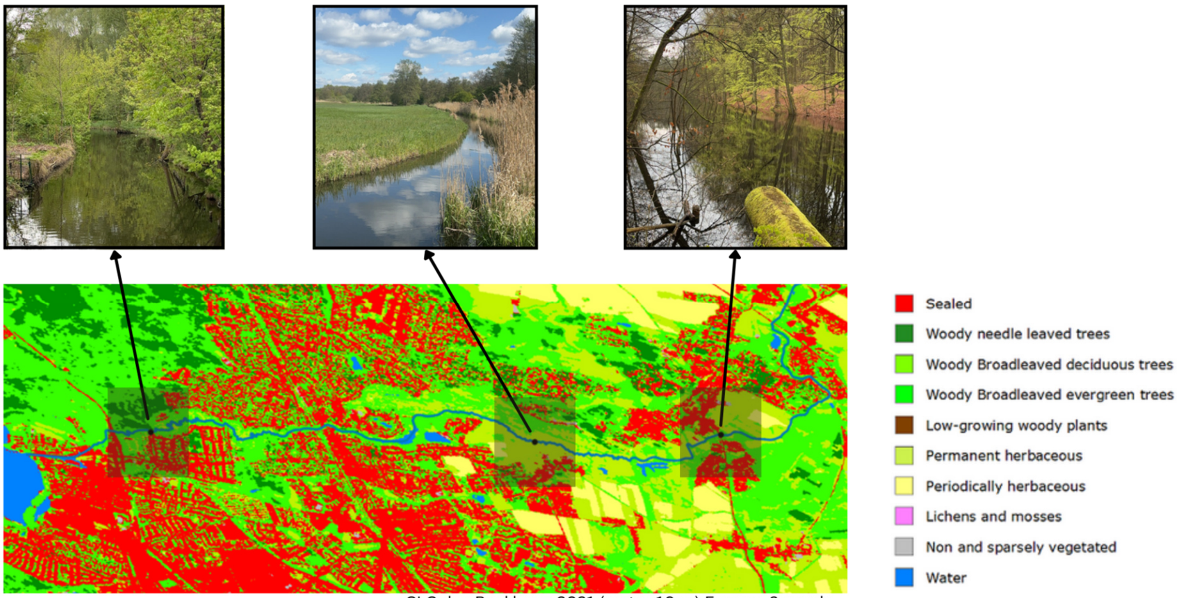
- *Figure below displays a representative ground-level photograph from each study location (Schildow, Luebars, and Tegel). These give readers a better understanding of the topography, dominating vegetation, and land use at each place.*
- *Land Cover Characterization in MODIS Pixels: We assessed land cover within the 500 m buffers for NDVI extraction at each station. This was accomplished using the CLCplus Backbone 2021 dataset (10 m resolution, published by the European Environment Agency), which contains high-resolution raster classifications of land cover across Europe.*

We also used the HCMGIS plugin in QGIS to visually assess high-resolution satellite imagery that was retrieved through the Google Satellite layer in order to support and validate these classifications. We were able to visually evaluate field structures, plant patterns, and built-up regions at each location thanks to this imagery, which is undated and without metadata but provides detailed, cloud-free mosaics assembled from several sources.

Our combined analyses indicated the following:

- *Luebars is mostly characterized by open agricultural fields and shrubland, with little tree cover. Based on visual assessment of field geometry and local land use patterns, these regions are most likely planted with seasonal crops, particularly cereals and fodder grasses, which are widely grown in the Berlin-Brandenburg area. However, as our sources do not contain crop-type metadata, we have treated this as a cautious interpretation and clarified this limitation in the revised text.*
- *Schildow has a more diverse landscape, with small-scale agriculture, tree cover, and patches of fen meadow and semi-natural flora. This structural variety could explain the somewhat buffered NDVI response seen here.*
- *Tegel, being a peri-urban environment, has a mix of vegetation, including managed green spaces, tree-lined urban infrastructure, and proximity to developed regions. Irrigation or municipal landscaping measures are most likely used to manage the plants in this area.*

These findings have been integrated into the revised manuscript, and the discussion has been updated to reflect how land cover variation across MODIS pixels can influence the strength and timing of NDVI responses to drought. We have also emphasized that, while NDVI can be a valuable proxy for vegetation state, it has limits, particularly when not combined with more precise ground-based or high-resolution vegetation structure data.



CLCplus Backbone 2021 (raster 10 m),Europe, 3-yearly

RC: Another critique is the usage of the SSWLI index. As the authors state mention by themselve, that “that this approach has a lower accuracy than traditional discharge based indices, as river levels (especially in smaller streams) are more susceptible to nonlinear behavior due to the river profile”, but they did not provide any estimate of how less accurate the chosen approach is. Considering that all gauges are along the same river, different resilience to drought can only be estimated by comparing discharges due to the impact of cross-section and streambed roughnesses variability at each station (see Manning-Strickler Formula). I suggest conducting a number of discharge measurements at each station and deriving a rating curve (https://en.wikipedia.org/wiki/Rating_curve) to convert the water level estimates to discharges.

AC: Thank you for this comment and constructive feedback regarding the use of SSWLI and the challenge of discharge measurements in our study. We appreciate the opportunity to clarify our methodological choices and provide additional context, including relevant literature and practical considerations.

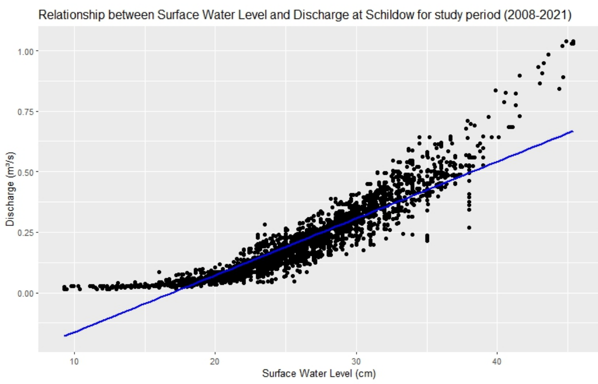
We fully recognize the importance of discharge-based indices for hydrological drought assessment and agree that, in principle, comparing discharges across stations would offer a more direct evaluation of drought resilience. We have considered performing a series of discharge measurements along the stream, but this was technically not feasible. The streambed is very wide near the level measurement stations, creating a very shallow profile and making it impossible to use standard flow measurement equipment. The installation of a weir or other permanent hydraulic structures is also not possible due to the stream's width and local regulations. Furthermore, continuous water level records, as used in our study, provide more robust insights into drought dynamics than isolated, punctual discharge measurements (Tetzlaff et al., 2017).

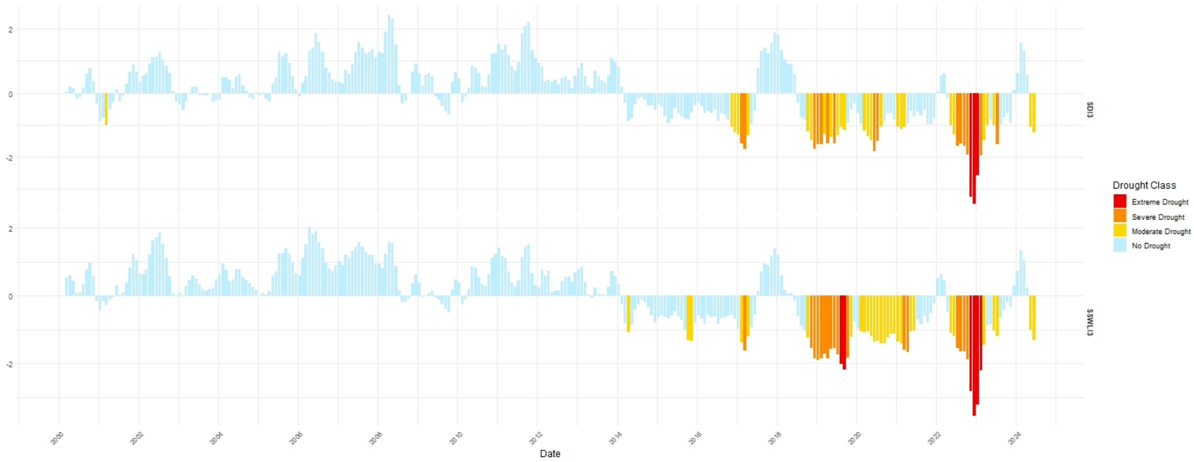
In catchments where discharge data is unreliable or unavailable, studies have used surface water proxies, such as water level, river width or surface water extent (often derived from remote sensing), to estimate flow dynamics, calibrate models or validate model output (Karamuz et al., 2016; Hulsman et al., 2018; de Vitry and Leitão, 2020; Mao et al., 2025). The use of surface water proxies is becoming increasingly common, particularly in ungauged or data-scarce regions. The growing adoption of satellite-based river monitoring further supports the use of level data as a practical alternative (Schwatke et al., 2015).

At the Schildow station, an artificial narrowing (called the Tegeler Fließ waterfall) of the otherwise largely natural course of the Tegeler Fluss was created to enable flow measurements. In addition to the water level, the discharge is also recorded close to this station. We compared the level and discharge data with each other and found that linearity between surface water level and discharge for levels above 20 cm (compare scatterplot below), supporting the use of water level as a reliable proxy in this range. Very low surface water *levels correspond to negligible discharge, which is consistent with hydrological expectations. The SSWLI index is more sensitive to low water levels than the discharge based index SDI. This makes it better suited for detecting drought conditions, especially during periods of minimal flow. Comparing SSWLI and SDI, the SSWLI often indicated more severe droughts, suggesting it is less likely to overlook critical low-water events (compare figure below).*

Our method is designed to be applicable in other catchments and regions where only water level data are available, which is the case for many rivers globally. There are far more stations equipped with water level sensors than with discharge measurement devices. Moreover, observation infrastructure for field measurements is declining over time, both in Germany and internationally (Vörösmarty et al., 2006; DFG, 2013; Rhodes et al., 2015). This trend underscores the need for methodologies that utilize existing data sources, as it is unlikely that new discharge stations will be widely installed in the future. As a result, there is a global shift towards satellite-derived data, which is becoming increasingly available, while in-situ discharge data are stagnating or decreasing.

We acknowledge the limitations of using water level as a proxy and have now provided an empirical comparison at Schildow to estimate the degree of correspondence between water level and discharge. We will include the scatterplot (rating curve) and comparative drought index classifications in the supplements of the revised manuscript to enhance transparency. We will also expand the discussion of methodological limitations and the implications for drought assessment in similar settings.





RC: Based on your map, there appear to be surface tributaries (e.g., Kindelfließ) and a lake (Hermsdorfer Lake) within the study area, which could influence the SSWLI index at the downstream gauges. How might these features impact the observed values? This could help explain the variations in drought conditions observed at each station (Figure 3).

AC:

We would expect a smoothing effect over the observed river level values from these tributaries. This however would be mostly apparent on a higher temporal resolution than our analysis. On this monthly scale we don't expect a strong impact.

For transparency we will mention these factors when discussing the SSWLI in the revised manuscript.

RC: One assumption of your approach is that the dynamics of groundwater and surface water levels can be directly attributed to drought conditions. However, groundwater levels, for example, may also be influenced by anthropogenic factors, such as the lowering of local groundwater levels due to construction activities or changes in the extent of sealed surfaces. Did you consider these potential impacts in your analysis?

AC:

We thank the reviewer for highlighting this important point. We agree that anthropogenic influences such as construction activities or increasing surface sealing can have a significant impact on groundwater dynamics, particularly in urban or peri-urban environments.

To assess this possibility, we used Google Earth to visually inspect high-resolution historical satellite imagery during the whole study period (2008-2021). This enabled us to evaluate potential land use changes and urban growth around the groundwater wells in Schildow, Luebars, and Tegel.

Our observations indicate that:

- *Despite being located in or near urban areas, Luebars and Schildow showed no significant construction activity in the immediate vicinity of the groundwater wells during the study period.*
- *Tegel, located in a more semi-urban region, also revealed no evident indicators of large-scale urban expansion, new infrastructure construction, or extensive surface sealing over the 13-year period.*

Based on this assessment, we concluded that the groundwater dynamics observed at the three sites are unlikely to be considerably influenced by anthropogenic land use changes throughout the research period. This gives us more confidence in interpreting the observed patterns as responses to climatic and hydrological drought stress.

However, we accept that more precise local data (such as subsurface pumping, development permits, or changes in stormwater infrastructure) would provide an additional layer of confidence. As such data were not available for this analysis, we have included a brief statement of this restriction in the revised manuscript to ensure full transparency.

We appreciate the reviewer's attention to this nuance and have included it in our revised discussion of uncertainty and site-specific variability.

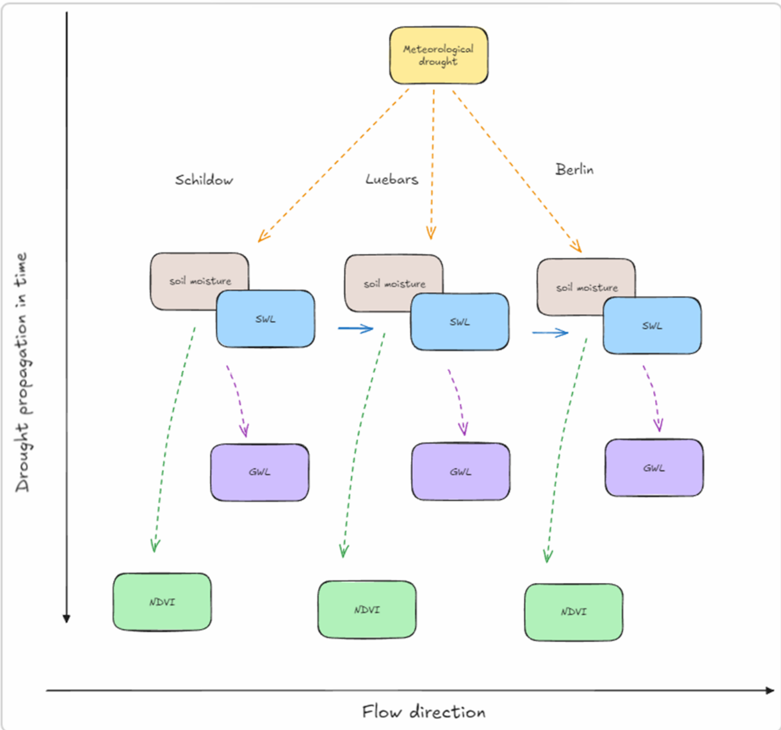
RC: Although the authors reference the drought index from the Helmholtz Centre, they did not analyze soil drought conditions, which are conceptually the link between meteorological drought and groundwater drought. If the authors choose not to include this analysis, they should provide a clear justification for this decision.

AC:

We have based our analysis on data that was available from the catchment, which was limited to surface water and groundwater observations. Soil moisture measurements in Berlin exist, but the available timeseries started after our investigation timeframe, and data only available from other catchments. Drought information for the site from the drought monitor would be based on model outputs and not observations.

Without this data we had to rely on surface water levels as a near-surface drought indicator. This of course impose a limitation on our analysis, which limitation we will acknowledge in the revised manuscript. A comparison of the surface water time series with available soil moisture data nearby (although outside from the catchment) show a similar temporal behavior in general (https://wasserportal.berlin.de).

We introduce a new concept figure in the revised manuscript to clarify the multi-domain drought propagation:



"2.3 Conceptual model of drought propagation

Fig. # presents the conceptual model of drought propagation within the catchment: across different domains and along the river as well. A meteorological drought could trigger droughts at the surface level, via the reduction of soil moisture content and streamflow (hydrological drought). Hydrological drought also propagates downstream along the river. Later in time, reduction in recharge can lead to drought in the groundwater, and the reduced available soil moisture (and in the case of deep rooted plants reduced groundwater table) can also lead to a drought in vegetation.

In the following, we will present how we analyze the drought propagation over the different domains based on the available datasets. We use the conceptual model as a roadmap to try to infer the relations of the different drought types at the different sites."

In the revised manuscript we will also extend the conclusions to emphasize the need for soil moisture observations to fully assess the multi-domain drought dynamics within the catchment.

RC: The study would benefit from a discussion on the applicable scale of your analysis and I feel there is a lack of mentioned studies which did similar works, please provide some more references. Are there limitations to using individual MODIS cells for the indices, considering the sensitivity of values to neighboring cells? Addressing this potential issue could be valuable and could be incorporated into the introduction section.

AC: We thank the reviewer for this thoughtful observation. We agree that a discussion of scale is critical when working with remotely sensed indices-especially in a relatively small and heterogeneous catchment like the Tegeler Fließ.

Regarding MODIS NDVI, we acknowledge that working with 250m resolution data has certain limitations at the sub-catchment level. Particularly in transitional peri-urban settings, each pixel may represent a variety of vegetation types, surfaces, or land use. This creates "mixed-pixel" problems, where signals from developed infrastructure, tree cover, and agricultural fields may be combined in a single cell. Due to this heterogeneity, studies like Vienneau et al. (2016) have demonstrated how coarser resolutions might weaken the NDVI's spectral response in urban and semi-urban areas. Similarly, Ordoyne and Freidl (2008) showed that NDVI varied significantly between plant groups that were only a few meters away, even in small wetlands.

In our case, the three NDVI points selected were deliberately placed in areas where vegetation was dominant within the 500m buffer to reduce this problem. Nevertheless, we now more clearly acknowledge in the manuscript that neighbouring land cover and sub-pixel variation may influence NDVI values and may partially explain the inconsistent response of vegetation to drought conditions in our results.

We also updated the introduction to incorporate citations to pertinent research that has employed MODIS NDVI in catchment-scale drought studies in order to bolster this argument. These include seminal studies (e.g. Tucker 1979; Huete et al. 2002) and more recent research (Philip et al., 2021), which assessed forest drought stress in Germany by combining MODIS NDVI with drought indices like scPDSI and NDWI. We also cite research by Liao et al. (2025), who used MODIS indicators to investigate vegetation lag and drought timing in several climate areas.

RC: L18. The second part of the sentence is an empty phrase, as the planet naturally includes the Berlin/Brandenburg region.

AC: Thank you for pointing it out. We rephrased the sentence as follows to be more specific:

The phenomenon of global warming is having a significant impact on our planet, with observable consequences already affecting the Berlin/Brandenburg region.

RC: L44. SenMVKU reference (and some others) have no year, please update

AC: Thank you for pointing that out. The SenMVKU reference refers to a website that unfortunately does not contain any information about the year in which the information was made available. Following the NHESS submission guidelines examples for reference types for websites, we have therefore added the date at which the information was accessed and written out the abbreviation to make it easier for the reader to understand.

(Berliner Senatsverwaltung für Stadtentwicklung, 2025)

Berliner Senatsverwaltung für Stadtentwicklung (Eds.): Natura 2000-Gebiet Tegeler Fließtal. Retrieved from https://www.berlin.de/sen/uvk/natur-und-gruen/naturschutz/natura-2000/natura-gebiete/tegeler-fliesstal/. Last accessed 11 July 2025.

RC: L46. It is called Helmholtz Centre for Environmental Research and not Helmholtz Institut

AC: Thank you for the important note. We have adjusted the reference accordingly:

Helmholtz Centre for Environmental Research (2023). Drought Monitor Germany. Retrieved from https://www.ufz.de/index.php?en=37937. Last accessed 25 September 2024.

RC: L50 – L 68. I believe there is an overlap with Section 2.3. In the introduction, you should list different indices for precipitation, hydrometry, and groundwater resources, discussing their advantages and disadvantages one by one. The specific details of the technique used in this study can be covered in Section 2.3."

AC: Thank you for this valuable suggestion. We have adapted the paragraph in the introduction along with parts of section 2.3 as follows.

Introduction L50 - L68:

Given its multifaceted nature, a universal drought definition does not exist; rather, it depends on the perspective and objective of a study (Van Loon, 2015). To capture this complexity, a range of indices has been developed for different components of the hydrological cycle (Heim Jr, 2002; Smakhtin and Schipper, 2008; Yihdego et al., 2019).

Meteorological drought is commonly assessed using indices such as the Standardized Precipitation Index (SPI) and the Standardized Precipitation-Evapotranspiration Index (SPEI). The SPI, introduced by McKee et al. (1993), is valued for its simplicity and effectiveness. It only requires precipitation data and is widely used to detect meteorological droughts across various timescales. However, its main limitation is that it does not take into account the effects of temperature or evapotranspiration, which have become increasingly important factors in the context of climate change. In contrast, the SPEI developed by Vicente-Serrano et al. (2010) incorporates both precipitation and potential evapotranspiration, offering a more comprehensive assessment of drought severity. The added complexity and data requirements of the SPEI, however, can restrict its application in regions where such data are scarce. SPI and SPEI are both recommended by the World Meteorological Organization for drought assessment (Svoboda and Fuchs, 2016). Studies show that although SPEI is more accurate due to its consideration of temperature, SPI and SPEI often provide comparable results (Pei et al., 2020; Ohja et al., 2021; Abu Arra and Si¸sman, 2024).

Hydrological drought indices include the Streamflow Drought Index (SDI) and the Standardized River Stage Index (SRSI). The SDI, as described by Nalbantis and Tsakiris (2009), monitors droughts characterised by reduced cumulative streamflow volumes over predefined time intervals. Though its effectiveness depends on the availability of reliable discharge data, which may not be available in certain catchments. When discharge data is unavailable, the SRSI offers an alternative by utilizing level data as a proxy and thereby enabling drought monitoring in ungauged basins (Zhong et al., 2022). It is important to note that water level measurements can be influenced by local channel morphology and may not always accurately reflect actual flow conditions.

For groundwater droughts, which are slow to develop but often take the longest to recover from, the Groundwater Drought Index (GDI) is commonly used. This index standardizes groundwater level anomalies to assess hydro-geological droughts (Bloomfield and Marchant, 2013). While the GDI is valuable for quantifying groundwater drought, its application is often limited by the sparse availability of groundwater data.

Each of these indices provides a different perspective on drought, and their combined use enables a more holistic understanding of drought dynamics across meteorological, hydrological, and hydrogeological domains.

Methods Section:

2.3 Drought and Vegetation Indices Calculation

As outlined in the introduction, various indices exist for assessing different drought types. In this study, we define drought as a period of at least one month during which water availability is significantly below average, resulting in a deficit in its respective domain: precipitation deficit (meteorological drought), reduced surface water levels (hydrological drought), or reduced groundwater levels (hydrogeological drought). To capture these aspects, we applied the SPI for meteorological drought, the Standardized Surface Water Level

Index (SSWLI) as a proxy for hydrological drought, and the Standardized Groundwater Level Index (SGLI) for groundwater drought. Each of these indices was selected to provide a consistent, standardised assessment of drought conditions, facilitating direct comparison and enabling a comprehensive analysis of drought propagation. The following subsections describe the calculation procedure and application of the indices used in this study.

2.3.1 Standardised Precipitation Index (SPI)

SPI standardises precipitation anomalies by fitting a gamma distribution to the precipitation data and then transforming it to a normal distribution. The gamma distribution is chosen for its ability to model the typically skewed nature of precipitation data, especially in semi-arid regions (McKee et al., 1993). We aggregated daily precipitation data into monthly totals to calculate SPI over a three-month period (SPI3).

[rest of the paragraph unchanged]

2.3.2 Standardized Surface Water Level Index (SSWLI)

The SSWLI is calculated by standardizing deviations of observed surface water levels from their long-term mean, following the same statistical approach as the SPI using the SPEI R package, replacing the input with river levels. A three-month accumulation period (SSWLI3) was used to capture short-term hydrological droughts relevant for water resource management and ecosystem response.

Due to the unavailability of discharge data at two of the three monitoring stations along the Tegeler Fließ, we employed the SSWLI as a proxy for hydrological drought. A similar approach has been used by Zhong et al. (2022) for ungauged rivers with satellite data that was able to monitor drought propagation in river systems. By standardizing surface water levels, the resulting index allowed us to assess the intensity of hydrological drought in a way that is directly comparable to meteorological drought (Vicente-Serrano et al., 2012; Bloomfield and Marchant, 2013). Note however, that this can affect the accuracy of hydrological drought characterization, as river levels (especially in smaller streams) are more susceptible to nonlinear behavior.

2.3.3 Standarised Groundwater Level Index (SGLI)

The Standarised Groundwater Level Index (SGLI) extends the SPI methodology to groundwater level data, allowing for the assessment of hydrogeological drought. The SGLI, as applied in this study, is based on the index introduced by Bloomfield and Marchant (2013), which standardizes groundwater level time series to characterize groundwater droughts. This approach aligns with SPI and SSWLI. Standardising deviations from the historical mean allows the SGLI to provide a consistent metric for assessing periods of groundwater scarcity. Similarly to SPI3 and SWLI3, SGLI was also calculated over a three-month period (SGLI3).

Drought classes were used to categorise the severity of drought conditions based on standard thresholds of different drought indices (see Table 2). For this study, we classified drought events according to the standardised values of SPI3, SSWLI3 and SGLI3, following common classification systems in the drought literature (McKee et al., 1993; Nalbantis and Tsakiris, 2009). A drought index value of $\leq$-1 indicates at least moderate drought. This is when conditions are at least one standard deviation below the historical mean and significantly dryer than usual. Values between -1.5 and -2 classified as severe, while values below -2 as an extreme drought.

2.3.4 Normalized Difference Vegetation Index (NDVI)

[rest of the paragraph unchanged]

RC: L73. Provide reference to Granger Causility test

AC: The original work by Granger (1969), which introduced the Granger Causality test (GC), is introduced in greater detail in Section 2.4. However, since the first mention occurs in L 73, we have included the reference there and also provided additional citations to studies that apply GC in drought research within hydrology and environmental sciences:

By combining these drought indices with an examination of long-term trends and conducting Granger Causality tests (Granger, 1969), the research seeks to uncover trends in how droughts propagate and impact vegetation health (Zolghadr-Asli et al., 2021; Li et al., 2024; Xiao et al., 2025).

RC: Figure1: Surface Water instead of Surfacewater

AC: Thank you for pointing it out. We will adjust the figure accordingly in the revised manuscript.

RC: Section 2.2. Here you mentioned the Software “R” first. So here you should reference R and mention, which version did you use

AC: Thank you for pointing that out. We moved the sentence mentioning the software “R” to the beginning of the paragraph and added the version we used, so it is referenced before specific R packages are mentioned further on.

RC: L190: Why did time lag analysis fail? Perhaps you should not mention it here but using it in the discussion section

AC: That's a very good question! An earlier version of the manuscript was meant to include a time lag analysis. With the development of the script, however, we have deviated from this and should adapt that accordingly. Thank you for the important observation. We will remove the mention here in a revised version and address it in the discussion section instead.

RC: Figure3: The plot of absolute surface water levels and groundwater levels for comparing different stations provides limited insight. In Surface Water Hydrology, with a focus on water budgets rather than water levels, it is more informative to compare changes in discharge rather than stage, as stage is dependent on discharge and cross-sectional area. Since discharge data is not available, I recommend normalizing the stages and presenting them in a boxplot or histogram to visualize the distribution across all stations. For groundwater levels, the absolute values provide limited context for the reader. Instead, it is more meaningful to assess the distance from the surface, as this helps evaluate how accessible the groundwater is for vegetation, which is a central aspect of your study. Therefore, I suggest revising Section 3.1 to reflect these points, too.

AC: Thank you very much for your thoughtful and detailed feedback regarding Figure 3. We fully agree with your assessment that presenting normalized surface water levels and visualizing their distributions across stations with boxplots will provide more meaningful insights. We also appreciate your suggestion to present groundwater levels relative to the ground surface, as this will better contextualize groundwater accessibility for vegetation. We will revise Figure 3 and Section 3.1 accordingly in the revised manuscript, implementing your recommendations to improve the clarity and interpretability of our results.

Please note, as a response to multiple other comments, we will also move supplementary figure S2 into the main text, as it provides more detailed insight into the observed surface and groundwater dynamics.

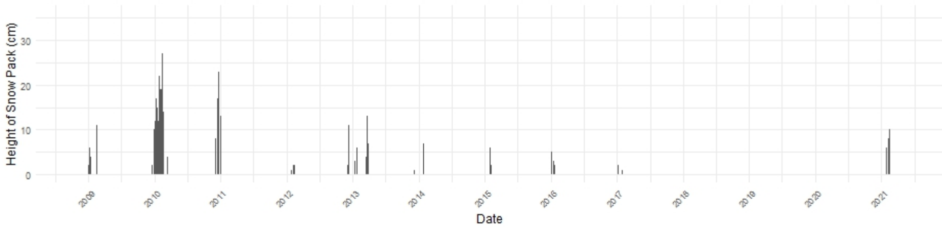
RC: Fig4: All plots should share the X Axis to make temporal comparison more easy.

AC: Thank you for your suggestion. This point was also raised by your fellow referee. We appreciate your feedback and will rearrange the plots in Figure 4 to make it easier to compare the different plots in the revised version of the manuscript.

RC: L285-295: The authors provide several potential explanations for the low NDVI values during the winter months between 2009 and 2011, such as reduced vegetation cover and lower growth rates. I believe the answer is very simple: There was a long lasting snow layer during this years. I recommend checking the DWD snow data to verify this and rephrasing the discussion to provide a more explicit and clear explanation, avoiding any ambiguity.

AC: Thank you for your helpful suggestion regarding the interpretation of low NDVI values during the winter months of 2009 to 2011. Following your recommendation, we examined the DWD snow data for Tegel station. The analysis (see figure below) confirms your observation. There was indeed substantial and long-lasting snow cover at Tegel, particularly in the winters of 2010 and 2011. While there are no direct DWD snow measurements at the stations Schildow and Luebars, the regional climate and landscape characteristics suggest that similar snow conditions likely prevailed at these sites as well. The proximity of the stations and the absence of significant topographic differences support this assumption.

We agree that the persistent snow cover is the most plausible and straightforward explanation for the observed low NDVI values during these winters. We will revise the discussion in the manuscript to explicitly state this finding and remove any ambiguity, making it clear that snow cover was the primary driver of the low winter NDVI values during this period.



RC: L313: “:. v”, improve language

AC: Thank you for pointing out the typo. We have corrected it accordingly.

RC: L379: If you say 13 years is not enough to capture longer-term climate cycles, why did you start in 2008 and end in 2021?

AC: The timeframe of our study was defined by the availability of complete datasets. Even though some measurements date as far back as 1945, the surface water level station at Luebars only started in 2008 and precipitation measurements were discontinued at the site in 2021.

References #1

Abu Arra, A., and Sisman, E. (2024). A comprehensive analysis and comparison of SPI and SPEI for spatiotemporal drought evaluation. Environmental Monitoring and Assessment, 196(10), 980. https://doi.org/10.1007/s10661-024-13127-7

Berliner Senatsverwaltung für Stadtentwicklung (Eds.): Natura 2000-Gebiet Tegeler Fließtal. Retrieved from https://www.berlin.de/sen/uvk/natur-und-gruen/naturschutz/natura-2000/natura-gebiete/tegeler-fliesstal/. Last accessed 11 July 2025.

Bloomfield, J. P., and Marchant, B. P. (2013). Analysis of groundwater drought building on the standardised precipitation index approach. Hydrology and Earth System Sciences, 17(12), 4769–4787. https://doi.org/10.5194/hess-17-4769-2013

de Vitry, M. M., & Leitão, J. P. (2020). The potential of proxy water level measurements for calibrating urban pluvial flood models. Water research, 175, 115669. https://doi.org/10.1016/j.watres.2020.115669

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, 424–438. https://doi.org/10.2307/1912791

Heim Jr, R. R. (2002). A review of twentieth-century drought indices used in the United States. Bulletin of the American Meteorological Society, 83(8), 1149–1166. https://doi.org/10.1175/1520-0477-83.8.1149

Helmholtz Centre for Environmental Research (2023). Drought Monitor Germany. Retrieved from https://www.ufz.de/index.php?en=37937. Last accessed 25 September 2024.

Huete, A. R., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment, 83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Hulsman, P., Bogaard, T. A., & Savenije, H. H. (2018). Rainfall-runoff modelling using river-stage time series in the absence of reliable discharge information: a case study in the semi-arid Mara River basin. Hydrology and Earth System Sciences, 22(10), 5081-5095. https://doi.org/10.5194/hess-22-5081-2018

Karamuz, E., Osuch, M., Romanowicz, R.J. (2016). The Influence of Rating Curve Uncertainty on Flow Conditions in the River Vistula in Warsaw. In: Rowiński, P., Marion, A. (eds) Hydrodynamic and Mass Transport at Freshwater Aquatic Interfaces. GeoPlanet: Earth and Planetary Sciences. Springer, Cham. https://doi.org/10.1007/978-3-319-27750-9_13

Liao, J., Xu, Y., Pi, J., Li, Y., Ke, C., Zhan, W., & Chen, J. (2025). Widespread sensitivity of Vegetation to the transition from Normal droughts to Flash droughts. *Geophysical Research Letters*, 52, e2024GL114321. https://doi.org/10.1029/2024GL114321

Li, Z., Sun, F., Wang, H., Wang, T., and Feng, Y. (2024). Detecting the interactions between vegetation greenness and drought globally. Atmospheric Research, 304, 107409. https://doi.org/10.1016/j.atmosres.2024.107409

Mao, F., Shanafield, M., Ouellet, V., Hannah, D. M., & Krause, S. (2025). Assessing river discharge dynamics through relative surface water extent changes in river basins. iScience, 28(1). https://doi.org/10.1016/j.isci.2024.111598

McKee, T. B., Doesken, N. J., and Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology (Vol. 17, No. 22, pp. 179–183).

Nalbantis, I., and Tsakiris, G. (2009). Assessment of hydrological drought revisited. Water Resources Management, 23, 881–897. https://doi.org/10.1007/s11269-008-9305-1

Ojha, S. S., Singh, V., and Roshni, T. (2021). Comparison of meteorological drought using SPI and SPEI. Civil Engineering Journal, 7(12), 2130–2149. https://doi.org/10.28991/cej-2021-03091783

Ordoyne, C., & Friedl, M. A. (2008). Using MODIS data to characterize seasonal inundation patterns in the Florida Everglades. *Remote Sensing of Environment, 112*(11), 4107–4119.https://doi.org/10.1016/j.rse.2007.08.027

Pei, Z., Fang, S., Wang, L., and Yang, W. (2020). Comparative analysis of drought indicated by the SPI and SPEI at various timescales in Inner Mongolia, China. Water, 12(7), 1925. https://doi.org/10.3390/w12071925

Philipp, M., Wegmann, M., & Kübert-Flock, C. (2021). Quantifying the Response of German Forests to Drought Events via Satellite Imagery. *Remote Sensing*, *13*(9), 1845. https://doi.org/10.3390/rs13091845

Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI–an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. Hydrology and Earth System Sciences, 19(10), 4345-4364. https://doi.org/10.5194/hess-19-4345-2015

Smakhtin, V. U., and Schipper, E. L. F. (2008). Droughts: The impact of semantics and perceptions. Water Policy, 10(2), 131–143. https://doi.org/10.2166/wp.2008.036

Svoboda, M. D., and Fuchs, B. A. (2016). Handbook of drought indicators and indices (Vol. 2). World Meteorological Organization, Geneva, Switzerland.

Tetzlaff, D., Carey, S. K., McNamara, J. P., Laudon, H., & Soulsby, C. (2017). The essential value of long-term experimental data for hydrology and water management. Water Resources Research, 53(4), 2598-2604. https://doi.org/10.1002/2017WR020838

Tucker, C.J. (1979) Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. Remote Sensing of Environment, 8, 127-150.https://doi.org/10.1016/0034-4257(79)90013-0

Van Loon, A. F. (2015). Hydrological drought explained. Wiley Interdisciplinary Reviews: Water, 2(4), 359–392. https://doi.org/10.1002/wat2.1085

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. Journal of Climate, 23(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., and Morán-Tejeda, E. (2012). Accurate computation of a streamflow drought index. Journal of Hydrologic Engineering, 17(2), 318–332. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000433

Vienneau, D., de Hoogh, K., Faeh, D., Kaufmann, M., & Röösli, M. (2017). More than clean air and tranquillity: Residential green is independently associated with decreasing mortality. *Environment International, 108*, 176–184. https://doi.org/10.1016/j.envint.2017.08.012

Xiao, L., Guo, C., Zheng, C., Lu, Z., Khan, M. A., Zhou, J., and Wu, X. (2025). Dense vegetation as a natural buffer: Mitigating and decelerating drought propagation. Journal of Environmental Management, 389, 126027. https://doi.org/10.1016/j.jenvman.2025.126027

Yihdego, Y., Vaheddoost, B., and Al-Weshah, R. A. (2019). Drought indices and indicators revisited. Arabian Journal of Geosciences, 12, 1–12. https://doi.org/10.1007/s12517-019-4237-z

Zhong, R., Zhao, T., Chen, X., and Jin, H. (2022). Monitoring drought in ungauged areas using satellite altimetry: The Standardized River Stage Index. Journal of Hydrology, 612, 128308. https://doi.org/10.1016/j.jhydrol.2022.128308

Zolghadr-Asli, B., Enayati, M., Pourghasemi, H. R., Naghdyzadegan Jahromi, M., and Tiefenbacher, J. P. (2021). Application of Granger-causality to study the climate change impacts on depletion patterns of inland water bodies. Hydrological Sciences Journal, 66(12), 1767-1776. https://doi.org/10.1080/02626667.2021.1944633