Review 1

I'm glad to have reviewed the manuscript. My opinion is that major revisions are necessary. Lack of references about important topics that authors discussed in the introduction is present. I have raised to the authors some conceptual questions, especially about the real goal of this study and the related methodology, which is still confused in the statement, from the abstract above all. Moreover, even if I appreciate the great efforts in the pretreatment and management of such amount of data with innovative methods, I think that at the local scale authors should accept the difficulty of the method in providing reliable information (and state it clearly in the text). Despite this, I believe that this work can be an excellent starting point for a computerized and indexed database of groundwater hydrographs on a global scale, useful for the first steps of groundwater resources management in coastal areas.

> We would like to thank reviewer 1 for the positive assessment of our study. We will respond to the valuable suggestions and requests for more clarity about our objectives, and the potential and limitations of our study, as detailed below in our responses (green). We also appreciate suggestions for improvement and corrections to individual words in the text, which we will carefully implement. ✓

Abstract

[Line 11]: Even if it's obvious, please, at least in the abstract and in the introduction, write the full word the first time you use the abbreviation GWL.

> We agree and will check the manuscript to ensure that abbreviations are always written out first. ✓

[Lines 13-17]: In my opinion this part of the abstract should be reviewed. The idea is maybe to provide to the reader a little preview of results obtained, but it is unclear because patterns, as well as site characteristics and factors, are still unknown. I don't know if it is better to say more or directly remove this part, focusing on the following one, when you discuss what your results mean in terms of GW management

> We understand that more background information on the complex methodology would be needed for the preview of the results in the abstract, so we plan to delete this part as recommended and instead focus more on the meaning of our findings. ✓

Introduction

[Lines 24-25]: Considering the following sentence and the sensitive topics listed, I would suggest to insert some references about recent studies. Especially the SWI is important because climate change and overexploitation are both responsible of groundwater quantity and quality degradation in coastal aquifers. Some suggestions here:

https://doi.org/10.3390/w11122467

https://doi.org/10.1038/s41598-020-66516-6

https://doi.org/10.7343/as-2019-373

https://doi.org/10.1016/j.scitotenv.2022.160697

https://doi.org/10.1016/j.watcyc.2023.05.002

https://doi.org/10.3390/w14152358

Thank you for this valuable recommendation and the references provided. We will add references to the topics listed.✓We added some of the references,

[Line 35]: Why do you talk about submarine processes? Until now, all the groundwater resources have been considered. You seem to be implying coastal aquifers despite not having mentioned them thus far.

We agree that the start of the introduction focuses on GWL dynamics analyses and assessments of all groundwater resources while the data set comprising the coastal aquifers is introduced further down. To talk about submarine processes is misleading and not of value here and will be removed. ✓

[Lines 67-69]: I don't understand (maybe I miss something) why do you choose a RF approach instead of other methods. Try to explain why and add some recent references on similar studies using RF on similar applications.

The RF approach is a robust choice for classifying groundwater dynamics, widely acknowledged in water science. With its capacity to capture non-linear dependencies and manage uncertainties such as unknown feature importance, overfitting, and outliers, RF provides a reliable tool for exploring complex interactions in natural processes (Tyralis et al., 2019). We will add an explanation accordingly. Furthermore, we plan to move the sentence "Such approaches of explainable machine learning have been rarely used, but are increasingly and successfully applied in hydrology (Worland et al., 2019; Yang and Chui, 2021; Wunsch et al., 2022; Liu et al., 2022; Haaf et al., 2023)." (lines 202-204) from the methods to the introduction section, to already provide an explanation for using SHAP values in the introduction. ✓

Methods

[Lines 94-95]: Even if you cite a definition from these two studies, I'm not convinced about this threshold. It seems to be not applicable for many coastal areas where mountain chains are not far from the sea. Some examples of these conditions are mostly in the Mediterranean Region. Is this your criteria to distinguish coastal aquifers from inland aquifers?

We thank the reviewer for raising this question. We agree that there are many possible definitions of coastal groundwater. We will clarify that our selection criteria focus on the environmental and economic importance of groundwater and management zones in a broader sense and do not aim to select aquifers that are directly associated with submarine processes such as SWI. ✓

[Line 102]: 1979-2019 ???

We will clarify: Data must be available for at least four complete calendar years within the years 1979 to 2019. ✓

[Line 147]: "hydrogeology" is not a bad word :) Since we are talking about groundwater, I expected to find it at the top of this list...

We of course agree and will place hydrogeology at the top of this list. ✓

[Lines 154-155]: The subsurface catchments are not easily definable with satellite or DEM data such as in the case of river catchments. It is "an invisible world" for which most of times we are obliged to find different methods and combine them (geological, hydrogeological, geophysical and chemical).

This is just to make it clear that creating a global dataset with the same precision as hydrographic data is very difficult. I think this should be focused more on your discussion.

> Thank you for this valuable recommendation. We already address this difficulty in the discussion and will check whether this discussion needs to be expanded. ✓ Sentence added to the discussion

[Lines 158-159]: I don't understand if you mean a specific decision of defining the domain of the study or if you simply are talking about the radius of influence of a pumping well. Please, explaine better.

> We are talking about taking a radius of the potential influence of the well which we will clarify. ✓

[Lines 164-174 / Table 1]: Please, insert this describtion in the text as it is a materials and methods section. The caption is too long and misleading in this way...

> Thanks for this suggestion. We will adapt the methods section accordingly. ✓

[Table 1]: please add also the resolution in km

> Thank you - we will. ✓

Results

[Figure 1]: Please don't start this section directly with this plot, but insert in the text after having explained it.

> We will rearrange the figure and the text accordingly. ✓

[Line 271 | Line 272]: BFS? Base Flow...? |

Base Flow Index I guess.. please state everything the first time you cite it.

> Thank you for pointing this out. We are talking here about the Base Flow Stability (BFS) and the Base Flow Index (BFI) as introduced by Heudorfer et al. 2019. We will make sure to state everything the first time we cite it. ✓

[Lines 324-327 | Lines 334-338]: these results were rather obvious because it is the hydrogeological nature of the contexts in which groundwater is present that makes the difference, anywhere in the world. Having reached this point of the study, which from the point of view of data collection and computational efforts is truly commendable, more is expected however from the point of view of the interpretation of the phenomena that occur or possibly of the different managements in the world and their effects on this data. |

Very Good! This is what I expected in two comments above. This should be your style and the basis of the entire discussion

> We agree that the results the reviewer is referring to are rather obvious, however important to mention for the completeness of the results. Further below in the results section, we break down the results further which is what reviewer 1 expected. ✓

[Lines 330-333 | Lines 345-348]: Please rephrase this sentence. It's measliding and confusing |

misleading sentence

> Thanks for pointing out the missing clarity in these sentences. We will revise the sentences. ✓

Discussion

[Lines 365-366]: Heterogeneity matters

We might include this correct statement in a revised version of the sentence. ✓ We did!

[Lines 367-369]: This is very important. You should stress more on this fact, also stating the ambition of this study in the abstract and in the introduction section.

Reviewer 1 is talking about our results including spatial patterns at local, regional, and global scales. We agree that we should state this ambition more clearly in the abstract and introduction section and we will revise these sections accordingly including the research questions stated. ✓

[Lines 383-384]: Very important!

Thank you. We agree. ✓

[Lines 388-390]: longer residence time in the aquifer? Or there is something else?

Yes, the cited findings from a previous study are referring to longer residence times. We will check the sentence for clarity. ✓ We have rephrased the sentence accordingly.

[Lines 397-399]: This aspect still remains the most important in the management of coastal aquifers, because of the combined effect of the increase in mean sea level and overexploitation due to growing urban areas along the coast.

Thanks for highlighting the importance of the described aspect. We will incorporate what our findings mean with regard to water management because reviewer 1 asks for this interpretation also in the abstract (see comment [Line 13-17]) and further down in the discussion and conclusions). ✓

[Lines 442-445 | Lines 514-517]: Can we understand if groundwater quantity is similarly affected in different coastal areas? Or differences are too low to be detected by this study? Are natural controls the total predominance? |

My question has now an answer. It is still a challenge. I hoped that this study could be more helpful in this sense.

Thank you for raising these important questions to which reviewer 1 found an answer at the end of the discussion section. Because there is a lot of text between the questions raised and the explanation, we see the potential to improve the clarity of the discussion by merging the passages and thus minimally restructuring the discussion. ✓

[Lines 450-451]: yet it is a theme that is quite relevant in the analysis of GWL time series. Several studies have shown the degradation of groundwater in quantitative and qualitative terms, with evident decreasing patterns. I wonder if the management of the indices in your study may have influenced the visibility of this phenomenon in some way.

https://doi.org/10.1016/j.jhydrol.2021.127238

https://www.mdpi.com/2073-4441/10/2/143

https://link.springer.com/article/10.1007/s10040-021-02448-3

Because of the question raised by reviewer 1, we plan to improve the discussion by stating how GWL dynamics classification results depend on the input data chosen (time series and

indices derived therefrom). Classification of specific groundwater processes requires input data specifically related to and known to be influential to these processes. To accurately predict the impact of climate change on GWLs without incorrectly accounting for water withdrawals remains a challenge in the field of machine learning, which is of great importance for sustainable groundwater management. More in-depth analyses of the individual indices would be needed to find out to what extent and in which indices anthropogenic influence is manifested and thus contributes to a categorization. However, this was outside the scope of our study. Furthermore, we already stated in the discussion, that accounting for SWI in GWL dynamics pattern analysis is best supported by pattern recognition or correction with groundwater chemistry, and high-resolution time series are generally required for analyzing the interaction of groundwater with the sea. We will also add some references to underline the effect pumping activities can have on groundwater quality and quantity. ✓

[Lines 477-479 | Lines 510-513]: We are therefore moving from a global to a local scale. At this point I would like to understand whether bringing a case study of which you are aware only serves to "calibrate" the reasonableness of data processing at a global level (in my opinion it is not enough) or whether, implicitly, it is being stated that in the end we must always study hydrogeological contexts at a local level, to understand their deeper dynamics and suggest management measures. |

This is the hydrogeology world! :) Most of times we must interpret what there is below us...I keep believing that this kind of studies, based on treatment of big data can be helpful for tracing general behaviors and identifying patterns on some effects of human actions and policies to be implemented, but they cannot and do not manage the deeper dynamics of hydrogeology, which often remain obscure even after in-depth analyzes local scale.

We appreciate the opportunity to clarify the purpose and significance of the presented case study in Northern Germany. The inclusion of this case study serves a dual purpose in our research. Firstly, it allows us to assess the robustness and validity of our proposed indices-clustering approach on different scales. By applying our methodology to a region with available hydrogeological maps and a dense well network, we demonstrate how data-driven approaches can detect (dis)similarity in groundwater observations. Secondly, the case study serves as a critical testbed for understanding the limitations and nuances of extrapolating explanations for patterns of GWL dynamics from the global to the local and regional scale. We acknowledge the importance of local hydrogeological context in providing a more nuanced interpretation at the finer scale. The case study is instrumental in highlighting the need for a balanced perspective, recognizing the strengths of data-driven global analyses while acknowledging the essential role of local hydrogeological knowledge for a comprehensive understanding of groundwater dynamics. We will clarify these aspects accordingly in the first paragraph of the chapter to ensure a better understanding of its purpose. ✓

Conclusions

[Lines 523]: hydraulic conductivity

With your comment, we have realized that this sentence is misleading. We will revise it and add that hydraulic conductivity is most important in distinguishing parts of the GWL dynamic patterns encountered. ✓

[Lines 526-527]: Stress more on this, please

Similarly to the comment in lines 367-369, reviewer 1 asks us again to stress more on the aspect that similar patterns are observed across different environments and climates globally, while specific cluster compositions vary among regions, suggesting complex interlinkages of controlling factors. We agree that this aspect is the focus of our study and also needs to be made more clear in the discussion and conclusions sections. We plan restructuring and additions to content in discussion and conclusion to better emphasize this aspect. ✓

[Lines 526-527]: I wouldn't go so far as to say this, it's risky to think like this

We agree and will remove pointing out the potential for unmonitored sites. ✓

[Lines 526-527]: I would say that an approach and methodology like that can be extremely useful as a starting point for then deepening studies at a local level on the specific coastal aquifer. What I appreciated most about this study is the desire to try to connect many datasets, all around the world, to try to provide a sort of global information archive on the dynamics of GWLs in coastal areas. I appreciated so much this efforts. I would focus much more on this aspect. I believe that, at an international policy level, it could be challenging for some large entities to provide a service of this type. But I would not go beyond the general interpretation, because at a local level the geological and hydrogeological characteristics can become difficult to interpret, even by the best statistical or machine learning model.

This comment is closely related to the comments of reviewer 1 in the discussion (lines 477-479 and lines 510-513), to whose response we refer. We agree with what reviewer 1 wrote in this comment and are very grateful for the appreciative and clarifying words. ✓

Appendix

[Figure A1 /DBSCAN methodology]: So, what is the solution in these cases? Do authors mean that some outliers are not possible to be detected in this way, using the DBSCAN?

Yes, like other tested outlier detection methods, DBSCAN does not allow us to detect all types of outliers and anomalies that we would expect should be removed to represent undisturbed GWL dynamics. With the parameters set, DBSCAN allows us to successfully detect density-dependent outliers, but many values are also incorrectly identified as outliers. Therefore, we also performed a visual inspection of all time series where DBSCAN identified potential outliers and used indices as additional quality checks. We will review the methods section and the figure caption for clarity. ✓

[Figure C6]: Is there a well in the sea? The red one in the left bottom area?

Thank you for pointing this out. It only looks like this because the wells have been minimally displaced to avoid overlapping wells being not visible. We have described this in the figure caption, but we will use a slightly smaller spacing for displacement to avoid wells in the sea and thus confusion. ✓

Review 2

The study is quite well structured and the approach used modern and innovative although it requires further bibliographical references with respect to the topics involved. I believe major revisions are needed to emphasise the real purpose of this study and the limitations it has in dealing with a global-scale dataset. This approach is valid in attempting to categorise groundwater level patterns, but it cannot be a tool for coastal aquifer management planning at the local scale, where geological, hydrogeological, structural and resource exploitation characteristics as well as climatic conditions influence the hydrogeological behaviour of the aquifers. I believe that by revising the article in this sense, it can be a valid starting point for the categorisation of GWLs at a global scale.

> We would like to thank reviewer 2 for the positive feedback and valuable suggestions and requests for more clarity about our objectives, and the potential and limitations of our study. We will revise our manuscript accordingly. Below, we respond (green) to individual comments made by the reviewer. We appreciate suggestions for improvement and corrections to individual words in the text, which we will carefully implement.

Abstract / General comments

Please, verify in the document to state everything before using acronyms. | [Line 11]: Please, provide the full word the first time you use GWL

> We agree and will check the manuscript to ensure that abbreviations are always written out first. ✓

Captions of figures and tables are often too long. Try to summarise them and include this information in the text

> Thanks for pointing out. We will check where descriptions from the captions should be in the text. ✓

Please, revise the abstract in order to better explain the main outcomes and limitations.

> We agree that the preview of the results in the abstract might be too ambitious and instead plan to focus more on the meanings of our findings including limitations. ✓

Methods

[Line 102]: in section 2.1 you stated that the dataset was compiled from 2019-2022. What does it means? please, clarify and revise

> In Section 2.1 we describe the period in which we had access to the datasets used, while here we describe the period from which we selected time series from the dataset. We will clarify accordingly. ✓

Results

[Lines 183-184]: Please explain this criterion and add a reference

> We will provide a short explanation of the criterion in the methods section. ✓ We already provided a reference in the methods section.

[Lines 330-332]: please, rephrase it

> Thanks for pointing out the missing clarity in the sentence. We will revise the sentence. ✓

Discussion

[Lines 383-384]: not only from a global perspective but also at local scale, affecting the establishment of an efficient monitoring and management strategy

We agree and will rephrase the sentence accordingly. ✓

[Lines 398-399]: in coastal aquifer, the qualitative characterisation is quite important and needs to be coupled with the quantitative one, especially in arid and semi-arid regions.

https://doi.org/10.1016/j.scitotenv.2022.160697

We thank reviewer 2 for this supporting formulation and the added reference. We plan to add to the discussion a more direct statement of how GWL dynamics classification results depend on the input data chosen (time series and indices derived therefrom). For further information, please refer to our response to reviewer 1's comment on lines 450-451. ✓

Conclusions

[Lines 529-530]: I feel that it can be a starting point but that groundwater management requires specific studies on a local or regional scale. There are complex systems that are often unequalled in the world and for which any form of large-scale generalisation may be a limitation rather than an advantage.

Thank you for your thoughtful comment. We acknowledge the importance of detailed, site-specific investigations, recognizing the complexity and uniqueness of various hydrogeological systems. The intention behind our approach is not to replace localized studies but rather to complement them. By analyzing data from diverse coastal aquifers around the world, we aim to identify hydrogeological (dis)similarities that can inform broader frameworks for groundwater management. As both reviewers have also pointed out in the main comment, we believe that we need to clarify our study objectives in the abstract and the introduction. Furthermore, we plan to provide more focus in the conclusions on how our global perspective can support groundwater management on various scales via a) identifying regions that may share common characteristics or face similar challenges, b) identifying hydrograph characteristics that are important to consider in global modeling frameworks, c) providing evidence on the currently limited possibilities to explain GWL dynamics using attributes that are available for the global scale, d) while recommending to make use of self-learning algorithms to better understand and predict GWL dynamics beyond the local scale. We also refer to our response to reviewer 1's comment on lines 477-479 and lines 510-513 regarding the case study. ✓ We have adjusted all named sections as indicated.

[Lines 534-535]: too strong as a statement, see previous comment

We agree and will remove pointing out the potential for unmonitored sites. ✓