Dear Ralf Loritz,

Following our recent peer review (Referee Comments RC1 and RC2 on manuscript ID egusphere-2025-5666), we are pleased to resubmit the revised version of our work entitled "Spectral Analysis of Groundwater Level Time Series for Robust Estimation of Aquifer Response Times."

We sincerely thank both reviewers for their time, thoughtful critique, and overall positive evaluation of our manuscript's scientific merits. Their feedback has been invaluable in strengthening our analysis and clarifying key technical points.

In response to the Referee concern about methodology value, we clarify that while SPI/SGI correlation captures vadose zone lag, our spectral method isolates the saturated aquifer's intrinsic buffering capacity. Using mHM-derived recharge we eliminate unsaturated transport delay and compute tc directly tied to physical properties T and S—distinguishing it from accumulation times which signal drought onset vs. tc determining recession rate and recovery potential.

Furthermore, we have addressed all other reviewer concerns:

Recharge: Detail mHM’s three-layer model origin and limitations of localized assumptions.

Trend removal: Acknowledge it suppresses low-frequency power but ensures stationarity.

Resilience: High tc enables sustained abstraction during dry years—yet caution against universal resilience due to long-memory depletion risks in prolonged Central European droughts.

Specifically, we have rewritten and expanded the Methods sections and the Discussion section to directly address all technical points raised by both reviewers.

We believe these revisions significantly strengthen our manuscript's scientific rigor, transparency, and impact on groundwater management practice.

Thank you again for the opportunity to submit this work for publication in Earth System Science Data.

Sincerely,

Timo Houben on behalf of all co-authors timo.houben@ufz.de Helmholtz-Centre for Environmental Research - UFZ Computational Hydrosystems

Author’s response to:

RC1: 'Comment on egusphere-2025-5666', Anonymous Referee #1, 27 Dec 2025 reply

Summary

The manuscript by Houben et al presents a methodology that combines time series analysis, GIS and analytical modelling to estimate the average response time of aquifers in a large, regional data set of groundwater level time series from observation wells in the upper Danube river basin. The approach rests on existing findings by (i.a. Houben et al., 2022; Liang & Zhang; Zhang & Schilling, 2004) regarding temporal scaling of groundwater head and relation to aquifer geometry, properties and recharge. Here, the focus was on estimating at each location the characteristic time scale, a single value that quantifies the rate at which an aquifer responds to an external stress. This value is then briefly framed as a characteristic to quantify aquifer vulnerability to drought and criteria for selection of aquifers for abstraction. While the approach generally is sound, my primary concern with this paper is **that the value of the analysis is not apparent when compared to other studies that require fewer data, assumptions, and less effort, such as the referenced study by Kumar et al. (2016) or Ebeling et al. (2025). What benefit does the characteristic time scale provide versus the current standard method in groundwater drought analysis using correlation times of SP(E)I vs SGI for example or e.g. cross-correlations (e.g., Bloomfield & Marchant, 2013; Ebeling et al., 2025)?** I think the authors need to reflect on - and clarify this before the manuscript can be considered for publication. Below find specific comments:

Thank you very much for the invested time, the valuable comments on our manuscript and the overall positive scientific evaluation! In the following, we answer the comments as precisely as possible and incorporate the overall motivation and value of the study and its approach. Please find our answers in italics below the respective comment and corresponding line references to adapted text passages in the track changes document.

1. The recharge data that is used needs more description for readers not familiar with mHM. What does it represent within the model context, is it the input to a groundwater component?

   Corresponding text adaptations at section: Time Series Preparation

   Thank you very much for the hint. We will include the following explanation in the methods section of the revised manuscript, where the modelled recharge product is mentioned:

   In the mHM model three subsurface storages are considered to simulate (i) fast interflow q2, (ii) slow interflow q3 and (iii) baseflow q4. Fast and slow interflow originate from the unsaturated zone, an upper storage, while the baseflow is fed by a lower saturated storage. The flow between the upper and the lower storage is estimated as a linear reservoir and is called groundwater recharge or percolation (Figure 2.3 from Kumar et al., 2010). In our study, a recharge product was taken were parts of q3 have been *attributed to the recharge as part of a bias correction in order to match estimated recharge volumes provided by the German Hydrogeological Atlas (von Jangiewicz et al, 2005, Marx et al., 2021).*

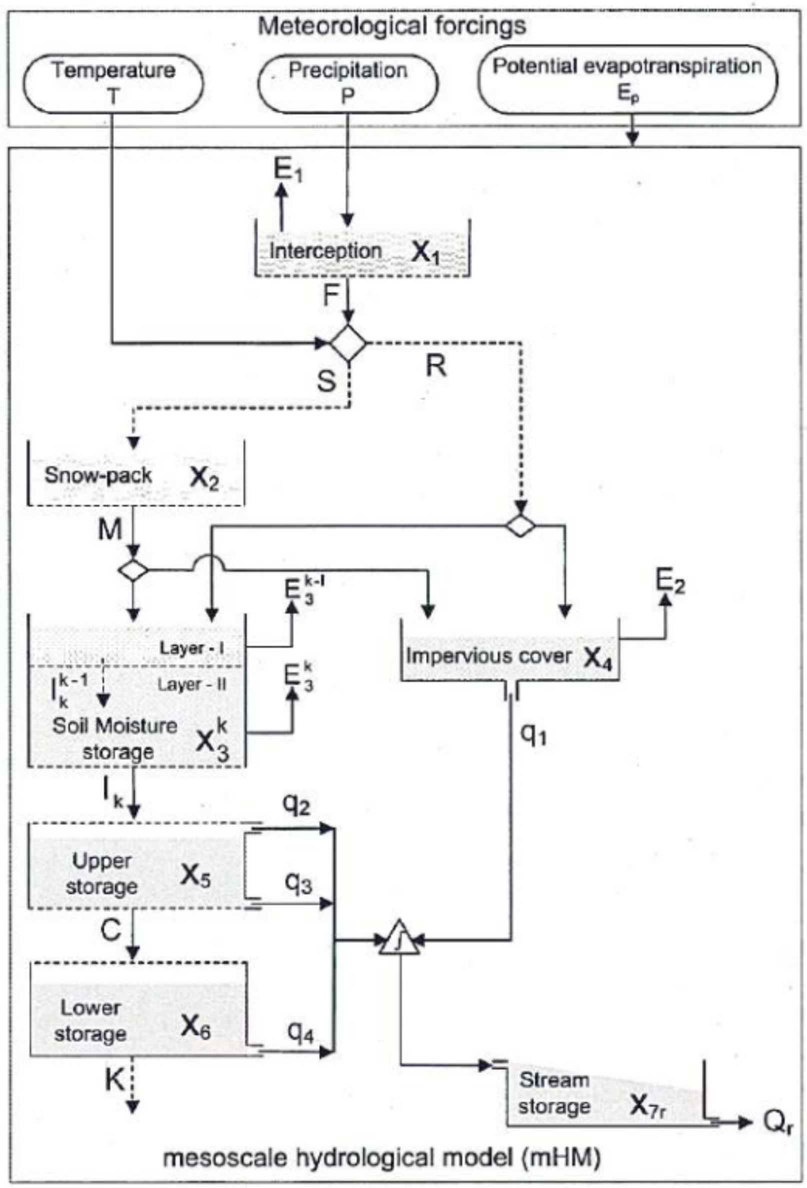

Figure 2.3: Schematic representation of different components accounted in mHM. Where, X ≡ state variable, E ≡ actual evapotranspiration, q ≡ component of runoff, S ≡ snow precipitation depth, R ≡ rain precipitation depth, F ≡ throughfall, I ≡ infiltration capacity, C ≡ percolation, K ≡ gain/loss flux in a leaking cell, Q_r ≡ net runoff produced at the outlet of a grid cell.

2. Is the recharge calibrated on observation data (lysimeter) or is it the residual from runoff and soil moisture? This distinction is important for understanding the reliability of the input data.

   Corresponding text adaptations at section: Time Series Preparation

   The mHM model was calibrated on observed stream discharge in the context of a multi-basin model calibration strategy, where small subsets of the total 201 basins all over Germany were calibrated individually and later evaluated against the full ensemble. The best performing parameters in terms of the median daily KGE over all 201 basins during a period of 1986-2005 (Boeing et al. 2022) were finally chosen for the model runs. In addition, the model was validated against soil moisture observations from 40 sites in Germany originating from single profile measurements, spatially distributed sensor networks, cosmic-ray neutron stations and lysimeters. High correlations in the vegetation active period (0.84 median correlation R) and lower in winder (0.50 median R) was achieved (Boeing et al. 2022).

   As mentioned in the previous answer, the recharge (C in Figure 2.3 from Kumar et al., 2010) is a residual from an upper storage and represented as a linear reservoir in the form:

$$C(t) = \beta_{22} X_5(t-1)$$

   Where X_5 is the state of the upper storage (water content), β_{22} is the percolation rate which is composed of the arithmetic mean $\mathcal{A}$ of the globally calibrated parameters γ through MPR and the saturated hydraulic conductivity $\mathcal{K}$ after Liang et al. (1994).

$$\beta_{22i} = \mathcal{A}\left(\gamma_{46}\left(1 + \left\|\mathcal{K}_j\right\| i\right)\right)$$

   This additional information will be provided in the revised manuscript.

3. Given the difficulty of estimating recharge on short temporal scales my expectation would be that this is a significant uncertainty for the study. Discuss the challenges and uncertainties associated with estimating recharge at short temporal scales, and how these uncertainties may affect the study's conclusions.

   Thank you for pointing that out. We agree that the estimated recharge introduces uncertainties to our study, but we evaluate the uncertainty of short temporal scales as minor compared to mid or long temporal scales. Short term recharge fluctuations (shorter than ~7 days) impose the spectrum to rise its right part, leading to less congruent fit of measured groundwater head and fitted spectrum. As long as this does not impact time scales for the mid to low frequency range (centrum and left part of spectrum) this will not influence the obtained hydrogeological parameters since the fitted spectrum remains on the same place (more weight has been given to the mid and low frequency part). As soon as the fitted spectrum remarkably deviates from the observed spectrum in the mid frequency range (30 – 90 days), we assume an increasing impact on the obtained *parameters. We categorize these cases as 'underestimated', potentially leading to higher or longer response times, since low frequency components are underrepresented in the fitted spectrum.*

Moreover, limitations of the complexity of the semi-analytical solution in representing complex slopes of the spectrum make it impossible to match the spectrum across the whole frequency range, especially when multiple slopes are present. Only the category 'good fit' might capture the precise characteristic response of the aquifer, although we consider the category 'underestimation' as still suitable for longer time scale estimations.

The present study and the categorization into four distinct goodness of fit groups provides the possibility to evaluate the spectral fit and consequently the uncertainty of the derived parameters due to the underlying recharge.

We understand that this should be emphasized throughout the results and discussion part. We will integrate the above elaborations in the reviewed manuscript.

4. Also related to recharge, in L150 you state that the recharge is extracted from multiple cells and spatially averaged. Why and how?

   Corresponding text adaptations at section: Time Series Preparation

   We averaged recharge over multiple grid cells to generate a regional signal that reflects the effective aquifer properties across a broader area. This approach assumes that the groundwater level represents a regional signal driven by an averaged, regional recharge connected laterally and hydraulically through the saturated zone in contrast to isolated soil-moisture variations in the vadose zone.

   Because the exact extent of each hydrological basin and the underlying hydraulic network are unknown without detailed watershed analysis, we adopted a standardized kernel approach. We applied a kernel size of 2, meaning that the recharge values of the central cell (where the well is located), its 8 immediate neighbors, and the 16 second-ring neighbors were averaged. Thus, each calculation incorporated 25 cells, each roughly 1 × 1 km, covering a 5 × 5 km (25 km²) area.

   Figure A3 (appendix) in the manuscript shows the resulting aquifer lengths (L) of the flow line length estimation. For the spectral-analysis parameter set used in the paper (t = 1000, fl = direct, s = 2, green in Figure A3), the maximum flow-line lengths (L) are around 5 km. Therefore, a kernel size of 2 was considered appropriate.

   Figure 2b (top) in the manuscript briefly illustrates this process. In the revised manuscript, we will add the above explanation to the methods section and reference Figure 2b (top) explicitly.

5. This also assumes localized recharge for all wells, which is a standard approach in 1D modelling approaches. You are however using a 2D conceptualization, where the

standard perceptual models consider recharge/intermediate/discharge zones (e.g. Tóthian basins or Winter's hydrologic landscapes) Clarify the assumption that localized recharge is appropriate for all wells. Discuss how this assumption fits and acknowledge any additional uncertainties introduced.

Corresponding text adaptations at section: Discussion - Regional Flow Connectivity and Temporal Recharge Uncertainty. A new chapter was introduced discussing the uncertainties of the applied recharge.

Thank you for pointing this out. Certainly, the localized recharge approach is an assumption that does not hold for every hydrogeological setting. This assumption was already introduced by Liang and Zhang (2013) as prerequisite for the stochastic derivation of the semi-analytical solution of the head spectrum. Their model considers a 2D-groundwater surface according to the Dupuit assumptions but a spatially consistent recharge rate. According to the assumptions of their derivation, we consequently were tight to these when we applied the solution. We searched for a way to find reasonable recharge estimate, representative for the full hillslope and finally developed the spatial averaging workflow explained in the previous answer.

Tóth (1963) distinguishes between three types of flow systems in small basins: local, intermediate, and regional. Within these systems, recharge and discharge areas alternate, meaning only specific portions of a hillslope contribute to the discharge of the main stream. Consequently, only a fraction of the recharge calculated for the entire hillslope may ultimately reach a specific well, while other wells may receive water from distant uphill recharge zones via regional flow systems.

This spatial variability can cause a mismatch between the applied recharge and observed groundwater fluctuations, leading to inaccurate response time estimates or a poor spectral fit. Ultimately, ignoring the connectivity of the basin may result in an incorrect estimation of t_c. To mitigate this limitation, we selected shallow wells (less than 100 m deep) for this analysis. However, even in shallow systems, uncertainties remain; for example, if an aquifer is confined, it may receive water from distal sources rather than local recharge. Therefore, this tool may be more effective for unconfined sediments than for fully confined aquifers. Through the goodness of fit evaluation in the manuscript, we are able to identify conditions where the approach works and where higher uncertainties remain. We will emphasize the difference between the categories in the revised work.

6. Improve the description of the groundwater level data set that was used. Currently it is not clear what temporal resolution the data has, or if it is regular.

Corresponding text adaptations at section: Time Series Preparation

The groundwater level time series had different sampling intervals and irregular lengths. Since spectral analyses requires equidistant time steps, the time series were interpolated to a daily time step. The derivation of the semi-analytical solution assumes a stationary process, which is not always the case for the analyzed groundwater level time series.

Roughly 25 % of the time series had severe jumps or inconsistencies in the recorded data due to changing sampling interval or other artefacts, thus these time series were screened and a representative period was selected for further processing. The final selected time series including highlighted and removed outliers and interpolated data are available in the Zenodo upload (Houben, 2026).

7. The outlier handling using the inter-quartile range is generally inappropriate for groundwater level time series due to their seasonality and thus potential for outliers that are not overall extreme values. Likely this is not important when the time series are transformed into the frequency domain. However, please elaborate on this issue.

Corresponding text adaptations at section: Time Series Preparation

We agree that the inter-quartile range is generally not always appropriate for the groundwater levels with seasonality. Though, we found this approach as valuable and practical and evaluated the results, i.e., removed outliers as successful, thus continued with the resulting time series. Only few very strong outliers exceeding the overall seasonality were finally removed. In future studies, we suggest using other techniques such as the Local Outlier Factor where time windows of data points can be considered in order to cope with seasonality in time series. Furthermore, we recommend using tools such as SaQC (Schmidt et al. 2023) for quality control of time series, data flagging and data correction.

8. The entire section LL183-193 is hard to follow and should be rewritten. Specific questions in these sections are:
    - L183: What is meant by tracing "back" the path of the water particle towards the river?
    - The sentence L189-190 is hard to understand. Is this the flow path from water divide to river?
    - L191 Why was the direct distance applied?

Corresponding text adaptations at section: Estimation of Flow Line Length

Thank your suggestions. Indeed, there is a small error in the explanation. We have rewritten the section and clarified the open questions. The new section reads:

Next, we approximate the flow lines (FL) by tracing the path of the water particle along the surface (DEM) starting at the observation well down-slope following a flow direction map towards the river (blue, Fig. 2). For the upper part, the part between the well and the water divide, the flow direction was reversely applied, i.e., the directional vector of each cell was rotated by 180 degrees, then followed from cell to cell and finally stopped when a summit was reached, leading to a unique uphill path (red, Fig. 2).
Real groundwater paths can be very complex and are usually unknown without detailed on-site experiments on the catchment scale. Since our study has a regional focus, we decided to use flow direction maps derived from the DEM, automate the process and obtain estimates of the FL for a few hundred wells without manual inspection of each

hydrogeological scenario. The used flow direction maps, though, are often very noise, in particular in flat regions. In order to remove noisy patterns, avoid unreasonably short flow lines and obtain robust estimates, we smoothed the flow direction map (Appendix A1).
The actual path of the water particle along the hill slope (from water divide, intersecting the well towards the river) represents the longest possible path to be obtained with our method which we called the 'arc length' (Fig. 2c), while the shortest and straight distance is called the 'direct' path. We considered the 'direct' path as more appropriate since the geomorphological surface introduced curves in the estimated flow paths which are presumably not present in the subsurface flow paths (Fig. 2), thus we considered the 'direct' path as better suited.

9. In LL242ff does this analysis only pertain to all time series or only good fits?

   Corresponding text adaptations at section: Intermediate Frequency Regime: Identification of Aquifer Response Times

   This analysis pertained to all analyzed spectra. We will make this clear in the text. We will exclude the irregular and overestimated spectra from the summary and further evaluation and also discuss the implications for the other categories in more detail.

10. In L290 you refer to intermediate frequencies as the "onset of filtering related to the groundwater response time". This description does not help with intuition. Can you clarify what this might mean process-wise?

    Corresponding text adaptations at section: Discussion

    Thank you for pointing this out. We try to clarify this further:

    As the aquifer acts as a low pass filter, low frequency components are mostly unaltered while the recharge signal passes the subsurface. The intermediate-frequency band corresponds to the break-point where the aquifer's low-pass filtering begins to attenuate the input signal, which, in turn, corresponds to the system's characteristic time.

11. Finally, (Ebeling et al., 2025; Kumar et al., 2016) carried out studies in the same regions and also link deeper wells to longer response times, although here responses of multiple years are identified. Reflect on the differences between t_c and accumulation times, and how t_c can be understood and used for groundwater drought analysis compared to current practice.

    Corresponding text adaptations at section: Discussion - Related Approaches. A new chapter was introduced discussing the uncertainties of the applied recharge.

    Thank you for stressing this important connection to studies from Kumar et al. (2016) and Ebeling (2025). Indeed, all studies (Kumar et al., 2016, Ebeling et al., 2025, *Bloomfield & Marchant, 2013, Houben et al. 2025) evaluate the response of groundwater table to external perturbation, though fundamental differences exist:*

1. *Kumar et al. (2016), Ebeling et al. (2025) and Bloomfield & Marchant (2013), examined the cross-correlation between the Standardized Precipitation Index (SPI) and the Standardized Groundwater Index (SGI), using the relationship to estimate how groundwater levels respond to meteorological drivers. Their analysis effectively incorporates the delay introduced by the unsaturated vadose zone, allowing for a response time evaluation that accounts for this lag. By contrast, our method bypasses the vadose zone entirely: we take recharge as a product from mHM (which already accounts for soil moisture and percolation) directly as the input and compute the aquifer's response to this input, thereby isolating the aquifer dynamics from the effects of the unsaturated zone. The resulting response times therefore characterize the aquifer itself excluding any effects from the vadose zone.*
   Comparing these two approaches, it becomes obvious why response times from Kumar et al. (2016) and Ebeling et al. (2025) exhibit longer times than the response times obtained by the spectral analysis approach. Furthermore, the correlation of longer times in relation to the depth of the well is more pronounced since the vadose zone plays a significant role.
2. *Our spectral analysis approach uses a full spectrum wide comparison over multiple frequencies and finds a best fitting spectrum, in contrast to Ebeling et al. (2025) and Kumar et al. (2016) where a best fitting accumulation period was calculated.*
3. *Our approach links the response time directly to physical properties of the aquifer since corresponding (and simplified) equations of groundwater flow are used.*

Current practices in groundwater drought analysis (e.g., Kumar et al., 2016; Ebeling et al., 2025) typically rely on accumulation times derived from the correlation between standardized indices (SPI/SPEI vs. SGI) and these metrics primarily provide insight into the "hydrological lag" which is the time required for a climatic deficit to propagate through the soil and vadose zone to reach the water table. In contrast, the characteristic response time t_c derived from spectral analysis characterizes how the saturated aquifer itself processes and moderates that stress once it has arrived.

Mathematically and physically, these two metrics describe different segments of the hydrological cycle. The accumulation time is a statistical window representing the system's total memory (soil + vadose zone + aquifer), whereas t_c is an intrinsic hydraulic parameter defined by the aquifer's dimensions and its ratio of storage to transmissivity. Consequently, high accumulation times do not necessarily imply long characteristic response times. A system may exhibit a multi-year accumulation lag due to a thick, slow-draining vadose zone, yet possess a short (e.g., < 50 days), indicating that the saturated zone has little capacity to buffer against high-frequency fluctuations or localized abstraction stress.

Implications for Drought Evaluation

In the context of drought evaluation and management, the following holds:

- *By using recharge time series (e.g., from mHM) as the spectral input, our approach cleans the signal. It separates the influence of the vadose zone from the saturated zone, allowing managers to identify whether a well's apparent resilience is due to surface-level delays or the intrinsic buffering capacity of the aquifer.*
- *While accumulation times tell us when a drought begins to manifest in the groundwater, t_c determines the recession rate and the recovery potential. A short t_c suggests a flashy system that may reach critical levels quickly during a recharge failure but can recover rapidly after a single wet season.*
- *For sustainable groundwater management, both values must be analyzed in conjunction. An aquifer with both long accumulation times and a high t_c represents a highly resilient strategic reserve. Conversely, an aquifer with a long accumulation lag but a short is deceptively vulnerable: it may appear stable during the early stages of a drought, but once the deficit reaches the water table, the system lacks the hydraulic capacity to dampen the impact, leading to rapid level declines.*

For the revision of the manuscript, we will integrate the extended elaboration from above in the discussion section and adapt according sections in the introduction.

Technical comments

- In abstract you write "nearly 200 … wells", but in section 2.2.2 it is 224.
  - Will be corrected to 'around 200 wells'
- L46ff: Clarify how variability in recession constants are related to the difficulty in defining drought extent? I challenge that defining drought extent is really that difficult, it is a factor of the systems hydraulic memory (Changnon, 1987).

Corresponding text adaptations at section: Introduction

Changnon et al. (1987) developed a practical framework for detecting the onset, severity, and termination of droughts by monitoring precipitation, soil-moisture levels, shallow groundwater levels, and streamflow. In the groundwater system, a drought is classified as moderate when groundwater levels fall more than 30 % below the long-term mean for at least three consecutive months, and as severe when the decline exceeds 55 % for at least twelve months. Identifying the end of a drought is more complex; it generally requires a rise in soil moisture accompanied by several months of excess precipitation.

Under the linear-reservoir approximation, the time needed for groundwater levels to drop below a drought-threshold is directly related to the recession constant. Hameed et al. (2023) applied this concept by analysing individual base-flow events from 1990–2019, extracting them from streamflow records with complementary precipitation data, and estimating the corresponding recession constants. Although the drought definition is straightforward in theory, inverse estimation of response times from *time-series data introduces uncertainties that can cause variations in the derived recession constants, even for a single hydrograph.*

Lee and Ajami (2023) further emphasized that baseflow drought should be viewed as a process rather than an instantaneous event. Because baseflow responds delayed and dampened to precipitation, they adopted the concept of an intermittent above-zero standardised baselow Index (SBI) during drought periods introduced by Perri et al. (2016).

In summary, while Changnon et al. (1987) provide a clear observational framework for drought definition, practical application, particularly when employing inverse time-series techniques (Lee & Ajami, 2023, Hameed et al., 2023, Perry et al., 2016), yields variability in recession constants. These challenges can be mitigated by adopting our spectral-analysis approach, which obviates the need to delineate start and end points and instead considers the entire frequency spectrum simultaneously.

- Fig1 Please elaborate on b,c in caption
    - b*) Spectrum of groundwater time series, c) corresponding groundwater time series*
- Fig 2 Please elaborate on a,b,c,d in the caption. Also the arc length and direct path are not clearly visible. Please improve. For a and b write descriptive titles for each step.
    - *Thank you for the suggested improvements. We will adapt the figure in review. The following will be added to the figure description:*
    - *a) processing of groundwater level time Series and transfer to frequency domain, b) extraction of recharge time series from spatio-temporal simulations of the mesoscale hydrological model mHM and spatially averaged for shaded region, c) DEM processing and consecutive flow line (FL) estimation. D) measured groundwater level spectrum fitted with semi-analytical solution with t_c and S subject to optimization*
- L187 noise --> noisy
    - *Will be corrected.*

References from Reviewers

Bloomfield, J. P., & Marchant, B. P. (2013). Analysis of groundwater drought building on the standardised precipitation index approach. *Hydrol. Earth Syst. Sci.*, *17*(12), 4769-4787. https://doi.org/10.5194/hess-17-4769-2013

Changnon, S. A. (1987). Detecting drought conditions in Illinois. *Circular no. 169*.

Ebeling, P., Musolff, A., Kumar, R., Hartmann, A., & Fleckenstein, J. H. (2025). Groundwater head responses to droughts across Germany. *Hydrol. Earth Syst. Sci.*, *29*(13), 2925-2950. https://doi.org/10.5194/hess-29-2925-2025

Houben, T., Pujades, E., Kalbacher, T., Dietrich, P., & Attinger, S. (2022). From Dynamic Groundwater Level Measurements to Regional Aquifer Parameters— Assessing the Power of Spectral Analysis. *Water Resources Research*, *58*(5). https://doi.org/10.1029/2021wr031289

Kumar, R., Musuuza, J. L., Van Loon, A. F., Teuling, A. J., Barthel, R., Ten Broek, J., Mai, J., Samaniego, L., & Attinger, S. (2016). Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator. Hydrol. Earth Syst. Sci., 20(3), 1117-1131. https://doi.org/10.5194/hess-20-1117-2016

Liang, X., & Zhang, Y.-K. (2013). Temporal and spatial variation and scaling of groundwater levels in a bounded unconfined aquifer. Journal of Hydrology, 479, 139-145. https://doi.org/10.1016/j.jhydrol.2012.11.044

References from Authors

Boeing et al., "High-resolution drought simulations and comparison to soil moisture observations in Germany," Hydrology and Earth System Sciences, vol. 26, no. 19, pp. 5137–5161, Oct. 2022, doi: 10.5194/hess-26-5137-2022.

M. Hameed, M. A. Nayak, and M. A. Ahanger, "Event-Based Recession Analysis for Estimation of Basin-Wide Characteristic Drainage Timescale and Groundwater Storage Trends," Water Resources Research, vol. 59, no. 12, Dec. 2023, doi: 10.1029/2023wr035829.

T. Houben, "Data accompanying the manuscript: Spectral Analysis of Groundwater Level Time Series for Robust Estimation of Aquifer Response Times." Zenodo, 2025. doi: 10.5281/ZENODO.17610314.

von Jankiewicz, P.; Neumann, J.; Duijnisveld, W. H. M. et al.: Abflusshöhe- Sickerwasserrate - Grundwasserneubildung - Drei Themen im Hydrologischen Atlas von Deutschland. In: Hydrologie und Wasserbewirtschaftung 49 (2005), Nr. 1, S. 2-13.

R. Kumar, "Distributed Hydrologic Model Parameterization: Application in a Mesoscale River Basin," 2010.

A. Marx et al., "Auswirkungen des Klimawandels auf Wasserbedarf und -dargebot," Wasserwirtschaft, 2021.

S. Lee and H. Ajami, "Comprehensive assessment of baseflow responses to long-term meteorological droughts across the United States," Journal of Hydrology, vol. 626, p. 130256, Nov. 2023, doi: 10.1016/j.jhydrol.2023.130256.

Xu Liang , Dennis P. Lettenmaier , Eric F. Wood , Stephen J. Burges, "A simple hydrologically based model of land surface water and energy fluxes for general circulation models," Journal of Geophysical Research: Atmospheres, vol. 99, no. D7, pp. 14415–14428, Jul. 1994, doi: 10.1029/94jd00483.

X. Liang and Y.-K. Zhang, “Temporal and spatial variation and scaling of groundwater levels in a bounded unconfined aquifer,” Journal of Hydrology, vol. 479, pp. 139–145, Feb. 2013, doi: 10.1016/j.jhydrol.2012.11.044.

S. Parry, C. Prudhomme, R. L. Wilby, and P. J. Wood, “Drought termination: Concept and characterisation,” Progress in Physical Geography: Earth and Environment, vol. 40, no. 6, pp. 743–767, Aug. 2016, doi: 10.1177/0309133316652801.

L. Schmidt et al., “System for automated Quality Control (SaQC) to enable traceable and reproducible data streams in environmental science,” Environmental Modelling & Software, vol. 169, p. 105809, Nov. 2023, doi: 10.1016/j.envsoft.2023.105809.

Zhang, Y. K., & Schilling, K. (2004). Temporal scaling of hydraulic head and river base flow and its implication for groundwater recharge. Water Resources Research, 40(3), W035041-W035049.

Author’s response to:

RC2: 'Comment on egusphere-2025-5666', Gunnar Lischeid, 09 Feb 2026 reply

Assessing the vulnerability of groundwater systems is one of the key challenges of groundwater research, not the least due to presumed increasing probability of extreme events. Most of the work done so far is based on modelling studies. Houben et al. followed a different approach based on earlier work by Zhang and Schilling. I agree that spectrum analysis has great and still massively underrated potential in this regard. It is to the credit of this work that it draws the attention of the scientific community to this. However, the devil is in the details.

Thank you very much for recognizing the potential of spectral analysis approaches and the value of the study in this realm. We tried to clarify your concerns as thoughtful as possible and hope that our clarifications resolve any remaining scepsis. Please find our answers in italics.

1. My major criticism is, that this study completely ignores the major source of low-pass filtering. The approach by Liang and Zhang (2013) works under the ideal conditions of tailored numeric experiments. Even then the effects on power spectra are fairly small and are restricted to small distances from the constant head boundary. In real world settings, though, these effects are hardly discernible due to the much stronger impact of vadose zone low-pass filtering. In contrast to the effect described by Liang and Zhang, however, the latter starts to manifest in the high frequency part. This phenomenon is closely related to the second law of thermodynamics (dissipation is the strongest for the high energy part of the spectrum). It applies likewise to soil temperature, electromagnetic waves, etc. Consequently, it doesn’t come as a surprise that the spectra are consistently underestimated for shallow groundwater sites and overestimated for deep groundwater sites (Fig. 7).

   Corresponding text adaptations at section: Discussion - Regional Flow Connectivity and Temporal Recharge Uncertainty. A new chapter was introduced discussing the uncertainties of the applied recharge.

   Thank you very much for pointing this out. We understand that this effect requires more thorough discussion and clarification:

   Since groundwater recharge equivalent product has been used as input, in theory, the filtering effect of the vadose zone should have been accounted for. Apparently, the majority of groundwater observation wells could be well fitted with the semi-analytical solution, indicating appropriate recharge estimation by mHM across all depths (see Figure 7, category ‘good’). Though, roughly one third of groundwater well spectra show incongruent fits, which indicate either a poorer estimation of the recharge or a more complex hydrological subsurface system where lateral inflow, older background water from unconsidered recharge zones far away from the groundwater well or abstraction has caused the analytical assumptions, including the boundary conditions, to fail. For cases where the recharge was under- or overestimated, this has to be attributed to the mHM modelled percolation C (=recharge in our study) and the conceptual *representation of it, being the residual of surface and interflow components (see answers to Reviewer comment 1, question 1 and 2), which might not be applicable to these groundwater time series.*

We would like to emphasize, that we do not consider the mHM model to fail in these cases, but rather the transfer from this concept of mHM to a recharge equivalent.

In the revised manuscript, we will clarify this issue and discuss the implications for the uncertainties of the derived parameters with respect to the different goodness of fit categories.

2. 58-62: What is perceived here as an interference signal is actually the main driver of the low-pass filtering effect: It is the vadose zone rather than the aquifer – have a look at Fig. 4, 5, and 6 in Tsypin et al. (2025, https://doi.org/10.1016/j.jhydrol.2025.133193) or at Liesch and Wunsch (2019, https://doi.org/10.1016/j.jhydrol.2019.02.060). The thicker the vadose zone, the more pronounced is the low-pass filtering effect. The pore space in the aquifer is fully saturated, and water is hardly compressible. Thus, how could the aquifer buffer any input (except by discharging to the constant head boundary)?

   Corresponding text adaptations at section: Introduction, Discussion – Related Approaches. A new chapter was introduced discussing the uncertainties of the applied recharge.

   We appreciate the opportunity to clarify this further. While the vadose zone is undoubtedly a primary and potentially the dominant driver of hydrological signal filtering (as demonstrated by Tsypin et al., 2025 and Liesch and Wunsch, 2019), the saturated aquifer independently functions as a low-pass filter governed by its intrinsic hydraulic properties: transmissivity T and storativity S. Although confined aquifers possess limited pore-space elasticity, resulting in low storativity, a concurrently low T can still yield significant characteristic response times ($t_c = L^2 S / T$). Physically, this buffering occurs through the elastic expansion of the aquifer skeleton and the vertical oscillation of the water table. Whereas the unsaturated zone primarily introduces a transport delay, the saturated zone attenuates the signal through hydraulic diffusion. Our methodology isolates this effect by utilizing mHM-derived recharge as a direct input, effectively bypassing the vadose zone's influence. It should be noted, however, that not all observation wells yielded a high-quality fit, suggesting that localized recharge estimations may be subject to greater uncertainty at specific sites.

3. 98-100, 220-223: Cut-off frequencies, that is, the change of slope is a nice index to compare different time series. But it comes with two major problems: Firstly, deep groundwater head dynamics usually exhibits pronounced persistency with Hurst coefficients exceeding 0.5. Thus, legacy effects will last for very long (Koutsoyiannis 2013, http://dx.doi.org/10.1080/02626667.2013.804626). Actually, there is no upper limit of response time at all for very fundamental reasons. Secondly, such clear breaks of slope are limited to time series with rather low degrees of low-pass filtering like stream discharge. In contrast, power spectra of deep groundwater head time series often do not

show any clear cut-off frequencies. See, e.g., observed groundwater head spectra (Fig. 4) in Zhang and Schilling (2004, https://doi.org/10.1029/2003WR002094) or Fig. 4 in Tsypin et al. (2025, https://doi.org/10.1016/j.jhydrol.2025.133193). There is no physical reason why low-pass filtering should stop at a certain frequency. It is just a matter of the intensity of the low-pass filtering process. Please comment.

Corresponding text adaptations at section: Results – Groundwater Level Spectra, Discussion - Regional Flow Connectivity and Temporal Recharge Uncertainty – Related Approaches. New subchapter were introduced discussing the uncertainties of the applied recharge.

We agree to your point in regarding the persistent nature of deep groundwater dynamics and the influence of the Hurst phenomenon. While it is theoretically true that hydrological legacy effects can exhibit long-range dependence without a strict upper limit, our application of the semi-analytical solution aims to identify an effective characteristic timescale relevant to resource management. We recognize that in highly damped systems (particularly deep aquifers) the power spectrum may not exhibit a visually distinct break or cut-off frequency within the available observational window as shown by Zhang and Schilling, 2004 and Tsypin et al., 2025. In these instances, the spectrum often follows a power-law decay. However, by fitting the full semi-analytical model rather than relying on a break-point analysis, we are able to constrain the parameter based on the overall spectral distribution, assuming a good fit over a wide frequency range. While this may not represent a complete cessation of low-pass filtering, it provides a physically grounded index for comparing the relative buffering capacities across the analyzed groundwater data.

4. 152-156: Removing linear trends distorts the power spectrum. Trends do not develop independent from the spectral signatures but are an inevitable consequence of the low-pass filtering process which results in increasingly more smooth time series. Consequently, the length of periods of monotonic rise or fall increases and thus the probability to identify significant trends (Koutsoyiannis 2006, https://doi.org/10.1016/j.jhydrol.2005.02.030; Lischeid et al. 2021, https://doi.org/10.1016/j.jhydrol.2021.126096). This holds even for random white noise time series.

Corresponding text adaptations at section: Methods – Time Series Preparations

We acknowledge your concerns regarding the removal of linear trends. As noted by Koutsoyiannis (2006) and Lischeid et al. (2021), apparent trends in groundwater time series are often an intrinsic consequence of the low-pass filtering process itself, representing the long-term persistence of the system rather than independent external forcing. Theoretically, removing these trends can suppress the power at the lowest frequencies and potentially lead to an underestimation of the characteristic response time. However, from a signal processing perspective, detrending is a standard pre-requisite for spectral analysis using the Fast Fourier Transform (FFT) to prevent spectral leakage and ensure the stationarity required for Welch's method and is typically *done in literature (e.g. Jiménez-Martínez et al. 2013). To balance these concerns, we focused our semi-analytical fitting on the frequency range between low- to mid-range frequencies leading to response times within a reasonable value range, thereby ignoring overarching trends in the time series. Our focus is on determining the prominent and dominant response time of the aquifer suitable for water management applications and not to uncover multi-decadal trends. A full estimation of the effect of trend removal, though, can only be experimentally evaluated.*

Besides that, we are interested in the aquifer's response time rather than the overarching response of groundwater levels to surface signals. Therefore, our goal is to isolate the response of the aquifer due to the recharge signal as best as possible, thus removing the very low frequency noise from the signal revealing the time intervals of interest and valuable for water management purposes.

5. 273-276: It is indicative that the paper does neither report on transmissivity nor on storage values. Other than the authors I would ascribe these uncertainties to disregarding the vadose zone effect on low-pass filtering (see above) rather than to uncertainties of flow line length determination. Please comment.

   Corresponding text adaptations at section: Results – High Frequency Regime: Filtering Behavior

   We acknowledge that the absence of site-specific transmissivity T and storativity S values introduces challenges in verifying the individual components of the characteristic response time. However, we argue that the primary source of uncertainty in decomposing t_c into its hydraulic constituents stems from the difficulty in precisely defining the drainage process and flow line lengths L in complex, heterogeneous terrain. As scales with the square of the aquifer length L^2, even minor variations in the assumed distance to the boundary significantly impact the calculation of T and S. While we recognize the reviewer's point that vadose zone filtering contributes to the overall signal attenuation, our use of model-derived recharge specifically targets the saturated zone's response. In this context t_c, serves as a robust, integrated index of aquifer resilience. Even if the individual values of T, S, L or are not independently resolved, their combined ratio (represented by t_c) accurately reflects the aquifer's physical capacity to buffer fluctuations in recharge.

6. 12-15, 300-306: The statement "Groundwater systems exhibiting longer response times are interpreted as more resilient to drought conditions and therefore potentially better suited for groundwater abstraction than aquifers with shorter response times" holds only in the short-term when the response time is long compared to the timescale of dry spells. But long-term buffering comes with the price of very slow and delayed recovery. Increasingly longer dry spells and extended periods of a monotonic decrease of groundwater head and total water storage which are now abundant in Central Europe (Xanke and Liesch 2022, https://doi.org/10.1007/s10040-021-02448-3; Wunsch et al. 2022, https://doi.org/10.1038/s41467-022-28770-2) question such recommendations. Please comment.

Corresponding text adaptations at section: Discussion – Implications for drought evaluation

Thank you very much for highlighting the importance of regarding the dual nature of groundwater response times. While aquifers with longer characteristic times t_c exhibit greater resistance to short-term meteorological droughts (effectively filtering out short to mid-term dry spells) this buffering capacity comes at the cost of delayed recovery. In the context of the increasingly frequent and prolonged drought periods observed in Central Europe (Wunsch et al., 2022; Xanke and Liesch, 2022), a long t_c may lead to a 'long-memory' depletion that persists well beyond the return of normal precipitation. Consequently, our interpretation of resilience must be adapted: systems with high t_c are better suited for sustaining abstraction during isolated dry years, but they require more strict long-term management to prevent entire depletion. Conversely, 'flashy' systems with short are highly vulnerable to immediate rainfall deficits but possess the advantage of rapid replenishment.

We will revise our discussion to emphasize that should be used to determine the temporal scale of management (seasonal vs. multi-annual) rather than a universal recommendation for abstraction suitability.

7. Figure 1: Please explain the b) and c) panels. It is a little bit unusual and confusing to have two panels b) and c) each.

   We will adapt the labels accordingly to match the standards of the Journal.

8. Figure 2: Correct "lenth" in the figure caption.

   Will be corrected in the revised version.

9. A3: I guess the term "spectral analysis" refers to the whole procedure rather than to the power spectra of groundwater head time series in particular, is that right? Otherwise, frequencies should always be < 1. Please clarify.

   Text adaptations: See figure caption A3.

   Thank you for pointing this out. We refer to the term 'spectral analysis' as the combination of estimating the power spectrum as well as the fitting with the analytical solution. Results of the 'spectral analysis' relates to the optimized parameters from the curve fitting process, i.e., t_c, T and S.

   Indeed, frequencies from the spectrum are always below 1. The term frequency in the histograms of figure A3 refers to the count of values for each bin. We will adapt the figure caption accordingly in order to avoid confusion.

10. A8: This spectrum seems to be indicative for a confined aquifer where short-term fluctuations are induced by atmospheric pressure variation, resulting in a remarkable atypical large share of variance in the high-frequency range.

Text adaptations: See figure caption A8.

Indeed, this is most likely a correct observation. The well (1937) is located in tertiary sediments from the Molasse basin (Landshut) and around 86 m deep (Table 1). The groundwater table depth varies around 10 m below subsurface. Multiple dominant frequencies can be observed: Long-term and multi-annual fluctuations with periods of decades (Figure 1), seasonal fluctuations with a period of a year and short-term fluctuations with periods of weeks to even days (Figure 2 and 3). The latter can be most likely attributed to barometric pressure changes in the atmosphere, as variations in summer are less than in winter times.
As shown in Figure A8 in the manuscript, the diverse set of present frequencies in the signal result in a complex shape of the spectrum with several breakpoints. Our simplified semi-analytical approach with a recharge product taken from a land surface model fails to capture the complexity of the processes of this groundwater level fluctuation and was consequently categorized as irregular and not subject to further analysis.

References from Authors

J. Jiménez-Martínez, L. Longuevergne, T. Le Borgne, P. Davy, A. Russian, and O. Bour, 2013, "Temporal and spatial scaling of hydraulic response to recharge in fractured aquifers: Insights from a frequency domain analysis," Water Resources Research, vol. 49, no. 5, pp. 3007–3023, May 2013, doi: 10.1002/wrcr.20260.

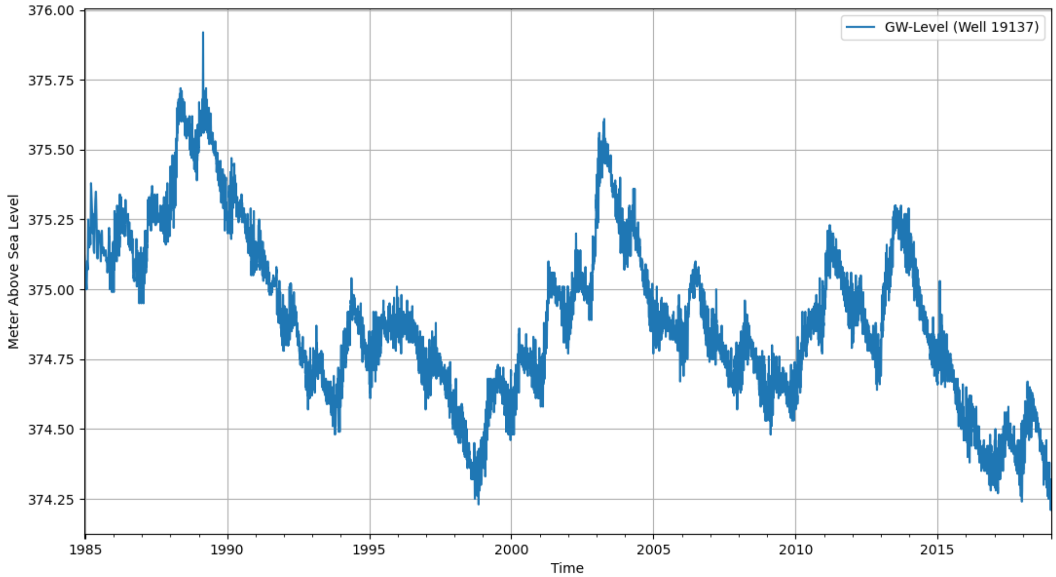

Figure 1 Complete Groundwater level time series from well 19137.

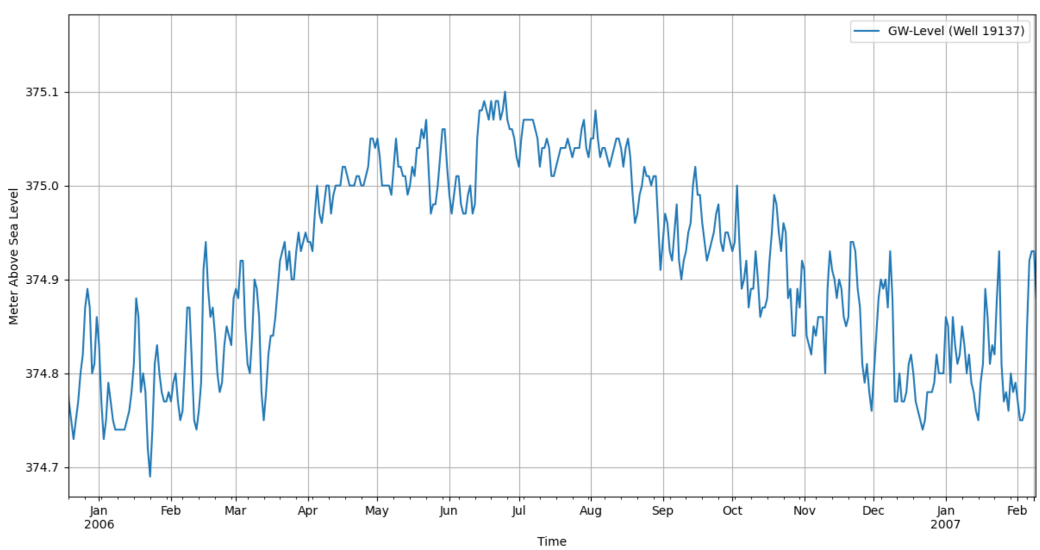

Figure 2 Groundwater level time series from well 19137 zoomed to the year 2006.

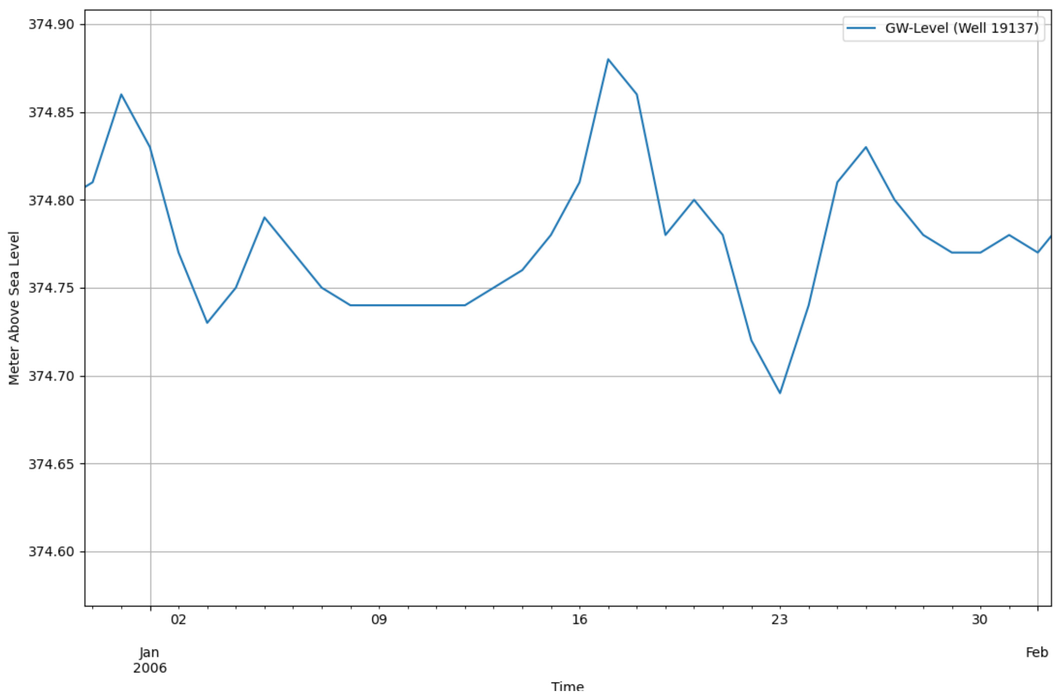

Figure 3 Groundwater level time series for well 19137 zoomed to January 2006.

Table 1: Metadata well information of well 19137

ID	Name short	Name	Date	east	north	county	Municipal	aquifer	depth	Surface elevation	Max GW below Surface	Low GW below surface	Mean GW below surface
19137	Landshut	DENGKOFEN964	1984	4533280.0	5400700.0	Dingolfing_Landau	Mengkofen	Tertiär	86.0	385.45	9.51	11.23	10.53