Interactive comment on 'Observed groundwater temperature response to recent climate change' by K. Menberg et al.

Review by H. Kooi

Reply to comments:

<u>**General comments**</u>

Comment #1: The work presented in this manuscript involves a case(s) study of the impact of climate warming (atmospheric temperature rise) on groundwater temperatures at depths of about 10-30 m. This topic is relevant to the hydrological community. The authors draw particular attention to the potential relevancy for stream/river temperatures and (associated) groundwater dependent ecosystems.

> **Reply #1:** This statement accurately captures the scope of this paper.

Comment #2: Novel and/or particularly interesting aspects of the work are the rather long time series of observational data on (pumped) groundwater temperature and the 'regime shift analysis' approach. The advantage of the latter is that it allows to establish in an elegant way a relation between local (diffused) 'shifts' in groundwater temperature and climatic regime shifts that can be recognized over very large spatial domains, even up to the global scale. It is interesting to see that it apparently works. However, applicability of the method to the 'diffused signals', which inherently do not include 'abrupt shifts' needs to be justified and findings need to be interpreted more carefully to avoid misinterpretations of air-ground temperature coupling (SAT-GST).

> **Reply #2:** We agree that the applicability of the analysis of 'abrupt shifts' in groundwater temperatures is certainly limited due to the diffusive heat transport in the subsurface. Abrupt shifts in the thermal signal at the surface will diminish within the subsurface and actually disappear in a certain depth, appearing rather as diffused signals. However, in the quite shallow GWT time series evaluated here, the breakpoints observed in the long term mean GWT are statistically significant, though exhibiting lower p-values than in the atmospheric time series. Additional information on this issue and the statistical requirements of the regime shift analysis are addressed in detail according to the specific comment #22. The interpretation of the observed regime shift in GWT regarding the damping of the thermal signals and the adjacent

discussion of the air-ground temperature coupling (SAT-GST) will be thoroughly revised according to these concerns and the comments #17, #20 and #22.

Comment #3: Where virtually all geothermal climate studies use borehole temperature logging data, here temperatures obtained by well pumping are used. This aspect (value of this type of data) deserve to be elaborated more comprehensively and better, because results basically confirm what is already known about propagation of surface temperature signals into the subsurface or air-ground temperature coupling.

> **Reply #3:** As also stated correctly in comment #12, the novelty of our study is the type of data used for the analysis of air-ground temperature coupling. The measured time series of groundwater temperature and the applied methods enable an evaluation of the temperature coupling over the last decades with a temporal resolution of a few years. Palaeoclimatic studies using deep borehole temperature profiles can track ground temperatures over hundreds of years before present, but with a much lower temporal resolution in the order of decades. A respective statement will be included in the manuscript to highlight these distinctions.
>
> Due to general knowledge about the propagation of thermal signals in the subsurface, the observed influence of atmospheric temperature signals on groundwater temperature might indeed be anticipated. However, according to our knowledge this is the first study that examines this short term coupling by statistical analysis of actual measured data and analytical modeling. Details about the newly derived implications of this short term coupling for groundwater dependent ecosystems are addressed with the associated comments #12 and #25.

Comment #4: The conclusions provided in the final conclusions section are generally sound. However, the forward modelling with the analytical solution, its results, and the discussion thereof is inadequate in several respects and causes unnecessary confusion as will be explained below. These aspects can be remedied fairly easily.

> **Reply #4:** Thank you for raising these concerns. They are addressed on a point by point basis below according to the comments #9, #10, #17 and #21-25.

Specific comments

Comment #5: GWT data. Compared with temperature logging in standing water in a well bore, interpretation of temperatures obtained via well pumping is subject to a large number of unknown influences. Firstly, the water represents a mix of water entering the well bore along

the whole! vertical of the well screen and hence different depths. The inflow can be fairly even, but can also have a dominant inflow near the base or the top of the screen depending on aquifer heterogeneity, backfill and screen clogging processes. This distribution would even depend on the pumping rate/ induced water level drop in the well bore during pumping. This is of fundamental importance for the Hardtwald wells which have very long screens.

> **Reply #5:** Thank you for noting the influence of the well screen depth and how heterogeneities can influence the vertical well capture zone. These comments are addressed in more detail below in the replies to comments #8-10 and #21. Based on your concerns, we will include a statement in the methods that describes why we now consider a range of depths in the analytical model.

Comment #6: Secondly, depending on the heterogeneity structure of the aquifer around the well bore, there could in principle even be a relatively large groundwater contribution from below the bottom of the well screen or above its top.

> **Reply #6**: We will employ this information to present a plausible reason for the measured Sinthern GWT not adhering to the predicted GWT trends.

Comment #7: Thirdly, the water flows upward from the pump through the pumping hose/tube which exchanges heat with the air in the well bore, air outside, and, a notorious factor is the impact of direct solar insolation of the hose which can heat up even fairly fast flowing water very quickly. The heat exchange very much depends on factors such as pumping rate and tube size, length and material, and ambient air temperature during sampling (season). A constant outflow temperature for a constant pumping rate does not guarantee the temperature is representative of the groundwater temperature. A simple check for seasonal trends in the data would be a minimum (for the mean parameters used for Dansweiler, for instance, at the depth of the screen (20 m), numerical modelling shows a seasonal GST change between 0 and 20°C corresponds to a groundwater temperature fluctuation of about 0.02°C). Much larger fluctuations point at 'contamination effects 'such as those mentioned above. The observed inter-annual fluctuations of several tenths of a degree for the Dansweiler well already indicate that such influences are significant. Furthermore, has the whole procedure (protocol and instrumentation) used for sampling been exactly the same over the 40 years? This is an additional potential cause of fluctuation or systematic changes. Such uncertainties should be acknowledged/considered when using the data.

> **Reply #7:** We agree that there are several potential reasons for the rather large fluctuations in the inter-annual GWT time series, which are not properly addressed in

the manuscript so far. We will make the according changes to reflect the uncertainties of this type of data and the subsequent implications for the interpretation. The major concerns are briefly addressed here.

Due to the length of the time series of up to 30 years, detailed information about the pumping rate during sampling, the used equipment or ambient air temperatures are not available for the whole time period. The GWT time series from all wells exhibit certain seasonality with temperature measurements in summer and autumn being on average 0.4°C higher than in winter in Dansweiler, and 0.5°C in the Hardtwald wells. These values are considerably higher than the predicted fluctuations of 0.02°C from the seasonal SAT changes, which means that there is most likely an impact of ambient air temperature on the groundwater in the hose during sampling. Yet, it should be noted that the accuracy of the measurements is ±0.1°C, which is close to the observed variations. Consequently, the sentence stating that the measured temperature is representative for the upper aquifer will be deleted. However, years with only one measurement are scarce, so that annual mean GWT is rarely influenced by these seasonal fluctuations. Individual outliers in the time series caused by such variations or due to other reasons are accounted for in the statistical analysis and do not influence the results of the regime shift analysis due to the sequential approach (see comment #22 on statistical requirements below).

The measurements were performed by the water authorities within the groundwater quality observation program. The measurement protocol, which is standardized by the environmental state agencies to assure good quality of the obtained data, has undergone no significant changes in the last decades. The instrumentation will have certainly changed within this rather long period, which will be addressed as a possible source for GWT fluctuations in the manuscript.

Comment #8: Modelling and its interpretation. Given the general groundwater flow behaviors for pumped wells mentioned above, it is conceptually inappropriate to compare the observed temperature time series with a model-generated time series for the depth corresponding to the water table. Most logical would be to generate a time series for the mean temperature (integral divided by length) along the depth of the screen. Even depth-weighted integrals could be considered in a sensitivity analysis for uneven inflow into the well. The 'cone of depression' (p. 3654, line 8) is not a concept which would justify the adopted water table depth approach.

Reply #8: We agree that our original approach was not ideal, and we thank you for raising this concern. We solved out the integral of the analytical solution, and the resultant equation was several lines long. Given the approximate nature of this study, we think it is best to rather simply use the midpoint of the screen depths for our 'best guess' results. A quick glance at Figure 4 indicates that the temperature results corresponding to halfway down the well screen are approximately halfway between the upper and lower limits of the temperature envelope. These upper and lower limits represent results for the top and bottom of the well screen, respectively.

Comment #9: For Dansweiler and Sinthern a single depth of about 20 or 21 m may be appropriate because of the short well screen. For the Hardtwald wells an integral approach is crucial; the water table depth (6 m) definitely is way too shallow to generate a meaningful time series. This most likely accounts for the inferred offset between ΔSAT and ΔGST for these wells, which therefore seems an artifact. The text of sections 2.3 and 3.2 should be modified accordingly.

Reply #9: We now consider depths that range from the top of the well screen to the bottom. As you suggest, this change to the manuscript eliminates any suggestion of an offset between ΔSAT and ΔGST. We have removed such statements and are pleased with how the manuscript has been improved based on your suggestion.

Comment #10: Presently, the predictive uncertainty of the model is captured in Figure 4 in the 'predicted GWT range'. However, this is due to uncertainty in thermal parameters only. The uncertainty caused by the screen length in combination with unknown inflow distribution is way larger. Point depths ranging between 15 and 25 m may be reasonable estimates for this uncertainty (or specified uneven inflow distributions with the integral approach).

Reply #10: Good suggestion. We now include uncertainties in the predicted envelope due to thermal properties AND depths. The maximum limit of the temperature envelope was obtained using the depth to the top of the well screen and the highest diffusivity and lowest heat capacity (heat capacity is in the U term in the solution), whereas the minimum limit of the temperature envelope was obtained using the depth to the bottom of the well screen, the lowest diffusivity, and the highest heat capacity.

Comment #11: Table 3 lists ranges of thermal parameter values. However, the combination of heat capacity and thermal diffusivity values in not clear from the way they are presented. Probably a small bulk heat capacity would correspond to a large diffusivity, otherwise thermal conductivities seem unrealistic. This should be clarified.

Reply #11: In order to clarify this issue, we will include the assumed literature values of the thermal conductivity of the saturated and unsaturated zone in Table 3. The given ranges in thermal conductivity and heat capacity account for varying water saturation of the porous media and for variation in the composition of the sedimentary material, i.e. different contents of gravel, sand, etc. Due to the interaction of these variations the small bulk heat capacities in Table 3 do not necessarily correspond to the larger diffusivities. The corresponding paragraph in the manuscript will be changed accordingly.

Other comments/corrections

Comment #12: p. 3638: line 35: Rather vague and in my opinion incorrect statement. In what sense are the implications of climate change for groundwater temperatures not comprehensively understood? The present study certainly does not add to or require changes in present understanding. What is shown (with corrections suggested) was predicable on forehand. What is new here is that it is shown that long temperature time series obtained from pumping wells can also be valuable to document and study climate impacts, in spite of its more 'contaminated' and vertically integrated signature.

Reply #12: We do not necessarily agree. There is, in fact, a poor understanding of how shallow groundwater temperature rise may impact ecologically important aquifers and rivers. This is manifested by the plethora of surface water temperature papers that consider stream warming due to climate change but do not consider the potential of groundwater temperature rise to influence stream warming. Surface water hydrologists are one of our target audiences for this paper, and this is partly why we have included so much ecohydrological content. In particular, our study shows that groundwater may warm rapidly and drastically in response to climate change, which are surprising results to some (although not necessarily to anyone who understands subsurface heat transport). The matter of predictability of our results and the novelty of the used data and methods is already described in the reply to comment #3.

Comment #13: line 10: Abrupt changes in groundwater temperature? Violates heat transport behavior.

Reply #13: We agree that this was not a good descriptor of the diffuse signal. We meant abrupt increases in terms of breakpoints in the long term mean. This was confusing; hence 'abrupt' has been removed from this sentence in the abstract as well as a related sentence in the conclusions.

Comment #14: p. 3642: lines 16/17: Variations of water table of 6m (and mean water table 6m below land surface beg for some explanation. Relevancy for the present study, and the magnitude in relation with the recharge of about 220 mm/yr. Is there a pumping station nearby? Irrigation extraction by farmers? How can this be consistent with a steady vertical advective heat flow (U) in the model?

> **Reply #14:** The variation in the depth of the water table stated here is the total variation, i.e. the mean depth is 7 m, with a maximum depth of 10 m and a minimum depth of 4 m, which occurred in individual years during observation time. We will rephrase the variation to 'maximum variations of ±3m' to make this clear.
>
> Variations in the depth of water table are relevant for the interpretation of the time lags of the shifts between atmospheric and groundwater temperature. In particular, a trend in the depth of the water table would impair the comparison of time lags of different shifts in one well. However, there is no obvious long term trend in the water level of the four observation wells over the last decades. The wells in the Hardtwald are located near a pumping station, which is likely to influence the water level. According statements will be included in the manuscript.
>
> Time-series for the annual groundwater recharge were unfortunately not available for the whole observation time. To account for the long-term variability an uncertainty of ±20% was assigned to the annual recharge values, which were also applied to the vertical advective heat flow U in the model (p. 3648, lines: 13-17).

Comment #15: p. 3646: section 2.3: Would be good to also explicitly state the model assumes (a) uniform and steady vertical groundwater flow over a depth range deeper than the well depths and (b) recharge temperature equals the average annual surface temperature. These assumptions, together with assumed heterogeneity of thermal properties for a variably saturated system, merit discussion in later sections in relation to conclusions drawn from the modelling.

> **Reply #15:** These are certainly assumptions associated with the governing conduction advection equation. We will note these, along with others, in the methods section. We will also include a new paragraph in the discussion that explains the shortcomings of this approach.

Comment #16: p. 3648: Equation (6): For sake of completeness mention that the contribution of each summation term only applies for $t >= t\_i$. Otherwise unwarranted cooling is

calculated before the relevant step change in surface temperature. U_z is not defined and appears to equal U.

> **Reply #16:** This is precisely what the Heaviside function indicates (i.e., the Heaviside function turns on and stays on when the value inside the Heaviside function is positive). Nonetheless, such a statement will be added. U_z should have been Uz as can be shown from an analysis of the units within the exponential term.

Comment #17: p. 3647, line 17-25: This is inappropriate reasoning. In the model initial GWT and hence GST are set equal to observed GWT. Potential offsets in SAT GST due to surface conditions in the real world system are subsequently of no consequence for the imposed step changes in annual GST (unless a step change in surface conditions (e.g. vegetation or snow regime) occurred at the same time, which is not the case).

> **Reply #17:** As other studies have shown (Kurylyk et al., 2013, HESS; Mellander et al, 2007, Clim. Change), a shift in SAT can produce a shift in snowpack conditions and/or deciduous vegetation, which produces decadal GST changes that do not necessarily follow SAT changes. However, this text has been removed from the manuscript based on the modifications we have made to the depths utilized in the analytical solution. In the approach of the modified manuscript we have set ΔGST = ΔSAT for all wells.

Comment #18: p. 3650, lines 3-5: This is a vague statement, in particular the 'up to 30 m' and 'significant'. It can be readily shown that for the well sites studied here variations due to inter-annual fluctuations of GST (or SAT) are much smaller than those observed in the data. Analytical solutions to quantify the damping of periodic GST fluctuations with depth (also with advection influences) can be used to show this. Or numerical solutions can be used.

> **Reply #18:** We agree that this was poorly worded. Stallman's (1965) equation could be used to demonstrate that intra-annual fluctuations should be completely damped at this depth even under high recharge rates. The sentence will be removed.

Comment #19: p. 3650: If the accuracy of each shift is +/- 1 year, then the accuracy of the difference between two shifts (lag) is less accurate than that.

> **Reply #19:** We agree. As the accuracy of ±1 year applies to the shift in SAT and to the succeeding shift in GWT, the overall accuracy of the difference is ±2 years. The values in Table 4 will be changed accordingly.

Comment #20: lines 18-20: What is the relevancy of this statement?

Reply #20: The point is that local SAT will not necessarily follow global SAT or even regional SAT changes. We will include the adjective 'local' and change 'yet' to 'furthermore' to clarify the purpose of this statement. This helps explain why the timing of regime shifts at different spatial scales may not completely overlap.

Comment #21: lines 23-25: This is not substantiated and not evident. The depth of the well screens may be more important (can be evaluated via sensitivity analysis).

Reply #21: This sentence will be deleted. The sensitivity of the GWT response to the well screen depth will be considered in the results 'envelope' in the analytical solution figure (Fig. 4).

Comment #22: p. 3651, lines 8-12: Indeed. Due to this slow and 'smoothed' response in the subsurface I would expect the regime shift method is NOT suited to determine the proper amplitude of the GWT and hence the GST shift, and overestimate its timing. The inferred amplitude step change of the diffused signal would depend on the length of the stable regime. The inferred amplitude can therefore NOT be used to draw conclusions regarding damping of GST change relative to SAT change. Aren't there statistical requirements of time series for regime shift analysis? And do diffused signals meet these requirements? The discussion of lines 25-28 seems inappropriate in this light.

Reply #22: We agree that the regime shift method can underestimate the groundwater rise in response to climate change (unless equilibrium has been met due to long stable regime). That is why, unlike Figura et al. (2011, GRL), we include the process-based modeling which demonstrates that GWT will eventually rise (with the same magnitude) in response to the regime shift in SAT or GST if the new climate regime shift lasted indefinitely. We agree that no conclusion regarding ΔGST damping in comparison to ΔSAT can be made. Such statements will be removed. We will explicitly state that in the absence of snowpack evolution or land cover changes, ΔGST should follow ΔSAT, and ΔGWT should in turn follow ΔGST if given enough time (see analytical solution discussion).

There are requirements regarding the length of a time series in order to detect the regime shifts within a certain level of confidence depending on the assigned length of a stable regime. With an assigned regime length of 10 years, as in our study, up to 3 statistical significant regime shifts could theoretically be identified in a time series of 40 years. In principle the regime shift analysis resembles the fitting of a step function to a time series, though with a limited number in the degrees of freedom. Not only is

the length of the identified regimes restricted by a minimum value (cut-off length), also the difference in the long term means between two subsequent has to be statistically significant according to a student's t-test. In this test also the variance of the input data is considered, which means that the discussed large fluctuations in GWT are accounted for. Regarding time series of completely diffuse signals with some kind of trend the test would certainly fail to find a significant regime shift. However, in the rather shallow GWT time series evaluated here the breakpoints observed in the long term mean GWT are still statistically significant (p-values < 0.01), despite their, to a certain degree, diffusive nature. Thus, we are certain that our data fulfil the mentioned requirements and that the results are statistically sound.

Comment #23: p. 3653 and 3654: See specific comments on modelling and its interpretation.

Reply #23: This section will be extensively modified. Please see our replies to the specific comments #8-10.

Comment #24: p. 3655, lines 17-20: This statement should be removed/modified. Results of the present study do not support this.

Reply #24: This sentence will be modified to remove any reference to GWT damping in comparison to SAT warming. This has been shown to be the case in other studies, but admittedly was not demonstrated in the present one.

Comment #25: Conclusions should be more geared to specific findings/result of the study and not repeat discussion items that are not true results / have not been explicitly demonstrated.

Reply #25: We agree. We will remove the sentences at the end of the conclusions that discussed GDE (this information will be retained in the discussion) and include a sentence that states that groundwater temperature evolution in response to climate change can be tracked by analyzing long term records of the temperature of pumped groundwater.

Comment #26: Referencing is generally fairly complete. A modelling study dealing with the same time period and various factors influencing subsurface signals, including groundwater flow, heterogeneity and surface influences that may be of some use: Earth and Planetary Science Letters, 270, 86-94**.**

Reply #26: Kooi (2008) will be added as a citation in the introduction and in the reference list.