General comments

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.

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 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).
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) deserves 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.

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.

Specific comments

<u>GWT data</u>. 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.
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.
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 interannual 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.

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.
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.

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).

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.

Other comments/corrections

p. 3638:
line 3-5: 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.

line 10: Abrupt changes in groundwater temperature? Violates heat transport behaviours.

3642:

lines 16/17: Variations of water table of 6 m (and mean water table 6 m 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?

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.

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.

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 of 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 (eg vegetation or snow regime) occurred at the same time, which is not the case).

3649:
line 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 interannual 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.

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

line 18-20: What is the relevancy of this statement?

line 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).

3651:
line 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.

3653 and 3654
See specific comments on modelling and its interpretation.

3655
line 17-20: This statement should be removed/modified. Results of the present study do not support this.

3656/57
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.

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.

Henk Kooi
Critical Zone Hydrology Group
Department of Earth Sciences
VU University Amsterdam