We thank the anonymous reviewer for his time spent in reviewing our manuscript. In the following we will give detailed response to his comments.

General comments

This paper describes a study within which long term datasets of groundwater level and spring discharge and potential recharge modeling results are interpreted to infer the sources of variability within the groundwater datasets. The source of the variability is assigned depending on whether patterns are observed in all datasets (ascribed to a climatic source), in just groundwater levels and spring data (ascribed to a landuse change source) or solely in groundwater levels (ascribed to an abstraction source). Whilst the basis of the interpretation is superficially attractive, it is underpinned by many questionable (largely implicit) assumptions which undermine the value the study to an unknown extent.

As far as we can interpret the specific comments of the reviewer, the "many questionable (largely implicit) assumptions" are all referring to the modeling part of our study, especially to our assumptions about soils, land use and runoff. We acknowledge the doubts of the reviewer and in the following we will try to explain our assumptions and make suggestions how to resolve the issues.

Specific comments

1) There appears to be a surprising presupposition in the Introduction that hydrogeologists doing climate impact studies do not first try and understand system behavior under current conditions (e.g. P7624, L1-3) before starting to assess future conditions. The development of a robust conceptual model of a system is a fundamental prerequisite for any impacts modeling. That the authors final conclusion (P7638, L26) that the focus should be "on extreme groundwater drought periods rather than analyzing annual averages" is disappointing in its unintentionally-implied criticism of hydrogeologists working in this subject area.

We are sorry if the reviewer got this impression, we did not intend this. We agree with the reviewer, that hydrologists and hydrogeologists do a great job in conceptualizing their models. However, we do think that it is legitimate do deliver additional information (as we intended with this study) which can be of benefit for future impact studies. Additionally, we think that it is also legitimate to point out some deficiencies made in past impact studies. The reviewer may disagree, but there are certainly some cases where modelers did not pay much attention why certain results came out of their models (for example reported in Blöschl and Montanari 2010, p.377).

2) Potential recharge modeling using a 1D MIKE SHE model is use to determine the influence of landuse change, since the recharge outputs from the model do not incorporate landuse change whereas the impacts of any landuse change on recharge are assumed to be present within the time series groundwater data. However, there are a number of assumptions within the modelling that weaken the presupposed relationship – in essence it is assumed that the model is a perfect representation of recharge

assuming no landuse change, so that any deviations between model and observations can be related to land use.

We do not see how this assumption weakens our "presupposed relationship". By neglecting the unknown land use changes in the modeling we try to distinguish between land use and climate introduced features. If it is not possible to relate the observed features with the climate signal (as produced by the recharge modeling) other processes (LU changes, pumping) have to be responsible.
We do not assume our model to be perfect. We discuss the model deficits in detail in the first part of section 4.2 (P7633; L1-19). It is clear that our assumptions have an influence on the absolute values of recharge. However, we assume that the model deficits (parameterization…) do not largely affect the changes and especially the long-term trends of recharge. For example we do not expect significant differences in the recharge trend between a forest on a sandy loam and a meadow on a loamy sand. The governing processes affecting the trends of the groundwater data are assumed to be climate, land use changes and pumping.

In particular, the model is parameterized with the landuse around the monitoring site (rather than over the source's catchment area); runoff is neglected, so any changes in the rainfall-runoff relationship associated with changing soil conditions (e.g. compaction) or rainfall intensity are ignored; uncertainty in the estimated reference ET is ignored; all sites are assumed to have identical soils (a Cambisol, which is not even a soil type but a broad soil grouping); and there is no calibration nor assessment of the plausibility of the model outputs.

The intention of the simplified modeling is not to produce a "solve-all-the-problems-at-once-and-forever-model" but to use it as a tool to distinguish between processes dominating the groundwater dynamics. The most interesting cases are those were model results and observations disagree. Those are the cases where secondary processes (land use changes, pumping) prevail. It would only be problematic (indicating highly uncertain model results) if for none of the observation sites good agreement would be reached. However, we found, that we got the best agreements for those observation sites (springs) where we also assumed the secondary processes to be least important.

The reviewer is further referring to many different points which we will discuss in detail:
1. Land use
As mentioned before we assume that the deficits in the land use parameterization are mainly affecting the absolute values and not the changes or trends (see also answer before). However, to acknowledge the doubts of the reviewer we suggest incorporating the outcomes of a sensitivity study analyzing how different land use types are affecting the model results. In this sensitivity study also different soil types and the effect of the runoff will be evaluated. We firmly believe that the results from this sensitivity study will confirm the conclusions from this paper: although absolute values are affected by soil properties and land use variations around the site, we do not expect that the trends will be affected by it.

2. Changing soil conditions
Changing soil conditions (e.g. compaction) are mainly related to changing land use. Therefore, soil compaction is handled as an effect of land use change. Land use changes are already part of our framework (as mentioned above).
3. Soils
As mentioned before we assume that the deficits in the soil parameterization are mainly affecting the absolute values and not the changes or trends. However, different soil types will be also part of the sensitivity study, which will be included in the modified version of the manuscript. We will also deliver the exact soil parameters (e.g. Van Genuchten) instead of soil types or classes.
4. Runoff
We neglected surface runoff, as most of our observations sites are not situated in heavily urbanized areas. Since Sklash and Farvolden (1979), several studies have shown that outside of urbanized or arid areas the major part of discharge originates from the groundwater, and hardly any surface runoff is generated. Especially Hortonian runoff is seldom generated in humid, planted area. Saturation excess may play a role but this runoff generation process is usually encountered next to the receiving streams. We selected only observations sites with quite a distance to streams, so also that process seems not to be very important. However, we will also evaluate the influence of runoff in the sensitivity study and report the results in the modified version of the manuscript.
5. Potential evapotranspiration
We used the widely used Penman-Monteith approach, which we assumed to give reliable and better estimates than more simplistic approaches. Again, we assume that the choice will mainly affect the absolute values rather than the trends. In particular, it is important to re-stress that according to the model calculations there is not a significant trend in recharge over the past (1930-2010). If we would consider uncertainty of potential ET, this would not affect this conclusion. In fact, if we consider uncertainty of potential ET it is less likely that we would find a significant trend as the uncertainty associated with the estimated trend line would increase. The conclusions could only change if we would have bias in estimating potential ET which would be different for the past than for present times. This is highly unlikely and not worth of serious consideration.
6. Plausibility
We did perform an assessment of the plausibility. Every model result is compared with observed data and correlation coefficients are calculated (Table 2, Figure 2/3). Besides, the match is quite good with linear correlation coefficients between observed and simulated data around 0.5 (averaged over all sites).

Rather than the authors statement that “if certain patterns of change are observable in the groundwater level and spring data, but not in the recharge calculatiuns, the observed patterns are likely to be related to land use changes” (P7632, L19-21), I beleive that they are as likely (or more likely) to be due to conceptualisation, model and data limitations.

As mentioned, we will perform a sensitivity study to evaluate the assumptions. Besides, the main features are not related to land use changes. We state, that the most dominant features are related to pumping activity. Additionally, for this study we selected unconfined, recharge dominated and predominantly shallow aquifers with quite easy *precipitation-recharge-groundwater response-relationships. Thus conceptual and parameter uncertainty should not be of great importance.*

3) In P7631, is it surprising that there should be correlation over space, given that it has been assumed that there is no runoff, that the soils are spatially identical, and differences in simulated percolation/potential recharge are driven solely by precipitation and actual ET? PET generally has large scale spatial structure, as do dominant modes of precipitation, whether convective, orographic or frontal? No information is given as to the main precipitation drivers in the region, which might be useful to interpret the findings.

We think that the spatial coherence of recharge over large distances is an interesting result of our study, especially as this result can be used as a decision help for the choice of an appropriate downscaling method. Additionally, to our knowledge, there hasn't been a study dealing with large scale patterns of recharge. Especially the fact, that not only the model results but also the observations show this pattern is an interesting feature.
We will gladly incorporate more information about the dominant precipitation modes in the studied region in a revised version of the manuscript.

4) The assumption of the perfect model is also apparent when the paper identifies abstraction impacts – "some groundwater levels show much more pronounced drought periods than model calculations or spring discharges: : :.. this feature would be associated with pumping activity" (P7633). If the actual soil at some sites/catchments deviates from that assumed in the modeling, this could have an important impact upon the actual ET, maximum soil moisture deficit and the timing of the onset of the autumn/ winter recharge. Whilst irrigation might be important as suggested by the text, no information is presented to substantiate this.

We disagree with the reviewer, implying that we assume our model to be perfect. As we pointed out before, we are aware of the uncertainty related to the modeling. In the specific case of the pronounced drawdowns, the reviewer neglects, that we compare the groundwater levels not only with model results but also with observations of springs. If the mentioned feature would be related to the type of the soil, than, this behavior should also be visible in at least one of the 15 spring observations.
However we accept the doubts of the reviewer concerning our modeling approach that is why we will also analyze this behavior in the sensitivity study mentioned before. We will also try to give more information about irrigation.

Technical corrections
The text on P7627, L29 to P7628, L5 is unclear

We thank the reviewer for the indication. We will reformulate the text.

Conclusion 1 – the "strong relationship between winter precipitation and groundwater droughts" hardly represents the importance of the "detailed representation of the temporal precipitation distribution"! Besides the importance of temporal distribution of precipitation has been shown in the results of a number of earlier climate impact studies.

It is true, that this important feature has been shown in earlier impact studies. We will gladly incorporate the literature. We disagree with the reviewer that this does not represent the importance of a detailed representation of the temporal precipitation distribution. As the reviewer has pointed out, the precipitation (besides ET) is the main driver of recharge. Therefore, if the distribution of precipitation is not correctly described in the impact studies (for example by underestimating future winter precipitation) this will lead to wrong projections of future recharge. Applying just general annual change factors is not sufficient as also the intra-annual distribution is changing. As mentioned in the manuscript this is also important when choosing a downscaling method. Nevertheless, we will modify the text to more clearly connect the importance of this conclusion with the downscaling method.

Conclusion 3 mixes the impacts of future climate change, non-climate/socio-economic change and adaptation responses.

We thank the reviewer for the indication. We will reformulate the text. However we think, that also non-climate/socio-economic changes should be considered (if the information is available) in climate change (or global change) impact studies, as these can have equal or even more severe consequences for the water demand.

Sensitivity study

Below you can find the results of the mentioned sensitivity study for the groundwater level observation site Niederstotzingen. The figures show the results of eight different model setups. We used four different soil types (from sand (ks = 10^{-4}m/s) to clay (ks = 10^{-7}m/s)), combined them with two land uses (forest, meadow) and allowed the model to produce surface runoff. Figure 1 shows the absolute annual recharge sums. As we assumed, large differences can be observed. While the forest-on-sand combination (green line) shows the highest recharge sums and no runoff, the combination meadow-on-clay (blue line) showed least recharge and most runoff. For this extreme combination, the model even reports negative recharge, due to evapotranspiration from the capillary fringe.

However, as it comes apparent in Figure 2, the differences are hardly visible if the recharge is normalized. The dynamic as well as the trends are almost identical for the different model combinations. Therefore, we think, that our modeling approach is appropriate to analyze the recharge dynamics and trends. The uncertainty related to our assumptions about soil types, land use and runoff seems to be relatively of limited importance. The misfit of the pronounced drawdowns cannot be explained by the assumptions about soil types, land use and runoff, as the reviewer assumed in his fourth comment. Therefore other processes (in our opinion: pumping) have to be responsible.

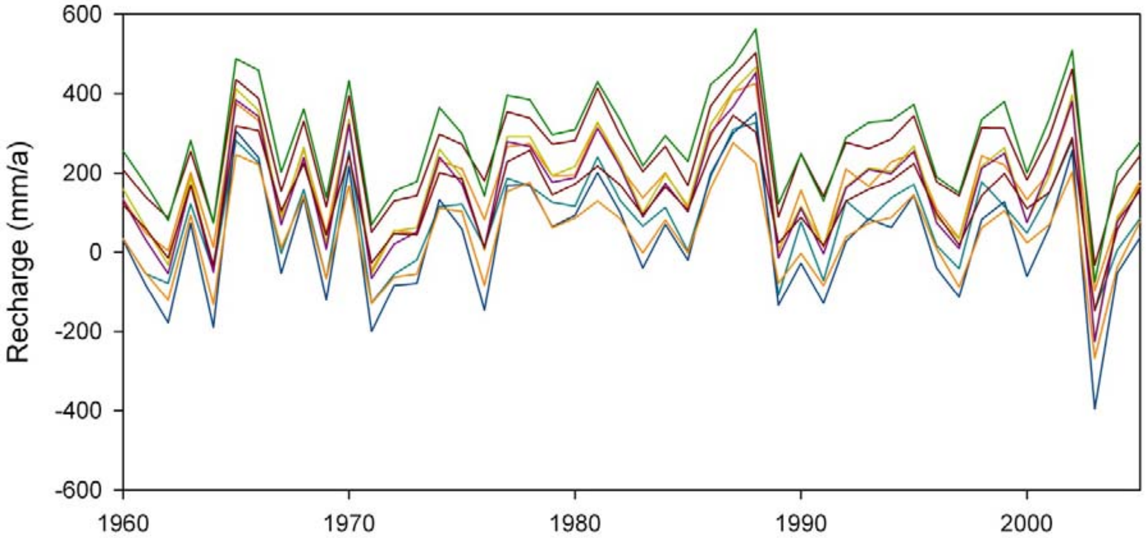

Fig. 1: Absolute annual recharge sums for different model combinations.

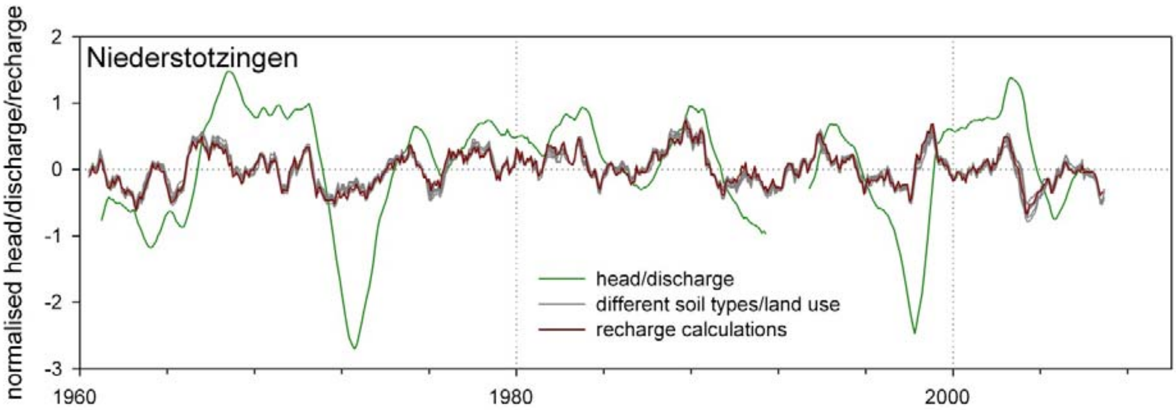

Fig. 2: Normalized monthly recharge sums for eight different model combinations (grey), the result from the original model setup (red) and the observed normalized groundwater level (green).

References:

Blöschl, G. and Montanari, A.: Climate change impacts-throwing the dice?, Hydrol. Process., 24, 374–381, doi:10.1002/hyp.7574, 2010.

Sklash, M. G., and Farvolden, R. N.: The role of groundwater in storm runoff, Journal of Hydrology, 43, 45-65, 1979.