We thank Professor Dr. Marc Bierkens for his time spent in reviewing our manuscript and appreciate the positive and constructive criticism. Hopefully we have answered the questions satisfyingly, otherwise we are available for further clarifications.

General remarks

The paper uses an integrated hydrological model (MIKE-SHE) and the results of 8 RCMs driven by various GCMs to assess climate change effects on groundwater levels. I feel that this is a very thorough study using state-of-the-art tools. I applaud the authors for being so very critical of their own results, basically coming to the conclusion that the quality of the GCM-RCM predictions, even when downscaled with various statistical downscaling methods, is not sufficient to be used in planning future water resources. The paper is well written and can be published with little modifications. I do have some issues though, which are the following:

- The first method, where only the change factors per month are applied seems a little naïve. Usually when such a correction is used, also the number of rain days is corrected (see e.g. Sperna-Weiland et al., Hydrol. Earth Syst. Sci., 14, 1595–1621, 2010, for a recent example). If this is done by finding a threshold to reproduce the number of observed rain days and keeping this threshold constant for future projections, one is also able to account for the changes in wet days. If I am correct, this cannot be done with the cpdf methods where P(wet days) remains constant. This is rather critical when analyzing hydrological effects. I do not know how much trouble it is to include such a downscaling method, i.e. change factors per month including number of rain days, but it would certainly make the paper more complete.

Answer:

We agree with the reviewer, that the Factor correction is a very simple approach. However, these kind of simple factor correction methods are still very common in the hydrological community (Anandhi et al. 2010) thus we wanted to evaluate one of them. We are aware of the limitations of the Factor correction that is why we have compared this method with other approaches which can for example handle the problem of the wet days. Both CDF methods are able to correct the number of wet days which becomes visible in Figure 5. If we would make the proposed improvements concerning the Factor Correction we would end up with three quite similar methods and we would gain little information about the consequences of the use of the factor correction methods.

- Related to that: the authors conclude that none of the corrections can do both: remove monthly bias and preserve inter-annual variability. This is precisely why people use weather generators to simulate future climate, i.e. where changes in the parameters of the weather generators are inferred from the GCM/RCM changes (see for instance Burton, Kilsby et al. Environ. Modell. Software 23, 1356–1369).

Answer:

We agree with the reviewer, that weather generators are possible alternatives to the used downscaling approaches. Weather generators are usually applied to downscale general future monthly climatic trends to submonthly resolution. As we used the complete daily time series (from ENSEMBLES) until 2100 this was not necessary for this study and did therefore not appear as a logical option. Intuitively, if complete time series from a climate model are available that

represent the interannual variations, including extreme events, and also the variability on smaller temporal scales, one would feel that with this information weather generators could be outperformed. Weather generators are known for reproducing not very well extreme events and interannual variability (e.g., Katz and Parlange, 1998; Journal of Climate). Nevertheless, weather generators could have been used to generate stochastic time series on the basis of global change statistics extracted from the ensemble.
In this study we limited ourselves to the presented methods. In the revised version we will argue that weather generators are a further alternative, and also shortly discuss why we did not use them in this study and what are the benefits and limitations of stochastic weather generators.

- I would like to have a more in depth analysis on why groundwater recharge and thus levels increase due to the climate projections. For instance, more precipitation in winter could have easily lead to more runoff and no noticeable increase of groundwater recharge if soils below the snow pack are frozen. Also if precipitation falls as snow, most of it will runoff if during melt the soil is saturated. The combination with increased temperature however may be the cause for less frozen soils (probably not accounted for in MIKE-SHE) and precipitation falling more as rainfall instead of snow causing a more gradual input of water and therefore increased groundwater recharge. Anyway, it would be good to show a full water balance (winter, summer and whole year) including precipitation, interception evaporation, transpiration, runoff and groundwater recharge for current climate and projected for 2100 to understand what happens.

A limitation of MIKE-SHE is that frozen soil is not accounted for. In the revised version of the manuscript this will be acknowledged. However, we expect that it has only a minor impact on the results. In this context it is very important to stress, that there is no surface runoff leaving the system. Precipitation either recharges or is converted to evapotranspiration in terms of the water balance. But the reviewer is right that still during a significant part of wintertime (DJF) soil is frozen. However, also in that case we expect only limited surface runoff. There are no visible creeks or small streams in the study area and if surface runoff is created, it infiltrates again in local depressions within this area (run-on). If the soil is frozen, during a longer time such local water filled depressions might appear during winter time. We visited once the area during/after a significant rainfall event in winter, with snow patches left and frozen soil conditions slightly below the soil surface. We observed surface runoff along the streets in the study area, but no surface water runoff along small creeks or streams. We saw local depressions in the landscape filled with some water. Finally, the role of frozen soil conditions would be that recharge would be overestimated, especially for the past and the current conditions, but this overestimation would be smaller for future conditions. In this respect, it does not change the main conclusion that these model simulations do not point to increased groundwater stress.
As there is no runoff leaving the system, all relevant changes in water balance (recharge, evapotranspiration and storage) relative to the reference (Table 5) are already provided in monthly resolution. The reviewer also asked for the evaluation of transpiration and interception. We decided not to split evapotranspiration because (as the reviewer mentioned himself) vegetation is represented rather simply in the study. Therefore changes in e.g. transpiration are not expected to give reliable additional information. The temporal changes in evapotranspiration will be equally divided over transpiration and evaporation.
The process of snow pack formation and melting is accounted for by MIKE-SHE, although it is implemented with a rather simple parameterization (degree day approach).

We will acknowledge the mentioned limitations in the study, add additional information where necessary and discuss it in more detail in the paper.

- When it comes to analyzing changes in groundwater recharge, the response of vegetation to climate change may very important. This does not only pertain to the physiological response of vegetation to CO2 increase, but also the effect of frequent and prolonged summer drought stress on vegetation density and biomass (LAI) and through this on transpiration and interception evaporation. For instance, from our own work (Brolsma et al., 2010, Water Resources Research 6, W11503, doi:10.1029/2009WR008782), it follows that depending on the severity of projected summer droughts, groundwater levels can either increase or decrease depending on whether vegetation density decreases or not.

Answer:
We completely agree with the reviewer, that vegetation plays a major role in assessing climate change impacts. In this study we focused on the effects of the downscaling process and the climate models. However, currently we are focusing more on the role of the representation of processes for climate change impact studies and we look at the role of land use changes (including vegetation), changes in the length of the phenological cycle etc.

We will also acknowledge in the paper this limitation of our study, including adequate references like the one that the reviewer provided. However, it is good also to re-stress at this place that concerning the impact of climate change on groundwater resources we already went beyond existing studies by using a coupled representation of the unsaturated zone- saturated zone, more physically based estimates of evaporation and transpiration, the representation of interception, surface runoff and snow pack accumulation/melt. In addition, a multi-model approach was used including different downscaling methods (on which we focus in this paper). But the reviewer is right that a dynamical representation of vegetation changes is another important topic.

Some small remarks:
- In the review about climate change effects on groundwater levels, perhaps the following paper is relevant: Roosmalen et al (2007), Regional differences in climate change impacts on groundwater and stream discharge in Denmark, Vadose Zone J., 6, 554–571, doi:10.2136/vzj2006.0093.

Answer:
We appreciate the hint and will add the paper in our review.

- The spikiness of observed groundwater heads at Girhalden (Fig. 3) that cannot be reproduced by the model cannot be from pumping because it shows head suddenly rising, not falling. So the explanation in the text is not satisfactory.

Answer:
By default, at Girhalden, the pump is running every day. The spikes occur when the pump is shut down (e.g. for maintenance). This will be clarified in the text.

- Page 2, column 1, line 11: change to “For instance, Fuhrer and Jaspen (2000) state that: : :

Answer:
We thank for the hint and we will change the text accordingly.

- The validation period is 12 years: is this sufficiently long to correctly estimate monthly biases of each GCM? Maybe this explains the poor performance of the cpdf downscaling methods?

The bias is estimated on the basis of the calibration period (1961-2000) which is much longer. The relatively small verification period (10 years) is just used to evaluate the performance of the downscaling methods. Despite this rather short evaluation period, significant and explainable differences between the downscaling methods could be identified.

References:

Anandhi, A., Frei, A., Pierson, D.C., Schneiderman, E.M., Zion, M.S., Lounsbury, D. and Matonse, A.H.: Examination of change factor methodologies for climate change impact assessment, Water Resour. Res., doi:10.1029/2010WR009104, in press, 2010.

Katz R. W. and Parlange, M.B.: Overdispersion phenomenon in stochastic modeling of precipitation, Journal of Climate 11, 591–601, 1998.