This manuscript describes the application of an existing hydrologic model for karst aquifers. The approach of evaluating the model calibration in terms of hydrologic exceedance rates appears to be a new and useful approach. Exceedance rates of projected hydrologic simulations are evaluated for human safety or the needs of species (e.g., https://pubs.er.usgs.gov/publication/sir20145089). Therefore, if a model is to be used for this purpose, it makes good sense to evaluate the model directly on the basis of exceedance frequencies.

We thank Dr Andy Long for his positive evaluation and his valuable and helpful comments.

Main comments:

1. *The abstract explains that the approach to simulate groundwater level frequency is novel. I would say that this is not the novel part, because the time-series records were simulated and simply converted into frequency distributions, which is a common way to summarize hydrologic time-series records. However, the novel part is that the model calibration is evaluated on the basis of frequency distributions, which I have not seen before, and I suggest presenting it that way. The tile is more accurate: "A percentile approach to evaluate simulated groundwater levels and frequencies. . ."*

We will change the title and the abstract accordingly.

2. *Section 5.2 discusses the possibility of focusing the calibration on high percentiles. I don't totally agree that longer time-series records would be needed to do this, and this section could benefit from further discussion of this idea. For example, an approach could be developed to evaluate the usefulness and data adequacy of such an endeavor. You could vary the weights within the observed time-series record for individual observations at different exceedances to tailor the calibration to a target percentile. It would be possible to calibrate to a different weighting scheme for each percentile. Further, an uncertainty analysis could be applied on each separate calibration run, and quantifying the presumed decrease in uncertainty as the percentile increases could be useful. Then, when you make predictions for different percentiles, you can also report the differences in uncertainty. This idea also applies to section 5.3, which discusses the prediction of increased drought.*

We will take these considerations into account and incorporate a more detailed discussion on high percentile calibration at the end of Section 5.2.

3. *The Introduction discusses risks to events such as groundwater flooding and drought. I suggest adding a short statement to this effect in the Abstract to emphasize the need for this study in terms of natural hazards.*

We will add a statement in the abstract.

Other comments:

1. *p. 2, lines 24-26: indicates that karst groundwater levels were simulated by lumped models only in a few instances, but see also Long and Derickson (1999), Long and Mahler (2013), and Pinault et al. (2001).*

We thank the referee for this useful suggestions. We will add Long and Mahler (2013) and (

Derickson and Long (1999) to the literature comparison. We could not find any simulation of groundwater levels in the work of Pinault et al. (2001) and we added the work of Kumar et al. (2016) instead.

2. *p. 3, lines 22-23: describes a new approach to show groundwater levels as frequency distributions. Showing hydrologic time-series data as frequency distributions is a common method. Please explain how this is new, or describe it differently.*
3. *p. 3, line 35: "PET" should be defined.*

We agree with both comments and we will change the text accordingly.

4. *p. 4, line 34: discusses a "weighting scheme." I think the calibration weights are applied to observations, but that should be explained here for clarity.*

A clearer description of the model and parameters estimation will be provided in the revised manuscript (see also specific comment 1 of the other referee).

5. *figure 7: what is the meaning of "manipulated" in the caption?*

This refers to the manipulation of our observed "baseline" data, see chapter 3.5. We will clarify this in the caption of figure 7 in the revised manuscript.

6. *table 5: I think these result apply to a particular model time step (e.g., daily), but I'm not sure. Please clarify.*

These are the mean model outputs (Qsim, AET) and exceedances per year in the simulation period 2070-2099 which was calculated on the basis of our model which runs at a daily time step. We will clarify this in the revised version.

References

Derickson, R.., Long, a. ., 1999. Linear systems analysis in a karst aquifer. J. Hydrol. 219, 206–217. doi:10.1016/S0022-1694(99)00058-X

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, 1117–1131. doi:10.5194/hess-20-1117-2016

Long, A.J., Mahler, B.J., 2013. Prediction, time variance, and classification of hydraulic response to recharge in two karst aquifers. Hydrol. Earth Syst. Sci. 17, 281–294. doi:10.5194/hess-17-281-2013

Pinault, J.-L., Pauwels, H., Cann, C., 2001. Inverse modeling of the hydrological and the hydrochemical behavior of hydrosystems: Application to nitrate transport and denitrification. Water Resour. Res. 37, 2179–2190.