Response to referee #2 on review of "Estimating catchment scale groundwater dynamics from recession analysis- enhanced constraining of hydrological models", by T. Skaugen and Z. Mengistu

First of all, we would like to thank the referees and the editor for taking their time to read closely and comment thoughtfully on our paper. Time for such tasks is hard to find so it is very appreciated.

Referee #2

General comments

1.They show that the performances of both models are comparable but that the new version of the model produces more realistic recessions. With these results the authors conclude that their new approach is another step towards simulation of ungauged basins.

Response: The point of this study was to see if we could estimate the parameters of the subsurface routine from observations apriori to calibration against streamflom (and precipitation and temperature) and still obtain reasonable hydrological simulations. Such a procedure will reduce the model parameters dependency on the input (precipitation and temperature) and reduce the equifinality problem (see response to first general comment of R#1). The fact that a reduction in free calibrations parameters will enhance the model's ability to predict in ungauged basins (see Seibert, 1999; Skaugen et al. 2015) is a pleasant consequence of a model with less parameter uncertainty. In Skaugen et al. (2015) all the DDD model parameters were shown to be significantly correlated with catchment characteristics.

2. First of all, I found the manuscript very hard to read. The theory is too long and the structure of the manuscript is structure confusing. This is particularly true for the explanation of the theory: I definetly recommend reordering the sections (2.1 Hydrological model, 2.2 Runoff dynamics, 2.3 Reformulation, 2.3.1 Estimating the mean storage, 2.4 Example..) The old calibrated model should be explained in much less detail supported by more detail in the appendix. The Reformulation should be structured in a better way and if there is no 2.3.2 there is no purpose in having a subsection 2.3.1, etc. Generally, discard all information that is not completely necessary.

Response: We agree that the theory part (section 2) is long. We propose to restructure as follows:
-We can move the part where we relate the distance distribution to the linear reservoirs to an appendix (P11135-L14—P11137-L6). This is with a heavy heart as we consider this a very important result of the paper, but is, perhaps, not directly relevant for the new subsurface routine.
-In subsection 2.3, the part where we justify the similarity in shape of the distributions of Λ and S, (P11139-L23 (The assumption..)—P11140-L24), is moved to the discussion.
-In subsection 2.4, the discussion around large sample hydrology (P11143-L13—P11144-L4) can be moved to the discussion.
- In a comment below (comment #4), R#2 suggest that we stay with V1, only one meteorological grid for our comparison. In that case a large part of section 2.4 will be deleted, and probably make the paper more easy to read.

3. Also, more focus should be put on the example where application, parameter estimation, evaluation and data should be explained in a structured way (more subsections). A lot of theory is also presented at the beginning of 2.4 (1st and 2nd paragraph), which should rather be moved to the discussion as the studies methods and the methods of other studies are quite mixed up now.

Response: Yes, see response above (for comment #2)

4. As new calibration is performed for DDD__M a strict evaluation of the observation derived parameters of the subsurface routine is difficult. I recommend rather remaining with all calibrated parameters of the

old surface routine with the old meteorological grid V1, which would allow estimating the skills of the new parameter estimation scheme without any calibration. Also this would make the approach simpler and the paper easier to read.

Response: We have done a preliminary study and get similar results. The difference in NSE is 0.02 (0.77 vs 0.75) and the difference in KGE is 0.01 (0.82-0.81), both in favour of the calibrated version. In our view the conclusions in this paper, that no precision is lost, still holds. The difference in precision is very small and one must bear in mind that such a comparison favours θ_M since it is optimized together with the other parameters for V1. The value of these parameters are not necessarily optimal for the new subroutine. However, we agree with the reviewer that readability will improve.

5. Finally, as the new parameter estimation scheme still requires discharge data I also do not see the real advantage in terms of simulating ungauged catchments. I agree with referee #1 that there is the need to apply the new approach for simulating catchments without discharge data as mentioned in the discussion. Most desirably at least first try should be part of the study indicating the advantages of this new approach.

Response: This comment is similar to the first general comment and to comment 12 of R#1, please see response to those. In addition, we can elaborate a bit further in the discussion on the PUB potential of the new structure of the DDD model. From a study of 84 Norwegian catchments we have found that the parameters of the distribution of Λ were highly correlated (r2- 0.97 and 0.98) to the parameters of λ. This is not a surprise since λ is derived from Λ (see eq. 12). The new model structure of DDD has hence effectively one parameter less (θm) to estimate from catchment characteristics for application to PUB, and Skaugen et al. (2015) show that the parameters of λ, and hence Λ, can be determined from catchment characteristics.

The focus of this paper is the introduction of the new subsurface routine with parameters estimated outside of the (classical) calibration. In our view, including a new PUB analysis would also include another scope (see also response to your first general comment).

Specific comment

1.p11131, L25.. mention importance of reliable estimation of storage –discharge relations as pointed out by Berghuijs et al.2016)

Response: We are not certain this is a correct place to have this reference, since we discuss linear reservoirs. It will fit very well, however, at P11132-L18.

2.p11138-L10: mention height of each storage level

Response: Yes this will be mentioned earlier

3. p11130-L9. mention variability of aquifer porosity

Response: Yes, a good point.

4.11144-L16 it is enough just to mention which data was used

Response: This part will be rewritten (see response to general comment #2)

5. 11144-L24 Can these parameters be assumed to be constant?, see studies Merz et al. 2011 and Berghuijs et al, 2014.

Response: Only partly, temperature threshold for snow melt should be 0 °C, but the others account for data and model structure errors. This part will anyway be deleted in the revised paper.

6. 11147-L15 This should be part of the study (PUB excersise)

Response: See response to general comment #1 and #5.

7.11148-L3: this makes sense even more in regions with heterogeneous subsurface, see studies (Refsgaard et al (2012) and Hartmann et al. 2015)

Response: Thank you for the references, they will be included.

8. 11148-L15 refer to TTD papers that make a similar point (Harman C.J 2014 and Kirchner, 2016.)

Response: Thank you for the references, they will be included.

9. Figure 4 Increase font size

Response: Yes

10. Have lambda symbol in Figure 6.

Response: Yes