Response to Report #1, Anonymous Referee #2

The authors have added model calibration/validation to the manuscript, which is clearly novel compared to the previous Zlotnik et al. publication. This addresses and at least partly resolves the most concerning issue I had with the last version.

We appreciate the reviewer taking the time to provide another detailed review of our manuscript. The major improvements to the manuscript include:

- Reorganization sections to begin with the calibration and follow with the sensitivity analysis. This significantly improves the flow of the manuscript.
- Details on the potential errors in the driving data are now provided and discussed. We state the resolution and/or accuracy of the measurements. We also provide standard errors for the parameter estimates from the calibration informing the reader how well constrained each parameter is and discuss their implications.
- It is now clearly stated how the current manuscript extends and expands on the results from Zlotnik et al. (2020).
- The role of the precipitation multiplier as not only a means to deal with rain gauge under catch, but also as a way to account for the inevitable runoff from rims into the center pond is clearly stated.
- It is now clearly explained that continuous local ET measurements are not available for our site over the 2013 thaw season. It is also explained that a large scale ET product is sufficient for our purposes given the continuous (albeit diurnal) low magnitude of ET, in contrast to the episodic precipitation record where local measurements are crucial.

Detailed responses to reviewer comments are provided below.

While the novelty aspect is improved, the manuscript needs restructuring before it can be considered for publication. The new paragraphs on calibration/validation have largely been inserted in the already established structure of the previous version, but this does not fit in my opinion. For example, Sects. 3.1 - 3.3 elaborate on the model sensitivity, but then Sect. 3.4 states that the simple model (which is the basis for 3.1-3.3) does not fit the measured data, but needs to be modified as in case 3 (I like the clear and unambiguous language on the model evaluation in 3.4!). This can be confusing for the reader, although most (or maybe even all) of the analysis in 3.1-3.3 still holds.

As the reviewer suspects, all the sensitivity analyses still hold because they are snapshots in time using nondimensional parameters. Therefore, the sensitivity analyses are not affected by the precipitation factor or the thaw-depth dependent Kz identified in the calibration. This is now clearly stated in section 3.1 “Calibration to field measurements:

“The calibration verifies that the model is able to capture ice-wedge polygon drainage characteristics. In the next sections, we perform sensitivity analyses using non-dimensional forms of this verified analytical solution to gain insights into ice-wedge polygon drainage characteristics. The use of non-dimensional solution snapshots eliminates the need to consider the precipitation multiplier and thaw-depth dependent vertical hydraulic conductivity explicitly. Instead, their effects are implicit in the relative differences between snapshots."

So my suggestions for restructuring and further analysis are:
1. Make it clearer in the Sections on study design that calibration/validation of the Zlotnik-model is one of the primary goals of the study. This part clearly goes beyond the Zlotnik et al.-paper, and that is what the revised manuscript must demonstrate. The authors write in their reply that the manuscript "rigorously confirms, fully fleshes out, and expands on" Zoltnik et al., which could be seen as a positive formulation for "the conclusions are essentially the same". So the authors should extend and deepen the new part and provide more extensive information how the cal/val model runs are set up. Also mention the spatial and time resolution and known or probable uncertainties/ error characteristics of the driving data sets (e.g. for evaporation) and how this affects the calibration procedure.

We agree that the paper is better organized in the way that the reviewer suggests, and we have switched the order throughout the manuscript. We are thankful for the suggestion. We also appreciate that the reviewer recognizes that the calibration/validation "clearly goes beyond the Zlotnik et al. paper". In Section 2.2 "Calibration approach", we provide a complete description of how the cal/val model runs are set up. We state that we use a Levenberg-Marquardt approach and define the objective function and the parameters and meta-parameters for each calibration case. This provides a comprehensive description of the cal/val model setup for readers.

We do not agree with the reviewer's characterization that "the conclusions are essentially the same" from the sensitivity analysis in this manuscript as what is in Zlotnik et al. (2020). Aside from the new information provided in the calibration, Figures 4-11 greatly expand on the preliminary scoping analysis in Zlotnik et al. (2020) providing new insights and information. This is clearly stated in the abstract:

"We also provide a comprehensive investigation of the effect of polygon aspect ratio and anisotropy on drainage timing and patterns expanding on previously published research. Our results indicate that polygons with large aspect ratios and high anisotropy will have the most distributed drainage, while polygons with large aspect ratios and low anisotropy will have their drainage most focused near their periphery and will drain most slowly. Polygons with small aspect ratios and high anisotropy will drain most quickly."

Also, it is readily apparent in Figures 7 and 10 that new information not provided in Zlotnik et al. (2020) is being presented in global sensitivity analyses of geometry and anisotropy which provides detailed non-linear nuances in drainage patterns due to combinations of geometry of anisotropy. The preliminary scoping analyses of Zlotnik et al. (2020) did not provide this information. This is indicated and elaborated on in the conclusions:

“We provide rigorous confirmation that the majority of drainage from inundated ice-wedge polygon centers occurs along an annular region along their radial periphery; however, polygon geometry and hydraulic conductivity anisotropy significantly impact the drainage pathways, as originally postulated by Zlotnik et al. (2020).”

The conclusions continue providing details on the global sensitivity insights derived from our manuscript that are not provided in Zlotnik et al. (2020).

The probable uncertainties/error characteristics of the driving data sets are now fully described in Section 2.4 “Acquisition of field data used in calibration” where the known accuracy or resolution of datasets are stated. We also now describe in detail why we used a large-scale ET product and the implications that has on our analysis:

“Due to a lack of continuous local evapotranspiration measurements, we obtained evapotranspiration data from NASA’s Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004). Given the continuous (albeit diurnally fluctuating), low magnitude evapotranspiration signal, its effect on our calibration is relatively insignificant compared to the sporadic precipitation events that drive large scale fluctuations in water levels. Therefore, in lieu of local evapotranspiration measurements, the GLDAS evapotranspiration is deemed sufficient for our purposes here.”

The probable uncertainties/error characteristics of the driving data sets are also now discussed in section 4.2 “Calibration implications”:

“The water level (Liljedahl and Wilson, 2016) and temperature measurements (Romanovsky et al., 2017) are well constrained with high degrees of resolution and accuracy. While the resolution of precipitation measurements is high (0.1~mm), there is the potential for under catch during windy precipitation events. More importantly with regard to ponded water levels, the precipitation measurements do not account for the runoff of water during precipitation events from rims into the polygon center, resulting in measured precipitation less than increases in pond water levels. This uncertainty is accounted for through the calibration of a precipitation multiplier, which effectively captures the effect of runoff from the polygon *watershed* into the center pond along with any potential under catch during windy precipitation events.”

2. Start Sect. 3 results with the cal/val section (now 3.4), and based on the findings motivate the sensitivity analyses 3.1-3.3 and why they are still meaningful in the light of the cal/val findings. I am also missing a closer analysis of case 3 and the sensitivity of the resulting parameters. I guess equifinality could be a significant problem here, so how well-defined is the minimum in RMSE? Are there other parameter combinations that result in a similar RMSE? Are some of the parameters better constrained by the analysis than others? For example is the value for the minimum hydraulic conductivity (about 5e-8 m/sec which is a reasonable value for poorly permeable silt) well-defined, or would a whole range of values, from e.g. 1e-6 to 1e-9 m/sec, provide a similar performance? How do uncertainties in input parameters, in particular

evaporation (which also leads to a drop in water levels) affect the calibration? The authors seem to use a large-scale ET product (which integrates over rims and centers?), but the model needs precipitation for a wet polygon center. Have the authors tried an evaporation multiplier?

The results section now begins with the cal/val section. The sensitivity analysis is now motivated at the end of the cal/val section as:

“The calibration verifies that the model is able to capture ice-wedge polygon drainage characteristics. In the next sections, we perform sensitivity analyses using non-dimensional forms of this verified analytical solution to gain insights into ice-wedge polygon drainage characteristics. The use of non-dimensional solution snapshots eliminates the need to consider the precipitation multiplier and thaw-depth dependent vertical hydraulic conductivity explicitly. Instead, their effects are implicit in the relative differences between snapshots.”

In order to address the question of “equifinality”, standard errors are now provided for each parameter in Table 1. The following discussion has been added to Section 3.1 as well:

“The standard errors of the calibrated parameters for calibration case 3 listed in Table 1 indicate how well constrained the parameters are by the calibration. It is apparent that the hydraulic conductivities (horizontal and minimum and maximum vertical) are not well constrained with relatively large standard errors. These parameters (or their meta-parameters) also have large covariances with each other indicating their correlated effect on the model. However, despite the lack of constraint of these parameters due to their correlated effect on the model, the calibration does identify reasonable values for them. The standard errors of the discharge conductance, initial polygon-center water level, and precipitation multiplier indicates that they are well constrained by the calibration.”

A large-scale ET product had to be used as continuous local ET measurements are not available at the site in 2013. Since ET is a relatively small, continuous (albeit diurnal) driver for our model, its affect is much less significant than sporadic precipitation events (refer to Figure 3). So, while ET does lead to a significant loss of water over the entire thaw season, it is not crucial in capturing fluctuations in water levels in the same way that precipitation events are. Therefore, the use of the large-scale ET product does not significantly affect the conclusions drawn from the calibration. Along the same lines, we do not feel that exploring an evaporation multiplier would provide any additional insights into polygon drainage in this case. We have added the following text in Section2.4 “Acquisition of field data used in calibration” to clarify this point for the reader:

“Given the continuous (albeit diurnally fluctuating), low magnitude evapotranspiration signal, its effect on our calibration is relatively insignificant compared to the sporadic precipitation events that drive large scale fluctuations in water levels. Therefore, in lieu of local evapotranspiration measurements, the GLDAS evapotranspiration is deemed sufficient for our purposes here.”

3. Restructure the Discussion accordingly, exactly same issue as for the Results section! Furthermore, the authors should discuss how clear model deficiencies (as raised in the earlier review report and confirmed in the author's reply) could affect the calibration and the hereof derived conclusions. In l. 521, the authors write rather boldly that the model "helps identify factors which need to be considered by any hydrologic model to simulate drainage from an inundated polygon center". Considering that the model does not account for obvious factors, e.g. different thaw depths in the rim and the center, it is important to substantiate this statement with a clear analysis of the model uncertainties and an assessment of how reasonable the resulting parameters are. A few parameters are discussed in Results, e.g. Kr, but this should be expanded to all parameters in Table 1. As an example, the precipitation multiplier of >2 is casually explained by undercatch, but is >100% undercatch possible/reasonable for summer rainfall (not winter snowfall) for the particular rain gauge type that it was measured with? The publication Pollock et al. (2018) cited by the authors reports an undercatch of 23% and discusses the effect of different rain gauge types in detail. If an undercatch of 100% was not reasonable, this would be an indication that the model produces the right results for the wrong reasons (also see my comment on an evaporation multiplier above), and this would need to be analyzed, discussed, and the conclusions adapted accordingly.

The discussion section has been reorganized accordingly. A more complete discussion of parameters is now included in the "Calibration implications" section:

"Due to covariance in the effects of the hydraulic conductivity parameters on water levels, the hydraulic conductivity parameters are loosely constrained by the calibration (based on local sensitivities). This indicates the importance of constraining the hydraulic conductivity parameters with field measurements if possible (field measurements are not available in our case). However, despite relatively large standard errors in the hydraulic conductivity parameter estimates, the calibration identifies physically realistic values. It should also be noted that despite vertical hydraulic conductivity being loosely constrained in the final calibration (calibration case 3), the model was unable to match water levels with a constant vertical hydraulic conductivity (calibration case 2). This indicates the importance of the thaw-depth dependent vertical hydraulic conductivity in the model. The other parameters (discharge conductance, initial polygon-center water level, and precipitation multiplier) are all well constrained by the analysis."

We have also added a paragraph in the "Calibration implications" section addressing data uncertainty:

"The water level (Liljedahl and Wilson, 2016) and temperature measurements (Romanovsky et al., 2017) are well constrained with high degrees of resolution and accuracy. While the resolution of precipitation measurements is high (0.1~mm), there is the potential for under catch during windy precipitation events. More importantly with regard to ponded water levels, the precipitation measurements do not account for the runoff of water during precipitation events from rims into the polygon center, resulting in measured precipitation less than increases in pond water levels. This uncertainty is accounted for through the calibration of a precipitation multiplier, which effectively captures the effect of runoff from the polygon *watershed* into the center pond along with any potential under catch during windy precipitation events.”

We agree that an under catch of >100% is not likely. However, the precipitation multiplier is not “casually explained by under catch” in the manuscript. In several places in the manuscript, we explain that the precipitation multiplier not only accounts for potential rain gauge under catch, but also accounts for the fact the pond in the center of the polygon will collect precipitation runoff from the surrounding rims and other high ground within the polygon center. The area of rims and other high ground can be a significant portion of the polygon *watershed*, resulting in the increase in the center pond height being much larger than the precipitation measured by a rain gauge. This was clearly stated in the previous manuscript and has now been elaborated on (see paragraph above) and clarified even further in the abstract:

“…accounts for runoff from rims into the ice-wedge polygon pond during precipitation events and possible rain gauge under catch…”

Section 2.2 “Calibration approach”:

“In calibration case 2, we add a precipitation multiplier M_P to case 1, where $\hat{P} = M_P P$ and $\hat{P}$ is the augmented precipitation accounting for runoff from microtopographic highs and potential rain gauge under catch.”

Section 3.1 “Calibration to field measurements”:

“Considering that precipitation will runoff from rims and collect in the polygon-center pond and that rain gauges may have under catch issues…”

And Section 4.1 “Calibration implications”:

“The first refinement is a precipitation multiplier and is based on a simple mass balance indicating that the measured precipitation cannot account for the total increase in ponded water levels after precipitation events. The precipitation multiplier accounts for the fact that precipitation will run-off from the rims into the center pond resulting in a larger increase in ponded water than rain gauge precipitation. The precipitation multiplier also accounts for any potential rain gauge under catch.”

These multiple discussions throughout the manuscript will clarify the role of the precipitation multiplier for readers.