Response to comments by Anonymous Referee #2 on "Modeling 25 years of spatio-temporal surface water and inundation dynamics on large river basin scale using time series of earth observation data" by V. Heimhuber et al.

We thank anonymous referee #2 for her/his recommendations and useful comments that helped us to improve the quality of our manuscript. In the following, we provide point-to-point answers to all of referee #2's comments. Questions raised by the reviewer are in bold face, our answers in regular face.

Heimhuber et al. produce a statistical method to model inundation extent of floodplains and wetlands. Run on a 10 km grid, the method predicts inundation extent on a given day as the inundation measured on the previous Landsat overpass, lagged discharge, and precipitation and evapotranspiration. The data-driven approach, and simplicity of the methods, as compared with large-scale physically-based simulations, make this an intriguing contribution. Some addition minor clarifications and discussion of physically-based inundation and hydrologic models would improve the manuscript.

1. Page 11852 line 15. Jung et al. also included lag times, though they were between measured elevation and inundated area. So while your statement is technically true, I think it's a little misleading. Please rephrase.

We changed the relevant section of the manuscript to account for the fact that Jung et al. are also applying lag times in a similar fashion.
Updated paragraph (Page 11852 line 15):
"Even though some of the larger study sites are divided into smaller sub-regions for modeling (Table 1), only two studies accounted for lag times between discharge recorded at the gauge (Westra and De Wulf, 2009) or water surface elevation at key points (Jung et al., 2011) and the correlated surface water extents in different areas of the sub-region."

2. Figure 2: It's impossible that there's 1 Gauge per 10 km cell, correct? Please clarify what is meant; maybe 1 gauge per zone?

As pointed out correctly, there is not one gauge per grid cell. What we tried to express here is that we used the river network to assign exactly one river gauge (the most suitable one) to each grid cell. In comparison to existing studies, we did however allow more than one gauge per EH-zone (e.g. Paroo). To clarify what is meant here, we changed the statement in Figure 2 from "One Gauge per Cell" to *"Most suitable gauge for each grid cell"*.

3. Page 11854, lines 14-21. How many gauges were used, in all? Would be great to see them mapped on Figure 1, if they will all fit.

We used 10 gauges for modeling and validating the 3 sub-regions that are presented in this paper. Since the basin-scale modeling application is not the focus of this study, we decided not to show the location of all gauges that are available for the Murray-Darling Basin (~100), since it is not directly relevant for understanding the methods, results and concepts that we present here. Therefore, we decided to show only the location of gauges that were used for modeling and validating the three sub-regions (see Figure 3).

4. Page 11856, lines 9-12. Do not need to list methods that were considered and discarded. Maybe just state the tradeoffs between different cell sizes, and the final size used.

We updated the relevant section based on the suggestions:

Original paragraph (Page 11856, lines 9-12):
"*For the definition of a suitable cell size for the sub-zonation, we considered edge lengths of 75, 50, 25 and 10 km. A finer grid leads to a decrease in the fraction of cells that contain any floodplain area and consequently also to an increase in cells that contain very small fractions of floodplain. Based on this consideration, we chose 10 km as the most suitable cell size for modeling as a trade-off between sufficient spatial detail for capturing the variability in SWD at a local scale and the suitability and spatial resolution of data for modeling the driver variables.*"

Updated paragraph:
"*A finer grid leads to a decrease in the fraction of cells that contain any floodplain area and consequently also to an increase in cells that contain very small fractions of floodplain. We chose 10 km as the most suitable cell size for modeling as a trade-off between sufficient spatial detail for capturing the variability in SWD at a local scale and the suitability and spatial resolution of data for modeling the driver variables.*"

5. Page 11856 lines 21-22 & lines 6-7 Page 11857 & Figure 2 & elsewhere: It's not necessary to provide the computer languages in which some of the processing steps were done, as the steps (such as spatial averaging) do not depend on any language-specific capabilities. Recommend removing the "R" and "Python" labels, and the outer boxes in Figure 2.

We agree that the processing steps undertaken in this analysis don't depend on a specific programming language and removed the corresponding references as suggested.

6. Page 11857 Equation 1: Differentiating using "SWD" (a quite nebulous term) vs. "SWE" is quite confusing, in my opinion. Both quantities are surface water extent, after all, it's just that one is predicted, and one is observed, and they are at different times. Can you maybe make it: SWE_t = nbeta_0 + nbeta_1 Lag(Q) + nbeta_2 SWE_{(t-1)}^{obs} : : : or something like that?

As pointed out correctly, the SWD(t) on the left side of the equation was wrong and was updated to SWE(t). However, SWE(t) is the dependent variable and can thus either be predicted (in the case of forecasting) or observed (in the case of model fitting). Therefore, we chose not to add a footnote in the form of (obs) to the SWE(t-1) term (since both SWE(t-1) and SWE(t) can be observations). The updated equation is:

$$SWE_t = \beta_0 + \beta_1 Lag(Q) + \beta_2 SWE_{(t-1)} + \beta_3 P + \beta_4 ET + \beta_5 SM + e$$

7. 11859 âˇAT Recommend changing "SWD" to "SWE", here. You're starting with your error statistics to refer to numerical predictions of SWE (although because Equation 1 is likely incorrect, it's tough to be totally sure what's going on!), rather than the more nebulous "surface water dynamics".

We agree that SWE is more adequate in Section 2.4.3 and changed it as suggested.

8. Page 11860, line 15 "discharges" should be "discharge values"

Changed as suggested.

9. Table 4: Here and elsewhere: recommend putting Discharge units into more standard units of m3/s. I find it difficult to think in terms of millions of liters in a day.

We changed discharge units from ML/day to m^3/s in the entire manuscript (running text, Figure 7, Figure 9, Table 3 and Table 4) as suggested.

10. Figure 4 helps to illustrate how the uniform 10 km square grid does not really respect the river structure. For example, different pages are calculated for different parts of the river when it is split in two parts along its centerline. The 10 km grid may be a necessary evil, rather than e.g. splitting up this area using the river as an organizing unit, and mapping units based on their distance to the outlet. Please justify the choice of the 10 km grid; pragmatism is ok, but some comments on this would be good.

We agree that the regular grid, despite being combined with hydrologically meaningful analysis units (EH-zones), doesn't really respect the structure of the river in many cases. One of the reasons why we chose a regular grid is because this was the only feasible way of modelling SWD at the scale of this study. One important step in our methodology is the separation between floodplain, floodplain-lake and non-floodplain surface water bodies, which despite the regular grid, makes our modeling units hydrologically meaningful as compared to modeling all surface water areas of a given spatial entity at once.

While in some cases, using the river as an organizing unit might be relatively straight forward, there are numerous vast and complex floodplains in the study area (e.g. Lower Paroo Floodplains) for which this approach would be very difficult and non-trivial to implement. Additionally, such an approach would not have been in accordance with the main objectives of this study. We wanted to develop a data driven, transferrable and highly automated modeling framework and approach and thus, limit steps that involve more complex site analysis and manual editing and data preparation to a minimum. Compared to existing approaches that modeled SWD on the level of the eco-hydrological zonation, we think that the grid along with the related Q-lag times is a big step forward towards more accurate statistical inundation models despite the fact that at times, this approach results in a less than ideal segmentation of the river and floodplain structure. Based on these considerations, we added a few remarks to the relevant paragraph of the manuscript.

Updated paragraph (11856 line 5-15) -additions underlined:

To enable similar modeling conditions across the study area and to identify local spatial patterns in the role of climate drivers (P, SM, ET) of SWD, we imposed a regular grid on top of the EH-zonation (Fig. 2c). Although the imposition of a regular grid at times led to a less than ideal spatial segmentation of the river and floodplain structure, it allowed us to quantify the relationship between SWD and hydrologic key parameters at a much finer scale compared to using only the regional EH-zonation. Another advantage of using a regular grid is that it is straight forward to implement on sub-continental or even global scales, and does not require a comprehensive analysis of the river structure. In regards to defining a suitable grid cell size, a finer grid leads to a decrease in the fraction of cells that contain any floodplain area and consequently also to an increase in cells that contain very small fractions of floodplain. We chose 10 km as the most suitable cell size for modeling as a trade-off between sufficient spatial detail for capturing the variability in SWD at a local scale and the suitability and spatial resolution of data for modeling the driver variables.

11. Page 11862 line 26 & elsewhere: What does "externally validated" mean, exactly?

In the case of the automated estimation of Q-lags, we refer to the fact that we used an entirely different methodology (manually comparing flow data from pairs of subsequent river gauges located along the same watercourse) to reveal flood travel times for a variety of river reaches (see Table 5). By using a different methodology and in most cases also different data (river gauges that are not used for modeling but only for validation), we aimed at validating our automatically estimated Q-lags "externally". Nevertheless, we decided that it is actually not relevant to mention "externally", since the way that we validated Q-lags is explained in detail and hence, it should be already clear to the reader that this was done by using a different method and different data compared to the automated estimation of Q-lags. Therefore, we dropped the term "externally" in the context of validation of Q-lags in all related sections of the manuscript.

12. Page 11865, line 28. What does it mean to "show similar SWD"? A similar temporal pattern? Spatial pattern? Please clarify. Maybe ": : : showed SWE timeseries similar to: : :"

We changed to: "showed SWE time series with similar temporal patterns to example cells Ex-A…"

13. Page 11868, line 10-12. Calculating a cross-correlation is just a couple of lines of code. No need to cite the language in which this was performed.

We removed as suggested.

14. Section 4.4. One thing not discussed is that the effect of a P or ET flux within a 16 day window is almost certainly a function of the given SWE. You might expect P to produce almost no SWE change if SWE is low, while during an event, or after significant P in previous time steps leading to saturated soils, the same P might well lead to a significant increase in SWE. Similar reasoning for ET. These dynamics are exactly what dynamic hydrologic models try to take into account.

We agree that the effect of P and ET flux within a 16 day window is partly a function of SWE or in other words, the current hydrological condition of the floodplain or floodplain lake area that is modeled. In the existing manuscript, we address this sort of variability in the introduction:
(11851, line 23-29) "*Even though the extent of floodplain inundation highly depends on the discharge and water level in the river, the hydrologic conditions of the floodplain as well as the local climate before, during and after a flood, play an important role in the flooding and drying behavior of floodplains. For many water bodies that are not connected to rivers, local rainfall (P) is the main source of inundation (Kingsford et al., 2001). Increased soil moisture (SM) prior to flooding usually leads to reduced transmission losses and thus to a larger flood extent and longer flood duration for a given flow level compared to dry antecedent conditions of the floodplain (Overton, 2005).*" Even though this explanation is more focused on the hydrologic condition of the floodplain prior to flooding, it gives the reader a first idea of the complex and dynamic relationship of all model variables.
Accounting for the variable effect of these driver variables on SWE would require for the model parameters to be flexible over time. Consequently, our approach of fitting a single dynamic regression equation (i.e. one set of regression coefficients) to the entire set of time series data was not suitable for explicitly capturing this variability. Nevertheless, including the previous SWE as a predictor into the model equation helped us to partly account for the current state of inundation on the modeling unit, so that our models captured the average of the variable effect of P and ET on SWE. Considering that in any case, the contribution of all the additional predictor variables (P, ET, SM) to explaining

SWE was small compared to discharge as the key driver for floodplain inundation, we argue that even a dynamic hydrologic model would hardly reveal significantly different effects of these driver variables on the basis of the available data. In the discussion section of the manuscript, we addressed this problem with the statements in line 17-22 on page 11877. Based on the above mentioned considerations, we added a few remarks to the original paragraph:

Updated paragraph (11877, line 12-25) -additions underlined:
"*Overall, the improvements in CV-RMSE after accounting for the APV were small compared to the improvements after accounting for the LDV in all sub-regions, which is likely a result of the complex relationship between SWD and the local climate drivers. This relationship is non-trivial to quantify due to the absence of spatially continuous information on inundation volumes and infiltration rates (missing variables) as well as spatial and temporal resolution limits of the datasets used here. A conceptual water balance model (i.e. a dynamic hydrologic model), as an alternative modeling framework, capable of accounting for all fluxes of water into and out of each spatial modeling unit, would likely be limited by the same factors. Such a modeling approach might, however, be more suitable for capturing the variable effect of P and ET on SWE over time, resulting from the fact that a certain strong rainfall event is more likely to influence the SWE during times of inundation as compared to dry conditions. Alternatively, in an event-based modeling approach, the APV could be used to characterize the antecedent conditions of floodplains for each modeled flood, which are known to have a strong influence on the magnitude and duration of inundation on floodplains. Such an event-based modeling approach may be useful for certain applications but is not suitable for modeling SWD continuously through cycles of flooding and drying as done in this study.*"

15. Page 11878, line 21-23. Well, you should here and elsewhere really be acknowledging the continental-scale inundation dynamics models. There are quite a number of these, and they are getting more and more skillful. I know you say that this analysis is easier to do than theirs, but additional distinguishing characteristics of this analysis vis-a-vis the existing physically-based models would be really useful. Here are a few obvious ones to cite:

Thank you for pointing out the advances in continental to global scale hydraulic modeling applications and why they are related to our SWD models. We included a brief section in the introduction, where we distinguish our models from these models (see below) and also added a few remarks about this in the discussion. The additional text blocks along with their location in the manuscript are:

Updated paragraph (11849, line 12-21) -additions underlined:
Consequently, there is an urgent need for improved management and restoration of terrestrial surface water resources, which requires cost effective methods for mapping and analyzing the distribution and dynamics of surface water across large spatial and temporal scales (Alsdorf et al., 2007; Bakker, 2012; Finlayson et al., 1999; Vörösmarty et al., 2015). Recent advances in the availability and spatial and temporal resolution of geospatial data along with improved processing capabilities enabled the development of various continental to global scale hydrodynamic models with improved representation of channel and floodplain inundation dynamics (Cauduro et al. 2013; Getirana et al. 2012; Neal et al. 2015; Neal et al. 2012; Sampson et al. 2015; Yamazaki et al. 2011). Although these models can potentially provide information about the distribution of surface water across extended areas and periods of time, they require complex parameterization, are computationally intensive and depend on the accuracy of DEMs with global coverage. As an alternative, Earth observation (EO) data combined with statistical modeling techniques represents a promising and cost-effective approach for systematic observation of surface water (Alsdorf et al., 2007; Overton, 2005). New

satellite, airborne and ground-based remote sensing data with high spatial, temporal and radiometric resolution are growing in size and variety at exceptional rates (Nativi et al., 2015).

Updated paragraph (11878, line 18-25) -additions underlined:
The empirical inundation models developed in this study provide unique insights into the inundation and retention behavior of surface water bodies and local driver combinations and can provide a valuable tool for improving water resource management in the MDB. Our modeling approach provides a data-driven and highly-automated alternative for analyzing the dynamics of surface water compared to the more complex framework of continental or global scale hydrodynamic models. Future work will focus on applying the presented methodology to the entire MDB, providing a first of its kind basin-wide and seasonally continuous inundation model.

References:

Cauduro, R. et al., 2013. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resource Research*, 49, pp.1226–1243.

Getirana, A.C.V. et al., 2012. The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon Basin. *Journal of Hydrometeorology*, pp.1641–1665.

Neal, J., Schumann, G. & Bates, P., 2012. A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. , 48(November), pp.1–16.

Neal, J.C. et al., 2015. Efficient incorporation of channel cross-section geometry uncertainty into regional and global scale flood inundation models. *JOURNAL OF HYDROLOGY*, 529, pp.169–183.

Sampson, C.C. et al., 2015. A high-resolution global flood hazard model. *Water Resources Research*, 51, pp.7358–7381.

Yamazaki, D. et al., 2011. A physically based description of floodplain inundation dynamics in a global river routing model. , 47(January), pp.1–21.