We thank the reviewer for insightful and constructive comments and provide answers to the comments below. The reviewer comments are in normal font, our replies are in bold font.

This manuscript addressed operational forecasting in the catchment located in Africa. The topic and contents of this study will be of interest to broad ranges of the scientific and engineering community especially because they achieved improved forecasting via combination of a well-known rainfall-runoff model, SWAT, and a basic data assimilation technique, Kalman filtering on a linear routing scheme. However, no innovation is found in their methodology on data assimilation and rainfall-runoff modeling compared to their previous publications in HESS (Michailovsky and Bauer-Gottwein, 2014) and WRR (Michailovsky et al., 2013). Although their focus seems to be on "operational" applications in "poorly" gauged basins, new methodology or finding is limited for these two targets. Therefore, major revision should be required for possible publication in HESS considering following comments:

It is correct that the hydrologic-hydrodynamic modeling and DA approach is the same as in the papers mentioned above. This is also stated in the manuscript (pages 11077, 11081). The new points in this paper are
- **Combination with real precipitation forecasts from a global weather forecasting system.**
- **Operational application of the forecasting system in real time. We are issuing daily forecasts as we are writing this.**
- **Extensive evaluation of the skill of the forecasts compared to persistence and (in the revised manuscript) climatology.**

Moreover, as part of TIGER-NET, the entire system has been implemented in an open-source GIS environment (OSGeo4W, Python). Installation and source code are available for download from the TIGER-NET webpage.

1. Methodology for operational forecasting
Even though values of this manuscript could be found in terms of "engineering", they didn't present any advanced method required for operational setting. Without innovation for these main keywords, this study could be mere applications using modeling techniques, developed by themselves in the previous publications, and operational input forcing. Therefore, I strongly suggest authors would provide additional methodology or analysis on operational applications for readers to have more confidence and understanding on their approach. For example, rainfall forecasts could be analyzed for varying lead times. If significant bias exists in operational forecasts, authors should re-think additional treatment such as bias correction or pre-processing, which shouldn't be remained as future study in such a case.

This basin is un-gauged in terms of precipitation. We do not have access to long-term in-situ precipitation records from any place in the basin. The "true" precipitation is therefore unknown. In order to address this issue, we compared NOAA-GFS with FEWS-RFE and we actually do implement a static bias correction (see page 11085). We will add an analysis of rainfall forecasts for varying lead times. However, since the "truth" is unknown, we can only compare the forecasts and evaluate them based on the quality of the discharge forecasts but cannot benchmark them against the truth. We believe that dynamic bias correction (e.g. based on near real-time satellite soil moisture or precip estimates) is beyond the scope of this revision, but definitely interesting for the future.

2. Data assimilation
2.1 It is interesting that the study utilizes the AR1-type model and runoff correlation matrix in the model noise specification. However, the impacts of the noise specification were not clearly verified in the manuscript. Please clarify how the error specification would affect the performance of discharge forecasts with additional evidence and analysis.

We will add forecasting scenarios with modified assumptions about temporal and spatial correlations of the runoff error for comparison. Because we do not have direct observations of runoff, we do not know the truth, and can only compare the suitability of the assumptions based on the resulting performance of the forecasts.

2.2 Although authors used several probabilistic measures such as coverage, sharpness, ISS, and CRPS, it was hard to find analysis on appropriateness of probabilistic forecasts. As DA is expected to reduce uncertainty range, only comparison between DA and open-loop could not justify appropriateness of probabilistic forecasts. It is recommended that authors should add evaluation and analysis of probabilistic forecasts based on their uncertainty assumption on observation and simulation (10% of standard deviation of discharge uncertainty seems to be underestimation especially in poorly gauged basins as well as in high-flow seasons). An additional measure such as predictive Q-Q plot could be useful to assess appropriateness of the probabilistic forecast.

We will address this comment following the suggestion by reviewer 1 and extend the comparison of our probabilistic forecasts to the persistence reference. We will also add a comparison with climatology. From these two comparisons we will be able to show how good our forecasts are compared to the two alternative forecasts that can be had without any effort, i.e. "last available observation" and "typical flow for this day of the year". We will also add a forecasting scenario with a higher assumed error on the in-situ observation and add q-q plots for forecasted flows.

2.3 Description on DA procedure is not enough for readers to understand and reproduce this study. Authors should revise the manuscript with additional description and equations about how noise specification would be applied and updated in Kalman filtering equations.

As the reviewer states above, the modeling and DA methodology is equivalent to what has been published in earlier papers. We therefore kept this part very short and referenced the existing papers.

2.4 Please clarify how the error model is applied in DA if observation is missing. It is not clear how confidence interval is estimated in open-loop simulations, either.

If no observations are available, the Kalman filter update equations are simply not applied, i.e. the model is propagated without updating. The AR1 error model for the runoff will still produce a model spread (confidence interval). In the open-loop simulations, there are no observations available, i.e. the model is run without any updating whatsoever.

3. Methodology for poorly gauged basins There is no doubt that the study area is a poorly gauged basin. But, there is no technical treatment or analysis for poorly gauged basins in the manuscript, although authors might think public-domain input forcing and models are solutions. Therefore, the current title could confuse readers who are finding new methodology for data-sparse regions. I wonder if it is proper to use the term "poorly gauged basins" in the title with the present content.

We are unsure how to address this comment. In principle, this system can run with no or minimal feed of in-situ data, which is a typical situation in many African river basins (the target region of TIGER-NET). For many of these basins (Zambezi, Chari-Logone, Kavango etc.) there presently do not exist any operational forecasting systems and the presented methodology can fill this gap and provide a first solution. Could the reviewer suggest alternative wording for the title which would help us understand the nature of the concern?

4. Revision of abstract
4.1 As an anonymous reviewer addressed, description on funding body in the abstract is not desirable especially because this manuscript covers limited parts of this project. Referencing the

website in the middle of the manuscript might be enough to show relationships of this study with the entire project.

This issue has been brought up by reviewer 1, see our answer there. Because this has now been identified as a problem by two reviewers, we will remove the reference from the abstract.

4.2 I couldn't find any evidence supporting the sentence "the value of the forecasts is greatest for intermediate lead times between 4 and 7 days". In the contrary, the accuracy of forecasts seem to degenerate gradually according to increasing lead times.

The support of this statement is table 4 and the comparison of our forecasts with the persistence reference. For 6 and 7 days, our forecasts perform better than the persistence reference, while for shorter lead times, their performance (according to the persistence index) is worse than persistence. Following the comments by reviewer 1, we have compared CRPS of persistence with CRPS of our forecasting runs. Preliminary results of this comparison show our forecasts outperforming persistence for lead times between 3 and 7 days. This information will be added to the revised manuscript.

5. Terminology
The authors used the term "hydrodynamic model" to indicate the Muskingum routing scheme, which might be different from general usage. As my limited knowledge, the hydrodynamic models usually refer to simulation models to represent the motion of the flow by momentum and continuity equations. I don't think the Muskingum routing scheme belongs to a range of hydrodynamic models. Instead, river routing scheme, as the authors used in their previous publications, would be a better term to indicate the Muskingum scheme through the manuscript.

The term "hydrologic-hydrodynamic modeling" is used in parts of the literature to refer to models addressing both the land phase and the channel flow phase of the runoff process. The Muskingum routing scheme solves a strongly simplified version of the full 1D Saint Venant equations for channel flow. We will adapt the terminology to avoid any potential misunderstanding.

6. Formulation of objective function
The authors included NSE and ME (Eq. (2)) in the objective function. However, ME varies in wider ranges compared to NSE. Please justify how two measures could be used having similar influence.

We realize that this has not been accurately reported in the paper: ME in Equation 2 is the mean error expressed as a fraction of the mean observed flow, i.e. the normalized relative error. The term 1-NSE and the normalized relative error have similar magnitude and thus a similar influence in the SCE objective function. This information will be added to the revised version of the paper.

7. Figure 7
It would help readers to understand discharge forecast more clearly to add hyetographs of (catchment-averaged) forecast input forcing for each lead time in Fig. 7.

We will consider this and add this information to the revised version of the paper (see also answer to your comment 1 above)