In the following, we present our responses to the referees.
Referees' comments are in blue.

Referee #2 – Report #2

Specific comment 1: Abstract

The authors addressed my general comments with:

Stochastic models in hydrology are very useful and widespread tools for making reliable probabilistic predictions. However, such models can be used for making predictions only if model parameters are first of all calibrated to measured data in a consistent framework such as the Bayesian one. Unfortunately, Bayesian parameter calibration, a.k.a. inference, with stochastic models, turns out to be an often computationally intractable problem. Therefore, the computational obstacle is often overcome by employing over-simplified error models, which lead to biased parameter estimates and unreliable predictions. Our goal in this work is to present a HMC algorithm that makes Bayesian parameter inference with stochastic models possible, from which hydrology can potentially take great advantages: a sound calibration of model parameters is essential for making robust probabilistic predictions, which can certainly be useful in planning and policy making. Discussing specific hydrological models or systems is outside the scope of our present work, and will be the focus of further studies.

This is a clearer, simpler version of the abstract and I suggest the authors use this instead. Something like:

Stochastic models in hydrology are very useful and widespread tools for making reliable probabilistic predictions. However, such models are only accurate at making predictions if model parameters are first of all calibrated to measured data in a consistent framework such as the Bayesian one. Unfortunately, Bayesian parameter calibration, a.k.a. inference, with stochastic models, is often a computationally intractable problem due to the expensive likelihood functions employed in traditional inference algorithms such as... . Therefore, the prohibitive computational cost is often overcome by employing over-simplified error models, which lead to biased parameter estimates and unreliable predictions. Our goal in this work is to present a very computationally efficient novel HMC algorithm which makes Bayesian parameter inference with stochastic models possible. We show that this approach is robust, by detailing a case study from urban hydrology which would normally be computationally prohibitive to calibrate. This work shows the potential of the approach; a sound calibration of model parameters is essential for making robust probabilistic predictions, which can certainly be useful in planning and policy making. Discussing specific hydrological models or systems is outside the scope of our present work and will be the focus of further studies.

We have replaced the original abstract with the version suggested by the referee.

Specific comment 2: Introduction

In the introduction authors have addressed previous comments which is good but now it is more muddied to read, and it was already a bit clunky and hard to follow. Suggest restructuring to make clearer to the reader what the actual goal of the paper is and why the inference algorithm is novel.

Suggested structure:

Cover concept of stochastic models in hydrology and their widespread use – what applications and why they are useful.

Cover concept of calibration aka inference in a Bayesian framework and why it is important for accuracy.

Introduce the concept of computationally prohibitive costs of traditional inference algorithms and what everyone else does – employ the use of over simplified error models and why this is bad.

Introduce novel inference HMC algorithm and why it is computationally efficient and accurate and for which applications it can be employed.

Introduce goal of the study – to prove the HMC is accurate, efficient and effective by employing in case study which is hard to model: urban hydrology and why it's hard (inaccurate and unrealiable rainfall) - could even introduce the concept of the "worst case scenario".

Then state implications – advantages in using this across a broad range of hydro applications.

We have restructured the introduction taking the referee's suggestion into account. We hope that this makes it clearer and easier to read.

Specific comment 2: Methods – still appears to be very long and not sure all details need to be in there.

In addressing my previous comments, the authors state:

In the present work, we apply an HMC method with a novel time-scale separation approach (from Albert et al., 2016) for the first time to a real-world hydrological case study, using real time-series of observed rainfall and outflows. Moreover, we also demonstrate for the first time the ability of the algorithm to reconstruct with great accuracy the unknown true average rainfall over the catchment using only prior knowledge and the observed outflow. The reconstructed precipitation is then used to calibrate the hydrological model parameters, which are thus protected against the degrading effect of the possible rainfall data inaccuracy.

Following on from this, it seems that the method has already been detailed in Albert et al., and thus only the additional information pertinent to the case study, e.g. priors, hydrological and rainfall models used etc., need to be detailed.

Perhaps there could be some movement of the HMC algorithm description to an appendix?

We have carefully considered this referee's comment, and we believe that only the information strictly pertinent to the specific case study under investigation is presented in detail in the manuscript (e.g., priors, models etc.). More generic non-specific aspects of the method already outlined in Albert et al, 2016, are not covered in this work. Therefore, we would be inclined to keep the description of the HMC algorithm in the main part of the manuscript. We hope to find an agreement with the referee on this point.

Specific comment 3: Conclusion – could be made stronger to convince the reader of the applicability of the HMC as an inference algorithm for stochastic hydrological models.

Authors note from addressing previous comments:

Our goal in this work is to present a HMC algorithm that makes Bayesian parameter inference with stochastic models possible, from which hydrology can potentially take great advantages: a sound calibration of model parameters is essential for making robust probabilistic predictions, which can certainly be useful in planning and policy making. Discussing specific hydrological models or systems is outside the scope of our present work, and will be the focus of further studies.

And:

In the present work, we apply an HMC method with a novel time-scale separation approach (from Albert et al., 2016) for the first time to a real-world hydrological case study, using real time-series of observed rainfall and outflows. Moreover, we also demonstrate for the first time the ability of the algorithm to reconstruct with great accuracy the unknown true average rainfall over the catchment using only prior knowledge and the observed outflow. The reconstructed precipitation is then used to calibrate the hydrological model parameters, which are thus protected against the degrading effect of the possible rainfall data inaccuracy. This considerably reduces the bias in the inferred parameters, thus leading to more realistic models and reliable runoff predictions.

Suggested structure:

Restate goal of paper and scope.

State novelty – first time to reconstruct...

State what was demonstrated and why biases in the inferred parameters were reduced.

State limitations of the study.

State future work.

State implications for hydrology and why broadscale use of the HMC is advantageous.

To include the above salient points.

We have tried to structure the conclusions following the referee’s suggestion. We hope that the current version meets the referee’s expectations.

Specific comment 4: Improve writing. Topics are introduced across two paragraphs or more (e.g. in the introduction). Sentences are too long and too complicated throughout.

Unfortunately, some of the key messages get lost in the unwieldy structure of the paragraphs and sentences. Suggest containing a topic per paragraph and to shorten sentences. Make sure key messages stand out – that they are simple, clear and succinct. The manuscript could be made more impactful by improving the writing.

We have revised the manuscript paying special attention to the style, especially in the introduction and conclusions, according to the referee’s suggestion. We hope that now the key messages stand out more clearly.

Referee #3 – Report #1

The authors have responded to all the questions and requests I made in the first round of reviews. Therefore, I recommend publication of this very interesting and well-written paper.

At the authors' discretion, I suggest that the paper could be further improved by addressing the following points:

- What is the mean value of the normal distributions used for drawing the momenta, eq. (23)? Is it unity?

The mean value is zero. We have added this detail to the text.

- If the transformation (17)-(19) was non-linear, would the authors have had to set those equations forward in the problem statement, including eq. (11)-(13), i.e., the Jacobian?

Yes, that's correct. Non-linearities in the transformations (17)-(19) would have to be considered in the calculations throughout the manuscript (e.g., Jacobians). However, since this is not essential to understand the key points of our work, we prefer not to discuss it.

- More than resorting to Meta-Dynamics, which is difficult to perform in high-dimensional spaces, shouldn't the author be more interested in parallel tempering? Wouldn't this also help in assessing what the effect of the specific values of the masses would be?

This is an interesting point. We added references to Parallel Tempering in the text.

- I would be curious to know why 6 evaluations of the derivatives of the posterior are required in the molecular dynamics part. Is the derivation of the energy six times more expensive than its evaluation because these are 6 evaluations, or is it actually 6*6=36 times more expensive?

There are 6 evaluations of the derivatives at each call of the algorithm because we have set the number of integration steps P=3, and each integration step requires 2 calculations of the derivatives (as explained in Albert et al., 2016).

The derivation of the energy is 6x6=36 times more expensive than its plain evaluation.

We have added comments to the text to clarify these interesting points.