

Dear editor,

As you requested, we have again sharpened the way we present and test our hypothesis and we have worked on the clarity of the presentation and solidity of the conclusions. While going through the paper in detail many ambiguities in terminology (as requested in the first review) have been removed and the structure has been revised.

Apparently we were not able to convince the 2nd round reviewer of the potential innovation of this new proxy for directly monitoring the soil moisture in the root zone of vegetation. The critical review has motivated us to do an even better job. But we decided not to change the paper in a way that addresses all the issues raised by the students in the first round. That would have changed the paper into a completely different one, made it unnecessary long and diverted the attention from the main question: "Is there potentially a new, more accurate and more direct way of monitoring the moisture content of the root zone, by using the already existing indicator NDII". If that appears to be a promising way, then that could be a real breakthrough in hydrological modelling, both in gauged and ungauged basins. We fear that the reviewer of the second round completely missed that point, although it was clearly mentioned in the revised paper. "In this paper we developed a novel approach to estimate the soil moisture storage deficit in the root zone of vegetation, by using the remotely sensed NDII in the Upper Ping River Basin in Northern Thailand"

We are very sorry that the reviewer did not appreciate the scientific significance of our paper. We disagree that the scientific significance of the paper is poor. In hydrology we should embrace ideas that provide a different perspective on the way we have been monitoring and describing hydrological processes. Sometimes such ideas open new ways to deal with a difficult issue by a simpler, more intuitive and more logical approach. This paper raised such an idea. Identifying an indicator that can monitor root zone moisture directly from space could become a breakthrough in hydrological modelling, both in gauged and ungauged basins. The scientific community developed the NDII as an indicator for leaf, or canopy, water content. It has been shown in numerous publications that this is indeed what the NDII does. It is also well documented in the literature that moisture content of leaves is connected to the suction pressure in the root zone. We give references for that. The next step to link the suction pressure in the root zone to the amount of moisture in the root zone is not only intuitive, it is also widely done in the literature. So the hypothesis is quite plausible. What we aimed to do is to test the hypothesis against independent hydrological modelling. There is no need to then request that more complex or more detailed or better validated models should be used. That is completely beside the point. If a simple model already indicates that the relationship is present, then that is even a more useful finding (allowing application in poorly gauged basins) than if a very detailed and sophisticated model were used. Moreover this is not the end of this research, it is merely the beginning of a completely new venue.

Many authors have been busy trying to relate the NDII to the total water content (TWC) of vegetation, because this is what the community wanted to know in order to subtract this signal from satellite derived signals that pretend to penetrate the ground to monitor soil moisture. As we all know, this has been rather unsuccessful, which led to the NDII being used merely as an indicator for drought.

The innovation here is that this ineffective and complicated detour (to filter out the water content of vegetation so as to be able to see the top layer moisture in the ground) may no longer be necessary. If the NDII indeed measures leaf water content, then through the hydraulic connection of the plant, it is directly connected to the suction pressure in the root zone. Hence, it reflects the moisture deficit in the root zone directly. There is no need to provide experimental proof that leaf water tension is related to moisture deficit in the root zone, it follows directly from Newtonian mechanics, and it has been published widely (e.g. [Rutter and Sands, 1958](#)). So this paper does the hydrological community a favour by testing if the NDII can be used as a direct proxy for what in conceptual models is represented by the moisture storage in the root zone.

Now the reviewers have insisted that a scientific paper always has a well-formulated hypothesis and that the authors should try by experiments and preferably measurements to test this hypothesis. This may be true, but is also a formality. Far more important is the communication of the possibility that we can circumvent the complex detour of filtering out the effect of vegetation on soil moisture sensing (in order to look into the ground to see how much moisture there is in the top layer and connect this to the moisture in the root zone), by observing root zone moisture deficit directly with an already existing product. What more is needed than to show that this is a promising way to go? Whether we demonstrate this with a lumped or semi-distributed conceptual model, or with a 'physically based distributed' model is irrelevant. It is also irrelevant if we validated the model before making the test. It also does not matter whether we demonstrate this in Thailand, in Europe or elsewhere. The point is that if it is a promising new proxy, then this opens opportunities for hydrologists to test the relationship between NDII and root zone storage with their own models and in their own catchments.

The main question is: Is the NDII a possibly promising new indicator that we can use to monitor a crucial hydrological variable directly, in a very simple and straightforward manner? The fact that we make this plausible is already a potential breakthrough.

We do admit that the first version of the paper may not have brought this issue to the fore clearly enough, although both student reviewer clearly understood and reflected the objective of the paper. We took the criticism of the reviewers at heart and revised the paper significantly addressing the major and minor issues raised by the student reviewers. In the process the paper got much better and we are grateful for that. By looking at the track changes of the first revision, one can see that the paper was substantially modified. But we did not do everything that the students asked. We did not consider it important to provide the traditional, but contested (see [Gharari et al. 2013](#)), approach of doing a split sample calibration and validation, simply because it is not the purpose of the paper to show how well the FLEX model is capable to mimic the runoff. There are several papers that already do that (e.g. [Fenizia et al., 2014, 2016](#)). We could have used any model. Rather this paper wants to show that the NDII may become a valuable proxy to monitor a key state variable and possibly to calibrate model parameters or develop better models.

The reviewer of the second round did not appreciate all the changes made, but rather revisited the first round of reviews and checked if all comments raised by the students were addressed. We don't think this should be the objective of an additional reviewer.

Just a note on the commendable effort of the MSc students to write a thorough critique on our paper. They really did a good job at it, and they should deserve a good mark for the review. But their comments are not a peer review. Rather the intention was to demonstrate their supervisor that they can write a well-founded critical review of a paper. We appreciated this and remarked on it in our reply. We also thankfully took most of the suggestions into account and tried to address them as much as possible without losing the focus of the paper. But we did not do everything that was suggested, purely because that would divert the attention from the main and most important issue. We have identified a promising, simple and logical way around the rather unsuccessful attempts to try to monitor soil moisture from space by looking through the vegetation. We don't prove the hypothesis, but we show it as an interesting venue and most likely a valuable new proxy. Unfortunately, in catchment hydrology, "proof" is virtually impossible. Identifying a hypothesis as 'very likely' is the best we can do in an uncertain and heterogeneous environment. So we find the grounds on which the reviewer rates the significance and quality of our paper as poor unfair. The quality may not be high, but the significance of having a new proxy for root zone moisture is potentially very high.

This second revision has been again major. By track changes you can quickly see the substantial changes made both in the text and the structure, and the way in which we have addressed the comments of the 2nd round reviewer. We also revisited the comments of the first round of student reviewers and made a couple of additional changes in response to their suggestions. We really tried to address everything, without losing focus and without the paper becoming unnecessarily long. We hope that this revision will be received more positively.

- Gharari, S., M. Hrachowitz, F. Fenicia, and H. H. G. Savenije, 2013. Moving beyond traditional model calibration or how to better identify realistic model parameters: sub-period calibration. *Hydrol. Earth Syst. Sci.*, 17, 149–161.
- Fenicia F., D. Kavetski, H.H.G. Savenije, M.P. Clark, G. Schoups, L. Pfister, J. Freer. 2014. Catchment properties, function, and conceptual model representation: is there a correspondence? *Hydrological Processes* 28, 2451–2467, doi: 10.1002/hyp.9726.
- Fenicia, F., D. Kavetski, H. H. G. Savenije, and L. Pfister, 2016. From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, *Water Resour. Res.*, 52, 954–989, doi:10.1002/2015WR017398.

Reply to reviewer

### General comments

I was not involved in the previous round of reviews of the manuscript, but I 'm impressed by the very thorough, insightful and constructive reviews provided by students of the MSc programme at Wageningen University. On the other hand, unfortunately and surprisingly, I do not find the response and revisions made by the authors as adequate and fully considering major points raised in the review. I like very much the idea and effort to find some proxies which can improve understanding of runoff generation processes and their modelling (particularly in the context of runoff predictions in data sparse regions). The response to reviews, however, is in many places incomplete or ignoring the suggestions for revision and often refers to some future studies. This I do not find balanced and appropriate. After my own reading of the revised manuscript I have really the opinion that in current form it still does not reach the level needed for a scientific paper in the prestigious HESS journal. I fully agree with most of the major concerns raised in the previous reviews. In the following part I will try to repeat and highlight the points which I find very important and where I did not find the response complete or adequate. Please consider these points very seriously in the revision. In the brackets I will recall the points raised in the review of the students.

1) (2.1 Use of model vs. observations): I fully agree with the reviews that **some kind of validation is needed** to justify the interpretations made. This point is strongly linked with the need to clearly formulate the hypothesis made, which needs to be tested and validated. Stating that the validation of the research question/hypothesis is not important is rather unusual for a scientific paper.

Our reply:

We fear the referee misunderstood the objective of the paper. This paper is not about calibration and validation of a simple HBV-like model. That is completely beside the point. If this was a modelling study, then the reviewer maybe would have a point, although the merit of the traditional calibration/validation approach has been criticized (e.g. Gharari et al., 2013). The connection between leaf water content and root zone moisture content is hardly contestable. This has been demonstrated in many studies, which we refer to. The point is: "Can the NDII become a useful proxy for directly monitoring root zone moisture content and be used to constrain hydrological models". And preferably simple hydrological models, so that the methodology can be used in poorly gauged or ungauged basins.

In the response the authors write "**Our method assumes that the NDII sees the moisture in the leaves of the vegetation, which is subject to the suction pressure of moisture in the root zone.**" But this is very difficult to see from the results which represent spatially and temporally lumped data.

Our reply:

This connection is not contested at all. There is sufficient literature that demonstrates that (e.g. [Rutter and Sands, 1958](#)).

**The statement “...we do not think root zone soil moisture is not well presented in the model” is very difficult to objectively accept without some validation.**

Our reply:

We do not understand this observation. In conceptual models the root zone storage reservoir is the part that represents root zone soil moisture. This is not something that we need to prove. What this paper wants to do is show that it is plausible to constrain this key parameter by a simple, and already widely existing RS product.

The suggestion to use a distributed model instead of a lumped model is very valid point. A typical application/potential of remote sensing products is to describe the spatial patterns (rather the absolute values). **Please consider to formulate the research hypothesis and present the results from a distributed modelling framework in this manuscript (not in a follow-up study). This will really add more credence to the results and interpretations.**

Our reply: Not only is this impossible because the data for such a model is not available, it is also not desirable. Testing the hypothesis with a lumped model is an even heavier test, than doing this with a distributed model. Moreover for applicability in poorly gauged basins, the use of distributed models is rather useless. Showing that the methodology may provide a constraint on lumped conceptual models, is a more valuable finding than if it would work for a distributed model.

The response to the point of **interpretation of no time lag** between NDII and Su is not very comprehensive. Some more reasoning would be very useful here. What are the effects of lumping over the 8-days and the entire basin on such interpretations? Please consider to be more precise with the formulation here.

Our reply: In the supplementary material we have shown the effect of including a time lag. A time lag reduced the correlation, which is a good indication that the relation is direct. Of course it would be nice if we could have used daily data of NDII, but this information is just not available. The 8-days averaging is necessary because the NDII product is available at this time scale, one reason being that it requires the use of cloudless pixels.

2) (2.3 Calibration, validation and correlation): I agree with the point that it is a pity that the **manuscript does not demonstrate the value of the relationship found for improving predictions in ungauged sites**. It will be very interesting to more clearly demonstrate (or **at least discuss**), how the scattered and less scattered relationship between NDII and Su can be used in PUB.

Our reply: This is a good point. We agree. Although in the first revision of the paper we emphasized the usefulness for prediction in ungauged basins (which the reviewer apparently missed) we have emphasized this further in the abstract, conclusion and body of the paper.

Moreover the comment on the **similarity between fitted exponential relationships** for wet and dry seasons **is not elaborated in the response and revision**.

Our reply: The exponential relationships are not intended to provide equations that should be used. They are merely illustrations of fit. Of course in a particular catchment the coefficients should be the same, although they may differ in different sub-catchments. We have emphasized both in the text and in the caption that the coefficients are purely meant for illustration and to help comparison.

I do not understand why there are separate fits for wet and dry season when these are at the end very similar.

Our reply: True, we agree. The same equations could apply, but the curves are purely used as a reference for the fit.

What do they demonstrate? **Where are the negative NDII values shown in the maps?** How to interpret these results when it seems that the spatial variability of NDII within the basins is larger than the temporal one? **Is the lumping useful here?** Maybe I missed some point here, but this part and the reasoning behind the dry and wet fits is completely not clear to me. **What is the take home message here?**

Our reply: It is to demonstrate that the poor correlation only occurs when the root zones are saturated.

3) (2.4 Lack of detail). The **comment on missing clearly formulated research hypothesis is not responded, but very important.** This is an essential part of scientific experiments.

Our reply: we are not sure if the referee looked at the revised version. We clearly followed up on the suggestion of the student reviewers in the first revision. In this 2nd revision we have even done it more rigorously, and added a paragraph in the discussion about the validity of the hypothesis (Section 5.2)

Also more thorough justification of the **choice of the model**, its reliability to simulated root zone soil moisture is not comprehensively discussed.

Our reply: The model has been extensively tested (Fenicia et al., 2014 and 2016). We don't think discussing this in a comprehensive way adds to the readability of the paper.

The comment on **differentiating of the type of vegetation** is also valid and relates to the evaluation of the spatial patterns (rather than lumped simulations) mentioned above.

Our reply: We agree that this would be interesting follow-up. We intend to do this using the landscape-based model. But again, more detailed models are beside the point of this paper.

4) (2.5 structure and outline): I fully agree with the comment that the **discussion is not very comprehensive** as it does not relate/link the results to previous studies. Only highlighting the problems and explaining reasons why some actions are not possible is not very balanced and does not clearly demonstrate the contribution of the paper.

Our reply: We have extended the discussion section substantially.

I also agree with the confusions coming from the **different definitions of root soil moisture in the paper**, which is not revised nor commented in the response.

Our reply: In the first revision we have followed this up, but we have in this second revision again revised the terminology and made it completely consistent. Summing up, **I found the comparison between lumped soil moisture simulations and lumped NDII index not enough to quantitatively support the interpretations and reasoning made**. Please consider to **more precisely formulate the scientific hypothesis and its validation**.

Our reply: We have reformulated and revised the hypothesis, including the title of the paper. We have also written a section on the validity of the hypothesis

I would suggest to significantly revise/extend the paper by using a distributed model and pattern evaluation or practical demonstration the value of the NDII for prediction in ungauged sites with some kind of validation.

Our reply: As said before, a distributed model is at this stage not desirable, particularly because that would inhibit the potential use in ungauged basins. The applicability in ungauged basins has been further emphasized.

Without a very serious consideration of previous reviewers comments, I would not suggest the manuscript for publication (in the current form).

Our reply: We have again revisited all the comments made. The paper has been modified and improved substantially. We are grateful for all the valuable suggestions, which have improved the paper. We would like to thank the reviewers of the first and second round of the review for there thorough and elaborate reviews.

#### References:

- Gharari, S., M. Hrachowitz, F. Fenicia, and H. H. G. Savenije, 2013. Moving beyond traditional model calibration or how to better identify realistic model parameters: sub-period calibration. *Hydrol. Earth Syst. Sci.*, 17, 149–161.
- Fenicia F., D. Kavetski, H.H.G. Savenije, M.P. Clark, G. Schoups, L. Pfister, J. Freer. 2014. Catchment properties, function, and conceptual model representation: is there a correspondence? *Hydrological Processes* 28, 2451–2467, doi: 10.1002/hyp.9726.
- Fenicia, F., D. Kavetski, H. H. G. Savenije, and L. Pfister, 2016. From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, *Water Resour. Res.*, 52, 954–989, doi:10.1002/2015WR017398.