Replies to author comments AC2 and AC3 by Chaopeng Shen for manuscript

'When ancient numerical demons meet physics-informed machine learning: adjoint-based gradients for implicit differentiable modeling'

By Y. Song et al.

Dear Chaopeng Shen, dear Authors, dear Editor,

Thanks for your detailed replies to my review. This is exactly what the HESS open discussion forum is made for. First of all, I would like to state once more that I think the authors, by introducing in their manuscript a method for integrating implicit solvers in modern ML-based model training via backpropagation, provide valuable research that is absolutely worth publishing. The reason I was (and am) recommending "reject with strong encouragement for resubmission" is that I suggest changes to the manuscript that will most likely require more time than usually assigned for major revisions. I will leave it to the Editor whether he thinks i) the changes I suggest are valid, and if yes, ii) how much time they would require.

Reading the replies (AC2, text and .pdf, and AC3) by Chaopeng Shen, four main points of discussion arise. I reply to them here in summarized form, rather than individually in each document:

- **The main purpose of the paper is to enable implicit schemes**. I agree with this statement by Chaopeng Shen, and I suggest focusing on this message. Therefore, I would support a revised paper, as a technical note, that introduces the method, not more and not less. In such a paper, there is no need for an in-depth discussion about if and when explicit schemes for conceptual hydrological models operating on daily data create problems, and no need for comparing implicit schemes vs. explicit schemes operated on higher-resolution data comparison.

- **Running small time steps (less than a day) with automatic differentiation creates problems (memory use, allowable window size)**. Chaopeng Shen suggests it is a bit unfair by me to ask the authors to demonstrate that implicit schemes solve a problem of explicit schemes. I see two options here. The first is to write a short technical note presenting only the main method innovation, see previous bullet point. The second is to keep it as a research paper, showing the method innovation and applications. In that case, as a reader I would expect a demonstration that the innovation solves an important problem present in the applications used in the paper (hydrological modeling on daily basis using conceptual models). That is, showing that the currently used explicit schemes introduce substantial numerical error. This does not need to be for the full set of catchments used, but could be done for a few representative catchments, and along the lines of the demonstrations in Clark and Kavetski (2010). If the editor thinks this is unnecessary detail, I would at least expect a more in-depth discussion about how the findings of Clark and Kavetski (2010) and Kavetski and Clark (2010) apply to the application in the manuscript.

- **Adaptive time stepping is difficult to realize in connection with AD, more specific in connection with parallel processing of minibatch optimization.** I never tested, but Chaopeng Chens explanation of this point makes sense to me. In my review, I never asked the authors to include such tests in the manuscript, therefore there is no disagreement here.

- **Discussion of HBV structural/functional changes (capillary rise) in the same paper**. In my review, I was mentioning that discussing structural/functional changes to the HBV model to solve an apparent model deficiency has little to do with the key message of the manuscript, and therefore suggested removing it. I still think that leaving this part away will help the paper to better convey its message. The reply by Chaopeng Shen - "Many article carry more than one stories and this is a beneficial (although not that major) improvements to the model. We do not want to write another article for this change." - has not convinced me otherwise. I will leave this decision to the editor.

Yours sincerely, Uwe Ehret