Hydrology and
Earth System
Sciences
Discussions

Interactive comment on "Evaluation of Random Forest for short-term daily streamflow forecast in rainfall and snowmelt driven watersheds" *by* Leo T. Pham et al.

Francesco Avanzi (Referee)

francesco.avanzi@cimafoundation.org

Received and published: 1 August 2020

Pham et al. developed a Random-Forest-based (RF) algorithm for day-ahead stream-flow forecasting. The algorithm employs 8 weather-snow-time features (see Table 2) and was tested across 86 watersheds in the Pacific Northwest (PNW), where it was compared with multilinear regression and previous-day streamflow as a minimum-information forecasting approach (so called Naïve method). Results show that RFs provide quite robust predictions across catchments with different climatology (rainfall dominated, snowfall dominated, or transient) and generally perform better than the two benchmark approaches – especially in rainfall-dominated and transient watersheds.

Drops in accuracy for RFs were correlated with watershed slope and sandiness.

This is an interesting, well-written, and concise paper about an emerging machine-learning technique in hydrology and its use for streamflow forecasting. While I have tangential knowledge of RF technical issues, I found the description of the algorithm clear and rigorous, which facilitates replicability and ultimately allows readers to learn more about RFs in general rather than only looking at is as a black box (note that these technical details are often bypassed or heavily summarized in other papers I have read on this matter). Also, RFs and machine-learning approaches in general are on the rise in hydrology, meaning I expect the paper to have some impact on the community. There are still some major and minor comments that I recommend for authors (see below), and I recommend the editor reconsider this manuscript after minor revisions.

MAJOR COMMENTS

1. The manuscript sometimes reads like a technical note, as it describes the algorithm and its implementation in great details but ultimately falls a little short on hydrological-process interpretation. I see that the main goal of the paper is testing an algorithm, and applied research is certainly within the scope of HESS. And yet, I feel like implementing RFs across 86 watersheds with different characteristics and 10 yrs with different climatology without looking more specifically at how performance changes across the landscape and between years with different characteristics is kind of a missed opportunity. For example, operational forecasters in the western US often calibrate multiple forecasting tools based on the concept of "water-year type" (for example, a set of parameters for dry yrs, another for wet yrs, etc.). Doing so here would allow authors to explore how predictive skills change between dry and wet yrs, or yrs with more or less snow or more or less rainfall, which would also have some future-climate implications. I ignore how demanding it would be to add an additional calibration experiment like this, so these were only examples. My bottom line is that I encourage authors to better explore their results from a hydrological-process perspective to add some process-based insights to an already interesting paper focused on an algorithm.

2. I was a little surprised by the choice of benchmark models and particularly by the fact that authors did not consider a full hydrologic model. I understand that authors would probably like to stay within the realm of data-driven models, but a Naïve approach looks very simplistic, especially at a daily time scale and in basins where rainfall and snowfall coexist. How can this approach predict, e.g., intense rain-on-snow events that are ubiquitous in the PNW? Most flood-forecasting tools I have been exposed to use full hydrologic models, and I encourage authors to at least discuss this matter in their manuscript.

3. Relatedly, I was also a little surprised that authors did not consider rainfall and snowfall as separate features in their model (see their Table 2). I am aware that PRISM only provides total precip, but it also provides temperature and relative humidity that could be employed to separate snowfall from rainfall. Perhaps considering SWE already makes up for this, but I encourage authors consider this at least for future work. This was again particularly puzzling to me given the well-known role of rain on snow in this region.

SPECIFIC COMMENTS

- Title: I have usually seen "rainfall-dominated" and "snowfall-dominated" being used, rather than rainfall and snowmelt driven. Consider revising.

- Line 3: I think "ease of application" might be relative, especially in ungauged areas or for users with limited computational capabilities. Consider revising or expanding.

- Line 8: transisent -> transient?

- Line 11: better than what?

- Line 24: I believe even ML algorithms need the formulation of some mathematical equations, although maybe not in a predictive role.

- Line 25: I am not sure if ML algorithms really require "fewer data" than, e.g., conceptual, minimal hydrologic models. Here again, a comparison with a hydrologic model

C3

would be a great addition to the paper.

- Line 38ff: maybe also mention glaciers here, although they might not be an important driver for hydrology in your study region.

- Line 56: indeed, statistical forecasting models are widely used across the western US to predict summer flow (e.g., April to July total runoff). I understand this is out of the scope of your paper, but maybe mention this application to provide broader framing to your work.

- Algorithm 1 Step 3: I believe the case $X_i = t$ is missing, contrary to Figure 1.

- Section 2.2: maybe redefine acronyms for MDA and MDI here since you introduce them in the Introduction. This will be greatly appreciated by diagonal readers.

- Section 2.3: see my major comment 2.

- Line 149: Knoben at al. (https://hess.copernicus.org/articles/23/4323/2019/) have recently pointed out that KGE = 0 has a different implication from NSE = 0, and so KGE = 0 should be used with caution. Please revise as relevant.

- Line 172 and Table 2: please use SI rather than customary units.

- Line 176: maybe some more quantitative climatology would be more appropriate here. For instance, replace "ample amount of winter precipitation" with statistics of winter precip for your watersheds. Same for "mild temperature". It would also be interesting to provide some statistics of mean-max SWE across the basins.

- Line 186: have you tried to impute missing values? What's the impact of gaps in your framework?

- Line 196ff: how "place-based" is this classification based on the day of the water year? I would have expected one to classify basins based on proportion of rainfall over total precipitation, which look more general to me.

C4

- Section 3.2.3: it has been reported that snow pillows have a certain bias in capturing the onset of snowmelt, basically because they isolate the overlying snowpack from the ground. I understand they are the only continuous-time data source to estimate snowmelt, but consider adding a warning about this bias here.

- Line 244: I believe at least some SNOTEL stations do measure soil moisture. Please specify this if relevant.

- Line 246: how were the validation and calibration period chosen? May this choice have played a role in your results? What were the climatological characteristics of these two periods? Please expand and support your choice here.

- Line 269: I might have missed this, but do you show any statistics of persistence for your catchments to support this statement? Again, I may be missing something here.

- Line 292: see my previous point regarding KGE

- Line 307: may these be due to rain on snow?

- 355: streamflow (typo)

- Figure 9: consider adding the scatter plot for slope

- Table 1: is there any reason why snowmelt-driven catchments have a larger range of drainage areas? Just out of my curiosity.

-Table 2: please add t or t -1 as relevant in the "predictors" column; you only did that for predictor 1.

- Table 5: what is the data source for these characteristics? Especially sandiness and forested area.