

We thank the reviewer for their detailed review, and for their comments which will improve the clarity of the manuscript. Reviewer comments are reproduced below in grey indented text and responses follow in black.

This Technical Note (TN) uses available streamflow simulations from several datasets to show that high NSE and KGE values for seasonal catchments do not necessarily translate in good model performances of interannual variability of streamflow. This is a relevant topic in the scope of HESS. I agree with all points raised in RC1 and provide a few comments below.

TITLE – the title could be more straightforward regarding highly seasonal streamflow regimes rather than specifying tropical, alpine, and polar catchments.

We will revise to title to: Technical Note: High Nash Sutcliffe Efficiencies conceal poor simulations of interannual variance in seasonal regimes

ABSTRACT – the short summary reads better than the abstract. The introduction of the abstract is too long. There should be only one opening sentence (e.g., “...common metrics used to evaluate hydrological models...”) followed by a sentence clarifying the scientific gap (e.g., “however, simulating interannual variability might be a problem...”). It should be made clear that the paper is mostly based on simulation available in the literature (i.e., the sentences “we show that hydrologic models...” and “we analyse 18 regional and global hydrologic models...” are quite ambiguous regarding the nature of this technical note).

We will revise the abstract to be clearer about the scientific gap and the models that we analysed.

L13 – is “irregular variance” the best term here?

The decomposition of time series into trend/interannual, seasonal, and irregular components is standard terminology (Persons, 1919). Other terms have also been used for the irregular component (including remainder, noise, residual, and error) but we feel that ‘irregular’ is the most appropriate term in a hydrological context. At line 32 we will add these other terms in parentheses.

L20 – how were “ecologically relevant” signatures determined?

These are based on the indicators of hydrologic alteration (IHA) in addition to some other signatures that have been used in the (eco)hydrology literature – see Table 1 for references. We will add the following underlined text at line 122:

“The ecologically-relevant signatures that we consider are the 32 indicators of hydrologic alteration proposed by Richter et al. (1996) in addition to...”

L21-23 – It would be nice to finalize the abstract with the important technical implications for hydrologic modeling (so what should we do now?) rather than a general comment about climate change and vulnerable regions (not really the core topic of this TN).

Thank you for the suggestion. We will include a recommendation about model evaluation. However, changes to modelling practices beyond model evaluation were not investigated here.

INTRODUCTION – The story is not clear. First 12 lines about streamflow and climate change. But this TN is about performance metrics. It seems that the most important paragraph starts at L50. This paragraph should be developed further to clarify the relevance of this TN.

We will revise the introduction to clarify the story. In the introduction (lines 31-50), we are building the argument that:

- (i) Interannual variance is separate from seasonal and irregular variance components and is driven by different processes (L31-35).
- (ii) Interannual variance is important because it has consequences for ecosystems and water management and because climate change is creating non-stationarities that increase interannual variance (L36-39).
- (iii) Accurately simulating historical interannual variability and change is a good indicator that models are suitable for climate change projection (L40-42).
- (iv) The KGE and NSE do not evaluate interannual variance separately (L42-47)

Points (i) and (iv) are important to make as they have not been previously made in the hydrologic literature, as far as we can tell. Points (ii) and (iii) illustrate why we should care about interannual variance, which is relevant to the reviewer’s later comment that “It should be clear in the intro why changes in hydrological signatures should be evaluated”. We will make the link between interannual variability and changes in hydrologic signatures clear here.

L37 & L56 – What about each of these references is interesting? Expand on it or cut it out.

L37: The references highlight different types of non-stationarity and how hydrologic signatures are used to detect and/or measure it. We will clarify the relevance of each reference to different types of hydrologic signature.

L56: These were simply examples of papers that compare their model performance to the climatological benchmark. We will remove the sentence.

METHODS –

Section 2.1 describes several data selection choices. Perhaps, moving Section 3 before Section 2 would be better.

Thanks for the suggestion. We prefer to describe the methods first, otherwise it is unclear for what the data will be used. The data selection choice (filtering to gauges with a minimum of 10 years of data with no missing days) is an important part of the methods, without which it is not possible to calculate the Fast Fourier Transform.

L74-77 – this is not completely clear.

We will expand the description of the decomposition, also in response to reviewer 1.

*Section 2.2 – this was done by running a model or using available simulations?
The language is ambiguous here.*

The climatological benchmark ‘model’ is not really a model in the way that most hydrologists think of models. It is simply the mean flow for each calendar day of the year. This is briefly explained at line 51 but we will revise to remind the reader of this fact in Section 2.2.

L91 – Isn’t the NSE using the average streamflow as a benchmark?

Yes, this is the benchmark inherent in the NSE. However, it may be an unreasonably naïve benchmark for seasonal catchments. To quote Schaefli & Gupta (2007):

“The use of the mean observed value as a reference can be a very poor predictor (e.g. for strongly seasonal time series), or a relatively good predictor (e.g. for time series that are essentially fluctuations around a relatively constant mean value).”

For this reason, the climatological benchmark performance NSE_{cb} is sometimes used as an alternative benchmark. We will change the section title to “Climatological benchmark performance.”

Section 2.3 – again using “Modelling” in the title is a bit ambiguous as to the methods.

We will revise. We suggest “**Representation of interannual and seasonal variability in hydrologic models**”.

L110 – Where do I_o and I_s come from?

These are the interannual components derived from time series decomposition on the observed and simulated time series (Section 2.1). We will introduce this notation in Section 2.1.

L128 – Why is this interesting? It should be clear in the intro why changes in hydrological signatures should be evaluated. There are several important references missing here.

In the introduction we briefly explain what interannual variability is, and why it is important. (paragraphs 1 and 2). We will revise these paragraphs to also explain how hydrologic signatures are used to characterize this interannual variability and how hydrologic signatures relate to (changing) hydrologic processes. We will include key references from the ecohydrological literature such as Poff et al (1997), Poff and Zimmerman (2010) and McMillan (2021)

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegard, K. L., Richter, B. D., Sparks, R. E., & Stromberg, J. C. (1997). The Natural Flow Regime. *BioScience*, 47(11), 769–784. <https://doi.org/10.2307/1313099>

Poff, N. L., & Zimmerman, J. K. H. (2010). Ecological responses to altered flow regimes: A literature review to inform the science and management of environmental flows. *Freshwater Biology*, 55(1), 194–205. <https://doi.org/10.1111/j.1365-2427.2009.02272.x>

McMillan, H. K. (2021). A review of hydrologic signatures and their applications. *WIREs Water*, 8(1), e1499. <https://doi.org/10.1002/wat2.1499>

We will also include references to other studies that demonstrate non-stationarity such as Gudmundsson et al (2021), Berghuijs et al (2025), Stewart et al (2005), and Wasko et al (2020):

Gudmundsson, L., Boulange, J., Do, H. X., Gosling, S. N., Grillakis, M. G., Koutroulis, A. G., Leonard, M., Liu, J., Müller Schmied, H., Papadimitriou, L., Pokhrel, Y., Seneviratne, S. I., Satoh, Y., Thiery, W., Westra, S., Zhang, X., & Zhao, F. (2021). Globally observed trends in mean and extreme river flow attributed to climate change. *Science*, 371(6534), 1159–1162. <https://doi.org/10.1126/science.aba3996>

Berghuijs, W. R., Hale, K., & Beria, H. (2025). Technical note: Streamflow seasonality using directional statistics. *Hydrology and Earth System Sciences*, 29(13), 2851–2862. <https://doi.org/10.5194/hess-29-2851-2025>

Wasko, C., Nathan, R., & Peel, M. C. (2020). Trends in Global Flood and Streamflow Timing Based on Local Water Year. *Water Resources Research*, 56(8), e2020WR027233. <https://doi.org/10.1029/2020WR027233>

Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2005). Changes toward Earlier Streamflow Timing across Western North America. *Journal of Climate*, 18(8), 1136–1155. <https://doi.org/10.1175/JCLI3321.1>

DATA – This data and simulation use should be clarified in the abstract and introduction sections.

We will revise and clarify these points.

RESULTS AND DISCUSSION

Section 4.1 – What is a highly seasonal catchment? What are the signature values that were used to classify the catchments?

In section 2.3.1 we define ‘highly seasonal’ to mean a seasonal variance fraction greater than 0.5. We will remind readers of this at the beginning of section 4.1.

L194-205 – A lot of climatological explanation here, but nothing about important hydrological catchment characteristics. What is the area of the chosen catchments? What is average annual rainfall? What is ET? Why were these three catchments selected?

These catchments are selected to illustrate the diverse drivers of interannual variability around the world. They are not exhaustive and we did not perform any formal analysis of how high interannual variability is controlled by catchment characteristics. We would prefer not to overload the reader with a long list of catchment characteristics and encourage the reader to look at the references, which provide much more detail about the locations in question.

We will add the following at L205: “The preceding examples are selected to illustrate the diverse drivers of interannual variability around the world. In-depth analysis of their hydrologic conditions is considered out of scope for this work.”

Section 4.3 The discussion here is not linear and difficult to follow. This section could be reduced considerably and the paragraphs should be grouped around main messages.

We will revise and condense this section.

L275 – Is this hypothesis exhaustive? Could you think about any other case where that would happen or any exception to this?

We will change “generally confirms our hypothesis” to “supports our hypothesis”. We do not claim to have *proven* that hydrological models will always simulate interannual variability more poorly in seasonal catchments, and this hypothesis is probably not provable in a traditional sense. Certainly there are many reasons for the behaviour, some of which we discuss in L316-338.

CONCLUSIONS AND RECOMMENDATIONS – This section is a bit convoluted and repetitive. The conclusions should strictly address the knowledge and recommendations without repeating the results section (e.g., “higher than 0.8. We observe, in Figure 3...”).

We will revise and condense.

L378-383 – a bit of repetition of the introduction.

We are a bit confused by this comment as L378-383 (reproduced below) are stating and interpreting results of the current study, and we don’t feel that any of the statements from L378-383 repeat information in the introduction.

L378: This is most evident in tropical, alpine, and polar regions, where most of the variance in streamflow is seasonal. Poor interannual performance in these regions (and in some cases almost complete failure to simulate year-to-year variability) raises concerns about the ability of these models to accurately simulate nonstationary hydrologic processes and responses to climate change. This is especially worrying because these regions may be some of the most vulnerable to climate change (Flores et al., 2024; Pepin et al., 2022; Rantanen et al., 2022) and are historically less-well studied regarding hydrologic extremes (Stein et al., 2024).

L387 – “Lastly...” Why is that? How much is enough?

As discussed from L349-L357 these regions are underrepresented in available datasets. Observational data are required to assess how well models perform in these regions, and to infer how well they might predict long-term changes. We will revise to tie this statement at L387 clearly to the discussion in L349-357.

We add here that this underrepresentation by climate zone masks an even greater geographic bias. The available data from the tropical rainforest zone are mostly from Brazil and the available data from the coldest climate zone (‘Polar Tundra’) are almost entirely from Europe and the Americas. In the gauges shown in Figure 2 there are, for example, no data from the world’s two largest islands, one of which is very cold (Greenland) and the other which is primarily tropical rainforest (New Guinea). There are data from a single gauge on the world’s third largest island, Borneo. There is also a single gauge each on Sulawesi and within

the Congo rainforest. There are no gauges in the Polar Tundra zone within the Hindu Kush or Karakoram ranges and only six in all the mountains of central Asia.

We will decline to give a precise number of how many long-term time series are 'enough' and only say that the numbers of stations in the aforementioned regions (zero to one) are low by any reasonable assessment.