Response to Referee #1

Responses are written in blue.

Changes made to the manuscript are written in red.

The focus of this study is on seasonality of forcings (i.e., watershed inputs) and streamflow (i.e., outputs) and how the former is translated into the latter through watersheds functioning. To understand the role of watersheds in dampening of forcings seasonality, authors develop two signatures (namely, the amplitude ratio and the phase shift) and show how combinations of linear models result in certain values for these two signatures. Subsequently, they calculate values for the same signatures using data from several watersheds in the UK and US and overlay the results on top of linear model findings. In this way, they could devise a perceptual model for a given watershed, e.g., two parallel linear reservoirs show to be suitable to model streamflow in some catchment. Finally, authors assess two hydrologic models to figure out whether or not they could properly reproduce expected variations of these two signatures. This task helps evaluate structural adequacy of a given model. The paper is really well-written, and has high quality presentations. Because this research also provides theoretical foundations for the analyses in this paper, I consider it a great contribution. I believe that the proposed methodology has many applications in the field of watershed modeling and water resources management. Still, I have a few comments that are provided below, which might help improve the quality of this interesting manuscript. I would recommend minor revision.

We thank reviewer #1 for the helpful and encouraging feedback.

Comments: Maybe my most major comment is about similarity in concepts between this study and previous studies. Authors themselves also point out that several previous research have essentially relayed the same type of information, but maybe using different techniques (such as unit hydrograph, transit time distributions, etc.). I still do not completely understand what the benefits of the proposed method are, and this requires a dedicated section in the paper. Basically, any other quantitative tools that highlight the differences between the time series characteristics of inputs and outputs could be used here too. For example, we could simply use lag time between forcings and streamflow time series, or maybe variance of these time series, to investigate watershed functioning. For instance, if the ratio between normalized variance of inputs and outputs is really small, watershed might be groundwater dominated. Such a situation would be actually the case with low amplitude ratio under the proposed method. My question is, ‘what makes this method unique or better in comparison to other methods?

Thank you for pointing that out. We have indeed pointed out similarities to other techniques, we however think that they do not necessarily relate to the same type of information. Transit times focus on the velocity of water particles and therefore yield different insights. Many other methods (unit hydrograph, lag time, variance of time series) focus on shorter time scales. We believe that the focus on seasonal dynamics can yield related yet additional information compared to methods focusing on event scales. Furthermore, we chose the approach because there are analytical solutions for how sine waves are propagated by linear systems. This allows for example to interpret the results in terms of configurations of linear reservoirs and to estimate their associated time constants. The suggested ratio of normalised variances will probably be related to the seasonal signatures, yet how exactly can such a number be interpreted beyond a qualitative statement like “this watershed might be groundwater dominated”? We will clarify the motivation for our approach in a revised manuscript.

We revised the introduction (in particular l.99-110 of the track-changes version) with the aim to clarify why we have chosen this approach rather than other approaches and with the aim to clarify the overall goals of this study.

Line 358-359: regarding limitations of this study, authors here mention that "In other climates with a less distinct seasonal pattern, or with two seasons per year our approach will not work". I would argue that there are other limiations that need to be mentioned here too. For example, the proposed method requires quite long records of data.

From the SI it can be seen that 10 years are enough to obtain a robust result for most places. But of course, we require at least a couple of years (i.e. seasonal cycles) to meaningfully fit a sine curve. We will add a sentence about data limitations.

We added a few sentences to Section 5.1 to clarify data limitations.

To robustly capture the average seasonal behaviour, we need relatively long time series. Comparing results from two different 10 year periods shows that the signatures are robust for the majority of catchments, i.e. their values do not differ substantially from one time period to the other (details are shown in the Supplement).

Authors claim that 'inference from observed values of the signatures' is a potential outcome of this method, but as I said, data is needed for this purpose, right?

The reviewer is correct that data is required for this purpose. We will clarify the sentence to make it clearer regarding what can be inferred from the signatures.

We added that data are needed for this purpose (l.644 of the track-changes version).

Moreover, most likely the method won't work for sub-annual time scales (because there are lots of hydrological non-linearities at smaller time scales.

We agree with the reviewer here. We decided to focus on the annual time scale because it has a clear physical meaning (see lines 106-110) and because the seasonal flow regime is of importance to many applications. We will emphasise that in a revised manuscript.

Note that SI 1.4 briefly investigates non-linear reservoirs.

We added a statement about the assumption of linearity (l.158-160 of the track-changes version) and we emphasised that we focus on the annual period (l.188-190 of the track-changes version).

Maybe, elaborate on different limitation aspects of this research in a separate section.

We will add a discussion of the limitations you mentioned to Section 5.1 and change the title of that section. We think that another separate section on limitations might not necessarily be helpful. For example, we discuss the limitations of the modelling exercise in Section 5.4 (line 507-519), where we think it fits best.

We revised Section 5.1 and tried to clarify the limitations elsewhere in the manuscript.

Other minor comments: Line 125: explain how multiple linear regression method will be used. I haven't seen any material so far that explains how linear regression could be useful.

We used multiple linear regression to fit sine waves to data. This is explained in SI 2.1.2. We will add a clearer reference to that in the text.

We tried to emphasise that the sine-fitting methods can be found in the SI (l.140-144 of the track-changes version). We also expanded the section on linear regression in the SI.

Line 546: 'reduce the need for calibration'. . .I don't think so. Maybe, signatures calculated in this research could be used as additional calibration metrics to improve the probability of getting the right answer for the right reasons. . .but not replacing the calibration process.

Once a certain arrangement of linear reservoirs is chosen, the signatures are associated with time constants of these reservoirs. For example, if we chose a model consisting of two reservoirs in series, the theory can be used to obtain the two time constants of the reservoirs. This might not replace the calibration process completely, but it could be used to limit parameter ranges or to fix certain parameters. Since we haven't tested that yet, we can't say whether that will be useful in practice. Yet in any case, as you have said, the signatures might be used as an additional calibration metric (which is also indicated by our modelling experiment). We will revise the paragraph to clarify this.

We revised the statement in the conclusion. It now says that the signatures *could be used as additional constraint in the calibration process* (l.649-650 of the track-changes version).

I have to say that, to me, the most interesting finding in this research is (lines 448-450: the attribute "fraction of highly productive fractured aquifers", which is a hydrogeological classification available for the UK, shows a much clearer pattern than any soil or geology attributes in the US.). This has great applications in model development for ungauged catchments.

Thank you. The question remains of how to get such a classification for other places than the UK.

Minor: Line 16: give a very brief meaning for the word 'seasonality'. . .later you use terms such as 'mean seasonal regime' or 'seasonal streamflow regime' or 'seasonal signatures', which will make more sense if a clear description of seasonality is provided at the beginning

We will revise the first paragraph to clarify the meaning of the word seasonality.

We changed the beginning of the first paragraph (l.21-24 of the track-changes version) to clarify what we mean by seasonality.

Line 44-45: Shafii and Tolson (2015) is another reference that needs to be cited here

We will add that reference.

We added the reference which we think fits well here.

Line 73-74: this sentence is a bit unclear: 'a signature describing how climate seasonality is translated into streamflow seasonaltiy adds a timing component with a focus on seasonal and thus slower dynamics.'

The obtained phase shift tells us how long – on average – the seasonal forcing peak is delayed before it becomes the seasonal streamflow peak. This time lag (e.g. 1 month) is what we mean by timing component. We will revise that sentence.

We reformulated that sentence and moved it to another paragraph (l.103-104 of the track-changes version).

Line 237: please explain what you mean by 'fast flow routing delay (1 to 5 days)'

We will add a more detailed description of the model parameters in the SI.

We added more details on the parameter ranges to the SI.

Thank you

Thank you for your review!

Response to Referee #2

Responses are written in blue.

Changes made to the manuscript are written in red.

Referee report on Hydrological signatures describing the translation of climate seasonality into stream flow seasonality by Gnann, Howden and Woods

In their manuscript the authors analyze how long term (seasonal) variations in precipitation time series translate into (long term) variations in stream flow. To do so the authors decompose the precipitation and corresponding stream flow time series into their Fourier modes and analyze the mode corresponding to the annual (seasonal) cycle.

The paper is well written and addresses the problem of signal and forcing from a point of view which is more common in electro-technical engineering than in hydrology. Thus the paper may help to stimulate the field by introducing new methods and alternative approaches to analyze the relation between input-output time series. Below some comments and suggestions which should help the authors to improve and strengthen their manuscript.

We thank reviewer #2 for the helpful and encouraging feedback.

Abstract: "We approximate [..] by sine waves." Input and output signals are not periodic per se, but show recurring patterns. In order to address this point the authors may simply rephrase the above statement with something like: "In order to analyze the seasonality relations between input [...] and output we represent the two time series by their seasonal (annual) Fourier mode." Such a formulation avoids the criticism that the signal itself periodic, while keeping all the rest of the analysis unchanged.

Thank you for the suggestion, we will revise the text accordingly.

We revised the abstract.

Sec 2.2.1: 1 year Fourier mode: It would be interesting to see for an example how the different Fourier modes are represented in the spectrum of the time series. Such a measure would show how "strong" the annual mode is compared to the other modes of the signal.

We did a quick analysis to check how strong the annual mode is in comparison to other modes. We calculated one-sided power spectra and extracted their maxima for all catchments. Two examples (following a copy of Figure 4 from the paper) are shown below (Fig. 2).

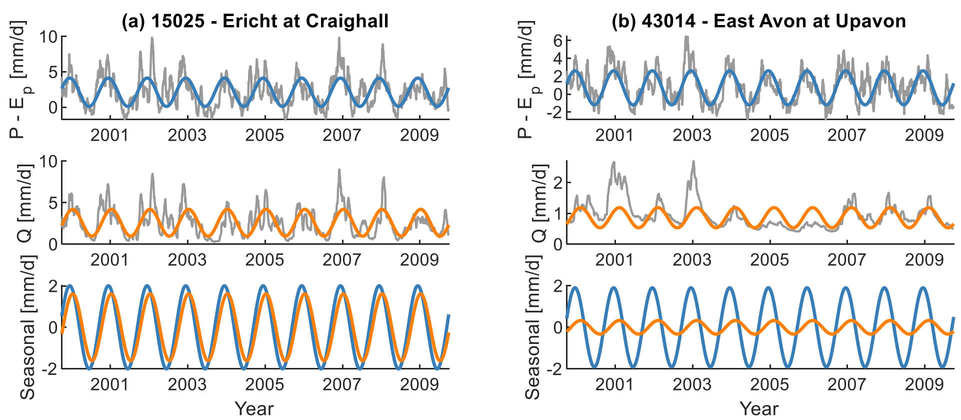

Fig 1. Climate input (P - E_p) and catchment output (Q) for two catchments in the UK, and their respective seasonal components. The time series are smoothed using a 30-day moving mean. The Ericht is a rather responsive catchment (BFI = 0.47), while the East Avon has a large baseflow component (BFI = 0.89). Note that for the bottom plots ("Seasonal") the mean values of the sine curves are set to zero. (Figure 4 in the manuscript.)

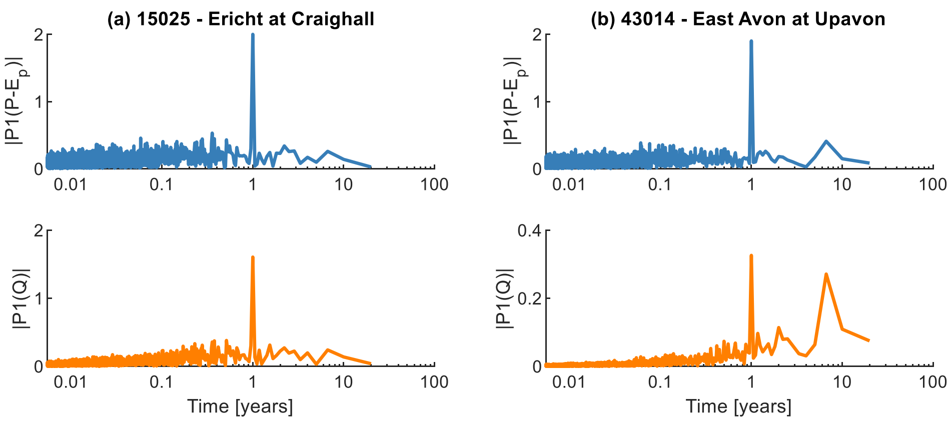

Fig 2. One-sided power spectra of climate input (P - E_p; blue) and catchment output (Q; orange) for two catchments in the UK.

For almost every catchment in our manuscript (~99%) the strongest forcing Fourier mode is the annual mode. For a few catchments in the US a 0.5y mode is the strongest, yet typically there is also a 1y mode present. Some of the streamflow data show strongest modes different from 1. Yet again, this doesn't mean that there is no annual mode present. For example, in panel (b) below we can see a strong multi-annual mode and the annual mode. We can also see that the groundwater dominated catchment (b) seems to act as a low-pass filter, dampening signals with shorter periods stronger than signals with longer periods. In principle, we could analyse more periods than the annual period and perhaps contrasting different periods might yield other interesting insights. But we have decided to focus on the annual time scale because it has a clear physical meaning (see lines 106-110). We will add the Fourier spectra to the SI.

We added the Fourier spectra to the SI and mentioned in the text that the annual mode is the dominant most in the majority of catchments (l.125-126 of the track-changes version).

Line 110: Although notation is an arbitrary choice, I would suggest the authors to use "PET" or at least "E_{PET}" in order to refer to Potential Evapo-Transpiration.

Thank you for the suggestion, but we would prefer to stick with our notation.

Reducing the in-/output signal by putting all weight of the time-series into the single (seasonal) Fourier mode may be problematic for analyzing real world data where: a. It is not per se clear that the overall dominant part of the signal. (Here as mentioned above the spectrum should give insight)

See above for an answer to that question and for Fourier spectra.

b. Additionally the different modes of the input signal do not necessarily need to be linearly coupled with modes of the same frequency in the output. Thus, it should be made clear that the description in section 2.1 relies on the assumption of a single wavelength forcing and a linear response system.

We will state these assumptions more clearly in a revised form of the manuscript.

We added some statements on the assumption of linearity (l.158-160 of the track-changes version).

Note: Due to linearity, all derivations presented in 2.1 should be valid for any Fourier component of the forcing function with F_n=A_n\exp(i*k*t) where A_n is the amplitude of the corresponding mode in the Fourier series.

Yes, the theory is not limited to the annual model. Yet as we've noted above, we focus on the annual mode as it is the dominant mode and as it has a clear physical driving force.

Figure 4: As mentioned before it would be interesting to see, how the blue and orange modes are represented in the corresponding spectra. If the seasonal modes are by far the most dominant frequencies in the signal it could help to justify for the single mode forcing model.

See above for an answer to that question and for Fourier spectra.

Sec. 4.2: Given the heterogeinity of natural systems it is not too surprising that a single linear (reservoir) model is not sufficient.

We agree on that, but we thought we start with rejecting the simplest model.

Fig. 6a and 7a: I would suggest the authors to use a two color divergent color scale to distinguish between negative and positive l_m (blue to white for neg. and white to red for pos values)

We originally intended to stick with the RGB colour schemes introduced by Knoben et al., 2018. We agree, however, that the colour scale is not the best choice in our case. We will change that accordingly.

We changed the colour scale in Figures 6a and 7a.

Another critique of Figs.6/7 is that the high point density can hide variabilities, especially when the points are plotted in a sorted manner, e.g. sorted by amplitude In order to avoid such a situation one could first randomize the sample with respect to the variable of the color bar.

At the moment, the points are plotted based on the list of catchments we've used. That is, neither completely random (the catchment list tends to follow geographical locations) nor sorted by anything specific such as amplitude ratio. We will check whether the plotting order influences the figure and improve the information content of the plot if possible.

We now plot the points in random order (we have also tried different random seeds), which didn't lead to a (significant) change in the patterns visible in Figures 6 and 7.

Section 4.3 requires some more details how the models were set up and parameters were varied/chosen (This can be added to the SI). Examples are: Line 333: Running IHACRES with 20 000 parameter sets. - Which are the parameters? - What are the parameter ranges that were varied? Line 335: The sentence "Plotting curves [...] produced by a certain set ..." needs some clarification. Questions which may arise here are: - How was the parameter set being chosen? - Was it always the same for all different catchments? - Did the authors perform a parameter sensitivity analysis?

Thanks for pointing out places where we were unclear in the modelling part. We will add more details on the modelling part to the SI.

We added more details on the modelling experiment to the SI and revised Section 2.3.

Line 343: "[...] with varying forcing.": Why do the authors introduce here the aridity index AI=PET/P=1-F/P as a nonlinear transformed quantity of F=P-PET rather than using their definition directly. Alternatively if the aridity dependence is the point to make here the authors should simply say this: "[...] does not vary substantially with varying AI=PET/P=1-F/P."

Thanks for pointing that out. Indeed, using the aridity index here is not necessary. The main purpose was to point out that each line corresponds to a different forcing input. We will change that to the moisture index I_m so that it's consistent with Figures 6 & 7.

We changed Figure 9 and used the moisture index as indicator for climate.

— I hope that the authors find my comments & suggestions useful to to improve the manuscript and strengthen their arguments.

Thanks again for reviewing our manuscript!

References

Knoben, W.J., Woods, R.A. and Freer, J.E., 2018. A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data. Water Resources Research, 54(7), pp.5088-5109.

Response to Referee #3

Responses are written in blue.

Changes made to the manuscript are written in red.

This review was prepared as part of graduate program course work at Wageningen University, and has been produced under supervision of dr Ryan Teuling by a student that prefers to stay anonymous. The review has been posted because of its good quality, and likely usefulness to the authors and editor. This review was not solicited by the journal.

Peer review on "Hydrological signatures describing the translation of climate seasonality into streamflow seasonality" by Gnann et al.

The manuscript "Hydrological signatures describing the translation of climate seasonality into streamflow seasonality" by Gnann et al. proposes two new hydrological signatures: the amplitude ratio and phase shift between the climatic forcing and the streamflow. The aim of this research is to use these signatures to quantify the catchment response to climatic forcing and use them for model evaluation. To determine the amplitude ratio and the phase shift, a sine function is fitted through both the climatic forcing and the streamflow. The climatic forcing is defined as the precipitation minus the potential evapotranspiration. The signatures are interpreted with the response (signatures) of linear reservoirs in series or parallel to climatic forcing. To test if the signature values are hydrologically interpretable, signatures for catchments in the UK and the US are defined and related to catchment characteristics to see if there is a pattern. Two models are discussed based on the signature range that they can produce. The authors conclude that the signatures can be used for model evaluation and to help model builders decide on the model configuration. The use of hydrological signatures to define a model configuration is a novelty, it would be interesting to look for other hydrological signatures and further investigate the abilities of this method. The phase shift is an interesting signature because it could quantify the time delay between climatic forcing and streamflow. However, my main concern is on the way the signatures are used here to evaluate models. The method is not appropriate, the model evaluation is not complete and no comparison is made with other evaluation methods. Furthermore, I also have some critical remarks on the proposed new signatures. They have a low accuracy and are not widely applicable. My last concern is about the conclusions, which are all based on visual interpretation instead of statistical analysis. Because of these reasons, I do not see the added value of this manuscript to the existing body of literature and therefore I recommend to reject the manuscript.

Thank you for your review and the feedback on our work.

To start with, I will explain my main concern on the model evaluation using the proposed method. In the paper a new way of model evaluation is proposed, namely looking at range of values of signatures (phase shift and amplitude ratio) that different models can produce. To test how large the range of produced signatures by the models is, a Monte Carlo sampling experiment is done. The authors state that this new method could be more meaningful and fit-for-purpose than already existing model evaluation methods: "Signatures rooted in hydrological theory offer a potentially more meaningful and fit-for purpose alternative to the typically used statistical metrics such as the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe". I do not agree with this argument, I will discuss the flaws of this method in the next paragraph.

The reviewer introduces a concern about the use of hydrological signatures for model evaluation, especially in exploring the range of hydrological responses that a model can produce. We do not make any claims about the novelty of the evaluation method itself. As stated in the manuscript, the idea of evaluating a model's response before calibration follows the idea of Vogel and Sankarasubramanian (2003).

First, I would like to raise attention to the fact that only two models are tested and no comparison is made with already existing model evaluation methods. I think a much more extensive approach is needed if they want to propose this as an alternative for the already existing model evaluation methods. More models need to be tested and the outcome of this evaluation method needs to be compared with outcomes of other model evaluation methods to see if they are in line and whether this method really gives more meaningful outcomes.

We would like to emphasise that the model evaluation is not the primary point of this paper, but the presentation of the seasonal signatures, which is one reason why we've kept the modelling part reasonably short. We will try to emphasise that more clearly in a revised version the manuscript.

We do not intend to present a full alternative to existing model evaluation methods. We primarily want to show how the signatures might be used as an additional source of information in model evaluation. We agree that a more extensive approach would be needed if the aim was a comparison to existing model evaluation methods, but this is not our intention.

We revised the introduction (l.48-54, l.99-110, and l.115-119 of the track-changes version) and other parts of the manuscript (e.g. the end of the abstract, and Section 2.3) to clarify the main aim of the paper.

First of all, in the manuscript only two figures show the results of the model evaluation with this new method. These figures alone, are not enough to evaluate the two models. Quantitative statements on the model functioning are needed, i.e. how well does the model predict the streamflow? All conclusions are based on visual interpretation, but graphs can sometimes be misleading, statistical analysis would be much more appropriate to compare different models on their functioning. In this manuscript only two models are tested, but if a lot of models need to be tested, numbers would make it easier to tell which model is best instead of comparing a lot of graphs.

We do not aim at evaluating whether streamflow is predicted well or not. In fact, we wouldn't expect streamflow to be predicted well based on the seasonal signatures alone, since they only aim at a certain aspect of the catchment response. The aim here is not to compare model runs from individual parameter sets with observed streamflow. We are primarily interested in the overall capabilities of the models. From Figure 6 we can see that GR4J (given the parameter ranges used) cannot reproduce what we observe. The question we try to answer here is not "which model is best". Rather, we want to test whether a certain model (given the parameter ranges used) is generally capable of producing the range of observed signatures, and thus cannot be rejected (see Vogel and Sankarasubramanian, 2003).

We revised Section 2.3 to clarify why we've chosen that approach to model evaluation.

Secondly, the choice of catchments used for model evaluation influences the outcome. For this experiment, 40 catchments in the UK are used. However, the UK catchments show better relationships between the signatures and catchment relationships (see figure 6 & 7), so the choice of using UK catchments instead of US catchments influences the outcome of testing this method.

We agree that the choice of catchments influences the outcome. But we think that choosing catchments in the UK is reasonable exactly because the seasonal signatures we propose are more robust in the mostly energy-limited UK. We will state our reasoning for using this subset of catchments more clearly in a revised version of the manuscript. We will also emphasise that the results of the modelling experiment are only valid for the UK.

We revised Section 2.3 to clarify why we've used this subset of catchments (l.260-261 of the track-changes version). We also added a sentence to emphasise that the results of the modelling experiment are not necessarily transferable to catchments outside the UK (l.601 of the track-changes version).

Thirdly, the number of parameters differs for the two models. Whether the difference between the signature space of the models is due to model functioning instead of the used range for different parameters, is questionable. My suspicion increases when reading line 481-483 "The actual reason...in Figure 2." and line 500-501 "Particularly the flow...than 60 days.", it seems that the signature output is determined by the parameter range instead of the model functioning, so how will this method evaluate models in an objective way then? The conclusion that the signatures are a diagnostic tool because GR4J is not capable of modelling the whole signature space (Line 427-428), is thus not valid in my opinion!

We agree that the results depend on the parameter ranges. Specifying parameter ranges always involves some subjective judgment. We mostly used the default ranges from the MARRMoT toolbox (Knoben et al., 2019), which are intended to be wide. We will have a look at recent literature on the parameter ranges. We will investigate whether broader ranges influence the results and we will update the parameter ranges if necessary. We will also try to emphasise the limitations of choosing certain parameter ranges more clearly.

We reran GR4J with wider parameter ranges. Details on the ranges can be found in the Supplement. The new results are shown in Section 4.5 and discussed in Section 5.4. While the wider parameter ranges do lead to wider signature ranges, our overall conclusions stay the same. We tried to focus the discussion more on diagnosing why GR4J leads to these results. We also changed Figure 9 and hope that the new figure helps to underline our message.

Lastly, only a small part of model predictability aspects is evaluated. Pechlivanidis et al. (2011) summarized different model evaluation methods, where they discuss different objective approaches. Objective functions are here defined as numerical measures of the difference between the model simulated output and the observed (measured) catchment output. The Nash-Sutcliffe Efficiency (NSE) and Kling and Gupta Efficiency (KGE) are examples of this approach. The proposed method here is an objective function as well, since produced signatures by models are compared with observed signatures of catchments. The KGE has been introduced to overcome some limitations of NSE, this method analyses the correlation, the bias, and a measure of relative variability in the simulated and observed values (Gupta et al., 2009). This method evaluates thus more aspects of model functioning than the new proposed method here, which only gives an indication of the ability of a model to attenuate the climate forcing into a streamflow signal with a right time delay (if the signatures are correct!), but not if the model can produce the right streamflow variability and mean, peak and low flows. The authors could improve the method by evaluating the models based on more hydrological signatures and quantify the model functioning. Furthermore, they could do test more models and compare the outcomes with other model evaluation methods.

We absolutely agree that in a "general" model evaluation, we should look at other aspects of the hydrograph, ideally by using multiple hydrologically interpretable signatures. Yet we did not intend to evaluate these two models in general, but only with respect to the proposed signatures. We also agree that the proposed signatures could be used as an objective function, we however decided against such an evaluation approach. Instead, we focused on the range of possible model responses. This doesn't mean that for individual catchments, the model would have to be rejected, but as a model for all the catchments investigated, it would have to be rejected (given the parameter ranges chosen). This might be particularly helpful for large sample studies, where often one or a few model structures are chosen a-priori for all catchments (see also Addor and Melsen, 2019, who show how models are often chosen based on legacy rather than adequacy) .

We revised Section 2.3 to clarify why we've chosen this evaluation approach.

They could also improve the transparency of this method by adding a table with the changed parameters and the range.

Thank you for the suggestion. We will add a table with the parameter ranges to the Supplement.

We added the parameter ranges to the Supplement.

Finally, they should argue why although different parameters of the two models are changed, the model outcomes can still be compared.

Regarding the chosen parameter ranges, we will add more information on that in a revised manuscript. Regarding the fact that different models have different parameters, we think that's inevitable when working with different models. Different models will have different parameters and sometimes even if they have the same name they might actually have a different meaning.

We added more information on the modelling experiment and the parameter ranges to the Supplement.

My second major concern is about the signatures, they have a low accuracy and are not widely applicable. First, I will address the accuracy of the signatures. To determine the phase shift and amplitude ratio, a sine function is fitted on the climate forcing (PETp) and the streamflow. The method of fitting a sine function through the forcing and streamflow time series does not seem adequate to me. Most of the catchment regimes do not show a clear sinusoidal yearly cycle. This is well visible in the 16 different regime types, after Weingartner and Aschwanden (1992). For example, catchments that show two discharge peaks in one year cannot be described well by a sine function, this will lead to an error in the phase shift. In the paper two examples are given were a sinusoidal function is fitted on the climate forcing and streamflow. The timing of the sine function (phase) on the forcing and streamflow seems to be quite good in these cases. However, the sine function does not follow peak discharge and low discharges. This is clearly visible in the middle figure of the East Avon at Upavon catchment, the discharge peaks in 2001 and 2003 are not represented in the sine function (discharge is double the fitted discharge!), also the sine function does not follow the discharge in 2005/2006 when there is a low discharge. This shows that the sine fitting leads to errors in the amplitude ratio. Since the signatures are used for model evaluation, these errors could also lead to errors in the outcome of the model evaluation.

Linear regression is a commonly applied technique to extract sinusoidal components (Fourier modes) from time series (see e.g. Kirchner, 2016). The comparison between the two techniques shown in the Supplement shows the robustness of the sine wave extraction. For the method to be applicable, the time series does not have "to look like a sine curve", the sine curve is rather a description of just the average seasonal behaviour. So, the fitted sine wave is not intended to represent all the variability. The extremely high and low peaks visible in the East Avon are mostly caused by a multi-annual mode (~7 years, see also Rust et al., 2019) and hence cannot be captured by a sine wave describing the annual mode. We also refer to Referee #2 here whose suggestion might help to clarify that: "In order to analyze the seasonality relations between input [...] and output we represent the two time series by their seasonal (annual) Fourier mode." We will try to clarify that in a revised version of the manuscript.

We revised a few passages of the manuscript to highlight that we are interested in the average seasonal behaviour of the time series, approximated by their annual Fourier modes (e.g. l.4-8 and l.88-89 of the track-changes version).

Furthermore, the use of the potential evapotranspiration (ETp) leads to errors (and thus lower accuracy) in the signatures for semi-arid and arid catchments, since the potential evapotranspiration deviates from the actual evapotranspiration in these areas. This problem is mentioned by the authors in line 380-385. This problem can be solved by including a model to estimate the actual evapotranspiration (also mentioned by the authors). This would also help interpreting the signatures with the catchment characteristics. For example, in line 436-438 the authors state that the signatures of the US catchments show a relation with the moisture index. This conclusion is made based on visual interpretation of figure 7a. But I think this conclusion is not valid because the signatures of the dry catchments on the left side have a large uncertainty because of the use of ETp instead of ETact.

We agree that the signatures are unreliable in arid catchments and we also state that in the manuscript. We will try to emphasise this limitation more clearly in a revised version of the manuscript.

We revised a few passages of the manuscript and stated this limitation more clearly (e.g. l.18-19 and l.625-629 of the track-changes version).

The other disadvantage of the signatures is that they are not widely applicable. The problem of using the signatures for arid and semi-arid areas is already mentioned, but this could be solved by using the actual evapotranspiration. However, these signatures are also not valid for catchments with precipitation falling as snow. Since catchments with precipitation as snow show a typically seasonal cycle, the need of leaving these out of consideration is a major lack of the proposed signatures.

Snow, while undeniably important, is a fundamentally different process and we want to avoid conflating different processes. The seasonal cycle of a snow-dominated catchment does have a distinct seasonal pattern, but would not be well modelled by the approach we have taken here. For an alternative, see Woods (2009).

Furthermore, the signatures are also not valid for climates with a less distinct seasonal pattern, so this will further limit the applicability of the signatures. Because the signatures can only be used for a certain type of catchments, it is the question whether they contain new information on the streamflow seasonality of these catchments. There are already hydrological signatures that describe the response of streamflow to climatic forcing, for example the flow duration curve. A steep slope in the flow duration curve indicates a fast response of the streamflow to climate forcing whereas a flatter curve indicates a relatively damped response and higher storage (Yadav, 2007). Only the timing component might add new information, but since the method of determining the phase shift is not accurate, I do not see the added value.

We accept that the signature is limited to particular climates. While universal signatures applicable to every catchment seem desirable, we don’t think that’s realistic. In practice, using a specific signature to target specific processes that occur in specific places seems unavoidable. If the proposed signatures help us to better understand humid, non-snowy catchments (e.g. most of the UK), they still have the potential to add valuable information.

The last thing I would like to point out is that all conclusions based on visual interpretation instead of statistical analysis. For the sine fitting method, I would like to see the goodness of it or the sum of squared errors (SS), to know how well the fit of the sine function to the climatic forcing and streamflow is.

As noted above, the purpose of the sine curve is not to capture all variability in the signals, just to extract the seasonal component. Comparing the extracted sine wave with the observed time series via a goodness of fit measure will only be of limited use. As described before, the sine wave is not (and it is not intended to be) a particularly good description of the whole hydrograph. So, in catchments where the seasonal mode will explain most of the variability, we will get a “good fit” and in catchments where the seasonal mode explains little of the variability, we will get a “bad fit”. But this will not tell us whether the extraction of the annual mode is robust or not. To test that, we have used two different methods and we have compared the results from two different time periods as shown in the Supplement.

For the relationships between the signatures and catchment characteristics, it would be better to calculate the correlation coefficient instead of only the visual interpretation, since this might be misleading. The same goes for the model evaluation method, it would be nice to have a quantitative statement on how well the model works. This would also make it easier to compare more models, as mentioned before.

Thank you for the suggestion. We will add tables with correlation coefficients for Figures 6 and 7. We primarily use figures as they can show us complex patterns between three variables and allow us to compare the observed signatures to the theoretical results from Figures 1-3. This would not be possible just with correlation coefficients, which have their own drawbacks (for example, the Pearson correlation as a measure of linear correlation cannot describe non-linear relationships).

We added tables with correlation coefficients to the manuscript (Tables 2 and 3).

Minor issues and typo’s:

Minor issue 1: Line 68-70 “All of these ... streamflow seasonality.” I am not convinced. For example the slope of the flow duration curve can say something about the translation of climatic forcing into streamflow seasonality. A steep slope in the flow duration curve indicates a fast response of the streamflow to precipitation inputs whereas a flatter curve indicates a relatively damped response and higher storage (Yadav, 2007).

Yes, we agree that the FDC can say something about the responsiveness of a catchment, but the FDC has its own limitations (see e.g. McMillan et al., 2017). It combines multiple hydrological processes which limits its interpretability and it doesn’t yield an explicit time scale such as the phase shift.

Minor issue 2: Line 94-96 “The amplitude might ... seasonal component alone.” Why stating this if it is not done for this research, is it a follow up research suggestion? Then it should be placed in the discussion.

This is just a comment that relates the signatures to other metrics existing in the literature. It is not essential, so we will remove it from the manuscript.

We deleted the sentence.

Minor issue 3: “catchment form” can better be replaced by catchment characteristics (For example in line 100). Catchment form suggest you are looking at the effect of a small river with a lot of branches or a stretched river.

Catchment form as defined in Wagener et al. (2007) relates to “drainage area, average basin slope, pedology, and geology”. It is a commonly used term and we would prefer to stick with it.

Minor issue 4: The aim could be stated much more clearly, “test whether the seasonal signatures are useful for modelling practice (line 101)” not specific enough.

We will revise that paragraph and add more details in a revised version of our manuscript.

We revised that paragraph (l.115-118 of the track-changes version).

Minor issue 5: Line 110/111 “We use Ep ... would be needed.” Not a valid argument, how much would the uncertainty increase if you add another model?

We do think that it is a valid argument. Another model would introduce uncertainty in both choosing the model and potentially choosing parameter values.

Minor issue 6: In line 124 a small remark is made on the method of the sine fitting. This could be elaborated a bit more. Why use the sine fitting method? Which methods did you compare and why did you choose for the linear regression method (it is now in the supplement, but I think it is better to include it in the text)?

We decided to report details on the sine wave fitting in the SI to make the methods section more concise. We will clarify the use of the fitting methods in a revised manuscript.

We tried to emphasise that the sine-fitting methods can be found in the SI (l.140-144 of the track-changes version). We also expanded the section on linear regression in the Supplement.

Minor issue 7: A reference is needed to support line 200 “The upper limit...shape parameter equal to 2.”

We will add a reference.

We added a reference (l.226 of the track-changes version).

Minor issue 8: About figure 3, could you explain the form of the curve when ?2 becomes larger and ?1 and fraction going to second reservoir are constant.

Let’s first look at the red line in Figure 3(a) from right (black line) to left. tau_1 is always 1d, the fraction going into the second reservoir is 0.3, and tau_2 starts with a value of 10d and then increases. So at first, both reservoirs are rather fast and we get a high amplitude ratio and a small phase shift for the outgoing sine wave (which is a mixture of the sine wave coming out of the first and the second reservoir, see Eq. 14 and 15). Then, the second reservoirs gets slower, leading to a decrease in amplitude ratio and an increase in phase shift. As the second reservoirs gets slower and slower, it will contribute less and less to the overall sine wave. For very high values of tau_2 (10000d), the sine wave coming out of the second reservoir is almost a straight line (the amplitude ratio is close to 0), so the overall sine wave is primarily consisting of the sine wave coming out of the fast reservoir. Since only 70% of the total input went into the first reservoir, we will get a sine wave that’s 0.7 times the original amplitude with a very small phase shift, as the first reservoir hardly attenuates the signal. We will try to clarify that in a revised version of the manuscript.

We revised the corresponding paragraph with the aim to clarify the shape of the lines in Figure 3 (l.243-251 of the track-changes version).

Minor issue 9: Line 235, explain the choice for Latin Hypercube sampling.

Latin Hypercube sampling is an efficient method (Cheng and Druzdzel, 2000) that assumes uniform prior parameter distributions, which we think is adequate for the present case.

We added that statement to the manuscript (l.275-276 of the track-changes version).

Minor issue 10: Table 1, add more information on range variables. For example for moisture index: -1= most arid and 1= most humid.

We will clarify how these indices have to be interpreted in a revised version of our manuscript.

We added a sentence that clarifies how these indices have to be interpreted (l.301-306 of the track-changes version).

Minor issue 11: Figure 4: add color indication to description, climatic forcing (blue) and streamflow (orange).

We will add colour indications to the figure caption.

We added colour indications to the caption.

Minor issue 12: Figure 5: Based on what criteria are the benchmark catchments chosen (grey dots)? Same goes for the two red dots, random or do they represent a certain type of catchments?

The benchmark catchments are described in Harrigan et al. (2018). The two red dots are chosen arbitrarily based on their contrasting streamflow regimes.

Minor issue 13: Line 284, missing reference to table 1. Catchment attributes

Minor issue 14: Line 300, missing reference to table 1. Catchment attributes

We will add the references.

We added the references to the table.

Minor issue 15: Line 304-305 “Yet generally, ...in figure 6).” Statement is not explained in discussion, why are the US phase shift larger than for the UK catchments?

We do discuss the extremely large phase shifts in lines 457-469. These phase shifts are unreliable because these catchments are very arid. For the other catchments in the US, we couldn’t find catchment attributes that could explain all the observed behaviour, which is discussed in lines 447-456. So the answer to that question is that we don’t know (yet).

Minor issue 16: Figure 9b, Higher probability for high BFI for GR4J than IHACRES, but GR4J lower phase shift (max 60 days)!! Why? I would expect a larger phase shift when a larger part of the flow is slow flow.

Yes, we agree here. We would expect high BFIs to be associated with small phase shifts, but that doesn’t seem to be the case here. It might have to do with the internal parametrisation of GR4J.

We added a few sentences on why that might be the case to the discussion (l.593-596 of the track-changes version).

Minor issue 17: Line 393-395 “Since the BFI... seasonal signatures.” I do not agree, the BFI cannot be used as a cause for observed patterns, but it can be related to the observed pattern. A higher base flow means more slow flow so this could be related to a larger phase shift.

Yes, the BFI cannot be seen as a cause for the observed patterns, and that’s what we’ve written in lines 393-395.

Typo’s: Line 17: sensitive Line 64: minimum Line 73: seasonality Line 278: reproduce Line 496: outputs

Thank you for pointing out these typos and thank you again for reviewing our manuscript!

We fixed the typos.

References

Addor, N. and Melsen, L.A., 2019. Legacy, rather than adequacy, drives the selection of hydrological models. Water Resources Research, 55(1), pp.378-390.

Cheng, J. and Druzdzel, M.J., 2000, May. Latin hypercube sampling in Bayesian networks. In FLAIRS Conference (pp. 287-292).

Harrigan, S., Hannaford, J., Muchan, K. and Marsh, T.J., 2017. Designation and trend analysis of the updated UK Benchmark Network of river flow stations: the UKBN2 dataset. Hydrology Research, 49(2), pp.552-567.

Kirchner, J.W., 2016. Aggregation in environmental systems–Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments. Hydrology and Earth System Sciences, 20(1), pp.279-297.

Knoben, W.J., Freer, J.E., Fowler, K.J., Peel, M.C. and Woods, R.A., 2019. Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1. 2: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations. Geoscientific Model Development, 12(6), pp.2463-2480.

McMillan, H., Westerberg, I. and Branger, F., 2017. Five guidelines for selecting hydrological signatures.

Rust, W., Holman, I., Bloomfield, J., Cuthbert, M. and Corstanje, R., 2019. Understanding the potential of climate teleconnections to project future groundwater drought.

Vogel, R.M. and Sankarasubramanian, A., 2003. Validation of a watershed model without calibration. Water Resources Research, 39(10).

Wagener, T., Sivapalan, M., Troch, P. and Woods, R., 2007. Catchment classification and hydrologic similarity. Geography compass, 1(4), pp.901-931.

Woods, R.A., 2009. Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. Advances in water resources, 32(10), pp.1465-1481.

Hydrological signatures describing the translation of climate seasonality into streamflow seasonality

Sebastian J. Gnann[1], Nicholas J. K. Howden[1], and Ross A. Woods[1]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK

Correspondence: Sebastian Gnann (sebastian.gnann@bristol.ac.uk)

Abstract. Seasonality is ubiquitous in nature, and it is closely linked to water quality, ecology, hydrological extremes, and water resources management. Hydrological signatures aim at extracting ~~relevant information about hydrological behaviour, and they can be used to better understand hydrological processes and to evaluate hydrological models~~information about certain aspects of hydrological behaviour. Commonly used seasonal hydro-climatological signatures consider climate or streamflow seasonality, but not how climate seasonality translates into streamflow seasonality. ~~We propose and test hydrological signatures based on the attenuation of the~~ In order to analyse the translation of seasonal climate input ~~by a catchment. We approximate the seasonality in the input~~ (precipitation minus potential evapotranspiration) ~~and the~~ into seasonal catchment output (streamflow)~~by~~ , we represent the two time series by their seasonal (annual) Fourier mode, i.e. by sine waves. A catchment alters the input sine wave by reducing its amplitude and by shifting its phase. We propose to use these quantities, the amplitude ratio and the phase shift, as seasonal hydrological signatures. We present analytical solutions describing the response of linear reservoirs to periodic forcing to interpret the seasonal signatures in terms of configurations of linear reservoirs. Using data from the UK and the US, we show that the seasonal signatures exhibit hydrologically interpretable patterns and that they are a function of both climate and catchment attributes. Wet, rather impermeable catchments hardly attenuate the seasonal climate input. Drier catchments, especially if underlain by a productive aquifer, strongly attenuate the input sine wave leading to phase shifts up to several months. ~~Finally~~As an example application, we test whether two commonly used hydrological models (IHACRES, GR4J) can reproduce the observed ranges of seasonal signatures in the UK. The results show that the seasonal signatures ~~can~~ have the potential to be useful for catchment classification, for predictions in ungauged catchments, and to aid model building and evaluation. The use of potential evapotranspiration in the input restricts the applicability of the signatures to energy-limited (humid) catchments.

1 Introduction

~~Seasonal patterns are ubiquitous in nature, and many streams have a distinct seasonal flow regime. Streamflow seasonality is primarily driven by climate seasonality and thus sentitive to changes~~ The annual course of the earth around the sun leads to seasonal cycles in climate in many places. Seasonal patterns in precipitation, evapotranspiration, and ~~snow fraction~~ snowfall, as well as the characteristics of the catchment a stream drains, often result in a distinct seasonal streamflow regime (Cayan et al., 1993; Regonda et al., 2005; Berghuijs et al., 2014). The seasonal flow regime is closely linked to water chemistry and water quality (DeWalle et al., 1997; Vega et al., 1998). Streamflow seasonality plays a crucial role for biological systems and ecosystems (Colwell, 1974; Poff et al., 1997; Poff and Zimmerman, 2010). Low flows are typically seasonal, and droughts – albeit a more general phenomenon than low flows – often occur during the low flow season and thus are to some degree predictable (Smakhtin, 2001; Peters et al., 2003). From a more applied point of view, the seasonal streamflow regime is crucial for water resources management, agriculture, and hydropower generation (Weingartner et al., 2013; Laaha et al., 2013; Svensson, 2016; Harrigan et al., 2018b). This is reflected in the increased application and development of seasonal forecasting methods (Shi et al., 2008; Svensson, 2016; Harrigan et al., 2018b). In summary, for many applications the mean seasonal regime is of high importance and thus deserves attention.

In this work we focus on the mean average seasonal hydrological response of snow-free catchments. We do not focus, for instance, on the seasonality of events (e.g. storms), noting, however, that the seasonal water balance can have an impact at event scales (Berghuijs et al., 2014). The In snow-free areas, the seasonality of the flow regime is primarily driven by the incoming forcing, that is, the seasonality of precipitation (water) and potential evapotranspiration (energy). Given a certain forcing, the flow regime of a catchment is determined by a catchment's form and function, that is, by how much water can infiltrate, how much water can be stored, and how slowly that water is being released. Since groundwater recharge and thus groundwater discharge are is often very seasonal (Jasechko et al., 2014), many hydrogeological studies focus on seasonality, or more specifically on how seasonal recharge is propagated through an aquifer system (Townley, 1995; Erskine and Papaioannou, 1997; Peters et al., 2003; Obergfell et al., 2019). Slowly responding, groundwater-dominated catchments closely resemble the aquifer system feeding the stream. Understanding the seasonal streamflow regime is therefore crucial particularly important for understanding slow (groundwater-driven) dynamics in catchments.

Different aspects of hydrological behaviour, such as streamflow seasonality, can be quantified by summarising metrics now mostly called hydrological signatures (McMillan et al., 2017). The use of such summarising metrics is not new, and they have been used extensively in ecohydrological studies (e.g. Clausen and Biggs, 2000; Olden and Poff, 2003) and hydrological studies (e.g. Jothityangkoon et al., 2001; Farmer et al., 2003). Hydrological signatures offer a way to quantify hydrologic similarity, which . This makes them useful for catchment classification (Wagener et al., 2007; Sawicz et al., 2011), for hydrological process exploration (McMillan et al., 2014), and for predictions in ungauged basins (Sivapalan et al., 2003; Wagener et al., 2007; Hrachowitz e . More recently, hydrological signatures have become more popular as a way (Yadav et al., 2007; Hrachowitz et al., 2013; Westerberg et al., . Hydrological signatures can also be used to guide diagnostic model evaluation (Gupta et al., 2008; Peel and Blöschl, 2011; Euser et al., 20 . Signatures rooted in hydrological theory (Gupta et al., 2008; Peel and Blöschl, 2011; Euser et al., 2013; Hrachowitz et al., 2014; Shafii an , as they offer a potentially more meaningful and fit-for-purpose alternative to the typically used statistical metrics such as the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970) or the Kling-Gupta efficiency (KGE; Gupta et al., 2009).

There are many hydrological signatures and we therefore need guidelines for signature selection (McMillan et al., 2017; Addor et al., 2018). Some of these guidelines refer to more technical aspects: the uncertainty in a signature should not be larger within a catchment than between catchments (identifiability), a signature should be insensitive to the data sources (robustness), and a signature should be comparable across (heterogeneous) catchments (consistency; McMillan et al., 2017). When using combinations of signatures, the different signatures should also contain different information, i.e. they should not be redundant (Olden and Poff, 2003; Addor et al., 2018). From a more hydrological perspective, a signature should be meaningful at the relevant scale (representativeness) and a signature should relate to and increase our knowledge of hydrological function (discriminatory power; McMillan et al., 2017). ~~The latter aspect was, for example, highlighted by Addor et al. (2018), who stated that "signatures directly related to the water balance are already well explained by climatic indices" and that other "poorly-predicted signatures deserve more attention". Besides modelling~~ Since (hydro-)climatic signatures such as the mean flow ~~correctly~~are already well understood, we should try to explain and use signatures that tell us more about catchment functioning (Addor et al., 2018), such as signatures that relate climate input to catchment output.

There is a multitude of hydrological signatures focusing on seasonality. Climate seasonality is accounted for by (hydro-)climatic signatures such as the (co-)seasonality of precipitation and potential evapotranspiration (Milly, 1994; Knoben et al., 2018). Streamflow seasonality can be characterised by the Pardé coefficients (Weingartner et al., 2013) or the regime curve, which is related to the slow flow component of the flow duration curve (FDC; Yokoo and Sivapalan, 2011). Seasonal signatures related to streamflow timing are the half flow date and the half flow interval (Court, 1962), and the date of each annual one-day maximum ~~(or mininum; Richter et al., 1996)~~(or minimum; Richter et al., 1996). Seasonal streamflow signatures focusing on low flows are for example the seasonality index, which measures the mean day of low flow occurrence and the intensity of seasonality, or the seasonality histogram, which shows the ~~occurence~~ occurrence of low flows in each month (Laaha and Blöschl, 2006). Colwell's predictability is another measure describing periodic signals (Colwell, 1974), mostly used in ecological studies. It consists of constancy (how variable the intra-annual flow regime is) and contingency (how persistent the inter-annual flow regime is). All of these signatures describe (parts of) the seasonality of either climate or streamflow, yet none of them look at how climate seasonality translates into streamflow seasonality. As the transformation of climate input into streamflow is, after all, what we are trying to understand, investigating the seasonal aspect of that seems worthwhile. Relating streamflow to climate input also removes the arbitrariness of picking a start date (e.g. by defining a water year), which is a limitation of many signatures that relate flows to a date (e.g. the half flow date). ~~Furthermore, a signature describing how climate seasonality is translated into streamflow seasonality adds a timing component with a focus on seasonal and thus slower dynamics. This might make it a valuable addition to other (slow flow) signatures such as the baseflow index (BFI), or the flow duration curve and parts thereof (e.g. Q_{95}), which focus on volumes and frequencies, respectively.~~

In this work, we propose ~~and test~~ the use of hydrological signatures based on how catchments attenuate the seasonal climate input (forcing). We approximate the input signal to a catchment (the forcing) by precipitation minus potential evapotranspiration and the output signal from a catchment by streamflow. We quantify the ~~seasonality in~~ seasonal components of both signals by fitting sine waves to them~~.~~, i.e. we extract their (annual) Fourier modes. As the period is fixed (one year), the incoming sine wave and the outgoing sine wave differ only in their amplitude, their phase and their mean. As the mean is rather a measure of the annual water balance, we are primarily interested in amplitude and phase. The differences in amplitude and phase are used as signatures describing the steady-state response of a catchment to periodic forcing. This idea is ~~not new. It is~~ similar to the approach of Peters et al. (2003) who investigated drought propagation through groundwater using sinusoidal recharge, and to the approach of Obergfell et al. (2019) who used the seasonal behaviour as an additional signature in recharge estimation. The approach is also similar to approaches in transit time modelling (e.g. McGuire and McDonnell, 2006; Kirchner, 2016).

Instead of focusing on the velocity of water particles, we, however, focus on the hydraulic response to periodic forcing, that is the celerity of the input "wave" of hydraulic potential (Harman, 2019). The proposed signatures are essentially also spectral domain signatures (Montanari and Toth, 2007), focusing only on a certain meaningful period – the annual period.

While there are other methods that quantify input-output relations, we propose the use of the seasonal signatures for several reasons. The seasonal signatures can be related to conceptual linear reservoirs (this will be outlined in Section 2), i.e. they can be interpreted in terms of simple conceptual model structures and parameter values (the reservoir time constants or response times). This gives them some hydrological interpretability (cf. discriminatory power; McMillan et al., 2017). These signatures Furthermore, by quantifying the delay between seasonal climate input and catchment output, we obtain a time scale that focuses on seasonal and thus rather slow dynamics. This might make it a valuable addition to methods focusing on event scales (e.g. recession analysis) and to other slow flow signatures such as the baseflow index (BFI), or the flow duration curve and parts thereof (e.g. Q_{95}), which focus on volumes and frequencies, respectively. Lastly, the signatures do not require any parameters, they can be estimated directly from precipitation, potential evapotranspiration and streamflow data. The signatures can be related to the predictability measures proposed by Colwell (1974). The amplitude might be compared to the constancy (how variable the signal is), and the goodness of fit of the sine wave to the contingency of a signal (how much of the variability can be explained by the seasonal component alone), which makes it straightforward to apply them to large samples of catchments.

In the following, we will first define the seasonal signatures, and we will present analytical solutions describing the response of linear reservoirs to periodic forcing (Section 2). Second, we will calculate the seasonal signatures for a range of catchments in the UK and in the US (Section 4, the data sources are presented in Section 3). We will explore how they relate to hydro-climatic forcing and catchment form, and we will interpret the underlying hydrological processes as well as limitations of the approach (Section 5). ThirdFinally, we will test whether the seasonal signatures are useful in modelling practice, i.e. we will investigate present an example application, in which we test whether two commonly used hydrological models (IHACRES, GR4J) can reproduce the seasonal signatures observed ranges of seasonal signatures in the UK. This modelling experiment aims at exploring whether the signatures can be used as an additional source of information in model evaluation (Section 5.4).

2 Methods

2.1 Extracting seasonal components from time series

2.1.1 Quantification of periodic components

To analyse the periodic components (Fourier modes) of time series we first need to quantify these components. While we could investigate the whole frequency spectrum of our time series and see how this is altered by a catchment (Montanari and Toth, 2007), we will focus on a period T of one year. The annual period has a clear physical meaning as it is the period the Earth moves in its orbit around the Sun, which is directly linked to the energy input to the Earth system. Furthermore, the annual mode is the strongest mode in the vast majority of catchments investigated here (see Supplement for further details). The input to a catchment, the forcing F, is approximated by precipitation P minus potential evapotranspiration E_p ($F = P - E_p$). We use E_p to avoid the need for a model or additional data which would be needed to obtain actual evapotranspiration E_a. This might be particularly problematic in water-limited catchments, where actual evapotranspiration is much smaller than potential evapotranspiration, and in catchments where precipitation and potential evapotranspiration are out of phase. We will discuss that in Section 5. The seasonal component of the forcing F_{sin} is given by (Milly, 1994):

$$F_{\text{sin}} = \bar{F}\left(1 + \delta_F \sin\left(\frac{2\pi}{T}t + \phi_F\right)\right) \tag{1}$$

where $\bar{F}$ is the mean, δ_F is the ratio between the amplitude and the mean (the dimensionless amplitude), and ϕ_F is the phase (with respect to a reference date) of the seasonal forcing component. The output from a catchment is approximated by streamflow Q. The seasonal component of streamflow Q_{sin} is given by:

$$Q_{\text{sin}} = \bar{Q}\left(1 + \delta_Q \sin\left(\frac{2\pi}{T}t + \phi_Q\right)\right) \tag{2}$$

where $\bar{Q}$ is the mean, δ_Q is the ratio between the amplitude and the mean, and ϕ_Q is the phase (with respect to the same reference date) of the seasonal streamflow component.

Since we know the period T of interest, we need to quantify the mean, the amplitude and the phase of the periodic components. There are different methods to fit a sine curve of a certain period to data~~.~~, i.e. to extract Fourier modes. We have compared two sine curve fitting~~methods leading~~, namely multiple linear regression and a method that makes use of the cross-covariance of two sine waves. Both methods lead to virtually the same results~~(see Supplementfor more information on sine curve fitting)~~. A description and a comparison of the methods is shown in the Supplement. For the rest of the analysis, we will use results obtained by means of multiple linear regression (details on the fitting method can be found in the Supplement).

2.1.2 Calculation of seasonal signatures

Once we have extracted the seasonal components from our time series (precipitation minus potential evapotranspiration, streamflow), we can quantify how the outgoing sine wave Q_{sin} has been altered by the catchment by comparing it to the incoming sine wave F_{sin}. We define two metrics, the amplitude ratio and the phase shift, which together we call seasonal signatures. The amplitude ratio A is the ratio between the seasonal streamflow amplitude $\delta_Q \bar{Q}$ and the seasonal forcing amplitude $\delta_F \bar{F}$:

$$A = \frac{\delta_Q \bar{Q}}{\delta_F \bar{F}} \tag{3}$$

Given a closed long-term water balance, the amplitude ratio should theoretically always be between zero and unity, that is, the streamflow amplitude cannot be larger than the forcing amplitude. The phase shift ϕ is the difference between the phase of the seasonal streamflow component ϕ_Q and the phase of the seasonal forcing component ϕ_F:

$$\phi = \phi_Q - \phi_F \tag{4}$$

The phase shift should theoretically always be positive (the input should lead the output) and smaller than one year.

2.2 Linear reservoir theory

The derivations presented here all rely on the assumption of a linear time-invariant system (see e.g. Dooge, 1973, for an overview of linear t . This implies that forcings of different wavelengths are not influencing each other. The assumption of linearity is invalid for most real systems, yet it is still widely made as it can yield useful insights.

A linear reservoir is described by:

$$Q = \frac{S}{\tau} \tag{5}$$

where Q [mm d^{-1}] is the outflow from the reservoir, S is storage [mm] and τ [d] is a time constant describing how fast (slow) the reservoir responds. Conservation of mass requires:

$$\frac{dS}{dt} = Q_{\text{in}} - Q \tag{6}$$

where Q_{in} is the inflow to the reservoir.

2.2.1 Periodic forcing of a linear reservoir

If we approximate the seasonal input to a linear reservoir by a sine wave of period T (e.g. one year), we can combine Equations (1), (5) and (6) to obtain:

$$\frac{dQ_{\text{sin}}}{dt} = \frac{\bar{F}}{\tau}\left(1 + \delta_F \sin\left(\frac{2\pi}{T}t + \phi_F\right)\right) - \frac{Q_{\text{sin}}}{\tau} \tag{7}$$

We might neglect the (initial) phase if we choose a starting time t that is aligned with the seasonal forcing component ($\phi_F = 0$). It can be shown that the steady state response of a linear reservoir to a sinusoidal input signal is a damped and phase shifted version of the input signal (see Supplement for a more detailed derivation; or Eriksson, 1971; Peters et al., 2003):

$$Q_{\text{sin}}(t) = \bar{F}\left(1 + \delta_F A \sin\left(\frac{2\pi}{T}t + \phi\right)\right) \tag{8}$$

where A is the amplitude ratio and ϕ is the phase shift induced by a single linear reservoir.

$$A = \frac{1}{\sqrt{1 + (2\pi\frac{\tau}{T})^2}} \tag{9}$$

$$\phi = \arccos\left(\frac{1}{\sqrt{1 + (2\pi\frac{\tau}{T})^2}}\right) = \arccos(A) \tag{10}$$

We can rewrite Equation (8) as follows:

$$Q(t) = \bar{Q}\left(1 + \delta_Q \sin\left(\frac{2\pi}{T}t + \phi\right)\right) \tag{11}$$

In a mass conserving system in steady-state, the mean of the output should equal the mean of the input. If the means obtained from data are different, either the forcing term is inaccurate (e.g. due to differences between actual and potential evapotranspiration) or the streamflow term is inaccurate (e.g. due to other losses or gains). The product of input amplitude and amplitude ratio equals the output amplitude ($\delta_F \bar{F} A = \delta_Q \bar{Q}$).

From Equations (9) and (10), we can see that the amplitude ratio and the phase shift are given by A and $\arccos(A)$, respectively. Since A is fully defined by the ratio between τ and T, and T is usually known (e.g. one year), we can theoretically use A to determine the time constant τ of the reservoir. This requires the identification of both the seasonal components of the input and output signal of that period (see Section 2.1), and assumes the system to behave as a single linear reservoir. In theory, we could also apply the theory to other periods than one year, but for the reasons stated above we only investigate the annual period.

The amplitude ratio A and the phase shift $\arccos(A)$ can be plotted against each other for various values of τ as shown in Figure 1. This results in a characteristic curve which captures the response of all single linear reservoirs. Different time constants τ (as proportions of the period, here one year) lead to different positions on the curve. For very fast reservoirs, the phase shift is close to 0 days and the amplitude ratio is close to unity (that is, the signal is not attenuated at all). For very slow reservoirs, the signal is phase shifted up to 91 days and the amplitude ratio approaches 0. The maximum phase shift of about 91 days corresponds to a quarter of a period (90 degrees). Mathematically, this can be explained by Equation (10), as the arccosine of a quantity between 0 and unity (such as A) ranges between 0 and 90 degrees.

Note the similarity of Figure 1 to Figure 3c in Kirchner (2016), which shows the relationship between phase shift and amplitude ratio for gamma-distributed catchment transit time distributions. An exponential transit time distribution (a special case of the gamma distribution) corresponds to a linear reservoir describing the velocity of particles. Similarly, a linear reservoir describing the impulse response (the linear reservoir from Equation (5)), i.e. the celerity of the incoming wave of hydraulic potential, corresponds to an exponential response time distribution or an exponential unit hydrograph (cf. Eriksson, 1971; Dooge, 1973).

2.2.2 Combinations of linear reservoirs

Linear systems (Dooge, 1973) have the advantage that it is relatively straightforward to add more components, that is, reservoirs. It is quite common to have serial and/or parallel combinations in rainfall-runoff models. In theory, we can find analytical solutions for the amplitude ratio and phase shift for all combinations of linear reservoirs ~~(cf. to the transfer function approach of Young, 199~~ (cf. to the transfer function approach of Young, 1998, who identifies combinations of reservoirs that fit the data best in an inductive way) . There are two basic arrangements, a serial arrangement of reservoirs and a parallel arrangement of reservoirs.

2.2.3 Linear reservoirs in series

Linear reservoirs in series can be conceptualised as follows. Every outflow is the inflow to the next reservoir. Hence, if the i-th reservoir has a time constant τ_i, the amplitude ratios A_i are multiplied and the phase shifts ϕ_i are added (see Supplement for a

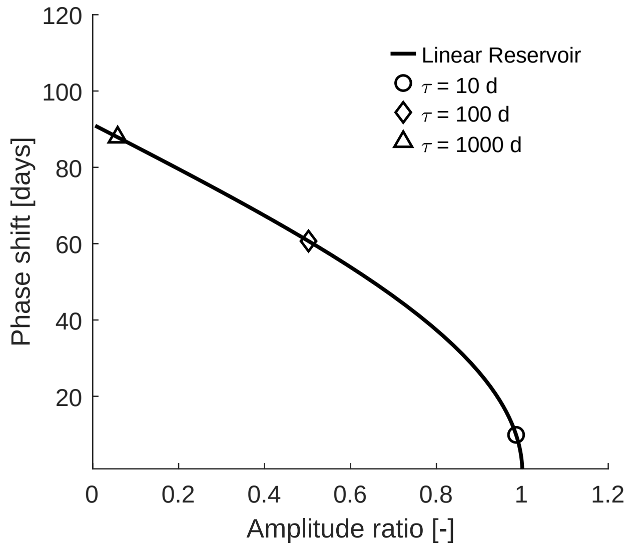

Figure 1. Amplitude ratio against phase shift for a single linear reservoir for varying time constants τ. Three example time constants are indicated by the symbols.

more detailed derivation):

$$A_{\text{tot}} = \prod_{i=1}^{n} A_i \tag{12}$$

$$\phi_{\text{tot}} = \sum_{i=1}^{n} \phi_i = \sum_{i=1}^{n} \arccos(A_i) \tag{13}$$

Figure 2 shows the amplitude ratio plotted against the phase shift similar to Figure 1, but now with two linear reservoirs in series. The different lines are examples with fixed time constants of the first reservoir. They all start from the black line (from the points marked by the symbols in Figure 1), the characteristic curve for a single linear reservoir, which is the lower limit. Then, as the time constant of the second reservoirs increases, the lines "move" left and upwards, which corresponds to a decrease in amplitude ratio and an increase in phase shift. For example, the red line ($\tau_1 = 10$ d) starts out with a phase shift of about 10 days, and ends at a phase shift of about 101 days, which is an increase of about 91 days, the maximum phase shift of the second reservoir. The lines cross each other as we allow τ_2 to be larger than τ_1. This implies that sometimes a faster reservoir is followed by a slower one, and sometimes a slower reservoir is followed by a faster one. The grey shaded area contains all possible combinations for two reservoirs in series. The lower limit is a single linear reservoir. The upper limit corresponds to two reservoirs with the same time constant (a two-reservoir Nash cascade), which equals a gamma distribution with a shape parameter equal to ~~2.~~ 2 (Nash, 1957).

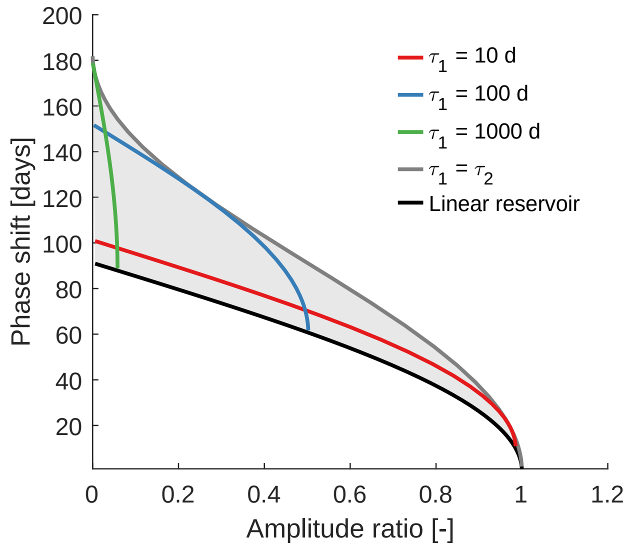

Figure 2. Amplitude ratio against phase shift for two linear reservoirs in series. Each line corresponds to a fixed time constant for the first reservoir (τ_1), while the time constant of the second reservoir varies (1 d $\leq \tau_2 \leq$ 10000 d; it is increasing from right to left). The black line indicates a single linear reservoir (the lower boundary). The grey line indicates the upper boundary where $\tau_1 = \tau_2$. The shaded area contains all possible combinations of amplitude ratio and phase shift for two linear reservoirs in series.

2.2.4 Linear reservoirs in parallel

Linear reservoirs in parallel result in a "mixture" of the outflows from each reservoir. The resulting flow is a combination of sine waves of the same period, weighted by the fraction p_i going into each reservoir. For the sake of simplicity, we only consider two reservoirs in parallel. We denote the fraction going into the second reservoir by p, and therefore the fraction going into the first reservoir by $1-p$. Thinking of the second reservoir as the slow one, p might be compared to the idea of the baseflow index (BFI), the volumetric ratio between baseflow and total streamflow (Institute of Hydrology, 1980). For two reservoirs in parallel we get (see Supplement for a more detailed derivation):

$$A_{\text{tot}} = \sqrt{\left[(1-p)A_1\cos\phi_1 + pA_2\cos\phi_2\right]^2 + \left[(1-p)A_1\sin\phi_1 + pA_2\sin\phi_2\right]^2} \quad (14)$$

$$\phi_{\text{tot}} = \arctan\left(\frac{(1-p)A_1\sin\phi_1 + pA_2\sin\phi_2}{(1-p)A_1\cos\phi_1 + pA_2\cos\phi_2}\right) \quad (15)$$

Figure 3 shows the amplitude ratio plotted against the phase shift similar to Figure 1, but now with two linear reservoirs in parallel. We show multiple plots to highlight the three degrees of freedom: the two reservoir time constants and the fraction going into each reservoir. The latter is highlighted in Figure 3d, but also visible in Figures 3a-c. ~~We can see that there is an "extreme case" where the output signal effectively comes from the first reservoir only as the second reservoir is so slow that it hardly contributes to the resulting sine wave. As only $1-p$ of the total input goes into the first reservoir, the amplitude will~~

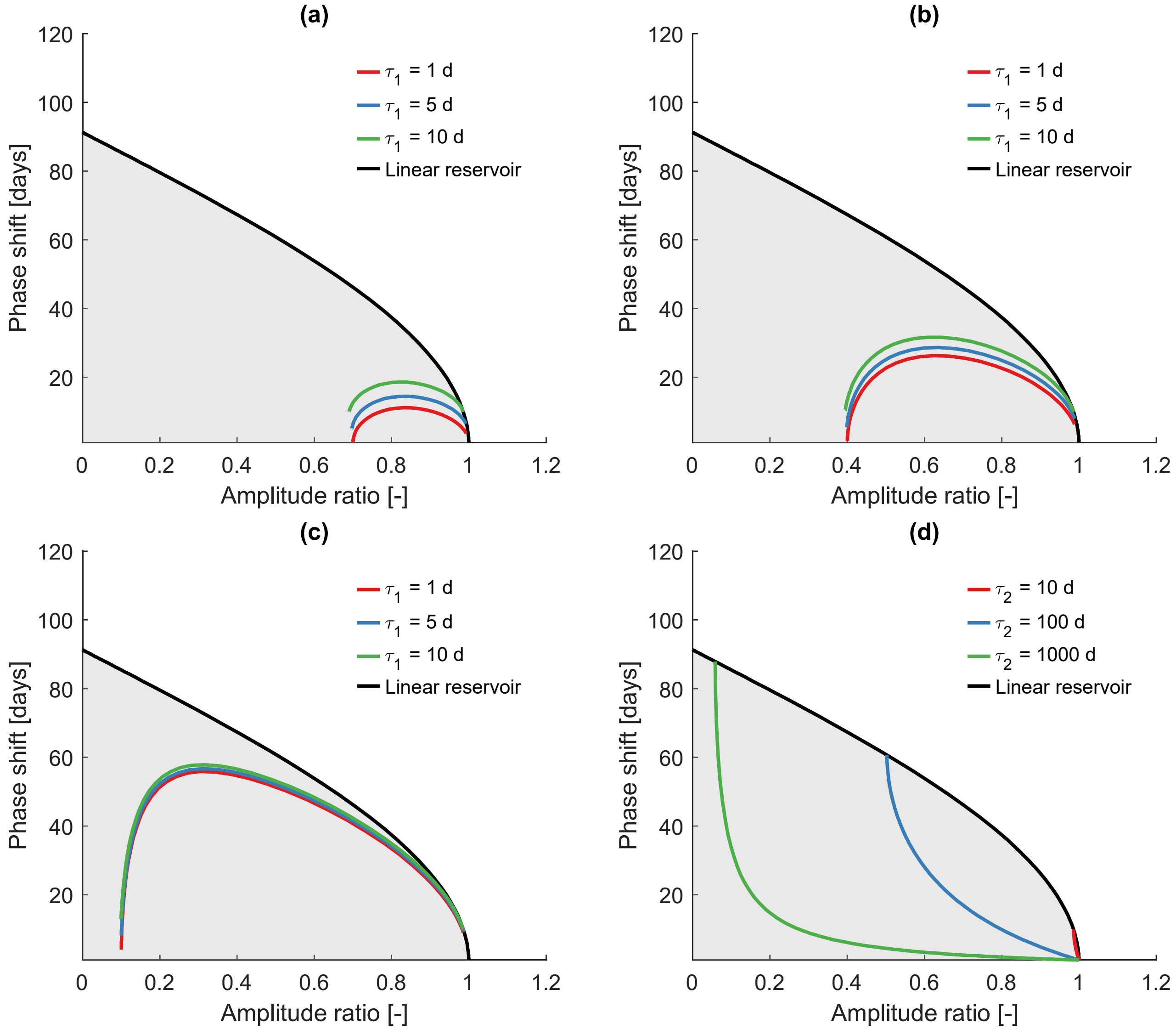

Figure 3. Amplitude ratio against phase shift for two linear reservoirs in parallel. **(a)** Each line has a fixed time constant for the first reservoir (τ_1), while the time constant of the second reservoir varies (10 d $\leq \tau_2 \leq$ 10000 d; it is increasing from right to left). The fraction p going into the second reservoir is 0.3. **(b)** Same as **(a)** with $p = 0.6$. **(c)** Same as **(a)** with $p = 0.9$. **(d)** Each line has a fixed time constant for the first reservoir ($\tau_1 = 1$ d), and for the second reservoir (τ_2). The fraction p going into the second reservoir is varied (it is increasing from right to left). The shaded area contains all the possible combinations of amplitude ratio and phase shift for two linear reservoirs in parallel.

be about $1-p$ times the input amplitude. The grey shaded area contains all the possible combinations for two reservoirs in parallel. The upper limit is a single linear reservoir. The lower limit is effectively given by the x- and the y-axis.

As an example, Figure 3a can be explained as follows: τ_1 is always 1 d, the fraction p going into the second reservoir is 0.3, and τ_2 starts with a value of 10 d and then increases. So at first, both reservoirs are rather fast and we get a high amplitude ratio and a small phase shift for the combined sine wave (see Equations (14) and (15)). Then, the second reservoir gets slower, leading to a decrease in amplitude ratio and an increase in phase shift. As the second reservoirs gets slower and slower, it will contribute less and less to the overall sine wave. For very high values of τ_2 (e.g. 10000 d), the sine wave coming out of the second reservoir is almost a straight line, so the combined sine wave is primarily consisting of the sine wave coming out of the first reservoir. Since only a fraction of $1-p=0.7$ of the total input has gone into the first reservoir, the amplitude of the combined sine wave is approximately 0.7 times the input amplitude with a very small phase shift, as the first reservoir hardly attenuates the signal.

2.3 Seasonal signatures as a diagnostic tool for evaluating hydrological models

We use two conceptual rainfall-runoff models and we use test whether the seasonal signatures can be used as a diagnostic tool to assess their model performance (Gupta et al., 2008). In particular, we test whether the models are generally capable of reproducing the range of observed seasonal signatures (cf. Vogel and Sankarasubramanian, 2003). We signatures without calibrating them to streamflow data (cf. Vogel and Sankarasubramanian, 2003). This modelling experiment is intended to test whether the proposed signatures have the potential to be a useful additional source of information in model building and evaluation. We do not intend (or suggest) that the presented evaluation approach can replace existing model evaluation methods. We limit the analysis to two models and 40 catchments to keep the computational demand manageable. The catchments are We also limit the model evaluation to catchments in the UK, as the seasonal signatures are unreliable in arid catchments (see Section 5). The subset of catchments is described in Section 3 and in the Supplement.

The first model is the IHACRES model. It is conceptually relatively similar to the considerations in Section 2. It has a soil moisture store (non-linear deficit store), and two parallel linear stores for fast flow and for slow flow (Croke and Jakeman, 2004). It has been used in many modelling studies in Australia (Post and Jakeman, 1999) and also in the UK (Sefton and Howarth, 1998). The second model is the GR4J model. It also has a parallel flow structure, but the internal parametrisation is different. It contains more non-linearities and it has fixed internal parameters. Additionally, it has a groundwater exchange parameter aimed at representing inter-catchment groundwater flows. It has been used in many modelling studies in France (Perrin et al., 2003), in the UK (Smith et al., 2019; Harrigan et al., 2018b) and in the US (Oudin et al., 2018). We use the implementations of the two models in the MARRMoT toolbox v1.2 (Knoben et al., 2019a), a Matlab toolbox containing many hydrological models aimed at model comparison studies. The pure delay function in the MARRMoT implementation of IHACRES is set to 0, making it (conceptually, not necessarily numerically) equal to the version used by Croke and Jakeman (2004). In our model evaluationmodelling experiment, IHACRES has therefore 6 parameters, and GR4J has 4 parameters. Detailed information on the parameter ranges and on model warm-up periods can be found in the Supplement.

To test which ranges of seasonal signatures the two models can reproduce, we run a Monte-Carlo sampling experiment. We sample parameter sets for both models using Latin Hypercube samplingand we test , an efficient sampling method (Cheng and Druzdzel, 2000) that assumes uniform prior parameter distributions. With the parameter sets obtained, we run both models for each of the 40 catchments, i.e. we use the same parameter sets for each catchment. To test for robustness, we sample an increasing number of parameter sets to see whether the results are robust (see Supplement). We mostly use the recommended parameter ranges from the MARRMoT toolbox, which are intended to be wide ranges. We use a narrower range for the fast flow routing delay (1 to 5 days), to ensure that it is indeed "fast flow"(20000 parameter sets are considered sufficient, see Supplement for more information). We then use the modelled streamflow time series to calculate three hydrological signatures per parameter set: the two seasonal signatures presented here and the baseflow index (BFI). These are The resulting modelled signatures are compared to observed signatures and explored in a rather general way, as we want to examine what the models can do without actually calibrating them to certain catchments (cf. Vogel and Sankarasubramanian, 2003) . streamflow data (cf. Vogel and Sankarasubramanian, 2003). That is, we are not interested in finding the "best" parameter set, but in whether a certain model (given certain parameter ranges) is generally capable of reproducing the signatures we observe.

3 Data

3.1 Data sources

We use catchment data from Great Britain and the United States. The data for the UK is are obtained from different sources. Daily streamflow data, catchment characteristics and catchment boundaries are obtained from the NRFA (National River Flow Archive, 2019), precipitation data from CEH-GEAR (Tanguy et al., 2016), and potential evapotranspiration data from CHESS-PE (Robinson et al., 2016). For the model evaluation we select catchments that are part of the UK Benchmark Network (Harrigan et al., 2018a), which describes catchments in the UK that are near-natural. The subset of catchments is chosen to be representative of the UK, details are shown in the Supplement. The data for the US is are obtained from the CAMELS dataset (Newman et al., 2015; Addor et al., 2017). CAMELS includes daily precipitation, potential evapotranspiration (we use Daymet forcing data) and streamflow data as well as a wide range of catchment attributes for 671 catchments in the contiguous US. We trim the daily data to contain only full water years (starting 1 October) and we analyse data from 1989 to 2009. We also remove catchments with missing records during that time period. While we need to pick a start date for the analysis, this date does not influence the results (e.g. using 1 January as starting date would result in the same phase shift).

3.2 Hydrological signatures and catchment attributes

We calculate different hydrological signatures and we use different catchment attributes, all summarised in Table 1. The climate indices from Knoben et al. (2018) are based on monthly averages and they need to be interpreted as follows. A moisture index I_m of 1 indicates the most humid (energy-limited) catchments, a moisture index of -1 indicates the most arid (water-limited) catchments. A moisture index seasonality $I_{m,r}$ of 0 indicates catchments where the climate stays constant throughout the year,

Table 1. Hydrological signatures and catchment attributes used in this study.

Name	Description	Unit	Range	Reference
Hydrological signatures				
BFI	Baseflow index	[-]	$[0,1]$	Institute of Hydrology (1980)
A	Amplitude ratio	[-]	$[0,1]$[1]	Equation (3)
ϕ	Phase shift	[d]	$[0,365]$[2]	Equation (4)
Catchment attributes				
~~E_p/P Aridity index - $[0,\infty]$~~ I_m	Moisture index	[-]	$[-1,1]$	Knoben et al. (2018)
$I_{m,r}$	Moisture index seasonality	[-]	$[0,2]$	Knoben et al. (2018)
f_s	Snow fraction	[-]	$[0,1]$	Knoben et al. (2018)
PROPWET	Catchment wetness index	[-]	$[0,1]$	National River Flow Archive (2019)
% fractured aquifer	Fraction of highly productive fractured aquifer	[%]	$[0,100]$	National River Flow Archive (2019)
% carbonate rock	Fraction of carbonate sedimentary rock	[%]	$[0,100]$	Addor et al. (2017)

[1] Should in theory be smaller than unity. [2] Should theoretically always be positive and in practice be smaller than one year. Further discussions on the possible ranges of the seasonal signatures can be found in the text.

a moisture index seasonality of 2 indicates catchments where the climate switches between fully arid and fully humid within the year.

4 Results

4.1 Extracting seasonal components from time series

First, we extract seasonal components from $P - E_p$ (forcing) and Q (streamflow) for all catchments. The resulting sine wave parameters are then used to calculate the amplitude ratios (Equation (3)) and phase shifts (Equation (4)), respectively. Figure 4 shows $P - E_p$ and Q for two catchments alongside their seasonal (sinusoidal) components. Both catchments experience a similar forcing, but their response is very different. The Ericht at Craighall, a rather responsive catchment, shows a seasonal streamflow component that is very similar to the seasonal forcing component. In contrast, the East Avon at Upavon, a groundwater-dominated catchment, shows a strongly attenuated seasonal streamflow component. For our seasonal signatures this would mean (a) that the responsive catchment has a high amplitude ratio, i.e. the streamflow amplitude is almost as large as the forcing amplitude, while the groundwater-dominated catchment has a low amplitude ratio. And (b) that the responsive catchment has a small phase shift, i.e. it responds quickly to the (seasonal) forcing, while the groundwater-dominated catchment has a large phase shift.

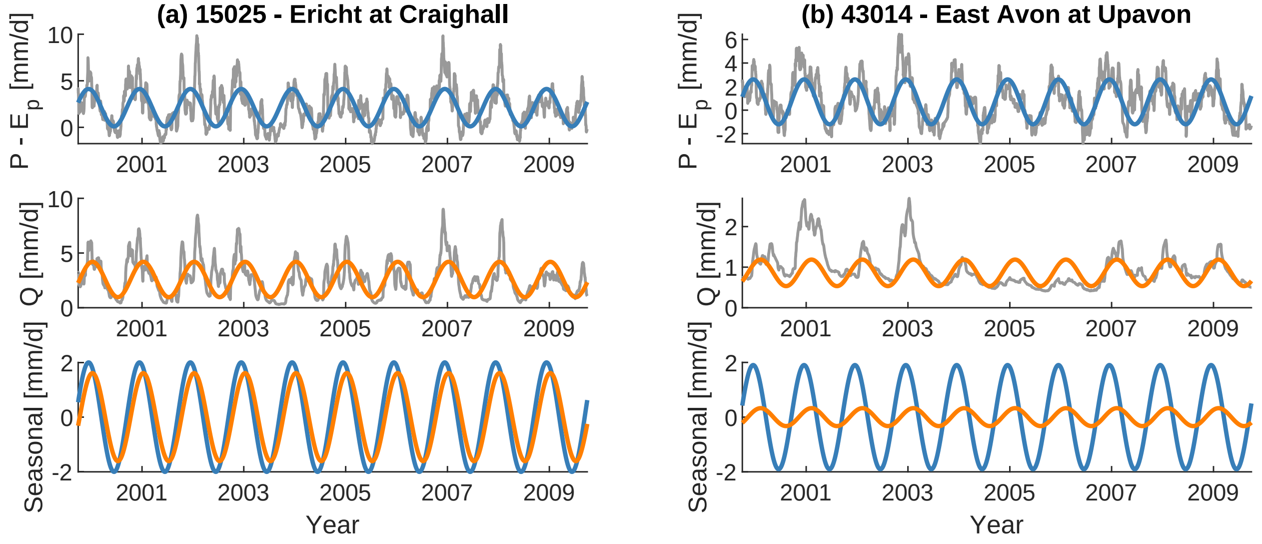

Figure 4. Climate input (P - E_p; blue) and catchment output (Q; orange) for two example catchments in the UK, and their respective seasonal components. The time series are smoothed using a 30-day moving mean. The Ericht is a rather responsive catchment (BFI = 0.47), while the East Avon has a large baseflow component (BFI = 0.89). Note that for the bottom plots ("Seasonal") the mean values of the sine curves are set to zero.

4.2 Seasonal signatures of observed catchment data

To visualise the seasonal signatures, we plot the amplitude ratios and phase shifts in a similar way as in Figures 1, 2, and 3. This is shown in Figure 5a for all UK catchments. These include catchments with human influences, such as groundwater abstractions, man-made reservoirs or water transfers. The overall pattern in Figure 5a is very similar to the pattern using benchmark catchments alone (grey dots). We therefore use all of the catchments, noting that a few catchments might be unsuitable for individual analyses.

Figure 5a shows that most of the catchments fall below the solid grey line, the line which indicates the type of response that could be simulated by a single linear reservoir (see Figure 5b). The area below the solid line can be simulated by two reservoirs in parallel. This would be the most parsimonious way to reproduce the observed behaviour if we decide to construct our model using linear reservoirs only. A few catchments plot above the solid line. For these catchments, the most parsimonious way to repdroduce reproduce the pair of observed amplitude ratio and phase shift would therefore be two reservoirs in series. Very few catchments have an amplitude ratio larger than unity. While this could be caused by various errors in the data, it is likely due to erroneous catchment areas and/or the presence of inter-catchment groundwater flows or water transfers. If a catchment receives more net rainfall than the surface catchment area suggests (runoff ratio > 1), the amplitude in the output signal (streamflow) can be larger than the amplitude in the (erroneous) input signal.

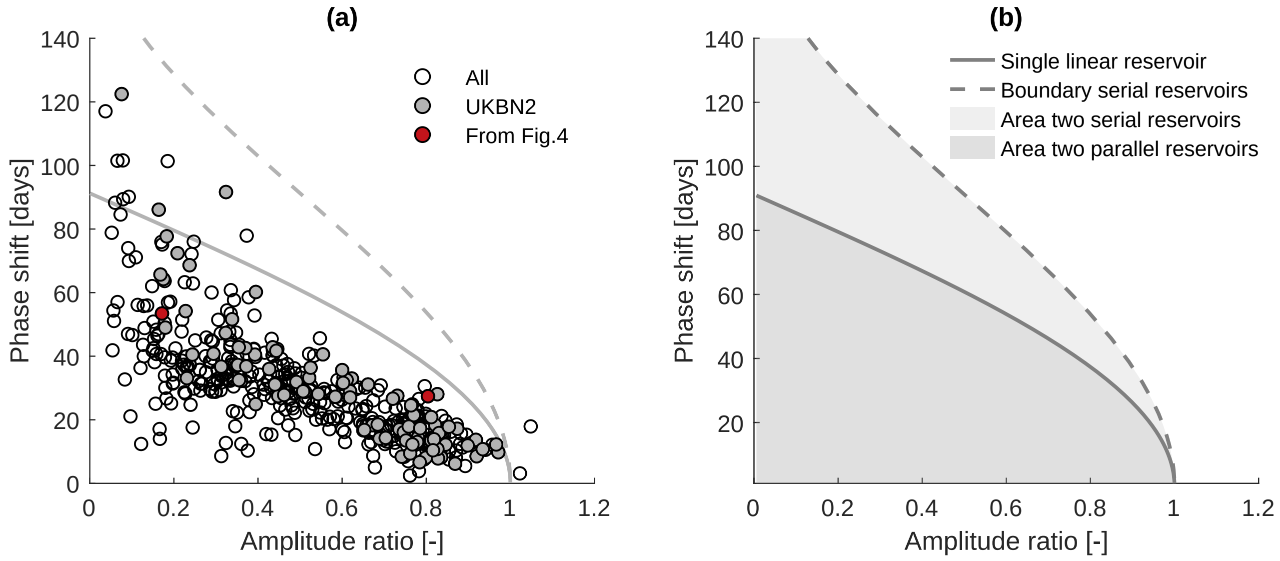

Figure 5. (a) Amplitude ratio against phase shift for UK catchments. Grey dots indicate benchmark catchments, red dots indicate the two catchments shown in Figure 4. Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Note that both axes are limited (two catchments are not shown). **(b)** Theoretical areas and limits for single linear reservoir, two reservoirs in series, and two reservoirs in parallel.

Table 2. Pearson and Spearman correlation coefficients between seasonal signatures and catchment attributes for UK catchments.

Pearson	Amplitude ratio	Phase shift	I_m	PROPWET	BFI	% fractured aquifer
Amplitude ratio	1.00	-0.60	0.80	0.74	-0.58	-0.49
Phase shift	-0.60	1.00	-0.49	-0.50	0.66	0.58
Spearman	**Amplitude ratio**	**Phase shift**	I_m	**PROPWET**	**BFI**	**% fractured aquifer**
Amplitude ratio	1.00	-0.80	0.82	0.78	-0.58	-0.51
Phase shift	-0.80	1.00	-0.76	-0.75	0.77	0.60

4.2.1 ~~Relationship between seasonal signatures and catchment attributes – UK~~

4.3 Relationship between seasonal signatures and catchment attributes – UK

Figure 6 shows pairs of amplitude ratios and phase shifts, coloured according to different hydrological signatures and catchment attributes, respectively ~~.~~ (explained in Table 1). Corresponding correlation coefficients can be found in Table 2. Figure 6a shows a clear pattern between the moisture index and the seasonal signatures. Generally, the less humid the catchments~~(indicated by a more intense red colour)~~, the lower the amplitude ratio and the larger the phase shift. In other words, drier catchments attenuate

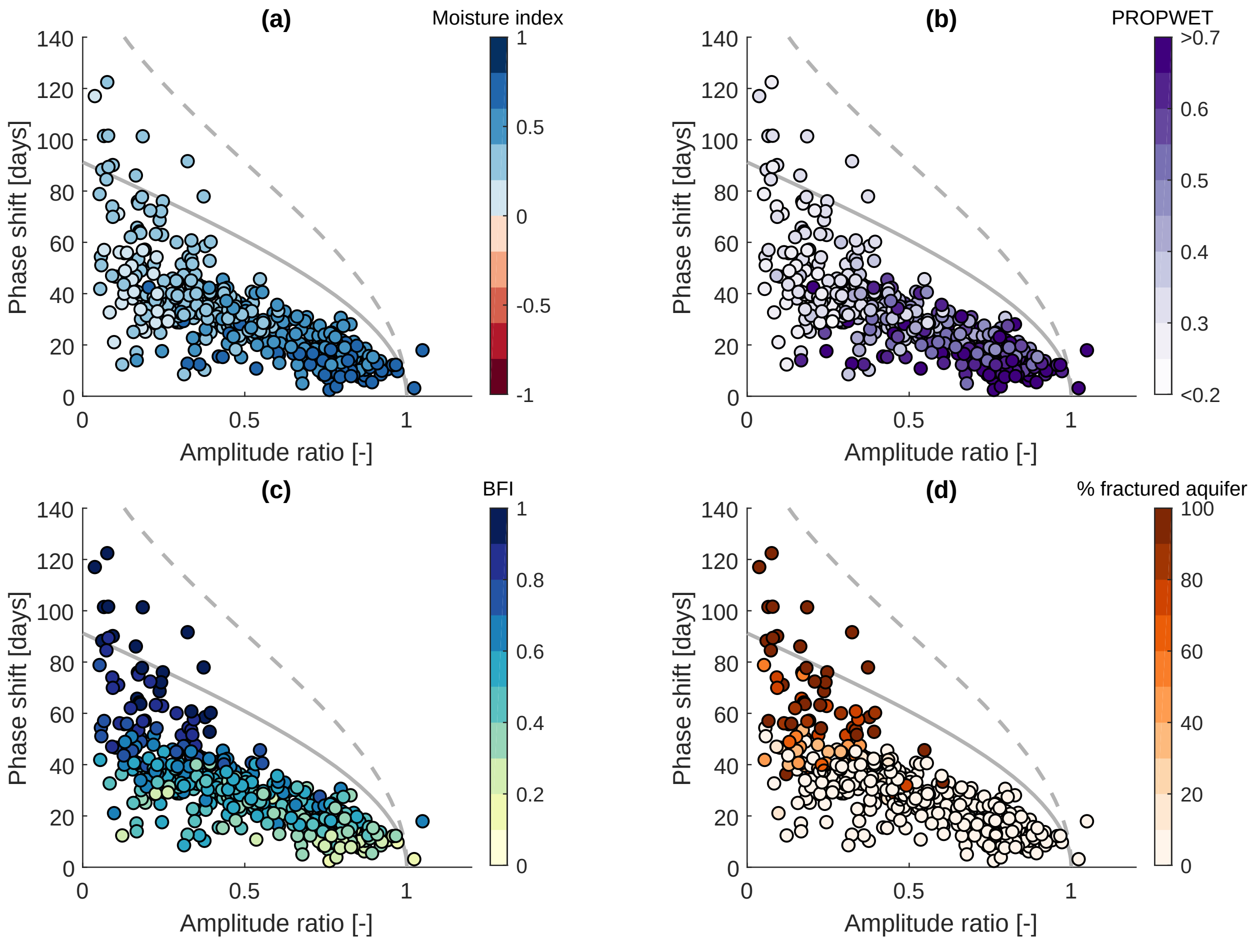

Figure 6. Amplitude ratio against phase shift for UK catchments. Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Colours indicate **(a)** the moisture index, **(b)** the catchment wetness index, **(c)** the baseflow index, and **(d)** the fraction of highly productive fractured aquifer. Note that both axes are limited (two catchments are not shown).

the incoming forcing signal more strongly. This might partly be because we use potential evapotranspiration as our forcing. Lower actual evapotranspiration than potential evapotranspiration leads to a decreased input amplitude and thus to a higher amplitude ratio. ~~Very~~ Most of the very humid catchments plot close together and the relationship between amplitude ratio and phase shift seems to be almost linear. Less humid catchments (note that in the UK none of the catchments are actually water-limited ~~on an~~ at the annual scale) show a larger spread, especially regarding the phase shift. Figure 6b shows a very similar pattern between the catchment wetness index and the seasonal signatures. Wetter catchments exhibit higher amplitude ratios and lower phase shifts, and vice versa. The catchment wetness index is strongly correlated with the moisture index (Spearman rank correlation of 0.94). Figure 6c shows a clear pattern between the baseflow index and the seasonal signatures. In contrast

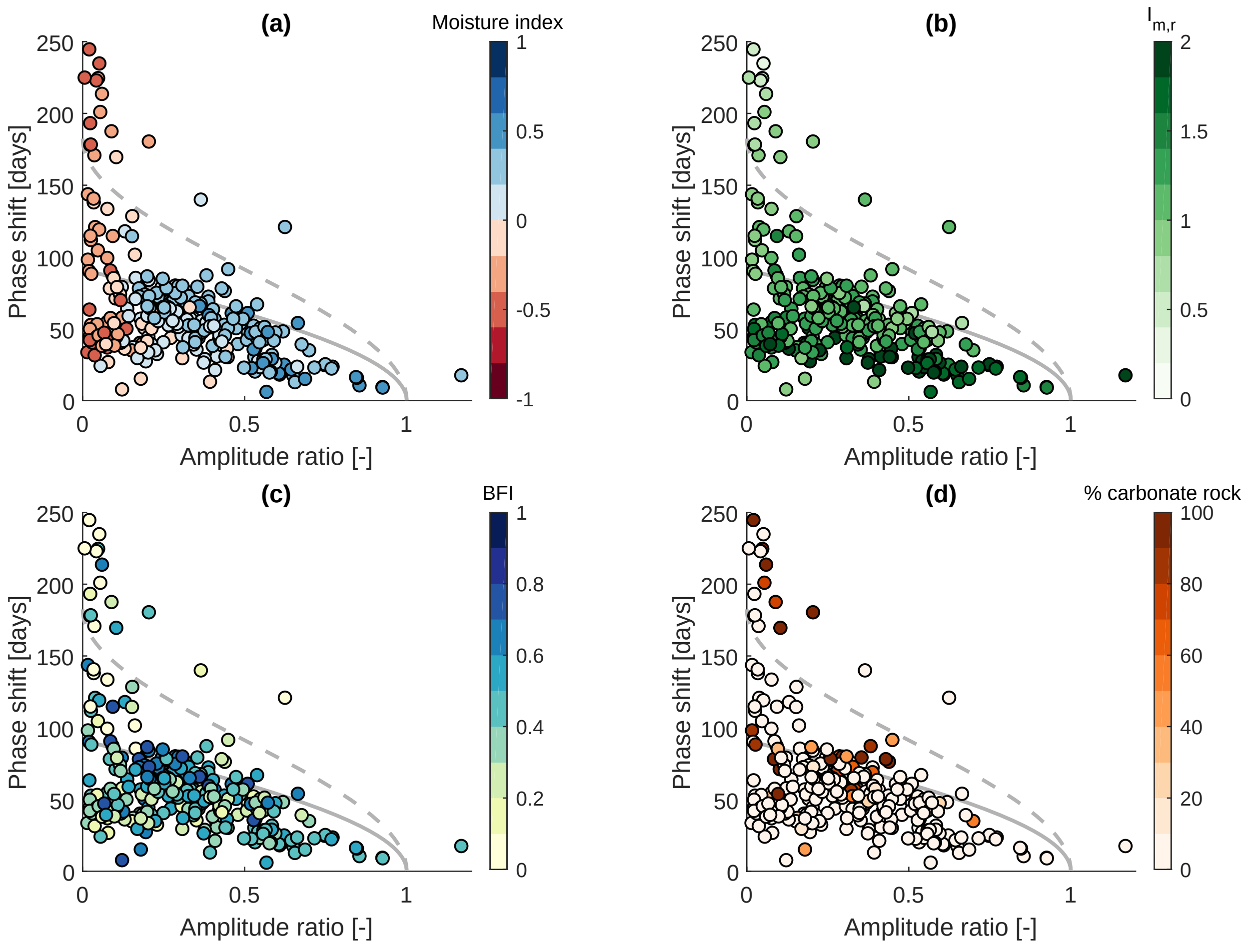

Figure 7. Amplitude ratio against phase shift for ~~UK~~ CAMELS catchments. Catchments with snow fraction ($f_s > 0.001$) are removed from the analysis. Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Colours indicate **(a)** the moisture index, **(b)** the ~~catchment wetness~~ moisture index seasonality, **(c)** the baseflow index, and **(d)** the fraction of ~~highly productive fractured aquifer~~carbonate sedimentary rock. Note that both axes are limited (~~two~~ 12 catchments are not shown) and that the range of the phase shift-axis is different from Figure 6.

to the moisture index, where the stratification follows mostly the x-axis (amplitude ratio), the stratification follows mostly the y-axis (phase shift). Catchments with high BFIs exhibit low amplitude ratios and large phase shifts, and vice versa. Finally, in Figure 6d we can see that catchments underlain by highly productive fractured aquifers exhibit (with a few exceptions) low amplitude ratios and large phase shifts.

Table 3. Pearson and Spearman correlation coefficients between seasonal signatures and catchment attributes for CAMELS catchments.

Pearson	Amplitude ratio	Phase shift	I_m	$I_{m,r}$	BFI	% carbonate rock
Amplitude ratio	1.00	-0.26	0.75	0.31	0.06	-0.16
Phase shift	-0.26	1.00	-0.39	-0.51	-0.14	0.26
Spearman	**Amplitude ratio**	**Phase shift**	I_m	$I_{m,r}$	**BFI**	**% carbonate rock**
Amplitude ratio	1.00	-0.46	0.78	0.23	0.04	-0.15
Phase shift	-0.46	1.00	-0.32	-0.64	0.06	0.36

4.4 Relationship between seasonal signatures and catchment attributes – US

4.4.1 Relationship between seasonal signatures and catchment attributes – US

Figure 7 shows pairs of amplitude ratios and phase shifts for the US, coloured according to different hydrological signatures and catchment attributes, respectively . (explained in Table 1). Corresponding correlation coefficients can be found in Table 3. Catchments with significant snow fraction ($f_s > 0.001$) are removed, as snow presents another hydrological process which is not the focus of this study. Generally, snow adds another storage process, and this is reflected in large phase shifts observed in snowy catchments (see Supplement for more information). The non-snowy catchments in the US show a similar trend to the catchments in the UK. Yet generally, the amplitude ratios are lower and the phase shifts larger compared to the UK (note that the y-axes in Figure 7 differ in their range from the y-axes in Figure 6). Humid catchments tend to have higher amplitude ratios and smaller phase shifts (Figure 7a). Climate seasonality, indicated by the moisture index seasonality (see Figure 7b), also influences the seasonal signatures. Catchments with a larger moisture index seasonality, i.e. a more variable monthly moisture index over the year, tend to have smaller phase shifts. The BFI (Figure 7c) does not show such a clear pattern as for the UK catchments (Figure 6c). Similarly, subsurface properties such as the fraction of carbonate sedimentary rock (Figure 7d; and other attributes not shown here) only show a weak relationship with the seasonal signatures. Catchments with larger fractions of carbonate sedimentary rocks tend to have lower amplitude ratios and larger phase shifts. The overall pattern, however, is rather scattered. Contrary to the UK, some of the catchments in the US plot outside the area that can be modelled by either two reservoirs in series or in parallel and some catchments have phase shifts larger than 182 days, the approximate limit for two reservoirs in series. These catchments are very arid and the low moisture seasonality index indicates that most the precipitation in these catchments falls when potential evapotranspiration is highest, i.e. in summer.

Amplitude ratio against phase shift for CAMELS catchments. Catchments with snow fraction ($f_s > 0.001$) are removed from the analysis. Grey solid line indicates a single linear reservoir, grey dashed line indicates the outer envelope for two reservoirs in parallel. Colours indicate **(a)** the moisture index, **(b)** the moisture index seasonality, **(c)** the baseflow index, and **(d)** the fraction of carbonate sedimentary rock. Note that both axes are limited (13 catchments are not shown) and that the range of the phase shift-axis is different from Figure 6.

4.5 Seasonal signatures as a diagnostic tool for evaluating hydrological models

In a similar fashion as for the observed catchment data, we now investigate the model runs using IHACRES and GR4J. Figure 8 shows the resulting amplitude ratios and phase shifts for all model runs, that is for 20000 parameter sets using data from a subset of 40 catchments in the UK. These plots show which combinations of the seasonal signatures (and BFI) can be obtained with each model, given the forcing of 40 different catchments covering most of the hydro-climatic variability of the UK, and given the parameter ranges chosen. They hence show the "signature space" of a model in the dimensions given by amplitude ratio and phase shift (and BFI).

IHACRES (Figure 8a) shows a pattern that covers the area that can be reproduced modelled by two reservoirs in parallel, and most a large fraction of the area that can be reproduced modelled by two reservoirs in series (see Figures 2 and 3). IHACRES sometimes yields The BFI spans the whole range from 0 to 1. IHACRES can reproduce the observed amplitude ratios and phase shifts, although one catchment sits just at the boundary of the point cloud. GR4J (Figure 8b) covers a different signature space. The phase shift never exceeds 110 days, the amplitude ratio often exceeds unity, and the BFI tends to be high. GR4J can reproduce most of the observed amplitude ratios and phase shifts, except for catchments with very large phase shifts. Furthermore, it struggles to simultaneously reproduce the observed phase shifts and BFIs.

Both models sometimes yield phase shifts that are close to one year (not shown here), which are effectively negative phase shifts. Negative implies that the periodic component of Q leads the periodic component of $P - E_p$. This can happen if actual evapotranspiration E_a differs considerably from potential evapotranspiration E_p, and hence most of the input seasonality stems from P (and not E_p). This can be observed in a few catchments in the US (not shown here). It is only observed once in the UK (in a catchment with a man-made reservoir, not shown here), and therefore we do not investigate these model runs further. GR4J (Figure 8b) covers a rather different signature space. The phase shift never exceeds 60 days and the amplitude ratio often exceeds unity. Amplitude ratios larger than unity are due to the groundwater exchange parameter, which allows the model to import water in addition to incoming P.

Figure 9 shows distributions ("one-dimensional signatures spaces") for three hydrological signatures from the two hydrological models. The plots are created as follows. For every catchment and for each of the signatures, we summarise the resulting signature values by a probability density function (PDF). For example, running IHACRES with 20000 parameter sets for one catchment leads to 20000 values for the BFI. We use these 20000 values to fit a PDF via kernel density estimation, resulting in one curve per signature per catchment (a proxy for the histogram). Plotting curves for all catchments on top of each other shows the range and likelihood of a signature value being produced by a certain model structure (using a certain set of parameter sets). How much the curves differ from each other gives an indication of how a change in forcing changes the resulting signature distribution. The forcing is indicated by the colour of a curve, which corresponds to a certain aridity index. These plots thus for the 40 catchments investigated here. These plots tell us which signature values a model structure tends to produce (given a certain sampling scheme)and how (much) a signature varies with varying forcing. Note that this evaluation is independent of observed streamflow, it just shows the theoretically possible hydrological response that a model can simulate. By comparing observed hydrological signatures to this theoretical signature space, we might be able to tell whether a certain model structure

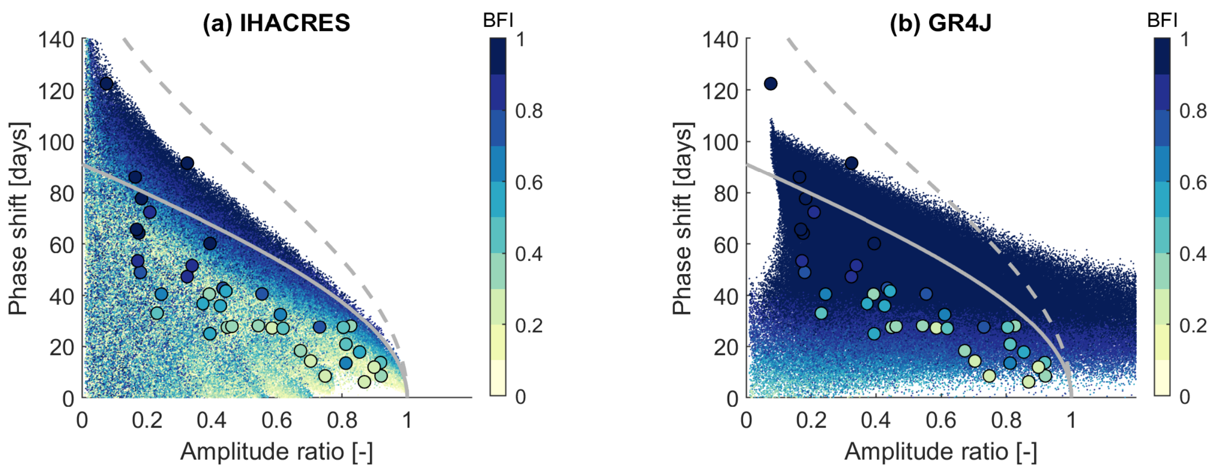

Figure 8. Amplitude ratio against phase shift for 40 catchments in the UK using 20000 parameter sets each for **(a)** IHACRES and **(b)** GR4J. The large dots show the observed signatures of the 40 catchments used for the modelling experiment. Colours indicate the BFI. Note that both axes are limited.

is generally able to simulate a certain set of catchments. , the ranges of signatures a model can reproduce (given the parameter ranges chosen), and how (much) a signature varies with varying forcing.

PDFs of different hydrological signatures ("signature space") resulting from the parameter sampling experiment for 40 catchments. **(a)** BFI IHACRES, **(b)** BFI GR4J, **(c)** amplitude ratio IHACRES, **(d)** amplitude ratio GR4J, **(e)** phase shift IHACRES, **(f)** phase shift GR4J. The colours indicate the aridity index of the catchments. Note that the x-axes are limited and that the y-axes are arbitrary and not to scale.

For the BFI (Figures 9a,b) we can see that IHACRES covers the whole possible space (0 to 1)relatively evenly, and it does not vary substantially with varying forcing. This means that the modelled BFI is essentially a function of the model (parameters) alone. GR4J tends to produce very high BFIs for almost every parameter set. BFI values below 0.5 are possible with GR4J, but rather rare (or unlikely). The forcing has a secondary influence on the BFI, with more humid catchments leading to lower BFIs. Figures 9 c-f are in line with Figure 8Figure 9 displays similar information as Figure 8, yet without considering interactions between the three signatures. IHACRES can produce amplitude ratios from 0 to 1 and phase shifts up to about 182 days , (which is the limit for two reservoirs in series.) and larger. GR4J can produce amplitude ratios that clearly exceed one, and cannot model phase shifts larger than 60 days – at least 110 days (given the parameter ranges chosen). For both models, more arid forcing leads to lower amplitude ratios and larger phase shifts, and vice versa. For the BFI (Figures 9c,f) we can see that IHACRES covers the whole possible space (0 to 1) relatively evenly. GR4J tends to produce very high BFIs for almost every parameter set. BFIs smaller than 0.5 are possible with GR4J, but rather rare (or unlikely).

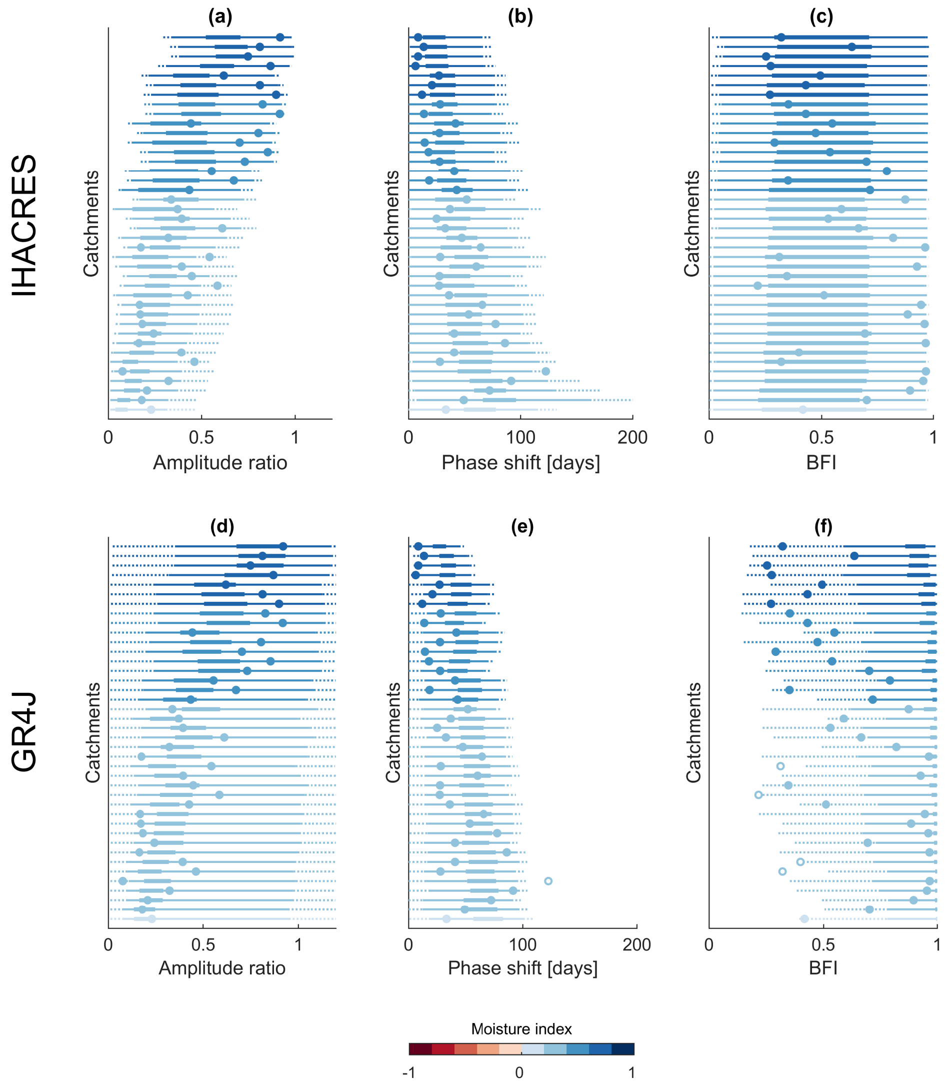

Figure 9. Distributions of different hydrological signatures resulting from the modelling experiment. Each line stands for one of the 40 catchments and the colours indicate the corresponding moisture index. The distributions of the modelled signatures are indicated by box-whisker-type plots. The thick line spans from the 25th to the 75th percentile. The thin line spans from the 1st (75th) to the 25th (99th) percentile. The dotted line indicates values below (above) the 1st (99th) percentile. The circles indicate the observed signature values, while filled circles indicate that the observed signature is inside the modelled signature space and vice versa. **(a)**, **(b)**, and **(c)** show results for IHACRES. **(d)**, **(e)**, and **(f)** show results for GR4J. Model runs with amplitude ratios lower than 0.01, amplitude ratios larger than 1.2, or phase shifts larger than 200 days have been removed.

5 Discussion

5.1 Representation of seasonal components by sine waves and limitations of the approach

A sine wave is a simple way of describing the seasonality of a signal. The results suggest that for most of the catchments investigated here, this approach is reasonable and efficient. Figure 4 shows that the average seasonal pattern is captured by the fitted sine waves. ~~Perhaps more importantly, the patterns visible in Figures 5, 6, and 7 indicate that the extracted seasonal components are meaningful and not just noise. Of course, as with every model, the chosen description of the seasonal components is imperfect.~~ Differences between years cannot be captured by our approach, as we fit a single sine wave to describe the average seasonal behaviour. To robustly capture the average seasonal behaviour, we need relatively long time series. Comparing results from two different 10 year periods shows that the signatures are robust for the majority of catchments, i.e. their values do not differ substantially from one time period to the other (details are shown in the Supplement).

The UK catchments and most of the US catchments exhibit a relatively strong unimodal (climate) seasonality (see e.g. Knoben et al., 2018). In other climates with a less distinct seasonal pattern, or with two seasons per year (Knoben et al., 2019b), our approach will not work. Semi-arid and arid catchments also tend to have a less smooth seasonal input, as water availability is more fragmented (Peters et al., 2003). ~~They also~~ Water-limited catchments can show a strong difference between potential evapotranspiration and actual evapotranspiration, which ~~might limit~~ limits the applicability of our approach (we will discuss that later in more detail). We exclude catchments where precipitation is falling as snow. While snowy catchments are typically also strongly seasonal (Schaefli, 2016), this seasonality is mostly a climate phenomenon. It is rather related to temperature seasonality and not to the response of a catchment to periodic forcing.

5.2 A perceptual model of the seasonal response of catchments in the UK

The results, in particular Figures 5 and 6 and Table 2, show clear patterns in the seasonal signatures. We can see that the seasonal response in the UK can be simulated by either two reservoirs in series or two reservoirs in parallel. This does not mean that there are no other configurations of more reservoirs leading to the same pairs of amplitude ratio and phase shift. Rather, two reservoirs in series and in parallel, respectively, are the most parsimonious reservoir configuration to reproduce the observed seasonal behaviour. Of course, two reservoirs in parallel and two reservoirs in series, respectively, might be seen as "special cases" of a soil reservoir followed by a fast and a slow reservoir, i.e. a three-reservoir arrangement. Furthermore, there might be concepts other than reservoirs which ~~are capable of explaining~~ can explain the observed behaviour. Still, the observed patterns, both where the catchments plot in the ~~amplitude-phase~~ amplitude ratio vs. phase shift plot (Figure 5) and how the catchment attributes relate to that (Figure 6), suggest that the seasonal signatures are indeed a window into catchment functioning (Berghuijs et al., 2014) and thus have discriminatory power (McMillan et al., 2017; Addor et al., 2018).

Figures 6a and 6b show how climate aridity and catchment wetness influence amplitude ratio and phase shift. The observation that more humid catchments respond more quickly to forcing (Figures 6a and 6b) concurs with our understanding of these catchments. Wetter and therefore more saturated catchments partition the incoming water mostly into fast flow. The hydrograph closely resembles the forcing, which can also be seen in Figure 4 for the responsive Ericht river. The drier the catchments become, the more water is able to infiltrate and subsurface properties become more important. This might explain why the spread becomes larger for less humid and hence less saturated catchments. In less humid catchments, actual evapotranspiration is more likely to deviate from potential evapotranspiration. This might be another reason for the greater attenuation in drier catchments, as the actual input ($P - E_a$) is lower than the theoretical one we compare to ($P - E_p$). In the UK the assumption that $E_a = E_p$ seems reasonable (see Supplement for further information). In more arid regions, such as parts of the US (see Section 5.3), this assumption might be is invalid.

The variability among UK catchments that cannot be explained by catchment wetness can mostly be explained by subsurface properties and the associated response time of a catchment. Catchments with high BFIs and thus large baseflow components show lower amplitude ratios and larger phase shifts, that is a more damped and lagged response (Peters et al., 2003). This can also be seen in Figure 4 for the groundwater-dominated East Avon river. The relationship between BFI and the seasonal signatures (Figure 6c) is not surprising, yet since the relationship is not unique, the seasonal signatures add another piece of information. In particular, the phase shift adds a timing componenttime scale, which quantifies how long – on average – the seasonal input is delayed to become the seasonal output. While the phase shift is only a few days for the most responsive catchments, in the slowest catchments the seasonal signal is shifted up to four months. Since the BFI is rather a consequence of a catchment's hydrological behaviour (as are the seasonal signatures) than an attribute of a catchment, the BFI cannot be seen as a cause for the observed patterns in the seasonal signatures. It cannot be used, for example, as a predictor in ungauged catchments. A qualitative attribute that could theoretically be available in ungauged catchments, the fraction of highly productive fractured aquifer, reinforces the influence of the subsurface (Figure 6d). Except for a few catchments, catchments underlain by such an aquifer exhibit very large phase shifts. In fact, all the catchments above the single reservoir line are underlain by highly productive aquifers. In these catchments, mostly underlain by Chalk, almost all the incoming water infiltrates into the aquifer, and the fast flow component often is negligible. This might explain why they do not behave like reservoirs in parallel, but rather like reservoirs in series, e.g. a soil reservoir (recharge) and a very slow groundwater reservoir. The few catchments which are underlain by highly productive aquifers, but do not exhibit large phase shifts, are typically overlain by rather impermeable drift, which stops water from infiltrating into the aquifer below.

Many models frequently used (and some of them developed) in the UK have a parallel flow structure, and catchments are usually conceptualised as having a fast and a slow component. While parametrisations and model structures vary between models, an overall parallel flow structure following a soil moisture module can be found in the PDM model (Moore, 2007), the TOPMODEL modelling concept (consisting of two fast flow responses; Beven and Kirkby, 1979), the IHACRES model (Croke and Jakeman, 2004), the GR4J model (Perrin et al., 2003), and many others. These or similar models have been applied to many catchments in the UK by various authors (e.g. Smith et al., 2019; ?; Coxon et al., 2019)(e.g. Smith et al., 2019; Lane et al., 2019; Coxon et . The seasonal signatures suggest that for most of the catchments, particularly if they are not underlain by a highly productive aquifer, a parallel model structure is a reasonable choice (at least for reproducing the response to seasonal forcing). For some groundwater-dominated catchments, however, the fast flow component seems to be rather unimportant. Many of these catchments, typically catchments underlain by Chalk, could only be poorly modelled in national-scale modelling studies (Smith et al., 2019; ?; Coxon et al., 2019)(Smith et al., 2019; Lane et al., 2019; Coxon et al., 2019). While this might partly be due to water balance problems (inter-catchment groundwater flows), it might also be due to an inadequate model structure or inadequate parameter ranges. The most parsimonious reservoir configuration to explain the seasonal behaviour of these catchments (phase shifts > 91 days) would be two reservoirs in series, e.g. a soil or unsaturated zone reservoir transforming the incoming forcing into recharge, and a (linear) groundwater reservoir. At least one of these reservoirs would need to be very slow to obtain such large phase shifts (cf. Figure 2). For these groundwater-dominated catchments, a serial structure as it is also used in simple lumped groundwater models (e.g. Peters et al., 2003; Obergfell et al., 2019), seems to be a reasonable choice (at least for reproducing the response to seasonal forcing). As mentioned before, two reservoirs in parallel and two reservoirs in series, respectively, might be seen as "special cases" of a soil reservoir followed by a fast and a slow reservoir. For example, some of the catchments underlain by a highly productive aquifer fall in the area that can be simulated by two reservoirs in parallel (see Figure 6d). Their large phase shifts and their proximity to the "single reservoir line" suggest, however, that the slow flow component is of particular importance and that large time constants (> 100 days) are required to model their behaviour.

In summary, the first control on the attenuation of the seasonal signal in the UK is the partitioning between fast flow and slow flow. More saturated catchments partition more rainfall into fast flow and hence lead to a higher amplitude ratio and to a smaller phase shift. The second control ~~is~~ are catchment subsurface properties, which determine the available storage and how slowly water leaves the system. The slower the catchment responds, the larger the phase shift and the lower the amplitude ratio. The Chalk catchments in the UK might be seen as an extreme case where almost all the water infiltrates, and hence the response time of a single slow reservoir (or perhaps two reservoirs in series) is the main control on the propagation of a periodic signal. On the other end of the spectrum, there are fully saturated, very responsive catchments mostly along the west coast of the UK, which behave almost like a single fast reservoir. Using conceptual reservoirs is only one way to interpret the seasonal signatures. It is useful as many hydrological models are built in that way. There might be, however, other possible ways of interpretation which we do not consider here.

5.3 A hydro-meteorologically more diverse set of catchments – the contiguous US

From Figure 7 and Table 3 it can be seen that for CAMELS catchments (US) the climate indices explain most of the variability in the seasonal response. Again, more humid catchments tend to create more fast flow, and hence they have high amplitude ratios and small phase shifts. Catchments with a larger moisture index seasonality tend to have smaller phase shifts. In these catchments precipitation and potential evapotranspiration are mostly out of phase. Therefore, precipitation falls in more humid months, which might lead to a ~~more flashy~~ flashier response. That means that both precipitation falling on wetter catchments and precipitation falling in wetter months will be attenuated less. The influence of catchment form is much less pronounced than in the climatically more homogeneous UK. Continental or global studies tend to identify climate as the dominant hydrological driving force (van Dijk, 2010; Beck et al., 2015), yet regional studies often show other attributes such as geology to be important (for baseflow, see e.g. Longobardi and Villani, 2008; Bloomfield et al., 2009). Our findings highlight anew that generalising from global to regional scale, or from regional to global scale, is not straightforward. Such scaling should ideally be done in a process-based way, or by analysing sub-climates, as the dominance of climate might mask the influence of other factors at large scales. We can also see that the attribute "fraction of highly productive fractured aquifers" (Figure 6d), which is a hydrogeological classification available for the UK, shows a much clearer pattern than any soil or geology attributes in the US (see e.g. Figure 7d which shows the fraction of carbonate sedimentary rock; the same is true if we use e.g. soil permeability for the UK). This might partly be due to the more heterogeneous US climate which masks the influence of subsurface properties to some degree. But it might also indicate that the soil or geology data used do not contain the hydrologically relevant soil or geology information. The hydrogeological classification based on expert judgement available for the UK, even though it is only categorical, might be more representative of the actual hydro(geo)logical processes at the scale of interest. We therefore cannot conclude that in the US catchment form does not play a role. We can merely say that the catchment attributes used do not show clear patterns at the continental scale.

Some of the rather arid catchments in the US plot outside the area that can be modelled by two reservoirs in series or in parallel (Figure 7). This either indicates that we would need another reservoir in series to model the observed phase shift (three reservoirs in series would result in a maximum phase shift of approximately 273 days), that (linear) reservoirs are not a good description of the hydrological processes, or that the ~~seasonal signature approach breaks down~~ proposed signatures are unreliable for these arid catchments. Since in water-limited catchments, actual evapotranspiration is typically much smaller than potential evapotranspiration, the input signal we use is very likely a poor proxy for the actual input signal. In very arid catchments ($I_m < -0.5$, dark red dots in Figure 7a), particularly with low moisture seasonality index (Figure 7b), the results should therefore be interpreted with care. It is unclear to what extent these large phase shifts are the result of a poorly approximated input signal or actual catchment function. This compromises the consistency (McMillan et al., 2017) of the seasonal signatures and makes them most suitable for energy-limited catchments. A way to overcome this limitation would be the use of modelled or measured actual evapotranspiration as input data. As this would require another modelling step or additional data, we leave this for future work (see Supplement for further information).

5.4 Can two common hydrological models reproduce the observed seasonal signatures?

The ensemble of IHACRES simulations ~~(Figure 8a)~~ covers the observed range of amplitude ratios and phase shifts~~(Figure 5a~~, although one catchment sits just at the boundary of the point cloud (Figure 8a, Figures 9a-c). The BFI pattern also roughly resembles the observed pattern (Figure 6c). Catchments with low BFIs tend to have high amplitude ratios and small phase shifts and vice versa. ~~Figure 8a also shows some patterns that correspond to the theoretical considerations in Section 2 (Figures 2 and 3). To explain this~~To explain the signature space of IHACRES, it is useful to recall the structure of the model. IHACRES consists of a soil moisture deficit store, followed by two parallel linear reservoirs. It thus approximately features the two examples introduced in Section 2, namely two reservoirs in series or in parallel.

If one of the parallel reservoirs in IHACRES receives very little water (due to an extremely high or low fraction p going into the slow reservoir), the whole system acts like two reservoirs in series. The only difference is that the first reservoir is not a single linear reservoir. It is a non-linear deficit store and thus different from the idealised linear reservoir. This might explain why the upper boundary looks similar to the grey dashed line indicating two linear reservoirs in series, yet not exactly the same. We did explore how non-linear reservoirs behave in terms of amplitude ratio and phase shift and they seem to behave similar to linear reservoirs (see Supplement). ~~The actual~~ Another reason for IHACRES not covering the whole area might be the parameters ranges (see Supplement for details). The parameters ranges used are ~~relatively~~ intended to be wide, yet especially the fast reservoir is (to be indeed fast) limited to ~~5~~ 10 days, which limits the theoretical space to be smaller than shown in Figure 2.

If the soil moisture reservoir transmits water relatively quickly without much attenuation, the whole system acts like two reservoirs in parallel. ~~Catchments with similar BFIs have a similar fraction p going into the slow reservoir, which is why we can see patterns (the "yellow stripes") similar to Figure 3a-c (the green, blue, and red stripes).~~ In summary, IHACRES is very similar to the idealised arrangement we introduced in Section 2 and this can be seen in the model output. It ~~would therefore be~~ is therefore likely that IHACRES is capable of reproducing the observed seasonal signatures for catchments in the UK (Figure 6) and for most of the catchments in the US (Figure 7). Whether IHACRES can reproduce the seasonal signatures, other hydrological signatures and achieve satisfactory statistical performance metrics simultaneously is to be explored and beyond the scope of this paper.

The ensemble of GR4J simulations ~~(Figure 8a) only covers a small range of~~ covers most of the amplitude ratios and phase shifts observed in the UK ~~and the US (Figure 6c~~(Figure 8b, Figures 9d-f). Many of the model runs lead to amplitude ratios higher than unity, which is caused by the groundwater exchange parameter, which allows the model to import water in addition to incoming P. While this is possible (and can in fact be observed; e.g. in Figure 6c the blue dot outside the grey boundaries is a catchment with water transfer from a neighbouring catchment), it is observed very rarely in the catchments investigated. Furthermore, a non-zero groundwater exchange parameter should ideally be associated with actual water inputs or ~~ouputs~~ outputs (e.g. inter-catchment groundwater flows), and these inputs or outputs are ~~often unknown.~~ usually unknown. It is worth noting that many model runs that lead to signature values at the boundaries of the signature space (e.g. low BFIs or large phase shifts) are associated with large (positive or negative) values for the groundwater exchange parameter (not shown here). This might further reduce the "realistic" signature space, as, for example, obtaining a low amplitude ratio by removing water might be seen as "the right answer for the wrong reason". No model run leads to a phase shift larger than about ~~60~~ 110 days. GR4J also has a soil moisture store followed by two parallel routing stores, i.e. the overall model structure is similar to IHACRES. The stores are, however, not linear reservoirs. ~~They also have fixed~~ In addition to that, GR4J has fixed internal parameter values, such as the ~~splitting between fast and slow routing . These internal fixed parameter values might limit the ability of GR4J to reproduce the seasonal signatures. Particularly the flow delay parameter, for which we use a typical parameter range, might be too narrow to~~ fraction of water going through the slow routing store, which is set to 0.9. This might explain why the BFI tends to be very high, as it can be seen from Figure 9f. Despite the tendency towards large BFIs, GR4J cannot produce phase shifts ~~longer than 60 days . From Figure 9 we can see that GR4J tends to produce high BFIs. The internal parameter values, in particular the fixed fraction of 0.9 going through the slow routing store, might limit the signature space of GR4J (note that the BFI clusters around 0.9~~larger than about 110 days given the parameter ranges used here. This might be due to a too narrow range of the flow delay parameter (maximum 15 d). So, to model both the phase shift and the BFI correctly, we might require a more flexible splitting between fast and slow routing and a means to produce larger phase shifts (e.g. via a wider range for the flow delay parameter).

Figure 9 also shows how the ~~BFI and the seasonal signatures~~ seasonal signatures (and the BFI) vary with different input (forcing). For both models, more humid catchments lead to higher amplitude ratios and smaller phase shifts, and vice versa. This trend, not necessarily the values themselves, agrees with the observed behaviour shown in Figures 6a and 7a.

This analysis is necessarily incomplete for (at least) two reasons. First, we only looked at 40 catchments in the UK to limit the computational demand. Therefore, the conclusions are not necessarily transferable to catchments outside the UK. More arid catchments (e.g. in the US) might show a different behaviour (e.g. the catchments showing phase shifts larger than 182 days, see Figure 7). Second, the sampling scheme (Latin Hypercube sampling) explores only a subspace of the actual parameter values (both because of the parameter ranges and because of the finite amount of parameter sets). We also made an a-priori decision of how to sample by choosing Latin Hypercube sampling in the first place. This is inevitably subjective, and other sampling schemes might lead to different results. This might especially affect the ~~peaks of the PDFs~~ distributions of the modelled signatures shown in Figure 9. ~~It is, however, rather unlikely that signature spaces that are highly underrepresented here (e.g. BFIs lower than 0.4 for GR4J) will be covered using different sampling schemes.~~ Wider parameter ranges might change the ranges of the resulting signature spaces. As we use rather wide ranges based on ~~literature values (see Supplement of Knoben et al., 2019a)~~recent literature (see Supplement for details), our results ~~are at leastin line with the practical use of the models, perhaps even more general~~should (at least) be representative of current modelling practice. This kind of analysis and the seasonal signatures can therefore help to select (or not select) models a-priori, without calibrating them to streamflow data (cf. Vogel and Sankarasubramanian, 2003). This might be particular helpful for large sample studies where often a certain model structure is chosen a-priori, even if it might be inadequate for the catchment sample investigated (Addor and Melsen, 2019).

6 Conclusions and outlook

We have tested seasonal hydrological signatures aimed at representing how climate seasonality is translated into streamflow seasonality, both approximated by sine waves. The damping (the amplitude ratio) and the phase shift of the incoming sine wave have been used to quantify how catchments respond to seasonal forcing. The presented signatures follow the guidelines of McMillan et al. (2017). The signatures are identifiable, robust, and consistent (see Supplement for further information). They are representative and have discriminatory power as they exhibit explicable, hydrologically interpretable patterns, particularly for energy-limited catchments (Figures 6 and 7)~~, particularly for energy-limited catchments~~. They can be related to conceptual model structures (arrangements of linear reservoirs, Figure 5), and the model evaluation (Figure 8) has shown that we can indeed observe this theoretical behaviour in model outputs. ~~We have also shown that the signatures can be used as a diagnostic tool since GR4J has been shown to be incapable of modelling the whole observed signature space.~~ As we use precipitation minus potential evapotranspiration as a proxy for the input to a catchment, the seasonal signatures ~~become less reliable~~ are unreliable for water-limited catchments. To use the seasonal signatures in water-limited catchments we would need to estimate actual evapotranspiration. The current approach is therefore ~~most~~ only suitable for energy-limited~~catchments~~, non-snowy catchments with a distinct unimodal seasonality, such as catchments in the UK.

We have found that the propagation of the seasonal input through a catchment depends both on climate and catchment form. Climate aridity and seasonality, and corresponding annual and seasonal catchment wetness, drive the partitioning of the incoming forcing into fast and slow flow. Catchment form, such as subsurface properties, influences how strongly the seasonal input gets attenuated. This is particularly visible in the UK, where the hydrogeological classification available (fraction of highly productive aquifer) can explain the very slow response of some catchments. The ~~more dominant~~ seemingly more dominant (and less clear) role of climate in the US highlights that scaling from regional to continental (or global) scale is not straightforward and requires thoughtful, ideally process-based approaches. Or in the words of Turner (1989), "conclusions or inferences regarding landscape patterns and processes must be drawn with an acute awareness of scale". Nonetheless, the clear link to climate and aquifer characteristics in the UK suggests that the signatures might be useful for catchment classification and for predictions in ungauged catchments, as long as potential evapotranspiration is an adequate proxy for actual evapotranspiration.

The model evaluation has shown that the signatures have the potential to be used as a diagnostic tool. GR4J could not reproduce the observed combinations of phase shift and BFI, pointing towards structural deficiencies of the model for certain catchments. As the seasonal signatures are relatable to conceptual model structures (arrangements of reservoirs), we could – given sufficient data – also build models based on inference from observed values of the signatures, and not just test existing model structures. This could be done in a stepwise fashion, starting with the seasonal time scale and then adding more complexity if needed (Jothityangkoon et al., 2001; Farmer et al., 2003; McMillan et al., 2011). It would be a step towards model structure identification based on hydrological reasoning, i.e. getting the right answers for the right reasons (Kirchner, 2006). If we decide on a certain model structure (e.g. two reservoirs in series), we can then use the presented theory to ~~calculate~~ estimate time constants of the reservoirs (the parameters). This ~~might reduce the need for calibration , or at least narrow down the ranges of parameter values~~could be used as additional constraint in the calibration process. If the time constants obtained from the seasonal signatures differ from time constants obtained by other means, e.g. by calibrating the model using a metric such as KGE, this might be indicative of limitations of typical modelling approaches (Fowler et al., 2018). It might be that the slower annual signal is exciting different parts of the catchments than events (individual peaks or recessions) do, which we typically calibrate to.

The idea of exploring a model's signature space (following the approach of Vogel and Sankarasubramanian, 2003) perhaps deserves more attention. It allows to explore models systematically and it can reveal whether a model ~~leads to a representative signature space. That is, whether the model has the tendency – or is capable at all – to produce hydrological signature distributions (ranges) that resemble the observed signature distributions (ranges) exhibited by~~ can simulate the ranges of hydrological signatures we obtain by analysing catchment data. Similar to sensitivity analysis, it allows us to explore and to better understand how a model works, which parameters are important for which signature, and what output behaviour a model can generate in general – without (and before) calibration. While we limited this analysis to a few signatures, in future studies we should focus on testing whether a model can simultaneously reproduce multiple signatures focusing on different aspects of the hydrological system (Euser et al., 2013; Hrachowitz et al., 2014).

Code and data availability. A repository with Matlab code used for the analysis and the resulting data is available from https://github.com/SebastianGnann/Seasonal_signatures_paper_public. Colours are based on www.ColorBrewer.org, by Cynthia A. Brewer, Penn State. The MARRMoT toolbox is available from https://github.com/wknoben/MARRMoT. The CAMELS dataset is available from https://ral.ucar.edu/solutions/products/camels. Information about the UK Benchmark Network can be obtained from https://nrfa.ceh.ac.uk/benchmark-network. Streamflow data and catchments attributes are available from https://nrfa.ceh.ac.uk. CEH-GEAR precipitation data are available from https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca. CHESS-PE potential evapotranspiration data are available from https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7.

Author contributions. SJG, NJKH and RAW conceptualised the research project. SJG performed the formal analysis. SJG prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is funded as part of the Water Informatics Science and Engineering Centre for Doctoral Training (WISE CDT) under a grant from the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/L016214/1. Parts of this work were carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - http://www.bris.ac.uk/acrc/. Thanks to Wouter Knoben for help with the MARRMoT toolbox, helpful discussions, and for helpful comments on an earlier version of this manuscript. Thanks to Gemma Coxon for assisting with the data. ~~This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol -~~ We also thank the Editor and three anonymous reviewers for their helpful feedback.

References

Addor, N. and Melsen, L. A.: Legacy, rather than adequacy, drives the selection of hydrological models, Water Resources Research, 55, 378–390, https://doi.org/10.1029/2018WR022958, 2019.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrology and Earth System Sciences, 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, Water Resources Research, 54, 8792–8812, https://doi.org/10.1029/2018WR022606, 2018.

Beck, H. E., de Roo, A., and van Dijk, A. I. J. M.: Global maps of streamflow characteristics based on observations from several thousand catchments, Journal of Hydrometeorology, 16, 1478–1501, https://doi.org/10.1175/JHM-D-14-0155.1, 2015.

Berghuijs, W. R., Sivapalan, M., Woods, R. A., and Savenije, H. H. G.: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales, Water Resources Research, 50, 5638–5661, https://doi.org/10.1002/2014WR015692, 2014.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrological Sciences Bulletin, 24, 43–69, https://doi.org/10.1080/02626667909491834, 1979.

Bloomfield, J. P., Allen, D. J., and Griffiths, K. J.: Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK, Journal of Hydrology, 373, 164–176, https://doi.org/10.1016/j.jhydrol.2009.04.025, 2009.

Cayan, D. R., Riddle, L. G., and Aguado, E.: The influence of precipitation and temperature on seasonal streamflow in California, Water Resources Research, 29, 1127–1140, https://doi.org/10.1029/92WR02802, 1993.

Cheng, J. and Druzdzel, M.: Latin hypercube sampling in Bayesian networks., Proceedings of the Thirteenth International Florida Artificial Intelligence Research Symposium Conference, pp. 287–292, http://www.aaai.org/Papers/FLAIRS/2000/FLAIRS00-054.pdf, 2000.

Clausen, B. and Biggs, B.: Flow variables for ecological studies in temperate streams: groupings based on covariance, Journal of Hydrology, 237, 184–197, https://doi.org/10.1016/S0022-1694(00)00306-1, 2000.

Colwell, R. K.: Predictability, constancy, and contingency of periodic phenomena, Ecology, 55, 1148–1153, https://doi.org/10.2307/1940366, 1974.

Court, A.: Measures of streamflow timing, Journal of Geophysical Research, 67, 4335–4339, https://doi.org/10.1029/JZ067i011p04335, 1962.

Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J. M., Howden, N. J. K., Quinn, N., Wagener, T., and Woods, R.: DECIPHeR v1: Dynamic fluxEs and ConnectIvity for Predictions of HydRology, Geoscientific Model Development, 12, 2285–2306, https://doi.org/10.5194/gmd-12-2285-2019, 2019.

Croke, B. F. and Jakeman, A. J.: A catchment moisture deficit module for the IHACRES rainfall-runoff model, Environmental Modelling and Software, 19, 1–5, https://doi.org/10.1016/j.envsoft.2003.09.001, 2004.

DeWalle, D. R., Edwards, P. J., Swistock, B. R., Aravena, R., and Drimmie, R. J.: Seasonal isotope hydrology of three Appalachian forest catchments, Hydrological Processes, 11, 1895–1906, https://doi.org/10.1002/(SICI)1099-1085(199712)11:15<1895::AID-HYP538>3.0.CO;2-%23, 1997.

Dooge, J.: Linear theory of hydrologic systems, 1468, Agricultural Research Service, US Department of Agriculture, 1973.

Eriksson, E.: Compartment models and reservoir theory, Annual Review of Ecology and Systematics, 2, 67–84, https://doi.org/10.1146/annurev.es.02.110171.000435, 1971.

Erskine, A. and Papaioannou, A.: The use of aquifer response rate in the assessment of groundwater resources, Journal of Hydrology, 202, 373–391, https://doi.org/10.1016/S0022-1694(97)00058-9, 1997.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrology and Earth System Sciences, 17, 1893–1912, https://doi.org/10.5194/hess-17-1893-2013, 2013.

Farmer, D., Sivapalan, M., and Jothityangkoon, C.: Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis, Water Resources Research, 39, 1–21, https://doi.org/10.1029/2001WR000328, 2003.

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., Woods, R. A., and Zhang, L.: Simulating runoff under changing climatic conditions: a framework for model improvement, Water Resources Research, 54, 9812–9832, https://doi.org/10.1029/2018WR023989, 2018.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22, 3802–3813, https://doi.org/10.1002/hyp.6989, 2008.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Harman, C. J.: Age-ranked storage-discharge relations – a unified description of spatially-lumped flow and water age in hydrologic systems, Water Resources Research, p. 2017WR022304, https://doi.org/10.1029/2017WR022304, 2019.

Harrigan, S., Hannaford, J., Muchan, K., and Marsh, T. J.: Designation and trend analysis of the updated UK Benchmark Network of river flow stations: the UKBN2 dataset, Hydrology Research, 49, 552–567, https://doi.org/10.2166/nh.2017.058, 2018a.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrology and Earth System Sciences, 22, 2023–2039, https://doi.org/10.5194/hess-22-2023-2018, 2018b.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) – a review, Hydrological Sciences Journal, 58, 1198–1255, https://doi.org/10.1080/02626667.2013.803183, 2013.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., and Gascuel-Odoux, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, Water Resources Research, 50, 7445–7469, https://doi.org/10.1002/2014WR015484, 2014.

Institute of Hydrology: Low Flow Studies Report No. 1: Research Report, Institute of Hydrology, 1980.

Jasechko, S., Birks, S. J., Gleeson, T., Wada, Y., Fawcett, P. J., Sharp, Z. D., McDonnell, J. J., and Welker, J. M.: The pronounced seasonality of global groundwater recharge, Water Resources Research, 50, 8845–8867, https://doi.org/10.1002/2014WR015809, 2014.

Jothityangkoon, C., Sivapalan, M., and Farmer, D.: Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development, Journal of Hydrology, 254, 174–198, https://doi.org/10.1016/S0022-1694(01)00496-6, 2001.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42, 1–5, https://doi.org/10.1029/2005WR004362, 2006.

Kirchner, J. W.: Aggregation in environmental systems – Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, Hydrology and Earth System Sciences, 20, 279–297, https://doi.org/10.5194/hess-20-279-2016, 2016.

Knoben, W. J. M., Woods, R. A., and Freer, J. E.: A quantitative hydrological climate classification evaluated with independent streamflow data, Water Resources Research, 54, 5088–5109, https://doi.org/10.1029/2018WR022913, 2018.

Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., and Woods, R. A.: Modular Assessment of Rainfall–Runoff Models Toolbox (MARRMoT) v1.2: an open-source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous state-space formulations, Geoscientific Model Development, 12, 2463–2480, https://doi.org/10.5194/gmd-12-2463-2019, 2019a.

Knoben, W. J. M., Woods, R. A., and Freer, J. E.: Global bimodal precipitation seasonality: A systematic overview, International Journal of Climatology, 39, 558–567, https://doi.org/10.1002/joc.5786, 2019b.

Laaha, G. and Blöschl, G.: Seasonality indices for regionalizing low flows, Hydrological Processes, 20, 3851–3878, https://doi.org/10.1002/hyp.6161, 2006.

Laaha, G., Demuth, S., Hisdal, H., Kroll, C. N., van Lanen, H. A. J., Nester, T., Rogger, M., Sauquet, E., Tallaksen, L. M., Woods, R. A., and Young, A.: Prediction of low flows in ungauged basins, in: Runoff Prediction in Ungauged Basins, edited by Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., 2011, pp. 163–188, Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9781139235761.011, 2013.

Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., Greene, S., Macleod, C. J. A., and Reaney, S. M.: Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain, Hydrology and Earth System Sciences, 23, 4011–4032, https://doi.org/10.5194/hess-23-4011-2019, https://www.hydrol-earth-syst-sci.net/23/4011/2019/, 2019.

Longobardi, A. and Villani, P.: Baseflow index regionalization analysis in a Mediterranean area and data scarcity context: Role of the catchment permeability index, Journal of Hydrology, 355, 63–75, https://doi.org/10.1016/j.jhydrol.2008.03.011, 2008.

McGuire, K. J. and McDonnell, J. J.: A review and evaluation of catchment transit time modeling, Journal of Hydrology, 330, 543–563, https://doi.org/10.1016/j.jhydrol.2006.04.020, 2006.

McMillan, H., Westerberg, I., and Branger, F.: Five guidelines for selecting hydrological signatures, Hydrological Processes, 31, 4757–4761, https://doi.org/10.1002/hyp.11300, 2017.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, Hydrological Processes, 25, 511–522, https://doi.org/10.1002/hyp.7841, 2011.

McMillan, H. K., Gueguen, M., Grimon, E., Woods, R. A., Clark, M. P., and Rupp, D. E.: Spatial variability of hydrological processes and model structure diagnostics in a 50km^2 catchment, Hydrological Processes, 28, 4896–4913, https://doi.org/10.1002/hyp.9988, 2014.

Milly, P. C. D.: Climate, soil water storage, and the average annual water balance, Water Resources Research, 30, 2143–2156, https://doi.org/10.1029/94WR00586, 1994.

Montanari, A. and Toth, E.: Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, Water Resources Research, 43, 1–10, https://doi.org/10.1029/2006WR005184, 2007.

Moore, R. J.: The PDM rainfall-runoff model, Hydrology and Earth System Sciences, 11, 483–499, https://doi.org/10.5194/hess-11-483-2007, 2007.

Nash, J.: The form of the instantaneous unit hydrograph, International Association of Scientific Hydrology, Publ, 3, 114–121, 1957.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I – A discussion of principles, Journal of Hydrology, 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

National River Flow Archive: https://nrfa.ceh.ac.uk, NERC CEH, Wallingford, https://nrfa.ceh.ac.uk, 2019.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, 2015.

Obergfell, C., Bakker, M., and Maas, K.: Estimation of average diffuse aquifer recharge using time series modeling of groundwater heads, Water Resources Research, 55, 2194–2210, https://doi.org/10.1029/2018WR024235, 2019.

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Research and Applications, 19, 101–121, https://doi.org/10.1002/rra.700, 2003.

Oudin, L., Salavati, B., Furusho-Percot, C., Ribstein, P., and Saadi, M.: Hydrological impacts of urbanization at the catchment scale, Journal of Hydrology, 559, 774–786, https://doi.org/10.1016/j.jhydrol.2018.02.064, 2018.

Peel, M. C. and Blöschl, G.: Hydrological modelling in a changing world, Progress in Physical Geography, 35, 249–261, https://doi.org/10.1177/0309133311402550, 2011.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275–289, https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Peters, E., Torfs, P. J., van Lanen, H. A., and Bier, G.: Propagation of drought through groundwater – A new approach using linear reservoir theory, Hydrological Processes, 17, 3023–3040, https://doi.org/10.1002/hyp.1274, 2003.

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

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., and Stromberg, J. C.: The Natural Flow Regime, BioScience, 47, 769–784, https://doi.org/10.2307/1313099, 1997.

Post, D. A. and Jakeman, A. J.: Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model, Ecological Modelling, 123, 91–104, https://doi.org/10.1016/S0304-3800(99)00125-8, 1999.

Regonda, S. K., Rajagopalan, B., Clark, M., and Pitlick, J.: Seasonal cycle shifts in hydroclimatology over the western United States, Journal of Climate, 18, 372–384, https://doi.org/10.1175/JCLI-3272.1, 2005.

Richter, B. D., Baumgartner, J. V., Powell, J., and Braun, D. P.: A method for assessing hydrologic alteration within ecosystems, Conservation Biology, 10, 1163–1174, https://doi.org/10.1046/j.1523-1739.1996.10041163.x, 1996.

Robinson, E., Blyth, E., Clark, D., Comyn-Platt, E., Finch, J., and Rudd, A.: Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (1961-2015) [CHESS-PE], https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7, 2016.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrology and Earth System Sciences, 15, 2895–2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

Schaefli, B.: Snow hydrology signatures for model identification within a limits-of-acceptability approach, Hydrological Processes, 30, 4019–4035, https://doi.org/10.1002/hyp.10972, 2016.

Sefton, C. and Howarth, S.: Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales, Journal of Hydrology, 211, 1–16, https://doi.org/10.1016/S0022-1694(98)00163-2, 1998.

Shafii, M. and Tolson, B. A.: Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives, Water Resources Research, 51, 3796–3814, https://doi.org/10.1002/2014WR016520, 2015.

Shi, X., Wood, A. W., and Lettenmaier, D. P.: How essential is hydrologic model calibration to seasonal streamflow forecasting?, Journal of Hydrometeorology, 9, 1350–1363, https://doi.org/10.1175/2008JHM1001.1, 2008.

Sivapalan, M., Blöschl, G., Zhang, L., and Vertessy, R.: Downward approach to hydrological prediction, Hydrological Processes, 17, 2101–2111, https://doi.org/10.1002/hyp.1425, 2003.

Smakhtin, V.: Low flow hydrology: a review, Journal of Hydrology, 240, 147–186, https://doi.org/10.1016/S0022-1694(00)00340-1, 2001.

Smith, K. A., Barker, L. J., Tanguy, M., Parry, S., Harrigan, S., Legg, T. P., Prudhomme, C., and Hannaford, J.: A multi-objective ensemble approach to hydrological modelling in the UK: an application to historic drought reconstruction, Hydrology and Earth System Sciences, 23, 3247–3268, https://doi.org/10.5194/hess-23-3247-2019, 2019.

Svensson, C.: Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues, Hydrological Sciences Journal, 61, 19–35, https://doi.org/10.1080/02626667.2014.992788, 2016.

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D. G., and Keller, V. D. J.: Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890-2015) [CEH-GEAR], https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca, 2016.

Townley, L. R.: The response of aquifers to periodic forcing, Advances in Water Resources, 18, 125–146, https://doi.org/10.1016/0309-1708(95)00008-7, 1995.

Turner, M. G.: Landscape ecology: the effect of pattern on process, Annual Review of Ecology and Systematics, 20, 171–197, https://doi.org/10.1146/annurev.es.20.110189.001131, 1989.

van Dijk, A. I. J. M.: Climate and terrain factors explaining streamflow response and recession in Australian catchments, Hydrology and Earth System Sciences, 14, 159–169, https://doi.org/10.5194/hess-14-159-2010, 2010.

Vega, M., Pardo, R., Barrado, E., and Debán, L.: Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis, Water Research, 32, 3581–3592, https://doi.org/10.1016/S0043-1354(98)00138-9, 1998.

Vogel, R. M. and Sankarasubramanian, A.: Validation of a watershed model without calibration, Water Resources Research, 39, 1–9, https://doi.org/10.1029/2002WR001940, 2003.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment classification and hydrologic similarity, Geography Compass, 1, 901–931, https://doi.org/10.1111/j.1749-8198.2007.00039.x, 2007.

Weingartner, R., Blöschl, G., Hannah, D. M., Marks, D. G., Parajka, J., Pearson, C. S., Rogger, M., Salinas, J. L., Sauquet, E., Srikanthan, R., Thompson, S. E., and Viglione, A.: Prediction of seasonal runoff in ungauged basins, in: Runoff Prediction in Ungauged Basins, edited by Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., 2011, pp. 102–134, Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9781139235761.009, 2013.

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., and Freer, J.: Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resources Research, 52, 1847–1865, https://doi.org/10.1002/2015WR017635, 2016.

Yadav, M., Wagener, T., and Gupta, H. V.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Advances in Water Resources, 30, 1756–1774, https://doi.org/10.1016/j.advwatres.2007.01.005, 2007.

Yokoo, Y. and Sivapalan, M.: Towards reconstruction of the flow duration curve: Development of a conceptual framework with a physical basis, Hydrology and Earth System Sciences, 15, 2805–2819, https://doi.org/10.5194/hess-15-2805-2011, 2011.

Young, P.: Data-based mechanistic modelling of environmental, ecological, economic and engineering systems, Environmental Modelling and Software, 13, 105–122, https://doi.org/10.1016/S1364-8152(98)00011-5, 1998.