The reviewer's comments are in italics and the responses in regular text.

In the manuscript "If a Rainfall-Runoff Model was a Hydrologist" by Ewen and O'Donnell, a set of parameterless rainfall-runoff models are developed in an experiment which aims to quantify the importance of the knowledge contained by the model itself. To make this knowledge explicit, the rainfall-runoff model is personified as a layman with an interest in the weather and river flows. The model is parameterless and relies on time-matching based on the similarity of the simulation day with a set of other days from the historical data record. The model performance of the developed KERR model is overall just slightly lower than the GR4J model performance for a set of UK catchments. A main finding of the paper is the strong relative importance of the temporal pattern of antecedent rainfall. This concrete model development example supports a broader and more philosophical discussion on hydrologic knowledge and laws within rainfall-runoff models.

The manuscript addresses relevant scientific questions on the role of hydrologic knowledge contained in rainfall-runoff models on model performance. However, at first read, I found the manuscript to be confusing in how it is structured and in its balance between the broader philosophical discussion and the very concrete, simple and specific modeling experiment. Which aspects of the specific example should or can we apply in more complex traditional rainfall-runoff modeling, is it the knowledge documentation through personification of the rainfall-runoff model?

I hope that the comments below will help improving the manuscript.

Thank you. The comments will be helpful in improving the manuscript.

(A) We accept that readers find the paper confusing on first reading. Some linking and summarising sentences will be added to brings out the scope and structure of the work and paper. The paper is unusual in that its scope is huge: broader philosophical questions are discussed, many definitions created, methods created, RR models developed, a knowledge experiment run (which is essentially a test for the hypothesis that **An RR model can be a model of a hydrologist**), and a concrete example created which gives a benchmark which can be used as a basis for generating hypotheses.

(B) The scope for what can be meant by "more complex traditional rainfall-runoff modelling" is huge. It extends from basic scientific studies, through model development and testing, to model use in a variety of ways and for an endless list of purposes. It is up to modellers to pick out what they find useful in the paper, given their own circumstances and requirements.

General comments:

1) The aim of the paper is not clearly stated in the abstract, I would suggest to explicitly add it. The aims are described in L70 and later L110 and in L174, however, throughout the manuscript it remains unclear what is exactly meant by 'corruption' of hydrologic knowledge flows within RR modelling. Could you clarify this further?

There is discussion of the scope and structure of the work in response para. A above. The aim is to create the concrete example (i.e. an RR model which is a model of a hydrologist), as indicated in the title and stated in L14-15 in the abstract and in L83. That aim needed a lot of support in the form of definitions and methods and discussion of broader philosophical questions.

(C) The general point about corruption is that the simulations and conclusions from RR modelling sometimes do not depend wholly or exactly on what the model developer knew or assumed when

creating the model. There are three parts to how a conclusion is reached: (1) the design of the model by the model developer; (2) how the model actually works; and (3) the model user drawing conclusions from simulations. Clearly, there can be corruption in part 2: e.g. missteps, approximations, and in translations when an algorithm is created from knowledge. The classic example of approximation is when numbers representing physical properties lose their meaning, such as when spatial differential equations for runoff are solved on a coarse grid. There can also be problems with parts 1 and 3, including problems which arise when model users create knowledge flows out of thin air. For example, KERR does not know about the physical reality of river catchments: the RR records lie at an outer boundary (see Fig.1) and are essentially a set of numbers. If in part 3 KERR is used to draw concussions relating explicitly in some way to realism then it will involve the misallocation of simulation responses to knowledge relating to realism (i.e. the model user creates a knowledge flow out of thin air). Parameter calibration is a confounding factor and can play a role in the general set of problems relating to the misallocation of simulation responses to hydrologic knowledge, and the misallocation of hydrologic knowledge to simulation responses. If and where that is seen as corrupting the flow of knowledge would depend on exactly what is done during calibration and exactly what conclusions are drawn from the results from calibration.

2) The manuscript does not contain a dedicated Conclusion section. Concluding remarks are provided in the Summary of Section 8. However, I think it would help the reader to include a dedicated conclusion section which specifically links back to the aims of the study. This would help to clarify the main message/focus of the manuscript.

See response para. A above.

3) The experiment was performed for a set of UK catchments. Could you discuss the application of the developed parameterless models and the conclusions drawn on the importance of wetness in the light of different climatic zones?

It would not make sense to study the importance of wetness using modelling that gives poor performance as a result of neglecting something that is very important in a region, such as snow or water resources management or drought.

4) How could the proposed methodology of quantifying the importance of hydrologic knowledge held by the MH on model performance be applied in more traditional rainfall-runoff modeling?

The method was created as a means to an end, for the knowledge experiment. This question therefore goes well beyond the scope of the paper. Given the nature of more traditional rainfall-runoff models, one way forward might be to use algorithmic differentiation to estimate the point sensitivities of performance to parameters (or the sensitivity of conclusions to parameters, if the conclusions are numerical or can be forced to be numerical).

5) Hydrological modeling is often used in practice to quantify the effect of change in a catchment (e.g. land use). In science, hydrological modeling is often used to increase our understanding of catchment functioning. Both would be difficult using the proposed approach of the parameterless model, could you please elaborate on this?

The parameterless modelling was created as a means to an end, for the knowledge experiment. This question therefore goes well beyond the scope of the paper. Off the shelf lumped RR models are of limited use if the aim is to understand catchment functioning. The best way forward is probably to work with black box models (e.g. deep learning RR models) and look for general laws or specific behaviours, but to do this taking into account the richness and value of what is already known. One conclusion which can be drawn from the paper is that any attempt to use lumped RR modelling to increase the understanding of catchment functioning must take very great care over the selection and use of performance measurement. Note that when measured in terms of the median NSE, UK catchments function exactly as laymen think they function.

6) The manuscript includes several references to the study of Jakeman and Hornberger (1993). However, a short summary of the main aspects of this paper in relation to the current paper seems to be lacking.

Agreed.

7) The way the work is presented is sometimes confusing. For example, in Section 4.2, the Trivial and Seasonal RR models are presented. Later in Section 6, an additional Wetness model is mentioned. In Table 6, also the KERR model is presented. Perhaps, it would be good to clarify this in Section 4.2 and in the Method section 3 so that the reader has a better understanding of the main approach.

See response para. A above.

Specific comments

L5: Could you add here why personification can also be instructive?

Yes.

L10: Simplification of complex systems is inherent to modeling, but I guess you want to quantify how and which of the knowledge contained in the model mostly affects model performance?

Yes.

L11: What do you mean by classic MH?

The classic structure for conceptual RR models comprises little more than a few leaky reservoirs: this structure is in widespread use and has been so for a few decades. This structure brings with it a need for parameter calibration. The combination of a few leaky reservoirs and a need for parameter calibration is the classic form for conceptual RR modelling.

L17: I found the sentence with "the relative importance is measured as 1 and 6" rather confusing. Do you mean: antecedent wetness is 6 times more important than seasons in simulating runoff in a time-matching modeling approach? I would suggest rephrasing this sentence (also in the Summary section).

You understood the meaning correctly. We will rephrase the sentence.

Figure 1: Although mentioned in the caption, it was at first read not entirely clear to me that the numbers refer to the knowledge statements of the different Tables, I would suggest rephrasing the caption to clarify.

Agreed.

L66-67: Could you elaborate further on this?

One interpretation of what Jakeman and Hornberger did was to create a compound performance measurement which included the NSE and a measure for the degree of overparameterization. They then manually adjusted the model structure to find good structures which maximise the compound measure. It is the fact that they used an appropriate compound measure which makes their conclusions about complexity useful.

L70: As mentioned before, what do you exactly mean by 'lost or corrupted' hydrologic knowledge flows within RR modeling?

See response para. C above.

L72: What do you exactly mean by science is an "activity and attitude"?

Baker (2017) should have been cited. Also see the response below (L73) .

L73: Could you elaborate further what you mean by the significant deficits, difficulties or dangers?

It can be quite daunting when an engineering decision has to be made based on scientific understanding. The best first step is often simply to force forward to get a concrete solution of some sort, even if it is for a simplified problem, and then work from there (it gives perspective). Getting to a solution involves paying attention to any deficits (especially perceived deficits in understanding), finding solutions to difficulties (e.g. not having a model to work with) and being aware of what might have gone wrong or might go wrong.

L85: It is not clear to me what you mean by "not well behaved", could you clarify?

It is not a simple, orderly flow. There can be loops and jumps.

L111: aim (3) is not entirely clear, could you elaborate on "the need to draw valid hydrologic conclusions"?

The reason why hydrologists use RR modelling is to reach some conclusion or other, whether it is simply about what the flow is or will be, or about some process or assumption, or the basis for some engineering decision. The aim and hope is that the conclusion is valid.

L134: In 'traditional' hydrological modeling, would you also recommend describing equations in the form of statements in everyday English? How does this relate with the more commonly provided model descriptions and equations?

KERR was designed for use in a prototype machine which analyses text and numerical tables. That is why everyday English was used as the mechanism for keeping control of the knowledge (rather than using equations or unnatural English which simply mimics logic sequences and equations). The approach was described as "extreme" in L182. Less extreme methods, with a mix of text and equations, would probably work just as well if there is no need for analysis using any specific type of tool.

L138: with "Here," do you mean: in the models being developed in this study?

"Here" means in this paper.

L141: could you specify conclusion 3 in Sect. 1, it is unclear to me to which point you are specifically referring to.

It is L34.

L174: when you mention "one of the aims", I would find it helpful to also have a recap of the other aims of the research.

Agreed.

L198: I am not sure to fully understand what you mean by "the MH should know why", could you clarify this part?

Knowledge is needed if the value of a constant has to be set, and that knowledge will clearly be of considerable importance if the value has to be set precisely. If a model was a hydrologist then the hydrologist would be acutely aware of that knowledge.

L201: "from such an experiment", do you mean an experiment without data fitting?

Yes, and with fixed constants that did not need to be set precisely.

L259: in contrast to the Trivial and Seasonal models, the Wetness model was not introduced earlier.

See response para. A above.

Fig2: Rain is a flux and should therefore have unit [L/T], I assume here it is mm/d. It is somewhat confusing to show negative values on the y-axis of the top panel. It would be clearer for the reader if rainfall pattern difference was also explained in the text describing Fig 2.

Rain is the depth accumulated in a day. The -50 will be changed to 50. A pointer to the figure legend will be added to the text.

In the paragraph 313-319, the horizontal alignments around 1996 are explained twice.

One mention is for alignment of blank spaces and the other for alignment of dots.

Table 6: is the KERR model a general name for the Trivial, Seasonal and Wetness models and a fourth model? Could you please clarify?

KERR is not a general name. It is the name for the model containing algorithm A4 (see Table 6).

L355: could you elaborate: "when drawing conclusions" on what?

When drawing conclusions about the median NSE and impacts on the median NSE.

Table 9: conclusion based on the third statement "Unpredictability" are not explained in the text of Section 7.1. This statement only comes back in the Summary of Section 8. Could you please elaborate on this finding already in Section 7.1?

Text will be added to Sect. 7.1. If there was perfect predictability then the NSE for all catchments could be 1.0, and if this was the case the mean rise in NSE for the 38 catchments would be 0.216 (see final row in Table 9). The importance of unpredictability is therefore 2 when compared to the 6 for wetness and 1 for seasonality.

L429: Here, I would suggest to explicitly mention "wetness, seasonality and unpredictability" to clarify: "the three pieces of hydrologic knowledge given in 8 and 9".

Agreed.

L449: Could you elaborate further on what you mean by the wetness kernel and discuss more in detail the related hydrologic law?

The notion of wetness linked to accumulated rainfall is a core idea that appears in several different forms in RR modelling, going back decades. In theory, it could be treated along the lines of Darcy's law and a simple classification process linked to the parameters for, say, leaky reservoirs or accumulations of rainfall. Note that, in the first instance, the underlying problem can be reduced to a cluster analysis.

Discussion: how important is personification of RR models? This seemed to be an important focus point at the start of the manuscript.

The concrete example is a hypothesis test for the concept of the MH (see response para. A above), so personification is a focus all the way through the paper, not just at the start. Personification gives

context to the benchmark created for the combination of the NSE and hydrologic knowledge, and that benchmark is generally useful (it is a benchmark for a layman). The notion of the personification of RR models does not, in itself, give a direct solution to anything. However, it may have an effect on what RR modellers think, how they think, and what they do. Until RR modelling is on a far better footing (e.g. using hydrologic laws at the catchment scale), it is difficult to know what will ultimately be found to be important.

Typos:

L14: a MH instead of an MH

L228: meteorological instead of metrological