GMD Submission by Coxon et al

DECIPHeR v1: Dynamic fluxEs and ConnectIvity for Predictions of HydRology

General Response

We thank the reviewers for taking the time to review the paper and their comments, which have greatly helped to improve this manuscript and the quality and clarity of the research.

The main comments from the reviewers focused on (1) better defining the novelty and originality of the proposed modelling framework, (2) clarity of the model structure and equations and (3) the model evaluation.

In response to these reviewer comments, we have substantially re-written sections of the abstract, introduction and key concepts to highlight the unique features of the modelling framework and how it compares to other modelling frameworks in hydrology. We have produced flow timeseries plots for a new figure in the paper that demonstrates the ability of the model to reproduce the observed flow timeseries for six catchments.

Due to the requested extra time the GMD editor kindly agreed to, we have also been able to implement a new, more computationally efficient and stable, analytical solution to the subsurface flow equations and have detailed this solution in a new appendix. The result of this has a) increased the novelty of the model because the solutions are a departure from those implemented in Dynamic TOPMODEL (Beven and Freer, 2001), b) resulted in improvements in the national results overall, thus all figures/results now reflect these new equations, c) sped up simulation times as the solution is no longer iterative and d) addressed the comments of Reviewer 2 who requested more detail about the implementation of our flux equations.

Detailed responses to all comments are provided below. Author responses are in **bold** and any modifications to the manuscript are in *italic* below each of the reviewer's comments. The reviewer comments are inserted as comments next to the relevant tracked changes in the main document.

Gemma Coxon, March 2019

Reviewer #1

This paper describes the development of the Dynamic fluxEs and Connectivity for Predictiosn of HydRology (DECIPHeR) framework for simulation of hydrology (especially river flow) at catchment to continental scales. The model is tested across the Great Britain at 1,366 gauges in the current study but the authors intend to expand the model domain and suggest that it can be applied at the continental scales. The framework appears to be efficient computationally but there are a number of issues that authors need to address before the manuscript can be considered for publication. I provide my specific comments below.

We thank Reviewer #1 for taking the time to review the paper. We appreciate their comments and provide our responses below.

(1) The authors should revise the introduction to clearly highlight the motivation behind and the need for such a framework in relation to numerous other ongoing model development efforts. For example, how does the proposed study advance hydrological modeling compared to the model presented by Chaney et al. (2016)? Further, there are a number of large-scale models that have the capability to simulate far more number of processes (e.g., groundwater dynamics, pumping, flood dynamics, human impacts) than those presented in the current framework (for example: Hanasaki et al. 2008; Ozdogan et al. 2010; Pokhrel et al. 2015; Wada et al. 2014). Certainly these models are intended for global/regional applications but there have been ongoing efforts to increase the spatial resolution (i.e., hyper-resolution models) for application of these models at smaller scales. Extensive review of these models is available in recent literature (Nazemi and Wheater 2015; Pokhrel et al. 2016; Wada et al. 2017). I suggest that the authors thoroughly revise the introduction including a discussion on these past/ongoing efforts. Note that most of these models use TOPMODEL to simulate some of the surface/sub-surface hydrologic processes.

We agree that we needed to make this clearer and have revised the abstract and introduction to clearly highlight the motivation behind the framework and it's unique features in response to comments from Reviewer 1 and 2.

(2) Since the framework is currently designed to primarily simulate river flow, it is also important to note studies on streamflow/flood simulations at local to continental scales (Bates et al. 2010; Miguez-Macho and Fan 2012; Yamazaki et al. 2011; Zhao et al. 2017). What is the rationale for having the new framework?

We do not believe the studies suggested by the reviewer are relevant to this study. Bates et al (2010) presents a new set of equations for floodplain inundation, Miguez-Macho and Fan (2012) are focused on the role of groundwater and Yamazaki et al (2011) and Zhao et al (2017) are primarily investigating flow routing schemes. While these are interesting studies, they are not focused on modelling frameworks that simulate the key hydrological processes (e.g. infiltration, runoff generation, subsurface flows etc.) at catchment-based scales in the generation of river flow and thus are fundamentally different to the modelling framework presented here.

(3) The above two issues are important because the authors' intent is to provide a framework for large-scale application.

We agree that this is an important point and address this comment in the two responses above.

(4) P4, L16-40: Why did the authors use HRUs instead of doing a fully-distributed model? Is it just the run time minimization? Is there a compromise in terms of adding new features such as groundwater flows and human water use? Again, I suggest adding a note on how this framework advances our capability to simulate the hydrology in comparison to numerous existing framework (see comments above)?

One clear benefit of using HRUs is minimising the run times of the model. However, the key benefit is the flexibility it gives you to modify the spatial complexity/scale of how spatial variability and hydrologic connectivity are represented. This flexibility means you can (1) run the model as a fully distributed, semi-distributed or lumped model, (2) have more/less spatial/process complexity where needed in the landscape and (3) represent point scale features in the landscape whilst still maintaining modelling efficiencies elsewhere. These features are hugely beneficial to having a pre-defined fully distributed model which cannot handle such occurrences. We have modified section 2.1 to clarify this point.

There isn't any compromise in terms of adding new features as each HRU is treated as a separate store in the model which can have different process conceptualisations and parameterisations. This means that more process complexity can be incorporated where needed to better suit local conditions e.g. to account for 'point-source' human influences or more complex hydrological processes such as surface-groundwater exchanges.

(5) P5, L15: "must contain no sinks": What if there are real inland sinks? There are too many across continents.

Sink filling is very common in digital terrain analyses when generating river and catchment layers (for example the SRTM DEM used for HydroSheds undergoes a sink filling process before it can be used to derive catchment basins). We agree that this will mean any real inland sinks in the digital elevation model will be filled. Currently the modelling framework is unable to account for these features (such as lakes), however, this is a feature we will be looking to include soon.

(6) Section 2.2.3: What is the routing scheme used? I find some description later in another section. Please consolidate the text and provide more details.

In Section 2.2.3 (now Section 2.2.2) we are describing the river routing data that is generated by the digital terrain analysis. These data (such as the river network connectivity and routing tables) provide the information for several different routing schemes. The river routing scheme is then described fully in the model structure (Section 2.3.4) as this is the current routing scheme implemented in the model.

We have modified section 2.2.2 to guide the reader better:

"From the river network and gauge locations, the river network connectivity is derived with each river section labelled with a unique river ID and a suite of routing tables so that each ID knows it's downstream connections and to allow multiple routing schemes to be configured (see section 2.3.4 for a description of the current routing scheme implemented in the modelling framework)."

(7) P7, L24: "potential evapotranspiration": first, this term is used here and then abbreviated several times later. Second, why is PET required for rainfall-runoff modeling? Is it to calculate the actual ET? If yes, where is such description provided?

We have removed all abbreviations of potential evapotranspiration in the document. Potential evapotranspiration is a common input for hydrological models and is used to calculate actual ET. The description of how this is calculated is given by Equation 2 in Section 2.3.4, but the model could also include other conceptualisations depending on the users requirements.

(8) P7, L40: why and how was the 1mm/day set?

The model needs a starting flow to initialise the storage deficits. Typically we take this from an observed flow time series but in some cases, particularly for ungauged flow points, there may be no flow time series available. In this case we define a starting flow of 1mm/day as a representative starting flow for most catchments. The choice of this initial starting flow only affects model flows during the initialisation period and has no effect once the flows are fully initialised. The model is always started with a 'spin up' period as would be normal standard practice.

(9) P7, L43: what are the "internal states"? Some examples should be provided.

We have modified this to *"model stores and fluxes"* **to better clarify this point. These are described in full in Section 2.3.4.**

(10) P7, L45: How are runoff generation, infiltration, and soil moisture movement modeled? Are they done in the same manner as in the original TOPMODEL?

These processes are described fully in Section 2.3.5 focused on the model structure. We have made clearer the differences between Dynamic TOPMODEL and DECIPHeR in Section 2.1 in response to Reviewer 2.

(11) P8, L15: What does the "multiple different" refer to?

In this case it refers to the model structures i.e. you can implement many different types of model structure within the model framework. We have removed the word 'different' to better clarify this point.

(12) P8, L24: How is SRmax determined?

SRmax is a parameter within the model that determines the soil root zone. The user can either set this to a default value or it can be sampled from parameter bounds as explained in section 2.3.3.

(13) P9, L6: "kinematic wave" formulation: is this sufficient when applying the model over large continents where backwater flow and other river-flood dynamics are important (see: Bates et al. 2010; Miguez-Macho and Fan 2012; Yamazaki et al. 2011; Zhao et al. 2017).

We believe the reviewer has mis-interpreted the routing used in the model. Channel river flow routing in the model is modelled using a set of time delay histogram. We agree that using a set of time delay histograms may not be appropriate where backwater flow and other river-flood dynamics may be important. However, we would like to stress that the model is not intended to be a flood inundation model and is not trying to compute full hydrodynamics. Computation times for such models are significantly longer than here (hours rather than minutes for simulations over 30-40 years).

The model is flexible to accommodate other flow routing schemes (as discussed in Section 2.2.2) and allow for variability in channel routing at the reach scale to recognise changes in local routing velocities. This will certainly be an area of future research to

improve the channel river flow routing. We have also recently coupled the model to LISFLOOD-FP to provide a better representation of river-flood dynamics in regions where this is important.

(14) P9, L42: “evapotranspiration losses are highest . . ..”: The figure shows PET, not the actual ET, and I believe high PET doesn’t necessarily mean high ET (in water limited regimes). I think this argument is not supported unless the actual ET is shown. Could the authors clarify this?

We have modified the text to clarify the data are potential evapotranspiration and not actual evapotranspiration.

(15) P10, L32-L42: Is the river network map described consistent with the topography data described in the previous paragraph? Isn’t it necessary to generate a river network map from the DEM used in the model?

Yes, we ensure consistency between the river and the DEM by producing the river network used by the model from the DEM during the digital terrain analysis. As described in section 3.2, we extract headwater cells from an external river map (the Ordnance Survey MasterMap Water Network Layer) and then route these cells downstream via the steepest slope so that the DEM and the calculated stream network are consistent for flow accumulations based on surface slope. Consequently, it is generated from the DEM used in the model and thus consistent. We have better clarified this point in Section 3.2 to avoid confusion.

(16) Section 3.3.1: Are the precip data used here same as those shown in Fig. 3.1?

We believe the reviewer is referring to Figure 4a here. The data used to derive the hydro-climatic characteristics are the same as the model forcing data described in Section 3.3.1. We have made this point clearer in the text.

(17) Section 3.3.2: What are the calibration and validation periods?

As described in Section 3.3.1, daily data of precipitation, potential evapotranspiration and discharge for a 55-year period from 01/01/1961–31/12/2015 were used to run and assess the model. The year 1961 was used as a warm-up period for the model; therefore no model evaluation was quantified in this period and the model was evaluated from 01/01/1962 – 31/12/2015. In this study we don’t use a split sample test with a calibration and validation period and instead choose to evaluate the model for the full time series available.

(18) P11, L14-25: what is the use of PET here? In fact, it was not clear to me on what the forcing variables are. Typical hydrological models use Precip, Temp, Radiation, Humidity, Wind etc. If such variables are used, is the PET consistent with those forcing variables?

Potential evapotranspiration is required as a forcing time-series for DECIPHeR to calculate actual evapo-transpiration. This is described in Section 2.3.2 which outlines the Data Pre-requisites of the model and Equation 2 in Section 2.3.5 which describes how potential evapotranspiration is used to calculate actual evapotranspiration.

We do not fully agree that ‘typical’ hydrological models use a range of data to construct PET internally. There are many hydrological models that use PET directly. Given there is considerable differences in how to construct PET then we often use more than one method to explore differences.

In this study (as described in Section 3.3.1), daily potential evapotranspiration (PET) data were obtained from the CEH Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (CHESS-PE) (Robinson et al., 2016). This dataset consists of 1km^2 gridded estimates of daily potential evapotranspiration for Great Britain from 1961 - 2015 calculated using the Penman-Monteith equation and data from the CHESS meteorology dataset (in this case air temperature, specific humidity, downward long- and shortwave radiation and surface air pressure). Consequently, potential evapotranspiration is calculated before being used as an input to DECIPHeR (as is common for many hydrological models).

(19) Section 3.4.3 (P13, L33): The authors should present the actual time streamflow time series. Since this is the only the variable simulated/discussed, I was surprised that authors are not showing the time series plots. I suggest selecting certain representative gauging stations with varying catchment area and those located in different climatic regions for such analysis (it could be a 20 stations for example).

Thank you for this suggestion. We have added a new figure (Figure 9) and text (Section 3.4.5) to the manuscript that shows the flow time series results for six gauging stations.

(20) Then, I also suggest showing the annual mean flow (rate or volume) as a scatter diagram for all gauging stations. Evaluation of high (Q5) and low (e.g., Q95) can also be presented similarly. Overall, the validation provided in the current version is not satisfactory/sufficient.

The model is evaluated against a large sample of catchments (1,366) for a number of different metrics capturing the annual flow rate (bias in runoff ratio), low flows (bias in low flow volume) and high flows (nash-sutcliffe efficiency). While we appreciate the reviewer's suggestions, we present results that already evaluate the model's ability to capture these aspects of the flow regime (see Figure 7). The main aim of the paper is to provide a description of DECIPHeR and more detailed model evaluation is outside the scope of this paper.

(21) P14, L23: "time series": where is this shown?

This is now shown in the new figure 9.

(22) P14, L39-45: The authors could discuss the appropriateness of different performance measures by referring some recent studies that have used a wide range of such performance measures (Veldkamp et al. 2018; Zaherpour et al. 2018). This comment is relevant to P12, L5-15 as well.

Thanks for the suggestions. We have added these references in section 4.1.

(23) P15, L23: "groundwater dynamics and human influences": Is the HRU-based representation a suitable choice for the representation of these missing factors? Would a fully distributed be required? Please also see a related comment earlier.

Please see response to comment 4 above.

(24) Finally, the authors should provide caveats in the current framework and the challenges in upscaling the framework to continental and possibly to global scales. The discussion regarding advancements compared to the existing models/ongoing efforts (e.g., the National Water Model) also becomes relevant here. A note on the use on the use HRUs, and not distributed grids, should also be made.

Section 4.2 and 4.3 discuss the limitations of the modelling framework as applied in this study and the areas for future research. These limitations are very relevant to applying the framework across continental scales and we have made this clearer in the discussion.

Minor/editorial issues:

(25) P2, L2: impact on “what”?

We have modified sentence

(26) P2, L8: some refs contain first names/initials

We have removed the first name and initial from this reference

(27) P11, L8: PET is abbreviated here but already used before.

We have modified to include only the abbreviation

(28) P12, L32: the catchment details are redundant with the information in Section 3.

We disagree and believe these catchment details are essential information.

References:

Bates, P. D., M. S. Horritt, and T. J. Fewtrell, 2010: A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. Journal of Hydrology, 387, 33-45.

Chaney, N. W., P. Metcalfe, and E. F. Wood, 2016: HydroBlocks: a fieldscale resolving land surface model for application over continental extents. Hydrol. Process., 30, 3543-3559.

Hanasaki, N., S. Kanae, T. Oki, K. Masuda, K. Motoya, N. Shirakawa, Y. Shen, and K. Tanaka, 2008: An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing. Hydrol. Earth Syst. Sci., 12, 1007-1025.

Miguez-Macho, G., and Y. Fan, 2012: The role of groundwater in the Amazon water cycle: 1. Influence on seasonal streamflow, flooding and wetlands. J. Geophys. Res. Atmos., 117, D15113.

Nazemi, A., and H. S. Wheater, 2015: On inclusion of water resource management in Earth system models – Part 1: Problem definition and representation of water demand. Hydrol. Earth Syst. Sci., 19, 33-61.

Ozdogan, M., M. Rodell, H. K. Beaudoing, and D. L. Toll, 2010: Simulating the Effects of Irrigation over the United States in a Land Surface Model Based on Satellite-Derived Agricultural Data. Journal of Hydrometeorology, 11, 171-184.

Pokhrel, Y., N. Hanasaki, Y. Wada, and H. Kim, 2016: Recent progresses in incorporating human land–water management into global land surface models toward their integration into Earth system models. WIREs Water, 3, 548-574.

Pokhrel, Y., S. Koirala, P. J. F. Yeh, N. Hanasaki, L. Longuevergne, S. Kanae, and T. C5 Oki, 2015: Incorporation of groundwater pumping in a global Land Surface Model with the representation of human impacts. Water Resources Research, 51, 78-96.

Veldkamp, T. I. E., F. Zhao, P. J. Ward, H. de Moel, J. C. Aerts, H. M. Schmied, F. T. Portmann, Y. Masaki, Y. Pokhrel, and X. Liu, 2018: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. Environmental Research Letters, 13, 055008.

Wada, Y., D. Wisser, and M. F. P. Bierkens, 2014: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. Earth Syst. Dynam., 5, 15-40.

Wada, Y., M. F. P. Bierkens, A. de Roo, P. A. Dirmeyer, J. S. Famiglietti, N. Hanasaki, M. Konar, J. Liu, H. Müller Schmied, T. Oki, Y. Pokhrel, M. Sivapalan, T. J. Troy, A. I. J. M. van Dijk, T. van Emmerik, M. H. J. Van Huijgevoort, H. A. J. Van Lanen, C. J. Vörösmarty, N. Wanders, and H. Wheater, 2017: Human–water interface in hydrological modelling: current status and future directions. Hydrol. Earth Syst. Sci., 21, 4169-4193.

Yamazaki, D., S. Kanae, H. Kim, and T. Oki, 2011: A physically based description of floodplain inundation dynamics in a global river routing model. Water Resources Research, 47, W04501.

Zaherpour, J., S. N. Gosling, N. Mount, H. M. Schmied, T. I. E. Veldkamp, R. Dankers, S. Eisner, D. Gerten, L. Gudmundsson, and I. Haddeland, 2018: Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environmental Research Letters.

Zhao, F., T. I. Veldkamp, K. Frieler, J. Schewe, S. Ostberg, S. Willner, B. Schauberger, S. N. Gosling, H. M. Schmied, F. T. Portmann, G. Leng, M. Huang, X. Liu, Q. Tang, N. Hanasaki, H. Biemans, D. gerten, Y. Satoh, Y. Pokhrel, T. stacke, P. ciais, J. Chang, A. Ducharne, M. Guimberteau, Y. Wada, H. kim, and D. Yamazaki, 2017: The critical role of the routing scheme in simulating peak river discharge in global hydrological models. Environmental Research Letters, 12, 075003.

Reviewer #2

General comments

The authors extended the concept and code of Dynamic TOPMODEL and developed an improved model termed DECIPHeR v1. They applied it to the entire Great Britain by calibrating and validating at 1366 gauges, and claimed that the performance was satisfactory. As a hydrological modeler who has developed open source code and applied it to an extensive study domain, I fully acknowledge the considerable efforts the authors made. The paper is overall readable for most parts but seems lacking some important statements particularly on the novelty and originality. The key characteristics and strengths of DECIPHeR should be clearly stated in comparison with existing catchment and global hydrological models (the current form of paper only compares DECIPHeR with the original Dynamic TOPMODEL). Also the value and significance of the model application to the entire Great Britain should be further discussed (the current form displays the performance scores without referring any earlier efforts).

We thank Reviewer #2 for taking the time to review the paper. We appreciate their comments and provide responses below.

Specific comments

Page 1 Line 15: “a new flexible model framework”: Make this part more specific. What is a flexible model framework (or what is an inflexible model)? Also, add the key strengths and characteristics of DECIPHeR compared to existing hydrological models.

We have modified the abstract to be more specific on the flexibility of the model framework (see response to comment below).

The key strengths and characteristics of DECIPHeR compared to existing hydrological models is now better discussed in the introduction in response to the comments of Reviewer #1 – we feel this discussion is more appropriate in the introduction rather than the abstract.

Page 1 Line 18: “modified to represent different levels of heterogeneity, connectivity and hydrological processes as needed”: Make this part more specific. All models can be “modified to represent” these in some extent. Add more concrete words in what sense DECIPHer is more adaptable compared with other models.

We have modified the abstract to be more specific on the flexibility/adaptability of the model framework.

“This paper presents DECIPHeR (Dynamic fluxEs and ConnectIvity for Predictions of HydRology); a new model framework that simulates and predicts hydrologic flows from spatial scales of small headwater catchments to entire continents. DECIPHeR can be adapted to specific hydrologic settings and to different levels of data availability. It is a flexible model framework which has the capability to (1) change its representation of spatial variability and hydrologic connectivity by implementing hydrological response units in any configuration, and (2) test different hypotheses of catchment behaviour by altering the model equations and parameters in different parts of the landscape.”

Page 2 Line 30 “the underlying model structures do not have the flexibility to represent different levels of complexity in different landscapes”: Quite unclear. Since this part is crucially important to identify the research needs/questions, discuss concretely what have been already achieved and what are still lacking by earlier models.

We agree this could be made clearer and have removed this sentence. We have significantly rewritten the introduction in response to comments from both reviewers to clarify the novelty of DECIPHeR and it's differences to other modelling frameworks.

Page 2 Line 42: "This is despite significant development of various modeling tools . . .": Again quite unclear. What have been already achieved and what are still lacking by earlier models?

See response to comment above

Page 3 Line 36 "builds on the code and key concepts of Dynamic TOPMODEL.": This sounds that DECIPHeR is an upgrade of Dynamic TOPMODEL. If this is the case, it is more readable to introduce the concept and formulations of Dynamic TOPMODEL first, then show the new functions and characteristics of DECIPHeR. Actually, the present form is hard to know what are same or different between two models.

We agree this could be made clearer. We have rewritten Section 2.1 to ensure this point is clarified. As suggested by the reviewer we now introduce the key concepts of Dynamic TOPMODEL first and then make clear the changes we have made to the model code.

Page 4 Line 18 "To realise this, DECIPHeR uses hydrological response units (HRUs)": It is hard to know whether the HRV concept has been already included in Dynamic TOPMODEL or not. I was confused similarly by many parts in this section. As mentioned earlier, please make it clear what are same or different between two models more clearly.

See response to comment above.

Page 6 Line 9 "In DECIPHeR, they provide the basis for river routing . . .". Ibid.

As now made clear in Section 2.1, the river routing code is completely new so this is unique to DECIPHeR.

Page 8 Line 12 "2.3.5 Model Structure": Unfortunately, I could hardly understand the model structure. Please describe all the equations for the terms in Figure 3 and the parameters in Table 1. At least describe where such full description of equations is available.

We have modified this section to provide a better description of all the key equations and parameters shown in Figure 3. We have also included the derivation of the new analytical solution for the subsurface zone in the appendix (see comments in general response).

Page 9 Line 9 "The parameter, SZM, sets..": This paragraph is particularly hard to follow. Please show the key equations how these parameters work.

We have modified this section and included the key equations for these parameters (see response to previous comment for modifications made to the manuscript).

Page 12 Line 44 "3.4.2 Overall model performance" and Figure 6: I am wondering why the parameters are so insensitive to the results (i.e. it is surprising that 90% of parameter sets yield NSE >0). I am also puzzled why the entire ensemble outperforms the behavioral ensemble (top 1% performance, if I understood correctly). Please elaborate these points.

We believe that Reviewer #2 has misinterpreted parts of these results. Figure 6 shows the percentage of catchments that meet the weaker and stricter performance thresholds for each catchment. Consequently, 90% of catchments yield NSE > 0, not 90% of the parameter sets (the number of parameter sets that achieve a score of NSE > 0 varies significantly between catchments). We have modified text in section 3.4.2 to make this clearer.

The 'best score' from the entire ensemble for any given metric is likely to outperform the best score from the behavioural ensemble as the behavioural ensemble is the top 1% based on the combined score of the four metrics. When creating a combined score of the four metrics, you would expect some trade offs between the different metrics as any simulation is unlikely to have the best score for all four metrics.

Page 14 Line 21 "We calculated four evaluation metrics for 10,000 model simulations for 1366 GB gauges. . .": Is this the first study to apply a hydrological to the entire Great Britain? If it is the case, clearly state so. If it is not, clearly refer the earlier efforts and compare the performance of them with this study.

This isn't the first study to apply a hydrological model to the entirety of Great Britain. However, it is the first to have such a comprehensive model evaluation against 1,366 gauges. We have made this clearer in the discussion and included a comparison of our model performance against other GB model evaluations in Section 4.1.

Technical comments

Page 7 Line 27 "a parameter file specifying set parameter bounds for Monte-Carlo sampling": Is "set" needed?

Agree. We will remove 'set'.

Page 7 Line 42 Q_SAT: I guess this term first appears. Define what this term is.

We have modified this to ***"used as the starting value for QSAT (subsurface flow)"***

Page 12 Line 37: "13,600,600" reads 13,660,000.

Thanks for spotting this. We have modified the text.

Page 13 Line 4"The vast majority of gauges (90% of the whole ensemble)": 90% of the gauges or 90% of the ensemble (i.e. 9000 simulations)?

Apologies that this was not clear. We meant 90% of the gauges and have modified the text (see response to Page 12 Line 44 above).

Page 14 Line 27: "is" reads in.

We have modified the text as suggested.

Page 34 figure 6: The caption says "weaker and stricter" while the figure says "upper and lower".

Thanks for spotting this. We have modified the legend in the figure so it says weaker and stricter thresholds.

DECIPHeR v1: Dynamic fluxEs and ConnectIvity for Predictions of HydRology

Gemma Coxon[1,2], Jim Freer[1,2], Rosanna Lane[1], Toby Dunne[1], Wouter J. M. Knoben[3], Nicholas J. K. Howden[2,3], Niall Quinn[4], Thorsten Wagener[2,3], Ross Woods[2,3]

[1]Geographical Sciences, University of Bristol, Bristol, United Kingdom, BS8 1SS
[2]Cabot Institute, University of Bristol, Bristol, United Kingdom, BS8 1UJ
[3]Department of Civil Engineering, University of Bristol, Bristol, United Kingdom, BS8 1TR
[4]Fathom Global, The Engine Shed, Station Approach, Bristol, United Kingdom, BS1 6QH

Correspondence to: Gemma Coxon (gemma.coxon@bristol.ac.uk)

Abstract. This paper presents DECIPHeR (Dynamic fluxEs and ConnectIvity for Predictions of HydRology); a new ~~flexible~~ model framework that simulates and predicts hydrologic flows from spatial scales of small headwater catchments to entire continents. DeECIPHeER can be adapted to specific hydrologic settings and to different levels of data availability ~~available data.~~ It is a flexible model framework which has the capability to (1) change its representation of spatial variability and hydrologic connectivity by implementing hydrological response units in any configuration, and (2) test different hypotheses of catchment behaviour by altering the model equations and parameters in different parts of the landscape. ~~and modified to represent different levels of heterogeneity, connectivity and hydrological processes as needed.~~ It has an automated build function that allows rapid set-up across ~~required~~ large model domains and is open source to help researchers and/or practitioners use the model. DEeCIPHeER is applied across Great Britain to demonstrate the model framework.~~, and~~ It is evaluated against daily flow time series from 1,366 gauges for four evaluation metrics to provide a benchmark of model performance. Results show the model performs well across a range of catchment characteristics but particularly in wetter catchments in the West and North of Great Britain. Future model developments will focus on adding modules to DECIPHeR to improve the representation of groundwater dynamics and human influences.

Commented [GC1]: RC2 Page 1 Line 15: "a new flexible model framework": Make this part more specific. What is a flexible model framework (or what is an inflexible model)?

RC2 Page 1 Line 18: "modified to represent different levels of heterogeneity, connectivity and hydrological processes as needed": Make this part more specific. All models can be "modified to represent" these in some extent. Add more concrete words in what sense DECIPHer is more adaptable compared with other models.

1 Introduction

Water resources require careful management to ensure adequate potable and industrial supply, to support the economic and recreational value of water, and to minimise the impacts of hydrological extremes such as droughts and floods on the economy, river ecosystems and human life. Robust simulations and predictions of river flows are increasingly needed across multiple temporal and spatial scales to support such management strategies (Wagener et al., 2010) that may range from the assessment of local field-scale flood mitigation measures to emerging water challenges at regional to continental scales (Archfield Stacey A. et al., 2015). Such approaches are particularly important, indeed mandated, given national and international policies on water management, such as the European Union's Water Framework Directive (EC, 2000) and Floods Directive (EC, 2007). Specifically (inter)national information on water resources, low and high flows is needed to underpin robust environmental management and policy decisions. This requires the effective integration of field observations and numerical modelling tools to provide tailored outputs at gauged and ungauged locations across a wide range of scales relevant to policy makers and societal needs. ~~Consequently there is a pressing need to develop new flexible modelling tools that can be applied to a range of space- and time- scales, and that are based on general hydrological principles applicable to a broad spectrum of different catchment types.~~

To address this need, a fundamental challenge for hydrologic sciences is to develop hydrological models that represent the complex drivers of catchment behaviour, such as space- and time- varying climate, land cover, human influence etc. (Blöschl and Sivapalan, 1995). The hydrologic community has made substantial investments to develop and apply hydrological models over the past 50 years to produce simulations and predictions of surface and groundwater flows, evaporation and soil moisture storage across ~~at~~ multiple scales. These include gridded approaches (e.g. PCR-GLOBWB, (Wada et al., 2014); VIC, Hamman et al., 2018; Liang et al., 1994; Grid-to-Grid, (Bell et al., 2007); Multiscale Hydrologic Model (Samaniego et al., 2010); DK-model, (Henriksen et al., 2003)), semi-distributed approaches that aggregate the landscape into hydrologic response units or sub-catchments (e.g. HYPE, (Lindström et al., 2010); SWAT, (Arnold et al., 1998); Topnet, Clark et al., 2008a) and many conceptual models applied at the catchment scale (Beven and Kirkby, 1979; Burnash, 1995; Coron et al., 2017; Leavesley et al., 1996; Lindström et al., 1997; Zhao, 1984). The current generation of hydrological models can represent a range of natural and anthropogenic ~~physical~~ processes and various levels of spatial complexity. Furthermore, there are significant ongoing efforts to represent spatial heterogeneity at finer scales over national-global scales (Bierkens et al., 2015; Wood et al., 2011) and build multi-model frameworks, to test competing hypotheses of catchment behaviour, such as FUSE (Clark et al., 2008a) and SUPERFLEX (Fenicia et al., 2011; Kavetski and Fenicia, 2011).

However, whilst these models have provided a wealth of useful insights and relevant outputs, ~~the underlying model structures do not have the flexibility to represent different levels of complexity in different landscapes. For example,~~ they either tend to: have a fixed representation of spatial variability (i.e. a single spatial resolution or a single spatial structure such as raster based); ~~have a~~ lack ~~of~~ spatial connectivity between hillslope-to-hillslope and hillslope and riverine components; be computationally expensive; and/or employ a single model structure applied homogenously across the model domain or nested catchment scale. This impacts our ability to apply models to a wide range of scales, places and water challenges, as different model representations of hydrological processes (i.e. model structure, parameterisations, hydrologic connectivity or spatial variability) are needed to capture

Commented [GC2]: RC1: The authors should revise the introduction to clearly highlight the motivation behind and the need for such a framework in relation to numerous other ongoing model development efforts. For example, how does the proposed study advance hydrological modeling compared to the model presented by Chaney et al. (2016)? Further, there are a number of large-scale models that have the capability to simulate far more number of processes (e.g., groundwater dynamics, pumping, flood dynamics, human impacts) than those presented in the current framework (for example: Hanasaki et al. 2008; Ozdogan et al. 2010; Pokhrel et al. 2015; Wada et al. 2014). Certainly these models are intended for global/regional applications but there have been ongoing efforts to increase the spatial resolution (i.e., hyper-resolution models) for application of these models at smaller scales. Extensive review of these models is available in recent literature (Nazemi and Wheater 2015; Pokhrel et al. 2016; Wada et al. 2017). I suggest that the authors thoroughly revise the introduction including a discussion on these past/ongoing efforts. Note that most of these models use TOPMODEL to simulate some of the surface/sub-surface hydrologic processes.

RC2: Page 2 Line 30 "the underlying model structures do not have the flexibility to represent different levels of complexity in different landscapes": Quite unclear. Since this part is crucially important to identify the research needs/questions, discuss concretely what have been already achieved and what are still lacking by earlier models.

RC2: Page 2 Line 42: "This is despite significant development of various modeling tools . . .": Again quite unclear. What have been already achieved and what are still lacking by earlier models?

Commented [GC3]: RC1 (25) P2, L2: impact on "what"?

heterogeneous hydrological responses and changing landscape connectivity, particularly for local conditions. Consequently, there is a pressing need to develop new spatially flexible modelling tools that can be applied to a range of space- and time- scales, and that are based on general hydrological principles applicable to a broad spectrum of different catchment types.

~~The key constraints to model development have been the lack of appropriate datasets and available computing power to run ensembles of complex simulations within reasonable timeframes. However, national and global open-source datasets are becoming readily available and significant gains in computing power have mitigated these constraints, to present the hydrological community with greater ability to build more flexible hydrological models that can resolve small-scale features in the landscape and be applied across continental scales to produce ensembles of flow simulations and predictions for a wide range of scales and water challenges. T~~ The need for such approaches is well documented in the literature (Clark et al., 2011, 2015; Mendoza et al., 2015) with calls for flexible hydrological modelling systems that can~~to~~: (1) incorporate different model structures and parameterisations in different parts of the landscape to represent a variety of processes; (2) change their spatial complexity, variability and/or hydrologic connectivity for hillslope elements and river network reaches (Beven and Freer, 2001; Mendoza et al., 2015); and, (3) be applied across a wide range of spatial and temporal scales, and across places (Blöschl et al, 2013). However, few such models exist.

In line with these requirements, ~~W~~we have ~~therefore~~ created a new ~~flexible~~ model framework, DECIPHeR (Dynamic fluxEs and ConnectIvity for Predictions of HydRology), ~~to be used~~ to simulate and predict hydrologic flows and connectivity from spatial scales of small headwater catchments to entire continents. The flexible modelling framework allows users to test ~~can be adapted to specific hydrologic settings and available data and can~~ different spatial resolutions, spatial configurations (i.e. gridded, semi-distributed or lumped), levels of hydrologic connectivity (i.e. representations of the lateral fluxes of water across model elements) and process representation (i.e. model structure and parameters). ~~be modified to represent different levels of heterogeneity, connectivity and hydrological processes as needed.~~ DECIPHeR has an automated build function that allows rapid set-up across required model domains with limited user input. The underlying code has been optimised to run large ensembles and enable model uncertainty to be fully explored. This is particularly important given inherent uncertainties in hydro-climatic datasets (Coxon et al., 2015; McMillan et al., 2012) and their impact on model calibration, regionalisation and evaluation (Freer et al., 2004; Kavetski et al., 2006; Kuczera et al., 2010; McMillan et al., 2010, 2011; Westerberg et al., 2016). We have specifically made the model code readable, reusable and open source to allow the broader community to learn from, verify and advance the work described here (Buytaert et al., 2008; Hutton et al., 2016).

In this paper, we: (1) describe the key capabilities and concepts that underpin DECIPHeR; (2) provide a detailed discussion of the model code and components; (3) demonstrate its application at the national scale to 1,366 catchments in Great Britain (GB); and, (4) discuss potential future model developments.

2 The DECIPHeR Modelling Framework

2.1 Key Concepts

The DECIPHeR modelling framework ~~builds on the code and~~ is based on the key concepts enshrined in Dynamic TOPMODEL ~~. Dynamic TOPMODEL was~~ originally introduced by

Commented [GC4]: RC2 Page 3 Line 36 "builds on the code and key concepts of Dynamic TOPMODEL.": This sounds that DECIPHeR is an upgrade of Dynamic TOPMODEL. If this is the case, it is more readable to introduce the concept and formulations of Dynamic TOPMODEL first, then show the new functions and characteristics of DECIPHeR. Actually, the present form is hard to know what are same or different between two models.

(Beven and Freer, 2001). Since its original development, ~~and has since~~Dynamic TOPMODEL has been applied in a wide range of studies (Freer et al., 2004; Liu et al., 2009, p.200; Metcalfe et al., 2017; Page et al., 2007; Younger et al., 2008) and integrated into other modelling frameworks (e.g. HydroBlocks, (Chaney et al., 2016). The core ideas of Dynamic TOPMODEL were three-fold (Beven and Freer, 2001); 1) to allow more flexibility in the definition of similarity in function for different points in the landscape, 2) to implement a non-linear routing of subsurface flow that simulates dynamically variable upslope subsurface contributing area and 3) to remain computationally efficient so that uncertainty in hydrological simulations can be estimated.

To realise this, Dynamic TOPMODEL uses hydrological response units (HRUs) to group raster-based information into non-contiguous spatial elements in the landscape that share similar characteristics (see Figure 1). Each HRU maintains hydrological connectivity in the landscape via weightings that determine the proportions of lateral subsurface flux from each HRU to all connected HRUs and flows to river cells. This solution offers key advantages in capability to traditional grid-based or lumped approaches employed by many hydrological models. Firstly, the user can split up the catchment using, for example, different landscape attributes (e.g. geology, land use) and/or spatially varying inputs (e.g. rainfall, evaporation, etc.) to define spatial similarity. This capability allows the user to modify the spatial complexity, resolution and/or hydrologic connectivity of hillslope elements and river network reaches in any configuration. Secondly, each HRU is treated as a separate functional unit in the model which can have different process conceptualisations and parameterisations. This means that more process complexity can be incorporated where needed to better suit local conditions (e.g. to account for 'point-source' human influences or more complex hydrological processes such as surface-groundwater exchanges). Finally, by grouping together similar parts of the landscape, HRUs minimise run times of the model compared to grid-based or fully distributed formulations, while still allowing model simulations to be mapped back into space.

While these key concepts that underpin Dynamic TOPMODEL address many of the challenges outlined in the introduction section, for the most part the model~~it~~ has only ever been applied to a single catchment or very simple nested ~~gauge networks~~catchments in headwater ~~catchments~~ basins (Peters et al., 2003). Consequently, we have completely restructured and rewritten the model code and added several new features to improve the flexibility and automation of the original Dynamic TOPMODEL code ~~made several key advances in flexibility and in automation~~ so the model can be applied from single small headwater catchments to regional, national and continental scales. These changes include:

1. Both legacy and new model ~~C~~code has been updated to a FORTRAN 2003 compliant version with new array and memory handling to allow~~ing~~ significantly larger and more~~,~~ complex gauging networks to be processed
2. The model build process is now fully automated to allow national/continental scale data to be easily and quickly processed, and to build and apply models in complex multi-catchment regions.
3. New model code and functions have been written to:
   a. Enable greater flexibility in the complexity and spatial characteristics of river network and routing properties. A newly developed river network scheme allows flow simulations to be produced for any gauged or ungauged point on a river network and segment river reaches into any length for individual hillslope-river flux contributions.

Commented [GC5]: RC2 Page 4 Line 18 "To realise this, DECIPHeR uses hydrological response units (HRUs)": It is hard to know whether the HRV concept has been already included in Dynamic TOPMODEL or not. I was confused similarly by many parts in this section. As mentioned earlier, please make it clear what are same or different between two models more clearly.

Commented [GC6]: RC1 (4) P4, L16-40: Why did the authors use HRUs instead of doing a fully-distributed model? Is it just the run time minimization? Is there a compromise in terms of adding new features such as groundwater flows and human water use? Again, I suggest adding a note on how this framework advances our capability to simulate the hydrology in comparison to numerous existing framework (see comments above)?

Commented [GC7]: RC2 Page 3 Line 36 "builds on the code and key concepts of Dynamic TOPMODEL.": This sounds that DECIPHeR is an upgrade of Dynamic TOPMODEL. If this is the case, it is more readable to introduce the concept and formulations of Dynamic TOPMODEL first, then show the new functions and characteristics of DECIPHeR. Actually, the present form is hard to know what are same or different between two models.

   b. Ensure that multiple points on the river network can be initialised via local storages and fluxes in each HRU successfully.
   c. Seamlessly facilitate DTA classification layers and results into rainfall-runoff model configuration that allows each individual HRU to have a different model structure, parameters, and climatic inputs.
4. A new analytical solution of the subsurface flow equations has been implemented, resulting in increased computational speed and numerical stability
5. The model can be easily adopted and adapted because it is open source, version controlled and includes a detailed user manual

The DECIPHeR modelling framework:

1. Can be easily applied to a single or multiple catchment(s), across large scales and complex river networks because it has an automated model building facility.
2. Can produce flow simulations for any gauged or ungauged point on a river network, and segment river reaches into any length for individual hillslope-river flux contributions.
3. Can (a) experiment with different landscape and climate attributes to characterise different spatial model structures and parameterisations and (b) modify the spatial scale/complexity of how spatial variability and hydrologic flow path connectivity are represented via its flexible modelling framework
4. Can easily allow different model hypotheses of catchment behaviour to be added because it is modular and extensible.
5. Can run large model ensembles to characterise model uncertainty because it is computationally efficient
6. Can be easily adopted and adapted because it is open source and includes a detailed user manual

DECIPHeR provides the capability to tackle fundamental hydrological modelling challenges and address general as well as site specific problems. To realise this , DECIPHeR uses hydrological response units (HRUs) to group together raster-based information into non-contiguous spatial elements in the landscape that share similar characteristics (see Figure 1). Importantly, these are not just fractions of the landscape with no explicit geographical location. Each HRU maintains hydrological connectivity in the landscape via weightings that determine the proportions of lateral subsurface flux from each HRU to all other connected HRUs, within itself and flows to river cells. The user can split up the catchment in any configuration to change the representation of spatial complexity and hydrologic connectivity using, for example, different landscape attributes (e.g. geology, land use) and/or spatially varying inputs (e.g. rainfall, evaporation, etc.) to define similarity. Each HRU is treated as a separate store in the model which can have different process conceptualisations and parameterisations so that more process complexity can be incorporated where needed to better suit local conditions (e.g. to account for 'point-source' human influences or more complex hydrological processes such as surface-groundwater exchanges). HRUs minimise run times of the model compared to grid or fully distributed based formulations and still allows model simulations to be mapped back into space.

HRUs are defined prior to rainfall-runoff modelling and DECIPHeR consists of two key steps where (1) digital terrain analyses are performed to define the gauge network, set up the river network and routing, discretise the catchment into HRUs and characterise the spatial variability and hydrologic connectivity in the landscape, and (2) HRUs are run in the rainfall runoff model to provide flow timeseries. These two steps are described in the following sections. More detailed descriptions of the input and output files, code workflows and codes can be found in the user manual.

2.2 Digital Terrain Analysis (DTA)

~~2.2.1 General Overview~~

The DTA in DECIPHeR constructs the spatial topology of the model components to define hillslope and riverine elements. The DTA defines the spatial extent of every HRU based upon multiple attributes, quantifies the connectivity between these HRU's in the landscape, determines the river network and all downstream routing properties, and determines the extent and where simulated output variables (i.e. discharge) should be produced (including gauged or ungauged locations) (see Figure 1).

~~While the DTA builds on the original Dynamic TOPMODEL code, several key advances have been made in the new version including:~~

1. ~~The process can be fully automated enabling multiple catchments and/or the processing of national scale data to be easily and quickly processed to build models in complex multi-catchment regions~~
2. ~~Code has been updated to a FORTRAN 2003 compliant version with new array and memory handling to allowing large, complex gauging networks to be processed in ways which is computationally efficient~~
3. ~~Greater flexibility in the complexity and spatial properties of river network and routing properties- separate flow pathway contributions of hillslope to river reach elements can be defined for any reach length~~

~~2.2.2~~2.2.1 Data Prerequisites

The minimum data requirement to run the DTA is a digital elevation model (DEM) and XY locations where flow time series are~~is~~ needed on the river network. The ~~digital elevation model~~DEM must contain no sinks or flat areas to ensure that the river network and catchments can be properly delineated.

Additional data can also be incorporated depending on data availability and modelling objectives. A river network can be supplied if the user wishes to specify headwater cells from a predefined river network and reference catchment areas and masks can be used to identify the best station location on the river network. Depending on user requirements, topographic, land use, geology, soils, anthropogenic and climate attributes can be supplied to define the spatial topology and thus differences in model inputs, structure and parameterisation.

~~2.2.3~~2.2.2 River Network, Catchment Identification and River Routing

DECIPHeR generates streamflow estimates at any point on the river network specified by the user. A river network is generated in DECIPHeR which matches the DEM flow direction and always connects to the boundary of the DEM or the sea. The river network is created from a list of headwater cells, which the functions can use/produce in three different ways depending on user requirements and/or data availability:

1. A list of pre-defined headwater (i.e. starting) river locations read into the DTA algorithms from a file
2. Headwater cells are found from a pre-defined river network

3. Where no pre-defined river network or headwater locations are available, then headwater cells are found from a river network which is derived from cells that meet thresholds of accumulated area and/or topographic index

Each headwater location is then routed downstream in a single flow direction via the steepest slope until reaching a sea outlet, other river or edge of the DEM, to construct a contiguous river network for the whole area of interest. Gauge locations are then generated on the river network from the point locations specified by the user. If a reference catchment mask or area is available, catchment masks are produced for candidate river cells found in a given radius and the catchment mask with the best fit to the reference mask or area is chosen as the gauge location. Otherwise the closest river cell is chosen as the gauge location.

Catchment masks are created from the final gauge list, with both individual masks for all the points specified on the river network and a combined catchment mask with the nested catchment masks created for use in the creation of the hydrological response units. From the river network and gauge locations, the river network connectivity is derived with each river section labelled with a unique river ID. A ~~and a~~ suite of routing tables is also produced so that each ID knows it's downstream connections and to allow multiple routing schemes to be configured (see section 2.3.4 for a description of the current routing scheme implemented in the modelling framework). These codes also provide the option of setting a river reach length where output time series can also be specified at different reach lengths between gauges (see Figure 1, HRU Setup D).

~~2.2.4~~2.2.3 Topographic Analysis

Topography, slope, accumulated area and topographic index are important properties of the landscape to aid the definition of hydrologic similarity and more dominant flow pathways. In DECIPHeR, they provide the basis for river routing and river network configuration and they also can be used to help determine the initial separation of landscape elements for defining hydrological similarity using percentiles of accumulated area, elevation and slope (in addition to alternative catchment attributes such as urban extent, geology, landuse, soils etc.).

Topographic index is calculated using the M8 multiple flow directional algorithm of (Quinn et al., 1995). The DTA calculates slope, accumulated area and topographic index for the whole domain. It uses the river mask to define the cells where accumulated area cannot accumulate downstream and the catchment mask to ensure accumulated area does not accumulate across nested catchment boundaries.

~~2.2.5~~2.2.4 Hydrological Response Units

The most critical aspect of running DECIPHeR is to define HRU's according to user requirements. The HRU configuration determines the spatial connectivity and complexity of model conceptualisation as well as the spatial variability of inputs and conceptual structure and parameters to be implemented in each part of the landscape. Any number of different spatial discretisations can be derived and subsequently applied in the DECIPHeR framework allowing the user to experiment with different model structures and parameterisations and modify representations of spatial variability and hydrologic connectivity.

In the DTA, hydrologically similar points in the landscape are grouped together so that each HRU is a unique combination of four different classification layers. These specify: (1) the initial separation of landscape elements from topographic information (e.g. slope, accumulated area and/or elevation); (2) inputs; (3) process conceptualisations; and (4) parameters implemented for each HRU store in the model (see Figure 2). These

Commented [GC8]: RC1 (6) Section 2.2.3: What is the routing scheme used? I find some description later in another section. Please consolidate the text and provide more details.

classification layers can be derived from climatic inputs, such as spatially varying rainfall and potential evapotranspiration, and landscape attributes such as geology, land use, anthropogenic impacts, soils data, slope, accumulated area. The simplest setup will consist of one HRU per catchment while the most complex can consist of a HRU for every grid cell (i.e. fully distributed).

To maintain hydrological connectivity in the landscape, the proportions of flow between the cells comprising each HRU are calculated based on accumulated area and slope. The flow fractions are then aggregated into a flow distribution matrix that summarises the proportions (weightings) of lateral subsurface flow from each HRU either to (1) itself, (2) another HRU or (3) a river reach. For *n* hydrological response units, the weights (*W*) are defined as:

$$W = \begin{pmatrix} w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,n} \end{pmatrix}$$

Equation 1~~1~~

Where each row defines how the HRU's output is distributed to other HRU's, any river reaches or itself and each column represents the total input to each HRU at every time step as the weighted sum of all the upstream outputs. Each row and column sum to one to ensure mass balance. The weights are detailed in a HRU flux file (which is fixed for a simulation) as a flow distribution matrix along with tabulated HRU attributes to provide information on which inputs, parameter and model structure type each HRU is using.

2.3 Rainfall-Runoff Modelling

~~2.3.1 General Overview~~

~~Once the DTA is complete, the rainfall-runoff modelling can then be executed. While DECIPHeR builds on the original Dynamic TOPMODEL code, several key advances were made in the new version including:~~

1. ~~Code has been updated to a FORTRAN 2003 compliant version with new array and memory handling to allow large datasets at any spatial scale and any complexity of gauging network to be processed to be commensurate with the DTA processing.~~
2. ~~New river routing algorithms added to allow streamflow estimates to be computed at any point on the river network and in any spatial configuration~~
3. ~~New initialisation code added to ensure that multiple points on the river network can be initialised via local storages and fluxes in each HRU successfully~~
   ~~New code to seamlessly facilitate DTA classification layers and results into rainfall runoff model configuration that allows each individual HRU to have a different model structure, parameters, and climatic inputs.~~

~~2.3.2~~2.3.1 Data Pre-requisites

To run the rainfall-runoff modelling component of DECIPHeR, time series forcing data of rainfall and potential evapotranspiration are required. Discharge data can also be provided for gauged locations and are used to initialise the model.

Besides forcing data, the model also needs, (1) the HRU flux file and routing files produced by the DTA, (2) a parameter file specifying ~~set~~ parameter bounds for Monte-Carlo sampling of parameters and (3) project/settings files specifying the number of parameter sets to run, which HRU and input file to use etc.

~~2.3.3~~2.3.2 Initialisation

Initialisation is an important step for any rainfall-runoff model. To ensure that subsurface flows, storages and the river discharge have all stabilised can be particularly problematic when modelling regionally over a large area as not all HRU's will initialise at the same rate (depending on size and slope characteristics).

A simple homogenous initialisation is currently implemented in DECIPHeR where the storage deficits for all HRU's are determined from an initial discharge. This is calculated as a mean area weighted discharge of the starting flows at timestep 1 for all output points on the river network. If a gauge does not have an initial flow, then the initial flow is either calculated from the mean of the data or set to a value of 1mm/day if no flow data is available. The initial discharge is assumed to be solely due to the subsurface drainage into the river so is used as the starting value for Q_{SAT} (subsurface flow) and to determine the associated storage and unsaturated zone fluxes. The model is then run for an initialisation period to allow its model stores and fluxes~~internal states~~ to fully stabilise with the catchment climatic information. Initialisation periods depend in part on~~to~~ the parameterisation of the model simulation run as well as the size and characteristics of the catchment being considered.

~~2.3.4~~2.3.3 Parameters

DECIPHeR can be run either using default parameter values or through Monte-Carlo sampling of parameters between set parameter bounds to produce ensembles of river flows. In the DTA, the user can set different parameter bounds for each HRU or sub-catchment thus specifying areas of the landscape where different parameter bounds may be~~are~~ needed. Alternatively, a single set of parameter bounds can be applied across the model domain.

For the model structure provided in the standard build and described below, there are seven parameters that can be sampled or set to default parameters. These parameters describe the transmissivity of the subsurface, the water holding capacity and permeability of soils and the channel routing velocity (see Table 1). More parameters can easily be added by the user if required for different model structures by changing the model source code.

~~2.3.5~~2.3.4 Model Structure

The description below details the model structure that is provided in the open source code (see Figure 3 and Table 1). While the code is built to be modular and extensible so that a user can easily implement multiple ~~different~~ model structures if so wished, the aim of this paper and the initial focus of the code development was on applying the model across large scales and beginning with a release that has relatively simple representations of the core processes. Thus, we provide a single model structure in the open source code that serves as a model benchmark to be built upon in future iterations.

The model structure consists of three stores defining the soil profile (Figure 4), which are implemented as lumped stores for each HRU. The first store is the root zone storage (S_{RZ}). Precipitation (P) is added to this store and then evapotranspiration (ET) is calculated and removed directly from the root zone. The maximum specific storage of S_{RZ} is determined by the parameter SR_{max}. Actual evapotranspiration from each HRU depends on the potential evapotranspiration (PET) rate supplied by the user and the root zone storage using a simple common formulation where evapotranspiration is removed at the full potential rate from saturated areas (i.e. if the root zone storage is full) and at a rate proportional to the root zone storage in unsaturated areas:

Commented [GC9]: RC2 Page 7 Line 42 Q_SAT: I guess this term first appears. Define what this term is.

Commented [GC10]: RC1 (9) P7, L43: what are the "internal states"? Some examples should be provided.

Commented [GC11]: Page 8 Line 12 "2.3.5 Model Structure": Unfortunately, I could hardly understand the model structure. Please describe all the equations for the terms in Figure 3 and the parameters in Table 1. At least describe where such full description of equations is available.

Commented [GC12]: RC1 (11) P8, L15: What does the "multiple different" refer to?

$$ET = PET * (S_{RZ}/SR_{max})$$

Equation 22

Once the root zone reaches maximum capacity (i.e. deficit of zero and conceptually analogous to field capacity), any excess rainfall input that is remaining is either added to the unsaturated zone (S_{uz}) where it is routed to the subsurface store or if this store is also full, Q_{EXUS} is added to the saturation excess storage (S_{EX}) and routed directly overland as saturated excess overland flow (Q_{OF}). The unsaturated zone links the S_{RZ} and saturated zones according to a linear function that includes a gravity drainage time delay parameter (Td) for vertical routing through the unsaturated zone. The drainage flux (Q_{uz}) from the unsaturated zone to the saturated zone is at a rate proportional to the ratio of unsaturated zone storage (S_{uz}) to storage deficit (S_D):

$$Q_{UZ} = S_{UZ}/(S_D * Td)$$

Equation 33

The dryness of the saturated zone is represented by the storage deficit. Changes to storage deficits for each HRU are dependent on recharge from S_{UZ} ~~S_{UZ}~~ (Q_{UZ} ~~Q_{UZ}~~), fluxes from upslope HRUs (Q_{IN}) and downslope flow out of each HRU (Q_{SAT} ~~Q_{SAT}~~), with subsurface flows for each HRU distributed according to the DTA flow distribution matrix described in section ~~4.2.5~~2.2.4.

$$\frac{dS_D}{dt} = Q_{SAT} - Q_{IN} - Q_{UZ}$$

Equation 44

~~Transfers between HRUs (and subsequently to the stream channel) are calculated using a kinematic wave formulation (see Beven and Freer, 2001; Li et al., 1975; Metcalfe et al., 2015) for downslope flow out of each HRU (Q_{SAT}) with both upslope (inputs) and local (for outputs) storages required.~~

~~The parameter, *SZM*, sets the form of the exponential decline in saturated zone hydraulic transmissivity with depth thereby controlling the shape of the recession curve in time. The parameter *ln(T0)* determines the lateral saturated hydraulic transmissivity at the point when the soil is saturated. The parameter, S_{max}, sets the saturated zone deficit threshold at which downslope flows between HRUs no longer occurs. If S_{max} has been reached then no downslope flow occurs and if the storage deficit is less than zero (the soil is at or above it's saturation capacity), then excess storage (Q_{EXS}) is added to saturation excess overland flow (Q_{OF}).~~

Where S_D is the current deficit in the saturated zone, Q_{SAT} is outflow from this HRU, Q_{IN} is inflow into the HRU representing drainage from the unsaturated zone of this HRU and Q_{UZ} is inflow into the HRU representing subsurface flow from other HRUs. This equation is solved sequentially for each HRU and provides values for the deficit S_D and outflow Q_{SAT} at time step *t* for each HRU. In DECIPHeR, this equation is solved analytically (see appendix for derivation of this solution), assuming a transmissivity profile that declines exponentially with depth and is truncated at depth S_{max} such that no flow is generated when the deficit is greater than S_{max} (Beven and Freer, 2001). The analytical solution provides better computational speed and increased numerical stability compared to the iterative 4-point numerical scheme described by Beven and Freer (2001).

Commented [GC13]: RC2 Page 9 Line 9 "The parameter, SZM, sets..": This paragraph is particularly hard to follow. Please show the key equations how these parameters work.

The exponential transmissivity profile takes the shape (Beven and Freer, 2001; eq. 6):

$$Q_{SAT} = T_0 \tan\beta \exp(-fz) = Q_0 exp(-\bar{S}/SZM)$$

Equation 5

The truncated exponential transmissivity profile takes the shape (rewritten from Beven and Freer, 2001; eq. 9):

$$Q_{SAT} = \begin{cases} Q_0 \cos\beta\,[exp(-\cos\beta\,\bar{S}/SZM) - exp(-\cos\beta\,S_{max}/SZM)] & \bar{S} \leq S_{max} \\ 0 & \bar{S} > S_{max} \end{cases}$$

Equation 6

Where β is the mean slope of the HRU and $\bar{S}$ is the average deficit across the HRU. The parameter, *SZM*, sets the rate of the exponential decline in saturated zone hydraulic transmissivity with depth thereby controlling the shape of the recession curve in time. The parameter, S_{max}, sets the saturated zone deficit threshold at which downslope flows between HRUs no longer occurs. If the storage deficit is less than zero (i.e. the soil is at or above its saturation capacity), then excess storage (Q_{EXS}) is added to saturation excess overland flow (Q_{OF}). Q_0 is the maximum rate of Q_{SAT} from a HRU when the HRU is at saturation and is calculated from:

$$Q_0 = \frac{T_0}{e^{\lambda}}$$

Equation 7

Where the parameter T_0 determines the lateral saturated hydraulic transmissivity at the point when the soil is saturated and λ is the average topographic index across the HRU.

Channel flow routing in DECIPHeR is modelled using a set of time delay histograms that are derived from the digital terrain analyses for the points where output is required. A fixed channel wave velocity (*CHV*) is applied throughout the network to account for delay and attenuation in the simulated flows (Q_{SIM}). DECIPHeR is a mass conserving model and therefore the model water balance always closes (subject to small rounding errors).

2.4 Model Implementation

The DECIPHeR model code is available on github (https://github.com/uob-hydrology/DECIPHeR) and is accompanied by a user manual which provides a detailed description of the file formats, how to run the codes and a code workflow. All the model code is written in FORTRAN for its speed, efficiency and ability to process large scale spatial datasets. Two additional bash scripts are provided as an example of calling the digital terrain analysis codes.

3 Great Britain National Model Implementation and Evaluation

While the modelling framework has a wide range of functionality, in this paper we wanted to demonstrate the ability of the model to be applied across a large domain to generate ensembles of flows at thousands of gauging stations and evaluate its current capability across large scales to guide future model developments. Consequently, we applied DECIPHeR to 1,366 gauges in Great Britain (GB) and in this section we describe the model setup, input data, evaluation criteria and model results.

3.1 Great Britain Hydrology

Catchments in Great Britain (GB) cover a wide hydrologic and climatic diversity. Hydro-climatic characteristics were derived from rainfall, potential evapotranspiration and flow data described in Section 3.3.1. Figure 4 shows the mean annual rainfall, mean annual potential evapotranspiration, runoff coefficient, ~~-~~and slope of the flow duration curve between the 30 and 70 flow percentiles for the 1,366 catchments in this study. Rainfall is highest in the West and North of GB and lowest in the East and South ranging from 540 to 3400 mm/year (Figure 4a), while potential evapotranspiration ~~losses are~~is highest in the East and South and lowest in the West and North ranging from 370 to 545 mm/year (Figure 4b). This regional divide of rainfall and ~~PET~~potential evapotranspiration is reflected in the runoff coefficients (Figure 4c) where generally runoff coefficients are lowest in the East and South and highest in the North and West. Slope of the flow duration curve (Figure 4d) is a more mixed picture across GB with lower values (i.e. a less variable flow regime) found in North-East Scotland, Midlands and patches of the South-East and higher values (i.e. a more variable flow regime) in the West, with the highest values for ephemeral and/or small streams in the South-East.

River flows vary seasonally with the highest totals generally occurring during the winter months when rainfall totals are highest and evapotranspiration totals are lowest, and the lowest totals during the summer months (April – September) resulting from lower precipitation totals and higher evapotranspiration losses due to seasonal variations in energy inputs. Snowmelt has little impact on river flows in GB except for some catchments in the Scottish Highlands where snowmelt contributions can impact the flows. River flow patterns are also heavily influenced by groundwater contributions from various regional aquifer systems. In catchments overlying the Chalk outcrop in the South-East of the GB, flow is groundwater-dominated with a predominantly seasonal hydrograph that responds less quickly to rainfall events. Land use and human influences also significantly impact river flows, with flows most heavily modified in the South-East and Midland regions of England due to high population densities.

3.2 Digital Terran Analyses for GB

To implement DECIPHeR across GB, the UK NEXTMAP 50m gridded digital elevation model was used as the basis of the Digital Terrain Analysis (Intermap, 2009). The first step was to ensure that the DEM contained no sinks or flat areas before being run through the DTA codes. Many freely available packages and codes exist to sink fill DEMs but for use with large national data sets, a two-stage process is often necessary to ensure no flat areas in the DEM and that important features, such as steep sided valleys, are not filled due to pinch points in the DEM. For this study, we first applied an optimised pit removal routine ((Soille, 2004), code available on github https://github.com/crwr/OptimizedPitRemoval). This tool uses a combination of cut and fill to remove all undesired pits while minimizing the net change in landscape elevation. We then applied a sink fill routine to ensure no flat areas remained in the DEM.

The inputs and outputs for the GB DTA is summarised in Figure 5. To build the river network, we first extracted headwater cells from ~~used~~ the Ordnance Survey MasterMap Water Network Layer; a dense national river vector dataset for GB. ~~This was used to extract headwater cells and a river network built by routing t~~These headwater cells were then routed downstream via the steepest slope to generate the river network used by the model. This ensures ~~so~~ that the DEM and the calculated stream network are consistent for flow accumulations based on surface slope. Locations of 1,366 National River Flow Archive gauges were used to define the gauging network and specify points on the river network

Commented [GC14]: RC1 (16) Section 3.3.1: Are the precip data used here same as those shown in Fig. 3.1?

Commented [GC15]: RC1 (14) P9, L42: "evapotranspiration losses are highest . . ..": The figure shows PET, not the actual ET, and I believe high PET doesn't necessarily mean high ET (in water limited regimes). I think this argument is not supported unless the actual ET is shown. Could the authors clarify this?

Commented [GC16]: RC1 (15) P10, L32-L42: Is the river network map described consistent with the topography data described in the previous paragraph? Isn't it necessary to generate a river network map from the DEM used in the model?

where output was required. We used NRFA catchment areas and masks as a reference guide to evaluate the best point for the gauge locations from potential river cell candidates within a local search area. Slope, accumulated area and the topographic index were then calculated for every grid cell and routing files produced.

Finally, we chose three classifiers to demonstrate the modelling framework while ensuring the number of HRUs was still computationally feasible for modelling across a large domain, these being:

1. The catchment boundaries for each gauge were used to ensure minimal fluxes across catchment boundaries.
2. A 5km grid for the rainfall and ~~PET~~potential evapotranspiration inputs was used to represent the spatial variability in climatic inputs across GB.
3. Three equal classes of slope and accumulated area were implemented resulting in HRU's that cascade downslope to the valley bottom.

3.3 Rainfall Runoff Modelling

3.3.1 Input and Evaluation Datasets

Daily data of precipitation, potential evapotranspiration ~~(PET)~~ and discharge for a 55-year period from 01/01/1961–31/12/2015 were used to run and assess the model. This period was chosen as an appropriate test for the model covering a range of climatic conditions and to demonstrate the model's ability to simulate long time periods within uncertainty analyses frameworks. The year 1961 was used as a warm-up period for the model; therefore no model evaluation was quantified in this period.

A national gridded rainfall and potential evapo-transpiration ~~(PET)~~ product was used as input into the model. Daily rainfall data were obtained from the CEH Gridded Estimates of Areal Rainfall dataset (CEH-GEAR) (Keller et al., 2015; Tanguy et al., 2016). This dataset consists of 1km^2 gridded estimates of daily rainfall from 1961 - 2015 for Great Britain and Northern Ireland derived from the Met Office UK rain gauge network. The observed precipitations from the rain gauge network are quality controlled and then natural neighbour interpolation is used to generate the daily rainfall grids. Daily potential evapotranspiration ~~(PET)~~ data were obtained from the CEH Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (CHESS-PE) (Robinson et al., 2016). This dataset consists of 1km^2 gridded estimates of daily ~~PET~~potential evapotranspiration for Great Britain from 1961 - 2015 calculated using the Penman-Monteith equation and data from the CHESS meteorology dataset. Both datasets were aggregated to a 5km grid as forcing for the national model run.

The model was evaluated against daily streamflow data for the 1366 gauges obtained from the National River Flow Archive (www.nrfa.ceh.ac.uk). This data is collected by measuring authorities including the Environment Agency (EA), Natural Resources Wales (NRW) and Scottish Environmental Protection Agency (SEPA) and then quality controlled before being uploaded to the NRFA site.

3.3.2 Model Structure and Parameters

To initially evaluate the model, DECIPHeR was run within a M~~m~~onte-C~~c~~arlo simulation framework whereby 10000 parameter sets were randomly sampled from a uniform prior distribution. This number of parameter sets was chosen to provide a reasonable sampling of the parameter space for demonstration purposes, however, for a full evaluation of the parameter space, more parameter sets would be needed.

These parameters were applied uniformly across the HRUs and used within a single model structure (as described in Section 2.3.4). Given the wide range of hydroclimatic conditions across GB, sampling of the feasible parameter space was ensured by using wide sampling ranges based on previous studies that have used Dynamic TOPMODEL (Beven and Freer, 2001; Freer et al., 2004; Page et al., 2007) (Table 2).

3.3.3 Model Evaluation

Daily time series of discharge for the 10,000 model simulations from each gauge were evaluated against daily observed flow for all 1,366 gauges. This is a challenging test for the model as these catchments cover a large range of hydrologic behaviour across GB and are impacted by a variety of climatic, geological and anthropogenic processes as outlined in Section 3.1. However, evaluating the model over such a large number of gauges acts as a benchmark of model performance and a means of identifying future areas for model development.

To benchmark model performance, we wanted to evaluate the model's ability to capture a range of hydrologic behaviour including maintaining overall water balance, capturing flow variability, reproducing low and high flows and the timing of flows. Consequently, multiple metrics, including hydrological signatures, standard hydrological model performance metrics and statistics of the flow time series were used to provide insights into model performance. Based on previous studies evaluating national scale models (McMillan et al., 2016) and considering a diagnostic approach to model evaluation (Coxon et al., 2014; Gupta et al., 2008; Yilmaz et al., 2008); four metrics were chosen which are summarised in Table 3 alongside their equations i) NSE (Nash and Sutcliffe, 1970), ii) Slope of the Flow Duration Curve (Yadav et al., 2007) iii) Bias in Runoff Ratio (Yilmaz et al., 2008) and iv) Low Flow Volume (Yilmaz et al., 2008).

These metrics are also used to determine a behavioural ensemble of parameter sets. The focus of this model application is to demonstrate the model can be run in a Monte Carlo framework. Consequently, while many different approaches could be used to determine a behavioural ensemble of parameter sets (see for example (Beven, 2006; Coxon et al., 2014; Krueger et al., 2010; Westerberg et al., 2011)), in this study we adopt a simple approach to produce ensembles of flows. The four metrics described above are combined and the behavioural ensemble was then taken as the top 1% of the model simulations according to this combined score. To calculate the combined score, each metric was ranked in turn, these ranks were summed, and all simulations sorted by the total combined rank. Weaker and stricter performance thresholds in NSE and bias metrics were also defined to further explore the performance of the ensembles against a common set of criteria (see Table 3). These were chosen based on previous studies and although subjective, the hydrological modelling community is yet to agree on benchmarks for the comparison of model performance (Seibert et al., 2018).

3.4 Results

3.4.1 Digital Terrain Analysis and Model Simulation

DECIPHeR was set up for GB covering a total catchment area of 154,763km^2 for 1366 gauges and 365 principal basins. Principal basin area ranged from 7.87km^2 to 9935km^2 with a median of 137km^2. Using the HRU classifiers specified in Section 3.2, the number of

HRUs contained within each principal basin ranged from 17 to 8978 with a median of 123 HRUs. HRU area ranged from 0.0025km^2 to 14.33km^2 with a median HRU area of 0.65km^2.

In total 13,6600,0~~6~~00 55 year time series, flow simulations were produced. One simulation over the 55 year time period for the largest river basin (9935km^2) with 8978 HRUs takes approximately 15 minutes to run on a standard CPU, outputting simulated discharge for all the 98 gauges that lie within the Thames at Kingston river basin. For the smallest river basin that has 17 HRUs and one river gauge, a single simulation over the 56 year time period on a standard CPU takes less than a second.

3.4.2 Overall Model Performance

Our first assessment of model performance is the overall model performance for the four performance metrics calculated from the 10000 simulated daily flow time series produced for each gauge. Figure 6 shows the percentage of catchments that met the stricter and weaker performance thresholds defined in Table 3 from the entire ensemble of 10000 model simulations and from the top 1% behavioural ensemble generated from the combined ranking of the four metrics. Our results show that most catchments are able to meet both the performance thresholds. The vast majority of catchments~~gauges~~ (92~~0~~%~~ for the whole ensemble~~) gain~~achieve~~ a NSE score greater than zero (i.e. better than mean climatology) and ~~many of the~~ 80% of the catchments ~~gauges (72% for the whole ensemble) achieve~~gain a NSE score greater than 0.5. The model does well in reproducing Low Flow Volumes and Slope of the Flow Duration Curve with most gauges (98~~5~~ and 96~~0~~% respectively) meeting the stricter performance threshold.

RRBIAS evaluates the model's ability to reproduce water balance in the catchment; the current implementation of the model has to maintain mass balance while many of the observed flow data for many of these catchments does not maintain mass balance either due to inter-catchment groundwater flows, anthropogenic influences such as surface and ground water abstractions, or data errors (this is further discussed in section 4.4.4). Consequently, RRBIAS is a more difficult metric for the model to capture and this is reflected by the fact that 75~~66~~% of the catchments meet the weaker threshold and just over 62~~50~~% meet the stricter threshold.

These numbers decrease slightly for the behavioural ensemble as expected due to trade-offs between the four metrics but the overall trends remain the same.

3.4.3 Spatial Model Performance

To analyse model performance spatially across GB, the four evaluation metrics for the best simulation (as defined by the combined rank across all four metrics) for each catchment is summarised in Figure 7.

For NSE, model performance is variable across the country but generally, better model performance is found in the wetter catchments in the North and West of GB, with poorer model performance in drier catchments in the South and East. Model performance is poor in groundwater dominated areas, particularly in the underlying chalk regions in the South East. This region has particularly low runoff coefficients (see Figure 4d) and does not maintain mass balance with large water losses. Consequently, results for RRBIAS shows that the model tends to over-estimate flows in the South-East. While bias in the runoff ratio shows the model is generally over-estimating flows, biases in the low flow volume is a more mixed picture with the model under-estimating low flows in some locations, particularly in the Midlands and North East Scotland. From Figure 4d, these areas are characterised by particularly low flow duration curve slopes suggesting strongly damped flow responses with

Commented [GC17]: RC2. Page 12 Line 37: "13,600,600" reads 13,660,000.

Commented [GC18]: RC2. Page 12 Line 44 "3.4.2 Overall model performance" and Figure 6: I am wondering why the parameters are so insensitive to the results (i.e. it is surprising that 90% of parameter sets yield NSE >0). I am also puzzled why the entire ensemble outperforms the behavioral ensemble (top 1% performance, if I understood correctly). Please elaborate these points.

RC2. Page 13 Line 4"The vast majority of gauges (90% of the whole ensemble)": 90% of the gauges or 90% of the ensemble (i.e. 9000 simulations)?

high baseflow. Flow in the Midlands region is heavily regulated by reservoirs which sustain low flows and could be a potential reason for over-estimating low flows in this area. The bias in slope of the flow duration curve shows DECIPHeR does well at reproducing the flow variability but tends to under-estimate the slope in Scotland and North Wales suggesting that the hydrographs in these catchments are too smooth and not sufficiently flashy.

3.4.4 Relationship Between Model Performance and Catchment Characteristics

To further analyse and understand the reasons for good/poor model performance, relationships between key catchment characteristics and model performance were further explored. Firstly, the catchments were grouped according to key catchment characteristics based on discharge; runoff coefficient and base flow index. The 5th, 50th and 95th percentiles of NSE and RRBIAS were calculated from the ensemble of runs for all catchments within each group to explore relationships between model performance and catchment characteristics (see Table 4). The relationship between runoff coefficient, wetness index and RRBIAS was also analysed to further explore the importance of water gains/losses on model performance.

There is a clear link between model performance and catchments with a low runoff coefficient. Table 4 highlights poor model performance in catchments where observed runoff coefficients are less than 0.2. In this group, the model always over-predicts (as shown by the RRBIAS result) and consequently leads to poor NSE scores. Figure 8 shows that for many catchments where the model over-predicts flows (and particularly for catchments with a runoff coefficient less than 0.2) observed potential evapotranspiration estimates are not high enough to account for water losses culminating in an over-estimation of flows. This is unsurprising given that currently the model maintains water balance and can't lose or gain water beyond the 'natural' conceptualisations of precipitation, discharge and evaporation dynamics. Consequently, we are either missing a process (such as water loss due to inter-catchment groundwater flows or anthropogenic impacts) or the data is wrong.

Poorer model performance is also found in high BFI catchments (Table 4), however, the results also show we can also gain very good simulations in these types of catchments (5th percentile has a NSE score of 0.83~~2~~), hence the challenge is to better understand water losses/gains in groundwater catchments ~~as the first step~~ to improve the representation of groundwater dynamics in the model.

3.4.5 Simulated Flow Time Series

Finally, we examined the simulated flow time series for six example catchments with different characteristics. Figure 9 shows the observed discharge, observed precipitation and the 5th-95th percentile uncertainty bounds of the behavioural simulations for six catchments with different characteristics (see Table 5) for a representative two-year period of the 55-year time series simulated. The 5th-95th percentile uncertainty bounds are generated from the likelihood-weighted distribution of the top 1% of the model simulations using the GLUE framework (Beven, 2006).

Our results show the model can capture a range of different hydrological dynamics from wetter catchments in the North-West (Figure 9a) to drier catchments in the South-East (Figure 9b). While model performance for groundwater catchments can be very good (Figure 9c and Table 5), it also shows that we need to incorporate additional model capability to simulate the dynamics of groundwater dominated catchments. Where we have a very low (for Great Britain) runoff coefficient, this is assumed to involve water losses into a more regional groundwater storage not expressed at the outlet and not yet represented in this

Commented [GC19]: RC1 (19) Section 3.4.3 (P13, L33): The authors should present the actual time streamflow time series. Since this is the only the variable simulated/discussed, I was surprised that authors are not showing the time series plots. I suggest selecting certain representative gauging stations with varying catchment area and those located in different climatic regions for such analysis (it could be a 20 stations for example).

version of the model (Figure 9d). While the catchments shown in Figure 9a-d are relatively un-impacted by human influences, the catchments shown in Figure 9e and 9f are heavily impacted by human influences and highlight the challenge of simulating flows nationally across catchments with diverse hydrological behaviour.

4 Outlook and Ongoing Developments

4.1 National Scale Model Evaluation

This is the first study to comprehensively benchmark hydrological model performance across GB. We calculated four evaluation metrics for 10,000 model simulations for 1,366 GB gauges to provide an initial benchmark of model performance. DECIPHeR generally performs well for the flow time series evaluated in this study, with better results in the West and North in wet catchments as compared to drier catchments in the South and East. This is a common finding for hydrological models, with many studies finding poor model performance and greatest water balance errors in drier catchments (Gosling and Arnell, 2011; McMillan et al., 2016; Newman et al., 2015; Pechlivanidis and Arheimer, 2015). These results are also reflected in other GB model evaluation studies. For example, (Coxon et al., 2014) applied FUSE to 24 GB catchments and found the best model performances in wet catchments compared to dry, chalk catchments, (Rudd et al., 2017) evaluated G2G for low flows across 61 GB catchments and found positive bias in low flow volumes in small catchments in the South-East of England and (Crooks et al., 2010) evaluated PDM across 120 GB catchments and found poorer model performance in groundwater dominated, drier catchments.

Commented [GC20]: RC2 Page 14 Line 21 “We calculated four evaluation metrics for 10,000 model simulations for 1366 GB gauges. . .”: Is this the first study to apply a hydrological to the entire Great Britain? If it is the case, clearly state so. If it is not, clearly refer the earlier efforts and compare the performance of them with this study.

Poor model performance in~~s~~ these catchments is partially due to some of the metrics chosen in this study, for example percent bias is most sensitive to small absolute biases in the driest catchments when compared to other metrics such as absolute bias. However, positive bias in the runoff ratio could be caused by a number of factors such as under-estimation of potential evapotranspiration (there are other UK gridded ~~PET~~potential evapotranspiration products which estimate much higher potential evapotranspiration), inter-catchment groundwater flows, and/or human influences such as water abstraction. Population density is much higher in the South and East compared to the North and West so this regional disparity in model performance could also be explained by a greater rate of abstractions and managed watercourses which alter the flow time series. For example, 55% of the effective rainfall in the Thames catchment is licensed for abstraction (Thames Water, 2017).

Commented [GC21]: RC2 Page 14 Line 27: “is” reads in

These results provide an initial test of DECIPHeR capabilities against a large sample of catchments, but this is only a first-order evaluation of model performance. A more rigorous evaluation would assess the model: over different seasons (Freer et al., 2004); under changing climatic conditions (Fowler et al., 2016); for different hydrological extremes (Coron et al., 2012; Veldkamp et al., 2018; Zaherpour et al., 2018); for multiple objectives simultaneously (Kollat et al., 2012); and, incorporate input and flow data uncertainty (Coxon et al., 2014; Kavetski et al., 2006; McMillan et al., 2010; Westerberg et al., 2016).

Commented [GC22]: RC1 (22) P14, L39-45: The authors could discuss the appropriateness of different performance measures by referring some recent studies that have used a wide range of such performance measures (Veldkamp et al. 2018; Zaherpour et al. 2018). This comment is relevant to P12, L5-15 as well.

4.2 Characterising Spatial Heterogeneity and Connectivity

The intended use of DECIPHeR is to determine how much spatial variability and complexity is required for a given set of modelling objectives. It can be run as a lumped model (1 HRU), semi-distributed (multiple HRUs) or fully gridded (HRU for every single grid cell). In this paper DECIPHeR was applied across 1,366 GB gauges, with catchment masks, 5 km input grids and three classes of accumulated area and slope as classifiers for the hydrological response units, resulting in a total of 133,286 HRUs. Future work needs to consider the appropriate spatial complexity and hydrologic connectivity needed to represent relevant processes (Andréassian et al., 2004; Blöschl and Sivapalan, 1995; Boyle et al., 2001; Chaney et al., 2016; Clark et al., 2015; Metcalfe et al., 2015; Wood et al., 1988). While this work highlights the clear potential of a computationally-efficient large-scale modelling framework that can run large ensembles, a balance is required to ensure computational efficiency when running large ensembles that also maintains sufficient spatial complexity to represent different hydrological processes.

4.3 Hypothesis Testing and Model Parameterisation

To demonstrate the modelling framework we implemented a single model structure, provided in the open source model code, in all HRUs across GB and did not experiment with different model structures in different parts of the landscape. This provides a good benchmark of DECIPHeR's ability at the national scale across GB, but the results suggest different model structures are needed to represent a greater heterogeneity of hydrological responses beyond the conceptual dynamics currently implemented in this simple model (as shown in Figure 9). We can gain new process understanding of regional differences in catchment behaviour by testing different model representations (Atkinson et al., 2002; Bai et al., 2009; Perrin et al., 2001). Future work will concentrate on adding modules to DECIPHeR to enhance performance across national and continental scales with a focus on improved representation of groundwater dynamic and human influences to address poor model performance in catchments with a low runoff coefficient. Furthermore, we have ensured the code is open-source and well-documented so that the hydrological community can contribute new/different conceptualisations of the processes shown in this paper. ~~We can gain new process understanding of regional differences in catchment behaviour by testing different model representations (Atkinson et al., 2002; Bai et al., 2009; Perrin et al., 2001).~~

Commented [GC23]: RC1 (24) Finally, the authors should provide caveats in the current framework and the challenges in upscaling the framework to continental and possibly to global scales. The discussion regarding advancements compared to the existing models/ongoing efforts (e.g., the National Water Model) also becomes relevant here. A note on the use on the use HRUs, and not distributed grids, should also be made.

It is challenging to parameterise a ~~national scale~~ hydrological model across large scales. Here we simply applied the same parameter set across each catchment. Using this basin-by-basin approach has the disadvantage of producing a "patchwork quilt" of parameter fields, with discontinuities in parameter values across catchment boundaries. This is only effective for gauged catchments (Archfield Stacey A. et al., 2015). Ongoing work aims to address these issues by implementing the multiscale parameter regionalisation (MPR) technique for DECIPHeR across GB. This technique links model parameters to geophysical catchment attributes through transfer functions applied at the finest possible resolution (Samaniego et al., 2010). The coefficients of the transfer functions are then calibrated, and parameters are upscaled to produce spatially consistent fields of model parameters at any resolution across the entire model domain. The MPR technique has been applied elsewhere, proving that it can produce seamless parameter fields across large domains and produce scale-invariant parameters (Kumar et al., 2013; Mizukami et al., 2017; Samaniego et al., 2017), which is ideal for a flexible framework such as DECIPHeR.

5 Conclusions

DECIPHeR is a new flexible modelling framework which can be applied from small catchment to continental scale for complex river basins resolving small-scale spatial heterogeneity and connectivity. The model is underpinned by a flexible, computationally efficient framework with a number of novel features:

1. *Spatial variability and connectivity* - ability to modify spatial variability and connectivity in the model via the specification of hydrological response units with different topographic, landscape, input layers

2. *Model structures and parameterisations* - ability to experiment with different model structures and parameterisations in different parts of the landscape
3. *Computationally efficient* - grouping of hydrologically similar points in the landscape into hydrological response units enables faster run times
4. *Automated bBuild* – to allow easy application over large scales
5. *Open source* - the open source model code is implemented in Fortran, with a user manual to help researchers and/or practitioners to use the model.

This paper describes the modelling framework and its key components and demonstrates the model's ability to be applied a large model domain. DECIPHeR is shown to be computationally efficient and perform well over large samples of gauges. This work highlights the potential for catchment to continental scale predictions, by making use of available big datasets, advances in flexible modelling frameworks and computing power.

Appendix A - Analytical Solution for Kinematic Subsurface Flow

1. Introduction

This appendix provides an analytical solution to the equations which were solved numerically by Beven and Freer (2001) in Dynamic TOPMODEL to route subsurface flows between HRUs. Here, we use calculus to integrate the relevant equations through time, as opposed to the finite volume scheme described by Beven and Freer (2001) for better computational speed and increased numerical stability.

This development starts from the kinematic wave description of flow (a partial differential equation), integrates that partial differential equation along the flow direction over the entire HRU to obtain an ordinary differential equation in time, and then integrates that ordinary differential equation in time to get an analytical solution which gives the flow and storage at the end of each timestep, in relation to the conditions at the start of the timestep, and the inflow from both upslope and from drainage. By first integrating the kinematic wave equation over the HRU, we have effectively chosen to model the flow in the HRU using a nonlinear reservoir, so there is no wave travelling across the HRU. This approach of integrating in space is the same as selected by Beven and Freer (2001), using their finite volume approach.

2. Flow in an HRU

Assume the HRU has area A, and that x is distance measured along the flow direction of the HRU. Define Q as the downslope flow rate [L^3/T] at some point x. Assume that the flow is kinematic, i.e. that Q depends only on S [L], the local storage deficit per unit area, and the HRU geometry. The drainage input from above is assumed to be r [L/T]. Assume the width of the HRU is $w(x)$, at distance x [L]. At any point x in the HRU we can write a partial differential equation for Q:

$$\frac{\partial Sw}{\partial t} = \frac{\partial Q}{\partial x} - rw \qquad \textbf{Equation A1}$$

This is a kinematic wave equation describing the subsurface flow at point x within an HRU. Note that both S and r have been multiplied by w, the width of the HRU, so that they can be compared with Q, which is the total flow through the HRU, at distance x.

To simplify the problem, we will now average over the entire HRU, along the flow direction, x, from the upslope end (x=0) to the downslope end (x=L). This will produce an equation describing how $\bar{S}$, the HRU-average of S, changes with time.

$$\frac{1}{L}\int_0^L \frac{\partial Sw}{\partial t}\mathrm{d}x = \frac{1}{L}\int_0^L \frac{\partial Q}{\partial x}\mathrm{d}x - \frac{1}{L}\int_0^L rw dx \qquad \textbf{Equation A2}$$

$$\frac{1}{L}\int_0^L \frac{\partial Sw}{\partial t}\mathrm{d}x = \frac{Q(L,t)-Q(0,t)}{L} - r\frac{1}{L}\int_0^L w dx \qquad \textbf{Equation A3}$$

The variables $Q(0,t)$ and $Q(L,t)$ refer to flows at the upslope and downslope ends of the HRU [L^3/T]

If we assume S and w are uncorrelated as x varies, and let $W = \frac{1}{\mathrm{L}}\int_0^{\mathrm{L}} \mathrm{w}dx$ then

$$W\frac{\partial \frac{1}{L}\int_0^L S\mathrm{d}x}{\partial t} = \frac{Q(L,t)-Q(0,t)}{L} - rW \qquad \textbf{Equation A4}$$

Dividing by W,

$$\frac{\partial \bar{S}}{\partial t} = \frac{Q(L,t)-Q(0,t)}{LW} - r \qquad \textbf{Equation A5}$$

Note that $A = \mathrm{LW}$ is the area of the HRU, so we can now define q=Q/A as flow per unit plan area [L/T] which is the same dimension as used by Beven and Freer (2001).

$$\frac{\partial \bar{S}}{\partial t} = q(L,t) - (q(0,t) + r) \qquad \textbf{Equation A6}$$

In equation 6, $q(0,t)$ and r are assumed to be known, and $q(L,t)$, the outflow from the HRU, is assumed to be a function of the mean deficit $\bar{S}$. Thus the HRU is being modelled as a nonlinear reservoir, where $\bar{S}$ is the state variable, the input is $q(0,t) + r$, and the outflow $q(L,t) = f(\bar{S}(t))$. Note that the inflow is now assumed to be applied as a spatially uniform flux within the HRU, rather than being applied at x=0. There is no representation of motion within the HRU.

Note that in the following equations, Q is equivalent to Q_{SAT} (eq. 4). Because no motion within the HRU is represented, Q_{IN} and Q_{UZ} (eq. 4) can be lumped together into a single term, here called r.

Analytical solutions for an exponential conductivity profile

There are several parsimonious descriptions of the vertical profile of saturated hydraulic conductivity which are hydrologically plausible. Here we consider the standard exponential profile and a profile truncated at finite depth. In each case we find the analytical solution for both $\bar{S}$ and $q(L,t)$ as functions of time. Analytical solutions are also possible for the parabolic and linear profiles given in Ambroise et al (1996).

Define $u = (q(0,t) + r)$ and $q = q(L,t)$

$$q = q_0 exp(-\bar{S}/m) \qquad \textbf{Equation A7}$$

$$\frac{\partial \bar{S}}{\partial t} = q - u \qquad \textbf{Equation A8}$$

If we substitute 7 into 8, and integrate 8 from $\bar{S}$ (0) at t=0 up to $\bar{S}$ (t), we obtain the intermediate result

$$exp(\bar{S}/m) = \frac{q_0}{u} + exp(\bar{S}(0)/m)\left(1 - \frac{q_0\, exp(-\bar{S}(0)/m)}{u}\right) exp\left(-\frac{ut}{m}\right) \qquad \textbf{Equation A9}$$

From this we can get expressions for both $\bar{S}$ (t) and q(t)

$$\bar{S}(t) = m\, log\left[\frac{q_0}{u} + \left(\frac{1}{exp(-\bar{S}(0)/m)} - \frac{q_0}{u}\right) exp\left(-\frac{ut}{m}\right)\right] \qquad \textbf{Equation A10}$$

$$q(t) = \left[\frac{1}{u} + \left(\frac{1}{q(0)} - \frac{1}{u}\right) exp\left(-\frac{ut}{m}\right)\right]^{-1} \qquad \textbf{Equation A11}$$

In the special case where u=0, we instead obtain

$$\bar{S}(t) = m \log\left(\exp(\bar{S}(0)/m) + \frac{q_0 t}{m}\right) \qquad \textbf{Equation A12}$$

$$q(t) = \left[\frac{1}{q(0)} + \frac{t}{m}\right]^{-1} \qquad \textbf{Equation A13}$$

Note that parameters m and q_0 in these equations are equivalent to SZM and Q_0 as defined in the paper.

Exponential truncated smoothly at Smax (Beven and Freer, 2001 equation 9)

$$q = \begin{cases} q_0 \cos\beta\, [exp(-\cos\beta\, \bar{S}/m) - exp(-\cos\beta\, S_{max}/m)] & \bar{S} \le S_{max} \\ 0 & \bar{S} > S_{max} \end{cases} \qquad \textbf{Equation A14}$$

$$q = \begin{cases} q_1 exp(-\cos\beta\, \bar{S}/m) - q_2 & \bar{S} \le S_{max} \\ 0 & \bar{S} > S_{max} \end{cases} \qquad \textbf{Equation A15}$$

Where $q_1 = q_0 \cos\beta$ and $q_2 = q_0 \cos\beta \exp(-\cos\beta\, S_{max}/m)$

Let's look first at the case where $\bar{S} \le S_{max}$

$$\frac{\partial \bar{S}}{\partial t} = q_1 exp(-\bar{S}/(m/cos\,\beta)) - (q_2 + u) \qquad \textbf{Equation A16}$$

If we let $m_2 = m/\cos\beta$ and $u_2 = q_2 + u$, then we can rewrite this as

$$\frac{\partial \bar{S}}{\partial t} = q_1 exp(-\bar{S}/m_2) - u_2 \qquad \textbf{Equation A17}$$

This is now exactly the same form as the exponential profile above, so the solution is formally identical: we just put q_1 instead of q_0, m_2 instead of m, and u_2 instead of u. The resulting equations are:

$$\bar{S}(t) = m_2 log\left[\frac{q_1}{u_2} + \left(\frac{1}{exp\left(-\frac{\bar{S}(0)}{m_2}\right)} - \frac{q_1}{u_2}\right) exp\left(-\frac{u_2 t}{m_2}\right)\right] \qquad \textbf{Equation A18}$$

$$q(t) = \left[\frac{1}{u_2} + \left(\frac{1}{q(0)} - \frac{1}{u_2}\right) exp\left(-\frac{u_2 t}{m_2}\right)\right]^{-1} \qquad \textbf{Equation A19}$$

This solution collapses to the standard exponential result if $\cos\beta = 1$ and $S_{max} = \infty$. Note that provided $S_{max} < \infty$, then $q_2 > 0$ so $u_2>0$, and there is no need to consider the case of zero forcing.

The deficit cannot go beyond S_{max} as a result of outflow; however deficits larger than S_{max} can arise through evaporation (this prepares for future developments, evaporation is not currently included in the conceptualisation of the saturated zone). Here we consider the case where $\bar{S} > S_{max}$, so q=0, but u>0, so the deficit is decreasing.

$$\frac{\partial \bar{S}}{\partial t} = -u \qquad \textbf{Equation A20}$$

This can be integrated to give

$$\bar{S}(t) = \bar{S}(0) - ut \qquad \textbf{Equation A21}$$

If $\bar{S}(t) \le S_{max}$ then we switch to the $\bar{S} \le S_{max}$ solution partway through the computational interval.

We use the equation for $\bar{S}$ (t) when $\bar{S} \le S_{max}$, because in that case the q(t) equation will lead to division by zero if it is started at q(0)=0.

Extra note on computational issues:

If u_2 is very small but not zero, numerical problems can arise in the calculation of $\bar{S}(t)$, because of loss of significance when subtracting two numbers of very different magnitudes. This can lead to calculating the logarithm of zero during calculation of $\bar{S}(t)$.

This can be avoided by making a Taylor series expansion of $\bar{S}$ (t) for small non-zero values of u_2. We obtain

$$\bar{S}(t) \cong m_2 log\left[\frac{q_1}{u_2} + \left(\frac{1}{exp\left(-\frac{\bar{S}(0)}{m_2}\right)} - \frac{q_1}{u_2}\right)\left(1 - \frac{u_2 t}{m_2} + \frac{1}{2!}\left(\frac{u_2 t}{m_2}\right)^2\right)\right] \qquad \textbf{Equation A22}$$

If we expand and then neglect terms in $u_2{}^2$ we obtain

$$\bar{S}(t) \cong m_2 log\left[\left(\frac{1}{exp\left(-\frac{\bar{S}(0)}{m_2}\right)} + \frac{q_1 t}{m_2}\right) - \frac{u_2 t}{m_2}\left(\frac{1}{exp\left(-\frac{\bar{S}(0)}{m_2}\right)} - \frac{1}{2}\frac{q_1 t}{m_2}\right)\right] \qquad \textbf{Equation A23}$$

We use this solution in cases where $\frac{u_2 t}{m_2} \ll 1$, currently implemented as $\frac{u_2 t}{m_2} < 10^{-10}$

Code Availability

The DECIPHeR model code is open source and freely available under the terms of the GNU General Public License version 3.0. The model code is written in fortran and is provided through a Github repository: https://github.com/uob-hydrology/DECIPHeR.

Persistent identifier: https://doi.org/10.5281/zenodo.2604120~~http://doi.org/10.5281/zenodo.1346159~~

Author Contribution

G Coxon and T Dunne wrote and modified the majority of the source code with contributions from R Lane, N Quinn and J Freer. J Freer provided overall oversight for the model development. R Woods and W Knoben derived the analytical solution for the subsurface zone equations. G Coxon prepared the input data and produced and evaluated the model simulations shown in this paper, with input from all co-authors on the experimental design. The manuscript was prepared by G Coxon with contributions from all co-authors.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

G Coxon, J Freer, T Wagener, R Woods and N Howden were supported by NERC MaRIUS: Managing the Risks, Impacts and Uncertainties of droughts and water Scarcity, grant number NE/L010399/1. Partial support for T Wagener comes from a Royal Society Wolfson Research Merit Award, and for N Howden from a University Research Fellowship. R Lane and W. Knoben were~~as~~ 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.

References

Andréassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C. and Loumagne, C.: Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using chimera watersheds, Water Resour. Res., 40(5), doi:10.1029/2003WR002854, 2004.

Archfield Stacey A., Clark Martyn, Arheimer Berit, Hay Lauren E., McMillan Hilary, Kiang Julie E., Seibert Jan, Hakala Kirsti, Bock Andrew, Wagener Thorsten, Farmer William H., Andréassian Vazken, Attinger Sabine, Viglione Alberto, Knight Rodney, Markstrom Steven and Over Thomas: Accelerating advances in continental domain hydrologic modeling, Water Resour. Res., 51(12), 10078–10091, doi:10.1002/2015WR017498, 2015.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R.: Large Area Hydrologic Modeling and Assessment Part I: Model Development1, JAWRA J. Am. Water Resour. Assoc., 34(1), 73–89, doi:10.1111/j.1752-1688.1998.tb05961.x, 1998.

Atkinson, S. E., Woods, R. A. and Sivapalan, M.: Climate and landscape controls on water balance model complexity over changing timescales, Water Resour. Res., 38(12), 50-1-50–17, doi:10.1029/2002WR001487, 2002.

Bai, Y., Wagener, T. and Reed, P.: A top-down framework for watershed model evaluation and selection under uncertainty, Environ. Model. Softw., 24(8), 901–916, doi:10.1016/j.envsoft.2008.12.012, 2009.

Bell, V. A., Kay, A. L., Jones, R. G. and Moore, R. J.: Development of a high resolution grid-based river flow model for use with regional climate model output, Hydrol Earth Syst Sci, 11(1), 532–549, doi:10.5194/hess-11-532-2007, 2007.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320(1–2), 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006.

Beven, K. and Freer, J.: A dynamic TOPMODEL, Hydrol. Process., 15(10), 1993–2011, doi:10.1002/hyp.252, 2001.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrol. Sci. Bull., 24(1), 43–69, doi:10.1080/02626667909491834, 1979.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., Roo, A. de, Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., Giesen, N. van de, Winsemius, H. and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, Hydrol. Process., 29(2), 310–320, doi:10.1002/hyp.10391, 2015.

Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, Hydrol. Process., 9(3–4), 251–290, doi:10.1002/hyp.3360090305, 1995.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z. and Smith, M.: Toward improved streamflow forecasts: value of semidistributed modeling, Water Resour. Res., 37(11), 2749–2759, doi:10.1029/2000WR000207, 2001.

Burnash, R. J. C.: The NWS River Forecast System-Catchment Modeling, in Computer Models of Watershed Hydrology, pp. 311–366, Water Resour. Publ., Littleton, Colo., 1995.

Chaney, N. W., Metcalfe, P. and Wood, E. F.: HydroBlocks: a field-scale resolving land surface model for application over continental extents, Hydrol. Process., 30(20), 3543–3559, doi:10.1002/hyp.10891, 2016.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T. and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44(12), n/a–n/a, doi:10.1029/2007WR006735, 2008a.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J. and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, Adv. Water Resour., 31(10), 1309–1324, doi:10.1016/j.advwatres.2008.06.005, 2008b.

Clark, M. P., Kavetski, D. and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47(9), W09301, doi:10.1029/2010WR009827, 2011.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J. and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51(4), 2498–2514, doi:10.1002/2015WR017198, 2015.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M. and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resour. Res., 48(5), W05552, doi:10.1029/2011WR011721, 2012.

Coron, L., Thirel, G., Delaigue, O., Perrin, C. and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environ. Model. Softw., 94, 166–171, doi:10.1016/j.envsoft.2017.05.002, 2017.

Coxon, G., Freer, J., Wagener, T., Odoni, N. A. and Clark, M.: Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, Hydrol. Process., 28(25), 6135–6150, doi:10.1002/hyp.10096, 2014.

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R. and Smith, P. J.: A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, Water Resour. Res., n/a-n/a, doi:10.1002/2014WR016532, 2015.

Crooks, S. M., Kay, A. L. and Reynard, N. S.: Regionalised impacts of climate change on flood flows: hydrological models, catchments and calibration. Milestone report 1, [online] Available from: http://randd.defra.gov.uk/Document.aspx?Document=FD2020_8853_TRP.pdf (Accessed 26 November 2018), 2010.

EC: Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. [online] Available from: http://data.europa.eu/eli/dir/2000/60/oj/eng (Accessed 30 May 2018), 2000.

EC: Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks (Text with EEA relevance). [online] Available from: http://data.europa.eu/eli/dir/2007/60/oj/eng (Accessed 17 July 2018), 2007.

Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47(11), doi:10.1029/2010WR010174, 2011.

Fowler, K. J. A., Peel, M. C., Western, A. W., Zhang, L. and Peterson, T. J.: Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models, Water Resour. Res., 52(3), 1820–1846, doi:10.1002/2015WR018068, 2016.

Freer, J. E., McMillan, H., McDonnell, J. J. and Beven, K. J.: Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, J. Hydrol., 291(3–4), 254–277, doi:10.1016/j.jhydrol.2003.12.037, 2004.

Gosling, S. N. and Arnell, N. W.: Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis, Hydrol. Process., 25(7), 1129–1145, doi:10.1002/hyp.7727, 2011.

Gupta, H. V., Wagener, T. and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22(18), 3802–3813, doi:10.1002/hyp.6989, 2008.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C. and Madsen, B.: Methodology for construction, calibration and validation of a national hydrological model for Denmark, J. Hydrol., 280(1), 52–71, doi:10.1016/S0022-1694(03)00186-0, 2003.

Kavetski, D. and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, Water Resour. Res., 47(11), n/a–n/a, doi:10.1029/2011WR010748, 2011.

Kavetski, D., Kuczera, G. and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, Water Resour. Res., 42(3), W03408, doi:10.1029/2005WR004376, 2006.

Keller, V. D. J., Tanguy, M., Prosdocimi, I., Terry, J. A., Hitt, O., Cole, S. J., Fry, M., Morris, D. G. and Dixon, H.: CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications, Earth Syst. Sci. Data, 7(1), 143–155, doi:https://doi.org/10.5194/essd-7-143-2015, 2015.

Kollat, J. B., Reed, P. M. and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, Water Resour. Res., 48(3), W03520, doi:10.1029/2011WR011534, 2012.

Krueger, T., Freer, J., Quinton, J. N., Macleod, C. J. A., Bilotta, G. S., Brazier, R. E., Butler, P. and Haygarth, P. M.: Ensemble evaluation of hydrological model hypotheses, Water Resour. Res., 46(7), n/a–n/a, doi:10.1029/2009WR007845, 2010.

Kuczera, G., Renard, B., Thyer, M. and Kavetski, D.: There are no hydrological monsters, just models and observations with large uncertainties!, Hydrol. Sci. J., 55(6), 980–991, doi:10.1080/02626667.2010.504677, 2010.

Kumar, R., Livneh, B. and Samaniego, L.: Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme, Water Resour. Res., 49(9), 5700–5714, doi:10.1002/wrcr.20431, 2013.

Leavesley, G. H., Markstrom, S. L., Brewer, M. S. and Viger, R. J.: The modular modeling system (MMS) — The physical process modeling component of a database-centered decision support system for water and power management, Water. Air. Soil Pollut., 90(1–2), 303–311, doi:10.1007/BF00619290, 1996.

Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201(1–4), 272–288, doi:10.1016/S0022-1694(97)00041-3, 1997.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J. and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, Hydrol. Res., 41(3–4), 295–319, doi:10.2166/nh.2010.007, 2010.

Liu, Y., Freer, J., Beven, K. and Matgen, P.: Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error, J. Hydrol., 367(1–2), 93–103, doi:10.1016/j.jhydrol.2009.01.016, 2009.

McMillan, H., Freer, J., Pappenberger, F., Krueger, T. and Clark, M.: Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions, Hydrol. Process., 24(10), 1270–1284, doi:10.1002/hyp.7587, 2010.

McMillan, H., Jackson, B., Clark, M., Kavetski, D. and Woods, R.: Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, J. Hydrol., 400(1–2), 83–94, doi:10.1016/j.jhydrol.2011.01.026, 2011.

McMillan, H., Krueger, T. and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol. Process., 26(26), 4078–4111, doi:10.1002/hyp.9384, 2012.

McMillan, H. K., Booker, D. J. and Cattoën, C.: Validation of a national hydrological model, J. Hydrol., 541, 800–815, doi:10.1016/j.jhydrol.2016.07.043, 2016.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G. and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, Water Resour. Res., 51(1), 716–728, doi:10.1002/2014WR015820, 2015.

Metcalfe, P., Beven, K. and Freer, J.: Dynamic TOPMODEL: A new implementation in R and its sensitivity to time and space steps, Environ. Model. Softw., 72, 155–172, doi:10.1016/j.envsoft.2015.06.010, 2015.

Metcalfe, P., Beven, K., Hankin, B. and Lamb, R.: A modelling framework for evaluation of the hydrological impacts of nature-based approaches to flood risk management, with application to in-channel interventions across a 29-km2 scale catchment in the United Kingdom, Hydrol. Process., 31(9), 1734–1748, doi:10.1002/hyp.11140, 2017.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O. and Samaniego, L.: Towards seamless large-domain parameter estimation for hydrologic models, Water Resour. Res., 53(9), 8020–8040, doi:10.1002/2017WR020401, 2017.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

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, Hydrol Earth Syst Sci, 19(1), 209–223, doi:10.5194/hess-19-209-2015, 2015.

Page, T., Beven, K. J., Freer, J. and Neal, C.: Modelling the chloride signal at Plynlimon, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty), Hydrol. Process., 21(3), 292–307, doi:10.1002/hyp.6186, 2007.

Pechlivanidis, I. G. and Arheimer, B.: Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case, Hydrol Earth Syst Sci, 19(11), 4559–4579, doi:10.5194/hess-19-4559-2015, 2015.

Perrin, C., Michel, C. and Andréassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, J. Hydrol., 242(3–4), 275–301, doi:10.1016/S0022-1694(00)00393-0, 2001.

Peters, N. E., Freer, J. and Beven, K.: Modelling hydrologic responses in a small forested catchment (Panola Mountain, Georgia, USA): a comparison of the original and a new dynamic TOPMODEL, Hydrol. Process., 17(2), 345–362, doi:10.1002/hyp.1128, 2003.

Quinn, P. F., Beven, K. J. and Lamb, R.: The in(a/tan/β) index: How to calculate it and how to use it within the topmodel framework, Hydrol. Process., 9(2), 161–182, doi:10.1002/hyp.3360090204, 1995.

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

Rudd, A. C., Bell, V. A. and Kay, A. L.: National-scale analysis of simulated hydrological droughts (1891–2015), J. Hydrol., 550, 368–385, doi:10.1016/j.jhydrol.2017.05.018, 2017.

Samaniego, L., Kumar, R. and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46(5), doi:10.1029/2008WR007327, 2010.

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., Eisner, S., Müller Schmied, H., Sutanudjaja, E. H., Warrach-Sagi, K. and Attinger, S.: Toward seamless hydrologic predictions across spatial scales, Hydrol Earth Syst Sci, 21(9), 4323–4346, doi:10.5194/hess-21-4323-2017, 2017.

Seibert, J., Vis, M. J. P., Lewis, E. and Meerveld, H. J. van: Upper and lower benchmarks in hydrological modelling, Hydrol. Process., 32(8), 1120–1125, doi:10.1002/hyp.11476, 2018.

Soille, P.: Optimal removal of spurious pits in grid digital elevation models, Water Resour. Res., 40(12), W12509, doi:10.1029/2004WR003060, 2004.

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], NERC Environ. Inf. Data Cent., doi:https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca, 2016.

Thames Water: Draft Water Resources Management Plan 2020 - 2100, Thames Water Utilities Ltd, Reading, UK., 2017.

Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C. J. H., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J. and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study, Environ. Res. Lett., 13(5), 055008, doi:10.1088/1748-9326/aab96f, 2018.

Wada, Y., Wisser, D. and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dyn., 5(1), 15–40, doi:10.5194/esd-5-15-2014, 2014.

Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N. B. and Wilson, J. S.: The future of hydrology: An evolving science for a changing world, Water Resour. Res., 46(5), doi:10.1029/2009WR008906, 2010.

Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E. and Xu, C.-Y.: Calibration of hydrological models using flow-duration curves, Hydrol Earth Syst Sci, 15(7), 2205–2227, doi:10.5194/hess-15-2205-2011, 2011.

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 Resour. Res., 52(3), 1847–1865, doi:10.1002/2015WR017635, 2016.

Wood, E. F., Sivapalan, M., Beven, K. and Band, L.: Effects of spatial variability and scale with implications to hydrologic modeling, J. Hydrol., 102(1), 29–47, doi:10.1016/0022-1694(88)90090-X, 1988.

Wood, E. F., Roundy, J. K., Troy, T. J., Beek, L. P. H. van, Bierkens, M. F. P., Blyth, E., Roo, A. de, Döll, P., Ek, M., Famiglietti, J., Gochis, D., Giesen, N. van de, Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A. and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water, Water Resour. Res., 47(5), doi:10.1029/2010WR010090, 2011.

Yadav, M., Wagener, T. and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Adv. Water Resour., 30(8), 1756–1774, doi:10.1016/j.advwatres.2007.01.005, 2007.

Yilmaz, K. K., Gupta, H. V. and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44(9), n/a–n/a, doi:10.1029/2007WR006716, 2008.

Younger, P. M., Gadian, A. M., Wang, C.-G., Freer, J. E. and Beven, K. J.: The usability of 250 m resolution data from the UK Meteorological Office Unified Model as input data for a hydrological model, Meteorol. Appl., 15(2), 207–217, doi:10.1002/met.46, 2008.

Zaherpour, J., Gosling, S. N., Mount, N., Schmied, H. M., Veldkamp, T. I. E., Rutger Dankers, Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanasaki, N., Hyungjun Kim, Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Jacob Schewe and Wada, Y.: Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts, Environ. Res. Lett., 13(6), 065015, doi:10.1088/1748-9326/aac547, 2018.

Zhao, R. J.: Watershed Hydrological Modelling, Water Resour. and Electr. Power Press, Beijing., 1984.

Table 1. Overview of DECIPHeR's stores, fluxes and parameters

Stores		
S_{RZ}	Root Zone Storage	m
S_{UZ}	Unsaturated Storage	m
S_{EX}	Saturation Excess Storage	m
S_D	Saturated Storage Deficit	m
Internal Fluxes		
Q_{UZ}	Drainage Flux	m ts^{-1}
Q_{IN}	Upslope Input Flow	m ts^{-1}
Q_{EXS}	Saturated Excess Flow	m ts^{-1}
Q_{EXUS}	Precipitation Excess Flow	m ts^{-1}
Q_{OF}	Overland Flow (sum of Q_{EXS} and Q_{EXUS})	m ts^{-1}
Q_{SAT}	Saturated Flow	m ts^{-1}
External Fluxes: Input		
P	Precipitation	m ts^{-1}
E	Potential Evapotranspiration	m ts^{-1}
Q_{obs}	Observed Discharge (for starting value of Q_{SAT})	m ts^{-1}
External Fluxes: Output		
Q_{sim}	Simulated Discharge	m ts^{-1}
Model Parameters		
SZM	Form of exponential decline in conductivity	m
SR_{max}	Maximum root zone storage	m
SR_{init}	Initial root zone storage	m
T_d	Unsaturated zone time delay	m ts^{-1}
CHV	Channel routing velocity	m ts^{-1}
$ln(T_0)$	Lateral saturated transmissivity	$\ln(m^2\ ts^{-1})$
S_{max}	Maximum effective deficit of saturated zone	m

Table 2. Parameter Ranges

Parameter	Units	Lower Bound	Upper Bound
SZM	m	0.001	0.15~~07~~
SR_{max}	m	0.005	0.3~~15~~
SR_{init}	m	0	0.01
T_d	m hr^{-1}	0.1	40
CHV	m hr^{-1}	10~~25~~0	4000
$ln(T_0)$	ln(m^2 hr^{-1})	-7	7~~5~~
S_{max}	m	0.3~~2~~	3

Table 3. Evaluation metrics used in the study

Evaluation Metric	Equation	Focus	Performance Threshold	
			Weaker	Stricter
Nash Sutcliffe Efficiency	$NSE = 1 - \frac{\sum_{i=1}^{n}(QO - QS)^2}{\sum_{i=1}^{n}(QO - \overline{QO})^2}$	High Flows, Timing	0	0.5
Bias in Runoff Ratio	$RRBias = \frac{\sum(QS - QO)}{\sum QO} * 100$	Water Balance	20	10
Bias in Low Flow Volume	$LFVBias = -100 * \frac{\sum_{p=70}^{95}(log\,(QS_p) - \log(QO_p))}{\sum_{p=70}^{95}(\log(QO_p))}$	Low Flows	20	10
Bias in Slope of the Flow Duration Curve between the 30th and 70th percentile	$SFDCBias = \frac{[\log(QS_{30}) - \log(QS_{70})] - [\log(QO_{30}) - \log(QO_{70})]}{[\log(QO_{30}) - \log(QO_{70})]} * 100$	Flow variability	20	10

Table 4. Summary statistics of DECIPHeR performance metrics for GB with catchments grouped by runoff coefficient and base flow index. Percentiles are taken from the behavioural ensemble from all catchments within each group. The column 'N' indicates the number of catchments in each group. Cells are coloured according to the thresholds outlined in section 4.3.3, green for the stricter threshold, yellow for the weaker threshold and red where it doesn't meet either of the thresholds.

	Runoff Coefficient							**Base Flow Index**						
	N	**NSE (-)**			**RRBias (%)**			**N**	**NSE**			**RRBias**		
		95th	Med	5th	95th	Med	5th		95th	Med	5th	95th	Med	5th
0-0.2	85	-73	-4.4	0.35	41	177	894	20	0.11	0.44	0.76	-31	0.54	134
0.2-0.4	362	-1.4	0.36	0.73	-0.5	22	123	320	-0.1	0.57	0.79	-12	1.4	100
0.4-0.6	348	0.12	0.54	0.81	-3.4	5.8	39	629	-0.1	0.54	0.80	-8.9	3.9	81
0.6-0.8	352	0.31	0.65	0.83	-10	0.14	14	257	-1.5	0.51	0.82	-10	8	113
>0.8	219	0.02	0.64	0.81	-41	-6	3.5	140	-37	0.04	0.83	-32	31	540

Table 5. Catchment characteristics and model performance for the six catchments shown in Figure 9. Baseflow index is a measure of the proportion of the river runoff that can be classified as baseflow and is derived from Marsh and Hannaford (2008). Water balance is calculated as mean annual rainfall minus mean annual discharge and potential evapotranspiration (as actual evapotranspiration is not available). NSE and BiasRR for the best ranked simulation according to the combined score described in Section 3.3.3 are shown for each catchment alongside the NSE and BiasRR derived from the mean of the behavioural ensemble.

Gauge Number	**River**	**Gauging Station**	**Catchment Area (km^2)**	**Mean Annual Rainfall (mm/year)**	**Mean Annual Potential Evapotranspiration (mm/year)**	**Mean Annual Discharge (mm/year)**	**Runoff Coefficient (-)**	**Water Balance (mm/year)**	**Baseflow Index (-)**	**Best Ranked Simulation**		**Ensemble Mean**	
										NSE (-)	**RRBias (%)**	**NSE (-)**	**RRBias (%)**
76014	Eden	Kirkby Stephen	69	1531	453	1230	0.8	-152	0.26	0.77	-2.6	0.79	-4.9
37005	Colne	Lexden	238	582	529	143	0.25	-91	0.52	0.63	18.8	0.43	21.3
43005	Avon	Amesbury	324	781	513	352	0.45	-84	0.91	0.91	-0.1	0.93	0.3
43004	Bourne	Laverstock	164	800	514	153	0.19	133	0.91	<0	147.4	<0	148
25023	Tees	Cow Green Reservoir	58	1696	446	1598	0.94	-348	0.57	0.10	-8.5	0.10	-12.9
39001	Thames	Kingston	9948	724	513	200	0.28	11	0.65	0.56	49	0.40	48.9

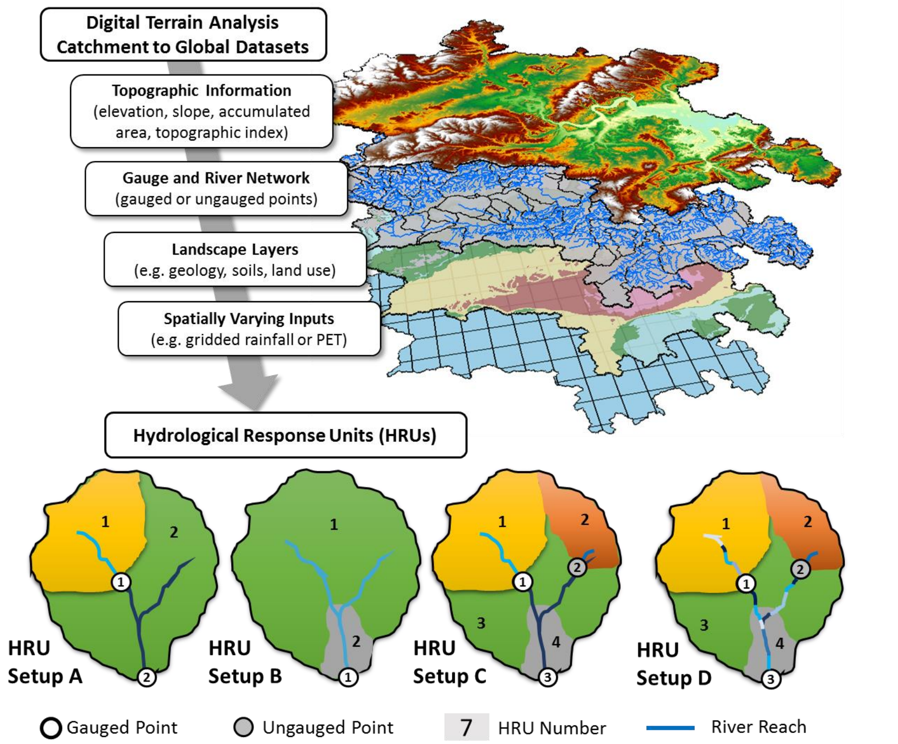


Figure 1. Digital Terrain Analysis and simplified examples of using classification layers to discretise a hypothetical catchment into Hydrological Response Units, from a) the gauge network, b) landscape layer with a chalk outcrop for HRU 2, c) the gauge network, ungauged flow point and landscape layer and d) same as c with individual river reach lengths specified

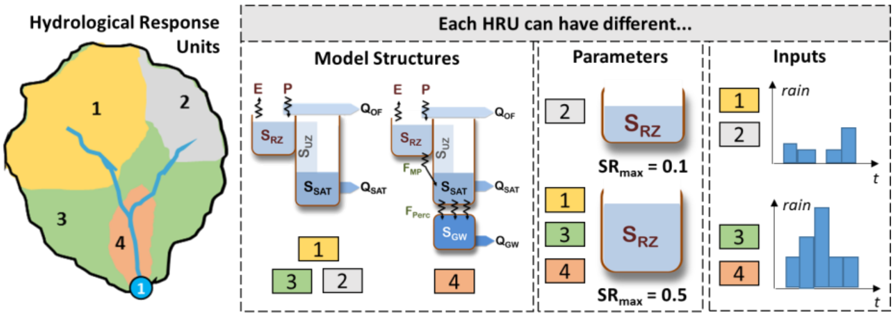


Figure 2. DECIPHeR represents spatial heterogeneity in the landscape through hydrological response units (HRUs). Each HRU can have a different model structure, parameters or inputs.

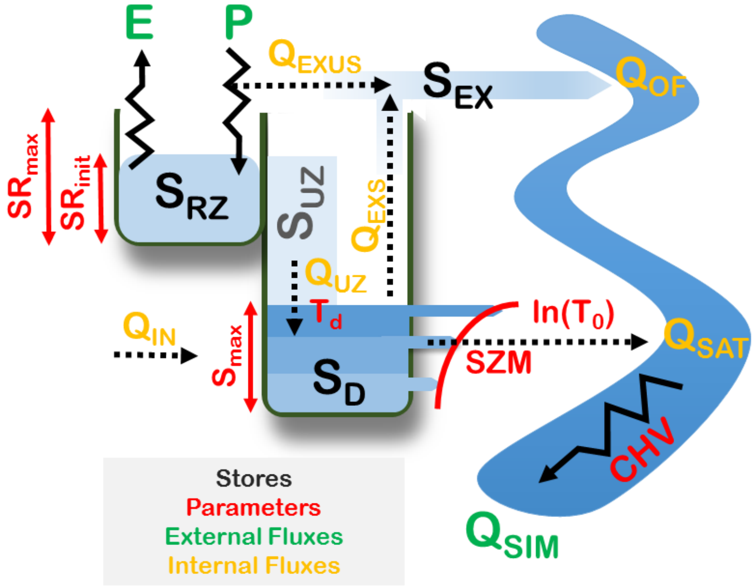


Figure 3. Simplified conceptual diagram of the model structure currently implemented in DECIPHeR. All scientific notations are described in Table 1.

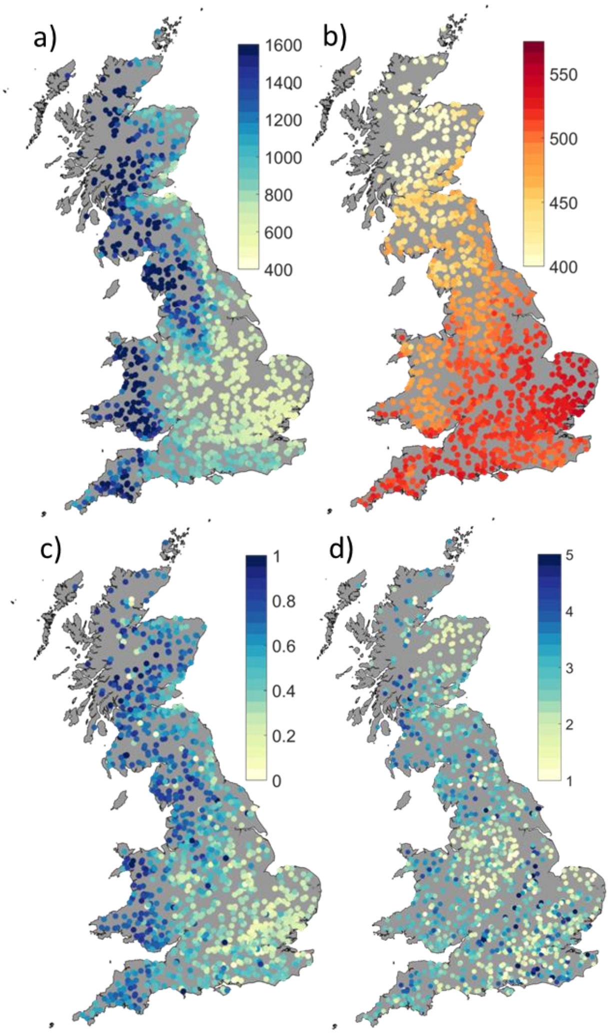


Figure 4. Hydro-climatic characteristics of 1366 GB catchments (a) Annual Rainfall (mm/year), (b) Annual ~~PET~~potential evapotranspiration (mm/year) (c) Runoff Coefficient (-), d) Slope of the Flow Duration Curve between the 30th and 70th percentiles (-). Min/max values on colorbars have been chosen to show clear differences between catchments.

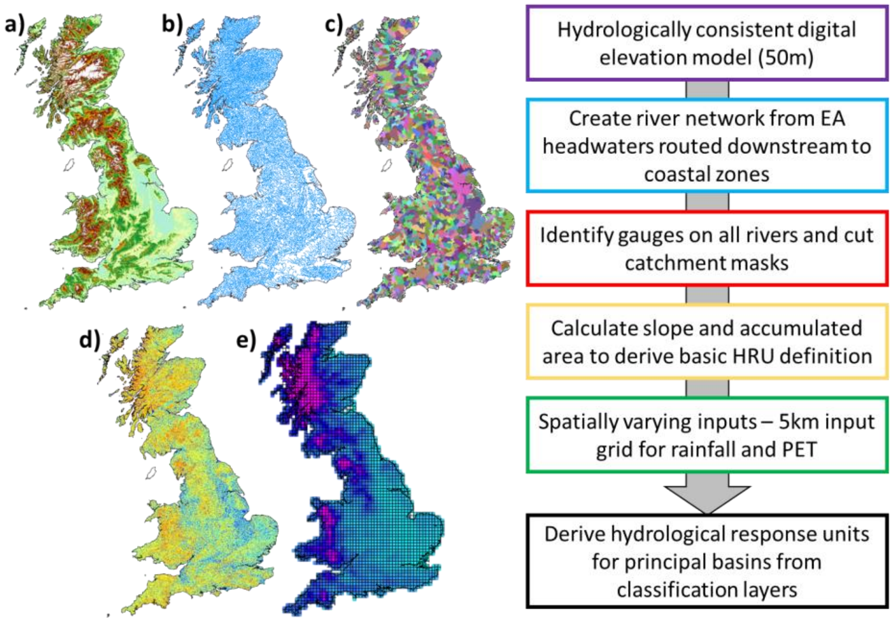
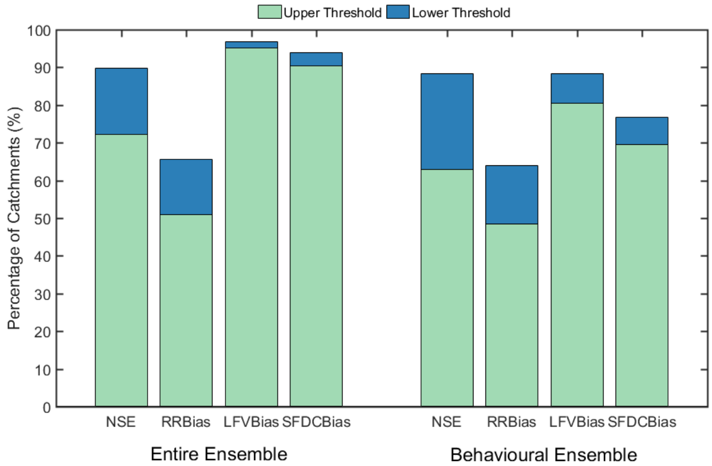


Figure 5. Inputs and Outputs of Digital Terrain Analyses for GB a) 50m Hydrologically Consistent Digital Elevation Model, b) DECIPHeR River Network, c) Nested Catchment Mask, d) Topographic Index, e) 5km input grid



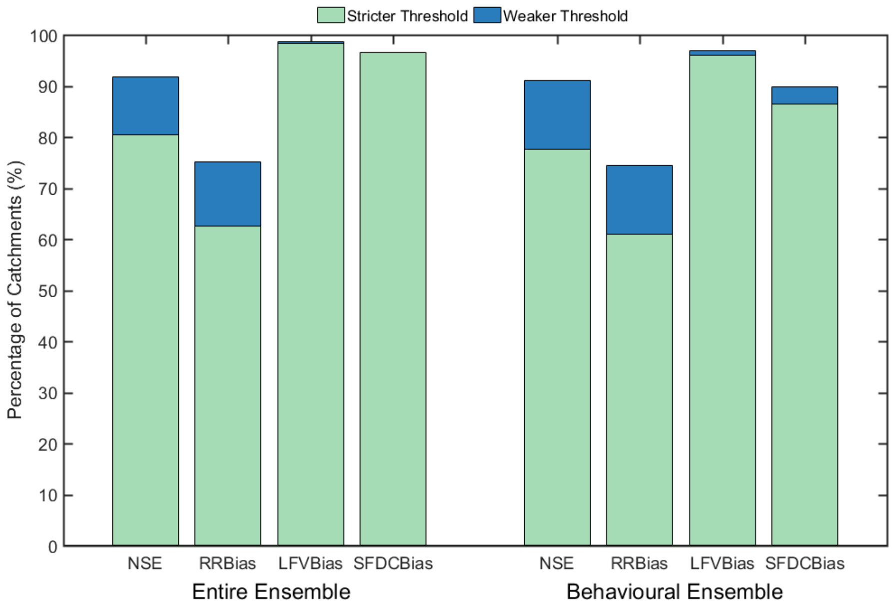


Figure 6. Percentage of catchments for each metric that meet the weaker and stricter performance thresholds for the entire ensemble of 10000 model simulations and from the top 1% behavioural ensemble of 100 model simulations generated from the combined ranking of the four metrics.

Commented [GC24]: RC2 Page 34 figure 6: The caption says “weaker and stricter” while the figure says “upper and lower”.

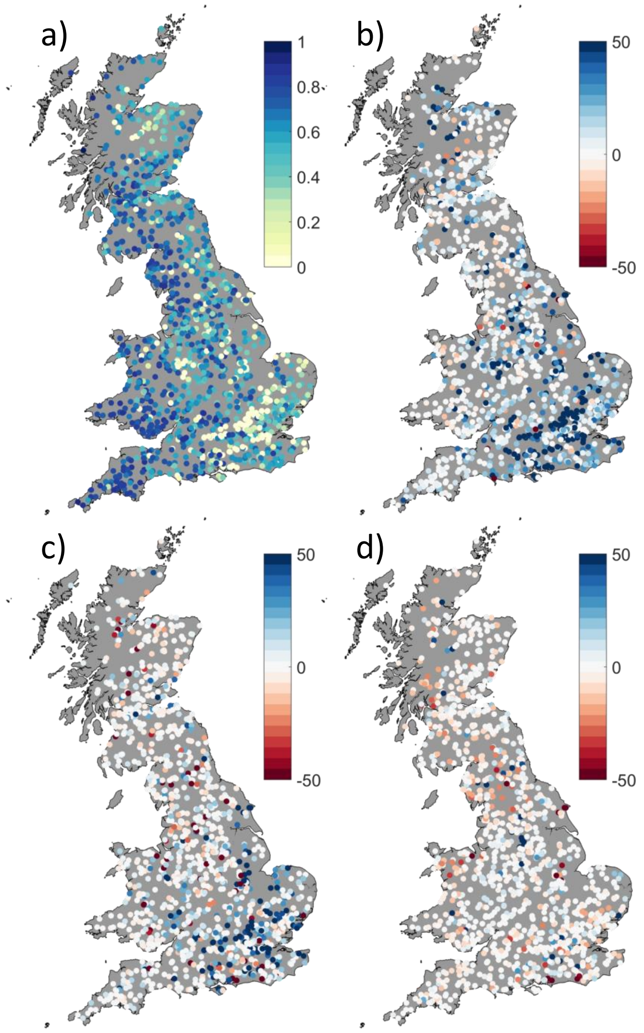


Figure 7. Model performance for the best simulation (as defined by the combined rank across all four metrics) for each evaluation metric a) NSE (-), b) Bias in Runoff Ratio (%), c) Bias in Low Flow Volume (%), and d) Bias in Slope of the Flow Duration Curve between the 30^{th} and 70^{th} percentil

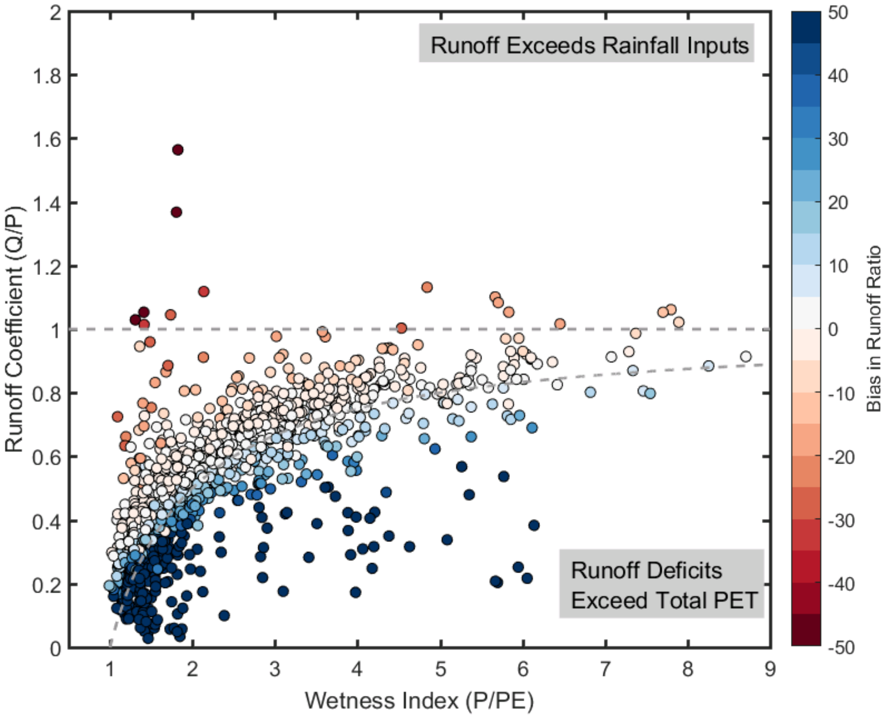


Figure 8. Scatter plot of wetness index (mean annual precipitation divided by mean annual potential evapotranspiration), runoff coefficient (mean annual discharge divided by mean annual precipitation) and bias in runoff ratio for each GB catchment evaluated in this study. Any points above the horizontal dotted line are where runoff exceeds total rainfall inputs in a catchment and any points below the curved line are where runoff deficits exceed total ~~PET~~potential evapotranspiration in a catchment.

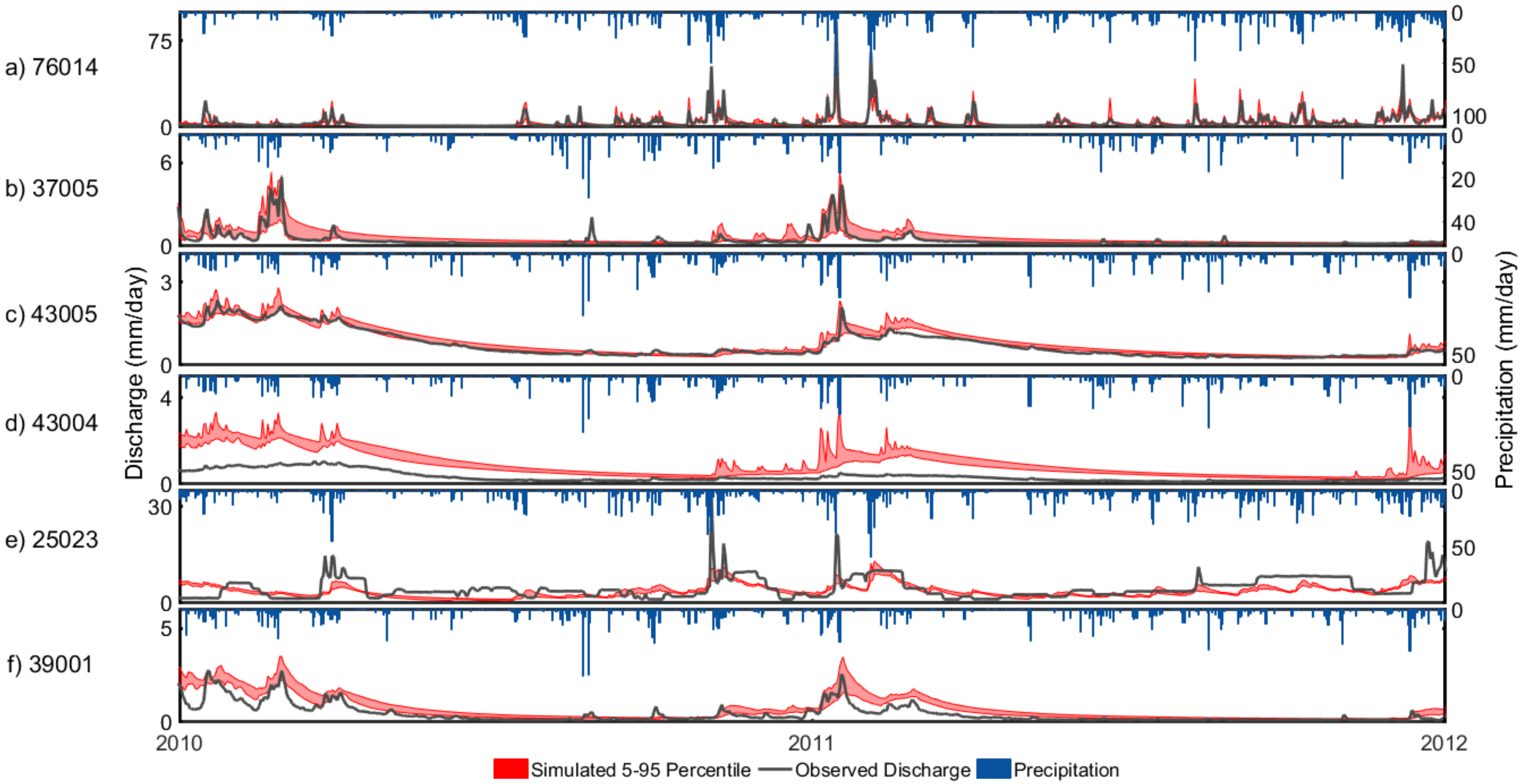


Figure 9. Observed discharge and uncertainty bounds for the behavioural simulations (5th and 95th percentile of the likelihood-weighted simulated discharge) for six catchments with different characteristics (shown in Table 5). The plots show a two year period (2010-2012) from the 55 year time series simulated.

Commented [GC25]: RC1 (19) Section 3.4.3 (P13, L33): The authors should present the actual time streamflow time series. Since this is the only the variable simulated/discussed, I was surprised that authors are not showing the time series plots. I suggest selecting certain representative gauging stations with varying catchment area and those located in different climatic regions for such analysis (it could be a 20 stations for example).