



Evaluation of overland flow modelling hypotheses with a multi-objective calibration using discharge and sediment data

Alban de Lavenne, Göran Lindström, Johan Strömqvist, Charlotta Pers,
Alena Bartosova, Berit Arheimer

► To cite this version:

Alban de Lavenne, Göran Lindström, Johan Strömqvist, Charlotta Pers, Alena Bartosova, et al..
Evaluation of overland flow modelling hypotheses with a multi-objective calibration using discharge
and sediment data. Hydrological Processes, 2022, 36 (12), 10.1002/hyp.14767 . hal-03992869

HAL Id: hal-03992869

<https://hal.inrae.fr/hal-03992869>

Submitted on 16 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

SPECIAL ISSUE PAPER

WILEY

Evaluation of overland flow modelling hypotheses with a multi-objective calibration using discharge and sediment data

Alban de Lavenne^{1,2}  | Göran Lindström¹ | Johan Strömqvist¹ | Charlotta Pers¹ | Alena Bartosova¹ | Berit Arheimer¹ 

¹Hydrology Research, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

²Université Paris-Saclay, INRAE, UR HYCAR, Antony, France

Correspondence

Alban de Lavenne, Université Paris-Saclay, INRAE, UR HYCAR, Antony, France.
Email: alban.delavenne@inrae.fr

Funding information

Horizon 2020 Framework Programme, Grant/Award Number: 776608; Joint Programming Initiative Climate; Swedish Environmental Protection Agency

Abstract

Conceptual hydrological models can move towards process-oriented modelling when addressing broader issues than discharge modelling alone. For instance, water quality modelling generally requires understanding of both pathways and travel times which might not be easily identified because observations at the outlet aggregate all processes at the catchment scale. In this study we tested if adding a second kind of observation, specifically sediment data, can help distinguish overland flow from total discharge. We applied a multi-objective calibration on both discharge and suspended sediment concentration simulation performance to the World-Wide Hydrological Predictions for the Environment (HYPE) model for 111 catchments spread over the USA. Results show that in comparison to two calibrations made one after the other, the multi-objective calibration leads to a significant improvement on the simulation performance of suspended sediments without a significant impact on the performance of discharge. New modelling hypotheses for overland flow calculations are proposed and resulted in similar discharge performances as the original one but with fewer parameters, which reduces equifinality and can prevent unwarranted model complexity in data-poor areas.

KEYWORDS

HYPE, multi-objective calibration, overland flow, surface runoff, suspended sediment

1 | INTRODUCTION

1.1 | Understanding the different flow paths

Conceptual rainfall-runoff models are generally designed to achieve one goal, that is, discharge prediction, with a limited number of parameters. Despite their efficiency for predicting discharge, they generally reach some limitations when it comes to describing water quality (see e.g., Fenicia et al., 2008), since water quality modelling generally relies on the understanding of flow path and travel time. One possible approach to directly address this issue is therefore to

introduce a higher level of description of hydrological processes allowing tracking the contaminants throughout the catchment. Following this idea, the Hydrological Predictions for the Environment (HYPE) model has originally been developed to address some of these limitations of the HBV conceptual model (Bergström, 1976; Lindström et al., 1997) when simulating water quality (Andersson et al., 2005; Arheimer et al., 2005; Arheimer & Brandt, 1998; Lindström et al., 2005). In the HYPE model, the water flow paths and the substances following them are simulated explicitly (Lindström et al., 2010; Pers et al., 2016). This more detailed flow path description also aims to improve the linkage to physiographic properties of catchments.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Hydrological Processes* published by John Wiley & Sons Ltd.

TABLE 1 Summary of modelling hypotheses for overland flow description.

Model hypothesis	Number of parameters	Description of overland flow
H0	3 (<i>srrate</i> , <i>mactrinf</i> , <i>mactrsm</i>)	Original formulation in HYPE v5.6.4
H1a	1 (β)	Function of soil moisture (Equation 2)
H1b	1 (β)	H1a with discrete formulation (Equation 4)
H2a	2 (α , γ)	Function of soil moisture (Equation 2) and rainfall intensity (Equation 3)
H2b	2 (α , γ)	H2a with discrete formulation (Equation 4)

The development of these 'process-based' models then relies on quantifying the contributions of the different hydrological compartments (such as surface flow, subsurface flow, groundwater contributions, etc.) for which different nutrient dynamics could be assigned. However, how to split the water flow into these compartments is often assessed indirectly through a comparison with the signal that aggregates them all, that is, the river discharge. Moreover, this additional description generally requires a calibration of new parameters. A lack of observation to support this level of complexity might lead to overparametrization, equifinality of parameter values, and to an increase of uncertainty (Beven, 1993; Guse et al., 2017; Her & Chaubey, 2015; Wagener et al., 2003).

1.2 | Addressing model complexity with additional data

Besides the amount of observation that is available, one key issue is also how much information about the involved processes these

observations contain, and how they are used to drive the model calibration (Gupta et al., 1998). Tracer and isotope data can for instance be used for a proper identification of flow partitioning (e.g., Didszun & Uhlenbrook, 2008; Klaus & McDonnell, 2013; Tetzlaff & Soulsby, 2008; Tonderski et al., 2017). However, their poor availability is limiting their use to a small number of places.

Satellite data is also often used, in addition to streamflow data, in order to better evaluate or calibrate rainfall-runoff models. It can be related to internal states of the models and to assess the realism of the modelled fluxes (e.g., Bouaziz et al., 2021; Nijzink et al., 2018; Rakovec et al., 2016). It could be for instance water storage anomalies extracted from GRACE satellites, soil moisture from the SMOS satellite, snow cover from MODIS images or evaporation rates from GLEAM. This type of data is widely available spatially and provides a useful understanding of the dynamics of each hydrologic compartment, but is difficult to use to actually follow a water particle along its path to the outlet.

Water quality observations can provide a useful complementary framework to address this issue, assuming that certain constituents preferentially follow certain flow paths. Despite the fact that they are generally non-conservative, water quality data can help to distinguish the contribution of different hydrological compartments (e.g., Bergström et al., 2002; van Griensven & Meixner, 2007).

Multi-objective optimisation thus appears as a relevant tool to make a profitable use of multiple data sets and to jointly optimize different processes that the model aims at describing (Efstratiadis & Koutsoyiannis, 2010; Gupta et al., 1998; Seibert, 2000; Yapo et al., 1998). By constraining the optimisation of parameter values, it can improve the description of flow partitioning, increase parameter identifiability and reduce uncertainty (Bergström et al., 2002; Her & Seong, 2018; Shafii et al., 2017).

In this paper we will focus only on the overland flow component, for which a good description is necessary for many water quality issues, as it is impacting the transport of several pollutants such as pesticides and phosphorus. We will make the hypothesis that in-

TABLE 2 Summary of the calibrated parameters for each modelling hypothesis (spatialised homogeneously over the sub-catchments).

Processes	Hypothesis	Parameter	Description	Range	Units
Overland flow	H0	<i>mactrinf</i>	Maximum infiltration capacity	[0; 100]	mm d ⁻¹
		<i>mactrsm</i>	Minimum soil wetness condition	[0; 1]	-
		<i>srrate</i>	Overland flow ratio	[0; 1]	-
	H1a, H1b	β	Effect of soil moisture	[0; 10]	-
	H2a, H2b	α	Effect of soil moisture	[0; 100]	-
		γ	Effect of rainfall intensity	[0; 2]	-
Macropore flow	All	<i>macrate</i>	Macropore flow ratio	[0; 1]	-
Soil storage	All	<i>wcfc1</i>	Water retention at field capacity of soil layer 1	[0; 1]	-
		<i>wcfc2</i>	Water retention at field capacity of soil layer 2	[0; 1]	-
Erosion and sediment transport	All	<i>erodindex</i>	Erosion scaling factor	[0; exp(10)]	-
		<i>erodexp</i>	Erosion precipitation dependent factor	[0; 5]	-
		<i>pprelexp</i>	Delay of sediment from overland flow	[0; 5]	-
		<i>pprelmax</i>	Delay of sediment from overland flow	[0; 100]	mm d ⁻¹

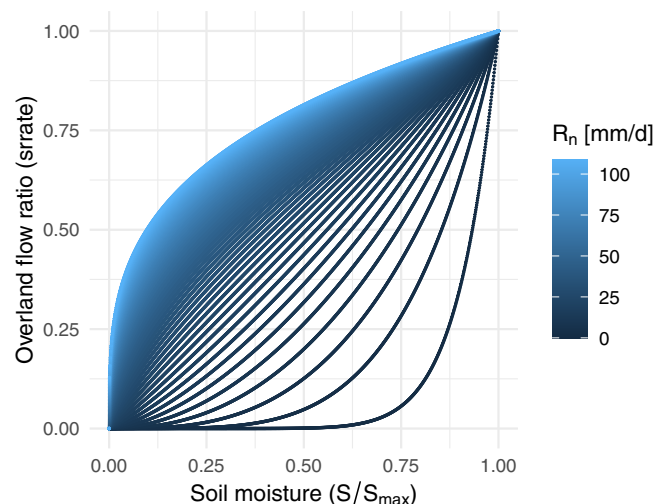
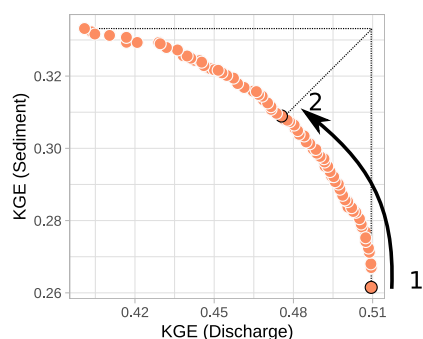
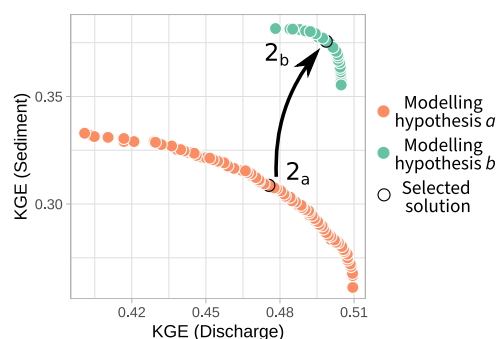


FIGURE 1 Change of the relation between overland flow ratio and soil moisture at catchment scale for different rainfall intensities in hypothesis 2 (for $\alpha = 10$ and $\gamma = 0.75$)



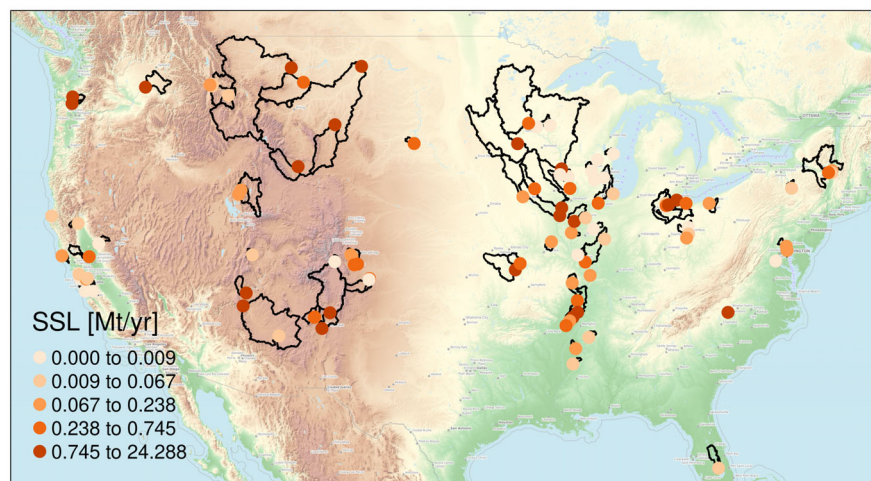
(a) Mono-objective (point 1) to multi-objective solution (point 2) for a given model



(b) Multi-objective solutions (points 2a and 2b) comparison between two modelling hypotheses

FIGURE 2 Illustration of how two different solutions are picked from the Pareto Front and how they are used for (a) calibration strategy comparison and (b) modelling hypotheses comparison. One multi-objective solution is selected using the shortest distance to the point defined by the maximum of the two objective functions

FIGURE 3 Map of the 111 USGS catchment boundaries with average observed suspended sediments loads (SSL) values illustrated at outlets (base map from OpenTopoMap)



stream total suspended sediment (TSS) concentrations are strongly influenced by overland flow (through erosion and remobilisation). By jointly optimizing the performance on sediment concentration and discharge, we aim to create a framework that allows a relevant evaluation of overland flow, as it is a process that is often not precisely identified in hydrological models (Beven, 2021).

A few studies have already tried multi-objective approaches in improving TSS predictions and discharge simultaneously. Sikorska et al. (2015) used a rainfall-runoff model (HydMod) with a build-up/wash-off model (BwMod) and shows that the multi-objective calibration provide more reliable TSS prediction with sharper uncertainty bounds. The SWAT model has also often been used to address this question (Bekele & Nicklow, 2007; Brighenti et al., 2019; Cheng et al., 2018; Muleta & Nicklow, 2005). The results generally lead to the conclusion that simultaneous calibration helps to achieve a more robust calibration compared to sequential calibration, and this improvement is particularly observed in the prediction of suspended sediments (Brighenti et al., 2019; Cheng et al., 2018).

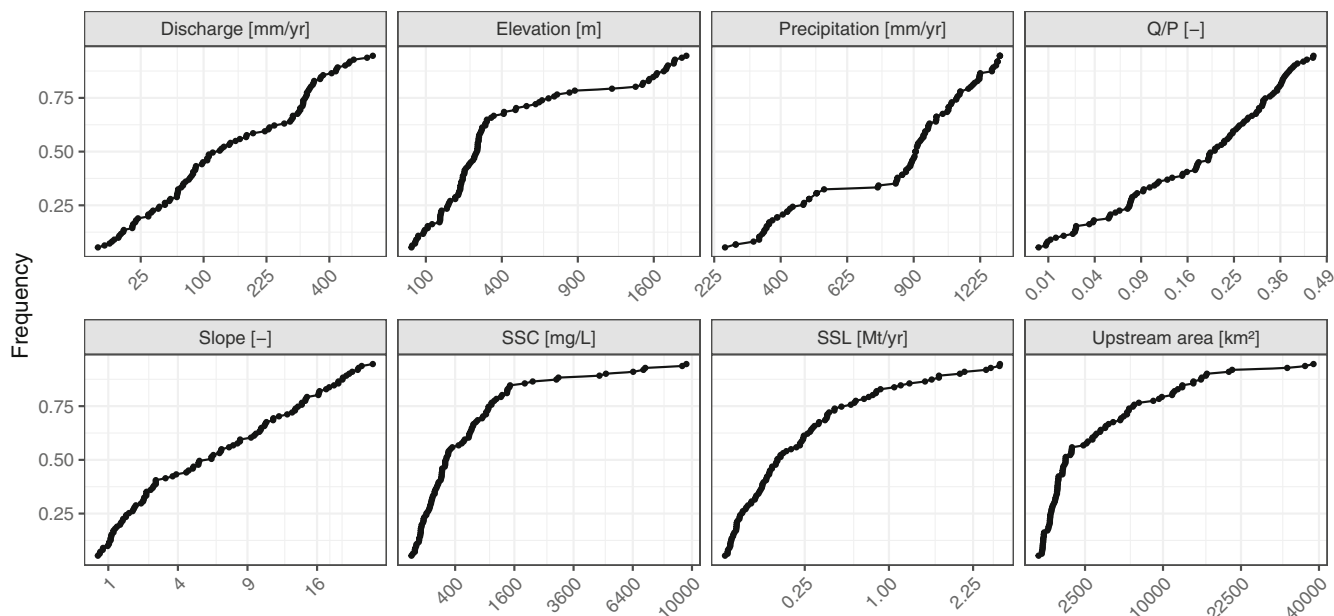


FIGURE 4 Catchments characteristic distribution. Annual statistics are computed on hydrological year starting on October 1st. Suspended sediment concentrations (SSC) and loads (SSL) are average values weighted by observed discharge

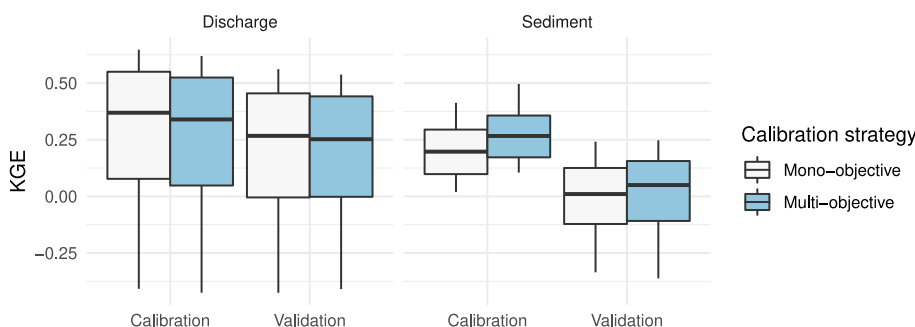


FIGURE 5 Boxplot comparison of KGE performances on sediment and discharge for mono-objective and multi-objective optimisation over the 111 catchments. The performances are for the modelling hypothesis H1a (for all the other hypotheses see Appendix A.2)

1.3 | General objectives

Based on this encouraging literature, we aim to use this strategy with two main objectives:

1. Study how a multi-objective calibration approach is affecting the performance of the HYPE model in comparison to a two-step calibration of overland flow and sediment.
2. Propose and assess new model equations for overland flow description by using this multi-objective calibration as a framework to discriminate modelling hypotheses.

2 | MATERIAL AND METHODS

2.1 | HYPE model set-up

The HYPE model is an open source semi-distributed hydrological model for small-scale and large-scale assessments of water resources

and water quality developed at the Swedish Meteorological and Hydrological Institute (SMHI). Different flow paths are described by the model: overland flow, macropore flow, tile drainage and outflow from each of the three soil layers. Rivers and lakes are described separately. A more extensive description of the model is provided by Lindström et al. (2010).

Regionalisation of this model is mostly based on the concept of Hydrological Response Units (HRU) that are defined according to the soil type, land use, altitude and climate (for more details, see Arheimer et al., 2020). However, some parameters, such as the ones calibrated in this study (see Table 2), are not regionalised by HRU but rather defined homogeneously over all the sub-catchments. Space is discretised into sub-catchments but different HRU can be implemented therein and contribute to the outflow.

We used HYPE version 5.6.4 and the same model set-up (catchment delineation, definition of HRU) and optimized parameter values of the recent calibration at global scale by Arheimer et al. (2020) called World-Wide HYPE (WWH). Only parameters and routines that could affect overland flow are updated for this study (see Section 2.3).

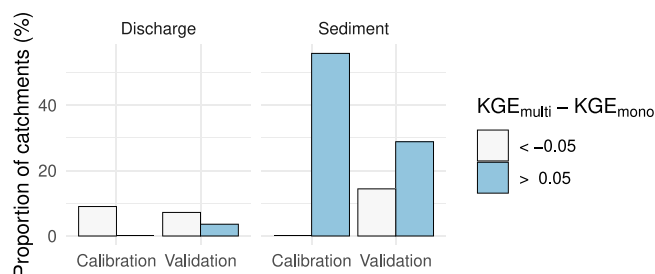


FIGURE 6 Proportion of the 111 catchments where a difference of KGE performances above 0.05 is observed between optimisation strategies for modelling hypothesis H1a: Blue highlights better performances of the multi-objective calibration and grey highlights better performances of mono-objective calibration (for all the other hypotheses and KGE components see Appendix A.3: Figures 17 and 18)

2.2 | Sediment modelling in HYPE

Sediment modelling in WWH is based on the HBV-SED model (Lidén, 1999; Lidén et al., 2001) which aims at describing soil erosion processes and transport. Soil mobilization is quantified through a power relation with rainfall intensity (*erodexp* parameter, Table 2) and its magnitude is adjusted according to soil, land use and slope (*erodin* parameter, Table 2). These mobilized sediments are then collected in a temporary storage. A potential amount of sediments that may be flushed from this storage is set by a power relation with total discharge (relation calibrated with *pprelexp* and *pprelmax* parameters, Table 2). Overland flow is the limiting factor in HYPE v5.6.4 that allows this potential amount of sediments to be actually transported to the stream (without overland flow there will be no sediments in the river discharge). A portion of these particles in the streams can settle out in the river bed or become re-suspended, and the rate of this process is a function of total discharge. Sedimentation in lakes depends on the concentration and a settling velocity parameter. These sediment transport processes are also simulated with WWH and HYPE in general.

In WWH, overland flow is thus an important process that will directly impact the simulation of suspended sediment. This study relies on this modelling hypothesis to discriminate among the different overland flow routines presented below. Thus, parameters affecting sediment transport processes in streams were not included in the recalibration efforts in this study, relying on parameter values obtained in the general WWH sediment calibration (Bartosova et al., 2021).

2.3 | New overland flow modelling hypotheses

Overland flow, as represented in WWH, takes place because of two main reasons: saturation excess (Dunne & Black, 1970) and infiltration excess (Horton, 1935). Both processes are described in WWH and will

therefore be taken into account in the first objective of this study. However, concerning the second objective of this study of proposing new descriptions of this process, we will not discuss overland flow that is triggered by the rise of the groundwater table, and will focus only on overland flow that is triggered by the dynamic and the spatial variability of the infiltration capacity.

2.3.1 | Hypothesis 0

In the original model version, called hypothesis 0 (H0), the overland flow process by infiltration excess is triggered when two thresholds are both exceeded during a time step: when rainfall intensity exceeds the maximum infiltration capacity (*mactrinf* parameter) and soil wetness condition exceed a minimum value (set by *mactrsm* parameter). This soil wetness is defined as the water content of the uppermost soil layers which mainly depends on one parameter (*wcfc1*) that is affecting this soil layer storage capacity. Overland flow Q_S then depends on a third parameter, the overland flow ratio *srrate*, that will define the amount of rainfall effectively reaching the stream and what will be stored by the soil following:

$$Q_S = srrate \cdot (R_n - mactrinf), \quad (1)$$

where, R_n is the sum of snowmelt and rainfall after interception.

2.3.2 | Hypothesis 1

This new modelling hypothesis 1 (H1) aims to eliminate the assumption of threshold behaviour at catchment scale and constant value of overland flow ratio (the proportion of rainfall which contributes to overland flow). It describes a smoother behaviour where the runoff ratio will dynamically change according to soil wetness condition. Soil wetness is defined as the relative water content of the two uppermost soil layers. It thus mainly depends on two parameters (*wcfc1* and *wcfc2*) that are affecting the storage capacity of both soil layers respectively (S_{max}). The fraction of the rainfall that will directly reach the stream can then be described by a power relation with the soil wetness, using only one parameter β (apart from the parameters used in all modelling hypotheses, Table 2) following:

$$Q_S = R_n \cdot \left(\frac{S}{S_{max}} \right)^\beta, \quad (2)$$

where, R_n is the sum of snowmelt and rainfall after interception, S/S_{max} is the soil moisture with S being the water content of the first two soil layers and S_{max} the water content capacity of these layers.

Within this formulation, for a given rainfall, we aim to generate more overland flow when the catchment is wet and less when it is dry. This hypothesis is closely related to the process of saturation

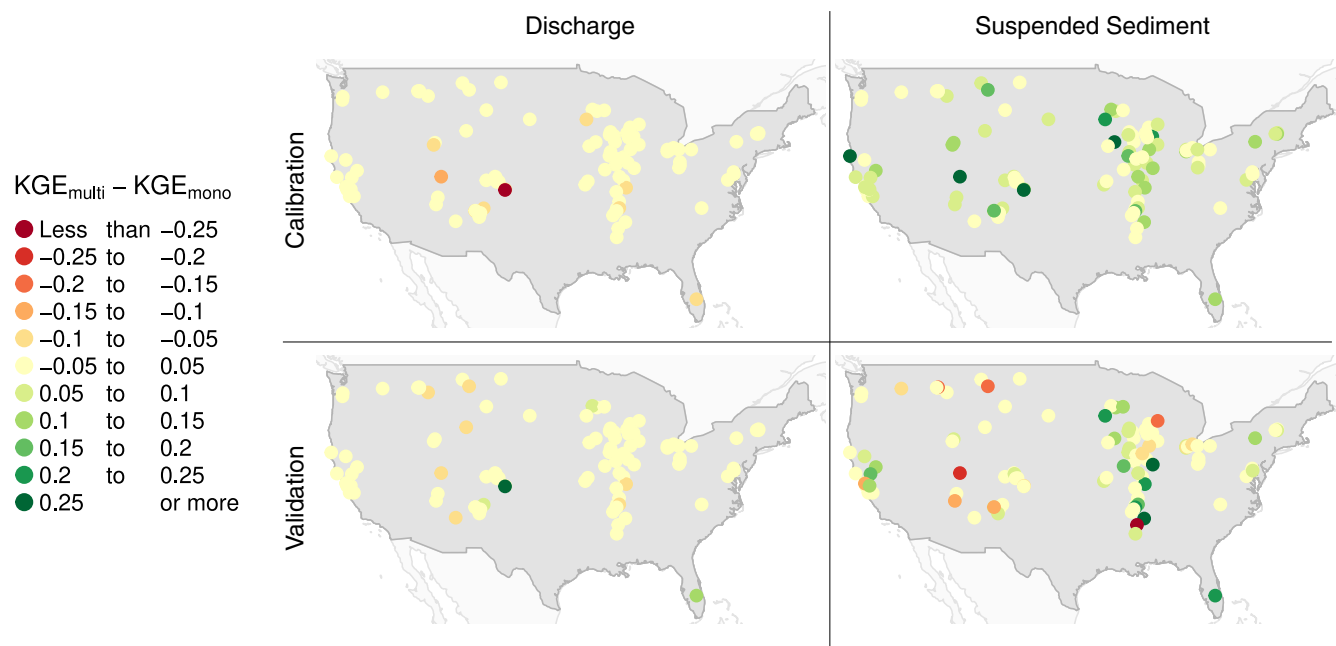


FIGURE 7 Map comparison of performances on sediment and discharge for mono-objective and multi-objective optimisation during calibration period. The performances are for the modelling hypothesis H1a (for all the other hypotheses see Appendix A.4)

excess, with the idea that a greater proportion of the watershed is saturated when the soil wetness index is high, leading to a greater volume of overland flow. It is interesting to notice that this relation is very similar to other models, such as the β curve parameter in the HBV model or the free water storage concept proposed by Zhao (1992) in the Xinanjiang model. However, it is used here to describe overland flow only instead of total runoff.

2.3.3 | Hypothesis 2

This hypothesis 2 (H2) aims to complete hypothesis 1 by accounting for the effect of rainfall intensity. Here we aim to give the ability of the model to increase the overland flow ratio with rainfall intensity for a given soil wetness. This is done by allowing the shape of the relation between overland flow ratio and soil moisture to change according to rainfall intensity (Figure 1). We propose to use the same Equation (2) but to dynamically change the exponent β with rainfall intensity following:

$$\beta = \frac{\alpha}{R_n^\gamma} \quad (3)$$

This modelling hypothesis thus requires two parameters: α describing the shape of the relation, and γ describing how much this relation can change with rainfall intensity. This hypothesis is closely related to the process of infiltration excess, with the idea that a greater intensity of rainfall will lead to a lower proportion of infiltration, thus leading to a greater volume of overland flow.

2.3.4 | Discrete formulations

Equations of hypotheses 1 and 2 are continuous-time formulations of overland flow. For a more explicit formulation of the water balance, they should be expressed at the time step of the model (Santos et al., 2018). Equation (2) at a discrete time step then becomes, with I corresponding to the infiltration rate:

$$I = S_{max} \cdot \left(1 - \left(\frac{S}{S_{max}}\right)^\beta\right) \cdot \tanh\left(\frac{R_n}{S_{max}}\right) / \left(1 + \left(\frac{S}{S_{max}}\right) \cdot \tanh\left(\frac{R_n}{S_{max}}\right)\right) \quad (4)$$

In this study we will explore a simple but approximated solution (Equation 2) and a more complex but exact solution (Equation 4) of the modelling hypothesis. Both mathematical formulations will be compared in order to evaluate the impact of modelling results. All modelling hypotheses are summarized in Table 1.

2.4 | Strategy for parameter's optimisation

2.4.1 | Model evaluation criteria

The models were calibrated and evaluated with the Kling-Gupta Efficiency (KGE, Gupta et al., 2009) criteria for both discharge and sediment concentrations at a daily time step. The three components of the KGE were also used to understand how the calibration strategy and the modelling hypotheses impact performances: the Pearson correlation r , the ratio of the mean simulated and observed value (μ_S/μ_O) and the ratio of their variance (σ_S/σ_O).

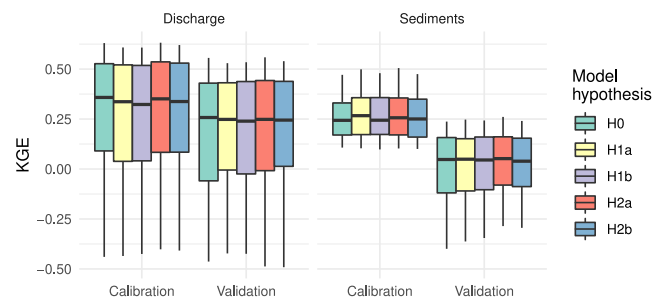


FIGURE 8 Boxplot comparison of KGE performances over the 111 catchments for the different modelling hypotheses when using the multi-objective solution on the Pareto front

$$KGE = 1 - \sqrt{(1-r)^2 + \left(1 - \frac{\mu_s}{\mu_o}\right)^2 + \left(1 - \frac{\sigma_s}{\sigma_o}\right)^2}. \quad (5)$$

Furthermore, in order to compare model assumptions not only with the aim of maximizing simulation performance but also with the aim of minimizing model complexity, we additionally used the AIC criterion (Akaike, 1973).

$$AIC = 2k - 2\ln(\hat{L}), \quad (6)$$

where, k is the number of free parameters (from 8 to 10, Table 2), and \hat{L} is the likelihood function defined as the root-mean-square error (RMSE) normalized by the average observation in order to facilitate comparison between catchments.

2.4.2 | A multi-objective framework

The multi-objective framework aims to optimize two objective functions: total discharge performance and TSS concentration (using KGE criteria described in Section 2.4). The search for the Pareto front is done using the caRamel algorithm (Monteil et al., 2020), a hybrid of the multi-objective evolutionary annealing simplex (MEAS) method and the non-dominated sorting genetic algorithm II (ϵ -NSGA-II).

As formulated at the end of the introduction, our first goal was to compare multi-objective calibration (simultaneous optimisation of both discharge and sediment modelling) with sequential calibration (optimisation of discharge followed by the optimisation of sediment). This was done by extracting two solutions from the Pareto Front (Figure 2a):

1. Mono-objective solution: maximizing the performance on discharge before sediment.
2. Multi-objective solution: finding a compromise on the Pareto front between both performances.

We used solution 1 as a wrap-up for sequential calibration: it is indeed theoretically the best solution for the discharge simulation and the best solution for sediment if parameters affecting discharge cannot be tuned anymore. By using the same parameter space

exploration for both solutions, we also avoided analysing differences that can be due to deficiencies of the optimisation algorithm in this highly dimensional space. Solution 2 was extracted by using the shortest Euclidean distance to the point whose coordinates were defined by the maximum values of both objective functions.

Moving from solution 1 to solution 2 will inevitably lead to a loss of performance on discharge for the calibration period. To evaluate the benefits of multi-objective calibration we thus performed a split-sample test (Klemeš, 1986) using two independent periods: the calibration period during which parameters were optimized and a validation period during which this optimized parameter set was used. The core idea was to evaluate if this multi-objective framework helps to identify a more robust solution (robustness evaluated through relative performance during validation) despite the inevitable loss of performance imposed by the trade-off during the calibration.

Calibration and validation periods were defined independently for each station according to the gauging period of both variables. This gauging period was split into two periods of equal length. The first one was used as a calibration period and the second one as a validation period. The two periods were then swapped and the second calibration/validation was performed. The average performance over the two calibration periods and the average performance over the two validation periods were finally used to avoid any bias due to potential differences of climate between periods. These calibration and validation evaluations were performed independently on each catchment. Compared to a calibration carried out over the entire domain (as done by Arheimer et al. (2020)), this catchment-by-catchment calibration aims to avoid dealing with the regionalisation strategy at the same time as assessing the modelling hypotheses.

This multi-optimisation framework was used to address our second objective of testing new modelling hypotheses. Each modelling hypothesis resulted in a Pareto front and the comparison between two fronts was done using solution 2 for each of them (Figure 2b). As previously explained, both calibration and validation performances were used to compare modelling hypotheses.

2.5 | Model inputs and catchment descriptors

Daily discharge and TSS concentration time series were extracted from the USGS Water Data portal (U.S. Geological Survey, 2016). Apart from the discharge and sediment data, the model inputs were the same as those that were used on a previous calibration at global scale (Arheimer et al., 2020). In particular, the climatic inputs were extracted from the Hydrological Global Forcing Data (HydroGFD; Berg et al., 2018). In this way, we aim to avoid the need to re-calibrate the full rainfall-runoff relation within the model and to focus only on the re-calibration of the overland flow routine. In some cases, this strategy may force the overland flow parameters to adjust for inadequacies in other processes.

The results will be analysed according to different catchment descriptors. The discharge data and climatic inputs of the model are used to define hydroclimatic descriptors. Similarly, other physiographic descriptors, such as catchment topography and land use, are

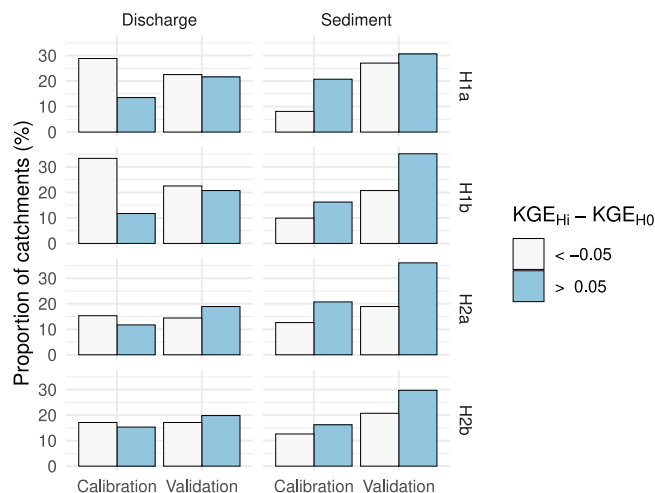


FIGURE 9 Proportion of the 111 catchments where a difference of KGE performances above 0.05 is observed between the original modelling hypothesis H0 and the new ones Hi (H1a, H1b, H2a, and H2b): Blue highlights better performances for the new hypotheses and grey highlights worse performances (for the KGE components see Appendix B.2)

derived from the definition of HRUs in the global model set up (Arheimer et al., 2020). According to this model set up, about 30% of the catchments are influenced by lakes and reservoirs.

2.6 | Study period and area

A total of 111 catchments widely distributed across the US were selected for this study. It allowed a wide range of hydro-climatic

characteristics to be explored (Figures 3 and 4), despite a lack of catchments in the middle precipitation range. Catchments size varies from about 70 to 180 000 km² (median 1400 km²), discharge from a few mm/yr to 1500 mm/yr (median 130 mm/yr) and precipitation from 220 to 2300 mm/yr (median 900 mm/yr). Stations were selected in order to have both discharge and sediment data available at the same location (see Section 2.5). This selection was also made according to the amount of data and the location of the USGS stations in comparison to WWH outlets: since discharge and sediment gauging can sometimes take place over different time periods, we imposed a minimum period of 3 years during which both observations are available jointly. The simulation period spreads from 1991/01/01 to 2018/12/31 with a warm-up period of 10 years (from 1981).

The model set-up of Arheimer et al. (2020) imposed the sub-catchments delineation that did not always fit exactly with the location of these USGS stations. These 111 catchments thus represent a selection of catchments that were respecting a reasonable spatial proximity between USGS stations and the WWH outlets: a maximum difference between areas of 50% and a maximum Euclidean distance between outlets of 50 km. After applying this matching, a HYPE outlet was found at less than 5 km, with a difference of catchment area below 2% and below 100 km² in median values.

3 | RESULTS AND DISCUSSION

3.1 | Mono-objective and multi-objective comparison

Moving from mono-objective to multi-objective calibration significantly improved the performances for sediment during the

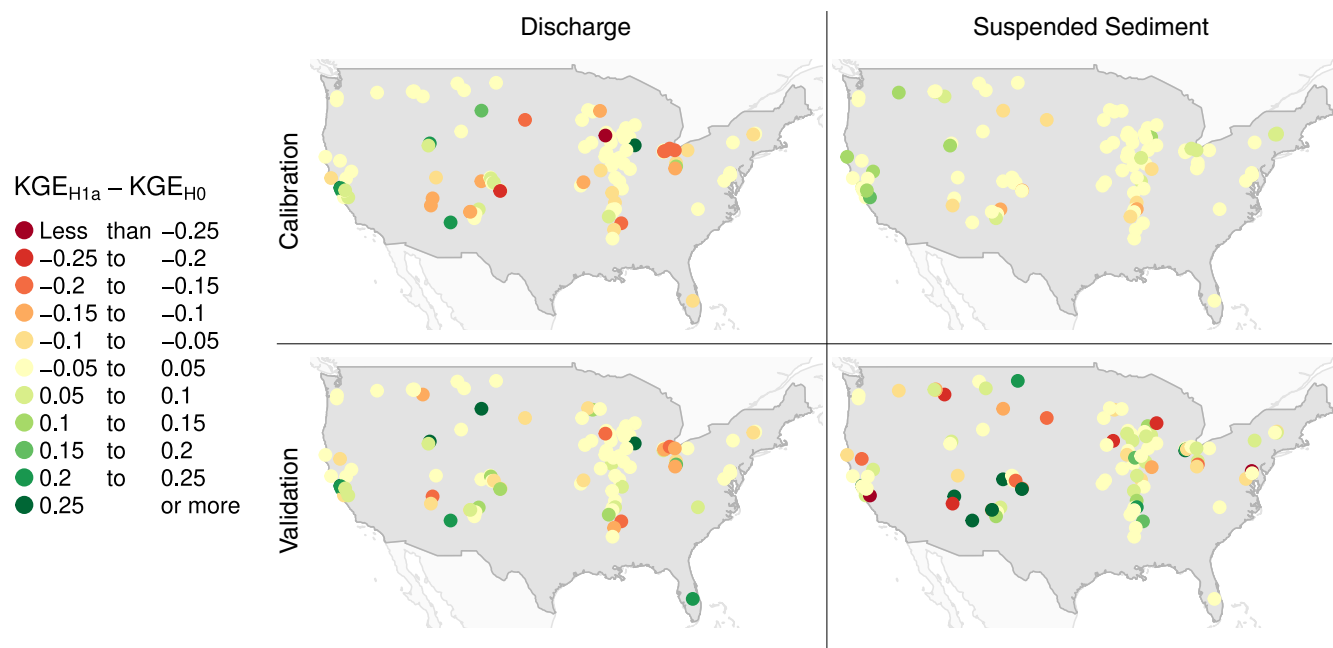


FIGURE 10 Map of model improvement between H0 and H1a during validation period

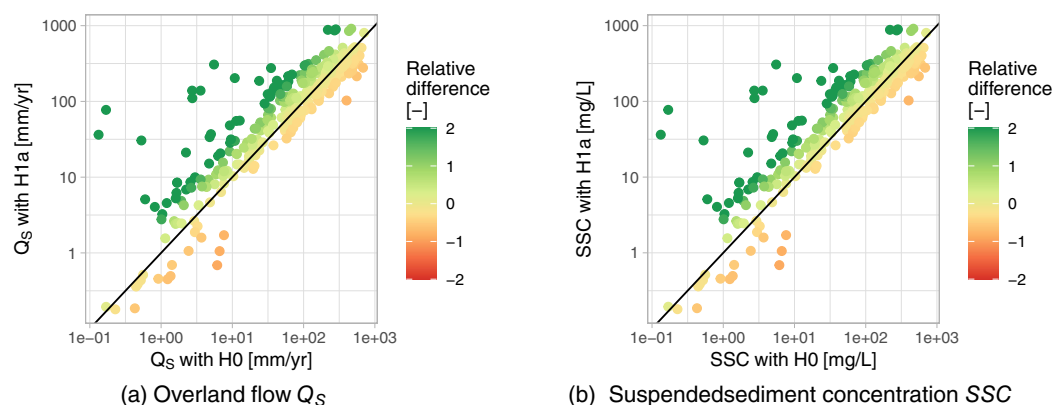
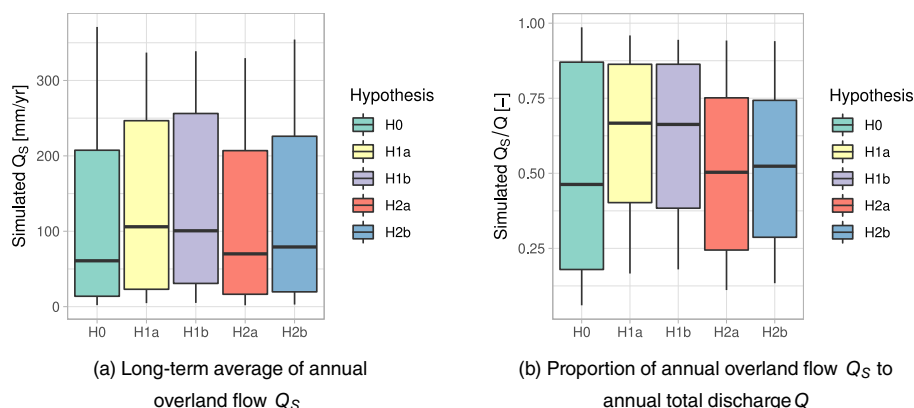


FIGURE 11 Comparison of mean annual simulated fluxes of overland flow (Q_S) and suspended sediment concentration (SSC) between H0 and H1a over the 111 catchments and the whole simulation period (1991–2018)

FIGURE 12 Comparison of annual fluxes of overland flow (Q_S) between modelling hypothesis during validation period. Q_S is computed as the sum of overland flow that has been generated over all the sub-catchments and Q is the river discharge at the outlet



calibration period (Figure 5). This highlights how much the mono-objective calibration was limiting the efficiency of the calibration of sediment routines. However, during validation period, the general improvement for sediment was not systematically kept, which highlights some lack of robustness of the sediment model (Figure 6). Only the new modelling hypothesis H2a appeared to keep the benefits of this significant improvement during validation period (results of Student *t*-test available in Appendix A.1: Table 3), demonstrating that the robustness of the sediment model can be potentially improved through new overland flow routines. This improvement concerned only the Pearson correlation component of the KGE, whereas the other components were generally not affected (details on KGE components are provided in Appendices A.1 and A.2: Figure 16). The hypothesis H1a was the only hypothesis where the improvement during calibration affected significantly each of the three components of the KGE. A map of the difference in performance between mono-objective and multi-objective calibration is presented for modelling hypothesis H1a in the text (Figure 7), and similar maps for the other modelling hypotheses are available in the Appendix A.4: Figures 19–23.

Discharge performances were generally less affected by this multi-objective calibration than sediment performances. As anticipated by Figure 2, the trade-off on the performance of discharge

irremediably leads to a slight loss of performance during calibration (verified by Figure 5). However, this loss of performance was non-significant according to a Student's *t*-test (Appendix A.1). Moreover, this loss was not systematically kept during validation: a few stations even achieved higher performances on discharge thanks to this multi-objective calibration (Figure 6). This demonstrates that more constraints on discharge calibration can potentially improve model robustness. This was particularly true for modelling hypothesis H1 where the highest improvement during validation was observed after accepting the highest loss of performance during calibration (Figure 7). However, this increase of model robustness due to multi-objective calibration was only found in a few stations.

3.2 | Comparison of modelling hypotheses

The second goal of this paper was to use this multi-objective calibration as a framework to evaluate new modelling hypotheses. It aims to detect which hypothesis could fit both objective functions and at the same time could potentially improve the model robustness through the constraint on two kinds of data.

Performances were generally very similar between modelling hypotheses for both discharge and sediment simulation according to

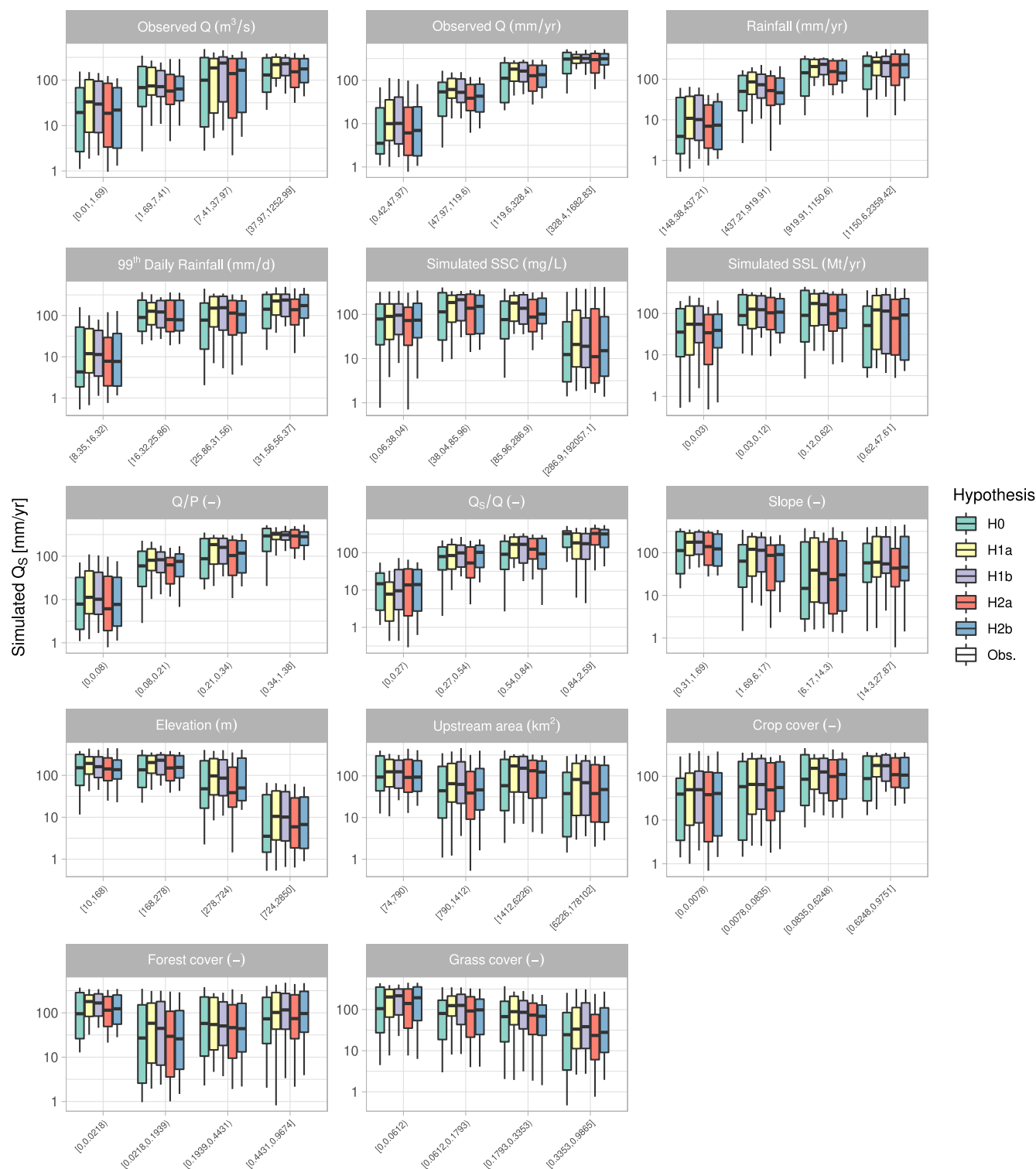


FIGURE 13 Statistical distribution of simulated overland flow (Q_s) by each modelling hypothesis during validation period and according to catchments hydroclimatic descriptors. Each hydroclimatic descriptor class is constructed on the quantile values of the 111 catchments, leading to approximately 27 catchments in each class. A similar graphic for overland flow ratio is provided in the Appendix B.6: Figure 34.

the KGE criteria (Figure 8): the slightly better performance, especially on the first quartile during validation, was not detected as a significant improvement by the Student t -test (Appendix B.1: Table 4). However, when looking at KGE components, the correlation criterion highlights the hypothesis H1 (H1a and H1b) as a significantly more

efficient model during validation period (Appendices B.1 and B.2: Figure 24).

The proportion of catchments where an improvement of KGE was observed with these new modelling hypotheses was generally higher compared to H0, especially with respect to the performance on

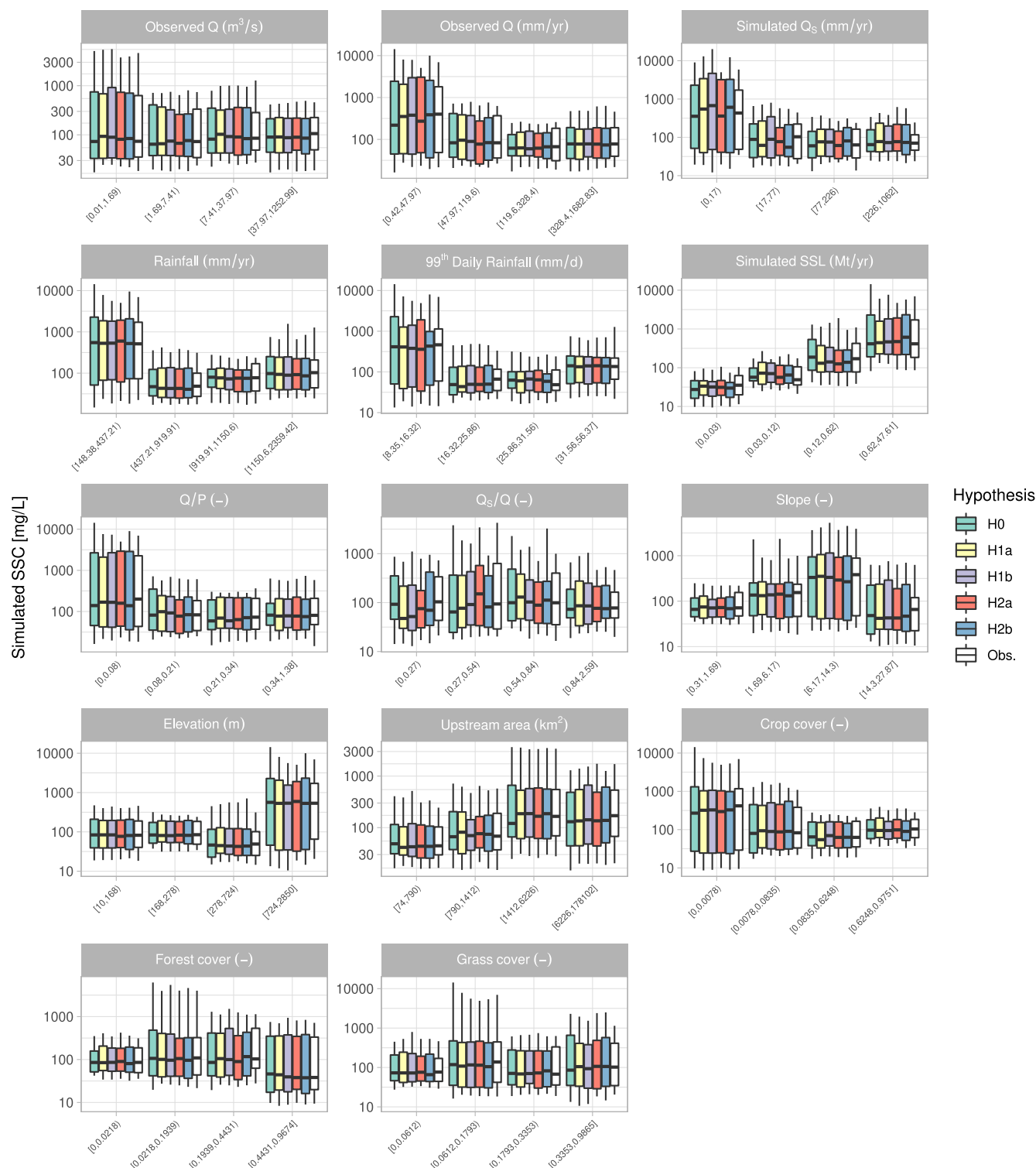


FIGURE 14 Statistical distribution of suspended sediment concentration SSC by each modelling hypothesis during validation period and according to catchments hydroclimatic descriptors. Each hydroclimatic descriptor class is constructed on the quantile values of the 111 catchments, leading to approximately 27 catchments in each class. A similar graphic for suspended sediment load SSL is provided in the Appendix B.7: Figures 35, 36.

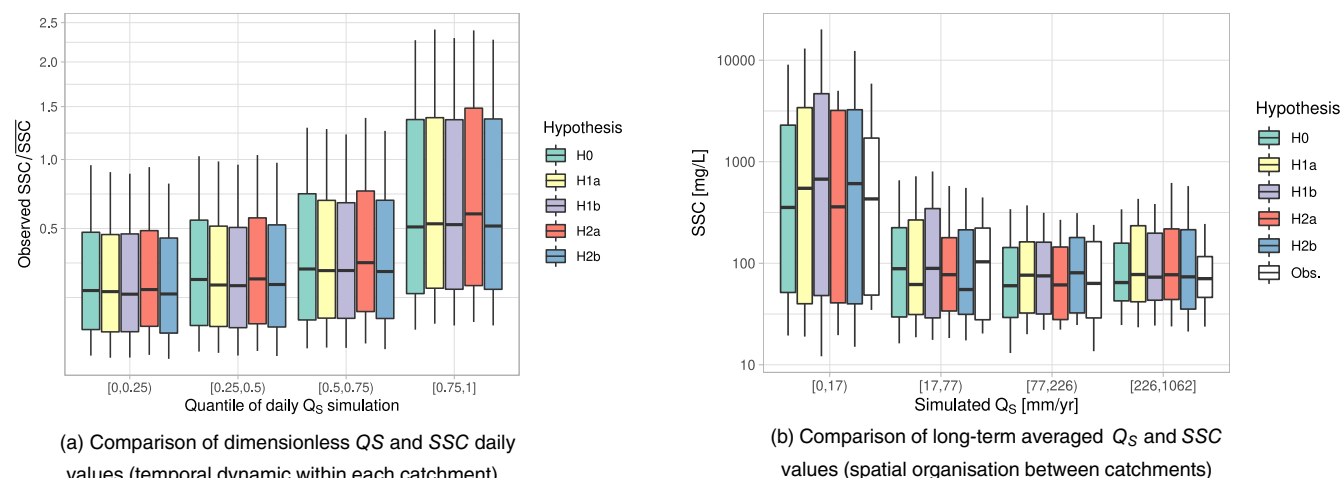


FIGURE 15 Relation between simulated overland flow (Q_S) and suspended sediment concentration (SSC) for each modelling hypothesis over the 111 catchments. Classes of Q_S are based on quantile values

sediment modelling (Figure 9). Hypothesis H2 appeared to be more robust than H1 for discharge modelling with fewer catchments where a decrease of performance is observed. Again, the overland flow routines affected the sediment simulation in a more unequivocal way than the discharge simulation (particularly during validation, Figure 9): suspended sediment concentration thus appears as a relevant observation to discriminate modelling hypotheses.

When considering model complexity in the evaluation (AIC, Equation 6), all new modelling hypotheses appear to be significantly more efficient (Appendix B.1). Indeed, despite almost similar performances, the number of parameters was lower leading therefore to a more parsimonious routine.

A spatial analysis of model improvement for modelling hypothesis H1a is available in Figure 10. Similar maps for the other hypotheses are provided in Appendix B.3: Figures 25–28. It highlights the mix of improvement and degradation of performances. Again, sediment modelling appears to be less robust than discharge modelling with a different spatial pattern between the calibration period and the validation period. No spatial patterns where one modelling hypothesis performs clearly better were found. However, when looking at catchments' characteristics, KGE improvements are generally observed for catchments with low sediment concentrations, high slopes and large catchment areas (Appendix B.4: Figure 29).

Apart from a comparison with observations, a comparison of the different simulated fluxes such as overland flow and sediment concentration also highlights the differences between the hypotheses. Figure 11 shows the differences between hypothesis H0 and hypothesis H1a (similar plots for the other hypotheses are available in Appendix B.5: Figures 30–33). All modelling hypotheses follow generally similar trends but differences can be locally quite high with more than 200 mm/year of overland flow and more than 500 mg/L of sediment which can represent more than twice the annual fluxes. This highlights that despite using two kinds of data for the calibration, uncertainty on these fluxes remains important.

Figure 12 presents the amount of simulated overland flow for each modelling hypothesis by summing overland flow that has been

generated over all the sub-catchments. The amount of simulated overland flow may seem relatively high. However, all the overland flow may not contribute to total river discharge Q because of subsequent infiltration, evaporation (e.g., from lakes), and exchanges between river discharge Q and groundwater. This could explain the relatively high values of overland flow ratio (Q_S/Q , Figure 12b).

Overall, the new modelling hypotheses H1 (H1a and H1b) were simulating more overland flow than the original one (Figure 12). Indeed, in the mathematical formulation of hypotheses H1 and H2, more hydro-meteorological conditions are likely to produce overland flow, as no threshold has to be exceeded. The modelling hypothesis H2 (H2a and H2b) produced an intermediate amount of overland flow between H0 and H1. This hypothesis is generating more overland flow when rainfall intensity is higher, so runoff generation is less smooth and regular in time, which leads to a lower annual volume according to Figure 12.

These differences among overland flow ratio (Q_S/Q , Figure 12b) highlight the equifinality issue that the paper aims to address: it is possible to reach very similar discharge performances through different simulations of overland flow. Without any additional data to evaluate this overland flow component, modelling hypothesis can hardly be ranked against each other in terms of physical realism. Evaluation on sediment concentrations aims to address this need of additional data to drive modelling hypothesis, as has proven useful in previous studies (see e.g., Bekele & Nicklow, 2007; Brighenti et al., 2019; Sikorska et al., 2015). When evaluating the behaviour of these hypotheses with respect to the catchment characteristics, drivers of overland flow and sediment generation can be further understood for each hypothesis.

The most influencing factors to drive overland flow appear to be rainfall, elevation and land use (Figure 13). Indeed, more overland flow is simulated when there is more rainfall, more discharge and a higher ratio of both (Q/P). The proportion of overland flow in total discharge follows the same trends (Appendix B.6: Figure 34). Overland flow is also well related to elevation, with less overland flow in higher elevation catchments. This tends to illustrate that lowlands are more likely to have larger humid areas (higher soil

wetness) that will generate more overland flow. Some trends are also detected with land uses: larger crop covers will generally produce larger overland flow, whereas grassland generally reduce the amount of overland flow.

Concerning the drivers of suspended sediment concentration, the most influencing factors appeared to be the average catchment slope, elevation, rainfall and drainage area (Figure 14). Sediment concentrations increased with the slope, but the steepest class has the lowest concentration. Indeed, higher slopes are generally more exposed to erosion, but in our database, steepest slopes are small catchments located in a relatively small area on the west coast, and often forested which could explain the behaviour of this last class. Sediment concentrations also tend to be higher for catchments located at higher elevations, with the most extreme concentrations observed in the last elevation class, which mainly correspond to the Rocky Mountains area. The higher concentrations are observed in dryer regions (with low amount of precipitation). Indeed, dry areas tend to have less vegetation, more soils exposed to erosion and smaller volume of runoff that lower dilution capacity. A gentle trend with catchments size is also observed with generally higher concentrations in larger catchments. These trends are generally well described by each modelling hypothesis.

3.3 | On the relation between overland flow and sediment concentration

The relationship between overland flow and suspended sediment concentration can be better understood in Figure 15: on a catchment-by-catchment basis (Figure 15a), more overland flow generally results in higher suspended sediment concentration, but when comparing catchments to each other (Figure 15b), more overland flow has lower sediment concentrations. In other words, suspended sediment concentration is informative in describing the temporal dynamics of overland flow for a given catchment, as exploited in this paper. However, this sediment information follows an opposite trend for understanding how overland flow is spatially distributed. This illustrates an effect of scale: more overland flow leads to more erosion (and thus higher loads) but also to more discharge which leads to lower concentrations if the increase in erosion is not as fast as the increase in discharge. This can be seen in Figure 14 with a decrease in suspended sediment concentration as the specific discharge is increased, as well as a relatively stable concentration between the different discharge ranges when expressed as volume per unit of time (m^3/s).

4 | LIMITATIONS AND PERSPECTIVES

Despite generally higher performances of the new modelling hypotheses, an understanding about the physical meaning of their parameters becomes reduced. Infiltration capacity and soil moisture threshold in hypothesis H0 are replaced by equation's exponents of a conceptual vision of overland flow at a catchment scale in the new hypotheses. While one could interpret the conceptual parameters as

representation of spatial variability in the physical parameters guiding infiltration and soil moisture processes, the improved performances of the more conceptual process descriptions also highlight that the threshold-based process in H0 is probably unlikely to reflect large spatial heterogeneity of the processes at a catchment scale. Indeed, this well-known scaling issue (Blöschl & Sivapalan, 1995) seems more easily addressed by the conceptual modelling hypotheses that avoid the use of threshold parameters that may not be applicable to sub-catchments of about 1000 km^2 .

Beyond model performance, the reduction of model complexity also makes the model calibration easier by reducing the required number of model runs for automatic calibration and the risk of equifinality. A further evaluation of these modelling hypotheses could rely on analysing the ease of regionalisation by linking parameter values to physiographic catchments characteristics. Following an HRU based calibration over the full domain (as in Arheimer et al., 2020) instead of a catchment-by-catchment calibration would allow further comparison to the original model and its ability to estimate their spatial distribution over a large domain.

The global model setup used for this study may affect the ability of this framework to identify differences among the individual model assumptions. With parameter values retrieved from a global scale regionalisation on monthly flows (Arheimer et al., 2020), model error might be locally important. Despite the recalibration of some sensitive parameters at a daily time step, as well as the catchment-by-catchment calibration strategy implemented in this study, this margin of error may be greater than the margin of improvement allowed by the new modelling hypotheses. The WWH model performance for the western US is notably low in the global model setup, such as in the Great Plains (Arheimer et al., 2020), and the partial recalibration in this study might not be able to fully address these issues. Theoretically, it is possible that this partial recalibration corrects for other processes and shifts water balance to higher overland flow. In addition, many anthropogenic impacts such as water transfers, municipal and industrial point sources, or flushing of sediments from reservoirs were not included in the WWH version used in this study. These may have large local impacts on sediment concentrations as well as stream discharge for certain catchments.

Similarly, the relatively coarse climatic inputs might also limit the evaluation of the impact of rainfall intensity on overland flow (comparison between hypothesis 1 and hypothesis 2). With a finer spatial and temporal resolution of rainfall inputs, the potential of the hypothesis 2, where it is explicitly taken into account, might be further explored. In spite of that potential limitation, accounting for rainfall intensity improved the model robustness: the higher performances on sediment modelling brought by multi-objective calibration were more often kept during validation period, and the balance in the proportion of catchments with improved and degraded performances was more favourable.

5 | CONCLUSION

As a first goal, this study evaluated the benefits of using a multi-objective calibration over a two-step mono-objective calibration for modelling overland flow and suspended sediment concentration.

Results show that multi-objective calibration brought a significant improvement of performances on sediments during calibration with an insignificant loss of performance on discharge. Hence, different parameter settings and internal model variables give similar results for discharge but different results for sediments. These results are in line with previous studies (e.g., Bergström et al., 2002; Her & Seong, 2018; Shafii et al., 2017) and call to move from sequential calibration to an integrated model calibration procedure that considers more aspects on model performance than just river discharge. It allows to avoid locking the parameter setting, which may be good for discharge but which is greatly limiting the performance of water quality modelling. This study thus encourages more joint work between water quantity and water quality modelling. However in our model, this improvement allowed by multi-objective calibration is reduced during validation period: simulations of total discharge and sediment simulations are improved only for a few catchments.

Overall, sediment concentration appears to be a good constraint for the calibration by being more sensitive to changes in overland flow modelling compared to total discharge. The multi-objective calibration is thus able to address some equifinality issues of the model where different parameter sets lead to similar performance on discharge: the use of a second kind of data, for example, sediment concentration such as in this study, helps to drive the optimisation by being more sensitive to the performance on the overland flow simulation.

However, even if a positive relation between the amount of overland flow and the sediment concentrations is observed when looking at the temporal dynamic of each catchment, an opposite trend in space (between catchments) is observed (with higher suspended sediment concentrations from catchments where there is less simulated overland flow). Sediment observation is thus also informative to understand non-linear relations between erosion rate and discharge at different scales, and thus open perspectives for regionalisation of overland flow routines.

As a second goal, we use this multi-objective framework to evaluate new modelling hypotheses for describing overland flow within the HYPE model. We propose to replace threshold parameters that are triggering overland flow in the HYPE model by a function that directly links the dynamics of overland flow to the dynamics of soil moisture and allows for a better description of the spatial variability in the processes. Performances on total discharges are not improved by the new modelling hypotheses, whereas performances on sediment concentrations appears to be significantly higher (according to the Pearson correlation coefficient). Beyond model performances, the new modelling hypotheses are also more parsimonious and are thus recommended in future set-ups of the HYPE model that would allow new evaluations of these hypotheses for other regions.

ACKNOWLEDGEMENTS

Detailed documentation and tutorials of the HYPE model is found at <https://hypeweb.smhi.se/model-water/>. This work was partially funded by several projects and thanks to a collective effort of the hydrological research unit (FoUh) at SMHI. The sediment modelling was developed by Bilateral Cooperation in South Africa funded by

Swedish Environmental Protection Agency as well as EU projects SPACE-O (H2020 GA 730005), DIRT-X (funded by JPI-AXIS, Grant No. 776608), and HYPOS (H2020 GA 870504).

FUNDING INFORMATION

Swedish Environmental Protection Agency with a Bilateral Cooperation in South Africa; EU projects: SPACE-O (H2020 GA 730005), DIRT-X (funded by JPI-AXIS, Grant No. 776608), and HYPOS (H2020 GA 870504).

DATA AVAILABILITY STATEMENT

The Hydrological Predictions for Environment (HYPE) model is available as open-source code under the Lesser GNU Public License. The code of the HYPE model is available at <https://sourceforge.net/projects/hype>. Discharge and sediment data are available from the USGS Water Data for the Nation website (U.S. Geological Survey, 2016) at <http://waterdata.usgs.gov/nwis/>.

ORCID

Alban de Lavenne  <https://orcid.org/0000-0002-9448-3490>

Berit Arheimer  <https://orcid.org/0000-0001-8314-0735>

REFERENCES

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd intr. symp. on information theory, Budapest*. Akademiai Kiado.
- Andersson, L., Rosberg, J., Pers, B. C., Olsson, J., & Arheimer, B. (2005). Estimating catchment nutrient flow with the HBV-NP model: Sensitivity to input data. *AMBIO: A Journal of the Human Environment*, 34, 521–532.
- Arheimer, B., & Brandt, M. (1998). Modelling nitrogen transport and retention in the catchments of southern Sweden. *Ambio*, 27, 471–480.
- Arheimer, B., Löwgren, M., Pers, B. C., & Rosberg, J. (2005). Integrated catchment modeling for nutrient reduction: Scenarios showing impacts, potential, and cost of measures. *AMBIO: A Journal of the Human Environment*, 34, 513–520.
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2020). Global catchment modelling using world-wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrology and Earth System Sciences*, 24, 535–559.
- Bartosova, A., Arheimer, B., de Lavenne, A., Capell, R., & Strömquist, J. (2021). Large-scale hydrological and sediment modeling in nested domains under current and changing climate. *Journal of Hydrologic Engineering*, 26, 05021009.
- Bekele, E. G., & Nicklow, J. W. (2007). Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology*, 341, 165–176.
- Berg, P., Donnelly, C., & Gustafsson, D. (2018). Near-real-time adjusted reanalysis forcing data for hydrology. *Hydrology and Earth System Sciences*, 22, 989–1000.
- Bergström, S. (1976). Development and application of a conceptual runoff model for scandinavian catchments. *Tech. Rep. 7*, SMHI Reports RHO, Norrköping.
- Bergström, S., Lindström, G., & Pettersson, A. (2002). Multi-variable parameter estimation to increase confidence in hydrological modelling. *Hydrological Processes*, 16, 413–421.
- Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16, 41–51.
- Beven, K. (2021). The era of infiltration. *Hydrology and Earth System Sciences*, 25, 851–866.
- Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, 9, 251–290.

- Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., Niel, J. D., Dewals, B. J., Drogue, G., Grelier, B., Melsen, L. A., Moustakas, S., Nossent, J., Pereira, F., Sprokkereef, E., Stam, J., Weerts, A. H., Willems, P., Savenije, H. H. G., & Hrachowitz, M. (2021). Behind the scenes of streamflow model performance. *Hydrology and Earth System Sciences*, 25, 1069–1095.
- Brighenti, T. M., Bonumá, N. B., Grison, F., de Almeida Mota, A., Kobiyama, M., & Chaffe, P. L. B. (2019). Two calibration methods for modeling streamflow and suspended sediment with the swat model. *Ecological Engineering*, 127, 103–113.
- Cheng, Q.-B., Chen, X., Wang, J., Zhang, Z.-C., Zhang, R.-R., Xie, Y.-Y., Reinhardt-Imjela, C., & Schulte, A. (2018). The use of river flow discharge and sediment load for multi-objective calibration of SWAT based on the bayesian inference. *Water*, 10, 1662.
- Didszun, J., & Uhlenbrook, S. (2008). Scaling of dominant runoff generation processes: Nested catchments approach using multiple tracers. *Water Resources Research*, 44, W02410.
- Dunne, T., & Black, R. D. (1970). Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research*, 6, 1296–1311.
- Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal*, 55, 58–78.
- Fenicia, F., McDonnell, J. J., & Savenije, H. H. G. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research*, 44, W06419.
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377, 80–91.
- Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34, 751–763.
- Guse, B., Pfannerstill, M., Gafurov, A., Kiesel, J., Lehr, C., & Fohrer, N. (2017). Identifying the connective strength between model parameters and performance criteria. *Hydrology and Earth System Sciences*, 21, 5663–5679.
- Her, Y., & Chaubey, I. (2015). Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty. *Hydrological Processes*, 29, 4220–4237.
- Her, Y., & Seong, C. (2018). Responses of hydrological model equifinality, uncertainty, and performance to multi-objective parameter calibration. *Journal of Hydroinformatics*, 20, 864–885.
- Horton, R. (1935). *Surface runoff phenomena. Part 1: Analysis of the hydrograph*. Edwards Bros. Horton Hydrological Laboratory.
- Klaus, J., & McDonnell, J. (2013). Hydrograph separation using stable isotopes: Review and evaluation. *Journal of Hydrology*, 505, 47–64.
- Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31, 13–24.
- Lidén, R. (1999). A new approach for estimation suspended sediment yield. *Hydrology and Earth System Sciences*, 3, 285–294.
- Lidén, R., Harlin, J., Karisson, M., & Rahmberg, M. (2001). Hydrological modelling of fine sediments in the Odzi River, Zimbabwe. *Water SA*, 27, 303–314.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201, 272–288.
- Lindström, G., Pers, C., Rosberg, J., Strömquist, J., & Arheimer, B. (2010). Development and testing of the HYPE (hydrological predictions for the environment) water quality model for different spatial scales. *Hydrology Research*, 41, 295–319.
- Lindström, G., Rosberg, J., & Arheimer, B. (2005). Parameter precision in the HBV-NP model and impacts on nitrogen scenario simulations in the Rönneå River, southern Sweden. *AMBIO: A Journal of the Human Environment*, 34, 533–537.
- Monteil, C., Zaoui, F., Moine, N. L., & Hendrickx, F. (2020). Multi-objective calibration by combination of stochastic and gradient-like parameter generation rules – The caRamel algorithm. *Hydrology and Earth System Sciences*, 24, 3189–3209.
- Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology*, 306, 127–145.
- Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., Nooijen, R. R. P., Savenije, H. H. G., & Hrachowitz, M. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, 54, 8332–8362.
- Pers, C., Temmerud, J., & Lindström, G. (2016). Modelling water, nutrients, and organic carbon in forested catchments: A HYPE application. *Hydrological Processes*, 30, 3252–3273.
- Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, 52, 7779–7792.
- Santos, L., Thirel, G., & Perrin, C. (2018). Continuous state-space representation of a bucket-type rainfall-runoff model: A case study with the GR4 model using state-space GR4 (version 1.0). *Geoscientific Model Development*, 11, 1591–1605.
- Seibert, J. (2000). Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, 4, 215–224.
- Shafii, M., Basu, N., Craig, J. R., Schiff, S. L., & Cappellen, P. V. (2017). A diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework. *Water Resources Research*, 53, 3279–3301.
- Sikorska, A., Giudice, D. D., Banasik, K., & Rieckermann, J. (2015). The value of streamflow data in improving TSS predictions – Bayesian multi-objective calibration. *Journal of Hydrology*, 530, 241–254.
- Tetzlaff, D., & Soulsby, C. (2008). Sources of baseflow in larger catchments – Using tracers to develop a holistic understanding of runoff generation. *Journal of Hydrology*, 359, 287–302.
- Tonderski, K., Andersson, L., Lindström, G., Cyr, R. S., Schönberg, R., & Taubald, H. (2017). Assessing the use of $\delta^{18}\text{O}$ in phosphate as a tracer for catchment phosphorus sources. *Science of the Total Environment*, 607–608, 1–10.
- U.S. Geological Survey. (2016). *National water information system data available on the World Wide Web (USGS water data for the nation)*. Accessed [April 28, 2020].
- van Griensven, A., & Meixner, T. (2007). A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models. *Journal of Hydroinformatics*, 9, 277–291.
- Wagener, T., McIntyre, N., Lees, M. J., Wheeler, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes*, 17, 455–476.
- Yapo, P. O., Gupta, H. V., & Sorooshian, S. (1998). Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, 204, 83–97.
- Zhao, R.-J. (1992). The Xinanjiang model applied in China. *Journal of Hydrology*, 135, 371–381.

How to cite this article: de Lavenne, A., Lindström, G., Strömquist, J., Pers, C., Bartosova, A., & Arheimer, B. (2022). Evaluation of overland flow modelling hypotheses with a multi-objective calibration using discharge and sediment data. *Hydrological Processes*, 36(12), e14767. <https://doi.org/10.1002/hyp.14767>

APPENDIX A: MULTI-OBJECTIVE VERSUS MONO-OBJECTIVE CALIBRATION PERFORMANCES FOR ALL MODELLING HYPOTHESES

A.1 | Student t-test results of statistical differences

Criteria	Hypothesis	Perf. loss on discharge		Perf. improvement on sediments	
		Calib.	Valid.	Calib.	Valid.
KGE	H0	No	No	Yes***	No
	H1a	No	No	Yes***	No
	H1b	No	No	Yes***	No
	H2a	No	No	Yes***	Yes**
	H2b	No	No	Yes***	No
<i>r</i>	H0	No	No	Yes***	No
	H1a	No	No	Yes***	No
	H1b	No	No	Yes***	No
	H2a	No	No	Yes***	Yes*
	H2b	No	No	Yes***	Yes*
μ_s/μ_o	H0	No	No	No	No
	H1a	No	No	Yes***	No
	H1b	No	No	No	No
	H2a	No	No	No	No
	H2b	No	No	No	No
σ_s/σ_o	H0	No	No	No	No
	H1a	No	No	Yes**	No
	H1b	No	No	No	No
	H2a	No	No	No	No
	H2b	No	No	No	No

Note: For discharge, the test is on a potential decrease of performance. For sediment, the test is on a potential increase of performance. The threshold on *p* value to distinguish yes/no is 0.05. Statistical significance: *For *p* value < 0.05; **for *p* value < 0.01, and ***for *p* value < 0.001.

TABLE 3 Is multi-objective calibration significantly impacting performance criteria compared to mono-objective calibration?

A.2 | Boxplots comparison of model performances

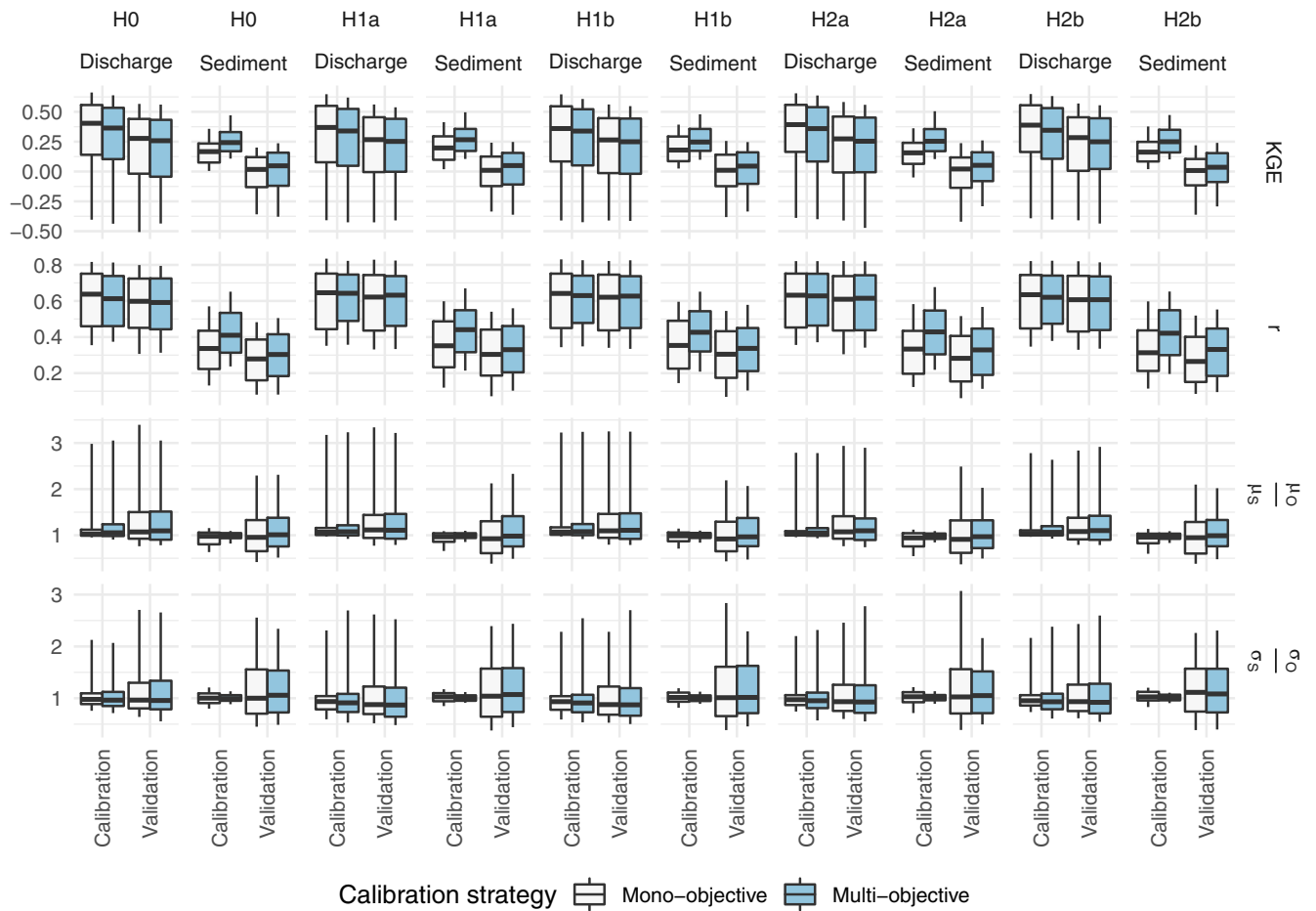


FIGURE 16 Boxplot comparison of performances on sediment and discharge for mono-objective and multi-objective optimisation for all modelling hypotheses and KGE components (Pearson's correlation r , mean ratio μ_S/μ_O , and variability ratio σ_S/σ_O)

A.3 | Proportion of catchments where performance is affected

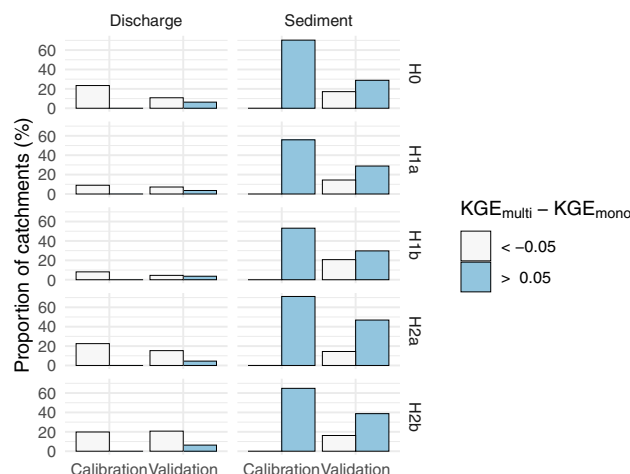


FIGURE 17 Proportion of the 111 catchments where a difference of KGE performances above 0.05 is observed between optimisation strategies for each modelling hypothesis: Blue highlights better performances of the multi-objective calibration and grey highlights better performances of mono-objective calibration

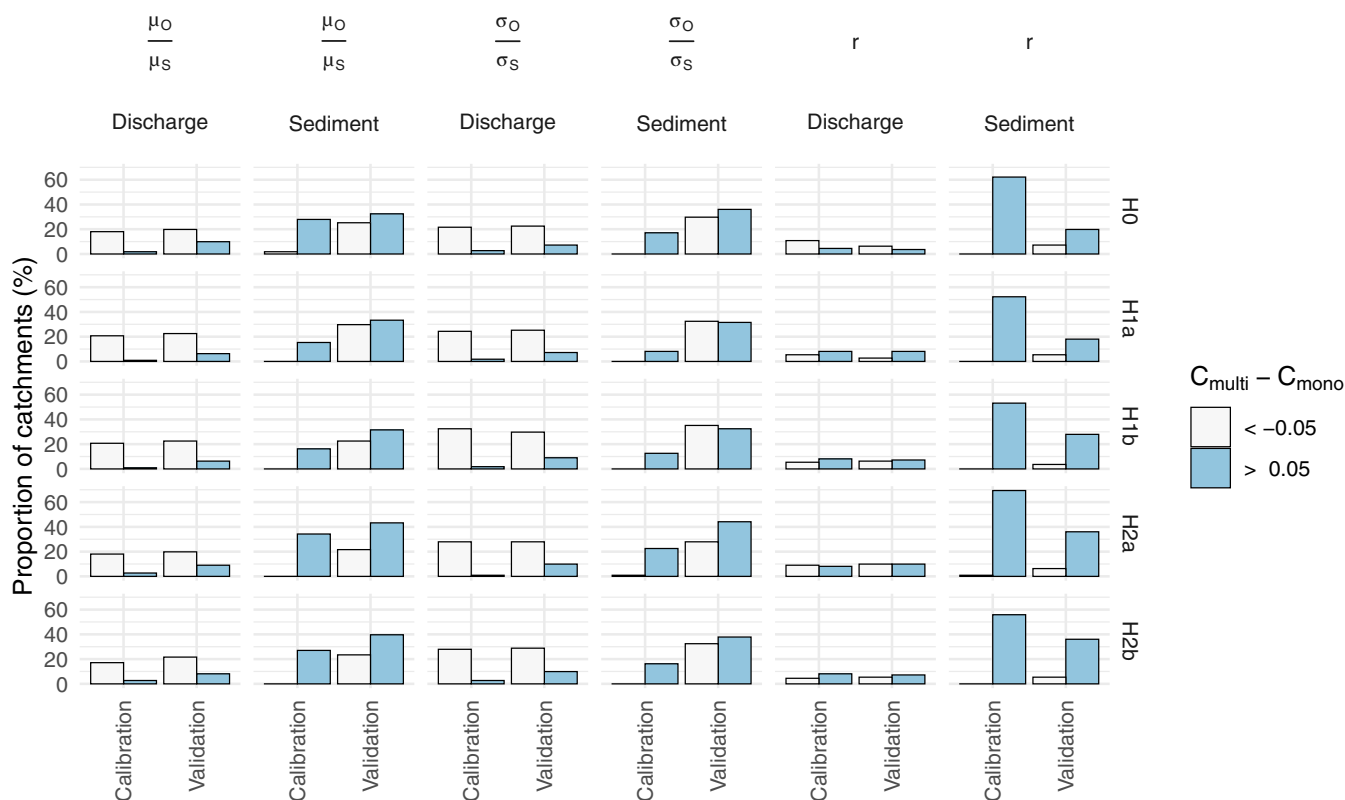


FIGURE 18 Proportion of the 111 catchments where a difference above 0.05 is observed between optimisation strategies for each KGE components C (Pearson correlation r , mean ratio μ_S/μ_O , and variability ratio σ_S/σ_O): Blue highlights better performances of the multi-objective calibration and grey highlights better performances of mono-objective calibration for each modelling hypothesis

A.4 | Spatial analysis of performance criteria

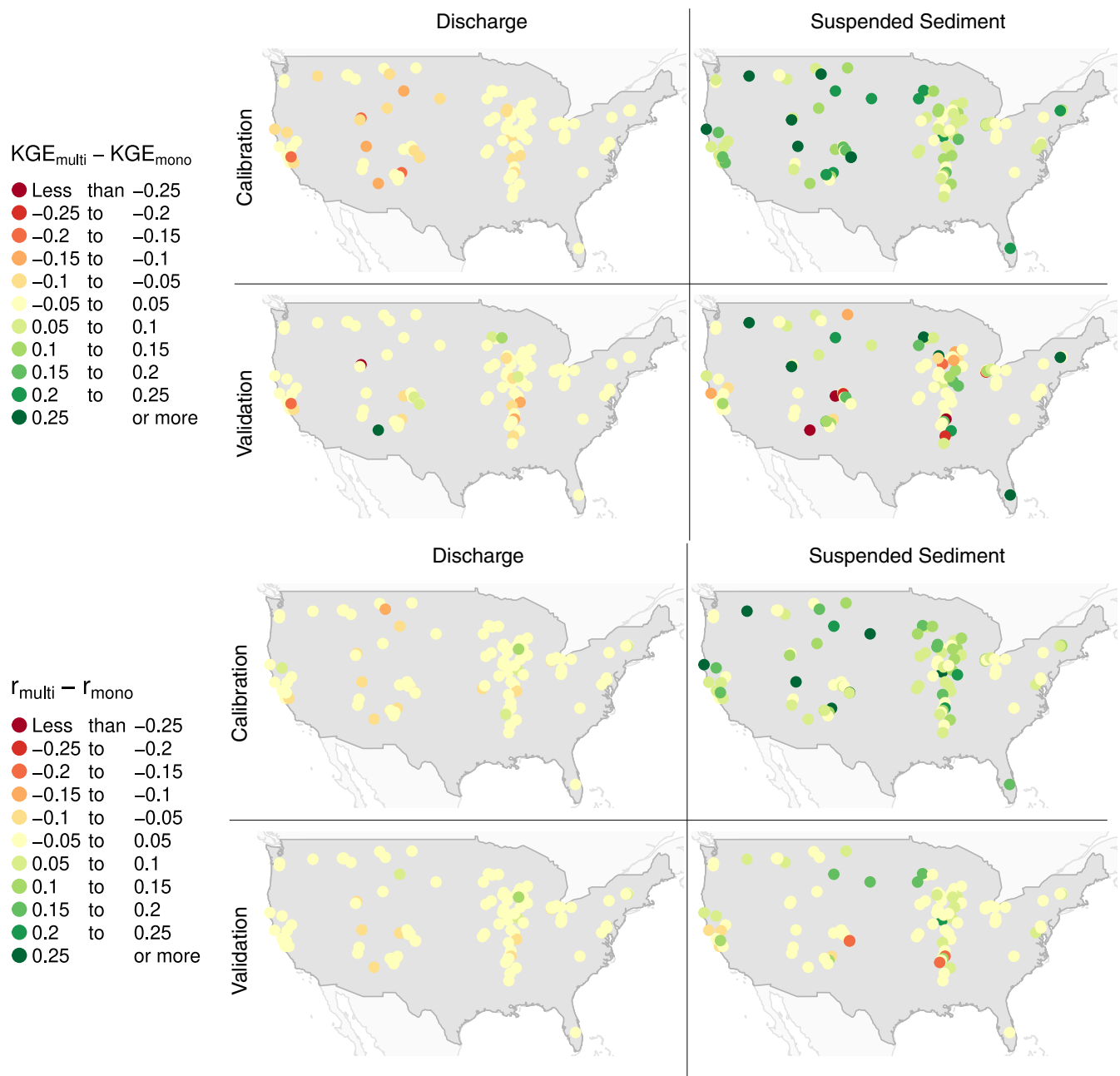


FIGURE 19 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between mono-objective and multi-objective calibration of modelling hypothesis H0

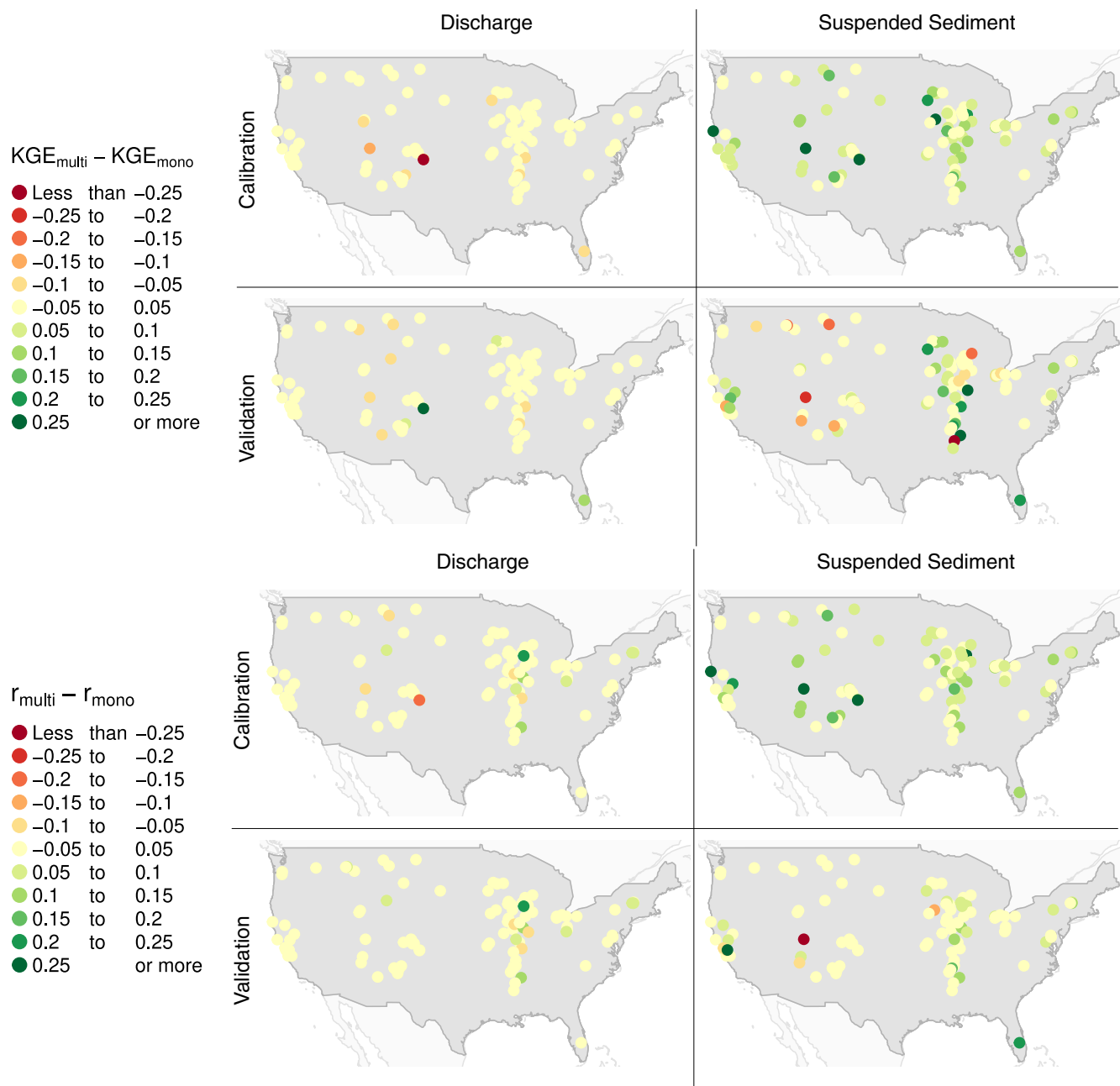


FIGURE 20 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between mono-objective and multi-objective calibration of modelling hypothesis H1a

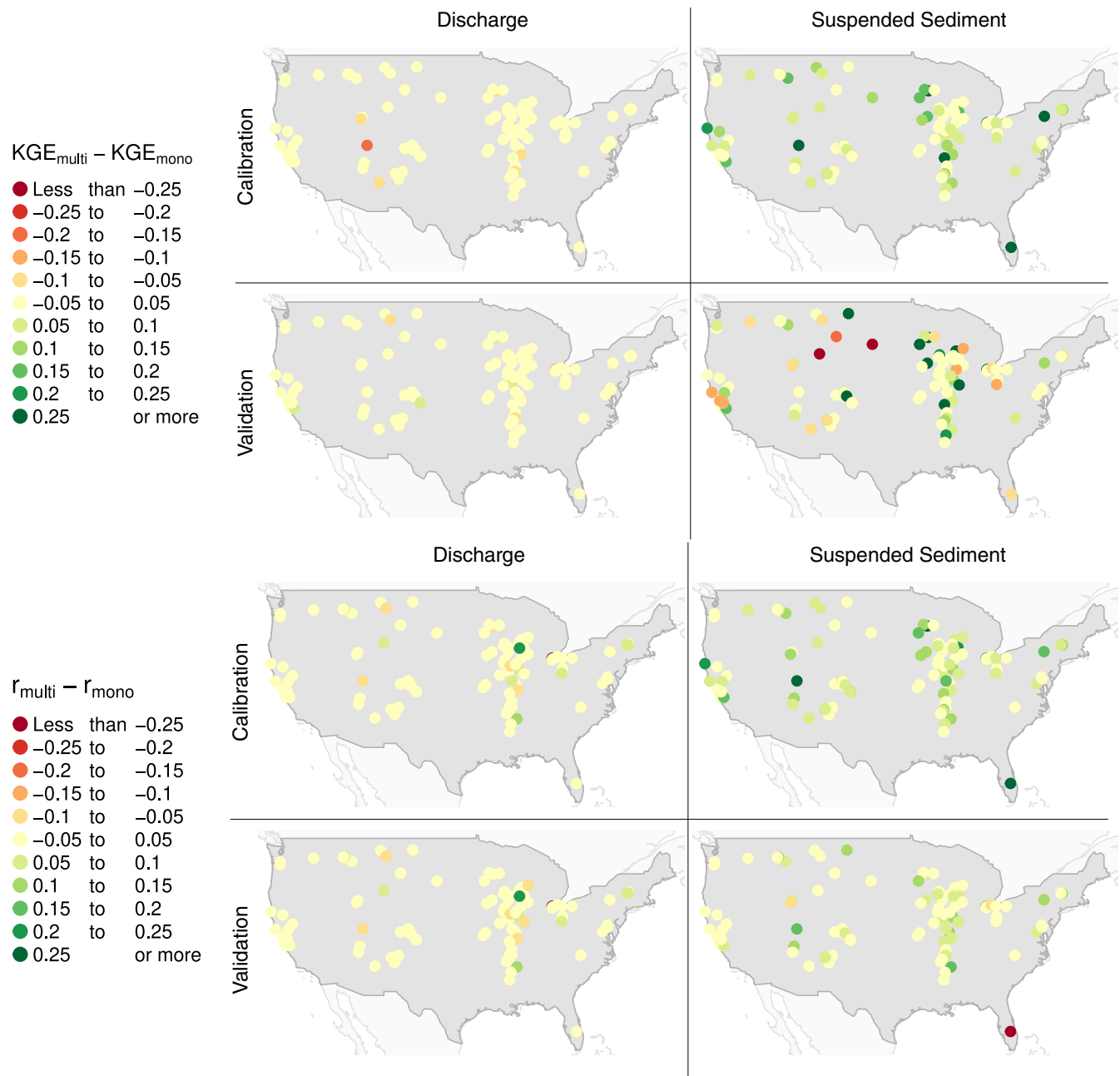


FIGURE 21 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between mono-objective and multi-objective calibration of modelling hypothesis H1b

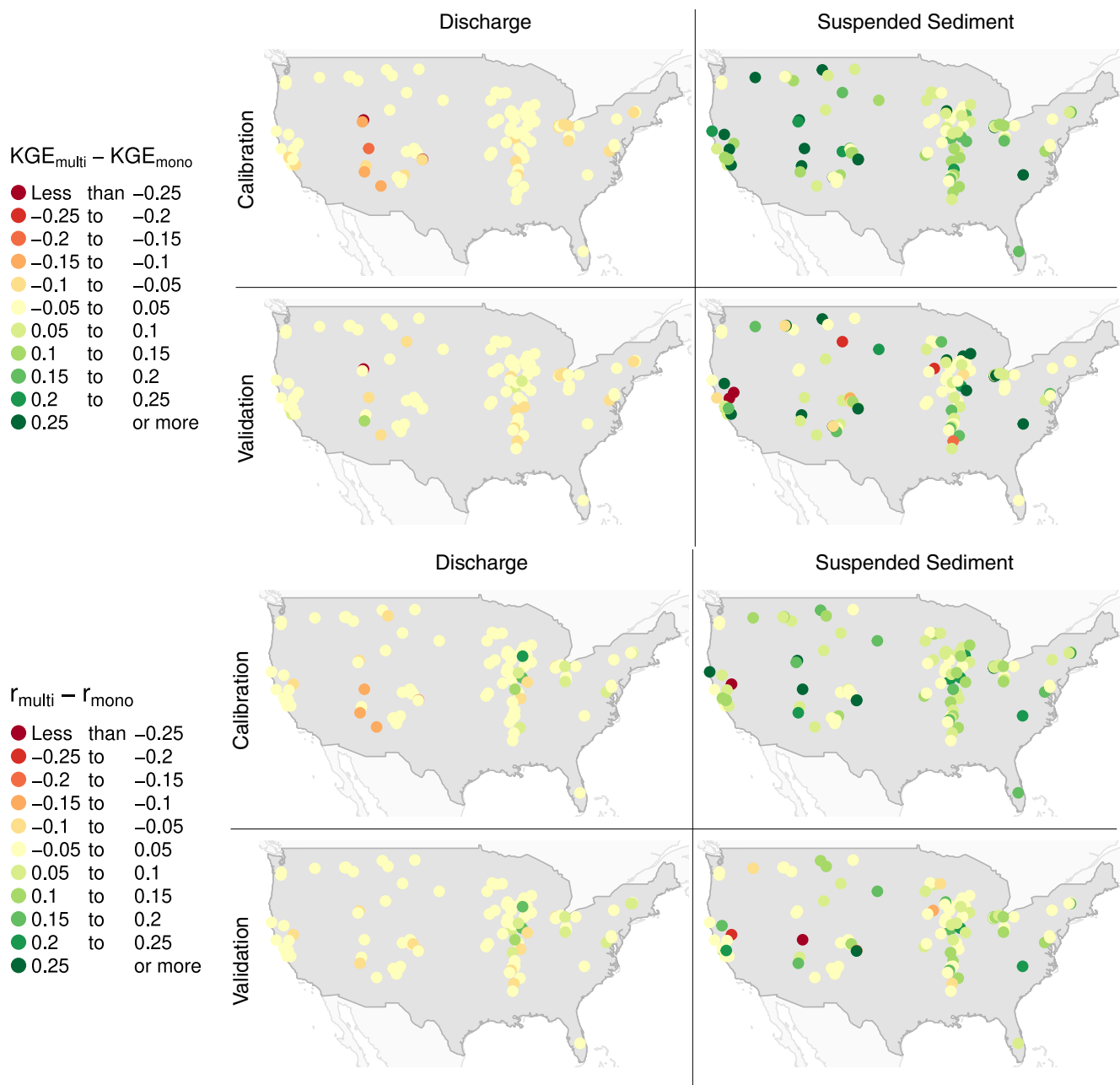


FIGURE 22 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between mono-objective and multi-objective calibration of modelling hypothesis H2a

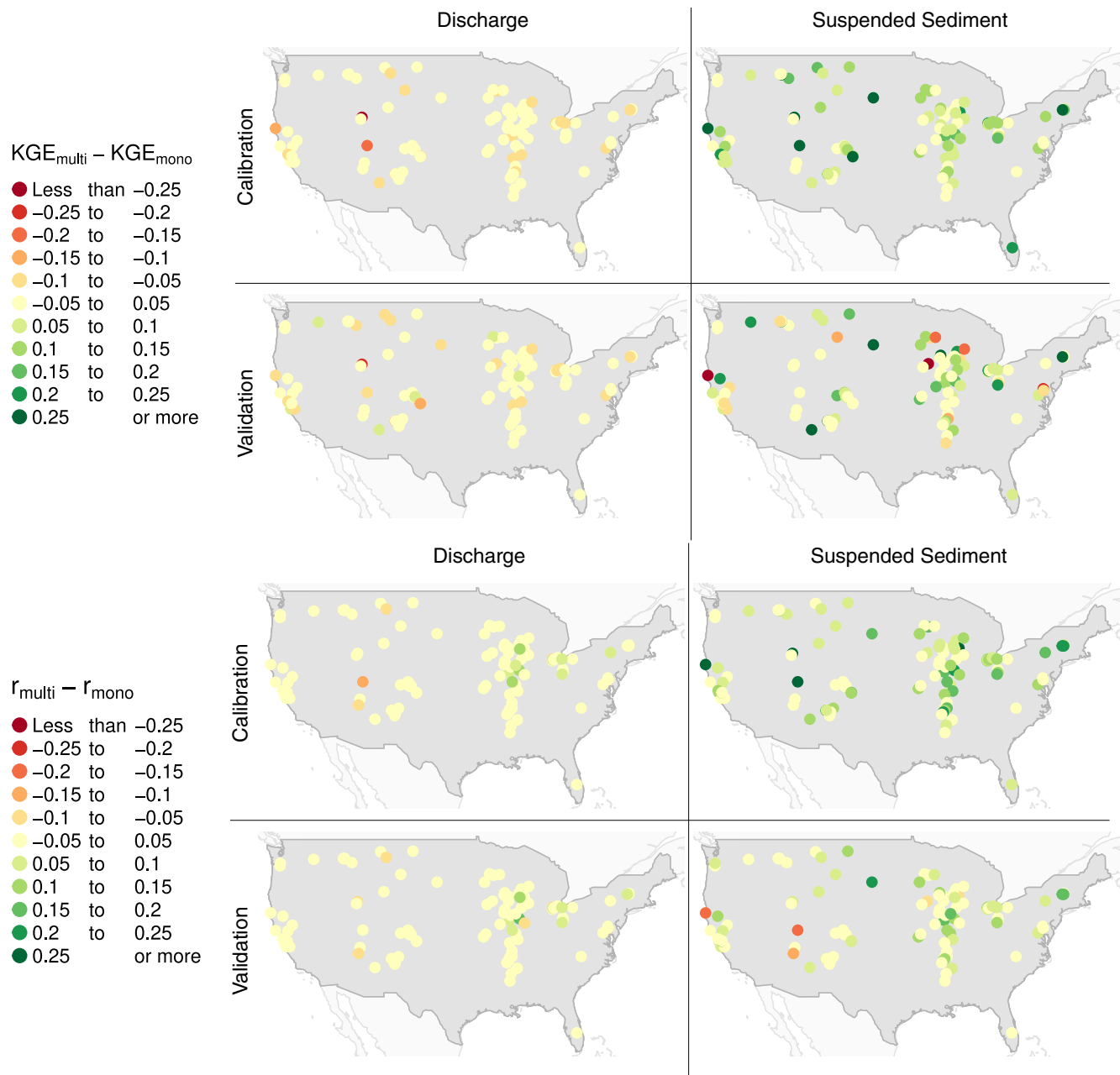


FIGURE 23 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between mono-objective and multi-objective calibration of modelling hypothesis H2b

APPENDIX B: COMPARISON OF THE PERFORMANCES OF THE NEW MODELLING HYPOTHESES WITH THE INITIAL H0 HYPOTHESIS

B.1 | Student t-test results of statistical differences

Criteria	Hypothesis	Perf. improvement on discharge		Perf. improvement on sediments	
		Calib.	Valid.	Calib.	Valid.
KGE	H1a	No	No	No	No
	H1b	No	No	No	No
	H2a	No	No	No	No
	H2b	No	No	No	No
<i>r</i>	H1a	No	No	No	Yes*
	H1b	No	No	No	Yes*
	H2a	No	No	No	No
	H2b	No	No	No	No
μ_s/μ_o	H1a	No	No	No	No
	H1b	No	No	No	No
	H2a	No	No	No	No
	H2b	No	No	No	No
σ_s/σ_o	H1a	No	No	No	No
	H1b	No	No	No	No
	H2a	No	No	No	No
	H2b	No	No	No	No
AIC	H1a	Yes***	Yes***	Yes***	Yes***
	H1b	Yes***	Yes***	Yes***	Yes***
	H2a	Yes***	Yes***	Yes***	Yes***
	H2b	Yes***	Yes***	Yes***	Yes***

Note: Calibration uses the multi-objective solution. The threshold on *p* value to distinguish yes/no is 0.05. Statistical significance: *For *p* value < 0.05, **for *p* value < 0.01 and ***for *p* value < 0.001.

TABLE 4 Are new modelling hypotheses significantly improving performance criteria compared to H0 hypothesis?

B.2 | Proportion of catchments for which performance is affected

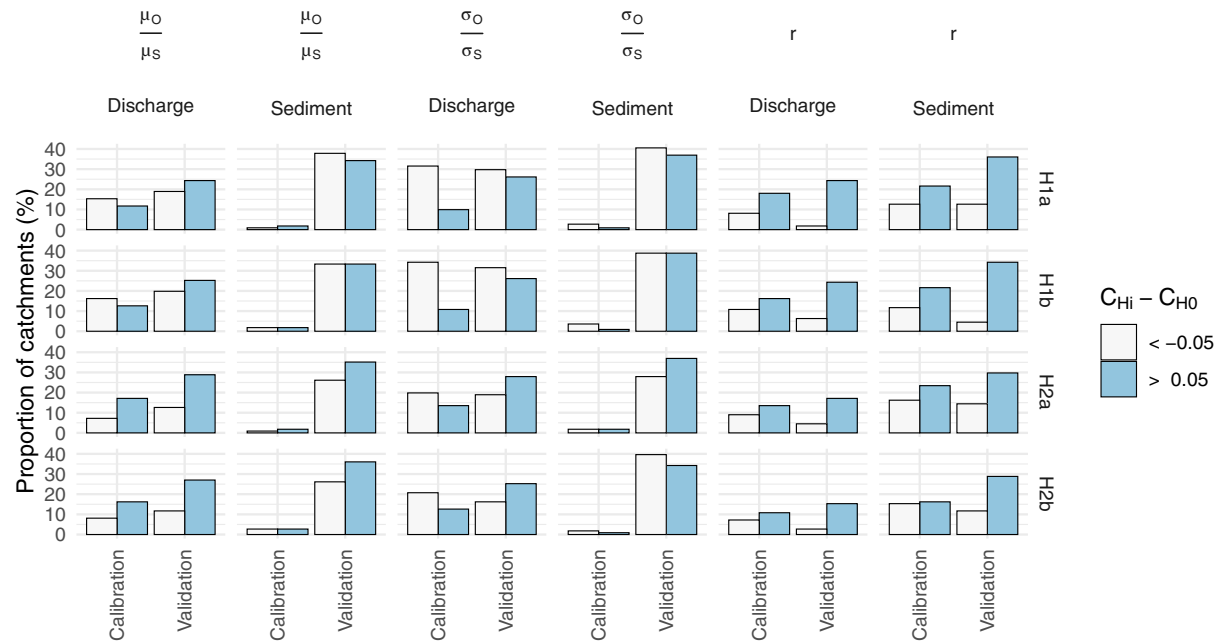


FIGURE 24 Proportion of the 111 catchments where a difference above 0.05 is observed between the original modelling hypothesis H0 and the new ones (H1a, H1b, H2a, and H2b) for each KGE components C (Pearson's correlation r , mean ratio μ_S/μ_O , and variability ratio σ_S/σ_O): Blue highlights better performances for the new hypotheses and grey highlights worse performances

B.3 | Spatial analysis of performance criteria

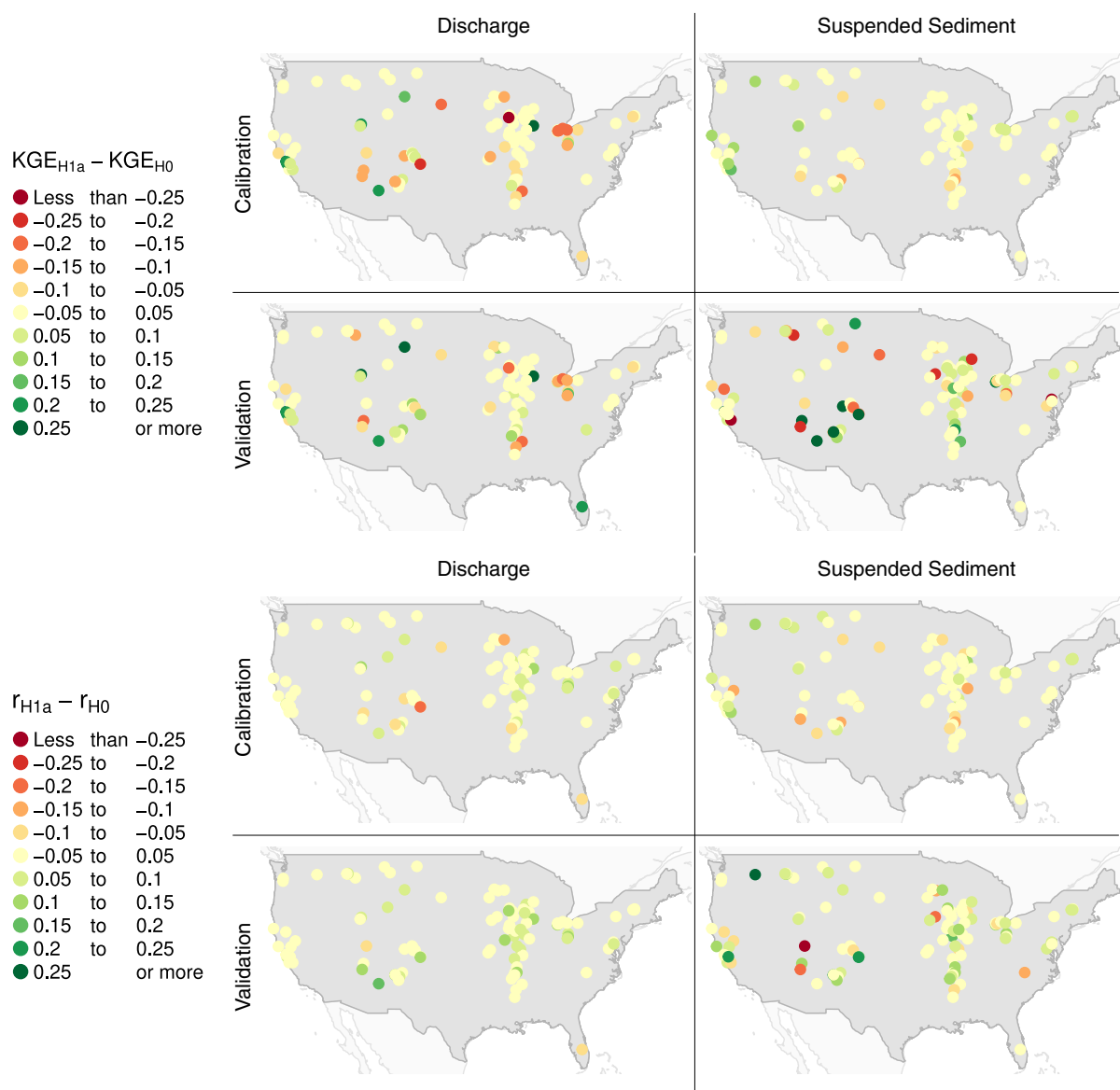


FIGURE 25 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between modelling hypothesis H1a and modelling hypothesis H0

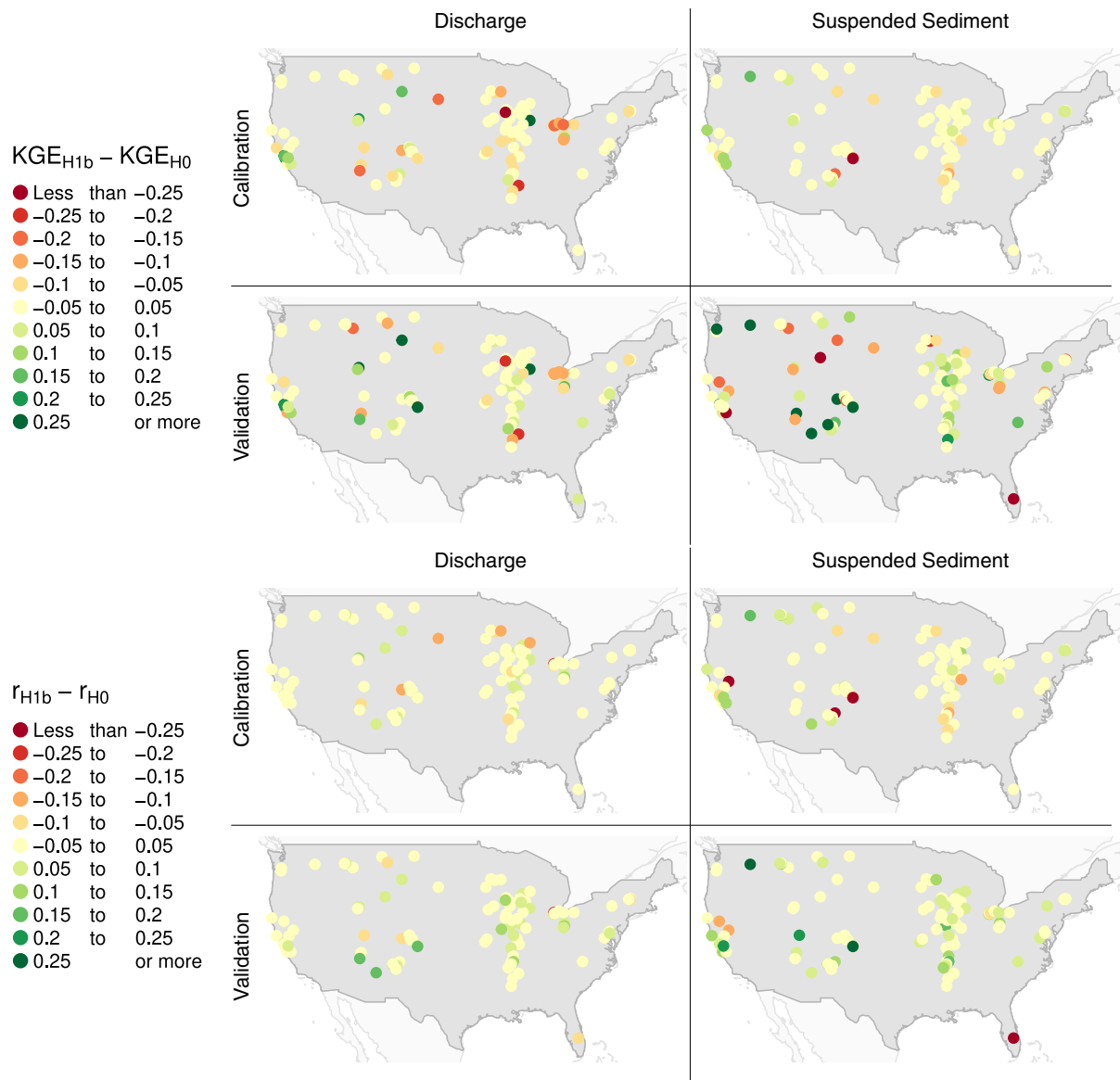


FIGURE 26 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between modelling hypothesis H1b and modelling hypothesis H0

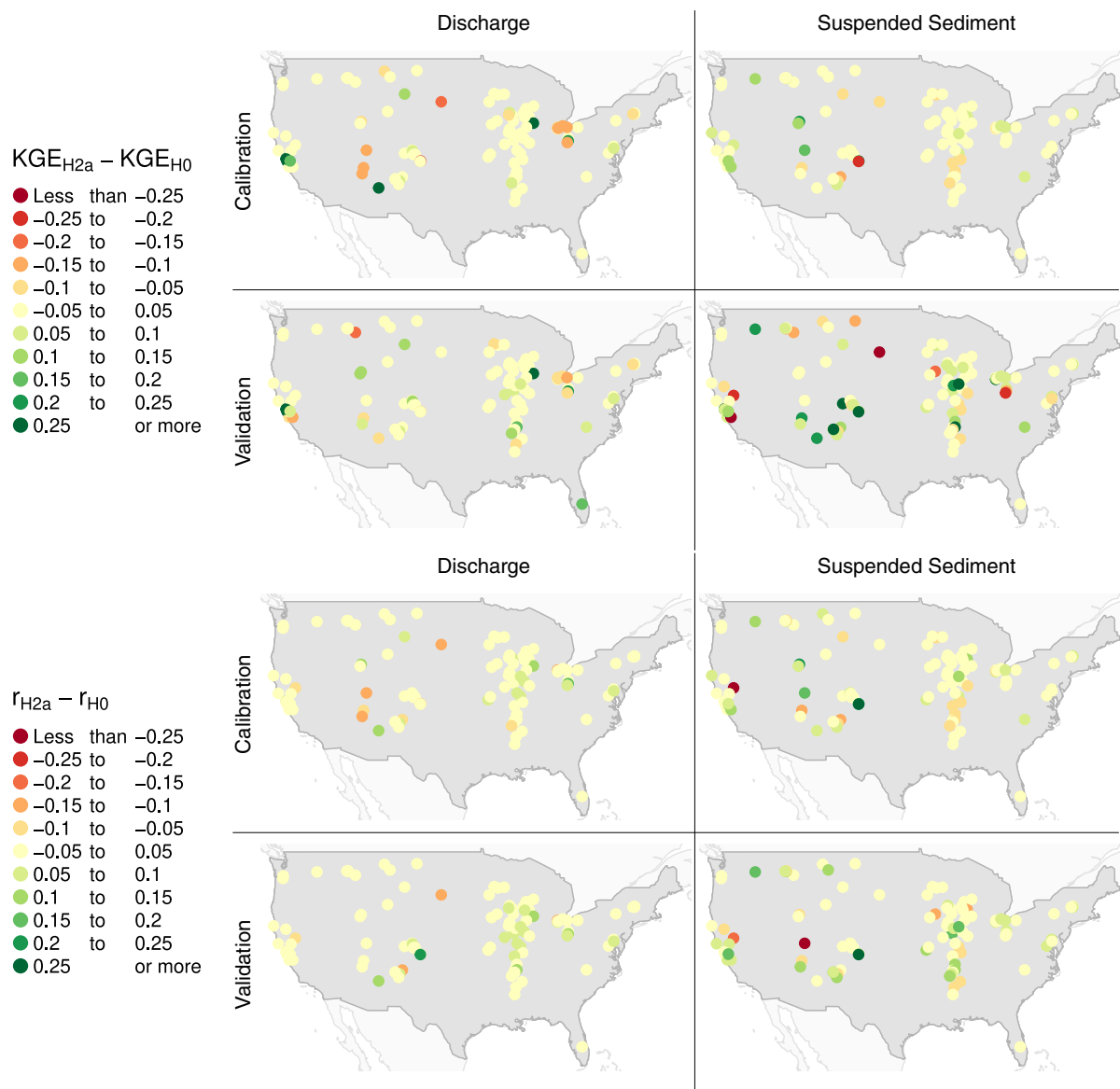


FIGURE 27 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between modelling hypothesis H2a and modelling hypothesis H0

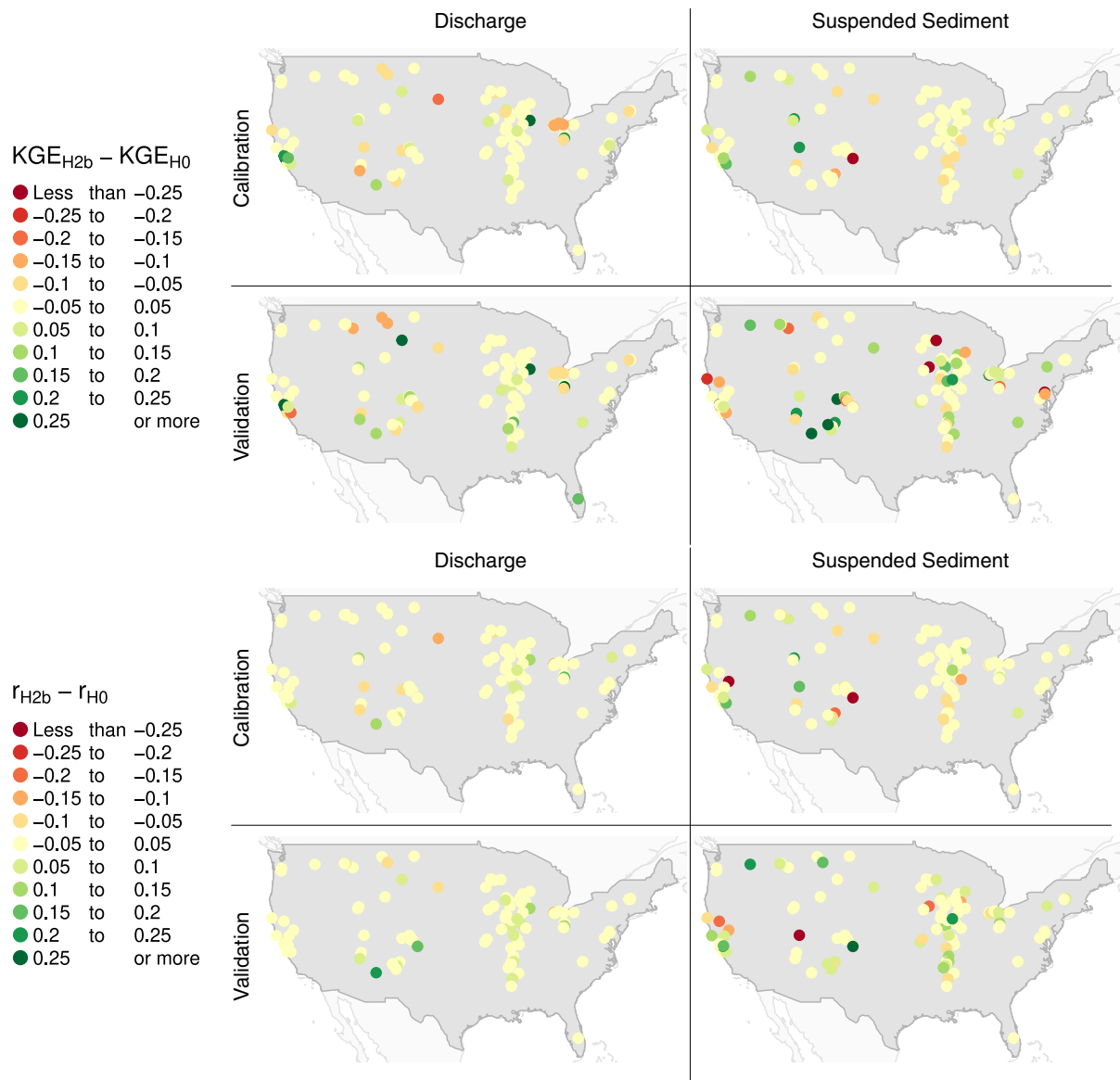


FIGURE 28 KGE and Pearson correlation (r) performance comparison on suspended sediment concentration and total discharge between modelling hypothesis H2b and modelling hypothesis H0

B.4 | Differences in simulation performance of suspended sediment concentrations according to catchment descriptors

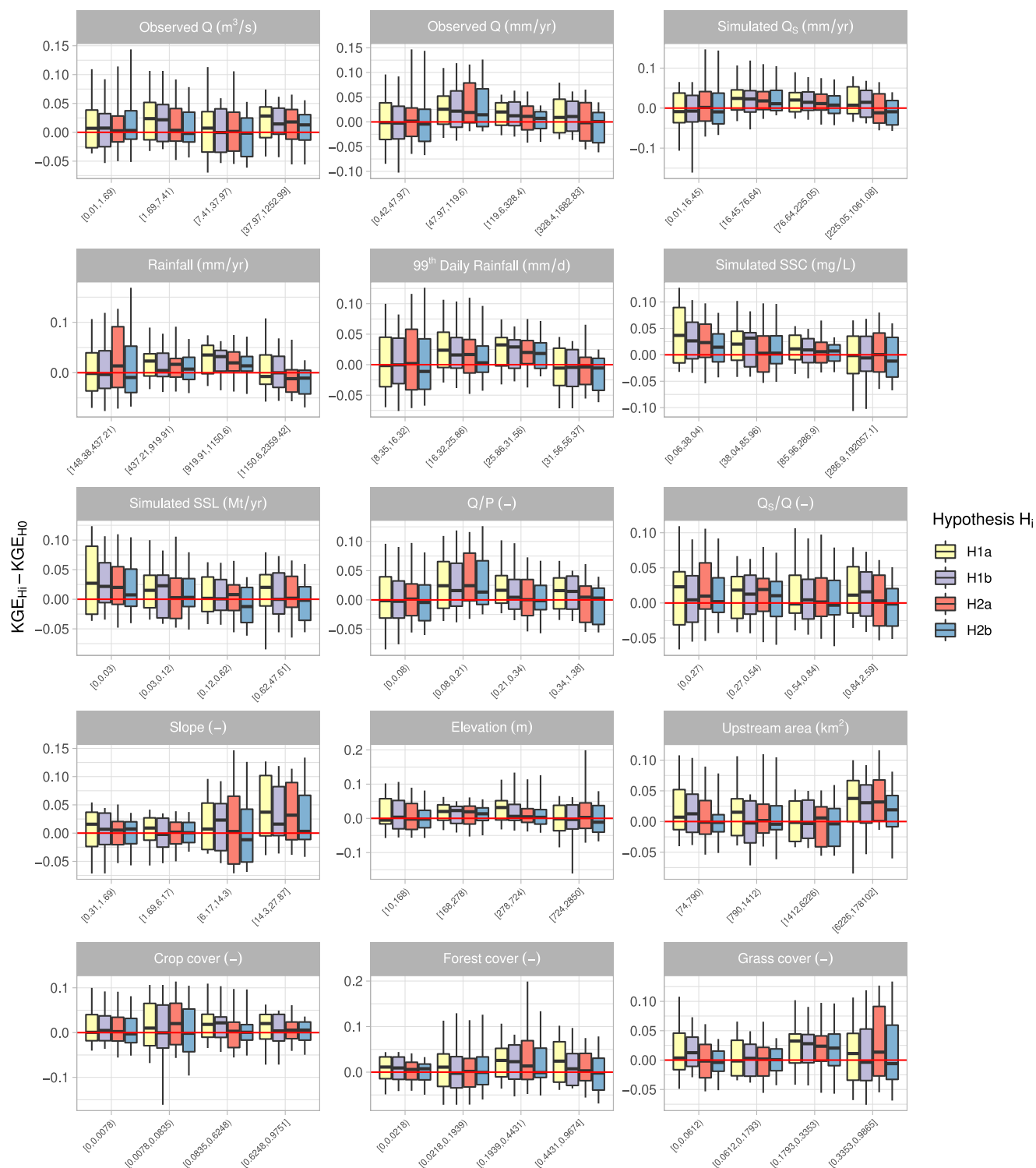


FIGURE 29 KGE performances comparison on suspended sediment concentration between each modelling hypothesis and the modelling hypothesis H0 during calibration period and according to catchments hydroclimatic descriptors. Each hydroclimatic descriptor class is constructed on the quantile values of the 111 catchments, leading to approximately 27 catchments in each class

B.5 | Comparison of annual fluxes

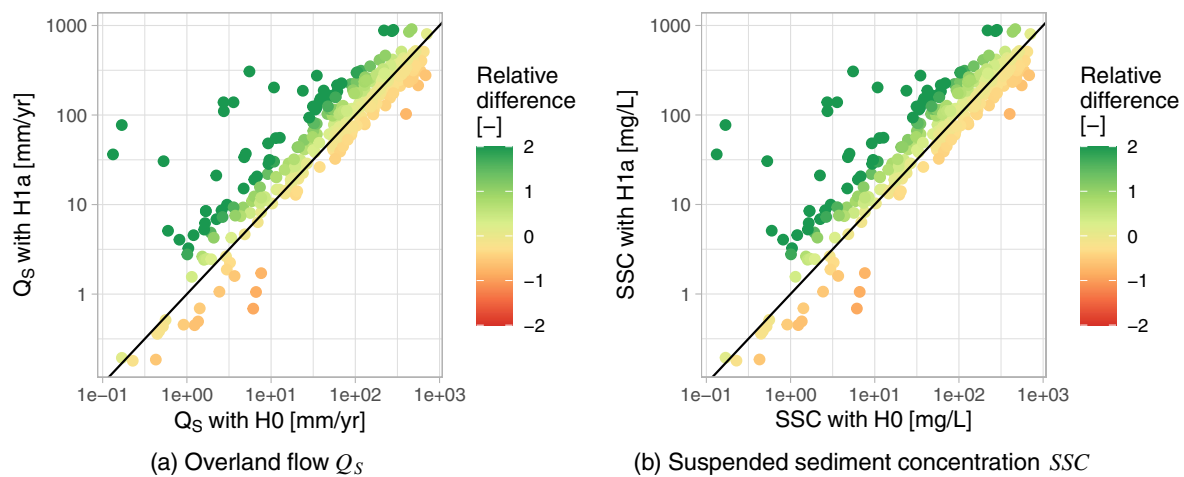


FIGURE 30 Comparison of annual fluxes of overland flow (Q_S) and suspended sediment concentration (SSC) between H0 and H1a

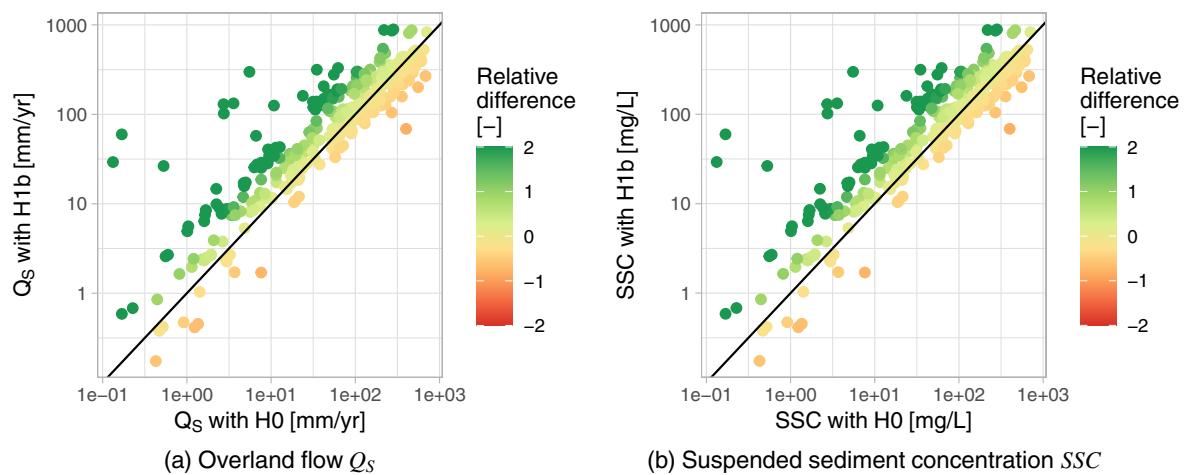


FIGURE 31 Comparison of annual fluxes of overland flow (Q_S) and suspended sediment concentration (SSC) between H0 and H1b

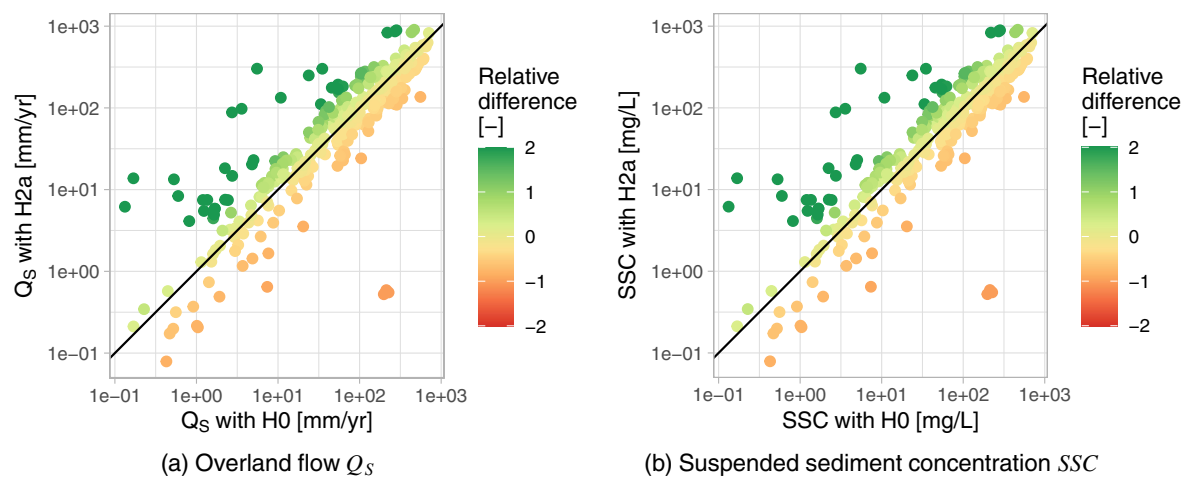


FIGURE 32 Comparison of annual fluxes of overland flow (Q_S) and suspended sediment concentration (SSC) between H0 and H2a

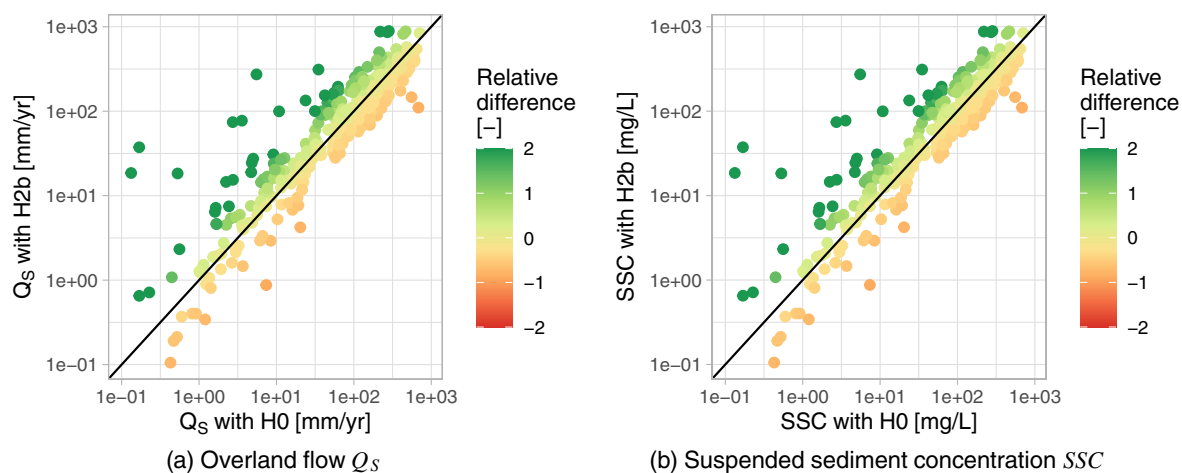


FIGURE 33 Comparison of annual fluxes of overland flow (Q_S) and suspended sediment concentration (SSC) between H0 and H2b

B.6 | Relation between catchment characteristics and overland flow ratio

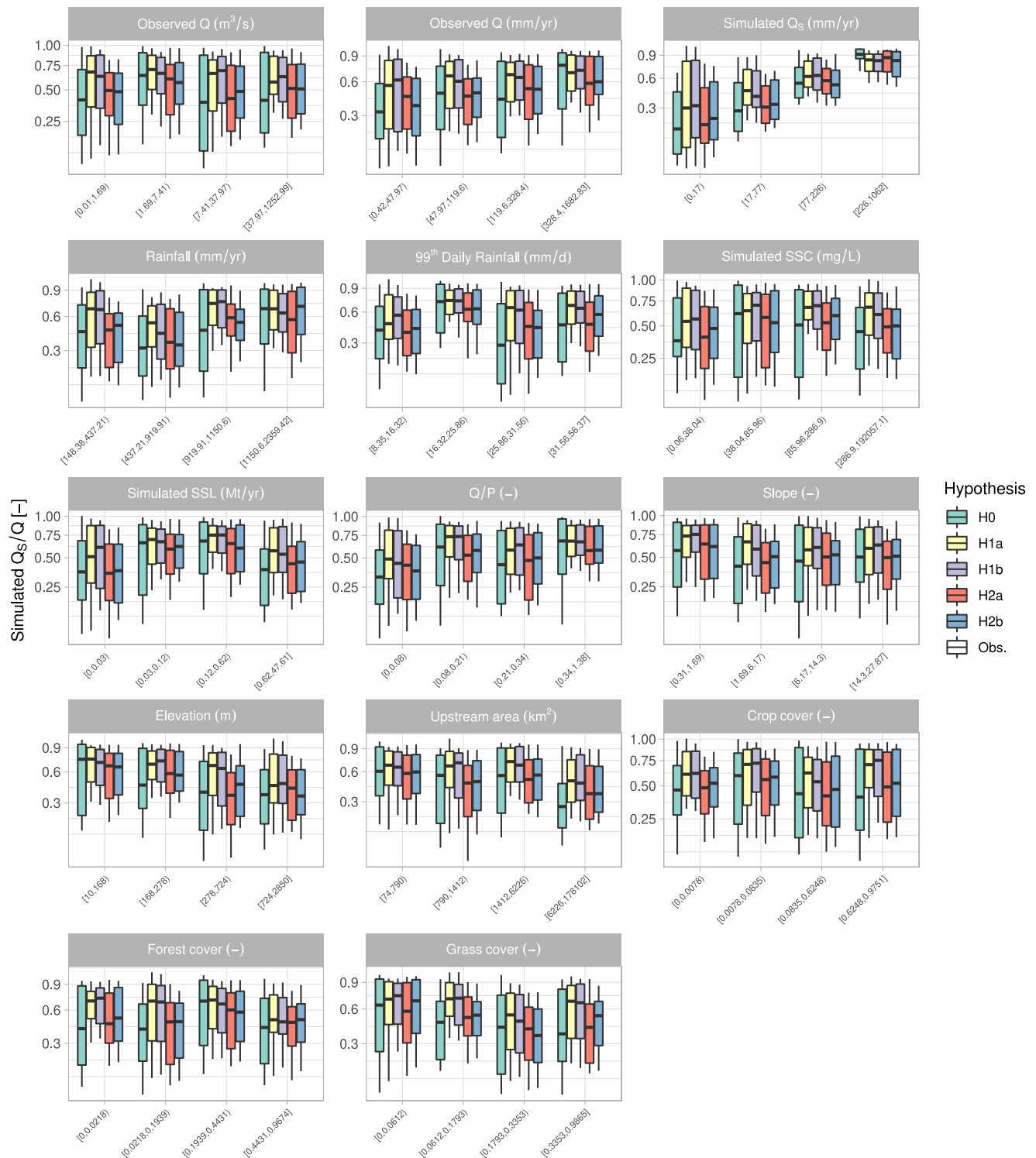


FIGURE 34 Statistical distribution of simulated overland flow ratio (Q_s/Q) by each modelling hypothesis during validation period and according to catchments hydroclimatic descriptors. Each hydroclimatic descriptor class is constructed on the quantile values of the 111 catchments, leading to approximately 27 catchments in each class

B.7 | Relation between catchment characteristics and suspended sediment load

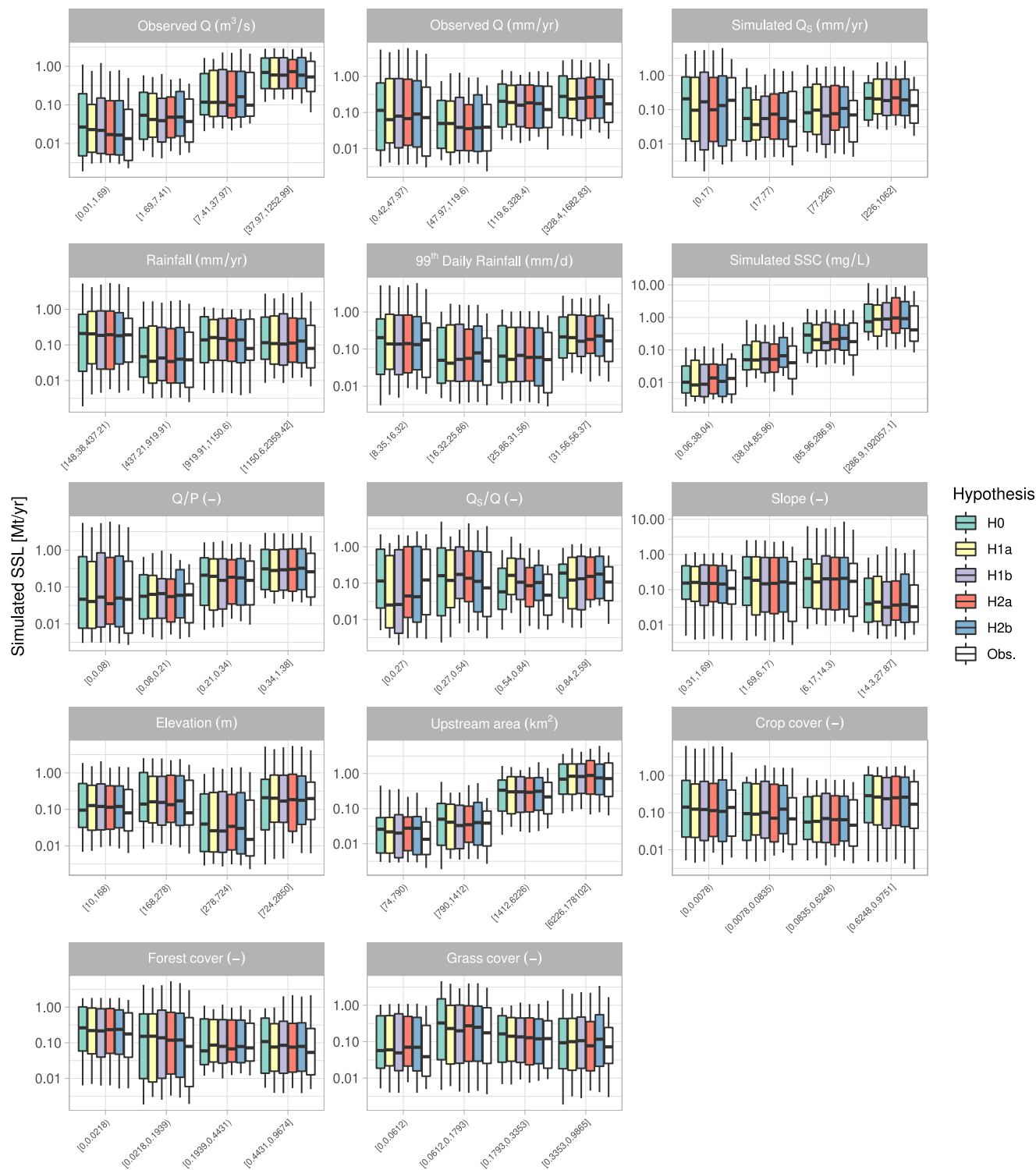


FIGURE 35 Statistical distribution of suspended sediment load SSL by each modelling hypothesis during validation period and according to catchments hydroclimatic descriptors. Each hydroclimatic descriptor class is constructed on the quantile values of the 111 catchments, leading to approximately 27 catchments in each class

B.8 | Spatial analysis of simulation performances for each modelling hypothesis with multi-objective calibration

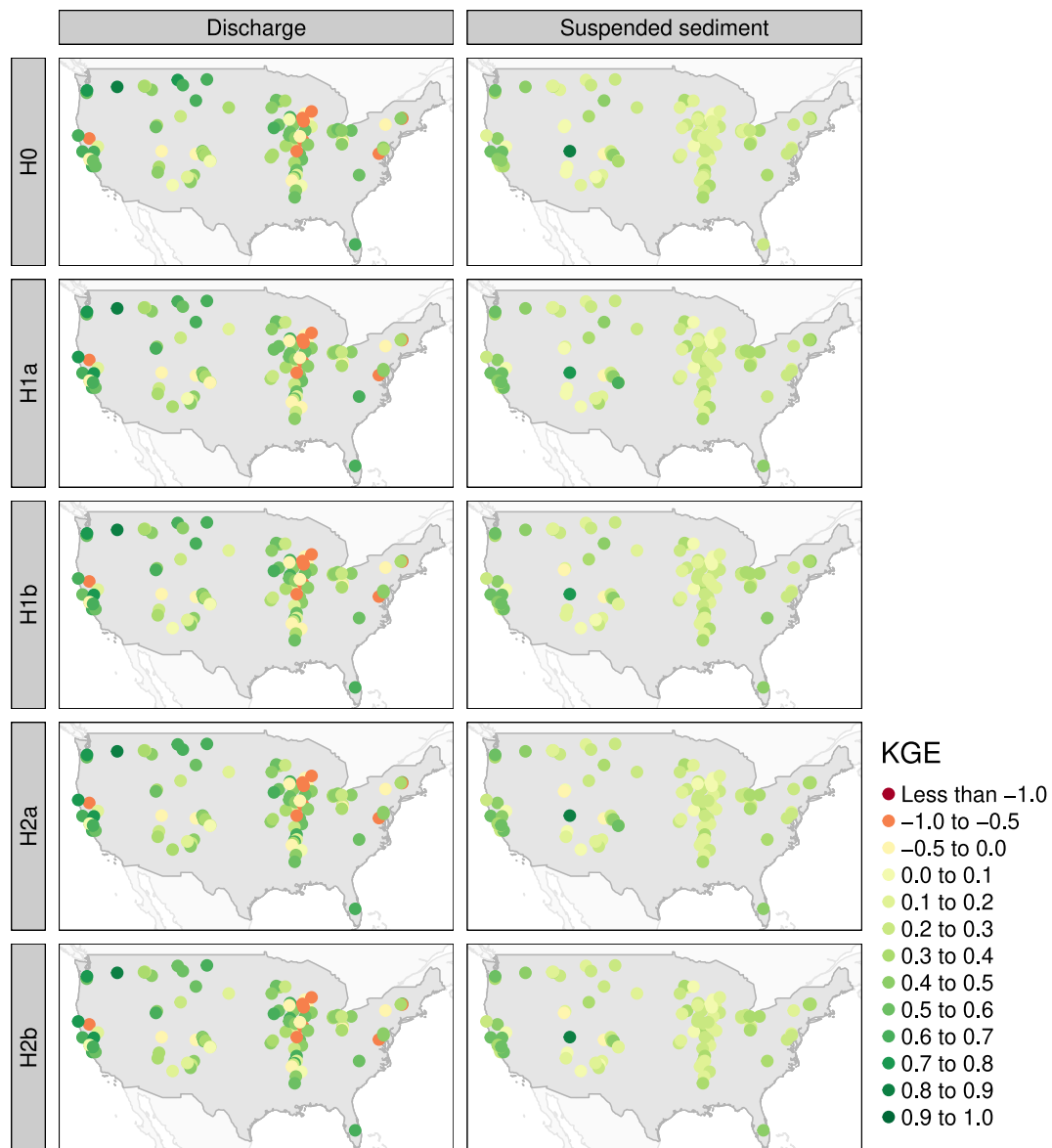


FIGURE 36 Performances of the two objectives functions used in the multi-objective calibration (KGE on total discharge Q and KGE on suspended sediment concentration SSC) for each modelling hypothesis