UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA CIVIL

UN ENFOQUE MULTIOBJETIVO PARA LIMITAR LAS INCERTIDUMBRES ESTRUCTURALES EN LAS EVALUACIONES DE IMPACTO CLIMATICO.

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN RECURSOS Y MEDIO AMBIENTE HÍDRICO

DANNY DANIEL SAAVEDRA ORE

PROFESORA GUÍA:

XIMENA VARGAS MESA

PROFESOR CO-GUÍA:

PABLO MENDOZA ZÚÑIGA

COMISIÓN:

HAROLD LLAUCA SOTO

SANTIAGO DE CHILE

2021

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE: Magíster en Ciencias de la Ingeniería, mención Recursos y Medio Ambiente Hídrico
POR: Danny Daniel Saavedra Ore
FECHA: Abril 2021
PROFESOR GUIA: Ximena Vargas Mesa

UN ENFOQUE MULTIOBJETIVO PARA LIMITAR LAS INCERTIDUMBRES ESTRUCTURALES EN LAS EVALUACIONES DE IMPACTO CLIMATICO.

La evaluación de los impactos del cambio climático sobre los recursos hídricos y el riesgo de inundación se ha abordado típicamente utilizando modelos hidrológicos calibrados y seleccionados en función de los registros de caudales observados. Sin embargo, los cambios en el clima rara vez se toman en cuenta, lo que hace que muchas estructuras de modelos no representen adecuadamente la hidrología de las cuencas en condiciones climáticas contrastadas. En este trabajo, se evalúa un esquema multiobjetivo simple para seleccionar un conjunto de estructuras de modelos hidrológicos calibrados en cuencas donde se han observado condiciones climáticas cambiantes. Para ello, se consideran 78 estructuras de modelos producidas con la plataforma de modelización modular Framework for Understanding Structural Error (FUSE, por sus siglas en inglés). El enfoque se basa en un esquema de Pareto y pretende encontrar estructuras de modelos que maximicen simultáneamente la eficiencia del modelo en períodos húmedos y secos. La metodología es aplicada en tres cuencas de tres diferentes vertientes hidrográficas de estudio en Perú (es decir, vertiente del Pacifico, Atlántico y Titicaca). En los resultados se encontró, que el enfoque permite encontrar estructuras con buen desempeño en términos de métricas de eficiencia, firmas hidrológicas y balances hídricos estacionales. Además, los resultados indican que algunas estructuras de modelo tienen bajo desempeño al pasar a un periodo climático muy diferente, y comportamientos erráticos en la simulación de estados y flujos internos. Este enfoque ayuda a reducir la incertidumbre estructural en la estimación de elasticidades con respecto a la precipitación, y de sensibilidades con respecto a cambios en la temperatura. En general, este trabajo demuestra el potencial de incorporar periodos con condiciones climáticas muy diferentes en un enfoque multiobjetivo para producir simulaciones robustas y creíbles, y para restringir la incertidumbre estructural en proyecciones hidrológicas.

A mis padres Daniel y Silvia, por su apoyo incondicional

A mi hermano Danilo, porque siempre ha estado ahí cuando lo necesitaba

A mi abuelo Cecilio, que siempre deseaste lo mejor para mi

AGRADECIMIENTOS

En primer lugar, agradezco al Programa Nacional de Becas y Crédito Educativo (PRONABEC) por el apoyo económico al otorgarme la Beca Presidente de la Republica convocatoria – 2018, para realizar mis estudios en el Magister de Ciencias de la Ingeniería, mención en Recursos y Medio Ambiente Hídrico en la Universidad de Chile. Esta Beca me han permitido fortalecer mis conocimientos y hacerme crecer como profesionalmente.

Al Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), que me han brindado la información de meteorológica e hidrológica necesaria para realizar este trabajo de Tesis, así mismo, agradezco a Harold por el apoyo y orientación durante el proceso de desarrollo de este trabajo de Tesis.

A mi profesora guía Ximena Vargas, por toda su confianza, apoyo y consejos enriquecedores brindados durante el tiempo que he estado cursando mis estudios de Magister, por incentivarme a investigar y a participar en las actividades de la Asamblea General de la EGU, agradecerle por sus enseñanzas y su disponibilidad de tiempo en todas mis dudas e inquietudes, siempre será una persona a quien respeto y aprecio.

Al profesor Pablo Mendoza, por todas sus enseñanzas y recomendaciones brindadas en las aulas universitarias, además, de su asesoramiento como profesor co-guía durante el proceso de elaboración del trabajo de Tesis y su publicación, por confiar en mis capacidades y exigirme a ser mejor, y por darme la oportunidad de conocer las ciencias hidrológicas.

Agradezco a Jacqueline y a todos mis profesores de la división de Recursos Hídricos y Medio Ambiente del Departamento de Ingeniería Civil de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, entre ellos a los profesores Yarko, Ximena, Pablo, Miguel, James, Marcelo, Yohann, Ana, Juvenal, Javier y Katherine, por su apoyo, comprensión y orientación en mi estadía en la Universidad.

A mis muy buenos amigos y compañeros Nery, Micol y Kathy, por su compañía, comprensión y apoyo durante estos dos años, gracias, mi estadía en Chile no hubiera sido igual sin ustedes. Así mismo, agradecer a todos mis compañeros de la División de Recursos Hídricos y Medio Ambiente.

A mis padres Silvia y Daniel, y a mi hermano Danilo por animarme a postular a PRONABEC y hacer el Magister en Chile. Agradezco a toda mi familia, por el apoyo constante desde Perú en el tiempo que estaba en Chile. Mil gracias a todos ustedes he podido culminar satisfactoriamente mis estudios.

A mi tía Lucia por siempre preocuparse por mi bienestar y apoyarme en todos mis nuevos proyectos.

Finalmente, agradecer a mis mejores amigos Juan, Yeral e Ingrid, por su sincera amistad, y que a pesar de la distancia hicieron que me sintiera acompañado. Gracias, me motivan a ser cada día mejor.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

1. INTRODUCCIÓN

El cambio climático está afectando en gran medida la economía y calidad de vida de las poblaciones en todo el planeta, ocasionando un aumento en la frecuencia e intensidad en la ocurrencia de eventos hidrológicos extremos como las sequías y las inundaciones observadas en estos últimos años (Alvarez-Garreton et al., 2021; Bhardwaj et al., 2020; Son et al., 2020; Haile et al., 2019; Shiru et al., 2019; Correa et al., 2017; Gavrilović et al., 2012), produciendo desafíos en la gestión de los recursos hídricos y riesgo de inundaciones a largo plazo. La comunidad hidrológica ha tomado conciencia de este problema (IPCC 2013), y en un esfuerzo colaborativo, se están desarrollando mejoras continuas en las metodologías utilizadas en la evaluación de los impactos del cambio climático.

La evaluación de los impactos del cambio climático en los recursos hídricos suele involucrar varias decisiones metodológicas, que incluyen la selección de escenarios de emisión, Modelos Climáticos Globales (GCM, por sus siglas en inglés), condiciones iniciales, método de escalamiento, estructura del modelo hidrológico y valores de los parámetros (e.g., Addor et al., 2014; Chegwidden et al., 2019; Chen et al., 2011; Wilby & Harris, 2006). Las decisiones anteriores tienen incertidumbres asociadas, cuya importancia relativa puede diferir según las condiciones hidroclimáticas específicas y las características de la cuenca (Clark et al., 2016). En particular, muchos autores han encontrado que la elección de la estructura del modelo hidrológico (es decir, la elección de los procesos explícitamente representados, las parametrizaciones del modelo, la arquitectura y la conectividad) y la elección de los parámetros (es decir, los coeficientes en las ecuaciones de los modelos, ya sean libres u observables) pueden tener grandes efectos en la caracterización de los impactos del cambio climático (e.g., Miller et al., 2012; Seiller et al., 2012; Vano et al., 2012; Brigode et al., 2013; Mendoza et al., 2015, 2016; Fowler et al., 2018a).

Los estudios de impacto de cambio climático sobre la hidrología de una cuenca, son conducidos mayormente usando modelos hidrológicos, los que primero son típicamente calibrados y evaluados en una línea base empleando datos históricos de caudal, y luego se corren con series temporales climáticas escaladas provenientes de Modelos de Circulación General. Con el fin de evaluar la estabilidad temporal del rendimiento de los modelos, la comunidad adoptó durante décadas las pruebas diferenciales de muestras divididas (DSST por sus siglas en inglés; Klemes, 1986) como práctica estándar. No obstante, muchos autores han demostrado que la habilidad de las simulaciones de modelos puede disminuir considerablemente cuando los valores de los parámetros que se encuentran en el proceso de calibración se aplican en condiciones climáticas muy diferentes (e.g., Brigode et al., 2013; Seiller et al., 2012; Merz et al., 2011; Vaze et al., 2010; Motavita et al., 2019; Pan et al., 2019; Stephens et al., 2019; Rau et al., 2018). Para abordar y comprender esta problemática, Stephens et al. (2019), desarrollaron tres experimentos: 1) transferibilidad temporal de los parámetros del modelo, 2) potencial para mejorar la simulación en condiciones climáticas futuras específicas, y 3) potencial para que los parámetros del modelo varíen según las condiciones climáticas en simulaciones mejoradas, que se llevaron a cabo utilizando el modelo conceptual GR4J (Perrin et al., 2003) en 164 cuencas australianas. Stephens et al. (2019) encontraron que, a través de experimentos numéricos, es posible mejorar el rendimiento de los modelos en condiciones climáticas contrastantes, aunque los beneficios fueron variables en los tres experimentos y en todas las cuencas analizadas. Coron et al. (2012) examinaron la capacidad de extrapolación de tres modelos hidrológicos para diferentes condiciones climáticas en 216 cuencas de Australia, proponiendo una prueba generalizada de muestreo dividido (GSST, por sus siglas en inglés) basada en la metodología DSST; sus resultados demostraron que la transferencia de los parámetros del

modelo en el tiempo puede introducir errores en las simulaciones y, por lo tanto, falta de robustez cuando los modelos se utilizan en un clima cambiante.

Fowler et al. (2016) re-visitaron el problema de la inestabilidad de los parámetros mediante la aplicación de un marco de Pareto, con el fin de encontrar conjuntos de parámetros que maximizaran simultáneamente la eficiencia de los modelos en un período húmedo y otro seco. Para ello, configuraron cinco modelos hidrológicos conceptuales en 85 cuencas situadas en Australia, comprobando que la inestabilidad temporal del rendimiento de los modelos conceptuales, reportada en estudios anteriores, podía atribuirse a deficientes estrategias de estimación de parámetros. Fowler et al. (2018a) evaluaron el rendimiento de los modelos hidrológicos en sequías históricas de varios años, llegando a la conclusión que, a menudo, tienen un rendimiento deficiente que podría atribuirse a errores en los datos, la estructura del modelo o los parámetros. Recientemente, Duethmann et al. (2020) analizaron las causas del bajo rendimiento de un modelo hidrológico semidistribuido en condiciones climáticas cambiantes en un gran número de cuencas de Australia, donde se centraron, principalmente, en: (1) problemas de datos, (2) problemas relacionados con la calibración del modelo y (3) deficiencias en la estructura del modelo. Duethmann et al. (2020) descubrieron que el deficiente rendimiento de los modelos se debe principalmente a que la mayoría de las estructuras de los modelos ignoran los cambios en la dinámica de la vegetación y la falta de homogeneidad en las entradas de las precipitaciones.

Los únicos efectos de la elección del modelo hidrológico - comúnmente basado en el legado, más que en la idoneidad (Addor & Melsen, 2019) – se han explorado ampliamente en el contexto de los impactos del cambio climático. Una gran parte del trabajo se ha basado en la selección de un pequeño conjunto de modelos hidrológicos (e.g., Bae et al., 2011; Jiang et al., 2007; Mendoza et al., 2015; Mizukami et al., 2016), mientras que unos pocos autores han propuesto cambios explícitos en las estructuras de los modelos para abordar este problema (Grigg & Hughes, 2018; Westra et al., 2014). La evaluación de la incertidumbre estructural se ve ahora facilitada por la aparición de plataformas modulares de modelización (FMMs, por sus siglas en inglés; e.g., Clark et al., 2008, 2015; Coxon et al., 2019; Craig et al., 2020; Kavetski & Fenicia, 2011; Knoben et al., 2019; Niu et al., 2011; Pomeroy et al., 2007), que permiten diseñar experimentos controlados para probar las hipótesis sobre el funcionamiento de la cuenca (Clark et al., 2011). Por supuesto, el gran número de opciones de modelización disponibles en los FMM plantea el reto de muestrear el espacio de modelos para evitar la sobreestimación de la incertidumbre estructural (Remmers et al., 2020), especialmente en escenarios de condiciones climáticas cambiantes.

Las sensibilidades hidrológicas a perturbaciones climáticas son atractivas debido a su formulación sencilla y porque permiten estimaciones rápidas de la producción de escorrentía en diferentes escenarios climáticos (Lehner et al., 2019; Milly et al., 2018; Vano et al., 2015; Vano & Lettenmaier, 2014). Vano et al. (2012) examinaron los efectos de la elección del modelo hidrológico en la estimación de las elasticidades de la precipitación y las sensibilidades de la temperatura en la cuenca del río Colorado, encontrando grandes diferencias que podrían ser atribuidas a la falta de calibración de los parámetros. Mendoza et al. (2015) compararon las diferencias entre los modelos en los cambios hidrológicos proyectados antes y después de realizar la estimación de los parámetros, concluyendo que las diferencias entre estimaciones provenientes de diferentes modelos persisten después de efectuada la calibración. La elección subjetiva de los modelos en dichos estudios, una mayor evaluación del modelo en períodos climáticos diferentes y el conjunto de trabajos a que se ha hecho referencia anteriormente, refuerzan la urgencia de: 1) mejorar la estimación de los parámetros y las estrategias de selección de los modelos, y 2)

seleccionar estructuras de modelo que permitan obtener resultados coherentes en condiciones climáticas cambiantes. En este trabajo de tesis, se combinan elementos de estudios recientes para re-visitar las conclusiones proporcionadas por Vano et al. (2012), abordándose las siguientes preguntas:

- ¿Es posible escoger estructuras de modelo mediante un esquema de Pareto simple para obtener simulaciones hidrológicamente consistentes en períodos climáticos contrastantes?
- ¿Existen características comunes entre las estructuras de modelos seleccionadas?
- ¿Es el esquema en cuestión capaz de reducir las incertidumbres en cuanto a la elasticidad de las precipitaciones y la sensibilidad a la temperatura?

El enfoque propuesto se basa en: 1) la selección de subperíodos secos y húmedos, 2) la calibración de los modelos hidrológicos en cada subperíodo, 3) la elección de las combinaciones de modelo hidrológico y conjunto de parámetros que maximicen el rendimiento en años húmedos y secos, 4) la evaluación de la consistencia hidrológica y 5) la cuantificación de la dispersión del conjunto en sensibilidades hidrológicas resultantes del subconjunto de modelos. El esquema se aplica en tres cuencas situadas en el Perú, y las esctructuras de modelo son generadas utilizando la plataforma Framework for Understanding Structural Errors (FUSE; Clark et al., 2008).

1.1. Objetivos

Los objetivos del presente trabajo de tesis son los siguientes:

1.1.1. Objetivos Generales

Proponer un enfoque multiobjetivo para reducir la incertidumbre proveniente de la elección de la estructura de modelos hidrológicos en la evaluación de los impactos del cambio climático en la hidrología.

1.1.2. Objetivos Específicos

- Obtener simulaciones hidrológicamente consistentes en periodos climáticos de características opuestas, mediante la elección de estructuras de modelo mediante un esquema simple de Pareto.

- Encontrar si existen características comunes entre las estructuras de modelos seleccionados.

- Comprender y disminuir la incertidumbre en cuanto a la elasticidad del caudal con respecto a la precipitación y la sensibilidad del caudal con respecto a cambios en la temperatura.

2. ARTÍCULO PARA PUBLICACIÓN

A continuación, se presenta el artículo titulado "*A multi-objective approach to constrain structural uncertainties for climate impact assessments*", que se encuentra en estado de revisión para ser publicado en la revista *Hydrological Processes* en una edición especial titulada: *"Towards more credible models in catchment hydrology to enhance hydrological process understanding":*

A multi-objective approach to constrain structural uncertainties for climate impact assessments

Danny Saavedra[1], Pablo A. Mendoza[1,2], Nans Addor[3], Harold Llauca[4], and Ximena Vargas[1]

[1]Department of Civil Engineering, Faculty of Physical and Mathematical Sciences, Universidad de Chile, Chile.
[2]Advanced Mining Technology Center (AMTC), Faculty of Physical and Mathematical Sciences, Universidad de Chile, Chile.
[3]Geography, College of Life and Environmental Sciences, University of Exeter, UK.
[4]Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), Lima, Perú.

Abstract

The assessment of climate change impacts on water resources and flood risk is typically approached using hydrological models that are calibrated and selected based on observed streamflow records. Nevertheless, changes in climate are rarely accounted for, causing the failure of many model structures to adequately represent catchment hydrology under contrasting climatic conditions. In this paper, we test a simple framework for selecting a suite of calibrated hydrological model structures in catchments where changing climatic conditions have been observed. We consider 78 model structures produced using the FUSE modular modelling framework. The approach is based on a Pareto scheme and aims to find model structures that simultaneously maximize model efficiency in wet and dry periods. After testing this approach in three case study basins in Peru, we found that it enables the identification of structures with good performance in terms of efficiency metrics, hydrological signatures and seasonal water balances, and reveals that some model structures have low performance in contrasting climates and poor or abnormal behaviors in internal states and fluxes. Further, this approach helps to reduce the spread in precipitation elasticities and temperature sensitivities estimated from alternative model structures. Overall, this work demonstrates the potential of using contrasting climatic conditions in a multi-objective framework to produce robust and credible simulations, and to constrain structural uncertainties in hydrological projections.

Keywords: Climate change; uncertainty; hydrologic model structures; Pareto scheme

1. INTRODUCTION

Climate change is greatly affecting the economy and quality of life of populations around the world, causing an increase in the frequency and intensity of extreme hydrological events such as droughts and floods observed in recent years (Alvarez-Garreton et al., 2021; Bhardwaj et al., 2020; Son et al., 2020; Haile et al., 2019; Shiru et al., 2019; Correa et al., 2017; Gavrilović et al., 2012), producing challenges in the water resources management and flood risk in the long term. The hydrological community has become aware of this used in the assessment of climate change

impacts. problem (IPCC 2013), and in a collaborative effort, continuous improvements are being developed in the methodologies used in the assessment of climate change impacts.

The assessment of climate change impacts on water resources commonly involves several methodological choices, which include the selection of emission scenarios, global climate models (GCM), initial conditions, downscaling method, hydrologic model structure and parameter values (e.g., Addor et al., 2014; Chegwidden et al., 2019; Chen et al., 2011; Wilby & Harris, 2006). The above decisions contribute with uncertainties, whose relative importance may differ depending on specific hydroclimatic conditions and basin characteristics (Clark et al., 2016). In particular, many authors have found that the choice of the hydrologic model structure (i.e., choice of processes explicitly represented, model parameterizations, architecture and connectivity) and the choice of parameters (i.e., the coefficients in model equations, either free or observable) may have large effects on the characterization of climate change impacts (e.g., Miller et al., 2012; Seiller et al., 2012; Vano et al., 2012; Brigode et al., 2013; Mendoza et al., 2015, 2016; Fowler et al., 2018a; Melsen et al., 2018).

The sole effects of hydrologic model choice – commonly based on legacy, rather than adequacy (Addor & Melsen, 2019) – have been widely explored in the context of climate change impacts. A large body of work has relied on the selection of a small ensemble of hydrological models (e.g., Bae et al., 2011; Jiang et al., 2007; Mendoza et al., 2015; Mizukami et al., 2016), while a few authors have proposed explicit changes in model structures to address this issue (Grigg & Hughes, 2018; Westra et al., 2014). The assessment of structural uncertainty is now facilitated by the emergence of modular modeling frameworks (MMFs; e.g., Clark et al., 2008, 2015; Coxon et al., 2019; Craig et al., 2020; Kavetski & Fenicia, 2011; Knoben et al., 2019; Niu et al., 2011; Pomeroy et al., 2007), which allow to design controlled experiments to test hypotheses about catchment functioning (Clark et al., 2011). Of course, the large number of modeling options available in MMFs raises the challenge to sample model space in order to avoid the overestimation of structural uncertainty (Remmers et al., 2020), especially under scenarios of changing climatic conditions.

During decades, the hydrology community adopted differential split-sample tests (DSST; Klemes 1986) as a standard practice to assess the temporal stability of model performance. However, many authors have reported decreased skill when a model is applied in very different climatic conditions compared to those used to infer the parameter values (e.g., Brigode et al., 2013; Seiller et al., 2012; Merz et al., 2011; Vaze et al., 2010; Motavita et al., 2019; Pan et al., 2019). For example, Coron et al. (2012) examined the extrapolation capacity of three hydrological models for different climatic conditions in 216 basins in Australia, proposing a generalized split-sample test (GSST) based on the DSST methodology. Their results demonstrated that the transfer of the model parameters in time can introduce errors in simulations, and therefore lack of robustness when the models are used in a changing climate. Stephens et al. (2019) developed three experiments to test the potential of (1) transferring model parameters in time, (2) improve the simulation in specific future climate conditions, and (3) varying model parameters according to climate conditions in improved simulations. Stephens et al. (2019) used the GR4J (Perrin et al., 2003) conceptual model to conduct three experiments over 164 Australian basins, obtaining mixed results in their experiments to improve model performance under contrasting climatic conditions. More recently, Duethmann et al. (2020) analyzed the causes of the low performance of a semi-distributed hydrological model under changing climate conditions over a large number of basins in Australia, where they mainly focused on: (1) data problems, (2) problems related to the model calibration and (3) model structure deficiencies. Duethmann et al. (2020) found that poor model performance is mainly because most

model structures ignore changes in vegetation dynamics and non-homogeneity in precipitation input.

Fowler et al. (2016) revisited the problem of temporal instability in model performance through the application of a Pareto framework – aimed to find parameter sets that simultaneously maximize model efficiency in a wet and a dry period – in 85 catchments located in Australia, using five conceptual hydrological models. They found that temporal instabilities reported in previous studies may be attributed to poor parameter estimation strategies, rather than model structural inadequacies. Fowler et al. (2018a) assessed hydrological model performance in historical multi-year droughts, finding that they often have poor performance that could be attributed to errors in the data, model structure or parameters. It should be noted that their Pareto scheme was used to discriminate parameter sets given a fixed model structure, rather than screening competing models with the same application purpose.

Hydrologic sensitivities to climate perturbations are attractive due to their simple formulation, and because they allow quick estimates of runoff production under different climate scenarios (Lehner et al., 2019; Milly et al., 2018; Vano et al., 2015; Vano & Lettenmaier, 2014). Vano et al. (2012) examined the effects of hydrologic model choice on the estimation of precipitation elasticities and temperature sensitivities in the Colorado River basin, finding large inter-model differences that could have been attributed to the lack of parameter calibration. Mendoza et al. (2015) compared inter-model differences in projected hydrologic changes before and after conducting parameter estimation, concluding that large differences remain between calibrated models. Nevertheless, the subjective choice of models in these studies, the lack of further model evaluation in contrasting climatic periods, and the body of work previously referred to reinforces the urgency to (1) improve parameter estimation and model selection strategies, and (2) conduct plausibility checks in model structures to obtain coherent results under changing climatic conditions. In this paper, we combine elements from recent studies to revisit the conclusions provided by Vano et al. (2012), for which we address the following questions:

- Is it possible to sample model space using a simple Pareto framework to obtain hydrologically consistent simulations in contrasting climatic periods?
- Are there common features among the selected model structures?
- Is such framework capable to reduce uncertainties in precipitation elasticities and temperature sensitivities?

The proposed approach is based on (1) selection of dry and wet sub-periods, (2) calibration of hydrological models in each sub-period, (3) choice of the combinations of hydrological model and parameter set that maximize performance in wet and dry years, (4) assessment of hydrological consistency and model screening, and (5) quantify the ensemble spread in hydrologic sensitivities resulting from the model subset. We apply this framework in three basins located in Peru, and model samples are produced using the Framework for Understanding Structural Errors (FUSE; Clark et al., 2008).

2. STUDY DOMAIN AND DATA

The Autoridad Nacional del Agua (ANA) of Peru has identified three main hydrographic slopes or regions (ANA 2009), called the Pacific slope, Atlantic or Amazon slope and Titicaca slope, characterized by their geomorphological differences, and climatic and hydrological behavior (see,

Llauca et al., 2021b; Heidinger et al., 2018; Rau et al., 2018; Lavado Casimiro et al., 2012). The Pacific slope is characterized by high water stress (i.e., arid conditions in the coastal zones and semi-arid conditions on the western flank of the Andes) due to its high water demand (densely populated) and low water availability. In contrast, the Atlantic and Titicaca slopes are characterized by high water availability (i.e., wet conditions) and low water demand (sparsely populated) (Llauca et al., 2021b; Lavado Casimiro et al., 2012).

In order to represent by means of a basin each of the hydrographic slopes, we select three case study basins (Figure 1): two of these are located in Peru (Vilcanota and Huancané), and the other is a transboundary catchment located between Peru and Ecuador (Puyango-Tumbes). The Vilcanota River basin belongs to the Atlantic slope in the Peruvian Andes, the Huancané River basin belongs to the Titicaca slope in the Peruvian Altiplano, and the Puyango-Tumbes River basin belongs to the Pacific slope. During recent years, these basins have seen an increase in flooding during the rainy season, and water shortages during the dry season (Sanabria et al. 2009; Lavado Casimiro et al. 2011; Rivas & Rivas, 2013; Andres et al. 2014; Zulkafli, 2014; Takahashi & Martínez Grimaldo, 2015; Martínez & Céspedes, 2017).

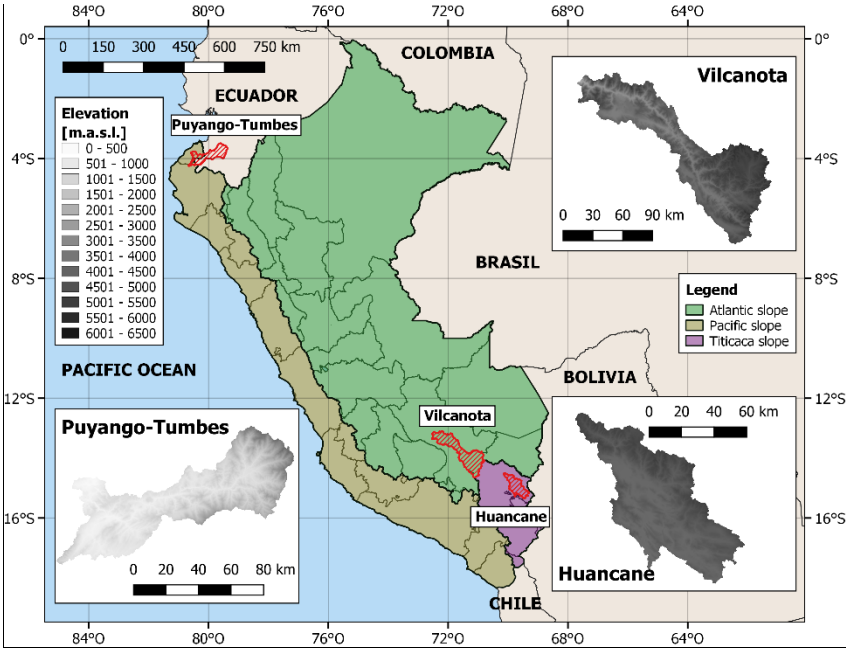

Figure 1: Location and elevation of the three case study basins

Table 1 summarizes the main physiographic and hydroclimatic characteristics of the three basins for a 30-year period of Sep/1986 – Aug/2016 (note that the hydrological year in Peru starts in the month of September (spring) and ends in the month of August), excepting the annual runoff coefficient (RC) and the mean annual runoff in the Puyango-Tumbes and Huancané basins due to the lack of streamflow records in some years. The Vilcanota and Huancané River basins (located in the southeast) have similar mean elevations (see Figure 1), while the Puyango-Tumbes River basin (located in the northwest, very close to equatorial line) has the lowest mean elevation (Figure 1), and higher values of mean annual precipitation, aridity index and runoff coefficient (0.999), which implies that a large part of the precipitation ends up draining, while the Huancané River basin has the lowest runoff coefficient (0.250), which implies that a small amount of the precipitation ends up draining. The Puyango-Tumbes River basin is an arid basin, since it presents a high value of aridity index in comparison with the other two basins, this value is due to the fact

that the basin has a great potential to evaporate, which means that evapotranspiration (ET) is an important variable in hydrological modeling.

Table 1: Characteristics of the three case study watersheds.

Basin	Area [km^2]	Mean basin elevation [m.a.s.l.]	Mean annual runoff [mm yr^{-1}]	Mean annual precipitation [mm yr^{-1}]	Mean annual PET [mm yr^{-1}]	Mean annual RC [R/P]	Mean annual AI [PET/P]
Vilcanota	9586	4279	398.29	742.10	813.02	0.537	1.096
Puyango-Tumbes	4694	1941	843.06*	851.07	1593.35	0.999*	1.872
Huancané	3545	4396	170.57**	674.44	694.01	0.250**	1.029

Note: PET is the mean annual evapotranspiration; RC is the annual runoff coefficient; AI is the aridity index.
*Period considered 24 years.
**Period considered 21 years.

Many authors have argued that the lack of meteorological observations in the Peruvian Andes and the Amazonia is the main limitation for climate change impact studies (Aybar et al., 2019; Rau et al., 2018; Lavado Casimiro et al., 2011; Andres et al., 2014; Zulkafli, 2014; Vegas Galdos et al., 2015; Llauca et al., 2021a, 2021b). To address this issue, Aybar et al. (2019) and Huerta et al. (2018) developed the Peruvian Interpolated data of SENAMHI's Climatological and hydrological Observations (PISCO) database, which provides daily time series of precipitation, minimum and maximum temperature for the period 1981-2016, with a 0.1° horizontal resolution. In particular, precipitation time series in PISCOp (Aybar et al., 2019) were obtained using geostatistical and deterministic interpolation methods that include three precipitation sources: (1) the quality-controlled and infilled national rain gauge dataset, (2) radar-gauge merged precipitation climatologies, and (3) the Climate Hazards Group Infrared Precipitation (CHIRP) estimates. Daily time series of maximum and minimum temperature in PISCOt (Huerta et al., 2018) were obtained from: (1) observed maximum and minimum temperature data, (2) soil temperature product from the MODIS sensor (Moderate Resolution Imaging Spectroradiometer), and (3) spatial statistical predictors (e.g., elevation, longitude, latitude and Topographic Dissection Index). The reader is referred to Huerta et al. (2018) for full descriptions of the development of temperature products. The PISCO product is free available in the IRI Data Library website: http://iridl.ldeo.columbia.edu/SOURCES/.SENAMHI/.HSR/.PISCO/.

Daily streamflow time series were obtained from the Peruvian National Meteorological and Hydrological Service (Servicio Nacional de Meteorología e Hidrología, SENAMHI). For the Vilcanota River basin, the Intihuatana Km-105 gauge-station is considered with data availability from 1985 – 2016; for the Puyango-Tumbes River basin, we collect data from the El Tigre gauge-station for the period 1981 - 2016, and the Puente Huancané gauge-station provides streamflow records for the Huancané basin during the period 1988 - 2016.

3. APPROACH

The proposed method to select model structures under changing climatic conditions is outlined in Figure 2, and includes the following steps: (1) selection of wet and dry periods, (2) calibration of all model structures in each basin, (3) sampling the model space based on temporally consistent performance skill, (4) assessment of hydrological consistency and model screening, and (5) quantifying the effects of model sampling on the spread in precipitation elasticities and temperature sensitivities. The step (3) builds upon the approach proposed by Fowler et al. (2018a), which aims

to maximize model performance in both wet and dry periods using a Pareto analysis scheme. Hence, the framework involves the selection of two 5-year periods with contrasting climates (section 3.1.1), and a one-year period for model spin-up. Once these subperiods are defined, we work sequentially towards a small multi-model ensemble that provides hydrologically consistent simulations through the following steps:

i. MM0-dry and MM0-wet: full FUSE model ensemble (i.e., 78 model structures) calibrated in dry and wet periods by minimizing RMSE (section 3.1.2).
ii. MMP-dry and MMP-wet: a small 5-member ensemble obtained after applying the Pareto scheme framework with the evaluation metrics listed in Table 4 (section 3.1.3).
iii. MMPS-dry and MMPS-wet: Same as 2, but after screening MMP-dry and MMP-wet based on the seasonal behavior of internal states and fluxes.

The details of each step are described in the following sub-sections.

3.1.1. Selection of analysis periods

The "differential split-sample test" (DSST) procedure formulated by Klemes (1986) was applied to select the analysis periods, using runoff and mean annual precipitation graphs (see Figure 2, box 1). In order to verify the contrasting periods (i.e. wet and dry period), the main hydroclimatic characteristics (e.g. mean annual runoff, mean annual precipitation, mean annual PET, runoff coefficient and aridity index), monthly average graphs, and the duration curve are used. The selected periods are displayed in Figure 3. For the Vilcanota River basin (Figure 3, left panels), we consider Sep/1988-Aug/1993 as dry period, and Sep/1999 - Aug/2004 as wet period; for the Puyango-Tumbes River basin (Figure 3, center panels), the selected periods are Sep/2002-Aug/2007 as dry period, and Sep/2007-Aug/2012 as wet period; and for the Huancané River basin (Figure 3, right panels), we consider Sep/2006-Aug/2011 as dry period, and Sep/1999-Aug/2004 as wet period. The main hydroclimatic characteristics of each period are summarized in Table 2, and the contrasting hydroclimates between the selected periods are reflected in mean monthly runoff, mean monthly precipitation and average monthly temperature (Figure 3, top panels), and the flow duration curves (Figure 3, bottom panels) for the three basins. For example, in the Vilcanota River basin the wet period has aridity index (IA) and runoff coefficient (RC) values of 1.00 and 0.58 respectively, compared to the dry period of 1.18 and 0.49 respectively. Interestingly, we observe that, in the Puyango-Tumbes River basin, the aridity index (see Table 2) has high values compared to the other two basins, this is due to the fact that along the Pacific coast, rainfall is higher in the north and decreases towards the south (Lavado Casimiro et al., 2012), and because the basin is close to the equatorial line, temperature values are higher, causing a high potential for evapotranspiration. Moreover, in the Puyango-Tumbes River basin, there is more runoff than precipitation between May and September (see Figure 3, upper panel), which suggests that the basin also receives groundwater contributions. According to the geoenvironmental study of the Puyango-Tumbes River basin conducted by Núñez Juárez & Zegarra Loo (2006), they identified aquifers located in the alluvial deposits of the Tumbes River and in the areas of the main streams, which are constantly recharged by the seasonal rains that occur in the upper part of the basin.

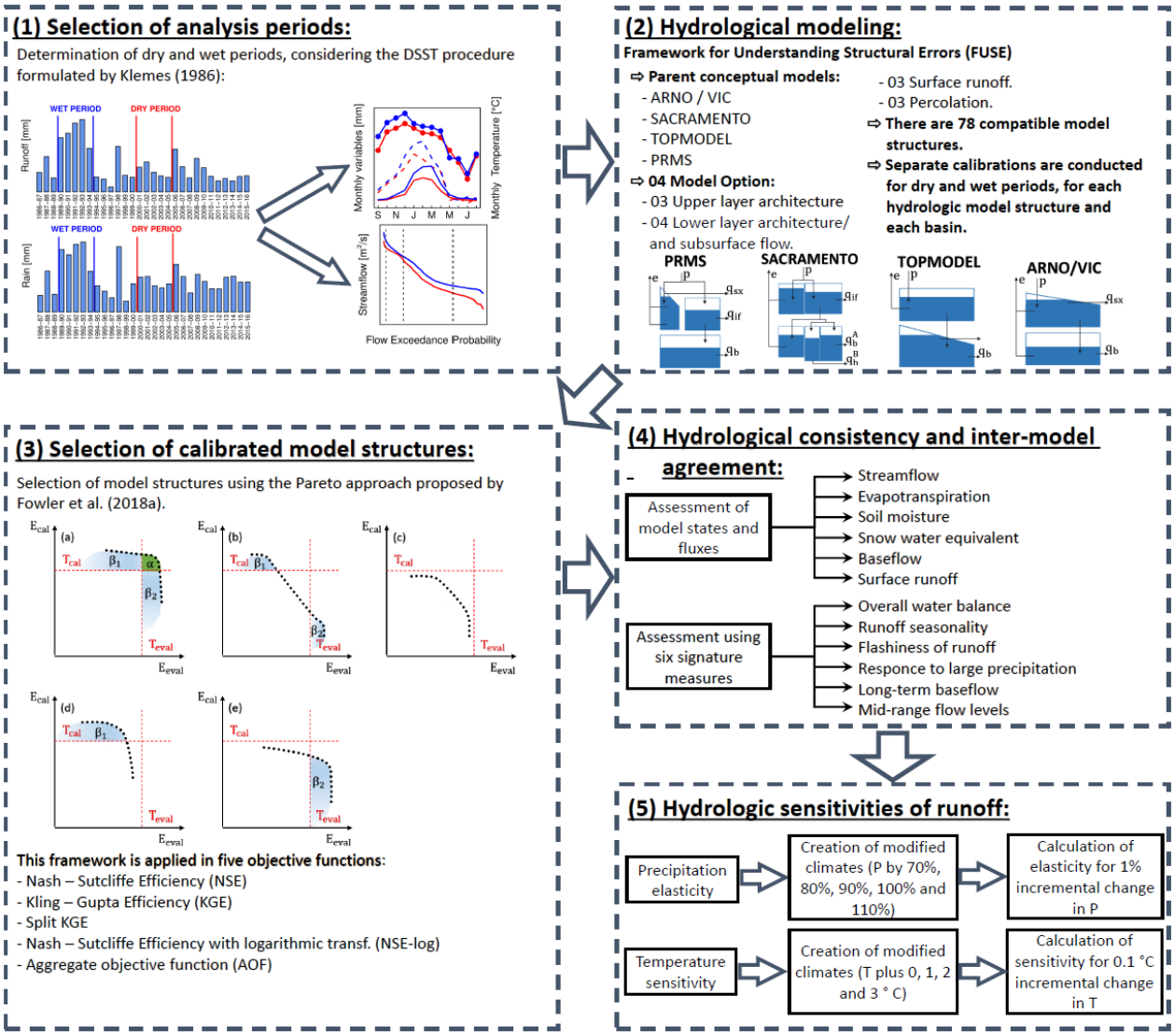

Figure 2: Methodological scheme

Table 2: Characteristics of the dry and wet periods for the three watersheds.

Basin	Period	Mean annual runoff [mm yr^{-1}]	Mean annual precipitation [mm yr^{-1}]	Mean annual PET [mm yr^{-1}]	Mean annual RC [R/P]	Mean annual AI [PET/P]
Vilcanota	**Dry**	327.77	665.88	787.93	0.49	1.18
	Wet	468.29	806.25	806.91	0.58	1.00
Puyango-Tumbes	**Dry**	486.81	527.62	1593.04	0.92	3.02
	Wet	914.72	991.63	1592.35	0.92	1.61
Huancané	**Dry**	114.43	605.21	705.62	0.19	1.17
	Wet	226.69	749.54	689.63	0.30	0.92

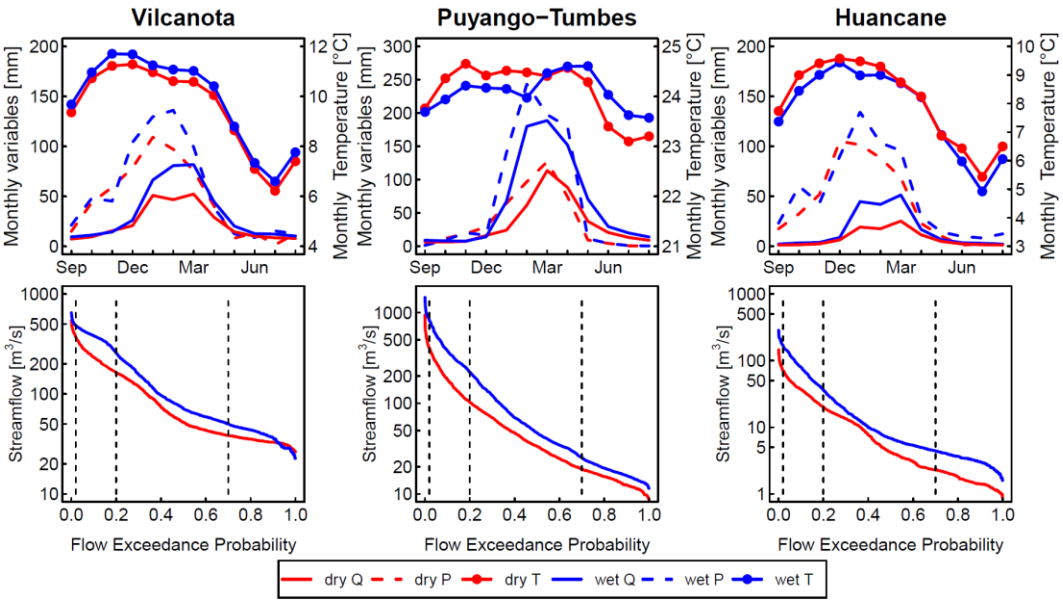

Figure 3: Average monthly runoff, average monthly precipitation and average monthly temperature (top panels), and flow duration curve (bottom panels), where the red line denotes the dry period and the blue line denotes the wet period in each of the three basins.

3.1.2. Hydrological modeling

In this study, we use the Framework for Understanding Structural Error (FUSE; Clark et al., 2008), which allows the implementation of an ensemble of hydrologic model structures that can be used to characterize structural uncertainty (e.g., Clark et al., 2011). Further, its modular functionality allows to diagnose inter-model differences in simulated states and fluxes through controlled experiments. FUSE discretizes the soil column along the vertical axis into an unsaturated zone (above the water table) and a saturated zone (below the water table). The major model-building decisions can be defined by the user, including the architecture of the upper and lower soil layers, and the parameterizations for simulating evaporation, surface runoff, percolation of water fluxes between soil layers, interflow, and baseflow. The multiple options available for each model building decision (Table 3) are drawn from four conceptual parent models: PRMS (Leavesley et al., 1984), Sacramento (Burnash, 1995), TOMODEL (Beven & Kirkby, 1979), and ARNO/VIC (Zhao, 1977). The formulation of the equations of state and fluxes parameterizations used for the construction of the model structures are shown in Appendix B.

It is important to note that FUSE does not perform any surface energy balance calculations, and neither does represent the canopy interception or the transpiration and evaporation from intercepted water (Clark et al., 2008). Snow is modeled using a Snow-17-based temperature index model (Anderson, 2006), which tracks snow water equivalent (SWE) based on precipitation and melt (see Henn et al., 2015 for further details). Despite all these simplifications, the models are designed to provide a relatively complete representation of the major hydrological fluxes in the subsurface (Clark et al., 2008).

Table 3: FUSE model decision options (modified from Clark et al. 2008; Staudinger et al. 2011).

Model structure	Model option	Existing model
Upper layer architecture "U"	Upper layer divided into tension and free storage	Sacramento
	Free storage plus tension storage sub-divided into recharge and excess	PRMS
	Upper layer defined by a single state variable	ARNO/VIC - TOPMODEL
Lower layer architecture and subsurface flow "L"	Tension storage combined with two parallel tanks	Sacramento
	Storage of unlimited size combined with linear fraction rate	PRMS
	Storage of unlimited size combined with power recession	TOPMODEL
	Storage of fixed size with non-linear storage function	ARNO/VIC
Surface runoff "SR"	ARNO/Xzang/VIC parametrization	ARNO/VIC
	PRMS variant; fraction of upper tension storage	PRMS
	TOPMODEL parametrization	TOPMODEL
Percolation "PE"	Water from field capacity to saturation available for percolation	PRMS
	Water from wilting point to saturation available for percolation	ARNO/VIC
	Percolation defined by moisture content in lower layer architecture	Sacramento

In this work, we configure 78 model structures for each basin using spatially lumped representations. Model simulations are conducted at a daily time step, requiring precipitation, temperature and potential evapotranspiration (PET) –computed with the formulation proposed by Oudin et al. (2005) – as meteorological forcings. All model structures are calibrated with the Shuffled Complex Evolution (SCE-UA; Duan et al. 1992) optimization algorithm to minimize the root mean squared error (RMSE) of simulated daily streamflow:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{QS_i} - \mathrm{QO_i})^2} \qquad (1)$$

Where, $\mathrm{QS_i}$ corresponds to the simulated runoff, $\mathrm{QO_i}$ corresponds to the observed runoff, and N is the number of days in the calibration period. We set a maximum of 10,000 iterations to ensure that the optimization process is complete. It should be noted that the choice of objective function is justified by the particular interest to evaluate the capability to simulate floods (i.e., peak flows) under contrasting climate scenarios. Therefore, separate calibrations are conducted for dry and wet periods, for each hydrologic model structure and each basin. The resulting combinations of model structures and parameters – obtained for dry (MM0-dry) and wet (MM0-wet) periods – serve as the basis of the model sampling framework described below.

3.1.3. Selection of calibrated model structures

In order to sample the model space, we adopt the Pareto approach proposed by Fowler et al. (2018a) to obtain temporally stable parameter sets for a given model structure, using performance acceptance thresholds in wet and dry hydroclimatic periods. Nevertheless, our study differs in that we focus on the temporal transferability of model structures under a common calibration framework, rather than model parameter sets for a fixed model structure. We apply the Pareto framework using five commonly used objective functions (Table 4) and consider 0.7 as the performance acceptance threshold. Hence, we select a small ensemble of six models from the original 78-member ensemble, aiming to satisfy the following conditions:

1. Model structure and parameter set with the best performance in terms of RMSE in the calibration period (smallest RMSE Cal).
2. Model structure and parameter set with the highest NSE (or, equivalently, smallest RMSE) in both calibration and evaluation periods (higher NSE Cal-Eval).
3. Model structure and parameter set with the highest KGE in both calibration and evaluation period (higher KGE Cal-Eval).
4. Model structure and parameter set with the highest split-KGE in both calibration and evaluation periods (higher splitKGE Cal-Eval).
5. Model structure and parameter set with the highest NSE-log in both calibration and evaluation periods (higher NSElog Cal-Eval).
6. Model structure and parameter set with the highest Aggregate Objective Function (AOF; Fowler et al. 2016) in both calibration and evaluation periods (higher AOF Cal-Eval).

It should be noted that alternative 1 has been the common practice in hydrology for decades. Alternatives 2-6 do not involve any re-calibration of model structures, but a model sampling based on evaluations with parameter values obtained from the calibration process (section 3.1.2). For example, the dry period calibration KGE (alternative 3) is the Kling-Gupta efficiency obtained with the parameters that result from model calibration conducted in the dry period, minimizing RMSE; similarly, the wet period evaluation KGE is the Kling-Gupta efficiency obtained with the same parameters, applied in the wet period. The result of this process are small (5-member) multi-model ensembles for a dry (MMP-dry) and a wet (MMP-wet) period, which provide combinations of flood-oriented models/parameters for (1) hydrologically consistent simulations, regardless of the climatic conditions, and (2) reducing the structural uncertainty in hydrologic sensitivities to climate change.

Table 4: Performance metrics used to sample the model space produced with FUSE.

Performance metric	Equation	Range	Emphasis	References
Nash – Sutcliffe Efficiency (NSE)	$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(QS_i - QO_i)^2}{\sum_{i=1}^{N}(QO_i - \overline{QO_i})^2}$	-∞ – 1	High flows and dynamic discharge.	Nash & Sutcliffe (1970)
Kling – Gupta Efficiency (KGE)	$\mathrm{KGE} = 1 - \mathrm{ED};$ $\mathrm{ED} = \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2};$ $\alpha = \frac{\sigma_S}{\sigma_O};\ \beta = \frac{\mu_S}{\mu_O};$ $\mathrm{rPearson} = \frac{\sum_{i=1}^{N}(QO_i - \overline{QO_i})(QS_i - \overline{QS_i})}{\sqrt{\sum_{i=1}^{N}(QO_i - \overline{QO_i})^2}\sqrt{\sum_{i=1}^{N}(QS_i - \overline{QS_i})^2}}$	0 – 1	High flows, Water balance, streamflow variability and correlation	Gupta et al. (2009)
Split KGE	$\mathrm{Split\ KGE} = \frac{\sum_{t=1}^{T}\mathrm{KGE}_t}{T}$	0 – 1	Base flows.	Gupta et al. (2009) Fowler et al. (2018b)
Nash – Sutcliffe Efficiency with logarithmic transf. (NSE-log)	$\mathrm{NSElog} = 1 - \frac{\sum_{i=1}^{N}(\log(QS_i) - \log(QO_i))^2}{\sum_{i=1}^{N}(\log(QO_i) - \log(\overline{QO_i}))^2}$	-∞ – 1	Low flows.	Nash & Sutcliffe (1970) Santos et al. (2018)

Aggregate objective function (AOF)	$AOF = \frac{AOF_{sig} + AOF_{gof}}{2}$ $AOF_{sig} = 1 - \sum_{q=1}^{8} \frac{Y_{q,o} - Y_{q,s}}{8\,\sigma_q}$ $AOF_{gof} = \frac{2B + R + R_{log}}{4}$	$-\infty - 1$	-	Beck et al. (2016)

Note: QO_i is the observed runoff; QS_i is the simulated runoff; $\overline{QO_i}$ is the mean of the observed runoff values; $\overline{QS_i}$ is the mean of the simulated runoff values; B and R represent, respectively, the bias and Pearson correlation coefficient computed between simulated and observed runoff; and R_{log} is the Pearson correlation coefficient computed between natural-log transformed simulated and observed runoff.

3.1.4. Hydrological consistency and inter-model agreement

We use six signature measures of hydrologic behavior (Yilmaz et al. 2008; Pokhrel et al. 2012; Mendoza et al. 2015) to assess the model capability to reproduce the water balance, runoff seasonality and hydrological signatures from the flow duration curve (FDC), whose mathematical formulation is presented in Appendix A. Further, we evaluate the capability of the model sampling approach to select model structures that produce hydrologically coherent simulations through the examination of monthly states and fluxes (i.e. streamflow, ET, soil moisture, snow water equivalent – SWE –, baseflow and surface runoff). Inter-model agreement in the simulation of seasonal cycles is quantified with the following equation:

$$Sd = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (2)$$

Where, x_i is the value of the variable of interest defined from the sample elements, and the subscript i corresponds to the i-th unit of the sample elements, $\bar{x}$ is the average value of the sample elements, and n is the number of sample elements.

Once the behavior of internal model states and fluxes in MMP-dry and MMP-wet is analyzed, we conduct a final screening procedure to discard potentially problematic model structures, obtaining final multi-model ensembles MMPS-dry and MMPS-wet from dry and wet calibration periods, respectively.

3.1.5. Hydrologic sensitivities of runoff

We compute hydrologic sensitivities for the period Sep/1986 – Aug/2016 to include both dry and wet periods. Following Vano et al. (2012), we create modified climates by using multiplicative perturbations in precipitation (70%, 80%, 90%, 100% and 110%), which are used to compute precipitation elasticities (ε), and additive perturbations in temperature (0°, 1°, 2°, 3°C) that are used for temperature sensitivities (S). Hydrologic sensitivities are estimated using 1% and 0.1°C incremental changes in precipitation and temperature, respectively, relative to each reference climate. We select these increments to be as small as possible to approximate the tangent (versus the secant), and reduce computational artifacts. We estimate ε as the fractional change in average annual runoff (Q) divided by the fractional change imposed on precipitation:

$$\varepsilon = \frac{\frac{Q_{ref+\Delta\%} - Q_{ref}}{Q_{ref}}}{\Delta\%} \qquad (3)$$

where Δ=1%. Temperature sensitivities are estimated by perturbing air temperature instead of PET, since the former variable is the most widely used in climate change impact assessments. We estimate S as the percent change in average annual runoff due to temperature changes as:

$$S = \frac{\frac{Q_{ref+\Delta} - Q_{ref}}{Q_{ref}}}{\Delta} \qquad (4)$$

where Δ=1°C.

4. RESULTS AND DISCUSSION

4.1. Choice of hydrological model structure

Figure 4 shows the coverage results from all calibrated hydrological models. In general, we observe that the Vilcanota River basin shows the best performance in contrasting climates, compared to the Huancané and Puyango-Tumbes River basins, where only a few model structures manage to meet the acceptable performance limits. Indeed, most model structures in the Vilcanota River basin (Figure 4, left panels) are in the shaded area, excepting the case where the Pareto scheme is applied with the NSE-log. For such criteria, only a few models meet the acceptance thresholds in both dry and wet periods since the calibration objective function (RMSE) is focused on high flows; notably, all the model structures selected in the two calibration periods (colored dots) are in the shaded area, excepting one structure in the NSE-log diagram for the wet calibration period (orange dot, split-KGE). In the Puyango-Tumbes River basin (Figure 4, center panels) only a few model structures fall within the shaded area, with the lowest performances obtained when the Pareto scheme is applied with the NSE-log and AOF. However, most of the selected models (colored dots) are in the shaded area.

Interestingly, the Huancané River basin (Figure 4, right panels) emerges as a challenging case study, since very few model structures fall within the shaded area, with the lowest performances obtained for NSE-log. In the diagrams for the calibration dry period, only one selected model is found mostly within the shaded area (i.e., selected model structure of NSE). But when the calibration is conducted during a wet period, few models fall within the shaded area, with evaluation metrics for which no model structures meet the acceptance thresholds (i.e., NSE, split-KGE). Overall, it is observed that the temporal transferability of model performance in the Huancané River basin is quite poor, regardless of the evaluation criteria, since most model structures have good performance in the calibration period, but not in the evaluation period. In other words, the parameter sets found from the optimization of RMSE provide acceptable performance in terms of other evaluation metrics within the calibration period (especially if this is wet), but not necessarily for the evaluation period. This can be explained by the simplifications adopted for some process representations (e.g., canopy interception, transpiration and others), and

possibly also by the level of spatial disaggregation of model structures considered in this case.

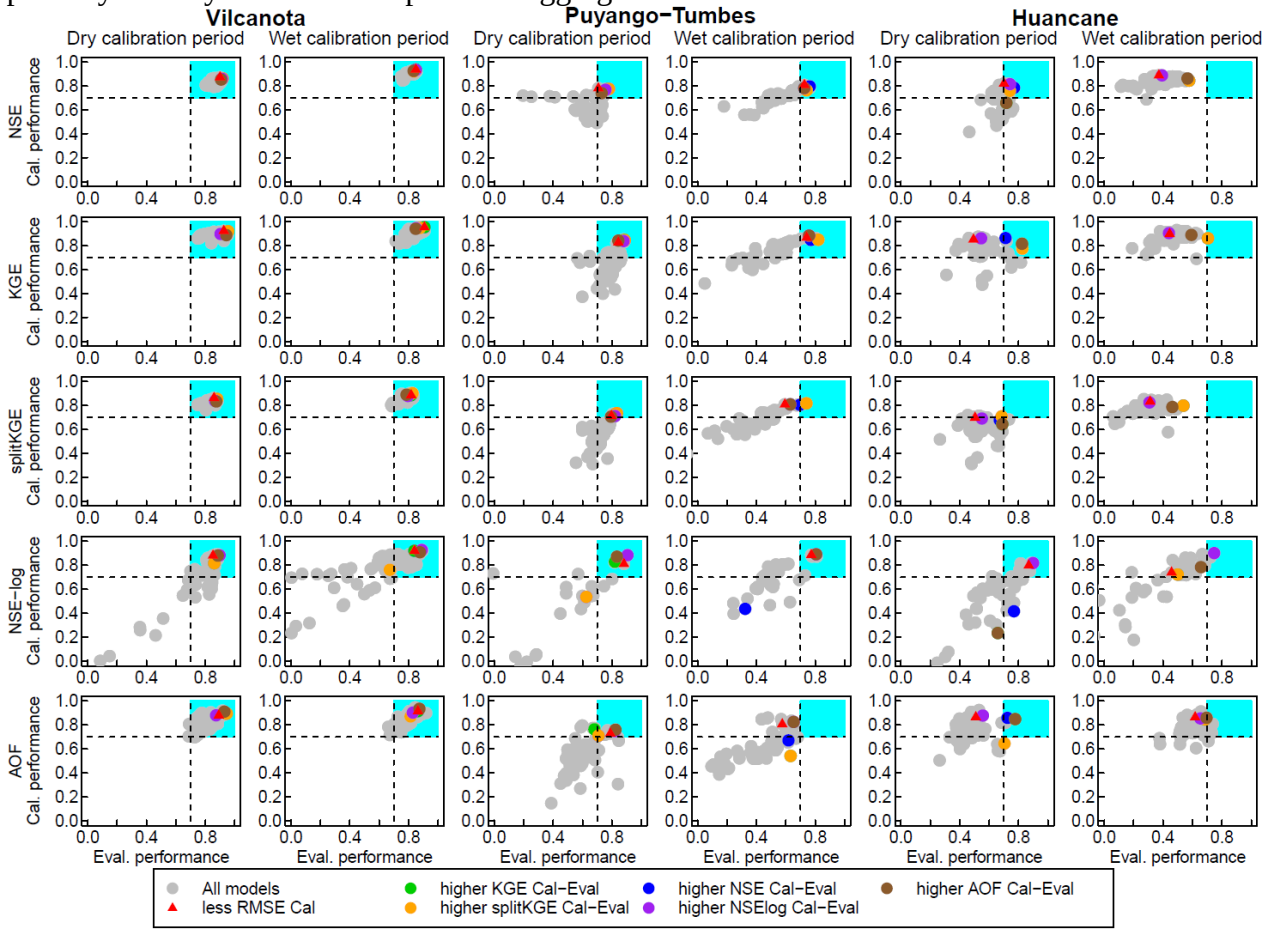

Figure 4: Coverage results from all calibrated model structures, for each basin and each calibration period (displayed along different columns). The horizontal and vertical dashed lines indicate performance acceptance thresholds, and the light blue region represents the region where temporally consistent performance is obtained. The red triangle represents the combination of model structure and parameter set that minimizes RMSE during the calibration period (i.e., the common practice); the colored dots represent the models that were selected using the criteria defined in section 3.1.3, and the remaining models are displayed as gray dots.

4.2. Hydrological consistency and inter-model differences

Figures 5 and 6 illustrate the performance of all model structures, for each basin and calibration period, in terms of signature measures of hydrologic behavior (Figure 5) and flow duration curve (Figure 6). In general, the model sampling approach provides an ensemble of structures with good performance, improving inter-model agreement in comparison to the original ensemble (grey symbols in Figure 5, and gray lines in Figure 6). Further, it is observed that model performance depends considerably on the calibration period (i.e., dry or wet).

Figure 5 shows that high performance in RR, CTR y FHV signatures and low performance in FMS, FLV, and FMM signatures is obtained in all basins, suggesting that many of the structures are unable to represent low and medium flows, which is also reflected in the flow duration curve (Figure 6). However, it should be noted that according to the duration curve graphs (Figure 6), there is an underestimation of the simulated values with respect to the observed values, but according to the hydrological signatures (Figure 5), there is an overestimation. This variation in

the results is due to the logarithmic transformation performed (see Appendix A) and the low values, which means that the results of the hydrological signatures for FMS, FLV and FMM indicate an underestimation but not an overestimation. The low performance of FMS, FLV and FMM signatures can be explained by the choice of RMSE as the objective calibration function and to the selected calibration period (i.e., dry or wet period). For example, in the Vilcanota River basin (Figure 5, left panels) there is more bias in FLV with EVAL W-> D (1488%), in comparison with EVAL D-> W (414%). This relative performance in wet and dry periods is also obtained in the Huancané River basin (Figure 5, right panels). However, in the Puyango-Tumbes River basin (Figure 5, center panels), there is more bias with EVAL D-> W (869%), in comparison with EVAL W-> D (758%).

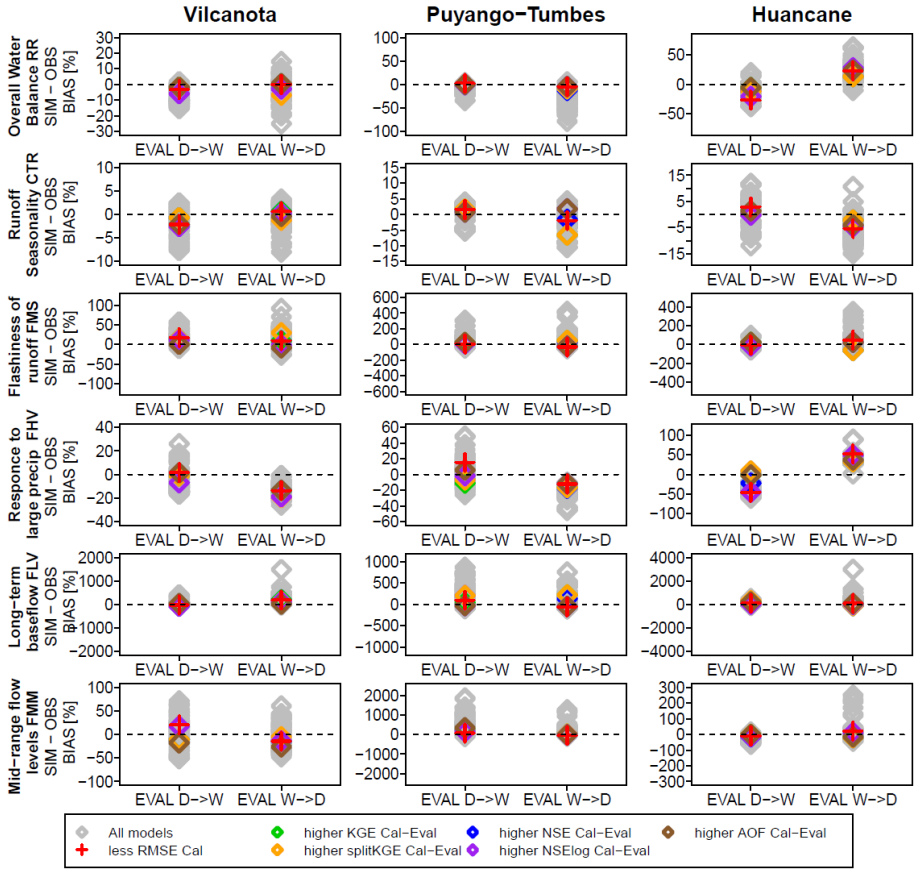

Figure 5: Percent biases in signature measures of hydrologic behavior (rows) for each basin and each calibration period (columns), where EVAL W->D (D->W) indicates model performance in a dry (wet) period with parameters calibrated in a wet (dry) period.

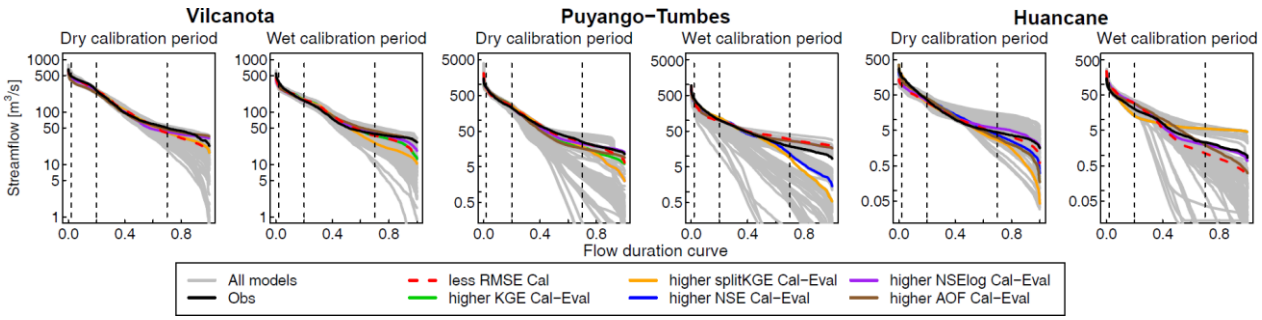

Figure 6: Flow duration curves for each basin and each calibration period. The black line represents observations, gray lines represent the full multi-model ensemble, and the model structures selected with different performance evaluation criteria are displayed in colored lines.

Figure 7 displays climatological averages (Sep/1986-Aug/2016) of monthly state variables and fluxes for each basin and calibration period. One can note that the largest inter-model differences are obtained for soil moisture, SWE, baseflow and surface runoff, even if there is good agreement in streamflow seasonality among the model structures of the full ensemble (e.g., Vilcanota and Puyango-Tumbes River basins). The largest dispersion of simulated states and fluxes is obtained during spring and summer – where most precipitation events occur –, and the smallest dispersion occurs during fall and winter, excepting ET at the Huancané River basin. Further, inter-model agreement improves among the selected model structures in comparison to the full 78-member ensemble provided by FUSE. For example, the multi-model selection scheme in the Vilcanota River basin for the dry period, allows a considerable reduction in the standard deviation of soil moisture from 291 mm to 187 mm (35.7%) with respect to the full model ensemble (gray color); nevertheless, such reduction is not achieved when the model structures are calibrated in a wet period, since here the selected model structures show an increase in the standard deviation of soil moisture from 200 mm to 298 mm (49%) with respect to all model structures. In the Huancané River basin, the opposite behavior is observed: the small multi-model reduces the original spread in simulated soil moisture when the calibration period is dry, but slightly increases the standard deviation if the models are calibrated in a wet period.

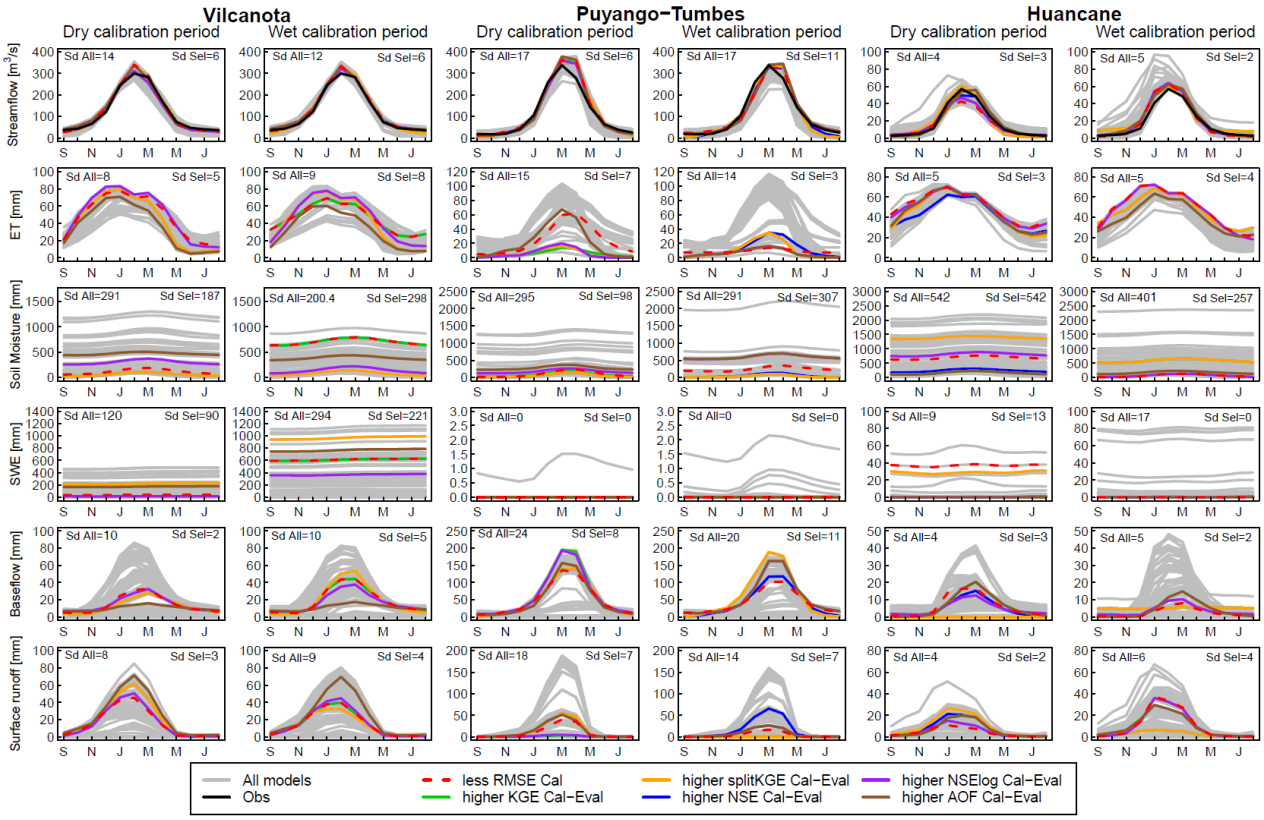

Figure 7: Monthly average fluxes and states for each basin and each calibration period, considering a 30-year period (September/1986 – August/2016). Inter-model agreement is quantified with an ensemble spread metric (equation 2), displayed at the top of each panel for the full ensemble (left) and the five model structures – represented by colored lines – selected with the Pareto scheme (right). The reference model structure that provides the lowest RMSE during the calibration period is displayed in dashed red, for comparison purposes.

The above analyses not only illustrate the potential of the model selection framework to reduce structural uncertainty in internal fluxes and states, but also highlight the need to examine simulated

internal fluxes and states to discard problematic model structures. Hence, we screen the selected models based on acceptance performance thresholds, and poor or abnormal behavior of internal states and fluxes (Figure 7). The results obtained after this model screening procedure are shown in Tables 5, 6 and 7. Note that the model structures are named independently of other papers (i.e., FUSE 1,2,3,..,78).

In the Vilcanota River basin, we dismiss FUSE 25 and FUSE 17 from the dry calibration period (MMP-dry) and FUSE 44, FUSE 21, FUSE 01 y FUSE 68 from the wet calibration period (MMP-wet), both due to an increasing trend in the SWE daily time series (not shown here) that has caused the abnormal behavior of the monthly SWE (Figure 7, left panels). In the Puyango-Tumbes River basin, we dismiss FUSE 44, FUSE 56, FUSE 59 and FUSE 62 from MMP-wet, because they do not meet the minimum performance threshold (Figure 4, center panels). For the Huancané River basin, we dismiss FUSE 03, FUSE 23, FUSE 34, FUSE 45 and FUSE 23 from MMP-dry because they do not meet the minimum performance threshold criteria. For the same basin, we discard FUSE 43, FUSE 69, FUSE 44 and FUSE 07 from MMP-wet, because they do not meet the minimum performance threshold. In the three basins, we interestingly found that all model structures from MMP-wet have not been able to meet the requirements of the screening procedure to discard models (MMPS-wet). In contrast, in the model structures from MMP-dry, we found that at least one model structure has passed the screening procedure (MMPS-dry, Vilcanota and Huancané River basins), this result shows that it is possible to obtain consistent simulations in the flows and states when model structures are calibrated in dry periods. For example, in the Huancané River basin, only the model structure (FUSE 77) from the dry calibration period was able to pass the screening procedure to discard models, where we highlight the discard of the model structure selected with the criterion of minimizing RMSE in MMP-dry and MMP-wet, suggesting that RMSE would not be a good calibration function in this basin.

Interestingly in Tables 5, 6 and 7 and without considering the discarding process, we found that some selected model components depend on both calibration period and climatic conditions in the basin. For example, the Percolation (PE) component during the dry calibration period in the Vilcanota River basin is mainly represented by the PRMS equation, but when the calibration period is wet, the mainly modeling choice for PE (ARNO/VIC or PRMS) depends on the evaluation metric selected to apply the Pareto scheme. In the Puyango-Tumbes River basin (with high temperatures and low mean elevation, Table 1), we show that the upper layer (U) and percolation (PE) components are fully represented by ARNO/VIC - TOPMODEL and ARNO/VIC equations, respectively, regardless of the calibration period or the performance evaluation metric used when applying the Pareto scheme. For the remaining components, it is observed that these depend on the calibration period and the performance evaluation metric used when applying the Pareto scheme: for example, the TOPMODEL equation is preferred for the Surface Runoff components (SR) when the calibration period is dry, and the ARNO/VIC equation is preferred when the calibration period is wet. Finally, we do not find a preferred modeling choice for any component across MMP-dry and MMP-wet in this catchment, with the exception of the upper layer component which is mostly represented by ARNO/VIC - TOPMODEL.

Finally, in the hydrographic slopes (i.e. Pacific, Atlantic and Titicaca) it is not possible to characterize any characteristic behavior between the components of the model structures selected with the Pareto scheme (MMP-dry and MMP-wet), because in this study only one basin per hydrographic slopes was considered, and this is not a significant sampling number that allow any characterization.

Table 5: Components of the hydrological model structures obtained from the application of the Pareto scheme in the Vilcanota River basin, for both dry (MMP-dry), and wet (MMP-wet) calibration periods. The reference model structure that provides the lowest RMSE during the calibration period is included for comparison purposes, and the model structures discarded due to abnormal behavior of states and/or fluxes are in italics and bold.

Calibration Period	Selection Criteria	Model structure name	Model structure components			
			U	L	SR	PE
Dry period	smallest RMSE Cal	FUSE 01	Sacramento	Sacramento	ARNO/VIC	PRMS
	higher NSE Cal-Eval	***FUSE 25***	***PRMS***	***PRMS***	***ARNO/VIC***	***PRMS***
	higher KGE Cal-Eval	***FUSE 25***	***PRMS***	***PRMS***	***ARNO/VIC***	***PRMS***
	higher splitKGE Cal-Eval	***FUSE 25***	***PRMS***	***PRMS***	***ARNO/VIC***	***PRMS***
	higher NSElog Cal-Eval	FUSE 43	ARNO/VIC - TOPMODEL	Sacramento	ARNO/VIC	PRMS
	higher AOF Cal-Eval	***FUSE 17***	***Sacramento***	***TOPMODEL***	***TOPMODEL***	***PRMS***
Wet period	***smallest RMSE Cal***	***FUSE 44***	***ARNO/VIC - TOPMODEL***	***Sacramento***	***ARNO/VIC***	***ARNO/VIC***
	higher NSE Cal-Eval	***FUSE 44***	***ARNO/VIC - TOPMODEL***	***Sacramento***	***ARNO/VIC***	***ARNO/VIC***
	higher KGE Cal-Eval	***FUSE 44***	***ARNO/VIC - TOPMODEL***	***Sacramento***	***ARNO/VIC***	***ARNO/VIC***
	higher splitKGE Cal-Eval	***FUSE 21***	***Sacramento***	***ARNO/VIC***	***PRMS***	***PRMS***
	higher NSElog Cal-Eval	***FUSE 01***	***Sacramento***	***Sacramento***	***ARNO/VIC***	***PRMS***
	higher AOF Cal-Eval	***FUSE 68***	***ARNO/VIC - TOPMODEL***	***TOPMODEL***	***TOPMODEL***	***ARNO/VIC***

Table 6: Same as Table 5, but for the Puyango-Tumbes River basin.

Calibration Period	Selection Criteria	Model structure name	Model structure components			
			U	L	SR	PE
Dry period	smallest RMSE Cal	FUSE 50	ARNO/VIC - TOPMODEL	Sacramento	TOPMODEL	ARNO/VIC
	higher NSE Cal-Eval	FUSE 59	ARNO/VIC - TOPMODEL	PRMS	TOPMODEL	ARNO/VIC
	higher KGE Cal-Eval	FUSE 77	ARNO/VIC - TOPMODEL	ARNO/VIC	TOPMODEL	ARNO/VIC
	higher splitKGE Cal-Eval	FUSE 59	ARNO/VIC - TOPMODEL	PRMS	TOPMODEL	ARNO/VIC
	higher NSElog Cal-Eval	FUSE 62	ARNO/VIC - TOPMODEL	TOPMODEL	ARNO/VIC	ARNO/VIC
	higher AOF Cal-Eval	FUSE 65	ARNO/VIC - TOPMODEL	TOPMODEL	PRMS	ARNO/VIC
Wet period	***smallest RMSE Cal***	***FUSE 44***	***ARNO/VIC – TOPMODEL***	***Sacramento***	***ARNO/VIC***	***ARNO/VIC***
	higher NSE Cal-Eval	***FUSE 56***	***ARNO/VIC – TOPMODEL***	***PRMS***	***PRMS***	***ARNO/VIC***
	higher KGE Cal-Eval	***FUSE 59***	***ARNO/VIC – TOPMODEL***	***PRMS***	***TOPMODEL***	***ARNO/VIC***
	higher splitKGE Cal-Eval	***FUSE 59***	***ARNO/VIC – TOPMODEL***	***PRMS***	***TOPMODEL***	***ARNO/VIC***
	higher NSElog Cal-Eval	***FUSE 62***	***ARNO/VIC – TOPMODEL***	***TOPMODEL***	***ARNO/VIC***	***ARNO/VIC***
	higher AOF Cal-Eval	***FUSE 62***	***ARNO/VIC - TOPMODEL***	***TOPMODEL***	***ARNO/VIC***	***ARNO/VIC***

Table 7: Same as Table 5, but for the Huancané River basin.

Calibration Period	Selection Criteria	Model structure name	Model structure components			
			U	L	SR	PE
Dry period	***smallest RMSE Cal***	***FUSE 03***	***Sacramento***	***Sacramento***	***PRMS***	***PRMS***
	higher NSE Cal-Eval	FUSE 77	ARNO/VIC - TOPMODEL	ARNO/VIC	TOPMODEL	ARNO/VIC
	higher KGE Cal-Eval	***FUSE 23***	***Sacramento***	***ARNO/VIC***	***TOPMODEL***	***PRMS***
	higher splitKGE Cal-Eval	***FUSE 34***	***PRMS***	***TOPMODEL***	***PRMS***	***Sacramento***
	higher NSElog Cal-Eval	***FUSE 45***	***ARNO/VIC – TOPMODEL***	***Sacramento***	***ARNO/VIC***	***Sacramento***
Wet period	***higher AOF Cal-Eval***	***FUSE 23***	***Sacramento***	***ARNO/VIC***	***TOPMODEL***	***PRMS***
	smallest RMSE Cal	***FUSE 43***	***ARNO/VIC – TOPMODEL***	***Sacramento***	***ARNO/VIC***	***PRMS***
	higher NSE Cal-Eval	***FUSE 69***	***ARNO/VIC – TOPMODEL***	***TOPMODEL***	***TOPMODEL***	***Sacramento***
	higher KGE Cal-Eval	***FUSE 69***	***ARNO/VIC – TOPMODEL***	***TOPMODEL***	***TOPMODEL***	***Sacramento***
	higher splitKGE Cal-Eval	***FUSE 69***	***ARNO/VIC – TOPMODEL***	***TOPMODEL***	***TOPMODEL***	***Sacramento***
	higher NSElog Cal-Eval	***FUSE 44***	***ARNO/VIC – TOPMODEL***	***Sacramento***	***ARNO/VIC***	***ARNO/VIC***
	higher AOF Cal-Eval	***FUSE 07***	***Sacramento***	***PRMS***	***ARNO/VIC***	***PRMS***

4.3. Hydrologic sensitivities.

Figure 8 illustrates the impacts that changes in precipitation and temperature produce on monthly streamflow values simulated by MM0-dry (gray lines) and MMP-dry (colored lines) in the Vilcanota River basin. One can note that precipitation increments produce Q and ET increases under the assumption that temperature remains unchanged, and these variations are more noticeable during spring and summer, where the highest values occur. Assuming that precipitation does not change, it is obtained that temperature increments produce increases in ET and a decrease in Q, and similar sensitivities are observed in the other basins, regardless of the calibration period (not shown).

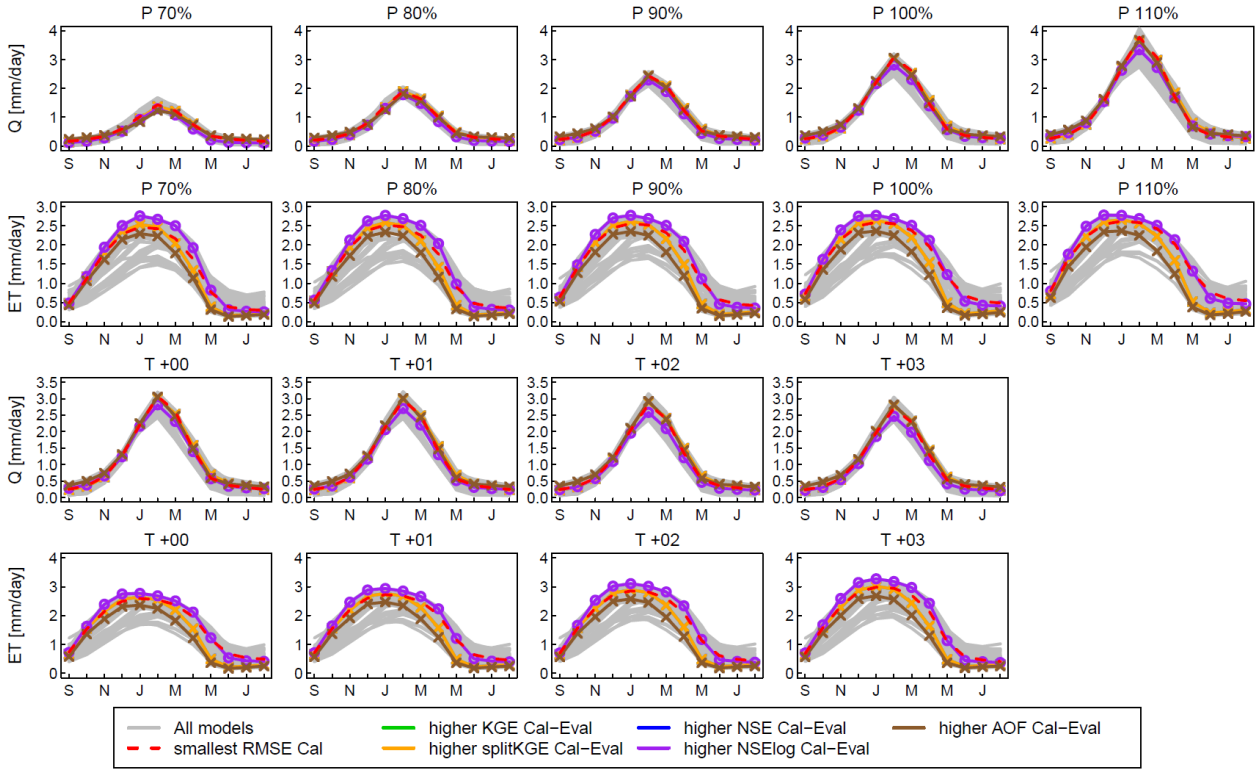

Figure 8: Climatological monthly averages (September/1986 – August/2016) of runoff and ET obtained with model parameters calibrated in a dry period. Results are displayed for (top) precipitation perturbations of 70%, 80%, 90% and 110%, and (bottom) temperature increases of 1°, 2° and 3°C. The gray lines show the results with the full ensemble (MM0-dry), and the colored lines show the results obtained with the multi-model ensemble obtained from the application of the Pareto scheme (MMP). The model structures discarded during the screening procedure are plotted with x, and the accepted model structures (MMS) are plotted with circles.

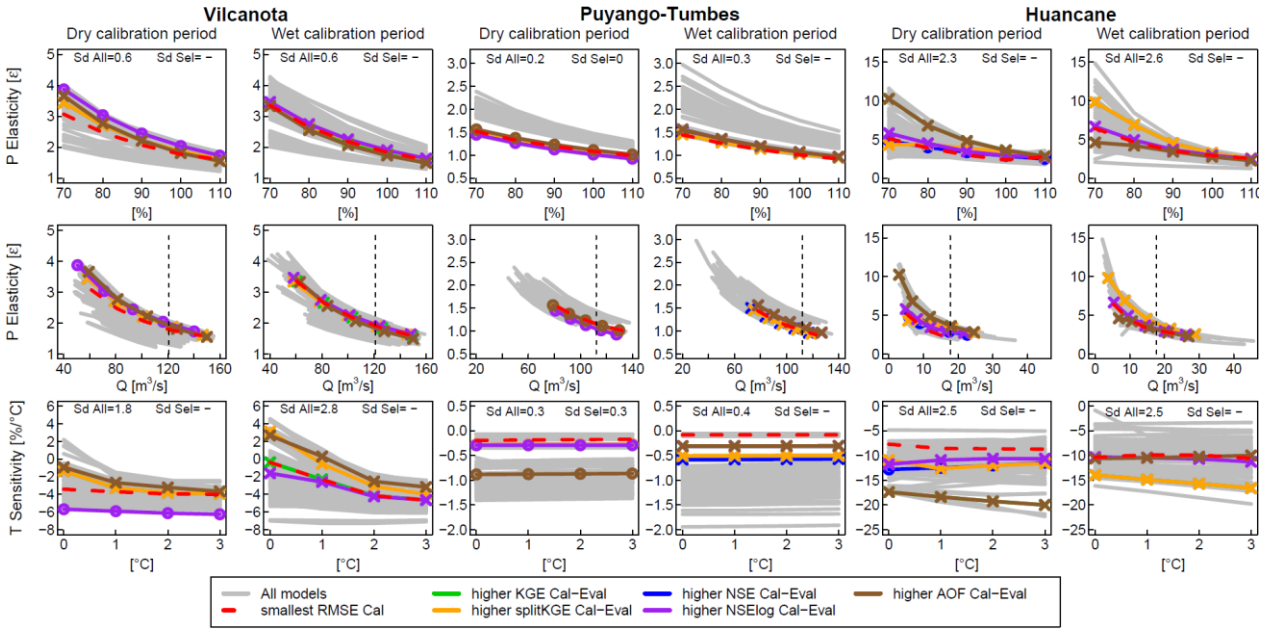

Figure 9: Precipitation elasticity (ε) and temperature sensitivities (S) for each basin and calibration period, computed for the 30-year period (September/1986 – August/2016). In the top panels, the x-axis represents the percent changes in precipitation from the reference climates; in the middle panels, the x-axis represents

the changes in mean annual runoff from the reference climates, and in the bottom panels the x-axis represents the additive temperature changes from the reference climates. The average monthly standard deviations obtained from the full ensemble (MM0) and the final multi-model ensemble (MMPS) are displayed at the top of each panel. Discarded model structures are represented with x, and accepted model structures are plotted with circles. The vertical dashed line is the observed mean annual streamflow.

We assess the spread in precipitation elasticities (ε) and temperature sensitivities (S) arising from model structure in each basin, calibration period and reference climate (Figure 9, top and middle panels). In agreement with the results reported by Vano et al. (2012), precipitation elasticities from all model structures (Figure 9, top panels) are non-linear and depend on the reference climate, with higher elasticities for drier conditions (i.e., -10%, -20% and -30%). Figure 9 (middle panels) shows that lower (higher) values of mean annual streamflow are related to larger (smaller) precipitation elasticities, and higher values of annual average streamflow are related to lower values of elasticity. The simulated temperature sensitivities (S, Figure 9, bottom panels) were largely negative with the exception of some model structures in the Vilcanota River basin, since as T increases, ET increases and Q decreases in agreement with the result of Vano et al. (2012) for the Colorado River Basin, USA. Moreover, we observe large inter-model differences in temperature sensitivities, and no clear relationships (i.e., trends) between S values and temperature perturbations.

Figure 9 shows that, among the three catchments, the Huancané River basin provides the highest ε and S values, which in turn leads to a larger spread arising from model structures. On the other hand, the lowest ε and S values are obtained at the Puyango-Tumbes River basin, which can be explained by its geographic location (very close to the equatorial line), where higher temperatures (sea monthly averages in Figure 3) and mean annual precipitation (Table 1) are observed; in this area, small temperature (0.1°C) or precipitation (1%) perturbations do not have much impact on streamflow in comparison to the other basins.

In general, we observe that the large dispersion in precipitation elasticities and temperature sensitivities arising from the original FUSE multi-model ensemble (MM0) decrease when a Pareto scheme is applied (MMP), for all basins and calibration periods. Further, a greater reduction of structural uncertainty is obtained after conducting a model screening step based on the examination of hydrological signatures and model states and fluxes (MMPS). For example, the ensemble spread in ε and S reduces 100% (i.e., 0.2 to 0) and 0% (i.e., 0.3 to 0.3), respectively, in the Puyango-Tumbes basin for a dry calibration period. Larger reductions are obtained in the Vilcanota and Huancané River basins in the dry calibration period (MMPS-dry), decreasing to 100% the structural uncertainty, since they have only one model structure, and therefore, the standard deviation cannot be calculated. Finally, by applying this Pareto scheme (MMP) in the three selected basins of the three hydrographic slopes, the uncertainty in the sensitivities is considerably reduced, highlighting the potential of the methodology proposed in this article.

5. CONCLUSIONS

In this paper, we test the capability of a simple framework to sample hydrological model structures in order to (1) provide hydrologically consistent simulations under contrasting climatic conditions, and (2) reduce the uncertainty arising from hydrologic model choice in precipitation elasticities and temperature sensitivities. The analyses are conducted in three case study basins in Peru of three different hydrographic slopes that are susceptible to flood occurrence. We configure and calibrate 78 FUSE models in dry and wet periods, obtain a sample of model structures using a Pareto scheme,

refine the selection based on model diagnostics, assess hydrological consistency and quantify hydrologic sensitivities. Further, we examine possible similarities between the selected model structures. The main conclusions are as follows:

- The approach allows to identify structures that are capable to simulate the catchment-scale hydrology under different climatic conditions, and these models provide good performances for various efficiency metrics, hydrological signatures and seasonal water balances.
- Some selected model components depend on the calibration period and hydroclimatic characteristics of the basin, and other model components do not present any dependence, as is the case of all the components of the Huancané River basin.
- The spread in precipitation elasticities and temperature sensitivities is reduced with the chosen model structures in comparison to the full, 78-member model ensemble.
- By discarding the model structures that do not meet the minimum performance threshold and coherent behavior of states and fluxes in the three basins of the three hydrographic slopes, we obtain a larger reduction in the spread of precipitation elasticities and temperature sensitivities. In addition, we find that it is mostly possible to obtain consistent simulations of flows and states when model structures are calibrated in dry periods.

The results presented here reinforce the idea that inter-model agreement in climate impact metrics does not necessarily improve if traditional objective functions are used for parameter estimation (Mendoza et al., 2015), and that further evaluation under contrasting climatic conditions, together with assessments of hydrological consistency, may help to guide the selection of hydrological models for climate impact studies. Also, our results highlight the challenge of designing model sampling strategies that provide a coherent model ensemble in terms of process representations, especially in basins that are 'problematic' (e.g., Huancané). Future studies could address this problem by incorporating other sources of information (Nemri & Kinnard, 2020; Nijzink et al., 2018; Sleziak et al., 2020; Széles et al., 2020) to find behavioral combinations of model structures and parameters that can be incorporated in the Pareto scheme.

ACKNOWLEDGEMENTS

The authors thank the authors thank the anonymous reviewers for their valuable comments and suggestions. The authors also thank the Servicio Nacional de Meteorología e Hidrología del Peru for providing the hydrometeorological data necessary for the research.

3. CONCLUSIONES

En este trabajo de tesis, se evalúa un esquema simple para seleccionar estructuras de modelos hidrológicos que proporcionen simulaciones coherentes en condiciones climáticas distintas. La metodología es aplicada en tres cuencas de estudio de caso del Perú de tres diferentes vertientes hidrográficas que son susceptibles de sufrir inundaciones, y el enfoque tiene por objeto disminuir la propagación de las sensibilidades hidrológicas a los cambios en la precipitación y la temperatura. Para este fin, se ha calibrado y configurado 78 modelos FUSE en períodos secos y húmedos, obteniendo una muestra de las estructuras de los modelos utilizando un esquema de Pareto, se refina la selección basada en los diagnósticos de los modelos y se evalúa la consistencia hidrológica y cuantifican las sensibilidades hidrológicas. También se analizó las posibles similitudes entre las estructuras de los modelos seleccionados. Las principales conclusiones son las siguientes:

- El enfoque permite identificar estructuras capaces de simular la hidrología a escala de cuenca en diferentes condiciones climáticas, y que estas estructuras tienen buenos resultados en diferentes métricas de eficiencia, firmas hidrológicas y balances hídricos estacionales.
- Algunos componentes del modelo dependen del período de calibración y de las características hidroclimáticas de la cuenca, pero también hay componentes del modelo que no presentan ninguna dependencia, como es el caso de todos los componentes de la cuenca del río Huancané.
- La dispersión de las elasticidades de las precipitaciones y las sensibilidades a la temperatura se reduce con las estructuras de modelo elegidas, en comparación con el conjunto completo del modelo de 78 miembros.
- Al descartar las estructuras de los modelos que no cumplen con el umbral mínimo de rendimiento y el comportamiento coherente de los estados y flujos en las tres cuencas de las tres vertientes hidrográficas, se obtiene una mayor reducción en la propagación de las elasticidades de la precipitación y las sensibilidades a la temperatura, lo que se demuestra con la desviación estándar. Además, se encuentra que mayormente es posible obtener simulaciones consistentes en los flujos y estados, cuando las estructuras de modelos son calibradas en periodos secos.

Los resultados presentados aquí refuerzan la idea de que la concordancia entre modelos en las métricas de impacto climático no mejora necesariamente si se utilizan funciones objetivo tradicionales para la estimación de parámetros (Mendoza et al., 2015), y que una mayor evaluación en condiciones climáticas contrastadas, junto con evaluaciones de la coherencia hidrológica, puede ayudar a orientar la selección de modelos hidrológicos para los estudios de impacto climático. Asimismo, estos resultados ponen de manifiesto el reto de diseñar estrategias de muestreo de modelos que proporcionen un conjunto de modelos coherente en términos de representaciones de procesos, especialmente en cuencas que son "problemáticas" (por ejemplo, Huancané). Los estudios futuros podrían abordar este problema incorporando otras fuentes de información (Nemri & Kinnard, 2020; Nijzink et al., 2018; Sleziak et al., 2020; Széles et al., 2020) para encontrar combinaciones de comportamiento de las estructuras y parámetros del modelo que puedan incorporarse al esquema de Pareto. Además, de evaluar la posibilidad de utilizar otras funciones objetivas y otras plataformas de modelación.

4. BIBLIOGRAFÍA

Addor, N., & Melsen, L. A. (2019). Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models. *Water Resources Research*, *55*(1), 378–390. https://doi.org/10.1029/2018WR022958

Addor, Nans, Rössler, O., Köplin, N., Huss, M., Weingartner, R., & Seibert, J. (2014). Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments. *Water Resources Research*, *50*(10), 7541–7562. https://doi.org/10.1002/2014WR015549

Alvarez-Garreton, C., Boisier, J. P., Garreaud, R., Seibert, J., & Vis, M. (2021). Progressive water deficits during multiyear droughts in basins with long hydrological memory in Chile. *Hydrology and Earth System Sciences*, *25*(1), 429-446. https://doi.org/10.5194/hess-25-429-2021

ANA. (2009). *Demarcación y Delimitación de las Autoridades Administrativas del Agua*. https://sinia.minam.gob.pe/modsinia/public/docs/2826.pdf

Anderson, E. A. (2006). Snow accumulation and ablation model–SNOW-17. *US National Weather Service, Silver Spring, MD*. http://www.nws.noaa.gov/oh/hrl/nwsrfs/users_manual/part2/_pdf/22snow17.pdf.

Andres, N., Vegas Galdos, F., Casimiro, W. G. S., & Zappa M. (2014). Water resources and climate change impact modelling on a daily time scale in the Peruvian Andes. *Journal des Sciences Hydrologiques*, 59, 2043-2059. http://dx.doi.org/10.1080/02626667.2013.862336.

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., & Obando, O. F. (2019). Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, 1-16. https://doi.org/10.1080/02626667.2019.1649411

Bae, D.-H., Jung, I.-W., & Lettenmaier, D. P. (2011). Hydrologic uncertainties in climate change from IPCC AR4 GCM simulations of the Chungju Basin, Korea. *Journal of Hydrology*, *401*(1–2), 90–105. https://doi.org/10.1016/j.jhydrol.2011.02.012

Beck, H. E., van Dijk, A. I., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, *52*(5), 3599-3622. https://doi.org/10.1002/2015WR018247

Beven, K., & Kirkby, M. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*., 24, 43-69. doi:10.1080/02626667909491834

Bhardwaj, K., Shah, D., Aadhar, S., & Mishra, V. (2020). Propagation of meteorological to hydrological droughts in India. *Journal of Geophysical Research: Atmospheres*, *125*(22), e2020JD033455. https://doi.org/10.1029/2020JD033455

Brigode, P., Oudin, L., & Perrin, C. (2013). Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?. *Journal of Hydrology*, *476*, 410-425. https://doi.org/10.1016/j.jhydrol.2012.11.012

Burnash, R. J. C. (1995). The NWS river forecast system-catchment modeling. *Computer models of watershed hydrology*, 311-366.

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., et al. (2019). How do modeling decisions affect the spread among hydrologic climate change projections? Exploring a large ensemble of simulations across a diversity of hydroclimates. *Earth's Future*, *7*, 623–637. https://doi.org/10.1029/2018EF001047

Chen, J., Brissette, F. P., Poulin, A., & Leconte, R. (2011). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resources Research*, *47*, W12509. https://doi.org/10.1029/2011WR010602

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., et al. (2008).

Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. *Water Resources Research*, *44*, W00B02. https://doi.org/10.1029/2007WR006735

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*, W09301. https://doi.org/10.1029/2010WR009827

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., et al. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*. https://doi.org/10.1002/2015WR017198

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., ... & Brekke, L. D. (2016). Characterizing uncertainty of the hydrologic impacts of climate change. *Current Climate Change Reports*, *2*(2), 55-64. DOI 10.1007/s40641-016-0034-x

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

Correa, S. W., de Paiva, R. C. D., Espinoza, J. C., & Collischonn, W. (2017). Multi-decadal Hydrological Retrospective: Case study of Amazon floods and droughts. *Journal of Hydrology*, *549*, 667-684. https://doi.org/10.1016/j.jhydrol.2017.04.019

Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J. M., Howden, N. J. K., et al. (2019). DECIPHeR v1: Dynamic fluxEs and ConnectIvity for Predictions of HydRology. *Geoscientific Model Development*, *12*(6), 2285–2306. https://doi.org/10.5194/gmd-12-2285-2019

Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, R. W., Jost, G., Lee, K., et al. (2020). Flexible watershed simulation with the Raven hydrological modelling framework. *Environmental Modelling and Software*, *129*(December 2019), 104728. https://doi.org/10.1016/j.envsoft.2020.104728

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research*, *28*(4), 1015-1031. https://doi.org/10.1029/91WR02985

Duethmann, D., Blöschl, G., & Parajka, J. (2020). Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change?. *Hydrology and Earth System Sciences*, 24(7), 3493-3511. https://doi.org/10.5194/hess-24-3493-2020

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., et al. (2018a). Simulating Runoff Under Changing Climatic Conditions: A Framework for Model Improvement. *Water Resources Research*, *54*(12), 9812–9832. https://doi.org/10.1029/2018WR023989

Fowler, K., Peel, M., Western, A., & Zhang, L. (2018b). Improved rainfall-runoff calibration for drying climate: Choice of objective function. *Water Resources Research*, *54*(5), 3392-3408. https://doi.org/10.1029/2017WR022466

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

Gavrilović, L., Milanović Pešić, A., & Urošev, M. (2012). A hydrological analysis of the greatest floods in Serbia in the 1960–2010 period. *Carpathian Journal of Earth and Environmental Sciences*, *7*(4), 107-116.

Grigg, A. H., & Hughes, J. D. (2018). Nonstationarity driven by multidecadal change in catchment groundwater storage: A test of modifications to a common rainfall–run-off

model. *Hydrological Processes*, *32*(24), 3675–3688. https://doi.org/10.1002/hyp.13282

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*(1-2), 80-91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haile, G. G., Tang, Q., Sun, S., Huang, Z., Zhang, X., & Liu, X. (2019). Droughts in East Africa: Causes, impacts and resilience. *Earth-science reviews*, *193*, 146-161. https://doi.org/10.1016/j.earscirev.2019.04.015

Heidinger, H., Carvalho, L., Jones, C., Posadas, A., & Quiroz, R. (2018). A new assessment in total and extreme rainfall trends over central and southern Peruvian Andes during 1965–2010. *International Journal of Climatology*, *38*, e998-e1015. https://doi.org/10.1002/joc.5427

Henn, B., Clark, M. P., Kavetski, D., & Lundquist, J. D. (2015). Estimating mountain basin-mean precipitation from streamflow using B ayesian inference. *Water Resources Research*, *51*(10), 8012-8033. https://doi.org/10.1002/2014WR016736

Huerta, A., Aybar, C., & Lavado-Casimiro. (2018). W. PISCO temperatura v.1.1. SENAMHI – DHI – 2018, Lima – Perú. http://iridl.ldeo.columbia.edu/documentation/.pisco/.PISCOt_report.pdf

IPCC. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF et al. (eds). Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 1535.

Jiang, T., Chen, Y. D., Xu, C., Chen, X., Chen, X., & Singh, V. P. (2007). Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. *Journal of Hydrology*, *336*(3–4), 316–333. https://doi.org/10.1016/j.jhydrol.2007.01.010

Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, *47*(11), 1–19. https://doi.org/10.1029/2011WR010748

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13-24. https://doi.org/10.1080/02626668609491024

Knoben, W. J. M., Freer, J. E., Fowler, K. J. A., Peel, M. C., & Woods, R. A. (2019). Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) v1.0: an open- source, extendable framework providing implementations of 46 conceptual hydrologic models as continuous space-state formulations. *Geoscientific Model Development Discussions*, 1–26. https://doi.org/10.5194/gmd-2018-332

Lavado Casimiro, W. G. S., Labat D., Guyot, J. L., & Bardin, S. A. (2011). Assessment of climate change impacts on the hydrology of the Peruvian Amazon–Andes basin. *Hydrological Processes*, 25, 3721-3734. https://doi.org/10.1002/hyp.8097

Lavado Casimiro, W. S., Ronchail, J., Labat, D., Espinoza, J. C., & Guyot, J. L. (2012). Basin-scale analysis of rainfall and runoff in Peru (1969–2004): Pacific, Titicaca and Amazonas drainages. *Hydrological Sciences Journal*, *57*(4), 625-642. https://doi.org/10.1080/02626667.2012.672985

Leavesley, G. H. (1984). *Precipitation-runoff modeling system: User's manual* (Vol. 83, No. 4238). US Department of the Interior.

Lehner, F., Wood, A. W., Vano, J. A., Lawrence, D. M., Clark, M. P., & Mankin, J. S. (2019). The potential to reduce uncertainty in regional runoff projections from climate models. *Nature Climate Change*, *9*(12), 926–933. https://doi.org/10.1038/s41558-019-0639-x

Llauca, H., Lavado-Casimiro, W., León, K., Jimenez, J., Traverso, K., & Rau, P. (2021a). Assessing Near Real-Time Satellite Precipitation Products for Flood Simulations at Sub-Daily Scales in a Sparsely Gauged Watershed in Peruvian Andes. *Remote Sensing*, *13*(4), 826. https://doi.org/10.3390/rs13040826

Llauca, H., Lavado-Casimiro, W., Montesinos, C., Santini, W., & Rau, P. (2021b). PISCO_HyM_GR2M: A Model of Monthly Water Balance in Peru (1981–2020). *Water*, *13*(8), 1048. https://doi.org/10.3390/w13081048

Martínez, A., & Céspedes, L. (2017). Estudio de la vulnerabilidad presente y futura ante el cambio climático en la región Tumbes. Informe Técnico Especial.

Melsen, L. A., Addor, N., Mizukami, N., Newman, A. J., Torfs, P. J., Clark, M. P., ... & Teuling, A. J. (2018). Mapping (dis) agreement in hydrologic projections. *Hydrology and Earth System Sciences*, *22*(3), 1775-1791. https://doi.org/10.5194/hess-22-1775-2018

Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A., Barlage, M., Gutmann, E., et al. (2015). Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. *Journal of Hydrometeorology*, *16*(2), 762–780. https://doi.org/10.1175/JHM-D-14-0104.1

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., & Rajagopalan, B. (2016). How do hydrologic modeling decisions affect the portrayal of climate change impacts?. *Hydrological Processes*, *30*(7), 1071-1095. https://doi.org/10.1002/hyp.10684

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, *47*(2), W02531. https://doi.org/10.1029/2010WR009505

Miller, W. P., Butler, R. A., Piechota, T., Prairie, J., Grantz, K., & DeRosa, G. (2012). Water management decisions using multiple hydrologic models within the San Juan River basin under changing climate conditions. *Journal of Water Resources Planning and Management*, *138*(5), 412-420. doi:10.1061/(ASCE)WR.1943-5452.0000237

Milly, P. C., Kam, J., & Dunne, K. A. (2018). On the sensitivity of annual streamflow to air temperature. *Water Resources Research*, *54*(4), 2624-2641. https://doi.org/10.1002/2017WR021970

Mizukami, N., Clark, M. P., Gutmann, E. D., Mendoza, P. A., Newman, A. J., Nijssen, B., et al. (2016). Implications of the Methodological Choices for Hydrologic Portrayals of Climate Change over the Contiguous United States: Statistically Downscaled Forcing Data and Hydrologic Models. *Journal of Hydrometeorology*, *17*(1), 73–98. https://doi.org/10.1175/JHM-D-14-0187.1

Motavita, D. F., Chow, R., Guthke, A., & Nowak, W. (2019). The comprehensive differential split-sample test: A stress-test for hydrological model robustness under climate variability. *Journal of Hydrology*, *573*, 501-515. https://doi.org/10.1016/j.jhydrol.2019.03.054

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

Nemri, S., & Kinnard, C. (2020). Comparing calibration strategies of a conceptual snow hydrology model and their impact on model performance and parameter identifiability. *Journal of Hydrology*, *582*, 124474. https://doi.org/10.1016/j.jhydrol.2019.124474

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., ... & Hrachowitz, M. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, *54*(10), 8332-8362.

https://doi.org/10.1029/2017WR021895
Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12), D12109. https://doi.org/10.1029/2010JD015139
Núñez Juárez, S., & Zegarra Loo, J. (2006). Estudio geoambiental de la cuenca del río Puyango-Tumbes-[Boletín C 32]. *1560-9928*. https://hdl.handle.net/20.500.12544/264
Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, *303*(1-4), 290-306. https://doi.org/10.1016/j.jhydrol.2004.08.026
Pan, Z., Liu, P., Gao, S., Xia, J., Chen, J., & Cheng, L. (2019). Improving hydrological projection performance under contrasting climatic conditions using spatial coherence through a hierarchical Bayesian regression framework. *Hydrology and Earth System Sciences*, *23*(8), 3405-3421. https://doi.org/10.5194/hess-23-3405-2019
Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, *279*(1-4), 275-289. https://doi.org/10.1016/S0022-1694(03)00225-7
Pokhrel, P., Yilmaz, K. K., & Gupta, H. V. (2012). Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *Journal of Hydrology*, *418*, 49-60. https://doi.org/10.1016/j.jhydrol.2008.12.004
Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W. L., Granger, R. J., & Carey, S. K. (2007). The cold regions hydrological model : a platform for basing process representation and model structure on physical evidence. *Hydrological Processes*, *21*, 2650–2667. https://doi.org/10.1002/hyp
Rau, P., Bourrel, L., Labat, D., Ruelland, D., Frappart, F., Lavado, W., ... & Felipe, O. (2018). Assessing multidecadal runoff (1970–2010) using regional hydrological modelling under data and water scarcity conditions in Peruvian Pacific catchments. *Hydrological Processes*, *33*(1), 20-35. https://doi.org/10.1002/hyp.13318
Remmers, J. O. E., Teuling, A. J., & Melsen, L. A. (2020). Can model structure families be inferred from model output? *Environmental Modelling and Software*, *133*(July), 104817. https://doi.org/10.1016/j.envsoft.2020.104817
Rivas, A. T., & Rivas, J. M. T. (2013). Estrategias de adaptación frente al cambio climático en familias rurales del altiplano puneño: estudio de caso en el centro poblado de Huancho-Huancané-Perú. *Comuni@cción: Revista de Investigación en Comunicación y Desarrollo*, *4*(1), 57-73. https://www.redalyc.org/pdf/4498/449844866006.pdf
Sanabria, J., Marengo, J., & Valverde, M. (2009). Escenarios de Cambio Climático con modelos regionales sobre el Altiplano Peruano (Departamento de Puno). Climate change scenarios using regional models for the Peruvian Altiplano (Departament of Puno). *Revista Peruana Geo-Atmosférica*, *1*, 134-149.
Santos, L., Thirel, G., Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, 22, 4583–4591. https://doi.org/10.5194/hess-22-4583-2018
Seiller, G., Anctil, F., Perrin, C., 2012. Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrol. Earth Syst. Sci.* 16 (4), 1171–1189. http://dx.doi.org/10.5194/hess-16-1171-2012

Shiru, M. S., Shahid, S., Chung, E. S., & Alias, N. (2019). Changing characteristics of meteorological droughts in Nigeria during 1901–2010. *Atmospheric Research*, *223*, 60-73. https://doi.org/10.1016/j.atmosres.2019.03.010

Sleziak, P., Szolgay, J., Hlavčová, K., Danko, M., & Parajka, J. (2020). The effect of the snow weighting on the temporal stability of hydrologic model efficiency and parameters. *Journal of Hydrology*, *583*, 124639. https://doi.org/10.1016/j.jhydrol.2020.124639

Son, R., Wang, S. Y. S., Tseng, W. L., Schuler, C. W. B., Becker, E., & Yoon, J. H. (2020). Climate diagnostics of the extreme floods in Peru during early 2017. *Climate Dynamics*, *54*(1), 935-945. https://doi.org/10.1007/s00382-019-05038-y

Staudinger, M., Stahl, K., Seibert, J., Clark, M. P., & Tallaksen, L. M. (2011). Comparison of hydrological model structures based on recession and low flow simulations. *Hydrology and Earth System Sciences*, *15*(11), 3447-3459. https://doi.org/10.5194/hess-15-3447-2011

Stephens, C. M., Marshall, L. A., & Johnson, F. M. (2019). Investigating strategies to improve hydrologic model performance in a changing climate. *Journal of Hydrology*, *579*, 124219. https://doi.org/10.1016/j.jhydrol.2019.124219

Széles, B., Parajka, J., Hogan, P., Silasari, R., Pavlin, L., Strauss, P., & Blöschl, G. (2020). The added value of different data types for calibrating and testing a hydrologic model in a small catchment. *Water resources research*, *56*(10), e2019WR026153. https://doi.org/10.1029/2019WR026153

Takahashi, K., & Martínez Grimaldo, A. (2015). Impacto de la variabilidad y cambio climático en el ecosistema de Manglares de Tumbes, Perú.

Vano, J. A., Das, T., & Lettenmaier, D. P. (2012). Hydrologic sensitivities of Colorado River runoff to changes in precipitation and temperature. *Journal of Hydrometeorology*, *13*(3), 932-949. https://doi.org/10.1175/JHM-D-11-069.1

Vano, J. A., & Lettenmaier, D. P. (2014). A sensitivity-based approach to evaluating future changes in Colorado River discharge. *Climatic Change*, *122*(4), 621–634. https://doi.org/10.1007/s10584-013-1023-x

Vano, J. A., Kim, J. B., Rupp, D. E., & Mote, P. W. (2015). Selecting climate change scenarios using impact-relevant sensitivities. *Geophysical Research Letters*, *42*(13), 5516–5525. https://doi.org/10.1002/2015GL063208

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J.-M., Viney, N. R., & Teng, J. (2010). Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies. *Journal of Hydrology*, *394*(3–4), 447–457. https://doi.org/10.1016/j.jhydrol.2010.09.018

Vegas Galdos, F., Andres, N., Zappa, M., Casimiro, W. S. L., & Hilker, N. (2015), Simulation and characterization of natural regime of daily discharge in the southern Andes of Peru; Apurimac and Cusco regions. *Revista Peruana Geo-Atmosferica RPGA*, 4, 1-18. www.senamhi.gob.pe\rpga

Westra, S., Thyer, M., Leonard, M., Kavetski, D., & Lambert, M. (2014). A strategy for diagnosing and interpreting hydrological model nonstationarity. *Water Resources Research*, *50*(6), 5090–5113. https://doi.org/10.1002/2013WR014719

Wilby, R. L., & Harris, I. (2006). A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. *Water Resources Research*, *42*(2), W02419. https://doi.org/10.1029/2005WR004065

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(9). https://doi.org/10.1029/2007WR006716

Zhao, R. (1977). Flood forecasting method for humid regions of China. *East China Institute of*

Hydraulic Engineering, Nanjing, China.
Zulkafli, Z. D. (2014). *The hydrology of the Peruvian Amazon river and its sensitivity to climate change* (Doctoral dissertation, Imperial College London). https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.650691

5. ANEXOS

Anexo A

This appendix presents the mathematical formulation of the signature measures used in this study. These measures are based on several previous studies (e.g., Mendoza et al. 2015; Yilmaz et al. 2008; Pokhrel et al. 2012):

The diagnostic signature measure for water balance is the percentage bias in the overall runoff rate of the simulation and observations (%Bias RR).

$$\%\ \text{Bias RR} = \frac{\text{RS} - \text{RO}}{\text{RO}} \text{x}100 \qquad (5)$$

Where RO and RS is the mean annual runoff ratio observed and simulated basin average respectively.

The measure for the seasonality of runoff is determined by the percentage bias of the centroid of the daily hydrograph for an average year (%Bias CTR), or the central time of runoff.

$$\%\ \text{Bias CTR} = \frac{(\sum_{j=1}^{N} t_j QS_j / \sum_{j=1}^{N} QS_j) - (\sum_{j=1}^{N} t_j QO_j / \sum_{j=1}^{N} QO_j)}{(\sum_{j=1}^{N} t_j QO_j / \sum_{j=1}^{N} QO_j)} \text{x}100 \qquad (6)$$

From the above equation, QO_j and QS_j is the observed and simulated streamflow respectively associated with t_j, and N is the total number of days in a hydrological year. The number of days in the hydrological year in our case begins on September 1.

The diagnostic signature measures for vertical redistribution are the percent bias in FDC mis-segment slope (% Bias FMS) and the percent bias in FDC high-segment volume (%Bias FHV). The first is computed as:

$$\%\ \text{Bias FMS} = \frac{[\log(QS_{m1}) - \log(QS_{m2})] - [\log(QO_{m1}) - \log(QO_{m2})]}{[\log(QO_{m1}) - \log(QO_{m2})]} \text{x}100 \qquad (7)$$

Where m1 = 0.2 and m2 = 0.7, therefore, Qm1 and Qm2 are flows with probability of exceedance of 0.2 and 0.7 respectively. FMS represents the variability or response of the magnitude of the flow, a steep slope indicates a greater response of the flow of the current, while a flatter curve indicates a relatively damped response and greater storage. The second is computed as:

$$\%\ \text{Bias FHV} = \frac{\sum_{h=1}^{H}(QS_h - QO_h)}{\sum_{h=1}^{H} QO_h} \text{x}100 \qquad (8)$$

Where h=1,2,...,H are the flow indices in the flow matrix with probability of exceedance less than 0.02. FHV represents a measure of basin response to high precipitation and snowmelt events.

The diagnostic signature measure for long-term baseflow is the percent bias in FDC low-segment volume (%Bias FLV):

$$\%\ \text{Bias FLV} = \frac{\sum_{l=1}^{L}[\log(QS_l) - \log(QS_L)] - \sum_{l=1}^{L}[\log(QO_l) - \log(QO_L)]}{\sum_{l=1}^{L}[\log(QO_l) - \log(QO_L)]} \text{x}100 \qquad (9)$$

Where l=1.2,....,L is the index within the set of flow values located within the low flow segment (probability of exceeding 0.7 - 1.0) of the duration curve, and L is the index for the minimum flow.

The signature measure %Bias FMM was calculated using the median value of the observed (QO_{med}) and simulated (QS_{med}) flows as an index:

$$\%\ \text{Bias FMM} = \frac{\log(QS_{med}) - \log(QO_{med})}{\log(QO_{med})} \text{x}100 \qquad (10)$$

This median is selected because it is less sensitive to a biased distribution than the mean of the flow time series.

Anexo B

FUSE Formulation

The following tables describe the equations of state and fluxes parameterizations used for the construction of the model structures (Clark et al., 2008):

Table B. 1: State Variables (Clark et al., 2008).

Variable	Descripción	Unidad
S_1	Total water content in the upper soil layer	mm
S_1^T	Tension water content in the upper soil layer	mm
S_1^{TA}	Primary tension water content in the upper soil layer	mm
S_1^{TB}	Secondary tension eater content in the upper soil layer	mm
S_1^F	Free water content in the upper soil layer	mm
S_2	Total water content in the lower soil layer	mm
S_2^T	Tension water content in the lower soil layer	mm
S_2^{FA}	Free water content in the primary base flow reservoir	mm
S_2^{FB}	Free water content in the secondary base flow reservoir	mm

Table B. 2: Model Fluxes (Clark et al., 2008).

Variable	Descripción	Unidad
p	Precipitation	mm d^{-1}
pet	Potential evapotranspiration	mm d^{-1}
e_1	Evaporation from the upper soil layer	mm d^{-1}
e_2	Evaporation from the lower soil layer	mm d^{-1}
e_1^A	Evaporation from the primary tension store	mm d^{-1}
e_1^B	Evaporation from the secondary tension store	mm d^{-1}
q_{sx}	Surface runoff	mm d^{-1}
q_{12}	Percolation of water from the upper to the lower layer	mm d^{-1}
q_{if}	Interflow	mm d^{-1}
q_b	Base flow	mm d^{-1}
q_b^A	Base flow from the primary reservoir	mm d^{-1}
q_b^B	Base flow from the secondary reservoir	mm d^{-1}
q_{urof}	Overflow of water from the primary tension store in the upper soil layer	mm d^{-1}
q_{utof}	Overflow of water from tension storage in the upper soil layer	mm d^{-1}
q_{ufof}	Overflow of water from free storage in the upper soil layer	mm d^{-1}
q_{stof}	Overflow of water from tension storage in the lower soil layer	mm d^{-1}
q_{sfof}	Overflow of water from free storage in the lower soil layer	mm d^{-1}
q_{sfofa}	Overflow of water from primary base flow storage in the lower soil layer	mm d^{-1}
q_{sfofb}	Overflow of water from secondary base flow storage in the lower soil layer	mm d^{-1}

Table B. 3: Formulation of FUSE model decision options (Clark et al., 2008).

	Sacramento	**PRMS**	**TOPMODEL**	**ARNO/VIC**
U	$\frac{dS_1^T}{dt} = (p - q_{sx}) - e_1 - q_{utof}$ $\frac{dS_1^F}{dt} = q_{utof} - q_{12} - q_{if} - q_{ufof}$	$\frac{dS_1^{TA}}{dt} = (p - q_{sx}) - e_1^A - q_{urof}$ $\frac{dS_1^{TB}}{dt} = q_{urof} - e_1^B - q_{utof}$ $\frac{dS_1^F}{dt} = q_{utof} - q_{12} - q_{if} - q_{ufof}$	$\frac{dS_1}{dt} = (p - q_{sx}) - e_1 - q_{12} - q_{if} - q_{ufof}$	
L	$\frac{dS_1^{TA}}{dt} = (p - q_{sx}) - e_1^A - q_{urof}$ $\frac{dS_1^{TB}}{dt} = q_{urof} - e_1^B - q_{utof}$ $\frac{dS_1^F}{dt} = q_{utof} - q_{12} - q_{if} - q_{ufof}$	$\frac{dS_2}{dt} = q_{12} - q_b$	$\frac{dS_2}{dt} = q_{12} - q_b$	$\frac{dS_2}{dt} = q_{12} - e_2 - q_b - q_{sfof}$
SR	$q_{sx} = A_c\, p$ $A_c = 1 - \left(1 - \frac{S_1}{S_{1,max}}\right)^b$	$q_{sx} = A_c\, p$ $A_c = \frac{S_1^T}{S_{1,max}^T} A_{c,max}$	$q_{sx} = A_c\, p$ $A_c = \int_{\zeta^{crit}}^{oo} f(\zeta)d\zeta$	
PE	$q_{12} = q_0 d_{lz} \left(\frac{S_1^F}{S_{1,max}^F}\right)$ $d_{lz} = 1 - \alpha \left(\frac{S_2^F}{S_{2,max}^F}\right)^{\psi}$	$q_{12} = k_u \left(\frac{S_1^F}{S_{1,max}^F}\right)^c$		$q_{12} = k_u \left(\frac{S_1}{S_{1,max}}\right)^c$