UNIVERSIDAD DE CANTABRIA

E.T.S. INGENIEROS DE CAMINOS, CANALES Y PUERTOS

Dpto. de Ciencias y Técnicas del Agua y del Medio Ambiente



DOCTORAL THESIS

Desarrollo de metodologías para desagregar, cuantificar y reducir la incertidumbre en las predicciones en cuencas sin aforos (PUBs)

Development of Methodologies to Disaggregate, Quantify and Reduce the Uncertainty in Predictions in Ungauged Basins (PUBs)

Author: CRISTINA PRIETO SIERRA

PhD Thesis Supervisors: Dr. Raúl Medina Santamaría Dr. Nataliya Le Vine Dr. Eduardo García Alonso

Environmental Hydraulics Institute "IHCantabria"

Santander, Spain. March, 2017

To my Daddy - my ALL, my beloved Mom (Ana), Sister (Tati), 'Brother' (Flavio) and Niece (Dani)

ACKNOWLEDGEMENTS

Many people have helped me making these almost three and a half years of PhD studies a transcendent experience. First of all, I would like to express my gratitude to my thesis directors. My deepest thanks goes to Full Professor and Director of the Environmental Hydraulics Institute "IH Cantabria" Dr. Eng. **Raúl Medina** for giving me the opportunity to carry out this challenge and embarking with me on this exciting adventure. He provided an inestimable amount of professional and personal support, guide, and helped me to remain calm at critical moments. Raúl, I have no words to express my gratitude. I would also like to express my thanks to Dr. **Nataliya Le Vine** for being the scientific "alma mater" of this thesis. I was tremendously fortunate to have you as my tutor at Imperial College (London) and then as a thesis director. Thank you for opening my eyes to the amazing Bayesian world, providing guidance and training, and supervising all stages of this thesis project. I would also like to thank Dr. Eng. **Eduardo Garcia**, thesis director, colleague, and friend: Thank you for always being there for me; encouraging my work, even when not in your area of expertise; providing leadership during the early stages of this project; and offering key insights into hydrology.

I would also like to express my gratitude for my colleagues and friends who also contributed to this project. A special thanks goes to Professor and Head of the Hydraulic Engineering group of the Environmental Hydraulics Institute "IH Cantabria" Dr. Eng. César Álvarez for helping me finish this project in a peaceful manner, believing and trusting in me, and providing the necessary resources. My sincere thanks also go to Professor and Head of the Water and Environmental Research Group of Bristol Dr. Eng. Thorsten Wagener for providing excellent guidance and advice, allowing me to collaborate with him and his team, and always making time for me. I would also like to thanks Dr. Susana Almeida and Dr. Chris Hutton for great discusions about methodical issues. Moreover, thanks to the Distinguished Professor Dr. Keith Beven (Lancaster University, U.K) for sparking insightful discussions and debates about uncertainty in the Uncertainty in Environmental Modelling course (Uppsala, Sweden), thank you Keith also for discussing such aspects with me and for offering me the opportunity to work with you. In addition, I would like to express my gratitude to Professor Dr. Dimitri Kavetski (Adelaida University, Australia), for being one of the "alma maters" of this thesis, offering me "lifechanging" courses at LIST (Luxembourg Institute of Science and Technology) and EAWAG (Swiss Federal Institute of Aquatic Science and Technology) and providing hospitality together with Dr. Fabrizio Fenizia (leader of the hydrological modelling group at EAWAG) and the Dr. Benjamin Renard (IRSTEA, National Research Institute of Science and Technology for Environmental and

Agriculture, France). Equally, thank to Fabrizio for providing relevant hydrological modelling knowledge and sharing research with me, and thank you Ben for your important comments on priors and quick responses. My conversations about uncertainty and statistics with the Professor Dr. Peter Reichert (Head of the SIMS group at EAWAG), and Dr. Carlo Albert and Dr. Andreas Scheidegger (from EAWAG too) were also high benefitial. Next, I would like to thank Dr. Claudia Vitolo (ECMWF, European Centre of Medium Range Weather Forecast, Reding) for valuable contributing to this project by running FUSE. As well as to Dr. Nans Addor and Dr. Martyn Clark from the National Center of Atmospheric Researc (NCAR, Boulder, U.S) for this collaboration with NCAR and our great biweekly Skype discussions about some of the topics presented in this thesis. Professor Dr. Jasper Vrugt (Irvine University, California) also made key contributions to this project in the form of prompt answers and insights. Thank you very much Jasper. I would also like to thank Professor Dr. Hoshin Gupta (Tucson University, Arizona) for sparking key thoughts about information content. Furthermore, I would like to thank Dr. Manuel Herrera (Bath University, U.K.) for being always encouraging, willing to help, and full of statistical knowledge and Professor Dr. Han Dawei (Bristol, U.K.) for the practical advices, cheerful attitude, and hospitality.

Finally, I would like to thank **IHCantabria**, which been more to me than simply an institution and workplace. Since I joined this "great family" in 2008, it has been "my home" as well.

I would also like to thank my family, and especially my mother Ana for her support and encouragement that made this project possible. Shortly after I began this project, the worst thing that could possibly happen to me occurred: I lost my **father (Chema)**, with whom I shared a very strong bond. **Mom (Ana)**, you have not only supported me in every way but also endlessly encouraged me, even if the midst of your own suffering. I would also like to thank my **sister Tati** for her work to ensure I finished this project. My **brother in law Flavio** has likewise been a true brother to me, always encouraging. Finally, my **niece Dani** has changed my life for the better. Also, without my father, this project would not have been possible. He and my mom worked immensely hard so that I could complete this project, and this thesis is dedicated to them. Daddy, you will always be with me.

ABSTRACT

A catchment is a complex system where a multitude of interrelated energy, water and vegetation processes occur at different temporal and spatial scales. A rainfall-runoff model is a simplified representation of the system, and serves as a hypothesis about catchment inner working. In predictions for ungauged basins, a common practice is to use a pre-selected model structure for a catchment, while there is usually no justification for its suitability (due to the lack of observed flows).

The overall objective of this thesis is to advance flow prediction capabilities for ungauged basins by developing methodologies for identifying, quantifying, and reducing the uncertainty associated with the various related sources of error.

This thesis moves towards overcoming some of the major problems in hydrological modelling and focuses in the 'one size fits all' problem for predictions in ungagued basins: 1) a traditional assumption that a pre-selected model (which includes structure, inputs and parameter selection) is perfect; 2) an assumption that a pre-selected form of regionalization is suitable to represent behavior of an ungauged catchment; and 3) a way to identify dominant hydrological mechanisms to be represented in a system's model.

First, new developments in selection of hydrological indices (based on Principal Components Analysis) and regression methods (Random Forests) are incorporated into the hydrological regionalization procedure that provides information constraints in a Bayesian approach. Second, two metrics are proposed to assess suitability and adequacy of a selected model based on a) how well the model reproduces regionalized information, b) knowledge gain from considering the model over what is known from regionalization alone. Third, dominant hydrological mechanisms (to be included into a model) are identified using the regionalized information via Bayesian approach. And fourth, the research investigates sensitivity of the mechanisms identification to: (1) regionalization quality, (2) model ensemble used, and (3) the information content.

The methodological developments are applied to basins in northern Spain with varied hydroclimatological regimes. The results show that prediction quality is sensitive to model (or ensemble) error, quality of regionalized information, and available information content. [Prieto et al., 2017].

V

CONTENTS

ACKNOWLEDGEMENTS

ABSTRACT		
LISTA DE FIGURA	S DEL RESUMEN EN CASTELLANO	IX
LIST OF FIGURES		IX
LISTA DE TABLAS	DEL RESUMEN EN CASTELLANO	XII
LIST OF TABLES		XIII

3
•

CHAPTER 1: INTRODUCTION AND BACKGROUND	79
1.1. CONTEXT	81
1.1.1. PREDICTIONS IN UNGAUGED CATCHMENTS	82
1.1.2. REGIONALIZATION METHODS FOR PREDICTING HYDROGRAPHS IN UNGAUGED BASINS	83
1.1.3. SOURCES OF UNCERTAINTY IN PUBS	85
1.1.4. DESCRIBING UNCERTAIN KNOWLEDGE	89
1.1.5. BAYESIAN UNCERTAINTY QUANTIFICATION	96
1.1.6. BAYESIAN HYPOTHESIS TESTING (MODEL COMPARISON)	100
1.2 AIMS AND OBJECTIVES	101
1.3 Thesis Outline	103
1.4 REFERENCES IN CHAPTER 1	105

CHAPTER 2: STATISTICAL METRICS TO ASSESS REGIONALIZATION AND MODEL ADEQUACY IN

JBS11
ABSTRACT
2.1 INTRODUCTION
2.2 CATCHMENTS AND DATA DESCRIPTION
2.3 METHOD
2.3.1 Hydrological indices and catchment descriptors
2.3.2 REGIONALIZATION USING RANDOM FORESTS
2.3.3 Posterior distribution
2.3.4 Suitability test
2.3.5 Adequacy metrics
2.3.6 Case studies
2.3.7 Prediction quality assessment
2.4 MODEL DESCRIPTION AND CONDITIONING
2.5 RESULTS AND DISCUSSION
2.5.1 SELECTION OF THE PRINCIPAL COMPONENTS
2.5.2 REGIONALIZATION AND LIKELIHOOD FUNCTION
2.5.3 Prediction quality assessment
2.6 CONCLUSIONS
2.7 APPENDIX A
2.8 APPENDIX B
2.9 REFERENCES IN CHAPTER 2

CHAPTER 3: IDENTIFICATION OF DOMINANT HYDROLOGICAL MECHANISMS FOR UNGAUGED	
BASINS VIA BAYESIAN APPROACH	151
Abstract	153
3.1. INTRODUCTION	153
3.2 CATCHMENTS AND DATA DESCRIPTION	155
3.3 Hydrologic Models	159
3.4 Method	
3.4.1 HYDROLOGICAL INDICES AND CATCHMENT DESCRIPTORS	
3.4.2 REGIONALIZATION BY RANDOM FORESTS	
3.4.3 POSTERIOR DISTRIBUTION OF MECHANISMS	
3.4.4 IDENTIFICATION OF DOMINANT MECHANISMS	
3.4.5 SENSITIVITY ANALYSIS: CASES	164
3.5 RESULTS AND DISCUSSION	
3.5.1 REGIONALIZATION AND LIKELIHOOD DEFINITIONS	165
3.5.2 PREDICTION QUALITY	
3.6 CONCLUSIONS	171
3.7 REFERENCES IN CHAPTER 3.	

CHAPTER 4: CONTRIBUTIONS, CONCLUSIONS AND FUTURE RESEARCH	. 177
4.1 CONTRIBUTIONS	. 179
4.2 CONCLUSIONS	. 180
4.2.1 Statistical Metrics To Assess Regionalization And Model Adequacy In Pubs	. 180
4.2.2 Identification OF Dominant Hydrological Mechanisms For Ungauged Basing	s Via
Bayesian Approach	. 181
4.3 FUTURE RESEARCH	. 182
4.3.1 DISINFORMATION DUE TO MISINTERPRETATION OF DATA	. 182
4.3.2 INFORMATION CONTENT	. 190
4.3.3 Model structure uncertainty	. 193
A) Model Selection And Combination	. 193
B) IMPACT OF THE PRIOR DISTRIBUTION ON HYPOTHESIS SELECTION	. 198
C) Model Structure Error	. 201
4.3.4 Preliminary conclusions	. 202
4.4 REFERENCES IN THE CHAPTER 4	. 204

PUBLISHED WORK IN CONFERENCE PROCEEDINGS

CERTIFICATES:

RESEARCH STAYS AND COLLABORATIONS WITH INTERNATIONAL RESEARCH INSTITUTIONS INTERNATIONAL EXPERT REVIEWERS OF THE PHD THESIS

FIGURA 1: CASOS DE ESTUDIO Y CLIMATOLOGÍA
Figura 2: Regionalización Mediante Bosques Aleatorios
Figura 3: Distribuciones Marginales E Histogramas Empíricos (Emp.) De Los Residuos De La Regionalización
FIGURA 4: MÉTRICAS DE APTITUD Y ADECUABILIDAD: (A) P-VALORES Y (B) BFS
FIGURA 5: A) NSPROB PARA LOS CASOS BC Y R, B) NSPROB PARA LOS CASOS M Y MR, C) 95%CI PARA LOS CASOS BC Y R, Y D) 95%CI PARA LOS CASOS M Y MR
Figura 6: Fuse
FIGURA 7: CONTENIDO DE INFORMACIÓN EN LOS PRIMEROS CUATRO PCS PARA LOS CAUDALES OBSERVADOS Y Para Los Caudales Sintéticos
FIGURA 8: MECANISMOS IDENTIFICADOS (A) Y CORRECTAMENTE IDENTIFICADOS (B) PARA CADA CASO 46
FIGURA 9: CASO BC: MECANISMOS IDENTIFICADOS PARA LOS PROCESOS A) ENRUTAMIENTO, B) EVAPORACIÓN, C) FLUJO LATERAL, D) ESCORRENTÍA Y E) PROCESOS EN LA CAPA SUPERIOR DEL SUELO 47
FIGURA 10: CASO MRI: MECANISMOS IDENTIFICADOS PARA LOS PROCESOS A) ENRUTAMIENTO, B) EVAPORACIÓN, C) FLUJO LATERAL, D) ESCORRENTÍA, E) PROCESOS EN LA CAPA INFERIOR DEL SUELO, F) INFILTRACIÓN Y G) PROCESOS EN LA CAPA SUPERIOR DEL SUELO
Figura 11: Elevación De Los Pluviómetros En La Función De Distribución Acumulada Del Modelo Digital De Elevación De La Cuenca56
Figura 12: Elevación De Los Termómetros En La Función De Distribución Acumulada Del Modelo Digital De Elevación De La Cuenca57
FIGURA 13: GRÁFICO DE COORDENADAS PARALELAS DE LOS DESCRIPTORES DE LAS CUENCAS. EJE X: DESCRIPTORES DE CUENCA. EJE Y: VALOR DE LOS DESCRIPTORES DE CUENCA
Figura 14: Gráfico De Coordenadas Paralelas De Los Índices Hidrológicos De Las Cuencas. Eje X: Índices Hidrológicos. Eje Y: Valor De Los Índices Hidrológicos
FIGURA 15: ESPACIO DE LOS DATOS. EJEMPLO DEL COEFICIENTE DE ESCORRENTÍA58
FIGURA 16: ESPACIO DE LOS PARÁMETROS. DYNIA. TIPO DE ERROR: EVENTOS DE CAUDAL NO REGISTRADOS.
Figura 17: Sensibilidad Del Proceso De Enrutamiento A La Información De Cada Pc Individual. Ejemplo De Una Cuenca Dónde Se Identifica Un Mecanismo Dominante

LIST OF FIGURES

FIGURE 1: CASE STUDIES
FIGURE 2: PDM CONCEPTUAL MODEL
FIGURE 3: MARGINAL DISTRIBUTIONS AND EMPIRICAL HISTOGRAMS (EMP) FOR REGIONALIZED RESIDUALS FOR A SELECTED CATCHMENT
FIGURE 4: REGIONALIZED RESIDUALS FOR PC2 AND PC3 FOR A SELECTED CATCHMENT
FIGURE 5: ADEQUACY AND SUITABILITY METRICS: (A) P-VALUES AND (B) BFS
FIGURE 6: A) NS ^{prob} For Cases BC And R, B) NS ^{prob} For Cases M And Mr, C) 95%CI For Cases BC And R, And D) 95% CI For M And MR
FIGURE 7: BASIN C8z1
FIGURE 8: BASIN X1353
FIGURE 9: BASIN AN439
FIGURE 10: BASIN X9040
FIGURE 11: THE SELECTED SET OF 16 BASINS
FIGURE 12: FUSE
FIGURE 13: MARGINAL DISTRIBUTIONS AND EMPIRICAL HISTOGRAMS (EMP) FOR REGIONALIZED RESIDUALS FOR SELECTED CATCHMENT
FIGURE 14: REGIONALIZED RESIDUALS FOR PC2 AND PC3 FOR A SELECTED CATCHMENT
FIGURE 15: INFORMATION CONTENT IN THE FIRST FOUR PCs FOR THE OBSERVED AND SYNTHETIC FLOWS
FIGURE 16: IDENTIFIED DOMINANT MECHANISMS (A) AND CORRECTLY IDENTIFIED DOMINANT MECHANISMS (B) FOR EACH CASE
FIGURE 17: CASE BC - IDENTIFIED MECHANISMS FOR THE PROCESSES: A) ROUTING, B) EVAPORATION, C) INTERFLOW, D) RUNOFF AND E) PROCESS IN THE LOWER SOIL
FIGURE 18: CASE MRI - IDENTIFIED AND CORRECTLY IDENTIFIED MECHANISMS FOR THE PROCESSES: A) ROUTING, B) EVAPORATION, C) INTERFLOW, D) RUNOFF, E) PROCESS IN THE LOWER SOIL, F) PERCOLATION AND G) PROCESS IN THE UPPER SOIL
FIGURE 19: ELEVATION OF THE RAINGAUGES IN THE CUMULATIVE DISTRIBUTION FUNCTION (CDF) OF THE DIGITAL ELEVATION MODEL OF THE CATCHMENTS
FIGURE 20: ELEVATION OF THE THERMOMETERS IN THE CUMULATIVE DISTRIBUTION FUNCTION (CDF) OF THE DIGITAL ELEVATION MODEL OF THE CATCHMENTS

FIGURE 21: PARALLEL COORDINATE PLOTS OF CATCHMENT DESCRIPTORS FOR CATCHMENTS IN NORTHERN SPAIN
Figure 22: Parallel Coordinate Plots Of Hydrological Indices For Catchments In Northern Spain
FIGURE 23: DATA SPACE: EXAMPLE OF THE EVENT-BASED RAINFALL RUNOFF COEFFICIENT FOR A CONSTANT BIAS IN THE FLOW SERIES
FIGURE 24: PARAMETER SPACE: EXAMPLE OF DYNIA [WAGENER, 2003] FOR MISSING EVENTS OF FLOW
Figure 25: Sensitivity Of The Routing Process To The Information In The Individual PCs. A Dominant Mechanism Is Identified
Figure 26: Sensitivity Of The Routing Process To The Information In The Individual PCs. A Dominant Mechanism Is Not Identified
FIGURE 27: 95% CI FOR CASES A) BC, B) R, C) M AND D) MR196
FIGURE 28: PRECISION FOR CASES A) BC, B) R, C) M AND D) MR
FIGURE 29: NS ^{prob} For Cases A) BC, B) R, C) M AND D) MR197
FIGURE 30: 95% CI FOR A HIGH-PERFORMING CATCHMENT FOR CASES BC, R, M AND MR
FIGURE 31: 95% CI FOR A POORLY PERFORMING CATCHMENT FOR CASES BC, R, M AND MR

LISTA DE TABLAS DEL RESUMEN EN CASTELLANO

TABLA 1: CARACTERÍSTICAS DE LAS CUENCAS	. 24
TABLA 2: ESCALA PARA INTERPRETAR EL FACTOR DE BAYES [JEFFREYS, 1961]	. 29
TABLA 3: CALIDAD MEDIA DE LA REGIONALIZACIÓN PARA CADA PC	. 31
TABLA 4: MECANISMOS INCLUIDOS EN LAS 624 ESTRUCTURAS	. 41
TABLA 5: TIPOS DE ERRORES INTRODUCIDOS EN CADA UNA DE LAS CUENCAS.	. 55

LIST OF TABLES

TABLE 1: CATCHMENT CHARACTERISTICS 1	26
TABLE 2: SCALE FOR BAYES' FACTOR INTERPRETATION [JEFFREYS, 1961]	131
TABLE 3: PARAMETER RANGES1	.34
TABLE 4: REGIONALIZATION QUALITY FOR INDIVIDUAL PCS FOR A SELECTED CATCHMENT	134
TABLE 5: CATCHMENT CHARACTERISTICS 1	.58
TABLE 6: PROCESSES AND MECHANISMS IN THE 624 FUSE MODELS 1	160
TABLE 7: REGIONALIZATION QUALITY FOR INDIVIDUAL PCS FOR A SELECTED CATCHMENT	166
TABLE 8: INTRODUCED ERRORS FOR ONE OF THE CATCHMENTS	183

Everything is Uncertain [1] and Everything Changes [2]

[1]"I think it's much more interesting to live not knowing than to have answers which might be wrong. I have approximate answers and possible beliefs and different degrees of certainty about different things. But I'm not absolutely sure of anything, and there are many things I don't know anything about, such as whether it means anything to ask why we're here, and what the question might mean. I might think about it a little bit; if I can't figure it out, then I go onto something else. But I don't have to know an answer. I don't feel frightened by not knowing things, by being lost in the mysterious universe without having any purpose, which is the way it really is, as far as I can tell. It doesn't frighten me. [Richard Feynman]"

[2]"Everything changes and nothing remains still [Heraclito, 550-480 a.c.]"

Resumen en castellano

De acuerdo con la normativa de estudios de doctorado de la Universidad de Cantabria en relación a los requerimientos exigidos para aquellas tesis redactadas en un idioma diferente al español, aprobada por Junta de Gobierno de 12 de marzo de 1999 y actualizada a 18 de diciembre de 2013, a continuación se presenta un resumen suficientemente extenso del documento original redactado en inglés.

1.1 CONTEXTO

El agua ha sido elemento crítico para el desarrollo de la vida y el sistema socio-económico desde el origen de las civilizaciones. El cada vez más escaso recurso hidrológico condiciona aspectos de la vida cotidiana como el abastecimiento, la producción hidroeléctrica, la agricultura, el turismo, el comercio y los ecosistemas naturales. Además, los fenómenos hidrológicos extremos, como sequías e inundaciones, originan la pérdida de vidas y bienes.

La gestión eficiente del agua y, por tanto, la economía basada en ella, dependen en gran medida de nuestra capacidad de hacer predicciones fiables del recurso hídrico.

Desde un punto de vista científico, las predicciones de las variables hidrológicas son importantes para entender el funcionamiento de las cuencas y los distintos procesos que originan el recurso. Desde un punto de vista operacional, las predicciones de las variables hidrológicas se usan, por ejemplo, para diseñar estructuras hidráulicas; hacer planificación hidrológica y establecer prioridades frente a las demandas de agua (e.j. agrícolas, industriales, domésticas, ambientales, recreativas, etc); calcular inundaciones y desarrollar sistemas de alerta temprana (cuando se hacen predicciones en tiempo real); gestionar embalses para operaciones hidroeléctricas; y hacer gestión integrada de cuencas incluyendo cómo los cambios en el medio ambiente (e.j. cambio en el clima o en los usos del suelo) afectan al régimen de caudales [Almeida, 2014].

Tales predicciones se obtienen mediante el uso de modelos hidrológicos. Estos modelos tratan de representar de forma muy simplificada los procesos que ocurren en la cuenca. Intentando reproducir cómo estos sistemas complejos reparten, transmiten, almacenan y liberan el agua [Wagener et al., 2007]. La cuenca es un sistema abierto (con condiciones de contorno mal conocidas) donde una multitud de procesos (en su mayoría no lineales) interconectados ocurren a distintas escalas temporales y espaciales [Beven, 2009]. Un modelo nunca podrá reproducir completamente los procesos que ocurren en la naturaleza [Westerberg, 2011]. Sin embargo, pese a que todos son una representación imperfecta de la realidad, algunos son útiles [Box, 1979] como herramienta para extrapolar caudales en el espacio y tiempo.

El reto del modelado para la predicción de caudales comprende tanto las cuencas con aforos como las cuencas sin instrumentar. En cuencas con suficientes datos históricos de caudales (calidad y cantidad) los parámetros del modelo pueden calibrarse por medio de técnicas de calibración habituales [Gupta et al., 2005]. En cambio, en aquellas sin datos es necesario usar métodos alternativos para suplir la ausencia de observaciones. Este es el caso de la gran mayoría de las cuencas del planeta [Almeida, 2014]. Ya sea porque tienen pocos datos (calidad y/o cantidad) para la escala espacial y/o temporal de interés [Sivapalan et al., 2003], o porque no

cuentan con ninguna medida [Blosch et al., 2013]. Además en muchas cuencas en las que se disponía de aforos, éstos están dejando de medir por las restricciones derivadas de la crisis económica global. Por ejemplo, aplicando los criterios requeridos para generar la base de datos del Experimento de Estimación de Parámetros de Modelos (MOPEX¹, por sus siglas en inglés) en España, de unas 1000 cuencas naturales, sólo en aproximadamente 100 se cumple el requisito de número mínimo de pluviómetros y más de 10 años de series diarias sincronizados de precipitación, caudal, temperatura y evapotranspiración potencial [Schaake et al., 2000; 2006]. La cantidad de datos es una condición necesaria pero no suficiente pues falta evaluar su calidad, por lo que 100 cuencas es el número máximo de 'cuencas aforadas' en España, cumpliendo los criterios del MOPEX.

La predicción fiable de caudales en cuencas sin aforos (en adelante **PUBs**) fue el leitmotiv de la Asociación Internacional de Ciencias Hidrológicas (IAHS por sus siglas en inglés) en la década 2003-2012. Para lograr este objetivo era prioritario reducir la incertidumbre en las predicciones y hacer un uso más eficiente de la información disponible. A pesar de los grandes avances realizados, la fiabilidad de las predicciones proporcionadas por los modelos hidrológicos es todavía muy pobre y PUBs es aún uno de los mayores retos del paradigma hidrológico [Hrachowitz et al., 2013].

Esta tesis está orientada a desarrollar metodologías que permitan obtener predicciones más fiables, exactas y precisas en las cuencas sin aforos. Para ello, a continuación se expone una revisión del estado del Arte y se plantean los objetivos de esta tesis (Apartado 2).

1.2 PREDICCIÓN EN CUENCAS SIN AFOROS

Una *cuenca sin aforos* es aquella donde bien no hay datos de caudal, o bien estos son escasos y/o no son adecuados (calidad y/o cantidad), para el cálculo de las variables de interés (en esta tesis caudales) a la escala (espacial y/o temporal) apropiada, y/o con la precisión requerida para aplicaciones prácticas. *El objetivo de PUBs* es obtener los caudales y su incertidumbre asociada en dichas cuencas, partiendo de los datos de clima, suelo, vegetación, geología y topografía, incluyendo cambios futuros en el clima y usos del suelo [Sivapalan et al., 2003]. PUBs incluye tanto la extrapolación espacial de lo inferido en cuencas con datos a cuencas sin medidas, como la extrapolación temporal de lo observado en periodos ricos en datos a periodos sin medidas. Esta transferencia de conocimiento en espacio y/o tiempo, es llamada *regionalización*. La

¹ http://www.nws.noaa.gov/oh/mopex/index.html

regionalización se basa en la hipótesis de que cuencas con similares características climáticas y fisiográficas tendrán similares regímenes hidrológicos.

Por supuesto, el problema planteado en PUBs no es nuevo (sus orígenes se remontan al siglo XIX con el método racional [Chow et al., 1988]), si bien en 2003 adquiere un nuevo enfoque cuando la IAHS lanza la década de PUBs. PUBs era una iniciativa conjunta y coordinada con el objetivo global de hacer predicciones fiables en cuencas sin aforos.

Para ello, PUBs se ha centrado en la cuantificación de la incertidumbre total en las predicciones, tratando de desagregar el impacto de cada una de las fuentes de error y desarrollando metodologías para reducirla, mediante un mejor uso de la información disponible [Sivapalan et al., 2003; Hrachowitz et al., 2014].

Además, PUBs adoptó la visión de la hidrología comparativa [Wagener et al., 2007] para entender y conocer las relaciones entre el comportamiento de las cuencas y sus atributos [e.j. Singh et al., 2014], permitiendo una clasificación basada en el funcionamiento de las cuencas [Wagener et al., 2007] y ayudando a identificar los principios de organización de las mismas. Las cuencas son sistemas complejos donde clima, suelo, topografía y seres humanos evolucionan conjuntamente, por lo que es necesario ampliar el concepto de ciclo hidrológico a un enfoque que refleje el sistema real [Gupta et al., 2008]. La aproximación comparativa implica el desarrollo de nuevas capacidades en hidrología: leer, interpretar y aprender de los patrones, realizar estudios de casos, sustituir la variable espacio por tiempo y modelar las interacciones de los procesos. 'Sobre todo, debe de ser educada la capacidad de análisis y síntesis de la nueva generación de hidrólogos' [Wagener, 2010].

Para valorar el éxito de las predicciones y poder contrastar los métodos desarrollados por distintos grupos, en PUBs se adoptaron una serie de métricas para evaluar la incertidumbre predictiva. Sin embargo, a pesar de todos los avances realizados en la década PUBs (la mayoría en cuencas con aforos), la prediccción en cuencas sin aforos sigue siendo uno de los grandes desafíos de la hidrología [Hrachowitz et al., 2013].

1.3 MÉTODOS DE REGIONALIZACIÓN EN PUBS

Tradicionalmente, los métodos de regionalización para el cálculo de hidrogramas en cuencas no aforadas tenían por objeto transferir los parámetros de un modelo hidrológico calibrado en cuencas con aforos a las cuencas sin medidas [Wagener et al. 2004]. En la cuenca sin medidas, estos parámetros eran usados junto con los datos climáticos para obtener las predicciones de caudal [Bardossy, 2007]. Sin embargo, los métodos basados en transferencia de parámetros

tienen varios inconvenientes, por ejemplo, dependen del modelo hidrológico usado, tienen el problema de la identificabilidad de parámetros y no tienen en cuenta sus interdependencias [McIntyre et al., 2005].

Recientemente, este enfoque basado en parámetros ha sido sustituido por un enfoque que busca predecir el comportamiento dinámico de las cuencas en base a sus características [Wagener and Montanari, 2011]. Dicho comportamiento queda reflejado en una serie de índices hidrológicos [por ejemplo, Yadav et al., 2007; Bulygina et al., 2009, 2011; Sawicz et al., 2011; Alemida, 2014; Almeida et al., 2016] que son calculados en las cuencas con aforos a partir de las observaciones. Una vez transferidos los índices a las cuencas sin aforos, estos se usan para condicionar las predicciones dadas por un modelo hidrológico (o un conjunto de ellos).

Tanto si se transfieren parámetros como si se transfieren índices, los métodos de regionalización pueden dividirse en 1) aquellos que usan algún tipo de distancia (ya sea geográfica –proximidad espacial, o funcional) y 2) aquellos que usan regresión [He et al., 2011].

Los métodos que usan proximidad espacial aplican la hipótesis de que las cuencas cercanas son similares en cuanto a características físicas, hidrológicas y climáticas. Sin embargo, debido a la singularidad del lugar, cuencas relativamente cercanas espacialmente pueden tener comportamientos muy distintos [Beven, 2000]. Los métodos basados en similitud funcional asumen que las cuencas con descriptores similares se comportarán de forma parecida. Sin embargo, no hay unanimidad acerca de qué medida usar para definir similitud, ni sobre qué descriptores de cuenca incluir. Asimismo, la revisión dada por Almeida [2014] concluye que en lugares con muchos aforos el método basado en distancia espacial tiende a dar mejores resultados al ser las distancias de interpolación pequeñas.

En cuanto a los métodos basados en regresión, aquellos que transfieren parámetros siguen el siguiente esquema: 1) el modelo hidrológico es seleccionado y calibrado en las cuencas con medidas; 2) el conjunto de parámetros con mejor comportamiento, respecto a un criterio de ajuste, es retenido y relacionado con los descriptores de las cuencas mediante modelos estadísticos; 3) los parámetros son estimados en las cuencas sin medidas, aplicando el modelo de regresión con los descriptores de estas cuencas; y 4) introduciendo los parámetros en el modelo hidrológico los caudales son predecidos. A pesar de que la regionalización de parámetros usando métodos de regresión fue muy utilizada en el pasado para PUBs [por ejemplo, Bulygina et al., 2012] ya que, además de los problemas de los métodos basados en parámetros mencionados [McIntyre et al., 2005], las relaciones entre los parámetros de un modelo hidrológico específico - que es una simplificación de la realidad); y las características de las cuencas no están claras - si es que estas relaciones pudieran llegar a estarlo.

En el caso de regresión de índices hidrológicos, la metodología puede resumirse en los siguientes pasos: 1) cálculo de los índices hidrológicos en las cuencas con medidas a partir de los datos observados de caudal; 2) establecer modelos de regresión entre estos índices y los descriptores de las cuencas; 3) obtención de los índices en la cuenca sin medidas, aplicando la ecuación de regresión con los descriptores de dicha cuenca; y 4) usar estos índices para condicionar las predicciones generadas por un modelo hidrológico (o un conjunto de ellos).

Este enfoque que combina la información a priori del modelo con la información acerca del comportamiento dinámico de la cuenca ha demostrado ser preferible y ventajosa tanto para PUBs como para avanzar en nuestro conocimiento acerca del funcionamiento del sistema [por ejemplo, Wagener and Montanari, 2011; Yadav et al., 2007; Zhang et al., 2008; Almeida et al., 2013; 2016; Bulygina et al., 2009; 2011; 2012; Le Vine, 2016].

1.4 FUENTES DE INCERTIDUMBRE EN PUBS

Hoy en día es ampliamente admitido que sólo se pueden hacer predicciones fiables cuando todas las fuentes de incertidumbre han sido tenidas en cuenta. En PUBs, además de las tres fuentes de incertidumbre presentes en todo modelado, es decir, incertidumbre acerca de los datos, incertidumbre acerca de la estructura del modelo e incertidumbre acerca de los parámetros del modelo, existe una cuarta fuente debida a la regionalización.

Cada uno de estos tipos de error está afectado por la incertidumbre aleatoria y la epistémica. La primera se debe a la aleatoriedad del proceso natural en sí y es por ende irreducible; representa el desconocimiento que tenemos cada vez que repetimos un experimento; y puede ser cuantificada mediante distribuciones de probabilidad - los errores tienen una estructura que puede caracterizarse estadísticamente [Beven and Smith, 2014]. La segunda es fruto de nuestras limitaciones a la hora de representar el sistema (por ejemplo, la falta de conocimiento o de recursos - datos y tiempo). E incluye tanto aquello que sabemos que no conocemos ' known unknowns', como aquello que no sabemos que no conocemos 'unknown unknowns' [Beven and Westerberg, 2011]. Y algunos autores proponen usar métodos no probabilistas para evaluarla (ver, Beven and Binley, 2014), alegando que los residuos no siguen una estructura estadística [Beven and Smith, 2014]. El análisis de incertidumbre trata de reducir la incertidumbre epistémica a aleatoria.

La incertidumbre en los datos se debe tanto a la calidad como a la cantidad de los datos de entrada, generalmente precipitación y evapotranspiración potencial; y de los datos usados para evaluar las predicciones, habitualmente datos de caudal. Aunque la incertidumbre asociada a la precipitación suele ser la dominante [Gupta et al., 2005]; también hay un error importante en

los datos de caudal [Westerberg et al., 2011]. Algunos errores típicos encontrados en series de precipitación y caudal son una desviación constante de las medidas, tormentas y/o eventos no medidos, autocorrelación de errores, residuos heterocedásticos e inversamente heterocedásticos, errores aleatorios, desfase temporal de las medidas, etc. Un catálogo de tipos de errores y sus magnitudes se encuentra en McMillan et al. [2012]. Por ejemplo, la incertidumbre en las precipitaciones obtenidas a partir de pluviómetros se debe principalmente a errores de interpolación. Aunque, la incertidumbre debida a la escala temporal puede ser también muy importante, cuando los datos se encuentran a una escala mayor a la que ocurren los procesos [Westerberg et al., 2011]. En el caso de los caudales obtenidos mediante curvas de tarado, gran parte de incertidumbre resulta de la extrapolación fuera del rango de la curva, de errores en los medidores de caudal, condiciones de flujo no estacionario, histéresis de la relación altura-caudal y cambios en el cauce del río [Renard et al., 2011; Kuczera, 1996]. Por ejemplo, McMillan et al., [2012] indican que para caudales bajos, el intervalo de error es de ±50 - 100%, ± 10 - 20% para caudales medio-altos en las llanuras y $\pm 40\%$ para caudales fuera de las llanuras. Esto da una idea de la dificultad de extraer información de los datos de caudales. Otros autores como Kavetski et al. [2002; 2006a; 2006b; 2006c] desarrollaron una metodología llamada analisis bayesiano del error total ('BATEA', por sus siglas en inglés) al objeto de cuantificar el impacto que la incertidumbre debida a los datos tenía en las predicciones.

Sin embargo, a pesar de estar ampliamente reconocido que la incertidumbre en los datos tiene un gran impacto y de larga duración en las predicciones [Beven and Westerberg, 2011; Beven and Smith, 2014], la práctica habitual en modelado hidrológico es considerar que los datos son 'perfectos'. Esto se debe esencialmente a la dificultad que entraña detectar aquellos períodos que son desinformativos, usando únicamente los propios datos. Es decir, independientemente del modelo hidrológico empleado; ya que los errores en los datos de entrada se propagan de forma no lineal a través del modelo interactuando con la incertidumbre asociada a la propia estructura del modelo [Gupta et al., 2005]. Y cada vez más autores reclaman una estrategia que permita evaluar la calidad de los datos de observaciones, tanto los forzamientos introducidos en los modelos como los datos usados para evaluar las predicciones que estos proporcionan [Kavetski et al., 2002 and 2006a; Renard et al., 2011; Beven and Westerberg, 2011; McMillan et al., 2012; Westerberg and Birkel, 2015].

Incertidumbre en la estructura del modelo. La hidrología de una cuenca es fruto de las interacciones complejas producidas por un número de procesos de agua, energía y vegetación fuertemente interrelacionados, que ocurren a distintas escalas espaciales y temporales [Gupta et al., 2005; Wagener et al., 2007; Westerberg et al., 2011]. Cualquier modelo es necesariamente

una simplificación de la realidad. Un modelo hidrológico es una combinación de hipótesis acerca de los mecanismos que gobiernan el funcionamiento de las cuencas [Beven, 2001]; estas hipótesis son asignadas diferentes grados de credibilidad [Clark et al., 2011a; Chamberlain, 1965]; al no poder ser validadas, pues no se conoce completamente el fenómeno natural [Wagener et al., 2004]. Distintos hidrólogos entenderán los procesos y su importancia de forma diferente, dependiendo de su conocimiento y experiencia, ésta subjetividad inherente en todo modelado [Westerberg, 2011], es lo que hace a un científico un experto en su campo [Rougier, 2010]. Por lo tanto, un diagnóstico o contraste de hipótesis acerca de la estructura del modelo debería de ser llevado a cabo en cada cuenca para verificar si éste es adecuado. Es decir, si proporciona los tipos de predicciones que son necesarias para su aplicación.

Debido a la escasez de datos, en el modelado hidrológico, generalmente, es típico emplear una única estructura pre-seleccionada para representar los mecanismos dominantes en todas las cuencas ('one-size-fits-all' [Fenizia et al., 2008]), bajo el supuesto de que el 'modelo es perfecto' [Wagener and Montanary, 2011]. Esto, a pesar de lo ampliamente confirmado que está el paradigma de 'la singularidad del lugar' [Beven, 2000]; de que es imposible que una única estructura proporcione una descripción adecuada del sistema hidrológico en todas las cuencas; y de que ignorar esta incertidumbre asociada a la estructura del modelo induce tanto una subestimación de la incertidumbre predictiva, como un sesgo en las predicciones. [Reicherte and Omlin, 1997].

En este sentido, los entornos de modelado flexible y adaptable son algunos avances que se han realizado para contrastar hipótesis de forma explícita. Por ejemplo, la herramienta de modelado Precipitación-Caudal (RRMT) [Wagener et al., 2001; 2004]; el Marco para Entender los Errores en la Estructura del modelo (FUSE) [Clark et al., 2008; 2011b]; FLEX y SUPERFLEX [Fenicia et al., 2006; 2008; 2011]; y la Estructura para Unificar Múltiples Alternativas de Modelado (SUMMA) 2015a; 2015b; 2015c]. El procedimiento requiere calcular la [Clark et al., probabilidad/credibilidad de cada hipótesis/estructura de modelo en función de lo bien que reproducen el comportamiento hidrológico del sistema, que en PUBs viene dado por la información regionalizada. Con este objeto, se usan métricas agregadas basadas en las discrepancias entre las predicciones dadas por los diferentes modelos/hipótesis y las observaciones (por ejemplo, el coeficiente de Nash-Sutcliffe [Nash and Sutcliffe, 1970]). Sin embargo, estas métricas son demasiado simplistas y poco realistas [Boyle et al., 2000; Wagener et al., 2003; Wagener et al., 2007; Gupta et al., 2008; Kirchner et al., 2008] y la necesidad de llevar a cabo un contraste de hipótesis riguroso es patente [Beven, 2001]. Una vez se han ordenado y seleccionado los modelos que compiten, los métodos de predicción basados en un

conjunto de estructuras [por ejemplo, Raftery, 2003; 2005; Vrugt et al., 2007; Duan et al., 2007] intentan afrontar algunos de los inconvenientes mencionados, combinando la información dada por los miembros seleccionados (cada estructura tiene sus fortalezas y debilidades para reproducir los procesos hidrológicos de las cuencas); siendo la estrategia de ponderación un aspecto clave diferenciador de cada método. Este conjunto de modelos trata de cubrir el espacio de las hipótesis, al cual la hipótesis verdadera pertenece. En consecuencia, representan el impacto que tiene en las predicciones la incertidumbre asociada a la estructura del modelo [Clark et al., 2008]. Nótese que es fundamental que al menos una de las estructuras incluidas en el conjunto se aproxime a la 'realidad'.

Sin embargo, a pesar de los avances realizados, sigue siendo uno de los grandes retos de la hidrología tanto entender y corregir la incertidumbre asociada a la estructura del modelo, como cuantificar su adecuabilidad y desarrollar una estrategia que permita identificar aquellas más adecuadas para un objetivo dado [Clark et al., 2008 and 2011a; Wagener et al., 2001; Fenicia et al., 2008; 2011]. Finalmente, una vez definido el modelo conceptual del sistema hidrológico que se quiere representar, el error debido a su implementación matemática, es decir, la identificación de la estructura de la ecuación del sistema, su discretización temporal y espacial, el esquema numérico usado para resolverla y los errores numéricos, también repercuten en la incertidumbre asociada a la estructura del modelo [Buligina and Gupta, 2010; Kavetski and Clark, 2010].

La incertidumbre de los parámetros da lugar a que no sea identificable un óptimo global de acuerdo a un criterio de evaluación. Esta incertidumbre puede estar inducida por el error de los datos y/o la estructura del modelo, la ausencia de datos en períodos largos de tiempo, la ineficiencia del algoritmo de optimización [Almeida, 2014], el hecho de que unos parámetros dependan de otros [Kuczera and Mroczkowski, 1998]. O simplemente, la inexistencia de un único conjunto óptimo de parámetros puede apoyar la necesidad de usar aproximaciones basadas en teoría de conjuntos, dónde todos los conjuntos adecuados son retenidos, hasta que no se demuestre lo contrario. Esta 'no-identificabilidad' de parámetros [Wagener et al., 2001], conduce a inferencias mal planteadas, a menos que se introduzca conocimiento a priori [Renard et al., 2011].

La incertidumbre de los parámetros ha sido considerablemente explorada [por ejemplo, ver Gupta et al., 2005]. Las estrategias tradicionales de calibración de parámetros estaban basadas en regresión de mínimos cuadrados. Sin embargo los rígidos axiomas que deben cumplir estos métodos no son satisfechos habitualmente en modelado hidrológico [Kavetski et al., 2006]. Las

aproximaciones basadas en mínimos cuadrados asumen que los errores de la variable de salida son independientes, están normalmente distribuidos, con media cero y varianza constante. En muchos casos en hidrología los residuos están correlacionados y/o no son Gausianos y/o son heterocedásticos [Beven, 2009]. Esto llevó a abandonar los métodos de inferencia de la estadística clásica y a utilizar otras aproximaciones como las basadas en métodos de simulación y muestreo. Entre estas aproximaciones destacan en hidrología la teoría de la probabilidad bayesiana [Kuczera and Parent, 1998], que hoy en día puede usarse de forma práctica gracias a los recientes avances en los Métodos de Montecarlo usando Cadenas de Markov (MCMC) y métodos numéricos, junto con la creciente capacidad computacional de los ordenadores [Kavetski et al., 2006]; el análisis de sensibilidad regional (RSA) [Spear and Hornberger, 1980]; 'Generalized Likelihood Uncertainty Estimation' (GLUE) [Beven and Binley 1992]; o el análisis de identificabilidad dinámico (DYNIA) [Wagener et al., 2003]. Estas aproximaciones han demostrado ser ventajosas en el modelado hidrológico [Wagener and Montanari, 2011].

Estas tres fuentes de incertidumbre, presentes en todo modelado, resultan en una equifinalidad de representaciones que son consistentes con las variables observadas e inducen incertidumbre en las predicciones dadas por el modelo [Beven, 1993; 2009]. Históricamente, solo se ha abordado la incertidumbre paramétrica, ignorando el resto [Beven, 2009]. Sin embargo, generalmente la incertidumbre debida a los datos y a la estructura del modelo dominan a la incertidumbre asociada a los parámetros [Kuczera et al., 2006]; siendo el error asociado a la estructura del modelo el más difícil de caracterizar – algunos investigadores indican que la incertidumbre de los datos debería poder estimarse analizando el muestreo y los aparatos de medida [Renard et al., 2010].

La cuarta fuente de error en PUBs es aquella debida a los índices hidrológicos [Almeida., 2014; Westerberg et al., 2011]. Algunas de las causas que la generan son: 1) incertidumbre debida a los índices calculados en las cuencas con medidas, e incluye la asociada a la cantidad y calidad de los datos; 2) incertidumbre sobre que el conocimiento adquirido en las cuencas aforadas sea extrapolable a las no aforadas; 3) incertidumbre ligada al concepto de regionalización, pues asume implícitamente i) comparabilidad de medidas entre cuencas, lo que no es estrictamente verdad ya que los errores en los datos varían espacialmente [McMillan et al., 2012]; y ii) que los descriptores de las cuencas representan de forma similar cuencas con y sin medidas; 4) incertidumbre asociada al procedimiento de regionalización empleado para traspasar la información; y 5) incertidumbre acerca de los descriptores usados en el método de regionalización.

En resumen, las predicciones en cuencas sin aforos están afectadas por la incertidumbre inherente al modelado y al proceso de regionalización, y reducir la incertidumbre total en las predicciones requiere separar las distintas fuentes de error que intervienen - al objeto de poder dar un tratamiento apropiado (cuantificar y reducir) a cada una de ellas. A pesar de que diferentes estudios han reiterado la importancia de lograr esta descomposición en modelado hidrológico [Kavetski et al., 2006a; 2006b; 2006c; Thyer et al., 2009; Renard et al., 2010; Kuczera et al., 2006; Sivapalan et al. 2003]; desarrollar una metodología estadística para lograrlo sigue siendo uno de los principales desafíos de la hidrología [Renard et al., 2010]. Además, cuando se emplea la información regionalizada para condicionar las predicciones, un primer objetivo consiste en separar la incertidumbre debida a la combinación estructura del modelo/datos de entrada/parámetros muestreados, de la incertidumbre debida al procedimiento de regionalización.

1.5 CUANTIFICACIÓN DE INCERTIDUMBRE

A la hora de cuantificar la incertidumbre en las predicciones hidrológicas, dos son los tipos de problemas a resolver: 1) el llamado problema inverso, es decir, cuantificar la incertidumbre del modelo y la incertidumbre de los parámetros; o en otras palabras, dadas las observaciones, obtener la probabilidad de que hayan sido generadas por cada una de las posibles configuraciones (modelo, datos de entrada y parámetros) que rigen el fenómeno; y 2) la propagación de todas las fuentes de incertidumbre a través del modelo, para predecir la incertidumbre total en la respuesta del sistema [Moradhkani and Soroshian, 2009]. La predicción en cuencas sin aforos incluye ambos, primero los parámetros son estimados con base en la información regionalizada y, luego, las predicciones de caudal son obtenidas propagando todas las incertidumbres a través del modelo.

Todos los métodos de cuantificación de incertidumbre se basan, implícita o explícitamente, en un conjunto de hipótesis acerca de las diferencias (errores o residuos) entre las predicciones/simulaciones y las observaciones. Cuando se puede definir una función de densidad de probabilidad de estas diferencias, la estadística bayesiana ha demostrado ser la aproximación más apropiada para cuantificar la incertidumbre (epistémica y aleatoria) en las predicciones (e.g. Engeland et al., 2005; Marshall et al., 2004; Renard et al., 2010; Bloch, 2013). A diferencia de la aproximación frecuentista que sólo se basa en la evidencia, el marco bayesiano proporciona un mecanismo formal para combinar la evidencia (información/datos) con el conocimiento a priori. Además, los métodos bayesianos están bien planteados para describir las no linealidades e incertidumbres del sistema [Kavetski et al., 2002; Vrugt et al., 2008; Renard et al., 2010; Bulygina et al., 2009; Almeida, 2014]. Sin embargo, a veces no es posible definir una distribución estadística formal adecuada para caracterizar los errores (por ejemplo, los residuos son muy difíciles de caracterizar, no son estacionarios, restricciones de tiempo para proveer las predicciones, etc.) o emplear una distribución demasiado sencilla conduce a una sobreestimación del contenido de información en los datos [Beven et al., 2008]. En estos casos, se han usado métodos no probabilistas (también llamados posibilistas) [Montanari et al., 2009] como GLUE [Beven and Binley, 1992], DYNIA [Wagener et al., 2003] o teoría de conjuntos dispersos [Beven, 2009]. Concretamente, estos métodos usan la filosofía del enfoque bayesiano pero no usan la teoría de la probabilidad, salvo casos puntuales (ver discusión [Mantovan and Todini, 2006; Todini and Mantovan, 2007; Beven 2006; 2008; Beven et al., 2008]).

1.5.1 Aproximación bayesiana

Dadas unas observaciones Y_{data} , la estructura de un modelo M, y unos datos de entrada I, la distribución posterior de los parámetros del modelo θ_M , viene dada por la Ley de Bayes:

$$p(\theta_M | Y_{data}, M, I) = \frac{L(\theta_M | Y_{data}, M, I) * p(\theta_M | M, I)}{\int_{\Theta_M} L(\theta_M | Y_{data}, M, I) * p(\theta_M | M, I) * d\theta_M}$$
(1)

Donde Y_{data} son los datos de observaciones/evidencia usados para inferir los parámetros del modelo. Por ejemplo, en el modelado hidrológico, Y_{data} suelen ser los datos de caudales o los índices hidrológicos. Sin embargo, en el contexto de PUBs, al no existir observaciones de la variable respuesta, esta evidencia puede venir dada por la información regionalizada usada para condicionar las predicciones.

 $L(\theta_M | Y_{data}, M, I)$ es la función de verosimilitud. Se define como una función de los parámetros del modelo θ_M , dada la evidencia Y_{data} . También puede interpretarse como la probabilidad de que las simulaciones, Y_{sim} dadas por el modelo M con parámetros θ_M representen la evidencia Y_{data} .

 $p(\theta_M)$ es la distribución a priori de los parámetros del modelo *M*. Representa la incertidumbre acerca de los parámetros antes de introducir la información dada por las observaciones/evidencia, Y_{data} . Cuando hay conocimiento previo acerca de la distribución de θ_M , la distribución a priori es llamada informativa. Cuando no hay información previa, suelen emplearse distribuciones a priori no informativas. En modelado hidrológico, es común usar una distribución uniforme en el espacio de los parámetros dentro de su rango admisible [Bulygina et al., 2011] para expresar la falta de conocimiento acerca de θ_M . Los límites de esta distribución uniforme deben de ser lo suficientemente anchos para incluir la zona donde la función de verosimilitud es apreciable y el intervalo donde la distribución a posteriori está definida (la distribución posterior es un subconjunto de la distribución a priori). Cuando se emplea una

distribución uniforme a priori, la calibración es conducida por los datos. Sin embargo, a veces, en entornos con pocos datos, cuando se quiere expresar la falta de conocimiento acerca de la respuesta del sistema, emplear una distribución uniforme en el espacio de los parámetros del modelo hidrológico no resulta adecuado, si ésta distribución no equivale también a una distribución uniforme en el espacio de la variable respuesta. Al objeto de formular esta ignorancia [Jeffreys, 1961], la distribución a priori de θ_M ha de ser definida tal que se traslade en una distribución uniforme en el espacio de la variable respuesta (por ejemplo, los índices hidrológicos, en el contexto de PUBs) [Almeida et al., 2013; Almeida, 2014].

El denominador de la ecuación (1) es la densidad de probabilidad marginal de las observaciones. Representa la probabilidad de obtener una observación bajo todas las posibilidades dadas por θ_M (es decir, integrada en el espacio de los parámetros). Luego para un modelo M, esta integral es igual para cada uno de los parámetros θ_M , pudiendo considerarse una constante de normalización para asegurar que la distribución de probabilidad posterior de los parámetros suma 1. La distribución de probabilidad marginal no es de especial relevancia a la hora de inferir los parámetros, sin embargo si lo es en el contraste de hipótesis o modelos, a través del factor de Bayes (BF, por sus siglas en inglés) [Gelman, 2013]. Los factores de Bayes (ecuación 2) permiten la comparación y selección de modelos/hipótesis [Raftery, 1993; Kass and Raftery, 1995].

$$BF = \frac{\int_{\Theta_{M1}} P(Y_{data}|\theta_{M1}, M1, I) * P(\theta_{M1}|M1, I) d\theta_{M1}}{\int_{\Theta_{M2}} P(Y_{data}|\theta_{M2}, M2, I) * P(\theta_{M2}|M2, I) d\theta_{M2}}$$
(2)

Fundados en la teoría de la probabilidad y deducidos a partir del Teorema de Bayes, los BFs tienen muchas propiedades interesantes, entre otras: permiten comparar muchas estructuras (o hipótesis) sin cambiar el método; las estructuras no necesitan estar anidadas; evitan el sobreajuste, favoreciendo aquellas más simples [Marshall et al., 2005]; y proporcionan una escala cuantitativa para clasificar e interpretar cómo de bien cada hipótesis apoya los datos [Jeffreys, 1961]. Así, para dos hipótesis o estructuras de modelo M1 y M2, BF (2) evalúa cuánto más es apoyada por la evidencia M1 (numerador), frente a M2 (denominador). Además, si para una misma estructura M, se quiere comparar cuánto más verosímil es el parámetro θ_1 que θ_2 a la hora de explicar Y_{data} , y todos los parámetros son igualmente probables a priori, es decir, $p(\theta_1 | M, I) \neq p(\theta_2 | M, I)$ son iguales. Entonces BF es simplemente la razón de verosimilitudes del parámetro θ_1 con respecto al parámetro θ_2 . $L(\theta_1 | M, Y_{data})/L(\theta_2 | M, Y_{data})$.

Una vez se han definido la función de verosimilitud, la distribución a priori y la distribución marginal de los datos, queda resolver la ecuación (1) para inferir la distribución posterior. Esta distribución sintetiza el conocimiento acerca de los parámetros θ_M actualizado por los datos Y_{data}. Cuando la distribución a priori es conjugada de la función de verosimilitud, la distribución a posteriori será del mismo tipo que la distribución a priori. En estos casos, la distribución posterior tiene solución cerrada, sin necesidad de integración numérica. Sin embargo, cuando se utilizan distribuciones a priori no conjugadas, generalmente las distribuciones a posteriori resultan multidimensionales y no siguen una distribución conocida, por lo que no pueden usarse las soluciones analíticas o numéricas comunes. Dicho de otro modo, la teoría que une los datos, Y_{data} , con los parámetros del modelo, θ_M , suele ser no lineal y la distribución de probabilidad en el espacio del modelo suele ser difícil de describir (multimodal, algunos momentos no están definidos, etc) [Beven, 2009]. Cómo consecuencia, en la mayoría de los casos la distribución posterior no se puede calcular analíticamente. Para lidiar con estas situaciones, se han desarrollado varias técnicas numéricas para aproximar la distribución posterior. Éstas se clasifican en 1) aquellas que requieren calcular la función de verosimilitud, y 2) aquellas que requieren muestrear de la función de verosimilitud. Las aproximaciones del primer tipo precisan evaluar la distribución posterior para muchos conjuntos de parámetros θ_M , que se generan numéricamente mediante métodos de Monte Carlo, como son el Muestreo por Importancia [Doucet, 2000] o los métodos de Monte Carlo basados en cadenas de Markov (MCMC) [Kuczera and Parent, 1998]. Estos métodos permiten aproximar la distribución posterior haciendo uso de una distribución alternativa de la que es fácil muestrear. Además, en MCMC la distribución estacionaria es la distribución posterior (ver Schoups and Vrugt, 2010; Engeland et al., 2005; Kuczera and parent, 1998; Steinshneider et al., 2012). Sin embargo, cuando no es posible calcular la función de verosimilitud pero es posible muestrear de ella (es decir, simular a partir del modelo), el segundo tipo de aproximación puede ser implementada. Entre este tipo de técnicas se encuentran los algoritmos de rechazo como el cálculo bayesiano aproximado (ABC, por sus siglas en inglés). El cálculo bayesiano aproximado reemplaza la evaluación de la función de verosimilitud por un paso que implica simular datos artificialmente, Y_{sim} , a partir de diferentes valores de los parámetros, θ_M , y comparar los estadísticos obtenidos en base a Y_{sim} con los estadísticos resumen obtenidos a partir de las observacioness, Y_{data}.Es decir, primero se muestrea un conjunto de parámetros, θ_{M_i} de la distribución a priori; segundo, usando θ_M , se simula a partir del modelo $L(\theta_M | Y_{data}, M, I)$; y tercero, θ_M es aceptado o rechazado en función de la distancia entre las métricas de resumen calculadas a partir de Y_{sim} y las obtenidas a partir de Y_{data} . Consecuentemente, combina un estimador de la función de verosimilitud con una distribución a priori para producir aproximadamente la distribución posterior [Fearnhead and Prangle, 2012]. Si bien ABC evita el cálculo de la función de verosimilitud, $L(\theta_M | Y_{data,}M, I)$, se requiere poder muestrear de ella, lo que implica que esta ha de definirse.

La aproximación que requiere la evaluación de la función de verosimilitud presenta dos dificultades principales: 1) muestrear las variables inciertas a partir de la distribución posterior $p(\theta_M | Y_{data}, M, I)$, a menos que sea una distribución normal multivariante [Kuczera and Parent, 1998], y 2) precisa de un gran número de simulaciones para asegurar la convergencia y que se ha muestreado densamente el espacio de los parámetros (posibles comportamientos de las cuencas). Al objeto de muestrear eficientemente a partir de una distribución de probabilidad, se han desarrollado técnicas de muestreo tales como el Muestreo por Hipercubo Latino (LHS, por sus siglas en inglés) [Mckay et al., 1979]. Las desventajas ligadas a la aproximación que requiere muestrear a partir de la función de verosimilitud están relacionadas: 1) con la tolerancia al rechazo y 2) con los estadísticos de resumen usados para definir los errores. Con respecto a 1), una la tolerancia nula asegura un resultado exacto pero hace los cálculos prohibitivamente caros (en términos computacionales). Sin embargo, usar valores mayores que cero introduce un sesgo en los resultados. En cuanto a 2), ya que habitualmente no se tiene acceso a todos los estadísticos resumen, suele emplearse un número reducido de ellos. Si bien, esto introduce un sesgo adicional debido a la pérdida de información.

Una vez se ha estimado la distribución posterior esta puede usarse para predecir los valores de la variable de interés. Por ejemplo, se pueden muestrear conjuntos de parámetros de la distribución posterior para simular un conjunto de caudales. De este modo, para cada paso de tiempo se tiene una distribución posterior de los caudales, que puede ser evaluada con respecto a las observaciones. Aspectos esenciales de las predicciones probabilistas son su fiabilidad, exactitud y precisión [Bulygina et al., 2011; Yadav et al., 2007]. La fiabilidad cuantifica la capacidad predictiva del modelo (el coste o beneficio de predecir una nueva observación), es decir, mide cuán confiable es que una observación sea la misma cuando el procedimiento se repite en las mismas condiciones. Consecuentemente, depende del método más que de la variable que se está prediciendo. Una predicción probabilista es fiable si la distribución predictiva de la variable de interés es aproximadamente igual a la de la verdadera. Por su parte, la exactitud cuantifica la distancia entre el valor predicho y el verdadero. Por lo tanto, está afectada principalmente por la incertidumbre epistémica. En cambio, la precisión caracteriza la proximidad entre las predicciones (realizadas en las mismas condiciones), sin necesidad de hacer referencia a un valor verdadero, dependiendo sólo de la distribución de errores aleatorios.
Se considera una predicción óptima a aquella que es fiable con la mínima dispersión - dos inferencias pueden proporcionar predicciones fiables pero con diferentes resoluciones [Renard et al., 2010]. La fiabilidad es prioritaria, no debe de sacrificarse fiabilidad por aumentar la precisión. A la hora de evaluar la calidad de las predicciones es común usar los percentiles de la distribución predictiva acumulada. Por ejemplo, los percentiles del 2.5 y 97.5% representan uno de los posibles intervalos de credibilidad (CI) del 95%, lo que significa que hay un 95% de probabilidad de que el intervalo contenga a la observación; y en base a la anchura del CI puede darse una estimación del nivel de precisión [Shrestha, 2009]. Así mismo, otras dos métricas propuestas para evaluar las predicciones son el coeficiente de Nash Sutcliffe probabilista [Buligyna et al., 2009], que permite estimar fiabilidad y precisión en una única métrica; y los gráficos Cuantil-Cuantil predictivos descritos por Laio and Tamea [2007], que permiten comparar gráficamente las distribuciones predictivas acumuladas y observadas. Si la distribución predictiva es fiable, las observaciones corresponden a realizaciones de esta distribución, y los p-valores de la distribución predictiva seguirán una distribución uniforme en el intervalo (U[0-1]). Estos gráficos pueden resumirse mediante los índices de fiabilidad y dispersión propuestos por Renard et al. [2010].

El estado del arte presentado ha subrayado los problemas que aparecen a la hora de hacer predicciones en cuencas sin aforos. Esta tesis está por tanto orientada a mejorar la capacidad de predecir caudales en el contexto de PUBs mediante el desarrollo de metodologías que permitan identificar, cuantificar y reducir la incertidumbre asociada a las diferentes fuentes de error involucradas. Cómo se ha enfatizado, algunos de los problemas principales que aparecen se deben 1) a la clásica hipótesis de que un modelo pre-seleccionado (que incluye su estructura, datos de entrada y parámetros muestreados) es perfecto; 2) a la hipótesis de que el procedimiento de regionalización pre-seleccionado es una representación aceptable y adecuada del comportamiento en una cuenca sin aforos; y 3) a la incapacidad de identificar los mecanismos que dominan el comportamiento hidrológico de las cuencas y que han de ser representados en el modelo del sistema. Esta tesis aborda estos tres problemas mediante los objetivos que se plantean en el Apartado 2.

19

2 OBJETIVOS DE LA TESIS

El objetivo general de esta tesis es contribuir a mejorar la capacidad de predecir caudales en cuencas no aforadas, mediante la identificación, cuantificación y reducción de la incertidumbre asociada a las distintas fuentes de error que intervienen. Esta tesis es un paso adelante en el marco bayesiano desarrollado por Bulygina et al. [2009, 2011], y continuado por Almeida [2014]. Esta tesis avanza algunos de los retos principales que aparecen en modelado hidrológico.

- 1) La clásica suposición de que un modelo pre-seleccionado (que incluye su estructura, datos de entrada y parámetros muestreados) es perfecto.
- 2) La hipótesis de que el procedimiento de regionalización pre-seleccionado es una representación aceptable y adecuada del comportamiento en una cuenca sin aforos.
- La identificación de los mecanismos que dominan el comportamiento hidrológico y que deben de ser representados en el modelo del sistema.

Los desarrollos metodológcicos realizados en esta tesis doctoral se aplican al caso práctico de predecir las variables hidrológicas en las cuencas del tercio norte de España. En particular, se pretende alcanzar los siguientes objetivos:

- Incorporar los nuevos desarrollos en selección de índices hidrológicos (basado en análisis de componentes principales) y métodos de regionalización mediante regresión (bosques aleatorios) usando un enfoque bayesiano. (Apartado 4 de este resumen, Capítulo 2 de la versión en inglés).
- 2. Proponer dos métricas estadísticas para evaluar la aptitud y adecuabilidad de un modelo y procedimiento de regionalización; y cuantificar el impacto que tienen en la predicción de caudales el error del modelo y la calidad de la regionalización, usando una aproximación bayesiana (Apartado 4 de este resumen, Capítulo 2 de la versión en inglés).
- 3. Diseñar un test estadístico (que emplea un conjunto de modelos hidrológicos) para identificar los mecanismos hidrológicos que gobiernan los procesos en las cuencas sin aforos, ayudando a definir una estructura mínima necesaria para un modelo hidrológico (Apartado 5 de este resumen, Capítulo 3 de la versión en inglés).
- 4. Analizar la sensibilidad del proceso de identificación de mecanismos a: i) la calidad de la regionalización, ii) el conjunto de modelos empleado y iii) el contenido de información disponible, usando un enfoque bayesiano. (Apartado 5 de este resumen, Capítulo 3 de la versión en inglés).

3 DESCRIPCIÓN DE LAS CUENCAS Y LOS DATOS

En la presente tesis se usan los datos de caudal y los descriptores de 92 cuencas situadas en el tercio norte de España (Figura 1) para obtener índices hidrológicos regionalizados. Se trata de 92 cuencas que van desde tamaños pequeños (2 km2) hasta los medianos (1038 km2), y constituyen una selección de las 156 utilizadas por Peñas et al. [2014a]. La elección de este subconjunto se ha realizado para asegurar su régimen hidrológico natural de acuerdo con la Directiva Marco del Agua (artículos 5 y 6, secciones de caracterización de las masas de agua y el análisis de presiones e impactos [http://servicios2.marm.es/sia /visualization/descargas/dma.jsp]). De ellas, 62 desembocan en Mar Cantábrico y las otras 30 en el Mar Mediterráneo.

Estas 92 cuencas cubren una amplia gama de geologías, suelos, topografía, usos del suelo y condiciones climáticas. Así, en base a la información extraída del Instituto Geológico y Minero de España ([http://www.igme.es/internet/default.asp), los principales grupos litológicos en las cuencas que drenan al Ebro están constituidos por areniscas, conglomerados y arcillas; siendo predominantes las rocas calizas, rocas calcáreas, evaporitas, margas y arcillas. En cambio, en las cuencas occidentales que desembocan en el Mar Cantábrico, los principales tipos son cuarcitas, pizarras, areniscas, rocas calcáreas, conglomerados y carbón; dominando en las orientales las dolomitas, rocas calcáreas y margas. Los tipos de suelos, por su parte, son en su mayoría arcillas consolidadas, suelos calizos y arenosos. Las altitudes medias de las cuencas oscilan entre 214 m a 2218 m, y la pendiente del canal principal de 9.3% al 62.9%. El área urbanizada en cada una de ellas es inferior al 8% y los usos del suelo son fundamentalmente bosques de hoja ancha, coníferas y pastos. La precipitación media anual varía de 450 mm a 1809 mm por año y la temperatura media de 5.5ºC a 15ºC, oscilando el índice de aridez entre 0.3 y 1.7 [Arora, 2002]. En la Figura 1 se muestra una clasificación de las cuencas en base a su índice de aridez y la temperatura media mensual mínima. Las condiciones climáticas son consideradas secas si el índice de aridez está por encima de 1, y húmedas en caso contrario [Arora, 2002]. Además, es probable que nieve en aquellas cuencas con una temperatura media mensual por debajo de 0ºC. El caudal medio anual oscila entre 17 mm y 1524 mm por año y el coeficiente de escorrentía varía entre 0.03 y 0.97. Los datos climáticos empleados se basan en las estimaciones mensuales de evapotranspiración potencial, temperatura y precipitación proporcionadas por el Centro de Estudios Hidrográficos para el periodo 1980-2005 (CEDEX, Ministerio de Obras Públicas y el Ministerio de Agricultura y Medio Ambiente, España); mientras que los datos de caudal se basan en series diarias de aforo con al menos ocho años de registro en el periodo 1976-2009 [Peñas et al., 2014a].

En esta tesis, los datos de las 92 cuencas se utilizan para seleccionar un conjunto ortogonal compacto de piezas de información que representan la variabilidad de los índices hidrológicos, así como para diseñar un modelo de regionalización para el conjunto ortogonal seleccionado.

Además, 16 cuencas de las 92 se usan para evaluar los resultados de las predicciones obtenidas al aplicar las metodologías desarrolladas en esta tesis. Por lo tanto, para este fin, serán considerada como 'cuencas no aforadas'. Estas 16 cuencas constituyen aquellas donde se dispone de series sincronizadas diarias de precipitación, caudal y evapotranspiración potencial mensual (con un período común de al menos 8 años) [Peñas et al., 2014a] (Figura 1) Los datos de precipitación diaria han sido proporcionados por la Agencia Estatal de Meteorología Española (AEMET) y las características de las 16 cuencas se muestran en la Tabla 1.



Figure 1: Casos de estudio y climatología

Tabla 1: Características de las cuencas (1) rocas calcáreas; (2) arcillas; (3) roca conglomerada, (4) arenas; (5) rocas sedimentarias; (6) esquistos (sedimentarias); (7) rocas silíceas; (8) pizarras; (9) rocas volcánicas. Ver Peñas et al. [2014a] para permeabilidad y geología

Geología (% Área ocupada)	(6)	Ē	0	0	0	0	0	7	0	0	0	1	0	0	0	0	7	4
	(8)	Ē	72	68	31	50	0	0	0	0	0	0	0	0	0	0	7	51
	(2)		1	0	35	0	0	9	0	0	0	0	0	0	0	10	0	0
	(9)		0	0	0	0	48	22	20	73	35	ы	30	33	20	21	37	18
	(2)	Ē	4	11	m	∞	0	Ч	0	1	0	7	11	0	0	10	0	0
	(4)	2	11	16	7	4	49	47	73	0	0	Ļ	9	55	80	0	28	21
	(3)	Ē	10	ъ	0	19	7	ß	0	m	0	0	0	0	0	7	0	0
	(2)		0	0	10	0	0	0	0	ы	0	7	0	0	0	0	m	1
	(1)	Ì	2	0	19	18	Ч	17	2	17	65	6	23	11	0	57	28	ы
	Permeabilidad		Baja	Muy Baja	Baja	Muy Baja	Baja	Baja	Baja	Baja	Baja	Baja	Alta	Baja	Baja	Alta	Alta	Baja
Mín. Temp	Mínima	Mensual (ºC)	1.1	0.2	4.0	-1.3	0.7	-2.7	-1.4	-1.1	-0.4	0.1	0.0	-1.8	-0.1	1.1	2.5	2.7
РЕТ	Anual	mm/ año	660	638	734	716	962	705	619	655	564	701	679	609	693	650	684	731
Precipitación	Anual	mm/ año	1459	1449	1342	1041	814	1356	1333	681	1134	1469	1376	1683	1371	1038	1794	1809
Coeficiente	occourse to	escolrentia	0.61	0.81	0.64	0.57	0.33	0.69	0.39	0.24	0.79	0.95	0.41	0.35	0.39	0.35	0.66	0.65
Área	2 mg	Ey	529	293	377	294	80	623	75	283	22	152	554	73	95	477	114	28
Nº máximo	al miómotroc	pinvioriietros	3	m	m	1	1	14	1	1	1	1	1	1	1	9	2	2
Ň	2 more	AIIOS	22	10	14	80	10	6	10	8	11	13	8	13	12	14	12	11
Estación			x1353	x1404	x1303	x1265	9257	x9040b	9269	9197	9221	AN439	AN433	AN520	AN530	AN313	c8z1	c7z1

24

4 MÉTRICAS ESTADÍSTICAS PARA EVALUAR LA APTITUD Y ADECUABILIDAD DE LA REGIONALIZACIÓN Y EL MODELO EN PUBS

Parte de este trabajo fue presentado en Viena en la Reunión de la Unión Geofísica Europea (EGU) 2015 [Prieto et al., 2015a; 2015b].

4.1 INTRODUCCIÓN

La predicción de caudales en cuencas no aforadas (PUBs) sigue siendo uno de los grandes retos de la hidrología [Almeida et al., 2016]. Entre las diferentes alternativas para cuantificar y reducir la incertidumbre en las predicciones, las aproximaciones bayesianas han demostrado ser las más adecuadas [Blöschl, 2013]. Estas aproximaciones permiten predecir los caudales en cuencas no aforadas en base a los índices hidrológicos regionalizados en dichas cuencas [Yadav et al., 2007; Bulygina et al., 2009; Singh, 2013]. Apoyándose en la teoría de la probabilidad, el procedimiento requiere de un número de elecciones. Entre las más importantes se encuentran: 1) la selección de los índices hidrológicos representativos del comportamiento general de las cuencas, que son calculados en las cuencas con aforos; 2) la selección del modelo de regionalización, que relaciona los índices hidrológicos con los descriptores de las cuencas (CDs, por sus siglas en inglés); y 3) la selección del modelo hidrológico.

La contribución de este capítulo es abordar los objetivos 1, y 2 del Apartado 2, progresando las prácticas tradicionales de predicción de caudales en cuencas sin aforos a) asimilando dentro del contexto de la predicción en cuencas sin aforos los nuevos avances en selección y regionalización de índices hidrológicos, y b) introduciendo métricas para evaluar la aptitud y adecuabilidad del modelo y procedimiento de regionalización seleccionados. La aptitud evalúa (i) cómo de bien el modelo reproduce la información regionalizada y (ii) la capacidad de la regionalización a la hora de reproducir la información hidrológica en una cuenca dada. La adecuabiliad evalúa la ganancia de información de (i) la regionalización y (ii) el modelo. Es decir, el valor de añadir información regionalizada o un modelo en las predicciones.

En primer lugar, se reduce un vasto número de índices hidrológicos a un conjunto ortogonal compacto de piezas de información (PCs, por sus siglas en inglés) mediante análisis de componentes principales (ACP). Segundo, estos PCs son regionalizados mediante bosques aleatorios y la información regionalizada se emplea para condicionar las predicciones hidrológicas en un marco bayesiano. Y tercero, se presentan dos métricas estadísticas formales para evaluar la aptitud y adecuabilidad. Estas métricas se basan en un nuevo test de 'aptitud' y los factores de Bayes (BFs, por sus siglas en inglés, Gelman, 2013). El procedimiento se ilustra

25

para la predicción de caudales en las cuencas del tercio norte de España (ver Apartado 3 de este resumen).

4.2 METODOLOGÍA

4.2.1 Índices hidrológicos y Descriptores de cuenca

En este trabajo 103 índices hidrológicos son transformados en un conjunto de PCs, mediante ACP. Y se usa el método 'Broken Stick' [Jackson, 1993; Peres-Neto et al., 2005] para seleccionar aquellos PCs que son más significativos. Además, a partir de 50 CDs [Peñas et al., 2014a], se eligen aquellos CDs con coeficiente de correlación menor de 0.7, lo que resulta en 16 CDs. Los PCs más significativos y estos 16 CDs, se usan para construir el procedimiento de regionalización.

4.2.2 Regionalización mediante bosques aleatorios

El algoritmo de bosques aleatorios hace un remuestreo con reemplazamiento del número de árboles de regresión para componer el Bosque. Cada árbol se construye tal que en cada uno de sus nodos se muestrean aleatoriamente un número de CDs y se escogen aquellos que dan la mejor división, desde el punto de vista del menor error medio cuadrático [Liaw et al., 2002; Snelder et al., 2012]. El PC regionalizado, *PC^{reg}*, viene dado por la media de las predicciones individuales dadas por cada árbol del bosque (Figura 2) [Snelder et al., 2011]. Para cada una de las 16 cuencas empleadas como cuencas no aforadas (ver Apartado 3), se utilizan las otras 91 para regionalizar sus *PC^{reg}*. La incertidumbre asociada al procedimiento de regionalización se estima mediante una estrategia de re-muestreo 'Jack-knife' con base en estas 91 cuencas. Así, la distribución conjunta de las diferencias entre los *PC^{reg}* y los PCs calculados a partir de los datos de observaciones, *PC^{data}*, caracterizan el error de la regionalización mediante bosques aleatorios.



Figure 2: Regionalización mediante bosques aleatorios

4.2.3 Inferencia bayesiana

Dados PC^{reg} la estructura de un modelo M, y unos datos de entrada I, la distribución posterior de los parámetros del modelo, θ , viene dada por la Ley de Bayes [Bulygina et al., 2009]:

$$p(\theta | PC^{Reg}, M, I) = \frac{L(PC^{sim}_{\theta, M, I} | PC^{Reg}) * p(\theta | M, I)}{p(PC^{Reg} | M, I)}$$
(3)

Donde $L(PC_{\theta,M,I}^{sim}|PC^{Reg})$ es la función de verosimilitud y $p(\theta|M,I)$ es la distribución a priori de los parámetros. $L(PC_{\theta,M,I}^{sim}|PC^{Reg})$ es proporcional a la distribución de residuos del procedimiento de regionalización descrita en el apartado anterior; y $p(\theta|M,I)$ es una distribución uniforme en el espacio de los parámetros. Como generalmente no es posible muestrear directamente de la distribución posterior, ésta es aproximada numéricamente usando Muestreo por Importancia [Doucet, 2000].

4.2.4 Test de Aptitud

Para determinar si un modelo es representativo de la información regionalizada PC^{reg} , se muestrean 10000 muestras de la distribución de PC^{reg} (10000 muestras de m-dimensionales PCreg), que se comparan con los PCs simulados $PC_{\theta,M,I}^{sim}$. Para ello, primero, se aproxima $p(PC_{\theta,M,I}^{sim})$ mediante una mezcla de N gausianas [Muller et al. 1996; Bulygina and Gupta, 2011] y se definen las elipses críticas de cada componente gausiana usando la distancia de Mahalanobis:

$$d_{i} = \sqrt{(PC - \mu_{i})^{t} * \Sigma_{i}^{-1} * (PC - \mu_{i})} \quad (4)$$

Para la componente gausiana ith (i=1,N), μ_i representa el mx1 valor medio y Σ_i la mxm matriz de covarianza . El cuadrado de la distancia de Mahalanobis d_i^2 para una distribución gausiana $N(\mu_i; \Sigma_i)$ sigue una distribución chi-cuadrado con m grados de libertad [Gallego et al. 2013]. Los contornos de igual densidad de cada componente gausiana se caracterizan por una distancia de Mahalanobis constante y son elipsoides centrados en μ_i con sus ejes alineados con los auto vectores de Σ_i [Friendly et al., 2013]. Estos contornos permiten definir la región más compacta al nivel de confianza [1- α_i] para una distribución Gausiana $N(\mu_i; \Sigma_i)$ [Gallego et al. 2013; Ribeiro, 2004].

Para la hipótesis nula de que el modelo es capaz de reproducir la información contenida en los primeros m PCs regionalizados, el correspondiente p-valor se obtiene promediando los p-valores de cada una de las 10000 muestras. El p-valor para cada muestra de PC^{reg} se define como el mayor valor de α_i , i=1,N. De este modo, hay al menos una región elíptica que contiene la muestra de PC^{reg} . Este test se denomina test de 'adecuabilidad'.

Un segundo test de aptitud evalúa la hipótesis nula de que el procedimiento de regionalización es capaz de reproducir los PCs basados en observaciones, PC^{data} . En este caso, se aproxima la distribución de residuos del modelo de regionalización mediante una mezcla de gausianas. De forma similar al caso anterior, las elipses críticas se calculan usando la distancia de Mahalanobis (4) y el p-valor para PC^{data} es el mayor valor de α_i , i=1,N. Este p-valor máximo caracteriza cómo de bien la distribución de PC^{reg} describe PC^{data} .

Un tercer test de aptitud evalúa la hipótesis nula de que el modelo es capaz de reproducir PC^{data} . En este caso se compara PC^{data} con $PC^{sim}_{\theta,M,I}$. Análogamente al primer test de aptitud, $p(PC^{sim}_{\theta,M,I})$ se aproxima mediante una mezcla de gausianas, las elipses críticas se calculan usando la distancia de Mahalanobis (4) y el p-valor máximo para PC^{data} se define cómo el mayor valor de α_i , i=1,N. En este caso, este máximo p-valor caracteriza cómo de bien la distribución de $PC^{sim}_{\theta,M,I}$ describe PC^{data} .

4.2.5 Adecuabilidad

4.2.5.1 Adecuabilidad de la Regionalización

La adecuabilidad del procedimiento de regionalización se define mediante los BFs [Gelman et al., 2013]. El factor de Bayes cuantifica la información que añade el proceso de regionalización con respecto al conocimiento a priori sobre *PC*^{data}.

$$BF = \frac{p(PC^{data}|PC^{reg})}{p(PC^{data})}$$
(5)

El numerador en la ecuación (5) es igual a la función de verosimilitud, que se calcula para la diferencia dada por ($PC^{data} - PC^{reg}$); y p (PC^{data}) se expresa mediante un producto de distribuciones uniformes para cada uno de los PCs. Los límites de esta distribución vienen dados por los valores mínimos y máximos de los PCs para el conjunto de las 92 cuencas. Se dice que el procedimiento de regionalización es adecuado si BF es mayor que 1. Los factores de Bayes sirven para interpretar la fuerza de la evidencia a favor de la hipótesis en el numerador con respecto a la hipótesis en el denominador [Jeffreys, 1961]. Además, se demuestra que la ecuación (5) es equivalente a cuanta información acerca PC^{data} es añadida por el procedimiento de regionalización con respecto a considerar únicamente la información dada por el modelo.

$$BF = \frac{p(PC^{data} | PC^{reg}, M, I)}{p(PC^{data} | M, I)}$$
(6)

BF	Fuerza de la evidencia			
<1	Negativa			
	(Apoya la hipótesis en el denominador)			
1 to 10^(0.5)	Apenas merece mencionarse			
10^0.5 to 10	Substancial			
10 to 10^3/2	Fuerte			
10^3/2 to 100	Muy Fuerte			
>100	Decisiva			

Tabla 2: Escala para interpretar el factor de Bayes [Jeffreys, 1961]

4.2.5.2 Adecuabilidad del Modelo Hidrológico

La adecuabilida de la elección del modelo (que incluye seleccionar su estructura, especificar sus datos de entrada y el muestreo de parámetros) también se evalúa mediante el BF [Gelman et al., 2013]. En este caso, BF cuantifica cuánta información es añadida por el modelo acerca del conocimiento a priori de *PC*^{data}.

$$BF = \frac{P(PC^{data} \mid M, I)}{p(PC^{data})}$$
(7)

El numerador de la ecuación (7) representa la probabilidad de obtener PC^{data} en la distribución de los PCs simulados por el modelo, $p(PC_{\theta,M,I}^{sim})$. $p(PC_{\theta,M,I}^{sim})$ se aproxima ajustando una mezcla de Gausianas a $PC_{\theta,M,I}^{sim}$. El denominador se expresa como un producto de distribuciones uniformes para cada uno de los PCs, como en la ecuación (5). El modelo se considera adecuado si BF es mayor que 1, y puede emplearse para interpretar cuánto es apoyado el modelo por la evidencia (Tabla 2). Se demuestra que la ecuación (7) equivale a cuánta información acerca de PC^{data} es añadida al introducir la información dada por el modelo y sus datos de entrada con respecto a considerar únicamente la regionalización:

$$BF = \frac{P(PC^{data} | PC^{reg}, M, I)}{P(PC^{data} | PC^{reg})}$$
(8)

4.2.6 Casos de estudio

Se consideran cuatro casos de estudio que difieren en el error del modelo y la calidad de la regionalización:

BC) El caso emplea las observaciones de caudal (considerando el error del modelo), y la regionalización obtenida mediante el método de bosques aleatorios.

R) El caso difiere de BC en que el procedimiento de regionalización es exacto y tiene una mayor precisión. Los errores de la regionalización siguen una distribución Ggausiana de media cero y desviación igual al 5% de la dispersión a priori de los PCs. Como los *PC^{reg}* se fijan igual al valor de los PCs calculados a partir de una serie de caudales, el p-valor del correspondiente test de aptitud es 1.

M) El caso difiere de BC en que el error asociado al modelo es nulo. Esto se logra sustituyendo los caudales observados por los caudales generados mediante un modelo con un conjunto de parámetros seleccionado, forzado con los datos de entrada observados. Las características de la regionalización son las mismas que en el caso BC. El conjunto de parámetros elegido para definir la serie sintética de flujo es aquel que minimiza la distancia normalizada entre $PC_{\theta MI}^{sim}$ y los

$$PC^{data}: \sqrt{\sum_{i=1}^{p} \left(\frac{PC_{i}^{sim} - PCdata}{std(PC_{i}^{sim})}\right)^{2}}.$$
 Dónde, std (.) denota la desviación estándar. El correspondiente

PC^{reg} se obtiene añadiendo el ruido de la regionalización (definido en el caso BC) a los PCs calculados a partir de los caudales sintéticos.

MR) El caso difiere del caso M en el procedimiento de regionalización, en MR el procedimiento de regionalización es igual al del caso R, luego el p-valor del test de aceptabilidad es 1.

4.2.7 Calidad de las predicciones de caudal

Dado un vector de series de caudales simulados, {Q_t}, y una serie de caudales observados, q_t^{obs} , en el tiempo *t*, la fiabilidad y precisión de las predicciones se evalúan mediante el coeficiente de eficiencia de Nash-Sutcliffe probabilista (*NS*^{prob}) [Bulygina et al., 2009] y los intervalos de credibilidad del 95% [Yadav et al., 2007]. En la ecuación (9), *E*[.] define el valor esperado, *Var*[.] la varianza y *T* es el número de instantes de tiempo. El primer término de la parte derecha de la ecuación (9) mide la exactitud y el segundo término la precisión.

$$NS^{Prob} = \left\{ 1 - \frac{\sum_{t=1}^{T} (E[Q_t] - q_t^{obs})^2}{\sum_{t=1}^{T} (q_t^{obs} - E[q_t^{obs}])^2} \right\} - \frac{\sum_{t=1}^{T} (Var[Q_t])}{\sum_{t=1}^{T} (q_t^{obs} - E[q_t^{obs}])^2}$$
(9)

Otro aspecto importante de una predicción probabilista es su fiabilidad. La fiabilidad cuantifica la consistencia estadística entre los datos observados y su distribución predictiva. El porcentaje de veces que los caudales observados caen en el intervalo de credibilidad del 95% (95% CI) [Yadav et al, 2007] se emplea para evaluar la fiabilidad de las predicciones.

4.3 RESULTADOS Y DISCUSIÓN

4.3.1 Calidad de las predicciones de caudal

Los PCs más significativos resultan ser los cuatro primeros PCs. Estos PCs explican un 87.6% de variabilidad total (63.3%, 12.4%, 6.9% y 4.8%). Por tanto, ACP reduce en gran medida el espacio de 103 dimensiones de los índices hidrológicos. Además, ACP maximiza el valor de la información, una tarea considerada de especial relevancia en PUBs [Wagener and Montanari, 2011].

4.3.2 Regionalización y función de verosimilitud

El algoritmo de bosques aleatorios se emplea para regionalizar los PCs. La Tabla 3 muestra que la calidad media de la regionalización varía entre el 9.8% y el 15.6%, del rango de los PCs calculado en base a las 92 cuencas. Para cada PC individual, la calidad se calcula como el cociente entre el parámetro de escala y el rango de dicho PC. El PC1 es el PC regionalizado con mayor precisión, pero es el más inexacto; en cambio el PC4 es el más exacto y el PC3 el menos preciso. Asimismo, se constata que la a incertidumbre de la regionalización llevada a cabo con bosques aleatorios es similar o mejor a la obtenida mediante regresión paso a paso en Almeida [2014, Tabla 7.6], dónde la precisión variaba entre 7 and 27%.

	Localización	Escala	Escala normalizada
PC1	2.42	3.69	10%
PC2	0.18	2.28	13%
PC3	0.05	1.7	16%
PC4	0	1.9	12%

Tabla 3: Calidad media de la regionalización para cada PC. El parámetro de escala normalizada se ha estimado dividiendo el parámetro de escala por el rango del correspondiente PC.

La función de verosimilitud en la ecuación (3) se estima mediante las distribuciones de residuos de los PCs regionalizados (es decir, las diferencias entre los PCs regionalizados y basados en datos de observaciones). El coeficiente de correlación linear de Pearson revela una correlación significativa entre los residuos de PC2 y PC3 (p-valor igual a 0.03). La distribución marginal de los residuos de PC1 sigue una distribución de extremos tipo I (EV1) en base al test chi-cuadrado. El test chi-cuadrado también indica que la distribución marginal de los residuos de PC4 sigue una distribución Gausiana. Además, el test chi-cuadrado y el test de Mardia indican que la distribución conjunta de los residuos de PC2 y PC3 sigue una distribución gausiana bidimensional. Todos los tests se realizan para el nivel de significancia del 5%). Por lo tanto, la función de verosimilitud cuatri-dimensional se estima mediante el producto de las distribuciones marginales de PC1 y PC4 y la distribución conjunta de PC2 y PC3. La Figura 3

proporciona las distribuciones marginales teóricas y los histogramas empíricos obtenidos empleando la técnica de dejar uno fuera.



Figura 3: Distribuciones marginales e histogramas empíricos (EMP.) de los residuos de la regionalización.

4.3.3 Evaluación de la calidad de las predicciones

Cuando las observaciones de caudal no están disponibles, el test de aptitud del modelo muestra si el modelo hidrológico es apto para representar la información regionalizada en la cuenca sin medidas. En 3 cuencas (X1353, AN433 y c8z1), el modelo hidrológico es apto (p-valor mayor que 0.05) para la hipótesis nula de que el modelo es capaz de reproducir la información regionalizada; la hipótesis de aptitud se rechaza en las otras 13 cuencas (Figura 4a, ítem Reg&Mod). Lo segundo significa que es probable que la estructura de modelo seleccionada no reproduzca las cuatro piezas de información regionalizada a la vez.

En el caso BC_{reg}, la hipótesis de que el procedimiento de regionalización es una representación aceptable de los PCs en base a las observaciones de caudal se acepta para todas las cuencas, el p-valor es mayor que 0.05 (Figura 4a, ítem BC_{reg}). Mientras que el test de adecuabilidad para el caso BC_{reg} muestra que el procedimiento de regionalización es adecuado en 12 cuencas (BF es mayor que 1 – Figura 4b, caso BC_{reg}) y es inadecuado en cuatro cuencas (X1404, AN433, X1353, y X9040). Así, en dichas 12 cuencas, la regionalización es a la vez apta y adecuada.

Sin embargo, el caso BC_M muestra que la estructura del modelo es una representación apta de *PC^{data}* (es decir, en base a las observaciones de caudal) en 6 cuencas (el p-valor es mayor que 0.05 – Figura 4a, caso BC_M) AN530, AN520, AN433, X9257, X1353 y C8Z1; y el modelo es adecuado (BF es mayor que 1 – Figura 4b, caso BC_M) en 3 cuencas (X1353, X9257 y C8Z1). Entonces el modelo es a la vez una descripción apta y adecuada de *PC^{data}* en tres cuencas (X1353, X9257 y C8Z1). Es importante señalar que el p-valor del test de aptitud (BC_M) es mayor que 0.05 para la cuenca AN520, pero el factor de Bayes es 1; luego el modelo no puede considerarse ni informativo ni desinformativo.

Consiguientemente, sólo en las cuencas C8Z1 y X1353, el modelo es a la vez (1) una representación apta de la información regionalizada (Figura 4a, ítem Reg&Mod), (2) una representación apta de los PCs observados en la cuenca (p-valor es mayor que 0.05 Figura 4a, caso BC_M), y (3) una representación adecuada de los PCs observados en la cuenca – es decir, el modelo añade suficiente información acerca de PC^{data} (BF is mayor que 1, Figura 4b caso BC_M) en las cuencas C8Z1 y X1353.

Por otra parte, únicamente en 2 cuencas de las 16 (C8Z1 y X9257), el procedimiento de regionalización y el modelo son una representación apta y adecuada de PC^{data} (Figura 4a y 4b, intersección de los resultados para los casos BC_M y BC_{reg}).

Además, simultáneamente (1) el modelo reproduce los PCs regionalizados con alta probabilidad (Figura 4a ítem Reg&Mod), (2) el procedimiento de regionalización es apto y adecuado (Figura 4a y 4b, caso BC_{reg}) y (3) el modelo es apto y adecuado (Figura 4a y 4b caso BC_{M}) sólo en la cuenca C8Z1.

Para el caso R_{reg}, los p-valores en la Figura 4a indican que la regionalización es aceptable para todas las cuencas (el p-valor es 1). Asimismo, los BFs en la Figura 4b revelan que el procedimiento de regionalización añade 53051 veces más de información en comparación con el conocimiento a priori (el valor es el mismo para cada cuenca, puesto que se basa en una distribución con un error fijo). Es decir, la fuerza de la evidencia es decisiva (Tabla 2). En el caso M_M el modelo es apto y adecuado en todas las cuencas (p-valor mayor que 0.05 y BF mayor que 1, Figuras 4a y 4b). Además, en el caso $M_{R_{reg}}$, mejorar la calidad de la regionalización conduce a valores de BF iguales a 53051 y p-valores iguales a 1 en todas las cuencas.

Para investigar el impacto del aptitud y adecuabilidad de la regionalización y del modelo en las predicciones de caudal, en la Figura 5 se muestran los valores posteriores de *NS^{prob}* y 95%CI para los casos BC, R, M y MR. Los gráficos a) y b) resumen el cambio en exactitud y precisión, que viene dado por *NS^{prob}* y los gráficos c) y d) muestran lo mismo para la fiabilidad, que se cuantifica

33

mediante el 95% CI. Las figuras muestran que para las cuencas C8Z1 y X1353, dónde el modelo es una representación apta y adecuada de la información regionalizada y una representación apta y adecuada de *PC^{data}*, mejorar la regionalización (caso R en la Figura 5a y 5c) conduce a un comportamiento similar o mejor en términos de *NS^{prob}* (de 0.78 a 0.71 en la cuenca C8Z1, y de 0.65 a 0.72 en la cuenca X1353) y 95% CI (de 70% a 74% en la cuenca C8Z1, y de 45% a 52% en la cuenca X1353). Sin embargo, en las otras 14 cuencas, mejorar sólo la regionalización (R) puede tanto mejorar como empeorar el comportamiento del modelo. *NS^{prob}* mejora en 8 cuencas de 14. El 95%CI mejora en 7 de las 14 cuencas, empeora en 6, y es igual en 1. El *NS^{prob}* promedio es 0.3 en el caso BC y es 0.32 en el caso R. Siendo el 95%CI promedio igual al 2% en el caso BC y 60% en el caso R. Esto es indicativo de que el error en el modelo impide una mejora de las predicciones cuando se mejora la regionalización.

La Figura 5 también muestra que los resultados mejoran para el caso MR con respecto al caso M en todas las cuencas - *NS*^{prob} y 95%CI en el caso M es 72%, mientras que en MR es 95%. Demostrando que cuando el modelo es capaz de reproducir los PCs, una mejora en la regionalización conduce a una mejora de las predicciones de caudal (exactitud, precisión y fiabilidad). En este caso, los esfuerzos han de enfocarse en mejorar la regionalización para mejorar las predicciones.



Figura 4: Métricas de aptitud y adecuabilidad: (a) p-valores y (b) BFs. Las métricas se evalúan para el procedimiento de regionalización y el modelo seleccionado. 'reg' indica las métricas relevantes al procedimiento de regionalización y 'M' indica las métricas

Para cuantificar si una mejora en las diferentes fuentes de conocimiento implica una diferencia estadísticamente significativa en la calidad de las predicciones, se ha llevado a cabo un conjunto de tests de análisis de varianza (ANOVA). Se han realizado las siguientes comparaciones (1) BC con R, que refleja si mejorar la calidad de la regionalización comporta un cambio estadísticamente significativo en la calidad de las predicciones (*NS*^{prob} y 95% CI) cuando el modelo es incorrecto, y (2) M con MR, que refleja si mejorar la calidad de las predicciones (*NS*^{prob} y 95% CI) cuando el modelo es tadísticamente significativo en la calidativo en la calidad de las predicciones (*NS*^{prob} y 95% CI) cuando el modelo es perfecto. Los resultados muestran que mejorar la exactitud y precisión de la regionalización cuando el modelo es perfecto comporta una diferencia estadísticamente significativa en la exactitud, precisión y fiabilidad de las predicciones. Para el caso MR, *NS*^{prob} y 95% CI mejoran con respecto a M. En el caso MR, *NS*^{prob} es mayor que 0.9 en 11 de las 16 cuencas y el 95% CI es mayor que el 90% en 13. Mientras que en el caso M, *NS*^{prob} es mayor que 0.9 en 6 cuencas (X9257, X1265, X9221, AN530, AN313, C8Z1).



Figura 5: a) NSprob para los casos BC y R, b) NSprob para los casos M y MR, c) 95%Cl para los casos BC y R, y d) 95%Cl para los casos M y MR.

4.4 CONCLUSIONES

En predicciones en cuencas sin aforos, una práctica común es usar un modelo y técnica de regionalización pre-seleccionados. Esto es así, a pesar de que no existe una justificación para su

aptitud (la capacidad de que el modelo y/o la regionalización reproduzcan la información disponible) y adecuabilidad (el conocimiento ganado al considerar el modelo y/o la regionalización).

Esta investigación propone tests estadísticos para cuantificar la aptitud y la adecuabilidad del modelo y/o regionalización y sugiere un procedimiento para desagregar las diferentes fuentes de incertidumbre y establecer prioridades al objeto de reducir la incertidumbre predictiva.

Este estudio supone un avance con respecto a las prácticas estándar para predecir caudales en cuencas sin aforos, cuando se emplean índices hidrológicos regionalizados para condicionar las predicciones de un modelo hidrológico usando un enfoque bayesiano. El trabajo presentado sintetiza la información hidrológica disponible mediante un análisis de componentes principales; seguidamente, las componentes principales más significativas se regionalizan mediante bosques aleatorios; y finalmente, esta información regionalizada actualiza la información dada por el modelo mediante un proceso de condicionamiento bayesiano.

Además, el trabajo contribuye con dos nuevos tests para evaluar la aptitud y adecuabilidad del modelo y la regionalización:

- El test de aptitud se basa en el concepto de inferencia simultánea y explícitamente aborda las siguientes hipótesis nulas: (a) un modelo es capaz de reproducir la información regionalizada (u observada, si está disponible) con alta probabilidad, (b) un procedimiento de regionalización elegido es capaz de reproducir la información disponible en una cuenca.
- 2. El test de adecuabilidad se basa en el factor de Bayes [Gelman, 2013] y cuantifica el conocimiento ganado (a) incluyendo el procedimiento de regionalización y (b) incluyendo el modelo.

Estos tests de aptitud y adecuabiliad permiten identificar la fuente de incertidumbre dominante (entre la regionalización y el modelo) y, por ende, dónde deben de concentrarse los esfuerzos para reducirla. La aplicación a un conjunto de 16 cuencas del tercio norte de España conduce a los siguientes hallazgos:

1. El algoritmo de bosques aleatorios demuestra ser un método de regionalización eficiente (calidad-tiempo). El error de la regionalización es pequeño y el método no requiere hacer hipótesis acerca de la forma de la función (lineal/no lineal), del número de predictores y de la naturaleza estadística de los errores. Por otra parte, el espacio de índices hidrológicos altamente dimensional reducido mediante ACP y el método de 'broken stick' proporcionan suficiente información para la regionalización mediante bosques aleatorios en el caso de las cuencas del norte de España.

- Cuando el modelo pasa el test de aptitud para la información regionalizada, así como los tests de aptitud y adecuabilidad para la información disponible, mejorar la calidad de la regionalización es el objetivo para obtener mejores predicciones (fiabilidad, exactitud y precisión).
- 3. Sin embargo, mejorar la regionalización cuando el modelo es erróneo, no conduce a mejores predicciones. Por lo tanto, la calidad de la regionalización no está directamente relacionada con la calidad de las predicciones. La prioridad debe de darse a la elección de una estructura de modelo representativa (que incluye los datos de entrada) y luego a una mejora de la regionalización. No al revés.
- 4. El procedimiento de regionalización mediante bosques aleatorios pasa el test de aptitud en todas las cuencas (es decir, la regionalización es probable que reproduzca la información disponible en todas las cuencas); y pasa el test de adecuabilidad en el 75% de las cuencas (es decir, la regionalización proporciona una ganancia de conocimiento).

El trabajo expuesto en este Apartado 4 da respuesta a los objetivos 1 y 2 indicados en el Apartado 2.

5 IDENTIFICACIÓN DE LOS MECANISMOS DOMINANTES EN PUBS VÍA APROXIMACIÓN BAYESIANA

Parte de este trabajo fue presentado Viena en la Reunión de la Unión Geofísica Europea (EGU) 2016 [Prieto et al., 2016a; 2016b].

5.1 INTRODUCCIÓN

Identificar los mecanismos que gobiernan el comportamiento de una cuenca es uno de los grandes retos de la hidrología, que afecta tanto a cuencas aforadas como no aforadas [Clark et al., 2008; Clark et al., 2011a; Coxon et al., 2014]; y cuya mayor aplicación se encuentra a la hora de seleccionar o identificar estructuras de modelos hidrológicos. Un modelo hidrológico combina diferentes hipótesis para cada uno de los procesos que rigen el funcionamiento de las cuencas [Beven, 2001]. A estas hipótesis se les asignan diferentes grados de credibilidad [Clark et al., 2011a]; ya que nunca pueden ser validadas al no conocerse completamente el fenómeno natural [Wagener et al., 2004].

Debido a la escasez de datos, la práctica habitual en hidrología es representar los procesos usando una única estructura de modelo previamente seleccionada, tanto en cuencas aforadas como no aforadas [Blöschl et al., 2013.; Wagener and Montanari, 2011; Ley et al., 2016]. Esto, a pesar de lo ampliamente evidenciado que está el paradigma de la singularidad del lugar [Beven, 2000], de que es imposible que una única estructura represente adecuadamente el sistema hidrológico [Clark et al., 2011a; 2011b; 2015a; 2015b; 2015c], y de que pasar por alto el error asociado a la estructura del modelo conduce a una subestimación de la incertidumbre predictiva y a un sesgo en las predicciones [Reicherte and Omlin, 1997].

Al objeto de mejorar esta situación se han desarrollado aproximaciones flexibles y adaptativas, que proporcionan diferentes mecanismos para cada proceso hidrológico, descomponiendo un modelo en un conjunto de hipótesis que pueden ser examinadas. Por ejemplo, el paquete de herramientas para el modelado de los procesos de transferencia lluvia-caudal (RRMT) [Wagener et al., 2001; 2004], el marco para entender los errores asociados a la estructura de los modelos hidrológicos (FUSE) [Clark et al., 2011a; 2011b], FLEX y SUPERFLEX [Fenicia et al., 2008], o la estructura para unificar las múltiples alternativas de modelado (SUMMA) [Clark et al., 2015a; 2015b; 2015c]. En cuencas con medidas, es típico contrastar tales hipótesis usando índices hidrológicos [Clark et al., 2011b; Coxon et al., 2014]; sin embargo, sigue siendo uno de los mayores retos en hidrología desarrollar una prueba estadística que además considere tanto la incertidumbre asociada a la estructura del modelo y sus datos de entrada, como la debida a los índices hidrológicos [Clark et al., 2011b; Boyle et al., 2000; Wagener et al., 2003; Wagener et al.,

2007; Kirchner et al., 2006]. Si bien la aproximación basada en índices es adecuada para ser extendida al contexto de PUBs, incorporando en el análisis la incertidumbre debida a la regionalización de la información.

La contribución de esta investigación es abordar los objetivos 3 y 4 expuestos en el Apartado 2, proponiendo una metodología bayesiana para condicionar un conjunto de estructuras de modelos hidrológicos usando la información regionalizada; y ayudar a identificar los mecanismos que dominan los procesos hidrológicos en las cuencas no aforadas. La metodología se fundamenta en la teoría de la probabilidad bayesiana, que está especialmente bien planteada tanto para contrastar hipótesis [Raftery, 1993], como para describir las no linealidades e incertidumbres del sistema [e.g., Kavetski et al., 2002; Vrugt et al., 2008; Renard et al., 2010; Bulygina et al., 2009; Almeida, 2014]. En primer lugar, la información contenida en una gran cantidad de índices hidrológicos es reducida mediante análisis de componentes principales. Segundo, en las cuencas sin aforos, esta información es regionalizada mediante bosques aleatorios. Tercero, en dichas cuencas, la información regionalizada se emplea para condicionar un conjunto de estructuras de modelos, usando una aproximación bayesiana. Y cuarto, se diseña un estadístico que permita contrastar hipótesis acerca de los mecanismos que dominan los procesos hidrológicos en las cuencas sin aforos. El método se demuestra para las 16 cuencas del tercio norte de España descritas en el Apartado 3. Los resultados muestran que identificar con éxito los mecanismos dominantes puede estar ligado a: 1) el error debido a la estructura del modelo, 2) la calidad de la regionalización, 3) el contenido de información en los índices hidrológicos, y 4) la cantidad de información necesaria para identificar los mecanismos que dominan los procesos hidrológicos en las cuencas no aforadas.

5.2 MODELOS HIDROLÓGICOS

A partir de cuatro estructuras de modelos padre (Sistema de modelado Iluvia-caudal - PRMS, por sus siglas en inglés; SACRAMENTO - Modelo Sacramento de Humedad del Suelo; TOPMODEL y ARNO/VIC), FUSE [Clark et al., 2008; 2011b] permite crear 1248 estructuras. De estas 1248, hemos usado 624, al no ser objeto de esta tesis considerar el error asociado a la precipitación.

Cada una de estas estructuras emplea distintos mecanismos para caracterizar: 1) los procesos en la capa superior del suelo, 2) los procesos en la capa inferior del suelo, 3) la generación de la escorrentía superficial, 4) la infiltración, 5) la evaporación, 6) el caudal lateral y 7) el enrutamiento (Figura 6, Tabla 4). Los datos de entrada empleados son los datos de observaciones diarias de precipitación y estimadores mensuales de la evapotranspiración potencial (Apartado 2).



Figura 6: FUSE

Tabla 4: Mecanismos incluidos en las 624 estructuras.

PROCESOS	MECANISMOS						
	1	Representados por una variable de estado					
	2	Representados por variables de estado separadas: agua sometida a tensión (por					
P1:		debajo de la capacidad de campo) y agua libre (por encima de la capacidad de campo).					
Procesos en la capa	3	Representados por variables de estado separadas: agua sometida a tensión (por					
superior del suelo		debajo de la capacidad de campo) y agua libre (por encima de la capacidad de campo).					
		El almacenamiento del agua sometida a tensión se representa desagregadamente en					
		depósitos en cascada.					
	1	Representados por un depósito de tamaño fijo.					
P2:	2	Representados por un depósito dónde el agua está sometido a tensión y dos tanques					
Procesos en la capa		en paralelo.					
inferior del suelo	3	Representados por un depósito de tamaño ilimitado (tasa lineal).					
	4	Representados por un depósito de tamaño ilimitado (recesión potencial).					
	1	El área saturada está relacionada con el almacenamiento en la zona no saturada					
D2.		mediante una distribución de Pareto.					
F3. Generación de	2	El área saturada es una función lineal del almacenamiento de agua sometido a tensión					
escorrentía		en la zona no saturada.					
escorrentia	3	El área saturada está relacionada con el almacenamiento en la zona no saturada					
		mediante el índice topográfico.					
	1	Caracterizada por la disponibilidad de agua desde la capacidad de campo a la					
₽ 4 ·	-	saturación.					
Infiltración	2	Depende de la demanda de infiltración desde la zona inferior del suelo.					
	3	Caracterizada por la disponibilidad de agua desde el punto de marchitez a la saturación					
		(la zona saturada domina).					
	1	La evaporación se calcula en base al porcentaje de raíz en cada capa del suelo.					
P5: Evaporación	2	Evaporación en la capa superior e inferior del suelo. La demanda que no es satisfecha					
		en la capa superior se satisface en la capa inferior.					
P6:	1	No hay flujo lateral.					
Flujo lateral	2	Es una función lineal del almacenamiento libre en la capa superior.					
P7: Enrutamiento	1	No hay enrutamiento					
	2	Hay enrutamiento					

5.3 METODOLOGÍA

5.3.1 Distribución posterior de los mecanismos

Cada proceso hidrológico p=1,P es representado por un mecanismo $mec_m^p, m = 1, N^p$ (Tabla 4); por tanto, una estructura de modelo M_k , $\{M_k\}$, k=1,N, se define mediante una combinación de mecanismos, tal que dicha combinación representa todos los procesos P considerados. Para cada proceso p, la probabilidad posterior de cada mecanismo mec_m^p en el conjunto de modelos $\{M_k\}$, es proporcional a la suma de las probabilidades posteriores de aquellas estructuras que representan el proceso p mediante el mecanismo mec_m^p considerado (en el conjunto hay N^p estructuras que tienen el mecanismo mec_m^p).

$$p(mec_m^p | PC^{Reg}, I, \{M_k\}) \propto \sum_{Mk:mec_m^p} p(M_k | PC^{Reg}, I)$$
(10)

Y la probabilidad posterior del modelo M_k , dados los componentes principales regionalizados, PC^{reg} (ver apartado 4.2.1) y datos de entrada *I*, puede calcularse mediante (11):

$$p(M_k|PC^{Reg}, I) \propto L(M_k|PC^{Reg}, I) * p(M_k|I)$$
 (11)

Dónde, $L(M_k/PC^{reg}, I)$, representa la probabilidad de obtener PC^{reg} con el modelo M_k y datos de entrada I; y $p(M_k/I)$ es la probabilidad a priori del modelo M_k . En este estudio se asume que todas las estructuras son equiprobables a priori, con probabilidad $\frac{1}{N}$. Una presunción clave aquí es que al menos una estructura M_k y sus datos de entrada son capaces de reproducir los PCs. Para tener encuenta que los distintos mecanismos no son muestreados con la misma frecuencia dentro del conjunto de modelos (para un proceso p, el conjunto de modelos puede incluir más mecanismos de un tipo que de otro), la probabilidad posterior del mecanismo mec_m^p es aproximada por la ecuación (10) ponderada por el inverso de la frecuencia con la que cada mecanismo es muestreado en el conjunto de modelos, $\frac{N_m^p}{N}$ (ver Muestreo por Importancia [Doucet et al. 2000]):

$$p(mec_m^p | PC^{Reg}, I, \{M_k\}) \propto \sum_{Mk:mec_m^p} p(M_k | PC^{Reg}, I) * \frac{1}{N_m^p}$$
(12)

Ya que las probabilidades posteriores (12) son variables aleatorias (debido a la finitud del conjunto de modelos seleccionado), para tener en cuenta la dispersión estadística de la probabilidad posterior en (12), se lleva a cabo un muestreo con reemplazamiento boot-strap. Para cada proceso p, se re-muestrean N_m^p modelos con mecanismo mec_m^p del conjunto de modelos con dicho mecanismo (hay N_m^p modelos con mecanismo mec_m^p en el conjunto de

modelos inicial); y se calcula la probabilidad posterior del mecanismo mec_m^p (12). El re-muestreo se realiza un gran número de veces para estimar una función de densidad de probabilidad (en adelante pdf) de la probabilidad posterior de cada mecanismo. De tal modo que es más independiente del conjunto de modelos seleccionado (ya que el conjunto de modelos seleccionado es finito, no es posible hacer estas pdfs completamente independiente del mismo).

5.3.2 Identificación de los mecanismos dominantes

Para cada cuenca y proceso, se hace un contraste de múltiples hipótesis, dónde se prueba la hipótesis nula de que un mecanismo mec_m^p no domina sobre los otros, es decir, $p(mec_m^p | PC^{Reg}, I) \leq \sum_{i \neq m} p(mec_i^p | PC^{Reg}, I)$, usando la corrección de Bonferroni [Yosef, 1988]. Como las probabilidades posteriores de cada mecanismo no se conocen con exactitud, y son descritas mediante re-muestreo (5.3.1.), el test de hipótesis compara múltiples muestras y calcula la fracción de veces que las muestras de la pdf posterior de mec_m^p excede la suma de las muestras de las pdfs de probabilidades posteriores de los otros que compiten. De modo que si el porcentaje es menor o igual que $\frac{\alpha}{N^p}$, se acepta la hipótesis nula (mec_m^p no es dominante), y se rechaza en otro caso (mec_m^p se considera dominante).

5.3.3 Análisis de sensibilidad: Casos

Para examinar a influencia de la calidad de la regionalización, el error asociado al modelo y la cantidad de información, se analizan cinco casos para cada una de las 16 cuencas:

- Caso BC: Los caudales son simulados usando el conjunto de modelos incluidos en FUSE forzados con los datos de observaciones de precipitación diaria y estimadores mensuales de evapotranspiración potencial descritos en el Apartado 2. Se selecciona el número de PCs que explican la mayor variabilidad mediante el método 'Brocken Stick'; y estos PCs se regionalizan empleando bosques aleatorios (4.2.1.).
- Caso R: Se mejora la calidad de la regionalización, la regionalización es exacta (los *PC*^{reg} reproducen exactamente los PC basados en observaciones) y de alta precisión.
- Caso M: Se excluye el error del modelo y la calidad de la regionalización es la misma que en el caso BC.
- Caso MR: Se mejora tanto la calidad de la regionalización como la del modelo, combinando los casos M y R. La mejora se logra sustituyendo los caudales observados por los caudales sintéticos generados por uno de los modelos del conjunto. Para que la calidad de los PCs regionalizados sea similar que en el caso BC, a los PCs calculados a partir de esta serie sintética se les añade el ruido calculado a partir de los PCs basados en observaciones y los PCs basados en los PCs regionalizados.

 Caso MRI: Es una modificación del caso MR que usa al menos un 90% de la información contenida en los PCs disponibles, tal que la cantidad de información no es menor que en el caso BC.

Para todos los casos y cuencas, se presenta el porcentaje de mecanismos identificados (A) y correctamente identificados (B). Nótese que los mecanismos verdaderos son desconocidos para los casos BC y R.

5.4 RESULTADOS

5.4.1 Contenido de información

La Figura 7 muestra el contenido de información en los primeros cuatro PCs calculados con base en las observaciones de caudal (BC y R) y en base a los caudales generados sintéticamente (M, MR y MRI).



Figura 7: Contenido de información en los primeros cuatro PCs para los caudales observados y para los caudales sintéticos.

5.4.2 Calidad de las predicciones

La Figura 8 muestra el número de mecanismos identificados (A) y correctamente identificados (B) en el conjunto de las 16 cuencas tratadas como 'no aforadas', para cada uno de los procesos descritos en la Tabla 3. Los mecanismos que dominan el enrutamiento, la evaporación, el flujo lateral y la escorrentía superficial son los más identificables para el caso basado en observaciones (BC). Detectándose en 16, 13, 11 y 9 de las 16 cuencas, respectivamente (Figura 9, imágenes BC a), BC b), BC c) y BC d)). Nótese que no es posible estimar el número de

mecanismos dominantes correctamente identificados en los casos BC y R, ya que no se tiene acceso al mecanismo verdadero. En cambio, los mecanismos que dominan el comportamiento en la capa superior del suelo y la percolación no se identifican en ninguna cuenca en el caso BC. Mejorar la regionalización (R) conduce a una mezcla de resultados, ya que se observa que el número de mecanismos dominantes identificados puede aumentar, disminuir o permanecer igual que en el caso BC. Esto puede deberse a la presencia del (mismo) error asociado a la estructura del modelo/datos de entrada en ambos casos. Entonces, mejorar la información regionalizada no conduce necesariamente a una mejora de las predicciones. La Figura 9 muestra los mecanismos hidrológicos identificados con mayor frecuencia en el caso BC. Es decir, con base en los cuatro PCs regionalizados a partir de la información observada en cuencas aforadas. Se trata de los mecanismos que dominan el enrutamiento, la evaporación, el caudal lateral y la escorrentía. En todas las cuencas se prefiere el mecanismo de enrutamiento de flujo; y la mayoría se describen mejor mediante la evaporación secuencial, sin flujo lateral y un mecanismo de generación de escorrentía basado en el índice topográfico. Así mismo, la capa inferior del suelo se caracteriza por un depósito de tamaño ilimitado con una tasa fraccionaria, en aquellas cuatro cuencas donde se identifica dicho proceso. Los mecanismos que ocurren en la capa superior del suelo y que dominan la percolación no se identifican en ninguna cuenca. Debido a la pequeña muestra de cuencas disponible (16 cuencas) y a la falta de variabilidad de mecanismos identificados, no es posible extraer conclusiones significativas acerca de las correspondencias entre los descriptores de las cuencas (ver Apartado 3 de este Resumen) y los mecanismos hidrológicos dominantes identificados.

Cuando se excluye la incertidumbre debida al modelo (M, MR y MRI), mejorar la calidad de la información regionalizada tiende a identificar los mecanismos dominantes con mayor fiabilidad (los mecanismos son correctamente identificados) (Figura 10). Además, añadir información de alta calidad aumenta la fiabilidad de los mecanismos dominantes identificados (comparar M y MR con respecto a MRI en la Figura 10).

Por otra parte, los mecanismos que dominan los procesos en la capa superior del suelo parecen ser los menos identificables (sólo en un 25% de las cuencas) para el mejor escenario (MRI). Asimismo, en ninguno de los procesos ocurre que se identifican los mecanismos dominantes correctamente en todas las cuencas. En un máximo de 10 cuencas, se identifican correctamente los mecanismos que rigen un proceso hidrológico. Esto podría estar causado por la imposibilidad de los 103 índices hidrológicos (los PCs derivados a partir de ellos) para capturar y representar los detalles de las dinámicas hidrológicas consideradas.

45

En los casos M, MR y MRI, se evalúa la exactitud con la que un mecanismo dominante es identificado. En estos casos se conoce el mecanismo verdadero. La Figura 10 muestra que cuando no hay error asociado al modelo y se emplea una gran cantidad de información regionalizada con una calidad alta (caso MRI), la identificación de los mecanismos hidrológicos dominantes es fiable. Es decir, si la metodología identifica un mecanismo dominante, la identificación es exacta. A pesar de que la variabilidad de los mecanismos identificados en el caso MRI (Figura 10) es mayor que en el caso BC (Figura 9), no es adecuado relacionar los mecanismos identificados con los descriptores de las cuencas, ya que el caso MRI se basa en caudales generados sintéticamente.



Figura 8: Mecanismos identificados (A) y correctamente identificados (B) para cada caso.



Figura 9: Caso BC: mecanismos identificados para los procesos a) enrutamiento, b) evaporación, c) flujo lateral, d) escorrentía y e) procesos en la capa superior del suelo.



Figura 10: Caso MRI: mecanismos identificados para los procesos a) enrutamiento, b) evaporación, c) flujo lateral, d) escorrentía, e) procesos en la capa inferior del suelo, f) infiltración y g) procesos en la capa superior del suelo.

5.5 CONCLUSIONES

Este trabajo propone una metodología estadística para identificar los mecanismos que gobiernan el comportamiento en cuencas sin aforos y plantea el camino para reducir la incertidumbre en su identificación. El estudio junta los desarrollos recientes en selección de índices hidrológicos y métodos de regresión con las fortalezas de la inferencia bayesiana. La investigación emplea un conjunto de estructuras de modelos hidrológicos (FUSE, Clark et al., 2008; 2011b), donde cada modelo se representa mediante una combinación de mecanismos. Los mecanismos se tratan como una muestra de un conjunto de mecanismos completo (pero inobservable). Esta información se combina con la información proporcionada por los PCs regionalizados mediante bosques aleatorios. La aplicación de los desarrollos metodológicos al conjunto de 16 cuencas del norte de España revela los siguientes hallazgos:

- Cuando existe error asociado al modelo, mejorar la regionalización no implica necesariamente una mejor identificación de los mecanismos dominantes. Sólo cuando el error del modelo es pequeño o eliminado y se emplea una gran cantidad de la información disponible, mejorar la calidad de la regionalización lleva a una mejor identificación de los mecanismos.
- Cuando no existe error asociado al modelo y hay disponible una gran cantidad de información bien regionalizada, el procedimiento propuesto para identificar los mecanismos dominantes es fiable.
- 3. El contenido de información en los primeros cuatro PCs permite identificar los mecanismos dominantes para los procesos de enrutamiento, evaporación, flujo lateral y escorrentía.
- 4. El extenso conjunto de índices hidrológicos típicamente considerados [Olden and Poff, 2003] puede no ser suficiente para identificar 1) los mecanismos que gobiernan el comportamiento en la capa superior del suelo y 2) los mecanismos que gobiernan el proceso de percolación.
- 5. La identificación de los mecanismos dominantes es sensible 1) al error del modelo, que incluye la estructura y sus datos de entrada, 2) a la calidad de la regionalización y 3) a la cantidad de información.

El trabajo expuesto en este Apartado 5 da respuesta a los objetivos 3 y 4 indicados en el Apartado 2.

6 RESUMEN DE LAS CONTRIBUCIONES, CONCLUSIONES GENERALES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

6.1 RESUMEN DE LAS CONTRIBUCIONES AL ESTADO DEL ARTE

El objetivo principal de esta tesis ha consistido en desarrollar metodologías para mejorar la fiabilidad en las predicciones de variables hidrológicas en cuencas no aforadas. El objetivo planteado se ha alcanzado por medio de la consecución de las siguientes metas de carácter específico:

- 1. Incorporar los nuevos avances en selección de índices hidrológicos y métodos de regionalización para la predicción de caudales en cuencas no aforadas. Para ello, primero, se ha reducido un vasto número de índices hidrológicos a un conjunto ortogonal compacto de piezas de información mediante análisis de componentes principales. Segundo, estas piezas de información se han regionalizado mediante bosques aleatorios. Y tercero, la información regionalizada se ha empleado para condicionar las predicciones hidrológicas usando un marco bayesiano.
- 2. Desarrollar dos métricas estadísticas para evaluar la aptitud y adecuabilidad de un modelo y procedimiento de regionalización seleccionados, usando una aproximación bayesiana, así como evaluar el impacto que tienen en la incertidumbre el modelo hidrológico y el procedimiento de regionalización seleccionados.

Las métricas desarrolladas se basan en a) cómo de bien el modelo y el procedimiento de regionalización reproducen la información hidrológica disponible en una cuenca, y b) la ganancia de información al considerar el modelo y la regionalización. Las métricas se definen empleando el concepto de inferencia simultánea y los factores de Bayes [Gelman, 2013].

- Desarrollar un test estadístico para identificar los mecanismos hidrológicos que gobiernan el comportamiento de las cuencas. Consecuentemente, los mecanismos identificados como dominantes deben de ser incluidos en la estructura del modelo hidrológico. El test desarrollado usa la corrección de Bonferroni para comparaciones múltiples [Yosef, 1988].
- Evaluar la sensibilidad de la identificación de mecanismos dominantes a a) la calidad de la regionalización, b) el error en la estructura del modelo y c) la cantidad de información disponible, usando estadística bayesiana

Estos desarrollos metodológicos se han aplicado al caso práctico de predecir las variables hidrológicas en las cuencas del tercio norte de España, las cuales cubren una gran variedad de

regímenes hidro-climáticos. Los resultados han revelado que el rendimiento de las predicciones es sensible al error en el modelo (o conjunto de modelos), la calidad de la información regionalizada y el contenido de información disponible.

6.2 CONCLUSIONES GENERALES

6.2.1 APARTADO 4: Métricas estadísticas para evaluar la aptitud y adecuabilidad de la Regionalización y el Modelo en PUBs

En predicciones en cuencas sin aforos, una práctica común es usar un modelo y técnica de regionalización preseleccionados. Esto es así, a pesar de que no exista una justificación para su aptitud (la capacidad de que el modelo y/o la regionalización reproduzcan la información disponible) y adecuabilidad (el conocimiento ganado al considerar el modelo y/o la regionalización).

Este estudio propone tests estadísticos para cuantificar la aptitud y la adecuabilidad del modelo y/o regionalización y sugiere un procedimiento para desagregar las diferentes fuentes de incertidumbre y establecer prioridades al objeto de reducir la incertidumbre predictiva.

Esta investigación supone un avance con respecto a las prácticas estándar para predecir caudales en cuencas sin aforos, cuando se emplean índices hidrológicos regionalizados para condicionar las predicciones de un modelo hidrológico usando un enfoque bayesiano. El trabajo presentado sintetiza la información hidrológica disponible mediante un análisis de componentes principales; seguidamente, las componentes principales más significativas se regionalizan mediante bosques aleatorios; y finalmente, esta información regionalizada actualiza la información dada por el modelo mediante un proceso de condicionamiento bayesiano.

Además, el trabajo contribuye con dos nuevos tests para evaluar la aptitud y adecuabilidad del modelo y la regionalización:

- El test de aptitud se basa en el concepto de inferencia simultánea y explícitamente aborda las siguientes hipótesis nulas: (a) un modelo es capaz de reproducir la información regionalizada (u observada, si está disponible) con alta probabilidad, (b) un procedimiento de regionalización elegido es capaz de reproducir la información disponible en una cuenca.
- El test de adecuabilidad se basa en el factor de Bayes [Gelman, 2013] y cuantifica el conocimiento ganado (a) incluyendo el procedimiento de regionalización y (b) incluyendo el modelo.

Estos tests de aptitud y adecuabiliad permiten identificar la fuente de incertidumbre dominante (entre la regionalización y el modelo), y por ende, dónde deben de concentrarse los esfuerzos

para reducirla. La aplicación a un conjunto de 16 cuencas del tercio norte de España conduce a los siguientes hallazgos:

- 1. El algoritmo de bosques aleatorios demuestra ser un método de regionalización eficiente (calidad-tiempo). El error de la regionalización es pequeño y el método no requiere hacer hipótesis acerca a) de la forma de la función (lineal/no lineal), que relaciona los descriptores de las cuencas con los índices hidrológicos; b) del número de predictores y c) de la naturaleza estadística de los errores. Por otra parte, el espacio de índices hidrológicos altamente dimensional reducido mediante ACP y el método de 'Broken Stick' proporciona suficiente información para la regionalización mediante bosques aleatorios en el caso de las cuencas del norte de España.
- Cuando el modelo pasa el test de aptitud para la información regionalizada, así como los tests de aptitud y adecuabilidad para la información disponible, mejorar la calidad de la regionalización es el objetivo para obtener mejores predicciones (fiabilidad, exactitud y precisión).
- 3. Sin embargo, mejorar la regionalización cuando el modelo es erróneo, no conduce a mejores predicciones. Por lo tanto, la calidad de la regionalización no está directamente relacionada con la calidad de las predicciones. La prioridad debe de darse a la elección de una estructura de modelo representativa (que incluye los datos de entrada) y luego a una mejora de la regionalización.
- 4. El procedimiento de regionalización mediante bosques aleatorios pasa el test de aptitud en todas las cuencas (es decir, la regionalización es probable que reproduzca la información disponible en todas las cuencas); y pasa el test de adecuabilidad en el 75% de las cuencas (es decir, la regionalización proporciona una ganancia de conocimiento).

6.2.2 APARTADO 5: IDENTIFICACIÓN DE LOS MECANISMOS DOMINANTES EN PUBS VÍA APROXIMACIÓN BAYESIANA

Este trabajo propone una metodología estadística para identificar los mecanismos que gobiernan el comportamiento en cuencas sin aforos y plantea el camino para reducir la incertidumbre en su identificación. El estudio junta los desarrollos recientes en selección de índices hidrológicos y métodos de regresión con las fortalezas de la inferencia bayesiana. La investigación emplea un conjunto de estructuras de modelos hidrológicos (FUSE, Clark et al., 2008; 2011b), dónde cada modelo se representa mediante una combinación de mecanismos. Los mecanismos se tratan como una muestra de un conjunto de mecanismos completo (pero inobservable). Esta información se combina con la información proporcionada por los PCs

regionalizados mediante bosques aleatorios. La aplicación de los desarrollos metodológicos al conjunto de 16 cuencas del norte de España revela los siguientes hallazgos:

- Cuando existe error asociado al modelo, mejorar la regionalización no implica necesariamente una mejor identificación de los mecanismos dominantes. Sólo cuando el error del modelo es pequeño o eliminado y se emplea una gran cantidad de la información disponible, mejorar la calidad de la regionalización lleva a una mejor identificación de los mecanismos.
- Cuando no existe error asociado al modelo y hay disponible una gran cantidad de información bien regionalizada, el procedimiento propuesto para identificar los mecanismos dominantes es fiable.
- 3. El contenido de información en los primeros cuatro PCs permite identificar los mecanismos dominantes para los procesos de enrutamiento, evaporación, flujo lateral y escorrentía.
- 4. El extenso conjunto de índices hidrológicos típicamente considerados [Olden and Poff, 2003] puede no ser suficiente para identificar 1) los mecanismos que gobiernan el comportamiento en la capa superior del suelo y 2) los mecanismos que gobiernan el proceso de percolación.

La identificación de los mecanismos dominantes es sensible 1) al error del modelo, que incluye la estructura y sus datos de entrada, 2) a la calidad de la regionalización y 3) a la cantidad de información

6.3 FUTURAS LÍNEAS

Cada uno de los capítulos de esta tesis, en especial los capítulos 2 y 3 de la versión en inglés que se corresponden con los apartados 4 y 5 de la versión en castellano, han dado lugar a unas preguntas de investigación que es necesario abordar. En concreto: desinformación debida a una mala interpretación de los datos, contenido de información, error asociado a la estructura del modelo y selección y combinación de modelos.

6.3.1 Desinformación debida a una mala interpretación de los datos

Una de las conclusiones señaladas en el apartado 4 (y corroborada en el resto del trabajo) es que la incertidumbre debida a la estructura del modelo/datos de entrada es mayor que la de la regionalización.

Este hallazgo ha hecho plantearse las siguientes preguntas que han de ser investigadas: ¿Es factible desagregar la incertidumbre asociada a los datos de la incertidumbre asociada a la estructura del modelo?, ¿Cuál es la contribución relativa de cada una de estas fuentes de error

en la incertidumbre total?, ¿Qué impacto tienen los períodos de datos desinformativos en las predicciones de caudal [e.g. Reichert and Mieleitner, 2009]?, ¿Cómo se puede cuantificar la cantidad de información asociada a cada tormenta y cómo se puede integrar dicha información en la función de verosimilitud? Cuestiones que están además en línea con los recientes desarrollos acerca de las bases filosóficas sobre incertidumbre hidrológica [Nearing et al., 2016].

Tales preguntas han conducido a la Autora de esta tesis a proponer una metodología para identificar aquellos períodos que son desinformativos en las series de precipitación y caudal, y estimar su impacto en la incertidumbre predictiva. Esta metodología se basa en una convergencia de aproximaciones para detectar tales períodos, extendiendo así mismo el análisis de identificabilidad dinámico propuesto por Wagener et al. [2003]. Con este objetivo, se han analizado las cuencas del tercio norte de España (Tabla 1) en términos de (1) distribución de los pluviómetros y termómetros en las cuencas (ver la función de densidad acumulada de la elevación de cada cuenca en la Figura 11 y Figura 12); (2) los gráficos de coordenadas paralelas de los descriptores de las cuencas y de los índices hidrológicos (Figura 13 y Figura 14); y (3) las curvas de duración de caudales. Este análisis de los datos ha puesto de manifiesto la mala calidad de los mismos. Asimismo, se ha comprobado que el criterio del MOPEX [Schaake et al., 2000, Schaake et al., 2006] con respecto al número de pluviómetros y termómetros en función del área, sólo se cumplía en una cuenca (C8Z1 – ver Tabla 1). El procedimiento propuesto pretende abordar la identificación de los períodos desinformativos en las series de precipitación y caudal a través de los siguientes pasos: Primero, se hace una clasificación de los tipos de error comúnmente encontrados en las series de datos (ver McMillan et al., 2012). La Tabla 4 muestra un ejemplo de esta clasificación para una de las cuencas analizadas. La primera columna indica el tipo de error y en la segunda columna el sufijo A indica precipitación y el B caudal. Segundo, se desarrolla un procedimiento de identificación de errores potenciales y se investiga cómo estos se manifiestan en el espacio de los datos, el espacio de los parámetros y el espacio de la estructura del modelo. Para ello:

Se genera una serie sintética 'verdadera' de caudal (Q_t) con un modelo hidrológico usando una serie real de precipitación (R_t) - la serie (R_t,Q_t) se denomina 'verdadera'. Aunque el modelo hidrológico empleado ha sido el 'Probability Distributed Model' (PDM, por sus siglas en inglés) [Moore, 2007], la metodología diseñada es genérica para ser aplicada con otros modelos:

 Para el caso 'verdadero' se calculan métricas/gráficos de comportamiento que sirven como referencia para identificar datos potencialmente erróneos. Estas métricas/gráficos son: a) el coeficiente de escorrentía basado en eventos (caracteriza el espacio de los datos); b) DYNIA [Wagener et al, 2003] dejando libre un parámetro del modelo hidrológico y fijos el resto (caracteriza el espacio de los parámetros). En primera instancia se ha dejado libre el parámetro que caracteriza la contribución de caudal rápido y lento del PDM [Moore, 2007]; y c) el conjunto de los modelos posibles [Beven and Binley, 1992] (caracteriza el espacio de la estructura del modelo).

- 2. Los errores mostrados en la primera columna de la tabla 4 se introducen en la serie 'verdadera' (Rt, Qt) y se comparan con las métricas/gráficos obtenidas en el caso 'verdadero' calculados en 1). Para ello, se dibujan los gráficos a, b y c para cada tipo de error. Los escenarios considerados se encuentran en las filas de la tabla 4.
- Adoptando el enfoque de la hidrología comparativa se propone investigar cómo estos tipos de error afectan a cuencas que cubren diferentes regiones climáticas, descriptores e índices hidrológicos (ver cuencas resaltadas en la Figura 13 y Figura 14).

Resultados preliminares

El análisis de las cuatro cuencas resaltadas en las Figuras 13 y 14 revela que los gráficos asociados al coeficiente de escorrentía basado en eventos sólo identifican períodos desinformativos cuando dicho coeficiente es mayor que 1. Luego es más fácil detectar eventos desinformativos en cuencas con grandes coeficientes de escorrentía y más difícil en cuencas con condiciones más secas. Esto ocurre tanto para una subestimación de la precipitación, como para una sobre estimación del caudal (por ejemplo, ver Figura 15). En otros casos se necesita tener un conocimiento acerca de la evapotranspiración y el almacenamiento al objeto de considerar cuando es plausible un coeficiente de escorrentía muy pequeño. Un descubrimiento potencialmente relevante es que DYNIA detecta los eventos de caudal y las tormentas que no se registran en las series de datos. Esto se ha obtenido empleando el parámetro que separa la contribución del caudal rápido y lento del PDM [Moore, 2007] (ver Figura 16). Es decir, DYNIA [Wagener et al., 2003] detecta si hubo una tormenta pero no se observa la respuesta en el caudal; o si hubo un hidrograma pero la tormenta que lo generó no se ve registrada en la serie de precipitación. Se propone continuar la metodología investigando la aproximación basada en DYNIA [Wagener et al., 2003] con otros parámetros del PDM [Moore, 2007]. Así como también investigar la manera en que la combinación de los diferentes errores impacta en la incertidumbre predictiva.
ESCENARIOS							
Sesgo Constante	1A	P-30% P	P-20% P	P-10% P	P+10% P	P+20% P	P+30% P
	1B	Q-30% Q	Q-20% Q	Q-10% Q	Q+10% Q	Q+20% Q	Q+30% Q
Eventos No Registrados	2A	Tormenta (4/3/06-12/3/06)	Tormenta (8/11/05-9/11/05)	Tormenta (27/12/05-5/1/06)	Tormenta (14/4/06-18/4/06)	Tormenta (5/1/05-6/1/0 5)	
	2В	Caudal (14/02/06-2/6/06)	Caudal (28/10/05-26/12/05) U (14/02/06-2/6/06)	Caudal (3/3/05-13/3/05)	Caudal (5/9/05-24/9/05)	Caudal (15/12/04-15/ 1/05)	
Error Heterocedástico	ЗA	0.5*P	0.75*P	1.25*P	1.5*P		
	ЗB	0.5*Q	0.75*Q	1.25*Q	1.5*Q		
Error Aleatorio Heterocedástico Autocorrelado	4A	Multiplicador Heterocedástico 0.2	Multiplicador Heterocedástico 0.4	Multiplicador Heterocedástico 0.6	Multiplicador Heterocedástico 0.8		
	4B	Multiplicador Heterocedástico 0.2	Multiplicador Heterocedástico 0.4	Multiplicador Heterocedástico 0.6	Multiplicador Heterocedástico 0.8		
Error Aleatorio en la precipitación	5A	Multiplicador Heterocedástico 0.2	Multiplicador Heterocedástico 0.4	Multiplicador Heterocedástico 0.6	Multiplicador Heterocedástico 0.8		
Error Aleatorio Inversamente Heterocedástico Autocorrelado	6A	Multiplicador Heterocedástico 0.2	Multiplicador Heterocedástico 0.4	Multiplicador Heterocedástico 0.6	Multiplicador Heterocedástico 0.8		
	6B	Multiplicador Heterocedástico 0.2	Multiplicador Heterocedástico 0.4	Multiplicador Heterocedástico 0.6	Multiplicador Heterocedástico 0.8		
Tormentas Desplazadas	7A	Tormenta (3/3/05-8/3/05) Desplazada a (2/3/05-7/3/05)	Tormenta (11/8/05) Desplazada a (10/8/05)	Desplazar todo 1 día	Tormenta (29/10/05-31/10/05) Desplazada a (28/10/05-30/10/05)	Torme (29/10/05-31 (9/11/05-18 Desplazz (28/10/05-30 (8/11/05-17	nta /10/05)& s/11/05) nda a s/10/05)& 7/11/05)

Tabla 5: Tipos de errores introducidos en cada una de las cuencas.



Figura 11: Elevación de los pluviómetros en la función de distribución acumulada del modelo digital de elevación de las cuenca.



Figura 12: Elevación de los termómetros en la función de distribución acumulada del modelo digital de elevación de las cuenca.



Figura 13: Gráfico de coordenadas paralelas de los descriptores de las cuencas. Eje x: descriptores de cuenca. Eje y: valor de los descriptores de cuenca.



Figura 15: Espacio de los datos. Ejemplo del coeficiente de escorrentía basado en eventos para un sesgo constante.

Eigura 15: Espacio de los datos. Ejemplo del coeficiente de escorrentia basado en eventos para un sesgo constante. Eje x: tiempo en días. Eje y izquierda: desviación introducida en la serie 'verdadera' de caudal. Eje y derecha: valor del coeficiente de escorrentía.



Figura 16: Espacio de los parámetros. DYNIA. Tipo de error: eventos de caudal no registrados. Eje x: tiempo en días. Eje y: valor del parámetro que representa el porcentaje de caudal rápido en el PDM [Moore, 2007].

Esta metodología para detectar períodos desinformativos hizo que surgieran las preguntas: ¿cómo de informativa es cada tormenta y cómo cuantificar esta información? Lo que condujo a la Autora de esta tesis a proponer el proyecto de artículo científico 'Diferentes formas de evaluar el contenido de información de cada tormenta'. A este respecto, la Autora ha propuesto aplicar la metodología para separar periodos informativos y desinformativos expuesta en el apartado anterior, y extender el marco creado por Beven and Smith [2014] para evaluar la información aportada por cada tormenta y definir la función de verosimilitud. Así como emplear los factores de Bayes [Gelman, 2013] para estimar el conocimiento dado por cada realización del modelo hidrológico acerca de cada evento.

6.3.2 Contenido de información

El Apartado 5 de esta tesis también ha demostrado que no todo el contenido de información es necesario para obtener predicciones fiables. Esto ha conducido a las siguientes preguntas de investigación: ¿Para cada cuenca, cuáles son los PCs que contienen la información necesaria para obtener predicciones fiables, exactas y precisas?, ¿ Cuál es la relación entre la información contenida en estos PCs (y, por ende, en los índices hidrológicos a partir de los que se derivan dichos PCs) y los mecanismos que gobiernan los procesos hidrológicos?, ¿ Cómo identificar que PCs deben de seleccionarse a priori?, ¿ Cuánta y qué información hidrológica, sintetizada en los PCs, puede ser asimilada en la estructura del modelo hidrológico? Conocer este compromiso entre la complejidad de la estructura del modelo y el contenido de información en los índices permitiría saber dónde han de concentrarse los esfuerzos a la hora de reducir la incertidumbre

total. La Autora de esta tesis doctoral está trabajando en desarrollar una metodología formal para abordar estas cuestiones. Dicha metodología se basa en un análisis de sensibilidad y se alinea bien con el reciente trabajo publicado por Markstrom et al. [2016].

Resultados preliminares

Una primera investigación ha indicado que en algunas cuencas se obtienen predicciones fiables con tan solo un 13% de la información disponible. Además, un análisis de sensibilidad regional [Spear and Hornberger, 1980] demostró que la causa de que en algunas cuencas no se identifique ningún mecanismo dominante es que hay diferentes estructuras de modelos hidrológicos que producen similares vectores (n-dimensionales) de PCs con distintos mecanismos. En otras palabras, el grado de separación en el espacio n-dimensional de los PCs es lo que permite identificar un mecanismo dominante. En la Figura 17 se muestra un ejemplo para una cuenca dónde se identifica un mecanismo dominante para el proceso de enrutamiento; y en la Figura 18 se muestra una cuenca dónde no se identifica ninguno. Al comparar ambas figuras se puede observar que las distribuciones posteriores de los mecanismos para el PC1 están desfasadas una con respecto a la otra en la Figura 17. La distancia de la distribución de probabilidad posterior del mecanismo de no enrutamiento al verdadero PC1 es mayor que la distancia de la distribución de probabilidad posterior del mecanismo de enrutamiento. Este comportamiento no se observa en la Figura 18. A pesar de que el análisis de sensibilidad regional [Spear and Hornberger, 1980] reveló que cada uno de los PCs es sensible (en un mayor o menor grado) al proceso de enrutamiento (Figura 18), se observa que hay estructuras con enrutamiento y sin enrutamiento que reproducen de forma similar las cuatro piezas de información (PCs) a la vez. Cosa que no ocurre en la Figura 17, dónde el grado de separación de las distribuciones posteriores para el vector cuatridimensional de PCs es mayor. Esto probablemente significa que hay una mayor separación en el espacio 4D, tal que el vector de PCs verdadero se encuentra muy distante de los PCs generados por las estructuras sin enrutamiento. Es decir, lo que importa es el grado de separación en el espacio 4D de los PCs, para cada mecanismo que compite. Se propone emplear otros tipos de análisis de sensibilidad global (ver Song et al., 2015) para analizar el impacto de los factores secundarios y sus interacciones.



Figura 17: Sensibilidad del proceso de enrutamiento a la información de cada PC individual. Ejemplo de una cuenca dónde se identifica un mecanismo dominante. Eje x: valor de cada PC. Eje y: $p(PC_i \leq X)$, i=1,4. Curva negra: probabilidad posterior de no enrutamiento.



Figura 18: Sensibilidad del proceso de enrutamiento a la información de cada PC individual. Ejemplo de una cuenca dónde no se identifica un mecanismo dominante. Eje x: valor de cada PC. Eje y: $p(PC_i \leq X)$, i=1,4. Curva negra: probabilidad posterior de no enrutamiento.

Además, la Autora de esta tesis doctoral es co-autora [Addor et al., 2016; 2017] de la metodología que se está diseñando en NCAR (Centro Nacional de Investigación Atmosférica de Estados Unidos, Boulder, Colorado, U.S.) junto a Martyn Clark y Nans Addor para seleccionar los índices hidrológicos en base a su contenido de información. La metodología está siendo desarrollada con 671 cuencas de los Estados Unidos. Se ha comenzado por explorar la posibilidad de obtener los índices hidrológicos únicamente a partir de los datos de descriptores de cuencas. Para ello, ambas variables se han relacionado mediante bosques aleatorios. Además, los índices hidrológicos obtenidos mediante bosques aleatorios se han comparado con los obtenidos usando un modelo hidrológico conceptual. Los hallazgos muestran que los descriptores climáticos tienen la mayor influencia a la hora de predecir los índices hidrológicos tanto con los bosques aleatorios como con el modelo hidrológico. En cambio, las características del suelo no han demostrado ser predictores significativos. Esto ha dado lugar a la cuestión de ¿qué información de los datos de suelos puede ser asimilada en el modelado hidrológico? Por otra parte, se ha observado que los índices con mayor variabilidad espacial son más difíciles de predecir. En base a esto, se propone hacer un ranking de índices basado en su contenido de información y emplear los índices más informativos para condicionar las predicciones hidrológicas y entender la similitud hidrológica.

6.3.3 Error en la estructura del modelo hidrológico

Cómo se señaló en el Apartado 5, ignorar la incertidumbre asociada a la estructura del modelo conduce a una subestimación de la incertidumbre e induce un sesgo en las predicciones. Con base en esto, se propone cuantificar e integrar en la función de verosimilitud el error asociado a la estructura del modelo y a la regionalización [Almeida, 2014]. Así mismo, la estocasticidad de los parámetros del modelo hidrológico y de la varianza de los índices hidrológicos, son otros dos aspectos que también deberían de ser integrados en la función de verosimilitud. Las metodologías desarrolladas en esta tesis están bien planteadas para ello. También, el asimilar en la función de verosimilitud la variabilidad temporal de los índices hidrológicos y/o usar la variable tiempo en lugar de espacio, permitiría conocer el impacto que tienen en las predicciones el efecto del cambio climático y la variabilidad temporal de los usos del suelo dentro del reciente paradigma 'Panta Rhei' (todo fluye) [Montanari et al., 2013].

Por otra parte, se propone aplicar las metodologías desarrolladas en el Apartado 4 para identificar los mecanismos que dominan los procesos hidrológicos a 'la Estructura para Unificar Múltiples Alternativas de Modelado' (SUMMA, Clark et al., 2015a; 2015b; 2015c), que permite representar los diferentes procesos físicos dentro de un conjunto común de ecuaciones de conservación.

6.3.4 Selección y combinación de modelos

Cuando se usan índices hidrológicos para condicionar las predicciones de caudales, es común asumir una distribución a priori uniforme en el espacio de los parámetros del modelo. Sin embargo, la elección de la distribución objetiva a priori que debiera ser empleada en modelado hidrológico, sigue siendo un reto para la comunidad científica [Renard et al., 2011]. Además, cuando se emplean los índices hidrológicos para condicionar las predicciones en entornos con pocos datos, Almeida et al. [2013] demostraron que en lugar de una distribución uniforme en el espacio de los parámetros, en algunas cuencas debería de emplearse una distribución a priori que se traslade en una distribución uniforme en el espacio de los índices hidrológicos (una distribución uniforme en el espacio de los parámetros puede no conducir a una distribución uniforme en el espacio de la variable respuesta - PCs o índices hidrológicos), haciendo que unos comportamientos del sistema hidrológico sean más probables que otros. Esto ha conducido a la cuestión: ¿cuánto más son apoyados los modelos por la información regionalizada cuando se emplea una distribución uniforme en el espacio de los PCs con respecto a cuándo se emplea una distribución uniforme en el espacio de los parámetros del modelo? O, lo que es lo mismo, ¿cuánta información añade cada estructura de modelo acerca de los PCs regionalizados cuando se usa una distribución uniforme en el espacio de los índices hidrológicos con respecto a cuándo se usa una distribución uniforme en el espacio de los parámetros? Para indagar en esto, la Autora de esta tesis doctoral ha propuesto usar BFs [Gelman, 2013; Kass and Raftery, 1995] cómo se describe en la ecuación (13):

$$BF_{uss,ups} = \frac{\iota \left(M_{k} \mid PC^{Reg}, I\right)_{uss}}{\iota \left(M_{k} \mid PC^{Reg}, I\right)_{ups}}$$
(13)
$$L \left(M_{k} \mid PC^{Reg}, I\right)_{uss} = \int_{\Theta_{k}} L \left(PC^{sim}_{\theta_{ik}} \mid PC^{Reg}, I\right) * p\left(PC^{sim}_{\theta_{i,k}} \mid I\right)_{uss} d\theta_{i,k}$$

$$L \left(M_{k} \mid PC^{Reg}, I\right)_{ups} = \int_{\Theta_{k}} L \left(PC^{sim}_{\theta_{ik}} \mid PC^{Reg}, I\right) * p\left(PC^{sim}_{\theta_{i,k}} \mid I\right)_{ups} d\theta_{i,k}$$

Dónde $L(PC_{\theta_{ik}}^{sim}|PC^{Reg},I)$ es la función de verosimilitud de las componentes principales simuladas con el modelo M_k con datos de entrada I y parámetros $\theta_{i,k}$, $p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{ups}$ es una distribución a priori en el espacio de los parámetros del modelo hidrológico M_k y $p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{uss}$ es una distribución a priori en el espacio de los PCs. Se propone usar una función de verosimilitud $L(PC_{\theta_{i,k}}^{sim}|PC^{Reg},I)$ proporcional a la distribución de residuos del procedimiento de regionalización, como se ha realizado a lo largo de esta tesis; una distribución a priori $p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{ups}$ uniforme en el espacio de los parámetros; y una distribución a priori $p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{uss}$ uniforme en el espacio de los PCs [Almeida et al. 2013]. Cómo en general no es posible muestrear directamente de $p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{uss}$, se plantea muestrear uniformemente $\theta_{i,k}$ y disminuir la frecuencia de aquellos conjuntos de parámetros que conducen a $PC_{\theta_{i,k}}^{sim}$ más probables mediante la ecuación (14) [Almeida et al., 2013].

$$p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{uss} \propto \frac{p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{ups}}{p\left(PC_{\theta_{i,k}}^{sim}|\{M_k\},\{\Theta_k\},I\right)}$$
(14)

En la ecuación (14), el denominador $p(PC_{\theta_{i,k}}^{sim}|\{M_k\},\{\Theta_k\},I)$ es la distribución de probabilidad conjunta de los PCs generados por todos los modelos $\{M_k\}$ con parámetros muestreados uniformemente de los correspondientes hipercubos $\{\Theta_k\}$ y datos de entrada *I*. Se propone aproximar esta distribución mediante una mezcla de gausianas [Muller et al. 1996], o histogramas multi-variantes.

Resultados preliminares

Un primer análisis de los BFs calculados de esta manera ha puesto de manifiesto que todas las estructuras de modelos están apoyadas de manera decisiva por la regionalización (ver Jeffreys [1961]) cuando se usa una distribución uniforme en el espacio de los PCs. Es decir, la probabilidad de obtener *PC*^{*Reg*} es más de 100 veces mayor cuando se usa una distribución uniforme en el espacio de los PCs que cuando se usa una distribución uniforme en el espacio de los PCs que cuando se usa una distribución uniforme en el espacio de los PCs que cuando se usa una distribución uniforme en el espacio de los PCs que cuando se usa una distribución uniforme en el espacio de los PCs a llevado a la pregunta: ¿puede hacerse una selección de modelos en base a los factores de Bayes [Jeffreys, 1961] y combinar los seleccionados usando un Promediado de Modelos Bayesiano (BMA, por sus siglas en inglés) [Raftery et al., 2005]? Para esto la Autora de esta tesis doctoral propone usar la ecuación (13) pero usando como referencia (en el denominador) una distribución a priori uniforme en el espacio de los parámetros, como se sugiere en (15). La idea es que esta distribución a priori representa lo que se conoce antes de considerar la información dada por los índices hidrológicos (o los PCs).

$$BF_{uss,ups} = \frac{\int_{\Theta_k} L(PC_{\theta_{i,k}}^{sim} | PC^{Reg}, I)_{*p}(PC_{\theta_{i,k}}^{sim} | I)_{uss}^{d\theta_{i,k}}}{\int_{\Theta_k} p(PC_{\theta_{i,k}}^{sim} | I)_{ups}^{d\theta_{i,k}}}$$
(15)

En base a la ecuación (15), se propone clasificar los modelos, seleccionar aquellos modelos que añaden un conocimiento substancial acerca de PC^{Reg} y combinar los modelos seleccionados usando BMA (16) [Raftery et al., 2005]:

$$p(Q|PC^{Reg}, I) = \sum_{k=1}^{K} p(Q|M_k, PC^{Reg}, I) * p(M_k|PC^{Reg}, I)$$
(16)

dónde $p(M_k | PC^{Reg}, I)$ es la probabilidad posterior de la estructura M_k del conjunto $\{M_k\}$, k=1, N y viene dado por la Ley de Bayes:

$$p(M_k | PC^{Reg}, I) \propto p(M_k | PC^{Reg}, I) * p(M_k | I)_{uss}$$

 $p(M_k | I)_{uss}$ es la distribución de probabilidad a priori del modelo M_k . Una hipótesis clave aquí es que, al menos, una estructura de modelo M_k y sus datos de entrada son capaces de reproducir los PCs verdaderos. La probabilidad posterior del modelo M_k puede re-escribirse como:

$$p(M_k | PC^{Reg}, I) \propto \int_{\Theta_k} L(PC^{sim}_{\theta_{i,k}} | PC^{Reg}, I) * p(PC^{sim}_{\theta_{i,k}} | I)_{uss} d\theta_{i,k} * \int_{\Theta_k} p(PC^{sim}_{\theta_{i,k}} | I)_{uss} d\theta_{i,k}$$

Finalmente, $p(Q|M_k, PC^{Reg}, I)$ es la distribución predictiva posterior de caudales, que se obtiene dados los parámetros y la información regionalizada. Esta distribución puede aproximarse mediante una mezcla de gausianas.

En el caso de que los BFs en (15) seleccionen muchas estructuras de modelos, la Autora de esta tesis sugiere emplear la siguiente estrategia [Vrugt et al., 2011] para seleccionar las estructuras que intervienen en el BMA: primero, muestrear aleatoriamente un número N a partir de las M estructuras de modelos; segundo, calcular el modelo BMA; tercero, obtener una métrica de comportamiento (por ejemplo, el error medio cuadrático); y cuarto, minimizar esta métrica seleccionando el número correcto de combinaciones de N modelos que proporcionen el mejor valor de dicha métrica.

Por último, vale la pena recordar lo siguiente [Jasper Vrugt, 2017, comunicación personal]: *el Promediado de Modelos Bayesiano no dice nada acerca del error epistémico en la estructura del modelo, es una solución de ingeniería a la incertidumbre conceptual. No da indicios sobre qué hay de incorrecto en las hipótesis acerca de dicha estructura. El Promediado de Modelos Bayesiano busca la combinación de estructuras que producen la mejor predicción. Por definición, BMA puede no incluir el mejor modelo. La idea es que la ponderación de modelos 'más débiles'* proporciona mejores simulaciones que la mejor estructura. Sin embargo, para ello, el conjunto de modelos considerado debe de incluir los datos de observaciones. El promediado es un método de interpolación, luego el modelo resultante siempre estará entre los extremos de las predicciones.

6.4 CONCLUSIONES PRELIMINARES OBTENIDAS

6.4.1 Desinformación debida a una mala interpretación de los datos

- Es más fácil detectar eventos desinformativos en cuencas con grandes coeficientes de escorrentía y más difícil en cuencas con condiciones más secas. Esto ocurre tanto para una subestimación de la precipitación, como para una sobre estimación del caudal.
- 2. DYNIA [Wagener et al., 2003] detecta si hubo una tormenta pero no se observa la respuesta en el caudal; o si hubo un hidrograma pero la tormenta que lo generó no se ve registrada en la serie de precipitación.
- La metodología ha de continuarse explorando la aproximación basada en DYNIA [Wagener et al., 2003] con otros parámetros del PDM [Moore, 2007].
- 4. Es necesario evaluar el contenido de información de cada tormenta y la Autora sugiere emplear el marco creado por Beven and Smith [2014].

6.4.2 Contenido de información

- La identificación de un mecanismo dominante es sensible al grado de separación en el espacio multidimensional (conjunto) de la información regionalizada (los PCs regionalizados en el contexto de esta tesis), no a la separación individual.
- 2. Los descriptores climáticos tienen la mayor influencia a la hora de predecir los índices hidrológicos.
- 3. La información en los datos disponibles de suelos no es capaz de ser asimilada en los modelos.
- 4. Los índices hidrológicos con mayor variabilidad espacial son más difíciles de predecir.
- 5. Es necesario establecer un ranking the índices hidrológicos que permita identificar aquellos más informativos a la hora de condicionar las predicciones.

6.4.3 Selección y combinación de modelos

- Los resultados indican que todas las estructuras de los modelos están apoyadas de manera decisiva por la regionalización [Jeffreys, 1961], cuando se usa una distribución univorme en el espacio de los PCs.
- 2. En base a la ecuación (15), se propone clasificar los modelos, seleccionar aquellos modelos que añaden un conocimiento substancial acerca de PC^{Reg} y combinar los

modelos seleccionados usando BMA (16) [Raftery et al., 2005]. Para seleccionar las estructuras que intervienen en el BMA, la autora de esta tesis sugiere emplear la siguiente estrategia [Vrugt et al., 2011]: primero, muestrear aleatoriamente un número N a partir de las M estructuras de modelos; segundo, calcular el modelo BMA; tercero, obtener una métrica de comportamiento; y cuarto, minimizar esta métrica seleccionando el número correcto de combinaciones de N modelos que proporcionen el mejor valor de dicha métrica.

7 REFERENCIAS AL RESUMEN EN CASTELLANO

- Almeida, S., Bulygina, N., McIntyre, N., Wagener, T., and Buytaert, W. (2013). Improving parameter priors for data-scarce estimation problems. Water Resources Research, 49(9), 6090-6095.
- Almeida, S. M. C. L. (2014). The Value of Regionalised Information for Hydrological Modelling. PhD thesis, Imperial College London, London, UK.
- Almeida, S., Le Vine, N., McIntyre, N., Wagener, T., and Buytaert, W. (2016). Accounting for dependencies in regionalized signatures for predictions in ungauged catchments, Hydrol. Earth Syst. Sci., 20, 887-901. doi:10.5194/hess-20-887-2016
- Arora, V. K. (2002). The use of the aridity index to assess climate change effect on annual runoff. Journal of Hydrology, 265(1), 164-177.
- Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. Hydrology and Earth System Sciences Discussions, 11(2), 703-710.
- Beven, K., and Binley, A. (1992). The future of distributed models model calibration and uncertainty prediction. Hydrological Processes, 6 (3), 279-298. doi: 10.1002/hyp.3360060305
- Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modeling. Advances in Water Resources, 16 (1), 41-51. doi: 10.1016/0309-1708(93)90028-E
- Beven, K. J. (2000). Uniqueness of place and process representations in hydrological modelling. Hydrology and Earth System Sciences Discussions, 4(2), 203-213.
- Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of hydrology, 249(1), 11-29.
- Beven, K. (2006), On undermining the science?. Hydrol. Process., 20: 3141–3146. doi:10.1002/hyp.6396
- Beven, K. (2008). On doing better hydrological science. Hydrological Processes, 22 (17), 3549-3553. doi: 10.1002/hyp.7108
- Beven, K. J., Smith, P. J., and Freer, J. E. (2008). So just why would a modeller choose to be incoherent? Journal of Hydrology, 354 (1-4), 15-32. doi: 10.1016/j.jhydrol.2008.02.007
- Beven, K. (2009). Environmental Modelling: An Uncertain Future? Oxon, U.K., Routledge.
- Beven, K., and Westerberg, I. (2011). On red herrings and real herrings: disinformation and information in hydrological inference. Hydrol. Process., 25: 1676–1680. doi: 10.1002/hyp.7963
- Beven, K. and Smith, P. (2014). "Concepts of Information Content and Likelihood in Parameter

Calibration for Hydrological Simulation Models." J. Hydrol. Eng., 10.1061/(ASCE)HE.1943-5584.0000991, A4014010

- Box, G.E.P. (1979). Robustness in the strategy of scientific model building. In: Robustness in Statistics. vol. Academic Press New York.
- Blöschl, G. (Ed.). (2013). Runoff prediction in ungauged basins: synthesis across processes, places and scales. Cambridge University Press.
- Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Research, 36(12), 3663-3674.
- Bulygina, N., and Gupta, H. (2010). How Bayesian data assimilation can be used to estimate the mathematical structure of a model. Stochastic Environmental Research and Risk Assessment, 24(6), 925-937.
- Bulygina, N., McIntyre, N., and Wheater, H. (2009). Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis.
 Hydrology and Earth System Sciences, 13 (6), 893-904. doi: 10.5194/hess-13-893-2009
- Bulygina, N., McIntyre, N. and Wheater, H. (2011). Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. Water Resources Research, 47 (2). doi: 10.1029/2010wr009240
- Bulygina, N., and Gupta, H. (2011). Correcting the mathematical structure of a hydrological model via Bayesian data assimilation. Water Resour. Res.,47, W05514, doi:10.1029/2010WR009614
- Bulygina, N., Ballard, C., McIntyre, N., O'Donnell, G., and Wheater, H. (2012). Integrating different types of information into hydrological model parameter estimation: application to ungauged catchments and land use scenario analysis. Water Resources Research, 48 (6). doi: 10.1029/2011wr011207
- Chow, V. T., Maidment, D. R., and Mays, L. W. (1988). Applied Hydrology, USA, McGraw-Hill.
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44 (12). doi: 10.1029/2007wr006735
- Clark, M. P., Kavetski, D., and Fenicia, F. (2011a). Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research, 47. doi: 10.1029/2010wr009827
- Clark, M. P., McMillan, H. K., Collins, D. B., Kavetski, D., and Woods, R. A. (2011b). Hydrological

field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses. Hydrological Processes, 25(4), 523-543.

- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M. (2015a): A unified approach for process-based hydrologic modeling: 1. Modeling concept. Water Resources Research, 51, 2498-2514. doi: 10.1002/2015wr017198
- Clark, M. P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., and Marks, D. G. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. Water Resources Research, 51, 2515-2542. doi: 10.1002/2015wr017200
- Clark, M. P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W.,, Brekke, L. D., Arnold, J. R., Gochis, D. J., Rasmussen, R.M., Tarboton, D.G., Mahat, V., Flerchinger, G.N., and Marks, D.G. (2015c). The Structure for Unifying Multiple Modeling Alternatives (SUMMA), version 1: Technical description. NCAR Technical Note NCAR/TN-514+ STR, 54 pp., National Center for Atmospheric Research, Boulder, Colo. doi: 10.5065/D6WQ01TD
- Coxon G., Freer J., Wagener T., Odoni N. A. and Clark M. (2014), Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, Hydrol. Process., 28, pages 6135–6150, doi: 10.1002/hyp.10096
- Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and computing, 10(3), 197-208.
- Duan, Q., Ajami, N. K., Gao, X., & Sorooshian, S. (2007). Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in Water Resources, 30(5), 1371-1386.
- Engeland, K., Xu, C.Y., and Gottschalk, L., (2005). Assessing uncertainties in a conceptual water balance model using Bayesian methodology. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 50(1): 45-63.
- Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L. (2006). Is the groundwater reservoir linear? Learning from data in hydrological modelling.Hydrology and Earth System Sciences Discussions, 10(1), 139-150.
- Fenicia, F., McDonnell, J. J., and Savenije, H. H. G. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. Water Resources Research, 44 (6). doi: 10.1029/2007wr006386

Fenicia, F., Kavetski, D., and Savenije, H. H. G. (2011). Elements of a flexible approach for

conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47, W11510. doi: 10.1029/2010WR010174

- Fearnhead, P., and Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74: 419–474. doi:10.1111/j.1467-9868.2011.01010.x
- Friendly M, Monette G, Fox J. Elliptical insights: understanding statistical methods through elliptical geometry. Stat Sci 2013; 28(1):1–39. http:// dx.doi.org/10.1214/12-STS402
- Gallego, G., Cuevas, C., Mohedano, R., & Garcia, N. (2013). On the Mahalanobis distance classification criterion for multidimensional normal distributions. Signal Processing, IEEE Transactions on, 61(17), 4387-4396.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. 2013.Bayesian Data Analysis, 3rd Edition. Chapman & Hall/CRC Press, London. [appendix on Stan]
- Gupta, H. V., Beven, K., and Wagener, T. (2005). Model Calibration and Uncertainty Estimation.
 In: Anderson, M. G. (ed.) Encyclopedia of hydrological Sciences. Hoboken, NJ, John Wiley
 & Sons.
- Gupta, H. V., Wagener, T., and Liu, Y. Q. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes, 22 (18), 3802-3813, doi: 10.1002/Hyp.6989.
- He, Y., Bárdossy, A., and Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. Hydrology and Earth System Sciences, 15 (11), 3539-3553, doi: 10.5194/hess-15-3539-2011.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W.,
 Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta,
 H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A.,
 Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.
 (2013). A decade of Predictions in Ungauged Basins (PUB) a review. Hydrological
 Sciences Journal, 58 (6), 1-58. doi: 10.1080/02626667.2013.803183
- Hochberg, Y. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance. Biometrika, 75(4), 800-802. doi:10.2307/2336325
- Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology, 2204-2214

Jeffreys, H. (1961). Theory of Probability, 3rd ed., Oxford, Oxford University Press.

Kass, R., & Raftery, A. (1995). Bayes Factors. Journal of the American Statistical

Association, 90(430), 773-795. doi:10.2307/2291091

- Kavetski, D., Franks, S.W., and Kuczera, G. (2002). Confronting input uncertainty in environmental modelling, in Calibration of Watershed Models, Water Sci. Appl., vol. 6, edited by Duan, Q., et al., pp. 49–68, AGU, Washington, D. C.
- Kavetski D, Kuczera G. and Franks SW (2006a) Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts, Journal of Hydrology (Special issue on The model parameter estimation experiment - MOPEX), 320(1-2), pp. 173-186.
- Kavetski D, Kuczera G and Franks, SW (2006b) Bayesian analysis of input uncertainty in hydrological modelling: 1. Theory, Water Resources Research, 42, W03407.
- Kavetski D, Kuczera G and Franks SW (2006c) Bayesian analysis of input uncertainty in hydrological modelling: 2. Application, Water Resources Research, 42, W03408.
- Kavetski, D., and Clark, M.P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. Water Resources Research, 46: W10511. doi: 10.1029/2009wr008896
- Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).
- Kuczera, G. (1996). Correlated rating curve error in flood frequency inference. Water Resources Research, 32(7): 2119-2127.
- Kuczera, G., and Mroczkowski, M. (1998). Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. Water Resources Research, 34 (6), 1481-1489. doi: 10.1029/98WR00496.
- Kuczera, G., and Parent, E. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. (1998). Journal of Hydrology, Volume 211, Issues 1–4, November 1998, Pages 69-85, ISSN 0022-1694, http://dx.doi.org/10.1016/S0022-1694(98)00198-X.
- Kuczera, G., Kavetski, D., Franks, S., and Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. Journal of Hydrology, 331(1), 161-177.
- Laio, F., and Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrology and Earth System Sciences, 11 (4), 1267-1277. doi: 10.5194/hess-11-12672007
- Le Vine, N. (2016). Combining information from multiple flood projections in a hierarchical Bayesian framework. Water resources research., 52. doi:10.1002/2015WR018143

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest.R news, 2(3), 18-

22.

- Mantovan, P., and Todini, E. (2006). Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. Journal of Hydrology, 330 (1-2), 368-381. doi: 10.1016/j.jhydrol.2006.04.046
- Markstrom, S. L., Hay, L. E., and Clark, M. P.: Towards simplification of hydrologic modeling: identification of dominant processes, Hydrol. Earth Syst. Sci., 20, 4655-4671, doi:10.5194/hess-20-4655-2016, 2016.
- Marshall, L., Nott, D., and Sharma, A. (2004). A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. Water Resources Research, 40(2).
- Marshall, L., Nott, D., and Sharma, A. (2005). Hydrological model selection: A Bayesian alternative. Water resources research, 41(10).
- McKay, M., Beckman, R., & Conover, W. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics, 21(2), 239-245. doi:10.2307/1268522
- McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T. (2005). Ensemble predictions of runoff in ungauged catchments. Water Resources Research, 41 (12). doi: 10.1029/2005wr004289
- McMillan, H., Krueger, T., and Freer, J. (2012). Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrol. Process., 26: 4078–4111. doi: 10.1002/hyp.9384
- Montanari, A., Shoemaker, C. A., and van de Giesen, N. (2009). Introduction to special section on uncertainty assessment in surface and subsurface hydrology: An overview of issues and challenges. Water Resources Research, 45 (12). doi: 10.1029/2009wr008471
- Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z. and Belyaev, V. (2013b). "Panta Rhei— Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022. Hydrological Sciences Journal, 58 (6), 1256-1275, doi: 10.1080/02626667.2013.809088.
- Moore, R. J. (2007). The PDM rainfall-runoff model. Hydrology and Earth System Sciences, 11 (1), 483-499, doi: 10.5194/hess-11-483-2007.
- Moradkhani, H., & Sorooshian, S. (2009). General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In Hydrological modelling and

the water cycle (pp. 1-24). Springer Berlin Heidelberg.

- Muller, P., Erkanli, A., & West, M. (1996). Bayesian Curve Fitting Using Multivariate Normal Mixtures. Biometrika, 83(1), 67-79. Retrieved from http://www.jstor.org/stable/2337433
- Pechlivanidis, I. G., McIntyre, N. R., and Wheater, H. S. (2010). Calibration of the semi-distributed PDM rainfall–runoff model in the Upper Lee catchment, UK. Journal of Hydrology, 386(1), 198-209.
- Peñas Silva, F. J. (2014a). Clasificación del régimen hidrológico natural y predicción de características hidroecológicas en el tercio norte de la Península Ibérica. Classification of the natural flow regime and prediction of hydroecological characteristics in the Northern third of the Iberian Peninsula.
- Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis, 49(4), 974-997.
- Prieto, C., García, E., Mínguez, R., Álvarez, C., and Medina, R. (2013a). PROPUESTA DE UN MODELO HIDROLÓGICO AGREGADO BASADO EN ECUACIONES GENERALES DE CRECIMIENTO. APLICACIÓN A LA CUENCA DEL RÍO TEIFI (UK), available online: http://www.ingenieriadelagua.com/2004/JIA/Jia2013/pdf/b48.pdf
- Prieto, C., García, E., Mínguez, R., Álvarez, C., and Medina, R. (2013b). Simulación hidrológica de cuencas basada en ecuaciones de crecimiento. Aplicación a la cuenca del río Teifi (UK).
- Prieto, C., Le-Vine, N., García, E., and Medina, R. (2015a) Improving standard practices for prediction in ungauged basins: Bayesian approach. Abstract for the European Geosciences UnionGeneral Assembly 2015, available online: http://meetingorganizer.copernicus.org/EGU2015/EGU2015-9832.pdf
- Prieto, C., Le-Vine, N., García, E., and Medina, R. (2015b). Improving standard practices for prediction in ungauged basins: Bayesian approach. Presentation for the European Geosciences UnionGeneral Assembly 2015, available online: http://presentations.copernicus.org/EGU2015-9832_presentation.pdf
- Prieto, C., Le-Vine, N., Vitolo, C., García, E., and Medina, R. (2016a). Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Abstract for the European Geosciences UnionGeneral Assembly 2016, available online: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-8580-2.pdf
- Prieto, C., Le-Vine, N., Vitolo, C., García, E., and Medina, R. (2016b). Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Presentation for the European Geosciences UnionGeneral Assembly 2016, available online:

http://presentations.copernicus.org/EGU2016-8580_presentation.pdf

- Prieto, C., Le-Vine, N., Vitolo, C., and Medina, R. (2016c). Dominant hydrolocial process identification for ungauged basins: Bayesian approach. Presentation for the European Geosciences UnionGeneral Assembly 2016, available online: https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/160806
- Raftery, A.E. (1993). Bayesian model selection in structural equation models. In Testing Stuctural Equation Models (K.A. Bollen and J.S. Long, eds.), Beverly Hills: Sage, pp. 163-180. Earlier version.
- Raftery, A.E., and Zheng, Y. (2003). Discussion: Performance of Bayesian Model Averaging. Journal of the American Statistical Association, 98, 931-938.
- Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review, 133(5), 1155-1174.
- Reichert, P., and Omlin, M. (1997). On the usefulness of overparameterized ecological models. Ecological Modelling, 95, 289-299. http://doi.org/10.1016/S0304-3800(96)00043-9
- Reichert, P., and J. Mieleitner (2009). Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, Water Resour. Res., 45, W10402, doi:10.1029/2009WR007814.
- Reichert, P., Langhans, S., Lienert, J., and Schuwirth, N. (2015)The Conceptual Foundation of Environmental Decision Support. Journal of Environmental Management 154, 316-332. doi:10.1016/j.jenvman.2015.01.053
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46, W05521. doi:10.1029/2009WR008328
- Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., and Franks, S. W. (2011). Toward a reliable decomposition of predictive uncertainty in hydrological modeling:
 Characterizing rainfall errors using conditional simulation, Water Resour. Res., 47, W11516. doi:10.1029/2011WR010643
- Ribeiro, M, I (2014). Gaussian Probability Density Functions: Properties and Error Characterization. Institute for Systems and Robotics.
- Rougier, J.C. (2010). Formal Bayes Methods for Model Calibration with Uncertainty. In: Applied Uncertainty Analysis for Flood Risk Management. vol. Imperial College Press / World Scientific London.
- Schaake, J., Cong, S., and Duan, Q. (2006). U.S. MOPEX data set. IAHS.

- Schaake, J., Duan, Q. Y., Smith, M., and Koren, V. (2000). Criteria to select basins for hydrologic model development and testing. In: Preprints, 15th Conference on Hydrology, Boston, MA: Amer. Meteorol. Soc, pp. 149-152.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G. (2011). Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. Hydrology and Earth System Sciences, 15 (9), 2895-2911. doi: 10.5194/hess-15-2895-2011
- Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research,46(10).
- Shrestha, D. (2009). Uncertainty Analysis in Rainfall-Runoff Modelling: Application of Machine Learning Techniques. PhD thesis. Delft University of Technology.
- Singh, R. (2013). An uncertainty framework for hydrologic projections in gauged and ungauged basins under non-stationary climate conditions (Doctoral dissertation, The Pennsylvania State University).
- Singh, R., Wagener, T., Crane, R., Mann, M. E., and Ning, L. (2014). A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds: Application to a watershed in Pennsylvania, USA, Water Resour. Res., 50, 3409–3427. doi:10.1002/2013WR014988
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X.,
 McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer,
 D., Uhlenbrook, S., and Zehe, E. (2003). IAHS decade on Predictions in Ungauged Basins
 (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. Hydrological
 Sciences Journal-Journal Des Sciences Hydrologiques, 48 (6), 857-880. doi: 10.1623/hysj.48.6.857.51421
- Snelder T. H., Lamouroux N., Pella H. 2011. Empirical modelling of large scale patterns in river bed surface grain size. Geomorphology, 127: 189-197. DOI: 10.1016/j.geomorph.2010.12.015
- Snelder T. H., Barquin Ortiz J., Booker D. J., Lamouroux N., Pella H., Shankar U. 2012. Can bottomup procedures improve the performance of stream classifications?Aquatic Sciences, 74: 45-59. DOI: 10.1007/s00027-011-0194-7.
- Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. Journal of hydrology, 523, 739-757.

Spear, R.C., and Hornberger, G.M. (1980). Eutrophication in Peel Inlet .2. Identification of Critical

Uncertainties Via Generalized Sensitivity Analysis. Water Research, 14(1): 43-49.

- Steinschneider, S., Polebitski, A., Brown, C., and Letcher, B. H. (2012). Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change. Water Resources Research, 48(11).
- Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., and Srikanthan, S. (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. Water Resources Research, 45: W00b14. doi: 10.1029/2008wr006825
- Todini, E., and Mantovan, P. (2007). Comment on: 'On undermining the science?' by Keith Beven. Hydrological Processes, 21(12), 1633-1638.
- Vrugt, J. A., & Robinson, B. A. (2007). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research, 43(1).
- Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research, 44(12).
- Vrugt, J. A. and Ter Braak, C. J. F. (2011): DREAM(D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems, Hydrol. Earth Syst. Sci., 15, 3701-3713, doi:10.5194/hess-15-3701-2011, 2011.
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S. (2001). A framework for development and application of hydrological models. Hydrology and Earth System Sciences, 5 (1), 13-26. doi: 10.5194/hess-5-13-2001
- Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., and Gupta, H.V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrological Processes, 17(2): 455-476.
- Wagener, T., Wheater, H. S., and Gupta, H. V. (2004). Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, London, Imperial College Press.
- Wagener, T., Sivapalan, M., Troch, P. and Woods, R. (2007). Catchment classification and hydrologic similarity. Geography Compass, 1 (4), 901-931. doi: 10.1111/j.1749-8198.2007.00039.x
- Wagener, T., Sivapalan, M., Troch, P.A., McGlynn, B.L., Harman, C.J., Gupta, H.V., Kumar, P., Rao,
 P.S.C., Basu, N.B., and Wilson, J.S., (2010). The future of hydrology: An evolving science for a changing world. Water Resources Research, 46: W05301. doi: 10.1029/2009wr008906

- Wagener, T. and Montanari, A. (2011). Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. Water Resources Research, 47 (6). doi: 10.1029/2010wr009469
- Westerberg, I.K. (2011). Observational Uncertainties in Water-Resources Modelling in Central America: Methods for Uncertainty Estimation and Model Evaluation. PhD Thesis, Department of Earth Sciences, Uppsala University. Acta Universitatis Upsaliensis, p. 75.
- Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C.-Y. (2011) Calibration of hydrological models using flow-duration curves, Hydrol. Earth Syst. Sci., 15, 2205-2227, doi:10.5194/hess-15-2205-2011.
- Westerberg, I. K., and Birkel, C. (2015), Observational uncertainties in hypothesis testing: investigating the hydrological functioning of a tropical catchment. Hydrol. Process., 29: 4863–4879. doi: 10.1002/hyp.10533.
- Yadav, M., Wagener, T., and Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources, 30 (8), 1756-1774. doi: 10.1016/j.advwatres.2007.01.005
- Zhang, Z., Wagener, T., Reed, P., and Bhushan, R. (2008). Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. Water Resources Research, 44(12)

Chapter 1: Introduction and Background

1.1. CONTEXT

Water has been a critical element for life and the development of our socio-economic system since the beginning of civilisation. Increasingly scarce water resources affect a number of aspects of everyday life, such as water supply, hydropower production, agriculture, tourism, trade, and natural ecosystems. Additionally, hydrological extremes such as droughts and floods can cause the loss of lives and property.

Efficient water management, and the economy based upon it, depends largely on our ability to make reliable predictions of available water resources.

From a scientific point of view, predictions help us to understand how basins function, as well as the different processes that influence the availability of water resources. From an operational point of view, predictions are needed, for example, to make plans for water use and to prioritize water demand (such as for agricultural, industrial, domestic, environmental and recreational use); to calculate the likelihood of flooding and to develop early warning systems (when predictions are made in real time); to design hydraulic structures; to manage dams (for hydroelectric operations); and to evaluate integrated river basin management, including assessments of how changes in the environment (such as changes in climate or land use) can affect the flow regime [Almeida, 2014].

Such predictions are obtained using hydrological models. These models provide a simplified representation of the processes occurring in a basin. They aim to replicate the ways in which these complex systems distribute, transmit, store, and release water [Wagener et al., 2007]. The basin is an open system (with poorly known boundary conditions) in which a multitude of interconnected processes (mostly non-linear) occur on different temporal and spatial scales [Beven, 2009]. A model can never fully replicate the processes that occur in nature [Westerberg, 2011]; however, while all models offer imperfect representations of reality [Box, 1979], some of them can be useful tools for extrapolating water flows in space and time.

Model flow predictions include both gauged and ungauged catchments. In gauged catchments with enough good quality and quantity data, model parameters can be calibrated using standard calibration techniques [Gupta et al., 2005]. However, in ungauged catchments it is necessary to use alternative methods to fill the data gaps – this is the case for the vast majority of the world's basins [Almeida, 2014], either because they do not have enough high quality data on the spatial and/or temporal scale of interest [Sivapalan et al., 2003], or because they do not have any measurements [Blösch et al., 2013]. An additional challenge is that, in many catchments where measurements have been available, governments have stopped collecting water management

data as a consequence resulting from budget cuts. For example, when applying the criteria required to generate the database Experiment Parameter Estimation Model (MOPEX) to Spanish catchments, only 100 out of the 1,000 available natural basins meet the data requirements (a minimum number of rain gauges and more than 10 years of synchronised series of daily precipitation, flow, temperature, and potential evapotranspiration) [Schaake et al., 2000; 2006]. Data quality is the other part of the story, because even with a sufficient quantity of data it may not be of a suitable quality; therefore, 100 basins would be the maximum number of 'basins with measures' in Spain that meet the MOPEX criteria.

To achieve reliable flow predictions in ungauged basins (hereinafter referred to as PUBs) was a key focus for the International Association of Hydrological Sciences (IAHS) in the decade from 2003 to 2012. Priority was given to reducing uncertainty in predictions and making more efficient use of the available information. Despite much progress, the reliability of the predictions provided by hydrological models is still very poor, and PUBs remain one of the biggest challenges within the hydrological paradigm [Hrachowitz et al., 2013].

By using regionalization for PUBs, this thesis represents a step towards more reliable, accurate, and precise predictions for ungauged basins. With that goal in mind, the following subsections review the state of the art, while the subsection 1.2 presents the thesis' objectives.

1.1.1. PREDICTIONS IN UNGAUGED CATCHMENTS

An **ungauged catchment** is a catchment where there is no flow data, data is scarce, or data is not suitable (in quality and/or quantity) for calculating the variable of interest [Almeida, 2014] to the appropriate scale (spatial and/or time), with a required precision for practical application. **The aim of PUBs** is to evaluate flow data and its associated uncertainties in such basins, based on data obtained from weather, climate, soil, vegetation, geology, and topography, including data relating to future changes in climate and land use [Sivapalan et al., 2003]. PUBs involve spatial and/or temporal extrapolation of data from gauged catchments to ungauged catchments, using observations from places and/or periods rich in data to infer data for places and/or periods without measurements. This transfer of knowledge in space and/or time is called **regionalization**. Regionalization is based on the hypothesis that basins with similar climatic and physiographic characteristics have similar hydrological regimes.

Of course, the problem of prediction in ungauged catchments is not new; its origins date back to the 19th century with the rational method [Kuichling, 1889]. However, in 2003 the IAHS provided a new perspective on the problem when it launched the PUBs decade. The Prediction PUBs initiative was a joint and coordinated effort of the International Association of Hydrological Sciences (IAHS) launched in 2003 and concluded by the PUB Symposium 2012 held in Delft (23– 25 October 2012), with the overall objective of making reliable predictions in ungauged catchments.

With this aim, PUBs has focused on quantifying the total uncertainty in predictions by trying to disaggregate the impact of each of the sources of error and developing methodologies to reduce errors through better use of the available data [Sivapalan et al., 2003; Hrachowitz et al., 2013]. Additionally, PUBs has adopted the vision of comparative hydrology [Wagener et al., 2007] to better understand the relationships between the behaviour of river basins and their attributes [Singh et al., 2014], allowing a classification based on the function of the catchment [Wagener et al., 2007] and helping to identify organisational principles. Catchments are systems where climate, soil, topography, and human systems evolve together, so it is necessary to extend the concept of the hydrological cycle to an approach that reflects the real system [Gupta et al., 2008]. The comparative approach involves the development of new capabilities in hydrology to read, interpret, and learn patterns; to develop case studies; and to replace the time variable with the space variable and model the interactions among processes – "Especially the new generation of hydrologists must be prepared to analyse and synthesise" [Wagener, 2010].

PUBs adopted different performance metrics to assess the uncertainty in predictions. In addition to analysing the success of the predictions, it also allows contrasting the methods developed by the different groups involved in the initiative. Despite all the progress made in the PUBs decade (albeit primarily in gauged catchments), PUBs remains one of the major challenges of hydrology [Hrachowitz et al., 2013].

1.1.2. REGIONALIZATION METHODS FOR PREDICTING HYDROGRAPHS IN UNGAUGED BASINS

Traditionally, regionalization methods for calculating flows in ungauged catchments were designed to transfer the parameters of a hydrological model calibrated in gauged basins to ungauged basins [Wagener et al., 2004]. In the ungauged catchments, these parameters were used, together with climate data, to calculate the flows [Bárdossy, 2007]. However, methods based on transferring parameters have several drawbacks; for example, these methods depend on the specific hydrological model used, have the problem of identifiability of parameters, and do not consider the interdependencies among parameters [McIntyre et al., 2005].

Recently, this parameter-based approach has been replaced by an approach that seeks to predict the dynamic behaviour of river basins based on their characteristics [Wagener and Montanari, 2011]. Such behaviour is reflected in a series of hydrological indices (also called

signatures) [Yadav et al., 2007; Bulygina et al., 2009, 2011; Sawicz et al., 2011; Almeida, 2014; Almeida et al.,2016], which are calculated in gauged catchments from the observed data. Once the hydrological indices have been transferred from gauged to ungauged catchments, they are used to constrain the predictions given by one hydrological model (or a collection of models).

Whether parameters or indices are transferred, regionalization methods can be divided into: 1) those which use some form of distance (either geographical – spatial proximity – or functional), and 2) those which use regression [He et al., 2011].

Methods using spatial proximity apply the hypothesis that nearby basins will be similar in physical, hydrological, and climatic characteristics. However, 'due to the uniqueness of the place', catchments that are relatively close in space can have very different behaviours [Beven, 2000]. Methods based on functional similarity assume that basins with similar descriptors will have equivalent behaviour. However, there is no agreement on what metric should be used in order to define the similarity, nor on the catchment descriptors that need to be included. In addition, a review by Almeida [2014] concluded that, for places with many flow gauges, the spatial distance method tends to give better results when the distances for interpolation are small.

Regression-based regionalization methods that transfer parameters utilise the following stages: 1) the hydrological model is selected and calibrated in gauged basins; 2) the best set of parameters, with respect to a performance metric, is retained and related to the catchment's descriptors using statistical models; 3) the parameters are estimated in the ungauged catchments by applying the regression model to its descriptors; and 4) these parameters are introduced into the hydrological model to predict flows in the ungauged catchment. Although the regionalization of parameters with regression methods has been widely used in the past for PUBs [Wagener et al., 2004; Pechlivanidis et al., 2010], this approach is not recommended [Bulygina et al., 2012] because aside from the problems associated with methods based on parameters [McIntyre et al., 2005], the relationships between the parameters of a specific hydrological model (which in itself is a simplification of reality) and the characteristics of the basins are unclear.

The methodology for regression of hydrological indices can be summarised in the following steps: 1) calculate the hydrological indices in gauged catchments using flow measurements; 2) establish regression models between these indices and the catchments descriptors in the gauged catchments; 3) obtain the hydrological indices in the ungauged catchments using the regression equation with the descriptors for the ungauged catchment; and 4) use these indices

for conditioning the predictions generated by a hydrological model. This approach, which combines *a priori* model information with information about the dynamic behaviour of the basin, has proven to be preferable and advantageous for PUBs and improving our knowledge about the functioning of the water system [Wagener and Montanari, 2011; Yadav et al., 2007; Zhang et al., 2008; Almeida et al., 2013, 2016; Bulygina et al., 2009, 2011, 2012; Le Vine , 2016].

1.1.3. SOURCES OF UNCERTAINTY IN PUBS

It is widely accepted that optimal predictions can only be made when all sources of uncertainty have been considered. In PUBs, aside from the three sources of uncertainty present in hydrologic modelling (that is, data, model structure, and parameters), there is a fourth source of uncertainty: regionalization.

Data uncertainty. This type of uncertainty is due to the quality and quantity of both input data (usually precipitation and potential evapotranspiration), and the data used to evaluate predictions (usually flow data). Of these, precipitation is usually considered the largest source of uncertainty [Gupta et al., 2005]; however, there is also significant potential for error in the flow data [Westerberg et al., 2011]. Some typical errors found in data for precipitation and flow series are a constant deviation in the measurements, storms and/or unmeasured events, heteroscedastic autocorrelation and inversely heteroscedastic errors, random errors, temporal shifts, and so on (a catalogue of types of errors and their magnitudes can be found in McMillan et al. [2012]). For example, the uncertainty in precipitation data obtained from rain gauges is dominated by an error in the interpolation. Also, temporal uncertainty can be very important when the available data is on a larger scale with respect to the scale at which the processes occur [Westerberg et al., 2011]. When runoff data is used to evaluate predictions, the forecasts may be affected by significant observational errors because of discharge gauging errors, extrapolation of rating curves, unsteady flow conditions, flow regime hysteresis, and temporal changes in the channel's properties [Renard et al., 2011; Kuczera, 1996]. For instance, McMillan et al. [2012] report that the error range is ±50 to 100% for low flows, ±10 to 20% for medium to high flows in the flood-plains, and ±40% for flows out of the flood plains. This gives an idea of the difficulty of extracting information from observational data. Kavetski et al. [2002; 2006a; 2006b; 2006c] developed a methodology called Bayesian total error analysis (BATEA) to quantify the impact of data uncertainty on the predictions chain.

However, despite it being widely recognised that data uncertainty has a large and long-term impact on flow predictions [Beven and Smith, 2014], the usual practice in hydrological modelling is to assume that the data is 'perfect'. This is mainly due to the difficulty of detecting inaccurate

periods based on the data itself (that is, before using any hydrological model), since errors in the input data are propagated non-linearly throughout the model, and then interact with errors in the model's structure [Gupta et al., 2005; Beven, 2008].

Model structure uncertainty. The hydrology of a catchment involves complex interactions produced by a number of strongly interrelated water, energy, and vegetation-based processes that occur at different spatial and temporal scales [Gupta et al., 2005; Wagener et al., 2007; Westerberg et al., 2011]. Any model is by necessity a simplification of reality. A hydrological model combines different hypotheses for each process that governs a given catchment's functionality [Beven, 2001]; these hypotheses are assigned different degrees of belief [Clark et al., 2011a], and can never be validated due to the lack of full access to the natural phenomenon [Wagener et al., 2004]. A diagnosis or hypothesis test to evaluate the adequacy of the model's structure should be carried out in each basin to verify the suitability of the model structure, that is, whether it provides the reliable and precise predictions that are necessary for its application. Different hydrologists understand processes and their importance differently, depending on their knowledge and experience. Therefore, there is an inherent subjectivity in all modelling [Westerberg, 2011] - and this subjectivity is what makes a scientist an expert in his or her field [Rougier, 2010]. Because of data scarcity, it is common practice to use an *a priori* single preselected model structure to represent the processes ('one-size-fits-all' [Fenicia et al., 2008]) under the assumption that the model is 'perfect' [Wagener and Montanari, 2011], regardless of the widely acknowledged 'uniqueness of the place' paradigm [Beven, 2000] and the inability of a single model structure to provide a suitable description of hydrological systems in all basins [Clark et al. 2011a; 2011b; 2015a; 2015b; 2015c]. Such ignorance of model structural error may lead to predictive uncertainty underestimation and leave predictions prone to bias [Reicherte and Omlin, 1997].

When there is little knowledge and data to support the selection of a single model structure, the use of different structures is critical for estimating uncertainty [Clark et al. 2011a; 2011b; 2015a; 2015b; 2015c]. There is no obvious justification for the selection of a single model as the 'true' description of a system. Typically, each different structure describes the system in an appropriate way, each with its own characteristics and limitations, and there are features of the real system that will not be covered by any of the hypotheses. All models are 'wrong', but some of them are useful [Box, 1979].

Flexible and adaptive approaches have been developed to improve on this situation (for example, the rainfall-runoff modelling toolkit, RRMT [Wagener et al., 2001; 2004]; the

framework for understanding structural errors, FUSE [Clark et al., 2008; 2011b]; FLEX and SUPERFLEX [Fenicia et al., 2006; 2008; 2011]; and the structure for unifying multiple modelling alternatives, SUMMA [Clark et al., 2015a; 2015b; 2015c]). These approaches provide multiple mechanisms for each hydrological process, breaking down a hydrological model into a set of testable hypotheses [Clark et al., 2011; Fenicia et al., 2008; Clark et al., 2015a; 2015c].

Multi-model ensemble-based forecasting methods overcome some of these deficiencies by combining the information from the hypothesis included in the ensemble (trying to cover the feasible hypothesis space, where the true hypothesis belongs) and, thus, representing the impact of the structural uncertainty in the flow predictions [Clark et al., 2008]. Each of the models in the ensemble is weighted according to its ability to reproduce the system's behaviour (synthetized in regionalized information in the ungauged basin). However, it is essential to ensure that the ensemble includes at least one hypothesis that approximates 'reality' within the range of the other sources of uncertainty – that is, data and regionalization uncertainty. Ensemble-based forecasting methods have become popular in recent decades in disciplines such as economics [Hoeting et al., 1999], sociology [Raftery et al., 1995], and weather forecasting [Raftery et al., 2003; Chandler, 2013] (the key differences to build the ensembles lie in the weighting strategy).

Finally, model-based uncertainty also results from its mathematical implementation, the temporal and spatial discretisation, and/or the numerical scheme used to solve the equations and numerical errors [Kavetski and Clark, 2010].

Parametric uncertainty. This type of uncertainty precludes identifying a global optimum (a nonidentifiability problem) or better set of parameters according to an evaluation criterion. It can be caused by data uncertainty, model structure uncertainty, the absence of data over long periods of time, the inefficiency of the optimisation algorithm [Almeida, 2014], the interdependence between parameters [Kuczera and Mroczkowski, 1998], or the fact that no single optimal set of parameters can support the need to use set theory-based approaches, in which all sets of suitable parameters are retained until proven otherwise. The inability to infer some or all quantities of interest from the available data is often referred to as 'no identifiability' [Wagener et al., 2001], and unless prior knowledge is available, no identifiability leads to an 'illposed' inference [Renard et al., 2011].

Parametric uncertainty has been extensively reported in the literature (for an example, see Gupta et al., 2005). Traditional strategies to calibrate parameters were based on least-squares regression; however, the strong axioms necessary for the least-squares approximation are rarely

satisfied in hydrological modelling [Kavetski et al., 2006]. The least-squares approximation assumes that the errors of the output variable are independent, and that they are normally distributed with zero mean and constant variance. In many cases in hydrology, one witnesses correlated errors – errors that do not follow a Gaussian distribution and whose variance is not constant [Beven, 2009]. This has led hydrologists to abandon the methods of classical statistical inference and the use of other approaches, such as those based on simulation and sampling (for example, Monte Carlo, MC; Latin hypercube sampling, LHS; and Markov chain Monte Carlo, MCMC). These approaches include Bayesian statistics [Kuczera and Parent, 1998]; Regional Sensitivity Analysis (RSA [Spear and Hornberger, 1980]); Generalised Likelihood Uncertainty Estimation (GLUE [Beven and Binley 1992]); Dynamic Identifiability Analysis (DYNIA [Wagener et al. 2003]); and fuzzy sets [Westbergerg et al., 2011]. These approaches have proven to be advantageous in hydrological modelling [Wagener and Montanari, 2011]. Thanks to recent advances in MC analysis, numerical methods and the increasing computational capacity of computers, Bayesian inference can be used efficiently nowadays [Kavetski et al., 2006].

These three sources of error, present in all modelling, result in an equifinality of representations that is consistent with the observed flow and induce uncertainty in the predictions given by the model [Beven, 1993; 2009]. Historically, the only source of uncertainty that has been addressed is parametric uncertainty, ignoring the other sources of uncertainty [Beven, 2009]. However, data and model structure uncertainty usually exceeds parametric uncertainty [Beven, 2009; Kuczera et al., 2006]. It is also recognised that structural error is harder to characterise than data uncertainty [Renard et al., 2010], because the latter can be estimated by analysing the model's sampling and measurement design [Refsgaard et al., 2009].

Regionalized hydrological indices uncertainty. The fourth source of uncertainty in PUBs is the uncertainty due to the regionalized hydrological indices [Almeida, 2014; Westerberg et al., 2016], which is generated by the following: 1) uncertainty in the indices calculated for the gauged catchments, which includes uncertainty due to the amount of data available for the gauged catchments, as well as uncertainty due to the quantity and quality of the information contained in those indices; 2) a lack of complete certainty that the knowledge acquired in the gauged catchments is applicable to ungauged catchments; 3) uncertainty from the regionalization concepts, since it implicitly assumes i) the comparability of measures between basins, which is not strictly true since errors in data vary spatially [McMillan et al., 2012], and ii) that catchment descriptors describe gauged and ungauged catchments similarly; 4) uncertainty from the regionalization procedure used to transpose information (for example, the formal

variant of the regression, when regression is used for indices regionalization); and 5) uncertainty stemming from the catchment descriptors used in the regionalization function.

In short, predictions in ungauged catchments are affected by uncertainty stemming from modelling (data – both input and output – model structure, and parameters) and regionalization.

Various studies have reiterated the importance of disaggregating the different sources of uncertainty in hydrological modelling [Kavetski et al., 2006a, 2006b; Thyer et al., 2009; Renard et al., 2010] in order to provide better solutions for each of them [Kuczera et al., 2006; Sivapalan et al., 2003]. Increasingly, authors claim the need for a strategy to quantify the suitability of the model structure for a given objective [Clark et al., 2008, 2011; Wagener et al., 2001; Bulygina and Gupta, 2010; Fenicia et al., 2008, 2011] and to assess the quality of observational data, both inputs introduced to the models and data of observations used to evaluate these models [Kavetski et al., 2002, 2006; Renard et al., 2011; Beven and Westerberg, 2011; McMillan et al., 2012; Westerberg and Birkel, 2015]. Reducing the total predictive uncertainty requires a robust quantitative understanding of each of its sources. In hydrology, robust characterisation of the uncertainties affecting rainfall-runoff models remains a major scientific and operational challenge, and there are still difficulties in developing statistical techniques for achieving uncertainty decomposition [Renard et al., 2010]. However, when regionalization is used for PUBs, a first objective must be to separate the uncertainty of the model structure/input/sampling from the uncertainty due to the process of regionalization.

1.1.4. DESCRIBING UNCERTAIN KNOWLEDGE

PHILOSOPHICAL BASIS OF USING PROBABILITIES TO DESCRIBE UNCERTAIN KNOWLEDGE

All the above sources of error are affected by aleatory and/or epistemic uncertainty. And these uncertainties are to be characterised and quantified via probability theory. This sub-section provides a summary of the different interpretations of the mathematical framework of probability across history as a philosophical basis of natural science and scientific reasoning. Such different approaches to probability have been due to the human being's desire to know with certainty the events that will happen. A usual classification of the interpretations is one that categorise them into I. objective, II. evidential, and III. intersubjective probabilities (see [Reichert et al. 2015]).

<u>I. The objective interpretation</u> uses probabilities to describe features of the material world that are independent from humans [Reichert et al. 2015]. Probabilities are associated with the randomness of the natural process itself (i.e. with random physical systems), in which an event

tends to occur at a persistent rate in a long run of trials and is therefore irreducible. Followers of the objective interpretation try to explain these stable relative frequencies. They are usually divided into frequentist [Von Mises, 1928] and propensity [Popper, 1959] approaches.

<u>The frequentist stream</u> uses aleatory probabilities [e.g. Fisher, 1922; 1950; Neyman, 1937; Kolmogorov, 1933]. They use probability as the limit of relative frequencies of events in a repeatable experiment (assuming that this limit exists). Probabilities are used to quantify the ignorance that exists every time an experiment is repeated, and it can be quantified by probability distributions – the discrepancies between the 'true' value and the obtained value follow a statistical structure that can be described by probability theory [Beven and Smith, 2014]. The main problem is that only a finite number of repetitions of the random process are available (it is of course impossible to perform the process an infinite number of times) to determine the probability of an event. Hence, erroneous relative frequencies will be obtained for each experiment (as these probabilities will be different from each other, but the true probability should be the same). These errors can only be characterised by probability density functions, which is what one is trying to obtain – therefore, it is a circular problem.

<u>Propensity or chance</u> [e.g. Popper, 1959] intends to make objective probabilities applicable to single events by emphasising the circumstances or causes of a single event that could make it repeatable [Reichert et al., 2015]. Propensity is a physical disposition of a physical situation to yield an outcome of a certain kind, or to yield a long-run relative frequency of such an outcome. Propensities are not relative frequencies, but they try to explain the assumed causes of the observed stable relative frequencies by appealing to the large numbers law, which allows that stable long-run frequencies are a manifestation of invariant single-case probabilities. Note that this differs from the frequentist approach, as relative frequencies only exist for a large ensemble of events, i.e. not for single ones. The main problem with this approach is defining propensity and then showing that propensity has the required properties; these two requirements have not been successfully met yet.

II. Followers of evidential probability use probabilities to represent the degree to which one statement is supported by the evidence. Evidential probability interpretations can be divided into logical [e.g. Keynes, 1921]; epistemic, inductive, or subjective [Ramsay, 1926; De Finetti, 1931; Cox, 1946]; Bayesian; and intersubjective [Gillies, 2000].

<u>Evidence probabilities</u> use the same mathematical probability theory as objectives to describe degrees of belief. Therefore, it is not related to the physical randomness; when is said that one model or hypothesis is true, this means that the evidence supports this hypothesis to a high
degree. These probabilities try to characterise our limited ability to represent the system, i.e. our lack of knowledge about the system or lack of resources such as data and time. It also includes what we know that we do not know (the 'known unknowns'), as well as what we do not know (the 'unknown unknowns') [Beven et al., 2011].

Logical theory states rational relations between propositions. They are degrees of coherent consequence and assume that given the same evidence, different individuals will independently come to the same degree of belief [e.g. Keynes, 1921].

Subjective probabilities: The fact that it is unlikely that all individuals will come to the same probability given the same knowledge led to the development of the **subjective probabilities** [e.g. Ramsey, 1926; De Finetti, 1931]. Here, probabilities describe individual degrees of belief, which can be different for different individuals given the same evidence.

Bayesian probabilities: Bayes [1763] and Laplace [1774] developed the Bayesian framework independently in the 18th century, although the former was more interested in the inverse probability and the latter in the data. Both developed what is known as Bayes' Law. However, in the 1980s there was a renaissance in research on and applications of Bayesian methods due to the Markov chain's Monte Carlo method and the interest in non-standard, complex applications; and some of the major contributors to 20th century Bayesian approaches were de Finetti [1937], Jeffreys [1939], Savage [1951] and Lindley [1953]. In this paradigm, probability quantifies the uncertainty or the degree of belief to which the hypothesis is supported by the evidence. These degrees of belief follow the laws of probability [Cox, 1946]. For Bayesians, both frequency probabilities (which are associated with random physical systems) and evidential probabilities are necessary. Bayesian probabilities can be assigned to any statement whatsoever, even when no random process is involved, as a way to represent its plausibility. The main differences with frequentists is that frequentists only recognise frequency probabilities; while parameters are fixed variables for frequentists, for Bayesians parameters have a probability structure as well.

The use of Bayesian probability involves specifying a prior probability. Bayesians defends that this prior state of the knowledge makes the problem well posed (Benjamin Renard [2016], personal communication). How to define the prior distribution has led to two schools: 'objectivist' Bayesians e.g. (Laplace, 1774; Lhoste, 1923; Jeffreys, 1939) and 'subjectivist' Bayesians (e.g. Keynes, 1921; Ramsey, 1926; De Finetti, 1937; Savage, 1951; Rubin, 1949; Berger, 1985; Bernardo and Smith, 2004).

However, it is worth mentioning that Bayes [1763] and Laplace [1774] were 'objectivist' Bayesians [Fienberg, 2005]. In this stream, probability objectively measures the plausibility of

propositions; the statistical analysis only depends on the model assumed and the data analysed. No subjective decisions need to be involved. This approach uses uniform priors, following Laplace's principle of insufficient reason (also called the principle of indifference), and is usually called inverse probability because it infers backwards from observations/evidence to parameters, or from consequences to causes [De Morgan, 1837]. It is remarkable that Pearson [1892] adopted the Laplace inverse probability, but Pearson also allowed for the role of experience in determining the priors. The question confronting Bayes and Laplace was how to choose the prior so as not to bias their inferences [Fienberg, 2005]. Laplace [1774] supported a flat (uniform) prior; however it was demonstrated that a uniform distribution of the prior parameters might yield a non-uniform density for some one-to-one functions of the parameters [Lhost, 1923; Jeffreys, 1939]. Obviously, not encoding ignorance in these non-uniform distributions (see the next section) - it makes little sense to have ignorance in one but not in the others. This led Lhost [1923] to develop vague prior distributions and similarly, it led Jefferys [1939] to develop the invariance principle. While much attention has been paid to improving the Laplace's work by developing methods for specifying 'objective' priors (default or ignorance), defining prior distributions reflecting ignorance is still 'the quest of the statistical holy grail' (see [Fienberg, 2005; Berger, 2006]). Jeffreys' work [1939; 1961] has been highly relevant for 'objective' Bayesians. Jeffreys [1939; 1961] laid out the inverse probability approach of updating degrees of beliefs in propositions by use of Bayes' theorem to learn from experience and data and used an information-like invariance approach to derive 'objective' prior distribution that expressed ignorance; and most of the applied Bayesian analyses follow Laplace-Jeffreys objective school [Berger, 2006].

In contrast, for 'subjective' Bayesians, probability quantifies a personal belief. They deny the possibility of fully objective analysis, claiming that it is misleading to say that any statistical analysis can truly be objective [Berger, 2006]. Goldstein [2006] tells that 'subjective' Bayesian approach is based on the following simple idea. 'You are uncertain about many things in the world. You can quantify your uncertainties as probabilities, for the quantities you are interested in, and conditional probabilities for observations you might make given the things you are interested in. When data arrive, Bayes' Law tells how to move from prior probabilities to new conditional probabilities for the quantities of interest. If you need to make decisions, then you may also specify a utility function, given which your preferred decision is that which maximises expected utility with respect to your conditional probability distribution'.

In the 'subjective' stream, Keynes [1921] stated that the degree of belief might not be numerically measurable. Ramsey [1926] employed the concept of expected utility to approach

subjective probability. De Finetti [1937] introduced the notion of exchangeability (which was exploited to develop hierarchical Bayesian methods decades later); and the implicit role of prior distributions (which has given later statisticians the tools to study the implications of the use of improper prior distributions). Further, as it was difficult to make detailed specifications of prior probabilistic beliefs, artificial conventional priors were used, but this priors had only a limited relation to the actual judgements. This led De Finetti [1974] to use expectation rather than probability for the subjectivist theory. Savage's work [1951] relied on ideas of personal probability from de Finetti's work; Savage [1951] developed a non-frequentist alternative to the Kolmogorov axioms, leading to the constructive methodology of maximizing expected utility. During the post-World War II period, one major research thrust was rational decision making; Rubin [1949] worked in this topic and wrote about subjective ideas.

However, 'subjective' priors are problematic for science, as well as for some practical applications. For example, decision-makers may lack the knowledge or time to specify an informed probability distribution. Moreover, it is difficult to move form principles to practice in the subjective approach [Goldstein, 2006]. Berger [2006] highlights that objective Bayesian analysis should be adopted in scenarios requiring 'objectivity', the author also notes that 'objective' Bayesian methods are promising to unify Bayesian and frequentist statistics. On the other hand, however, finding the right method for constructing 'objective' prior continues to be one of the main challenges for the research community [Renard et al. 2011]. Therefore, Bayesian statisticians usually need either to use informed priors using relevant expertise from previous data or to construct objective priors [Fienberg, 2005].

III. The intersubjective interpretation of probability extends Ramsey and De Finetti's argument to groups sharing a common goal [Gillies, 1991]. It claims a pluralist view of probability that uses different interpretations in different contexts [Gillies, 2000], using epistemic probabilities when describing individual beliefs and human behaviour (or the best available scientific knowledge) and objective probabilities in the natural sciences. In the review provided by Reichert et al. [2015] the authors advocate for using intersubjective probabilities, stating that the uncertainty of the outcome of a perfectly known system affected by randomness can be characterised by objective probabilities. Once the random event is realised, but the outcome has not yet been observed, uncertainty becomes epistemic. Bayesian inference provides a consistent description of an iterative learning process: updating prior information with evidence. As individual beliefs are uncertain and a group of beliefs may not reach consensus, these uncertainty intervals can be constructed from the replies. This information can be used to construct sets of probability

distributions also known as imprecise probabilities. When used as prior distributions in Bayesian inference, this leads to so-called robust Bayesian analysis.

NON-PROBABILISTIC APPROACHES TO DESCRIBE UNCERTAIN KNOWLEDGE

The use of probability theory as the most adequate framework for describing epistemic uncertainty has been challenged. The main criticism is that typically, the available knowledge is insufficient to define an appropriate distribution for characterising errors (for example, the joint probability distribution of the errors is very difficult to characterise, or they are not stationary). There may also be other practical issues such as time constraints for providing the predictions, and so on. In addition, using overly simple distributions leads to an overestimation of the information contained within the data [Beven et al., 2008]. A good review of alternative approaches to considering this ambiguity is found in Beven [2008]. In these cases, researchers have used non-probabilistic (also called possibility) methods [Montanari et al., 2009], such as evidence theory. Similar to imprecise probabilities, vague variables are defined in intervals in a way that can be used for aggregating multiple attributes within a decision tree or belief network. It allows for ignorance using two measures: a support measure and a plausibility measure. And the method combines evidence using different rules to arrive at a degree of belief (belief function), so this accounts for all the available evidence [Dempster, 1967; Shafer, 1976]. Possibility theory is based on fuzzy sets sets theory [e.g. Westerberg 2011], in which potential elements have a degree of membership between 0 and 1. Membership functions seem to be similar to probability densities, but they use a different normalisation and a different calculus. Interval analysis characterises unknown quantities by intervals, without specifying probability distributions within the intervals. These intervals are then propagated through functions allowing for all possible combinations of values in the intervals [Moore, 1979]. Finally, the nonprobabilistic method most used for uncertainty assessment in hydrology is the generalised likelihood uncertainty estimation (GLUE) [Beven et al., 1992]. Because of its long tradition in hydrology, a description of the method is provided below.

Although these methods use the philosophy of the Bayesian approach, none of these alternative theories has a similarly good axiomatic foundation as probability theory [see discussions in Mantovan and Todini, 2006; Todini and Mantovan, 2007; Beven, 2006; 2008; Beven et al. 2007; Beven et al. 2008], and therefore the predictions and uncertainty cannot be interpreted statistically. However, the argument that ambiguity about the correct probability distribution can hardly be represented by probabilities is recognised. For this reason, Reichert et al. [2015] suggest using intersubjective probabilities to describe scientific knowledge in the absence of significant ambiguity, and imprecise, intersubjective probabilities otherwise.

<u>GLUE</u>

The Generalized Likelihood Uncertainty Estimation method (GLUE) [Beven and Binley, 1992] is based on the concept of equifinality [Beven and Freer, 2001]. That is to say, within the same model structure, there are many possible sets of parameters that can provide acceptable representations of the system. Same applies for model structures, if more than one are employed. The procedure is carried out as follows:

- 1) Elect a hydrological model M (or more) with parameters θ .
- 2) Generate multiple parameter sets from the prior distributions (generally uniform distributions) by Monte Carlo sampling methods.
- 3) Introduce these parameters into the model M.
- 4) Compare the simulations given by each set of parameters with the observations.
- 5) Assign a weight to each of the parameters depending on how well the model reproduces the system response in terms of a criterion of goodness of fit (for example, by using Nash-Sutcliffe efficiency coefficient [Nash and Sutcliffe, 1970]).
- 6) Use these weights to build what Beven and Binley [1992] defined as a 'likelihood metric', which expresses the beliefs about the predictions given by a model/hypothesis; all simulations with a likelihood metric greater than a predefined threshold are retained, and the others are assigned a zero likelihood metric.
- 7) Rescale the likelihood metric associated with these parameters so that they add up to1.
- 8) Finally, use the retained models in predictions in which each simulation is assigned a weight equal to the rescaled likelihood metric, resulting in a cumulative distribution of predictions. From these distributions, it is possible to calculate uncertainty limits for the elected percentiles (conditional on the structure of the model M used, the ranges of the parameters, input and output data, and the likelihood metric used).

This method assumes that the errors associated with the retained models remain the same for predictions, so the 'likelihood metrics' obtained in the calibration can be used for prediction. This 'likelihood metric' (which is not a formal likelihood function) leads to prediction intervals without statistical interpretation, and has been criticised by many of the hydrological community [Mantovan and Todini, 2006; Todini and Mantovan, 2007; Stedinger et al., 2008], because in order to interpret the results in statistical terms, a formal likelihood function must be used. However, Beven and Smith [2014] point out that the methodology is general in the sense that, if it is not possible to define a likelihood function based on the residual distribution,

GLUE makes it possible to consider a wider range of likelihood metrics, such as fuzzy sets metrics, and it must not be given a statistical interpretation. The only requirement is that the 'likelihood metric' increases monotonically to improve the behaviour of the model, and those models that are not appropriate have a 0 likelihood metric.

GLUE is a widely used methodology due to the fact that it is easy to understand and implement. However, it requires a vast number of simulations to cover the multidimensional model space, which makes it very expensive computationally when the hydrological model has many parameters.

1.1.5. BAYESIAN UNCERTAINTY QUANTIFICATION

To support decisions in environmental management, one needs probabilistic predictions of the outcomes. There are two types of problems to solve in order to quantify the uncertainty in hydrological forecasts: 1) the so-called 'inverse problem', that is, model uncertainty quantification and parametric uncertainty quantification and 2) the propagation of the different sources of uncertainty through the model to predict the total uncertainty in the system response [Moradkhani, and Soroshian, 2009]. Predictions in ungauged catchments include both. First, parameters are estimated based on the regionalized information and, afterwards, flow predictions are obtained, propagating all sources of uncertainty through the model.

All methods of quantifying uncertainty are based, implicitly or explicitly, on a set of hypotheses about the discrepancies between predictions/simulations and observations. When it is possible to define a density function of these differences (residuals), Bayesian statistics has proven to be the most appropriate approach for describing a system's non-linearities and uncertainties quantification (epistemic and random) in the predictions [Engeland et al., 2005; Marshall et al., 2004; Renard et al., 2010; Bloch et al., 2013; Kavetski et al., 2002; Vrugt et al., 2008; Renard et al., 2010; Bulygina et al., 2009; Almeida, 2014]. As noted above, unlike the frequentist approach that is only based on the evidence, the Bayesian framework provides a formal mechanism for combining evidence with *a priori* knowledge about the parameters; or in other words, updating *a priori* knowledge with information/data.

Given the observed data *D*, a model structure *M*, and inputs *I*, the posterior distribution of the model parameters θ_M is provided by Bayes' theorem:

$$p(\theta_M | D, M, I) = \frac{L(\theta_M | D, M, I) * p(\theta_M | M, I)}{\int_{\Theta_M} L(\theta_M | D, M, I) * p(\theta_M | M, I) * d\theta_M}$$
(1.1)

where **<u>D</u>** is the evidence, information, data</u>, or observations used to infer the model parameters. In hydrological modelling, *D* usually represents the observations of the flows. As in

the PUBs context, there is no data for flows; the evidence is given by various types of information used to condition the predictions (i.e. the regionalized information).

 $L(\theta_M | D, M, I)$ is the likelihood function. It is a function of the parameters θ_M given D. It can be interpreted as the probability that the simulations Y_{sim} given by the model M with parameters θ_M represent D. In practice, the likelihood is expressed as a function of the residuals between observed and simulated values, $E(\vartheta)=D$ -Ysim, which only implies that the likelihood is shifted (but its values are not affected). To define this joint likelihood of the residuals (and thus parameters, ϑ), $L(E(\vartheta))$, the statistical distribution of the residuals must be estimated, or an assumption regarding its statistical distribution must be made [Vrugt and Massoud, 2017]. In an ideal case with a perfect model and inputs, it is expected that the residual distribution will perfectly match the distribution of the errors of the system response. However, the impact of model structural and inputs uncertainty is, in general, much larger than the error in observations [Renard et al., 2011]. This has led to the proposal of non-traditional residual distributions. Two main approaches to improving the handling of non-traditional residual distributions are found in the literature. The first approach characterises the likelihood by means of the probabilistic properties of the residuals. For example, Kavetski et al. [2006b; 2000c] and Renard et al. [2010] used nuisance variables to treat input uncertainty explicitly, and time-variable parameters [Reichert and Mieleitner, 2009] have been used to represent errors in the model structure and forcing data. Recently, in the PUBs context, Almeida [2014] derived the likelihood function based on the residual distribution of the regionalization model. The second approach uses the so-called likelihood-free defined by Vrugt and Sadegh [2013], which relies on Approximate Bayesian Computation (see below).

 $p(\theta_M)$ is the prior distribution of the model parameters; it represents the uncertainty about the parameters before introducing any information given by the observations (evidence). When there is prior knowledge, the prior distribution is called informative. If the prior distribution is conjugate to the likelihood function, the posterior distribution is of the same type as the prior. In these cases, the posterior possesses a closed solution without numerical integration. When there is no prior information, weak informative priors are commonly used, so the calibration is conducted by the data. In hydrological modelling, it is common to use a uniform prior distribution in the parameter space based on feasible model parameter ranges [Bulygina et al., 2011]. The limits of the prior must be wide enough to include the area where the likelihood function is appreciable and the interval where the posterior is defined; the posterior is a subset of the prior. Thus, under the assumption of a uniform prior distribution for the priors, different parameter samples are equally probable a priori. However, a uniform distribution on the

parameter space represents our lack of knowledge about the model structure, but not about the system (synthetized in the hydrological indices). Therefore, ignorance about the catchment behaviour is best defined in terms of a uniform prior distribution in the system's behaviour space [Almeida et al., 2013]. However, significant challenges remain in the development of statistical techniques for achieving the prior distribution knowledge necessary for a meaningful and wellposed inference [Renard et al., 2010] (see the difference between objectivist and subjectivist Bayesians in the previous section).

<u>The denominator</u> in equation (1.1) is the marginal distribution of observations, the so-called marginal likelihood or model evidence. This integral is a normalization constant (with respect to parameters), which ensures that the posterior distribution adds up to 1. It is of no particular interest for parameter estimation, but it is of crucial relevance in hypothesis or model testing (see the end of the section).

When the prior distribution is conjugated for the likelihood function, the posterior distribution is in the same family as the prior, which gives a closed-form expression for the posterior. However, generally, prior distributions are not conjugated, and posterior distributions are multidimensional and do not follow a known distribution, so the closed analytical solution cannot be used. In other words, the theory linking data with the model parameters is often nonlinear and the probability distribution in the model space is usually difficult to define (it is multimodal, and some points are not defined, for example (e.g. Beven, 2009). As a result, in most cases, the posterior cannot be calculated analytically and numerical integration is needed. In order to deal with these cases, various numerical techniques have been developed to approach the posterior distribution. They can be classified as 1) those that require the calculation of the density of the likelihood function, and 2) those that require sampling from the likelihood function. The approaches of the first type necessitate the evaluation of the posterior distribution for many sets of parameters θ_{M} , which are numerically generated by methods such as Importance Sampling [Kuczera and Parent, 1998] or Markov Monte Carlo Chains (MCMC) methods. However, when it is impossible to calculate the likelihood function but it is feasible to sample from it (that is, it is easy to simulate from the models), the second type of approach can be implemented; rejection algorithms such as Approximate Bayesian Computation (ABC) are among these kinds of techniques. ABC replaces calculations of the likelihood function with one step, which involves simulating artificial data Y_{sim} from different parameter values θ_M and comparing summary statistics of the simulated data with those of the observed data. In other words, the procedure requires that first, a sample set of parameters θ_M is acquired from the prior distribution; second, an output must be drawn from the model $L(\theta_M | D, M, I)$, given the

set of parameters θ_M ; and third, θ_M must be accepted or rejected based on how far the simulated data Y_{sim} is from the observed data D. Thus, an estimate of the likelihood function is combined with a prior distribution to produce an approximate posterior distribution [Fearnhead and Prangle, 2012]. Although ABC spares the modeler from evaluating the likelihood function $L(\theta_M | D, M, I)$, it still requires the ability to sample from the likelihood function, which entails that the likelihood function needs to be defined.

The approach that requires the calculation of the likelihood function has two major limitations: 1) it is difficult to sample the uncertainty variables of the posterior distribution $p(\theta_M | D)$ unless this is a multivariate normal distribution [Kuczera and Parent, 1998], and 2) it requires a large number of simulations (both to ensure the convergence criterion and to verify that it has sampled all the parameters (possible behaviours of the basin). To reduce the number of simulations, sampling techniques such as LHS [Mckay et al., 1979] and MCMC have been used. MCMC methods allow efficient sampling from a probability distribution based on building up a Markov chain that has the posterior distribution as a target (stationary) distribution (see discussions in Schoupus and Vrugt [2010], Engeland et al. [2005], Kuczera and Parent [1998] and Steinschneider et al. [2012]).

The disadvantages linked to the approach that requires sampling from the likelihood are related to the tolerance for rejection and the summary statistics used to define the residuals. Setting the tolerance parameter to zero ensures an exact result, but it makes computations prohibitively expensive, while using values larger than zero introduces a bias. With respect to the summary statistics (sufficient statistics that are needed to reduce the output dimension), these are typically not available, and instead, other summary statistics are used, which introduces an additional bias due to the loss of information.

Once the posterior distribution has been calculated, it can be used to estimate the predictive probability distribution of the variable of interest. For example, flows can be simulated by sampling sets of random parameters θ_M from the joint posterior probability density function, such that for each time step a posterior distribution is achieved, which can be evaluated against the observations. Verifying probabilistic predictions requires an examination of the reliability, accuracy, and sharpness of the probabilistic prediction [Yadav et al., 2007]. Reliability quantifies the predictive capacity of the model (that is, the benefit or cost of predicting unobserved data with the model). It measures whether the prediction will be the same when the experiment is repeated in the same conditions. Therefore, it depends more on the method than on the

variable that is being forecasted. A prediction is reliable if the predictive distribution of the variable approximates the data-based predictive distribution. Accuracy quantifies the distance between the predicted and true values. Thus, it is mainly affected by epistemic uncertainty. Finally, the sharpness or precision characterises the proximity among predictions (made under the same conditions). It does not require using a true value as reference; it relies only on the random error distribution. An optimal prediction will provide a reliable prediction with the minimum dispersion; two inferences can provide reliable predictions but with different resolutions [Renard et al., 2010]. Reliability is the priority; a forecast must not sacrifice reliability by precision. Although other quantitative measures could be used to quantify the reliability and sharpness of a predictive distribution, it is common to evaluate reliability based on the percentiles of the cumulative predictive distribution (for example, the percentiles of 2.5% and 97.5%, which are the credibility bounds of 95%; this means that there is a 95% of probability that such an interval contains the observation) or using Q-Q plots [Renard et al., 2010]. It is also possible to use a metric that combines accuracy and precision in one single metric, such as the probabilistic analogue of NS^{prob} [Bulygina et al., 2009]. Similarly, other common metric is the predictive QQ-plot [Laio and Tamea, 2007], which allows a graphic comparison of the predictive and observed cumulative predictive distributions. If the predictive distribution is reliable, the observations are realisations from this distribution and the p-values of the predictive distribution follow a uniform distribution in the interval [0-1]. Plotting the theoretical quantiles of the uniform distribution versus the quantiles of the observed p-values and calculating the deviations from the bisector 1:1 can examine this. This QQ-plot [Laio and Tamea, 2007] can be effectively summarised by the reliability and resolution indices [Renard et al., 2010].

1.1.6. BAYESIAN HYPOTHESIS TESTING (MODEL COMPARISON)

Bayesian methods are particularly appropriated for model comparison or hypothesis testing, i.e. how probable is model *M1* with respect to model *M2* given the data, p(M1|D, I)/ p(M1|D, I); and therefore for model selection [Raftery, 1993]. Bayesian model comparison requires calculation of the Bayes Factor (1.2) [Kass and Raftery, 1995]. Grounded in probability theory and derived from Bayes' Law, Bayes Factors have many desirable properties, including that they do not make the strong assumption that one model is true [Raftery, 1993]; they allow comparison between many model structures (or hypotheses) without a change in the method, and models are not required to be nested. Also, Bayes Factors protect against overfitting and model complexity, favouring simpler models automatically [Marshal et al., 2005]. They also provide a quantitative scale for classification and interpretation of how well each hypothesis (or model structure) supports the information [Jeffreys, 1961].

$$BF = \frac{\int_{\Theta_{M_1}} L(\theta_{M_1}|D, M_1, I) * p(\theta_{M_1}|M_1, I) * d\theta_{M_1}}{\int_{\Theta_{M_2}} L(\theta_{M_2}|D, M_2, I) * p(\theta_{M_2}|M_2, I) * d\theta_{M_2}}$$
(1.2)

Given two hypothesis or model structures, *M1* and *M2*, the numerator in equation (1.2) represents the evidence in favour of model *M1*, and the denominator the evidence in favour of model *M2*. Therefore, BF quantifies how well *M1* is supported by the data with respect to *M2*. The numerator and denominator are the marginal likelihoods or model evidence, which are equal to the denominator of the posterior distribution in (1.2) for each of the models. These integrals are the density of the un-normalised posterior distributions; as typically they do not have analytical solution, sampling methods, which approximate them numerically, are to be implemented (e.g. Importance Sampling [Doucet, 2000]).

One common approach in hydrology is the one-size-fits-all approach [Fenizia et al., 2008] (i.e. only one hypothesis or model structure is implemented for all the analysed catchments). The interest relies in comparing how much more likely one parameter θ_1 is with respect to θ_2 to explain *D*. In this case, BFs can be easily implemented, as equation (1.2) simplifies the likelihood ratio of the parameters $L(\theta_1|D)/L(\theta_2|D)$.

The overall objective of this thesis is to advance flow prediction capabilities for ungauged basins by developing methodologies for identifying, quantifying, and reducing the uncertainty associated with the various related sources of error. This thesis moves towards overcoming some of the major problems in hydrological modelling and focuses in the 'one size fits all' problem for predictions in ungagued basins. As this chapter has emphatized, some of the main sticking points in PUBs are due to 1) the traditional assumption that a pre-selected model (which includes structure, inputs and parameter selection) is perfect; 2) the assumption that a preselected form of regionalization is suitable to represent behavior of an ungauged catchment; and 3) inability to identify dominant hydrological mechanisms to be represented in a system's model. This thesis addresses these three challenges via the objetives that are posed in Section 1.2 of this Chapter.

1.2 AIMS AND OBJECTIVES

As stated in 1.1, the overall objective of this thesis is to advance flow prediction capabilities for ungauged basins by developing methodologies for identifying, quantifying, and reducing the uncertainty associated with the various related sources of error. This thesis advances the Bayesian framework developed in Bulygina et al. [2009, 2011], and further developed in Almeida

[2014] and Almeida et al. [2016]. This thesis moves towards overcoming some of the major problems in hydrological modelling:

- 1. A traditional assumption that a pre-selected model (which includes structure, inputs and parameter selection) is perfect.
- 2. An assumption that a pre-selected form of regionalization is suitable to represent the behavior of an ungauged catchment.
- A way to identify dominant hydrological mechanisms to be represented in a system's model.

The study illustrates the practical application of the new developments by examining a case study composed of basins in northern Spain. In particular, it seeks to achieve the following goals:

- To incorporate new developments in selection of hydrological indices (based on Principal Components Analysis) and regression methods (Random Forests) into the hydrological regionalization procedure that provides information constraints in a Bayesian approach (Chapter 2).
- To propose two statistical suitability and adequacy metrics for both a hydrological model and information regionalization procedure; and to quantify the effects of model and regionalization quality on flow predictions via Bayesian approach (Chapter 2).
- To devise a statistical test (that employs a hydrological model ensemble) for dominant hydrological mechanism identification in ungauged catchments, helping to define a minimal necessary structure for a hydrological model (Chapter 3).
- 4. And to analyze the sensitivity of the mechanisms identification to: (1) regionalization quality, (2) model ensemble used, and (3) the information content (Chapter 3).

1.3 THESIS OUTLINE

This dissertation contains four chapters, two of which are in paper format. Spanish law allows separate articles to be redacted and integrated into a single thesis, where each of the papers addresses one or more objectives.

The redaction of some of the chapters in this format means that the concepts in certain sections need to be reproduced in one of the chapters. This leaves the resultant chapters self-contained, easy to read, and focused on the aims of the study.

<u>Chapter 1</u> constitutes the introduction and describes the motivation behind this thesis. It provides an overview of the problem statement and analyses current regionalization approaches for ungauged basins, the most common sources of uncertainty in ungauged catchments, the philosophical basis for using (and for not using) probabilities to describe uncertain knowledge, and probabilistic methods for quantifying uncertainty, with a particular focus on Bayesian approaches. This background information makes the following chapters easier to understand and helps define the investigation's objectives.

Chapter 2 is in the format of an academic article. It focuses on the impact of the decisions to be made when regionalized information is used for model parameter estimation, including: 1) the selection of representative hydrological indices, 2) the selection of a functional form for the regionalization model, and 3) assumptions about models and regionalization's suitability and adequacy. The study merges new developments about indices selection and regionalization methodologies with the Bayesian prediction procedure, focusing on suitability and adequacy assumptions. This chapter introduces two statistical metrics for assessing suitability and adequacy using Bayesian statistics. The suitability metric evaluates the null hypotheses that the selected model and regionalization reproduce the available hydrological information for a given basin. The adequacy metric evaluates the knowledge gained from considering the hydrological model as compared to that provided by the regionalization alone, and the knowledge gained from the regionalization as compared to that provided by the model alone. It fulfills the first and second objectives of the thesis. Furthermore, the study analyzes the impact of the regionalization and hydrological model on the quality of the predictions.

Part of the work developed for this chapter has been presented at the EGU General Assembly 2015 [Prieto et al., 2015a; 2015b]:

• Prieto, C., Le-Vine, N., García, E., and Medina, R. (2015a) Improving standard practices for prediction in ungauged basins: Bayesian approach. Abstract for the European

Geosciences UnionGeneral Assembly 2015, available online: http://meetingorganizer.copernicus.org/EGU2015/EGU2015-9832.pdf

 Prieto, C., Le-Vine, N., García, E., and Medina, R. (2015b). Improving standard practices for prediction in ungauged basins: Bayesian approach. Presentation for the European Geosciences UnionGeneral Assembly 2015, available online: http://presentations.copernicus.org/EGU2015-9832_presentation.pdf

Chapter 3 is also in the format of an academic article. It focuses on the challenge of identifying dominant hydrological mechanisms in ungauged catchments. Using formal Bayesian probability theory, an ensemble of hydrological models, and regionalized information, it presents a statistical test for determining the dominant mechanisms in ungauged catchments. The methodology is based on the Bonferroni correction for the multiple hypotheses comparison problem. Furthermore, the study analyzes the sensitivity of the mechanism identification procedure to: (1) the quality of the regionalization procedure, (2) the model ensemble error, and (3) the quantity of information contained in the selected regionalized indices. It fulfills the third objective of the thesis.

Part of the work developed for this chapter has been presented at the EGU General Assembly 2016 [Prieto et al., 2016a; 2016b] and in the AGU Fall Meeting 2016 [Prieto et al., 2016b]:

- Prieto, C., Le-Vine, N., Vitolo, C., García, E., and Medina, R. (2016a). Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Abstract for the European Geosciences Union General Assembly 2016, available online: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-8580-2.pdf
- Prieto, C., Le-Vine, N., Vitolo, C., García, E., and Medina, R. (2016b). Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Presentation for the European Geosciences UnionGeneral Assembly 2016, available online: http://presentations.copernicus.org/EGU2016-8580_presentation.pdf
- Prieto, C., Le-Vine, N., Vitolo, C., and Medina, R. (2016c). Dominant hydrolocial process identification for ungauged basins: Bayesian approach. Presentation for the European Geosciences Union General Assembly 2016, available online: https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/160806

<u>Chapter 4</u> describes how the thesis contributes to the current state of the art, offers general conclusions, and provides recommendations for future research.

It is worth to mention that the main contributions of this thesis (which are shown in Chapter 4) will be presented at the EGU General Assembly 2017 [Prieto et al., 2017]. As well as part the work related to the proposed future research lines has already been presented at the Spanish Water Engineering Conference (JIA) 2013 [Prieto et al., 2013a; 2013b], JIA 2015 [Álvarez et al., 2015a; 2015b], in the AGU Fall Meeting 2016 [Addor et al., 2016] and will be presented at EGU Fall Meeting 2017 [Addor et al., 2017]:

- Prieto, C., Le Vine, N., Vitolo, C., and Medina, R. (2017). Addressing model structural uncertainty in PUBs via Bayesian Approach. Abstract for the European Geosciences Union General Assembly 2016, available online: http://meetingorganizer.copernicus.org/EGU2017/EGU2017-13157.pdf
- Addor, N., Clark, M., Prieto, C., Newman, A., Mizukami, N., Nearing, G., Le Vine, N. (2017) On the information content of hydrological signatures and their relationship to catchment attributes. Abstract for the European Geosciences Union General Assembly 2017, available online: http://meetingorganizer.copernicus.org/EGU2017/EGU2017-9718-1.pdf
- Addor, N., Clark, M., Prieto, C., Newman, A., Mizukami, N., Nearing, G., Le Vine, N.(2016) On the information content of hydrological signatures and their relationship to catchment attributes, Abstract for the Fall Meeting of the American Geophysical Union, San Fransisco, available online: https://aqu.confex.com/aqu/fm16/meetingapp.cgi/Paper/160052
- Prieto, C., García, E., Mínguez, R., Álvarez, C., and Medina, R. (2013a). Propuesta de un modelo hidrológico agregado basado en ecuaciones generales de crecimiento. Aplicación a la cuenca del río teifi (UK), available online: http://www.ingenieriadelagua.com/2004/JIA/Jia2013/pdf/b48.pdf
- Prieto, C., García, E., Mínguez, R., Álvarez, C., and Medina, R. (2013b). Simulación hidrológica de cuencas basada en ecuaciones de crecimiento. Aplicación a la cuenca del río Teifi (UK).
- Álvarez, C., García, E., Prieto, C., Rojo, J., and Tejerina., B. (2015a). Aplicación de un modelo logístico triparamétrico a la estimación de caudales diarios en la cuenca del río Nam Ngum (Laos), available online: http://www.uco.es/jia2015/ponencias/b/b008.pdf
- Álvarez, C., García, E., Prieto, C., Rojo, J., and Tejerina., B. (2015b). Aplicación de un modelo logístico triparamétrico a la estimación de caudales diarios en tres subcuencas tropicales de la cuenca del río Mekong en Laos.

1.4 REFERENCES IN CHAPTER 1

- Addor, N., Clark, M., Prieto, C., Newman, A., Mizukami, N., Nearing, G., Le Vine, N. (2017). On the information content of hydrological signatures and their relationship to catchment attributes. Abstract for the European Geosciences Union General Assembly 2016, available online: http://meetingorganizer.copernicus.org/EGU2017/EGU2017-9718-1.pdf
- Addor, N., Clark, M., Prieto, C., Newman, A., Mizukami, N., Nearing, G., Le Vine, N. (2016). On the information content of hydrological signatures and their relationship to catchment attributes, Abstract for the Fall Meeting of the American Geophysical Union, San Fransisco, available

online: https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/160052

Almeida, S., Bulygina, N., McIntyre, N., Wagener, T., and Buytaert, W. (2013). Improving parameter priors for data-scarce estimation problems. Water Resources Research, 49(9), 6090-6095.

- Almeida, S. M. C. L. (2014). The Value of Regionalized Information for Hydrological Modelling. PhD thesis, Imperial College London, London, UK.
- Almeida, S., Le Vine, N., McIntyre, N., Wagener, T., and Buytaert, W. (2016). Accounting for dependencies in regionalized signatures for predictions in ungauged catchments, Hydrol. Earth Syst. Sci., 20, 887-901. doi:10.5194/hess-20-887-2016
- Álvarez, C., García, E., Prieto, C., Rojo, J., and Tejerina., B. (2015a). Aplicación de un modelo logístico triparamétrico a la estimación de caudales diarios en la cuenca del río Nam Ngum (Laos), available online: http://www.uco.es/jia2015/ponencias/b/b008.pdf
- Álvarez, C., García, E., Prieto, C., Rojo, J., and Tejerina., B. (2015b). Aplicación de un modelo logístico triparamétrico a la estimación de caudales diarios en tres subcuencas tropicales de la cuenca del río Mekong en Laos.
- Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. Hydrology and Earth System Sciences Discussions, 11(2), 703-710.
- Bayes, T. (1763). An Essay Towards Solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London, 53:370– 418.
- Berger, J., (1985). Statistical Decision Theory and Bayesian Analysis. Springer Series in Statistics (second ed.) Springer-Verlag. ISBN 0-387-96098-8
- Bernardo, J. M. and Smith, A. F. M. (1994). Bayesian Theory. Chichester: Wiley.
- Berger, J. (2006). The case for objective Bayesian analysis. Bayesian analysis, 1(3), 385-402.
- Beven, K., and Binley, A. (1992). The future of distributed models model calibration and uncertainty prediction. Hydrological Processes, 6 (3), 279-298. doi: 10.1002/hyp.3360060305
- Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modeling. Advances in Water Resources, 16 (1), 41-51. doi: 10.1016/0309-1708(93)90028-E
- Beven, K. J. (2000). Uniqueness of place and process representations in hydrological modelling. Hydrology and Earth System Sciences Discussions, 4(2), 203-213.
- Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of hydrology, 249(1), 11-29.
- Beven, K. (2006), On undermining the science?. Hydrol. Process., 20: 3141–3146. doi:10.1002/hyp.6396
- Beven, K. J. (2007). Working towards integrated environmental models of everywhere: Uncertainty, data, and modelling as a learning process, HESS., 11(1), 460–467.
- Beven, K. (2008). On doing better hydrological science. Hydrological Processes, 22 (17), 3549-3553. doi: 10.1002/hyp.7108

- Beven, K. J., Smith, P. J., and Freer, J. E. (2008). So just why would a modeller choose to be incoherent? Journal of Hydrology, 354 (1-4), 15-32. doi: 10.1016/j.jhydrol.2008.02.007
- Beven, K. (2009). Environmental Modelling: An Uncertain Future? Oxon, U.K., Routledge.
- Beven, K., and Westerberg, I. (2011). On red herrings and real herrings: disinformation and information in hydrological inference. Hydrol. Process., 25: 1676–1680. doi: 10.1002/hyp.7963
- Beven, K. and Smith, P. (2014). Concepts of Information Content and Likelihood in Parameter Calibration for Hydrological Simulation Models. J. Hydrol. Eng., 10.1061/(ASCE)HE.1943-5584.0000991, A4014010
- Box, G.E.P. (1979). Robustness in the strategy of scientific model building. In: Robustness in Statistics. vol. Academic Press New York.
- Blöschl, G. (Ed.). (2013). Runoff prediction in ungauged basins: synthesis across processes, places and scales. Cambridge University Press.
- Bulygina, N., and Gupta, H. (2010). How Bayesian data assimilation can be used to estimate the mathematical structure of a model. Stochastic Environmental Research and Risk Assessment, 24(6), 925-937.
- Bulygina, N., McIntyre, N., and Wheater, H. (2009). Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis.
 Hydrology and Earth System Sciences, 13 (6), 893-904. doi: 10.5194/hess-13-893-2009
- Bulygina, N., McIntyre, N. and Wheater, H. (2011). Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. Water Resources Research, 47 (2). doi: 10.1029/2010wr009240
- Bulygina, N., and Gupta, H. (2011). Correcting the mathematical structure of a hydrological model via Bayesian data assimilation. Water Resour. Res.,47, W05514, doi:10.1029/2010WR009614
- Bulygina, N., Ballard, C., McIntyre, N., O'Donnell, G., and Wheater, H. (2012). Integrating different types of information into hydrological model parameter estimation: application to ungauged catchments and land use scenario analysis. Water Resources Research, 48 (6). doi: 10.1029/2011wr011207
- Chandler, R. E. (2013). Exploiting strength, discounting weakness: combining information from multiple climate simulators. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371, 20120388. doi:10.1098/rsta.2012.0388
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular

framework to diagnose differences between hydrological models. Water Resources Research, 44 (12). doi: 10.1029/2007wr006735

- Clark, M. P., Kavetski, D., and Fenicia, F. (2011a). Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research, 47. doi: 10.1029/2010wr009827
- Clark, M. P., McMillan, H. K., Collins, D. B., Kavetski, D., and Woods, R. A. (2011b). Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses. Hydrological Processes, 25(4), 523-543.
- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M. (2015a): A unified approach for process-based hydrologic modeling: 1. Modeling concept. Water Resources Research, 51, 2498-2514. doi: 10.1002/2015wr017198
- Clark, M. P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., and Marks, D. G. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. Water Resources Research, 51, 2515-2542. doi: 10.1002/2015wr017200
- Clark, M. P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W.,, Brekke, L. D., Arnold, J. R., Gochis, D. J., Rasmussen, R.M., Tarboton, D.G., Mahat, V., Flerchinger, G.N., and Marks, D.G. (2015c). The Structure for Unifying Multiple Modeling Alternatives (SUMMA), version 1: Technical description. NCAR Technical Note NCAR/TN-514+ STR, 54 pp., National Center for Atmospheric Research, Boulder, Colo. doi: 10.5065/D6WQ01TD
- Cox, R.T. (1946). Probability, frequency and reasonable expectation, Am. J. Phys. 14 (1), 1-13.
- De Finetti, B. (1931). On the Subjective Meaning of Probability. English Translation in De Finetti, B. Induction and Probability. Clueb, Bologna, 1993.
- de Finetti, B. (1937). La Prévision: Ses Lois Logiques, Ses Sources Subjectives. Annales de l'Institut Henri Poincar´e, 7: 1–68. Translated as Foresight: Its Logical Laws, Its Subjective Sources, in Kyburg, H. E. and Smokler, H. E. eds., (1964). Studies in Subjective Probability. Wiley, New York, 91–158. (Corrections appear in (de Finetti, 1972, p. xvii)).
- De Finetti, B. (1974). Theory of Probability, volume I. Wiley
- De Morgan, A. (1837). Review of Laplace Theorie Analytique des Probabilites. (3rd Edition). Dublin Review, 2, 3: 338–354, 237–248
- Dempster, A. P. (1967). Upper and lower probability inferences based on a sample from a finite univariate population. Biometrika, 54(3-4), 515-528.

- Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and computing, 10(3), 197-208.
- Engeland, K., Xu, C.Y., and Gottschalk, L., (2005). Assessing uncertainties in a conceptual water balance model using Bayesian methodology. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 50(1): 45-63.
- Fenicia, F., Kavetski, D., and Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47, W11510. doi: 10.1029/2010WR010174
- Fenicia, F., McDonnell, J. J., and Savenije, H. H. G. (2008). Learning from model improvement:
 On the contribution of complementary data to process understanding. Water Resources
 Research, 44 (6). doi: 10.1029/2007wr006386
- Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L. (2006). Is the groundwater reservoir linear? Learning from data in hydrological modelling.Hydrology and Earth System Sciences Discussions, 10(1), 139-150.
- Fienberg, S. A., (2005). When did bayesian inference become "bayesian"? Bayesian Analysis, 1(1):1–40, July 2005.
- Fearnhead, P., and Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74: 419–474. doi:10.1111/j.1467-9868.2011.01010.x
- Fisher, R.A. (1922). On the Mathematical Foundations of Theoretical Statistics . Philosophical Transactions of the Royal Society of London, Series A, 222: 309-368.
- Fisher, R. A. (1950). Statistical methods for research workers. Biological monographs and manuals. No. V. Statistical methods for research workers. Biological mono-graphs and manuals. No. V., (11th ed.).
- Gallego, G., Cuevas, C., Mohedano, R., & Garcia, N. (2013). On the Mahalanobis distance classification criterion for multidimensional normal distributions. Signal Processing, IEEE Transactions on, 61(17
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. 2013.Bayesian Data Analysis, 3rd Edition. Chapman & Hall/CRC Press, London. [appendix on Stan]
- Gillies, D. (1991). Intersubjectivity probability and Confirmation theory. Brit J. Phil. Sci. 42, 513-533.
- Gillies, D. (2000). Philosophical Theoruies of Probability. Routledge, London, UK.
- Goldstein, M. (2006). Subjective Bayesian analysis: principles and practice. *Bayesian Analysis*, 1(3), 403-420.

- Gupta, H. V., Beven, K., and Wagener, T. (2005). Model Calibration and Uncertainty Estimation.
 In: Anderson, M. G. (ed.) Encyclopedia of hydrological Sciences. Hoboken, NJ, John Wiley
 & Sons.
- Gupta, H. V., Wagener, T., and Liu, Y. Q. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. Hydrological Processes, 22 (18), 3802-3813, doi: 10.1002/Hyp.6989.
- He, Y., Bárdossy, A., and Zehe, E. (2011). A review of regionalization for continuous streamflow simulation. Hydrology and Earth System Sciences, 15 (11), 3539-3553, doi: 10.5194/hess-15-3539-2011.
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. Statistical science, 382-401.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W.,
 Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta,
 H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A.,
 Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.
 (2013). A decade of Predictions in Ungauged Basins (PUB) a review. Hydrological
 Sciences Journal, 58 (6), 1-58. doi: 10.1080/02626667.2013.803183
- Hsu, K. L., Moradkhani, H., and Sorooshian, S. (2009). A sequential Bayesian approach for hydrologic model selection and prediction. Water Resources Research, 45(12).
- Jeffreys, H. (1939). Theory of Probability. Oxford University Press
- Jeffreys, H. (1961). Theory of Probability, 3rd ed., Oxford, Oxford University Press.
- Kass, R., & Raftery, A. (1995). Bayes Factors. Journal of the American Statistical Association, 90(430), 773-795. doi:10.2307/2291091
- Kavetski, D., and Clark, M.P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. Water Resources Research, 46: W10511. doi: 10.1029/2009wr008896
- Kavetski, D., Franks, S.W., and Kuczera, G. (2002). Confronting input uncertainty in environmental modelling, in Calibration of Watershed Models, Water Sci. Appl., vol. 6, edited by Duan, Q., et al., pp. 49–68, AGU, Washington, D. C.
- Kavetski D, Kuczera G. and Franks SW (2006a) Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts, Journal of Hydrology (Special issue on The model parameter estimation experiment - MOPEX), 320(1-2), pp. 173-186.
- Kavetski D, Kuczera G and Franks, SW (2006b) Bayesian analysis of input uncertainty in hydrological modelling: 1. Theory, Water Resources Research, 42, W03407.
- Kavetski D, Kuczera G and Franks SW (2006c) Bayesian analysis of input uncertainty in

hydrological modelling: 2. Application, Water Resources Research, 42, W03408.

Keines, J.M. (1921). A Treatise on Probability. Macmillan, 1963.

- Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).
- Kolmogorov, A. N. (1933. Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer-Verlag.
- Kuczera, G. (1996). Correlated rating curve error in flood frequency inference. Water Resources Research, 32(7): 2119-2127.
- Kuczera, G., and Mroczkowski, M. (1998). Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. Water Resources Research, 34 (6), 1481-1489. doi: 10.1029/98WR00496.
- Kuczera, G., and Parent, E. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. (1998). Journal of Hydrology, Volume 211, Issues 1–4, November 1998, Pages 69-85, ISSN 0022-1694, http://dx.doi.org/10.1016/S0022-1694(98)00198-X.
- Kuczera, G., Kavetski, D., Franks, S., and Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. Journal of Hydrology, 331(1), 161-177.
- Kuichling, E. (1889). The relation between the rainfall and the discharge of sewers in populous districts. Transactions, American Society of Civil Engineers 20, 1-56.
- Laio, F., and Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrology and Earth System Sciences, 11 (4), 1267-1277. doi: 10.5194/hess-11-12672007
- Laplace, P.-S. (1774). Mémoires de Mathématique et de Physique, Presentés à l'Académie Royale desSciences, Par Divers Savans, & Lus dans ses Assemblées, 6:621–656
- Le Vine, N. (2016). Combining information from multiple flood projections in a hierarchical Bayesian framework. Water resources research., 52. doi:10.1002/2015WR018143
- Lindley, D. V. (1953). Statistical Inference. Journal of the Royal Statistical Society, Series B, 16: 30–76.
- Lhoste, E. (1923). Le Calcul des Probabilités Appliquée À L'artillerie. Revue D'artillerie, 91, 92: 405–423 and 516–532, 58–82 and 152–179.
- Mantovan, P., and Todini, E. (2006). Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. Journal of Hydrology, 330 (1-2), 368-381. doi: 10.1016/j.jhydrol.2006.04.046

Marshall, L., Nott, D., and Sharma, A. (2004). A comparative study of Markov chain Monte Carlo

methods for conceptual rainfall-runoff modeling. Water Resources Research, 40(2).

- Marshall, L., Nott, D., and Sharma, A. (2005). Hydrological model selection: A Bayesian alternative. Water resources research, 41(10).
- McKay, M., Beckman, R., & Conover, W. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics, 21(2), 239-245. doi:10.2307/1268522
- McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T. (2005). Ensemble predictions of runoff in ungauged catchments. Water Resources Research, 41 (12). doi: 10.1029/2005wr004289
- McMillan, H., Krueger, T., and Freer, J. (2012). Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrol. Process., 26: 4078–4111. doi: 10.1002/hyp.9384
- Montanari, A., Shoemaker, C. A., and van de Giesen, N. (2009). Introduction to special section on uncertainty assessment in surface and subsurface hydrology: An overview of issues and challenges. Water Resources Research, 45 (12). doi: 10.1029/2009wr008471
- Moore, R. E., & Bierbaum, F. (1979). Methods and applications of interval analysis (Vol. 2, pp. xi+-190). Philadelphia: Siam.
- Moradkhani, H., & Sorooshian, S. (2009). General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis. In Hydrological modelling and the water cycle (pp. 1-24). Springer Berlin Heidelberg.
- Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. Philosophical Transactions of the Royal Society, Series A, 236: 333–380.
- Pearson, K. (1892). The Grammar Of Science. Walter Scott
- Pechlivanidis, I. G., McIntyre, N. R., and Wheater, H. S. (2010). Calibration of the semi-distributed PDM rainfall–runoff model in the Upper Lee catchment, UK. Journal of Hydrology, 386(1), 198-209.
- Popper, K. R. (1959). The propensity interpretation of probability. Br. J. Ohilos. Sci. 10, 25-42.
- Prieto, C., García, E., Mínguez, R., Álvarez, C., and Medina, R. (2013a). PROPUESTA DE UN MODELO HIDROLÓGICO AGREGADO BASADO EN ECUACIONES GENERALES DE CRECIMIENTO. APLICACIÓN A LA CUENCA DEL RÍO TEIFI (UK), available online: http://www.ingenieriadelagua.com/2004/JIA/Jia2013/pdf/b48.pdf
- Prieto, C., García, E., Mínguez, R., Álvarez, C., and Medina, R. (2013b). Simulación hidrológica de cuencas basada en ecuaciones de crecimiento. Aplicación a la cuenca del río Teifi (UK).
- Prieto, C., Le-Vine, N., García, E., and Medina, R. (2015a) Improving standard practices for prediction in ungauged basins: Bayesian approach. Abstract for the European

Geosciences UnionGeneral Assembly 2015, available online: http://meetingorganizer.copernicus.org/EGU2015/EGU2015-9832.pdf

- Prieto, C., Le-Vine, N., García, E., and Medina, R. (2015b). Improving standard practices for prediction in ungauged basins: Bayesian approach. Presentation for the European Geosciences UnionGeneral Assembly 2015, available online: http://presentations.copernicus.org/EGU2015-9832_presentation.pdf
- Prieto, C., Le-Vine, N., Vitolo, C., García, E., and Medina, R. (2016a). Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Abstract for the European Geosciences UnionGeneral Assembly 2016, available online: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-8580-2.pdf
- Prieto, C., Le-Vine, N., Vitolo, C., García, E., and Medina, R. (2016b). Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Presentation for the European Geosciences UnionGeneral Assembly 2016, available online: http://presentations.copernicus.org/EGU2016-8580_presentation.pdf
- Prieto, C., Le-Vine, N., Vitolo, C., and Medina, R. (2016c). Dominant hydrolocial process identification for ungauged basins: Bayesian approach. Presentation for the European Geosciences UnionGeneral Assembly 2016, available online: https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/160806
- Prieto, C., Le Vine, N., Vitolo, C., García., E, Medina, R. (2016a), Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Abstract for the Meeting of the European Geophysical Union, Vienna, available online: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-8580-2.pdf
- Prieto, C., Le Vine, N., Vitolo, C., García., E, Medina, R. (2016a), Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Abstract for the Meeting of the European Geophysical Union, Vienna, available online: http://presentations.copernicus.org/EGU2016-8580_presentation.pdf
- Prieto, C., Le Vine, N., Vitolo, C., and Medina, R. (2017). Addressing model structural uncertainty in PUBs via Bayesian Approach. Abstract for the Meeting of the European Geophysical Union, Vienna, available online: http://meetingorganizer.copernicus.org/EGU2017/EGU2017-13157.pdf
- Raftery, A.E. (1993). Bayesian model selection in structural equation models. In Testing Stuctural Equation Models (K.A. Bollen and J.S. Long, eds.), Beverly Hills: Sage, pp. 163-180. Earlier version.
- Raftery, A.E., and Zheng, Y. (2003). Discussion: Performance of Bayesian Model Averaging. Journal of the American Statistical Association, 98, 931-938.

- Ramsay, F.P. (1926). Thruth and probability. In: Ramsey. F.P. (Ed.) The Fundations of Mathematics and Other Logical Essays. Routledge and Kegan Paul, 1931.
- Refsgaard, J.C., van der Sluijs, J.P., Brown, J., and van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. Advances in Water Resources, 29(11): 1586-1597.
- Reichert, P., and Omlin, M. (1997). On the usefulness of overparameterized ecological models. Ecological Modelling, 95, 289-299. http://doi.org/10.1016/S0304-3800(96)00043-9
- Reichert, P., and J. Mieleitner (2009). Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, Water Resour. Res., 45, W10402, doi:10.1029/2009WR007814.
- Reichert, P., Langhans, S., Lienert, J., and Schuwirth, N. (2015)The Conceptual Foundation of Environmental Decision Support. Journal of Environmental Management 154, 316-332. doi:10.1016/j.jenvman.2015.01.053
- Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., and Franks, S. W. (2011). Toward a reliable decomposition of predictive uncertainty in hydrological modeling:
 Characterizing rainfall errors using conditional simulation, Water Resour. Res., 47, W11516. doi:10.1029/2011WR010643
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46, W05521. doi:10.1029/2009WR008328
- Rougier, J.C. (2010). Formal Bayes Methods for Model Calibration with Uncertainty. In: Applied Uncertainty Analysis for Flood Risk Management. vol. Imperial College Press / World Scientific London.
- Rubin, H. (1949). The Existence of Measurable Utility and Psychological Probability. Statistics 331, Cowles Commission Discussion Paper. 5 pp.
- Savage, L. J. (1951). The Theory of Statistical Decision. Journal of the American Statistical Association, 46: 55–67. Reprinted in Savage, Leonard J. (1981). The Writings of Leonard Jimmie Savage: A Memorial Selection. American Statistical Association and Institute of Mathematical Statistics, Washington, DC, 201-213
- Schaake, J., Cong, S., and Duan, Q. (2006). U.S. MOPEX data set. IAHS.
- Schaake, J., Duan, Q. Y., Smith, M., and Koren, V. (2000). Criteria to select basins for hydrologic model development and testing. In: Preprints, 15th Conference on Hydrology, Boston, MA: Amer. Meteorol. Soc, pp. 149-152.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G. (2011). Catchment

classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. Hydrology and Earth System Sciences, 15 (9), 2895-2911. doi: 10.5194/hess-15-2895-2011

- Schoups, G., and Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Resources Research, 46 (10). doi: 10.1029/2009wr008933
- Shafer, G. (1976). A mathematical theory of evidence (Vol. 1, pp. xiii+-297). Princeton: Princeton university press.
- Singh, R., Wagener, T., Crane, R., Mann, M. E., and Ning, L. (2014). A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds: Application to a watershed in Pennsylvania, USA, Water Resour. Res., 50, 3409–3427. doi:10.1002/2013WR014988
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X.,
 McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer,
 D., Uhlenbrook, S., and Zehe, E. (2003). IAHS decade on Predictions in Ungauged Basins
 (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. Hydrological
 Sciences Journal-Journal Des Sciences Hydrologiques, 48 (6), 857-880. doi:
 10.1623/hysj.48.6.857.51421
- Spear, R.C., and Hornberger, G.M. (1980). Eutrophication in Peel Inlet .2. Identification of Critical Uncertainties Via Generalized Sensitivity Analysis. Water Research, 14(1): 43-49.
- Steinschneider, S., Polebitski, A., Brown, C., and Letcher, B. H. (2012). Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change. Water Resources Research, 48(11).
- Stedinger, J.R., Vogel, R.M., Lee, S.U., and Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research, 44: W00b06. doi: 10.1029/2008wr006822
- Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., and Srikanthan, S. (2009). Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis. Water Resources Research, 45: W00b14. doi: 10.1029/2008wr006825
- Todini, E., and Mantovan, P. (2007). Comment on: 'On undermining the science?' by Keith Beven. Hydrological Processes, 21(12), 1633-1638.
- Mises, V. (1928). Probability, Statistics, and Truth, second revised ed. Allen and Unwin. 1961.
- Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain

Monte Carlo simulation. Water Resources Research, 44(12).

- Vrugt, J. A., & Sadegh, M. (2013). Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. Water Resources Research, 49(7), 4335-4345.
- Vrugt and Massoud (2017), Uncertainty Quantification of Complex System Models: Bayesian Analysis. Submitted
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S. (2001). A framework for development and application of hydrological models. Hydrology and Earth System Sciences, 5 (1), 13-26. doi: 10.5194/hess-5-13-2001
- Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., and Gupta, H.V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrological Processes, 17(2): 455-476.
- Wagener, T., Wheater, H. S., and Gupta, H. V. (2004). Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, London, Imperial College Press.
- Wagener, T., Sivapalan, M., Troch, P. and Woods, R. (2007). Catchment classification and hydrologic similarity. Geography Compass, 1 (4), 901-931. doi: 10.1111/j.1749-8198.2007.00039.x
- Wagener, T., Sivapalan, M., Troch, P.A., McGlynn, B.L., Harman, C.J., Gupta, H.V., Kumar, P., Rao,
 P.S.C., Basu, N.B., and Wilson, J.S., (2010). The future of hydrology: An evolving science for a changing world. Water Resources Research, 46: W05301. doi: 10.1029/2009wr008906
- Wagener, T. and Montanari, A. (2011). Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. Water Resources Research, 47 (6). doi: 10.1029/2010wr009469
- Westerberg, I.K. (2011). Observational Uncertainties in Water-Resources Modelling in Central America: Methods for Uncertainty Estimation and Model Evaluation. PhD Thesis, Department of Earth Sciences, Uppsala University. Acta Universitatis Upsaliensis, p. 75.
- Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C.-Y. (2011) Calibration of hydrological models using flow-duration curves, Hydrol. Earth Syst. Sci., 15, 2205-2227, doi:10.5194/hess-15-2205-2011.
- Westerberg, I. K., and Birkel, C. (2015), Observational uncertainties in hypothesis testing: investigating the hydrological functioning of a tropical catchment. Hydrol. Process., 29: 4863–4879. doi: 10.1002/hyp.10533.
- Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., and Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resour. Res., 52, 1847–1865. doi:10.1002/2015WR017635

- Yadav, M., Wagener, T., and Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources, 30 (8), 1756-1774. doi: 10.1016/j.advwatres.2007.01.005
- Zhang, Z., Wagener, T., Reed, P., and Bhushan, R. (2008). Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. Water Resources Research, 44 (12)

Chapter 2: Statistical Metrics to Assess Regionalization and Model Adequacy in PUBs A number of the contributions presented in this Chapter has been presented at the EGU General Assembly 2015 [Prieto et al., 2015a; 2015b]:

Prieto, C., Le-Vine, N., García, E., and Medina, R. (2015a) Improving standard practices for prediction in ungauged basins: Bayesian approach. Abstract for the European Geosciences Union General Assembly 2015, available online: http://meetingorganizer.copernicus.org/EGU2015/EGU2015-9832.pdf

Prieto, C., Le-Vine, N., García, E., and Medina, R. (2015b). Improving standard practices for prediction in ungauged basins: Bayesian approach. Presentation for the European Geosciences Union General Assembly 2015, available online: http://presentations.copernicus.org/EGU2015-9832_presentation.pdf

ABSTRACT

This paper merges the strengths of Bayesian inference with new developments in indices selection and regionalization methods to predict flows in ungauged basins, and presents new statistical metrics to assess the model and regionalization suitability and adequacy conventions. First, an extensive set of available hydrological indices is reduced to a compact orthogonal set of pieces of information using principal component analysis. Second, the selected information is regionalized using functional form-assumption-free Random Forests. Third, two metrics are proposed to assess the suitability and adequacy of a selected model and a regionalization procedure. The suitability metrics evaluate if a selected model or a selected regionalization procedure reproduce available hydrological information for an ungauged basin. The adequacy metrics evaluate gains in knowledge from considering a selected model, or considering a selected regionalization method. Basins in northern Spain are selected as case studies. The advantages of this approach are: (1) a small number of hydrological indices regionalized via the Random Forests allows considerable performance improvements when there is no model error; (2) the proposed suitability and adequacy tests identify the main source of uncertainty and/or error between a model and regionalization method. Results show that model – including its structure and inputs, is the main source of error; only when there is not model error or is small, regionalization quality can be targeted to improve reliability, accuracy, and precision of model performance. Thus, priority should be given to the selection of a representative model followed by regionalization improvement, not the other way around.

2.1 INTRODUCTION

In hydrological modeling, the prediction of flows in ungauged basins (PUBs) remains a challenge [Smith et al., 2014]. A Bayesian framework is among the different alternatives to quantify and reduce the prediction uncertainty [Blöschl et al., 2013]; it allows flow prediction in ungauged basins based on regionalized hydrological indices (also called signatures) [Yadav et al., 2007; Bulygina et al., 2009; Singh, 2013]. The procedure relies on a number of selections: (1) selection of representative hydrological indices, (2) selection of regionalization procedure, and (3) hydrological model choice.

Among the variety of possible hydrological indices, the chosen indices must characterize the most important components of the flow regime. To confront this, Olden and Poff [2003] chose a large number of indices (171) that represent five meaningful aspects of the flow regime: annual and monthly flows, high and low flows, duration and frequency of high flows, rate of change in flows, and time of maximum and minimum flow events. They demonstrated the utility of principal components analysis (PCA) to compactly describe variability in the data. More recently, Yadav et al. [2007] considered 39 indices, which were divided into seven classes by means of linear and Spearman rank correlation coefficients.

To regionalize hydrological indices for ungauged basins, linear regression relationships are usually fitted between hydrological indices and catchment descriptors (CDs) that are available for gauged basins [Yadav et al., 2007; Zhang et al., 2008]. As the relationships between the hydrological indices and the CDs are usually non-linear [Snelder et al., 2009], Snelder et al. [2013] and Peñas et al. [2014a] used Random Forests (RFs) for regionalization. RFs extend the concept of one regression tree [Breiman et al., 1984] to a set of trees [Breiman, 2001]. Random Forests retain the advantages of a regression tree, as they are hypothesis-free with respect to the error distributions and the type of functional dependence. Meanwhile, RFs offer an improvement in accuracy over regression when compared to a regression tree and are more robust against random variations in the inputs [Snelder et al., 2013]. RFs have been successfully applied by the ecohydrological community to regionalize hydrological indices and to explain variations in hydrological patterns [e.g. Booker and Snelder, 2012; Booker, 2013; Snelder et al., 2013]. They have also been proven efficient compared with the other machine learning algorithms [Peñas et al., 2014b] or other more physically based approaches [Booker and Woods, 2014]. However, the previous studies did not aim to use the regionalized indices in a conditioning procedure to predict flows in ungauged basins.

Regionalized hydrological indices are used for probabilistic model parameter conditioning that relies on the following assumptions: (1) the chosen model is able to reproduce the regionalized values, (2) the regionalization procedure is able to reproduce the hydrological response in a given basin [Almeida, 2014] and (3) the chosen model is able to reproduce the hydrological response exactly and the observed model inputs are noise-free [Bulygina et al., 2009; Bulygina et al., 2011; Almeida et al., 2012]. An increasing body of literature suggests a need to assess model suitability and adequacy for a given task [Clark et al., 2008 and 2011; Wagener et al., 2001; Bulygina and Gupta, 2010; Fenicia et al., 2008 and 2011], as well as to assess the quality of available hydrological observations to drive and evaluate models [Kavetski et al., 2002 and 2006a; Renard et al., 2011; Beven and Westerberg, 2011; McMillan et al., 2012; Westerberg and Birkel, 2015]. When regionalization is used to constrain predictions, suitability and adequacy assessment must be extended to the regionalization procedure, as the extrapolation from a gauged to ungauged basin is affected by the uniqueness of an ungauged basin [Beven, 2000; Wagener and Montanari, 2011].

This work advances the standard practices for predicting flows in ungauged basins by (a) incorporating the new hydrological indices selection and regionalization methods into the context of flow conditioning in ungauged basins, and (b) introducing metrics to assess the suitability and adequacy of both a selected model and regionalization method. Suitability evaluates how well the model reproduces regionalized information, and the ability of the regionalization to reproduce the hydrological information in a given basin. Adequacy evaluates the gain of knowledge from the regionalization and the model, such that it is possible to assess the value of adding regionalized information, or a model to provide predictions. First, a large set of hydrological indices is reduced to a compact orthogonal set using PCA. Second, Random Forests are used to regionalize the selected information, which is then used to constrain hydrological predictions in a Bayesian framework. Third, suitability and adequacy are assessed using a new suitability test and Bayes' Factors [Gelman, 2013]. The procedure is illustrated for river flow predictions in northern Spain.

2.2 CATCHMENTS AND DATA DESCRIPTION

To regionalize hydrological indices, this study utilizes hydrological records and properties from a set of 92 small (2 km²) to medium-sized (1,038 km²) basins in northern Spain (Figure 1.). A subset of the 156 basins used by Peñas et al. [2014a] is employed, this analysis focuses on these basins to ensure a natural hydrological regime according to the Water Framework Directive (articles 5 and 6, sections on the "characterization of bodies of water" and the "analysis of pressure and impact" [http://servicios2.marm.es/sia/visualization/descargas/dma.jsp]). Of this study's subset of 92 basins, 62 basins drain into the Cantabrian Sea, with the remaining 30 basins draining into the Mediterranean Sea.



Figure 1: Case studies

The selected watersheds span a variety of geologies, soils, topographies, land uses, and climatic conditions. According to the Geological and Mining Institute of Spain (http://www.igme.es/internet/default.asp), the dominant lithological groups are clay, sand, and gravel in basins that drain into the Mediterranean Sea. However, the Pyrenees also contain some siliceous and calcareous rock. Moreover, the western basins that drain into the Cantabrian Sea are primarily composed of slates, while calcareous rock are the dominant class in the eastern basins. The main channels slope spans from 9% to 63%, while the average altitude ranges between 200 m and 2200 m. In each basin, urbanized zones comprise less than 8% of total land area, with pastures, broadleaf forests, and coniferous forests utilizing the greatest share of the land. In addition, average temperatures are between 5.5°C and 15°C, while annual average rainfall varies between 450 mm and 1800 mm. Likewise, annual average potential evapotranspiration (PET) ranges between 609mm and 962 mm, and the aridity index spans from 0.3 to 1.7 [Arora, 2002]. Figure 1 classifies the basins on the basis of (1) the aridity index and (2) minimum monthly average temperatures. It considers climatic conditions to be dry when for aridity index values are above 1 and conditions to be humid if otherwise [Arora, 2002]. When the mean monthly temperature is below 0°C, snowfall is likely. Furthermore, annual average flow ranges between 17 mm and 1,524 mm, and the runoff coefficient falls between 0.03 and 0.97.

The climatic values are based on monthly estimates of potential evapotranspiration, while the Centre for Hydrographic Studies (CEDEX, Ministry of Public works and Ministry of Agriculture and Environment, Spain) provided temperature and precipitation measurements for the period from 1980–2005. Daily flow time series determined the average flows [Peñas et al., 2014a], and at least eight years of data in the period from 1976–2009 were needed to complete this calculations.

The hydrological modeling treats a subset of 16 basins (out of 92 total basins) as ungauged, so as to test the proposed methodology. At minimum, for each of the subset basins, eight years of synchronized daily precipitation, daily flow, and monthly potential evapotranspiration data was available [Peñas et al., 2014a] (Figures 1 and Table 1). The Spanish Meteorological Agency (AEMET) supplied daily precipitation data for the basins.

Table 1: Catchment characteristics: (1) calcareous rocks, (2) clay, (3) conglomerates rocks, (4) sand, (5) sedimentary rocks, (6) shale (sedimentary) rocks, (7) siliceous rocks, (8) slates, (9) volcanic rocks. See Peñas et al. [2014b] for permeability and geology.

Chapter 2: Statistical Metrics to Assess Regionalization and Model Adequacy in PUBs

						lenna A	Maximum			Geo	logy	(% Ar	ea o	ccupi	ed)		
Flow Gauge Name	Nº years	N ^e max pluviometers	Area km²	RR Coefficient	rainfall mm/year	PET PET mm/year	monthly minimum T(ºC)	Permeability	(1)	(2)	(3)	(4)	(5)	(9)	(2	(8)	(6)
X1353	22	m	529	0.61	1459	660	1.1	low	7	0	10	11	4	0	-	72	0
X1404	10	ſ	293	0.81	1449	638	0.2	very low	0	0	ъ	16	11	0	0	68	0
X1303	14	ε	377	0.64	1342	734	4.0	low	19	10	0	2	ε	0	35	31	0
X1265	∞	1	294	0.57	1041	716	-1.3	very low	18	0	19	4	∞	0	0	50	0
X9257	10	1	80	0.33	814	962	0.7	low	Ч	0	7	49	0	48	0	0	0
X9040	6	14	623	0.69	1356	705	-2.7	low	17	0	Ŋ	47	Ч	22	9	0	7
X9269	10	1	75	0.39	1333	619	-1.4	low	7	0	0	73	0	20	0	0	0
X9197	∞	1	283	0.24	681	655	-1.1	low	17	ഹ	ŝ	0	Ч	73	0	0	0
X9221	11	1	22	0.79	1134	564	-0.4	low	65	0	0	0	0	35	0	0	0
AN439	13	1	152	0.95	1469	701	0.1	low	06	7	0	Ч	Ч	Ŋ	0	0	Ч
AN433	∞	1	554	0.41	1376	679	0.0	high	53	0	0	9	11	30	0	0	0
AN520	13	1	73	0.35	1683	609	-1.8	low	11	0	0	55	0	33	0	0	0
AN530	12	1	95	0.39	1371	693	-0.1	low	0	0	0	80	0	20	0	0	0
AN313	14	9	477	0.35	1038	650	1.1	high	57	0	Ч	0	10	21	10	0	0
c8z1	12	2	114	0.66	1794	684	2.5	high	28	ŝ	0	28	0	37	0	Ч	7
c7z1	11	2	28	0.65	1809	731	2.7	low	ഹ	Ч	0	21	0	18	0	51	4
2.3 METHOD

Regionalized principal components (PCs) are estimated for the ungauged basins by using Random Forests models, on the basis of their corresponding catchment descriptors (CDs); these regionalized PCs are used in a Bayesian framework to condition hydrological model parameters. In addition to the case based on observations, synthetic studies with different regionalization and model qualities are also considered.

Two new metrics to assess model and regionalization quality are introduced. Additionally, the performance of the predictions is evaluated via the probabilistic Nash-Sutcliffe coefficient and the prediction credibility intervals (CIs).

2.3.1 Hydrological indices and catchment descriptors

Adopting Olden and Poff's [2003] approach, this study utilizes flow records from the 92 basins to estimate 103 hydrological indices reflecting multiple flow regime elements. These indices indicate the mean and standard deviation of (1) annual and monthly flows, (2) high and low flows, (3) the duration and frequency of high flows, (4) the rate of change in the flows, and (5) the timing of maximum and minimum flow events. Next, the hydrological indices are transformed into uncorrelated (orthogonal) principal components (PCs) via principal component analysis. Afterwards, the broken stick method identifies those PCs explaining most of the variability in the hydrological indices [Jackson, 1993; Peres-Neto et al., 2005].

This evaluation utilizes 16 of the 50 available catchment descriptors (CDs) that Peñas et al. [2014a] selected. These CDs comprise the least correlated subset and have Pearson's correlation coefficients below 0.7. The selected features are the following: area, climate (mean annual precipitation and PET, ratio of minimum quarterly precipitation to maximum quarterly precipitation), topography (average catchment elevation and gradient), basin geometry (drainage density, number of river confluences), land use (area covered by agricultural land, broadleaf forest, coniferous forest, bare land, pasture, and urban areas), and geology (average rock density and permeability). To derive the climatic variables, monthly climate series are utilized. These climate series are a 1 km × 1 km grid map that the Centre for Hydrographic Studies (CEDEX, Ministry of Public Works and Ministry of Agriculture and Environment, Spain) developed. A 25 m digital elevation model is employed to derive the topography and basin geometry, while land use is derived from the Soil Occupancy Information System (in Spanish, SIOSE) at a 1:25,000 scale. The Spanish government's National Geographic Institute developed this information. Finally, geological variables are estimated using lithostratigraphic and

permeability maps created by the Spanish government's Geologic and Mining Institute (in Spanish, IGM). These maps were scaled at 1:200,000.

2.3.2 Regionalization using Random Forests

Regionalization using Random Forests has many theoretical and practical benefits [Breiman, 2001], such as:

- The model structure is assumption-free and does not require the preselection of a linear/non-linear model form and the number of predictors.
- Random Forests algorithm makes no assumptions about the model errors' statistical nature (i.e., normality, independency, and homoscedasticity), as required in multiple linear regressions.
- It is time efficient and quality efficient.

Random Forests are trained via the "randomForest v4.6.7" R package using the selected PCs (predictands) and the 16 CDs (predictors) corresponding to the 92 gauged basins [Liaw and Wiener, 2002]. The Random Forests algorithm resamples a set number of trees with replacement to create a forest. The trees grow so that multiple CDs are randomly sampled at each node. Those providing the best split (the lowest mean square error) are then selected [Liaw et al., 2002; Snelder et al., 2012]. The resampling and the use of a random subset of predictors introduce randomness to the way a tree is built in Random Forests [Breiman, 2001] as compared to a single regression tree [Breiman, 1984]. A regionalized value for an ungauged basin is estimated as the mean of all individual predictions from each tree in the forests [Snelder et al., 2011].

In this study, those 16 basins with synchronized hydrometeorological daily data constitute "ungauged" sites. For each of the 16 basins, the remaining 91 basins are employed to regionalize the PCs (the "leave-one-out" technique). Uncertainty in the regionalization model structure (e.g., uncertainty in the selection of the CDs, errors in the exact CD values, and the fact that the available gauged basins only represents a sample of the entire population) impacts the regionalized PCs [Almeida, 2014]. A jack-knife (or "leave-one-out") strategy is adopted to assess this regionalization uncertainty. Each time, a single catchment is excluded, with the remaining 90 catchments used to estimate regionalized PCs in the catchment set aside. And the next step is computing the residuals between the regionalized PCs and the observation-based PCs. For each catchment, this process is then repeated. The result is a set of 91 residuals, and a joint

probability density function (JPDF) is then fitted to it. The JPDF characterizes the regionalization error for Random Forests, and it also demarcates the regionalization model likelihoods.

2.3.3 Posterior distribution

Given regionalized principal components *PC^{reg}*, a model structure *M*, and inputs *I*, the posterior distribution of model parameters is provided by Bayes' law [Bulygina et al., 2009]:

$$p(\theta | PC^{Reg}, M, I) = \frac{L(PC^{sim}_{\theta, M, I} | PC^{Reg}) * p(\theta | M, I)}{p(PC^{Reg} | M, I)}$$
(2.1)

Where $L(PC_{\theta,M,I}^{sim}|PC^{Reg})$ is the likelihood of principal components $PC_{\theta,M,I}^{sim}$ simulated by model M with input I and parameters θ , $p(\theta|M,I)$ is the prior parameter distribution, and the denominator serves as a normalizing constant (with respect to parameters θ). A key assumption here is that the model and its inputs are capable of representing hydrological flow dynamics. The likelihood function $L(PC_{\theta,M,I}^{sim}|PC^{Reg})$ is taken as proportional to the residuals distribution of the PCs regionalization procedure described in section 2.3.2, and the prior is defined as a uniform distribution based on feasible model parameter ranges. As it is generally not possible to sample directly from the posterior distribution (2.1.), an importance sampling is utilized to approximate the distributions numerically [Doucet et al., 2000].

2.3.4 Suitability test

To determine if a model is representative of regionalized information in the form of a vector of regionalized PCs, PC^{reg} , 10,000 samples of PC^{reg} are drawn from the regionalized PCs distribution (10,000 of m-dimensional PC^{reg} draws), and compared to the probability density function of the simulated principal component $p(PC^{sim})$ as follows. First, $p(PC^{sim})$ is approximated via a mixture of N Gaussians [Muller et al. 1996; Bulygina and Gupta, 2011], and critical ellipses are defined for each Gaussian component based on the Mahalanobis distance:

$$d_{i} = \sqrt{(PC - \mu_{i})^{t} * \Sigma_{i}^{-1} * (PC - \mu_{i})} \quad (2.2)$$

where μ_i denotes m x 1 mean value and Σ_i denotes m x m covariance matrix for the ith Gaussian component (i=1,N). The squared Mahalanobis distance d_i^2 for a Gaussian distribution $N(\mu_i; \Sigma_i)$ follows a chi-squared distribution with *m* degrees of freedom [Gallego et al. 2013]. The equidensity contours for each Gaussian component characterized by constant Mahalanobis distance are ellipsoids centered on μ_i , with their axes aligned with the eigenvectors of Σ_i [Friendly et al. 2013], and allow the most compact $[1-\alpha_i]$ confidence set to be defined for a Gaussian distribution $N(\mu_i; \Sigma_i)$ [Gallego et al. 2013; Ribeiro, 2004].

When a null hypothesis is defined as a model being able to reproduce information contained in the first *m* regionalized PCs, a corresponding test p-value is based on averaging the p-values for each of the 10,000 samples. A p-value for each sample PC^{reg} is defined as the largest value of α_i , i=1,N so that there is at least one $[1-\alpha_i]$ elliptical region that contains the sample PC^{reg} . The test is termed a model suitability test.

A second suitability test evaluates the null hypothesis that the regionalization procedure is likely to reproduce the exact values of PCs (Note: As these values are not typically available, observation-based PCs are used instead in the calculations). In this case, the difference between the PC based on observations and the regionalized PC is compared with the residuals distribution in the regionalization model. For this, the residuals distribution of the regionalization model $p(PC - PC^{reg})$ is first approximated via a mixture of Gaussians [Muller et al. 1996]; second, critical ellipses are defined based on the Mahalanobis distance (2.2); and third, similar to the first suitability test, the maximum p-value for *PC* is defined as the largest value of α_i , i=1,N so that there is at least one [1- α_i] elliptical region that contains *PC* calculated for a flow time series. This maximum p-value characterizes how well the resulting distribution of regionalized PCs describes PC.

A third suitability test evaluates the null hypothesis that the model is able to reproduce information contained in the first *m* observed PCs with high probability. The test is similar to the first suitability test, except for only one draw PC^{obs} is compared with the simulated PC probability density function $p(PC^{sim})$.

2.3.5 Adequacy metrics

REGIONALIZATION ADEQUACY

The regionalization procedure adequacy is defined via a Bayes' Factor (BF) [Gelman et al., 2013] that quantifies how much information is added by the *PC^{reg}* compared to the prior knowledge about the PCs:

$$BF_{reg} = \frac{P(PC|PC^{reg})}{P(PC)} \quad (2.3.)$$

Where $p(PC|PC^{reg})$ is equal to the likelihood (see section 2.3.2) calculated at the difference given by $(PC - PC^{reg})$ and p(PC) is expressed as a product of uniform distributions for each of the PCs based on the minimum and maximum PC values from the set of 92 basins. The

regionalization procedure can be seen as adequate if the Bayes' Factor is above 1, and the BF can be used to interpret the strength of the evidence in favor of the hypothesis in the numerator against the hypothesis in the denominator as shown in Table 2 [Jeffreys, 1961]. It can be shown (Appendix A) that equation 2.3 is equivalent to how much information about the exact values of PCs is added by the regionalization procedure over the model:

$$BF_{reg} = \frac{P(PC|PC^{reg}, M, I)}{P(PC|M, I)}$$
 (2.4.)

BF	Strength of the evidence
<1	negative
	(supports the hypothesis in the denominator)
1 to 10^(0.5)	barely worth mentioning
10^0.5 to 10	substantial
10 to 10^3/2	strong
10^3/2 to 100	very strong
>100	decisive

Table 2: Scale for Bayes' Factor interpretation [Jeffreys, 1961]

Hydrological model adequacy

The adequacy of a hydrological model choice (please note, 'model choice' includes model structure selection, model input specification, and model parameter sampling) is also evaluated via the Bayes' Factor [Gelman et al., 2013] that quantifies how much information is added by a model over the prior knowledge about the PCs:

$$BF_M = \frac{p(PC|M,I)}{p(PC)} (2.5)$$

where p(PC|M,I) is equal to the probability of the observed PC in the distribution of the simulated PCs $p(PC_{\theta,M,I}^{sim})$, which is given by a mixture of Gaussians fitted to the simulated PCs and p(PC) is expressed as a product of uniform distributions for each of the PCs. A model can be seen as adequate if the Bayes' Factor is above 1, and Table 2. provides a quantitative scale to interpret to what extent the model is supported by the evidence (i.e. by the observed PCs). It can be shown (see Appendix B) that equation 2.5 is equivalent to how much information about the exact PCs is added by a model and its inputs over the regionalization:

$$BF_M = \frac{p(PC|PC^{reg}, M, I)}{p(PC|PC^{reg})} (2.6)$$

2.3.6 Case studies

Four case studies are considered, differing in the model error size and quality of PCs regionalization.

BC: This case is based on observed flows (with model error) and regionalization derived via the RF method.

R: This case differs from case BC in the regionalization procedure that is defined as accurate and more precise, so that PC regionalization errors follow a zero-mean Gaussian distribution with a standard deviation fixed at 5% of the prior PCs spread. *PC^{reg}* are set to the PCs values calculated from a corresponding flow time series, so that the p-value for the regionalization suitability test is 1.

M: This case differs from case BC due to the model error set to zero. This is achieved by substituting the observed flows with flows generated by a model with a selected parameter set, driven with the observed inputs. The regionalization characteristics are the same as in case BC. A model parameter set defining the synthetic flow time series is selected to minimize the normalized distance between p principal components calculated from the corresponding

simulated (PC_i^{sim}) and observed flows (PC_i) : $\sqrt{\sum_{i=1}^{p} \left(\frac{PC_i^{sim} - PC}{std(PC_i^{sim})}\right)^2}$, where std(.) denotes standard deviation. The corresponding PC^{Reg} are derived by adding the regionalization noise (defined in case BC) to the PCs calculated for the synthetic flows.

MR: This case differs from case M in the regionalization procedure that equals that of case R. Similar to case R, the p-value for the regionalization suitability test is 1.

2.3.7 Prediction quality assessment

The probabilistic analogue of Nash-Sutcliffe efficiency (*NS^{prob}*)[Bulygina et al., 2009] is utilized to assess the quality of probabilistic flow predictions:

$$NS^{Prob} = \left\{ 1 - \frac{\sum_{t=1}^{T} (E[Q_t] - q_t^{obs})^2}{\sum_{t=1}^{T} (q_t^{obs} - E[q_t^{obs}])^2} \right\} - \frac{\sum_{t=1}^{T} (Var[Q_t])}{\sum_{t=1}^{T} (q_t^{obs} - E[q_t^{obs}])^2}$$
(2.7)

where $\{Q_t\}$ is the simulated time series of flow at time t, q_t^{obs} is observation flow at time t, E[.] denotes expected value, *Var* [.] denotes variance, and *T* is the number of time steps. The first term on the right-hand side of (2.7.) measures the prediction accuracy and the second term measures the prediction precision.

Another important aspect of a probabilistic prediction is its reliability. The reliability quantifies the statistical consistency between the observed data and its predictive distribution. The percentage of times the observed flows fall into the 95% credibility intervals (95%CI) [Yadav et al., 2007] is used to evaluate the prediction reliability.

2.4 MODEL DESCRIPTION AND CONDITIONING

The PDM conceptual model with two parallel linear routing stores (Figure 2) is employed for flow simulation [Moore, 2007]. This model is chosen due to its structural simplicity and its wide applications around the world: UK [Lee et al., 2005; Pechlivanidis et al., 2010], Belgium [Cabus, 2008; Willems et al., 2014], Europe [Arnell, 1999], US [Kollat et al., 2012], Southeast Asia [Thompson et al., 2013], Southern Africa [MacKellar et al., 2013], and Australia [Srikanthan et al., 2007].



Figure 2: PDM conceptual model.

The PDM model employed uses a Pareto distribution to represent the spatial variability of the soil moisture storage capacity and two linear reservoirs in parallel for the effective runoff routing. The Pareto distribution is defined by two parameters: parameter *Cmax*, representing the maximum storage capacity in a catchment, and shape parameter *b*, controlling the spatial variability of storage capacity in a catchment. The actual evaporation is calculated as the potential evaporation multiplied by the relative saturation of the catchment, and the effective rainfall is equal to the soil moisture excess. Furthermore, the effective rainfall is split into two parts: quick and slow flows using a constant proportion parameter α . The flows are routed via the two parallel linear reservoirs so that the total flow is the sum of these two components. The routing has two parameters: quick-flow reservoir residence time *Tq* and slow-flow reservoir residence time *Ts* [Moore, 2007].

Prior ranges of the five parameters are adapted from Kollat et al. [2012] (Table 3). The initial soil moisture storage is assumed to be zero, and the first year of the simulations is used as a warming-up period. To approximate the posterior distribution (2.1), 1,000 parameter sets are sampled from the prior uniform distributions using the Latin Hypercube method, the PDM model (with fixed observed inputs) is run to simulate 1,000 flow time series, and 1,000 sets of hydrological indices are calculated from the flows. Then, each parameter set is assigned a weight proportional to the parameter likelihood in (2.1) so that the sampled parameter sets and the assigned weights approximate the posterior parameter distribution (2.1).

Table 3: Parameter Ranges

C max(mm)	В(-)	Tq(days)	Ts(days)	Alpha(-)	
0-2000	0-4	0-7	7-20,000	0-1	

2.5 RESULTS AND DISCUSSION

2.5.1 Selection of the principal components

This analysis estimates 103 hydrological indices for each of the 92 basins. The broken-stick method is employed to identify the four PCs explaining 63.3%, 12.4%, 6.9%, and 4.8% of variability, respectively. In sum, those four PCs explain 87.6% of total variance. Thus, the PCA extensively reduces the 103-dimensional (or 92-dimensional, if based on the number of basins) indices space. Moreover, it maximizes the value of the information, a task that researchers have classified as particularly pertinent for PUBs [Wagener and Montanari, 2011].

2.5.2 Regionalization and likelihood function

The Random Forests method is employed to regionalize the PCs. As Table 4 demonstrates, average regionalization quality values (calculated for each PC as the ratio of its distribution scale parameter to its range) fall between 10% and 16%. The regionalization of PC1 is the least accurate but the most precise. Moreover, the regionalization of PC3 is the least precise, while the regionalization of PC4 is the most accurate. The obtained regionalization quality is comparable to, or better than, the values provided by a step-wise regression regionalization [Almeida, 2014]. In that study, regionalization precision ranged from 7% to 27%.

Table 4: Regionalization quality for individual PCs for a selected catchment. The normalized scale is derived by dividing the scale by the corresponding PC range.

	Location	Scale	Normalized scale
PC1	2.42	3.69	0.1
PC2	0.18	2.28	0.13
PC3	0.05	1.7	0.16
PC4	0	1.9	0.12

The likelihood functions in equation 2.1. are estimated using the residual distributions for the regionalized PCs (i.e., the differences between the regionalized and observation-based PCs). Pearson's linear correlation test reveals a significant correlation between the residuals of PC2 and PC3 (*p*-value=0.03). Therefore, the four-dimensional likelihood is estimated as a product of the marginal residual distributions for PC1 and PC4 and as a joint distribution for PC2 and PC3. The marginal distribution of the PC1 residuals represents an extreme value type I distribution (EV1) based on the chi-square test. Moreover, the chi-square test indicates that the marginal distribution of the PC4 residuals constitutes a Gaussian distribution. Finally, the joint distribution of the PC2 and PC3 residuals represents a two-dimensional Gaussian distribution, as indicated by chi-square and Mardia tests (all tests significant at the 5% level). Figures 3 and 4 provide estimates of the marginal and joint distributions for a selected basin (values computed using the leave-one-out technique).



Figure 3: Marginal distributions and empirical histograms (EMP) for regionalized residuals for selected catchment



Figure 4: Regionalized residuals for PC2 and PC3 for a selected catchment: (a) a fitted joint Gaussian distribution, and (b) a scatterplot.

2.5.3 Prediction quality assessment

When flow observations are not available, the suitability test for the model shows whether the hydrological model is a suitable representation of the regionalized information in the ungauged catchment. The hydrological model is found suitable (p-value is greater than 0.05) for the null hypothesis that the model is able to represent the regionalized PCs for 3 catchments, catchments X1353, AN433, and C8Z1; the model suitability hypothesis based on the regionalized PCs is rejected for the other 13 basins (Figure 5a, item Reg&Mod). The latter means that the selected model is not likely to reproduce the four pieces of regionalized information (PCs) at the same time.

In the case of BC_{reg} , the hypothesis that the derived regionalization procedure is a suitable representation of the PCs based on observations of flows is accepted for all basins, as p-values are greater than 0.05 (Figure 5a, item BC_{reg}). While the adequacy test for the case BC_{reg} shows that the regionalization procedure is adequate in 12 basins (Bayes' Factor is over 1 – Figure 5b, case BC_{reg}) and is inadequate for four basins (X1404, AN433, X1353, and X9040). Therefore, in these 12 catchments, the regionalization passes the suitability test and is adequate at the same time.

However, the BC_M case shows that model structure is a suitable representation of the four pieces of information observed in the catchment (i.e., based on observations of flows) in six catchments (p-value is over 0.05 - Figure 5a, case BC_M) AN530, AN520, AN433, X9257, X1353, and C8Z1. The model is only adequate (Bayes' Factor is over 1 - Figure 5b case BC_M) in three basins (X1353,

X9257 and C8Z1). Therefore, the model is at the same time a suitable and adequate representation of PCs in three catchments (X1353, X9257, and C8Z1). It is also worth noting that the p-value for the model suitability test (BC_M) is above 0.05 for basin AN520, but the Bayes' Factor is 1; therefore, the model cannot be considered either informative or disinformative.

Then, the model is simultaneously (1) a suitable representation of the regionalized PCs (Figure 2.5a item Reg&Mod), (2) a suitable representation of the PCs observed in the catchment (p-value over 0.05 Figure 2.5a case BC_M), and (3) an adequate representation of the PCs observed in the catchment – i.e., the model adds enough information about the PCs based on data (BFs over 1, Figure 5b case BC_M) in catchments C8Z1 and X1353.

The regionalization procedure and model are a suitable and adequate representation of the PCs based on observations for only two basins (out of 16), basins C8Z1 and X9257 (Figure 5a and 5b, intersection of results for cases BC_M and BC_{reg}).

Furthermore, (1) the model reproduces the regionalized PCs with high probability (Figure 2.5a case Reg&Mod), (2) the regionalization procedure is adequate (Figure 5a and 5b case BC_{reg}), and (3) the model is adequate (Figure 5a and 5b case BC_{M}) simultaneously only in catchment C8Z1.

For case R_{reg} , BFs in Figure 5b reveal that the strength of evidence is decisive (Table 2), as the regionalization procedure adds 53,051 times more information compared to prior knowledge (the value is the same for each basin, as it is based on a fixed error distribution). Notice that the regionalization is also suitable for all the catchments, as the p-value is 1. Further, Figure 5a and 5b show that the model is suitable (p-values over 0.05) and adequate (BF over 1) in all the basins for the case M_{M} . Additionally, in the case MR_{reg} , improving the regionalization quality leads to very high BFs equal to 53,051 and p-values equal to 1 for all catchments.

To further investigate the impact of regionalization and model suitability and adequacy in the flow predictions, posterior values of *NS*^{prob} and 95% CI are shown in the Figure 6. The parcels a) and b) in this figure summarize the change in accuracy and precision, which is given by *NS*^{prob}, for the cases BC, R, M, and MR; and the parcels c) and d) do the same for reliability, which is quantified by 95% CI for the cases BC, R, M, and MR. The figures show that for basins C8Z1 and X1353, where the model is a suitable representation of the regionalized information and a suitable and adequate representation of the PCs based on observations (see comment above), improving regionalization (case R in figure 6a and 6c) leads to an improvement in the performance of *NS*^{prob} (from 0.68 to 0.71 in catchment C8Z1, and from 0.65 to 0.72 in catchment X1353) and 95% CI (from 70% to 74% in catchment C8Z1, and from 45% to 52% in catchment X1353). However, in the remaining 14 catchments, improving only the regionalization procedure

(R) might either improve or worsen model performance. *NS*^{prob} improves in 8 out of the 14 catchments. 95% CI improves in 7 out of the 14, worsens in 6, and stays the same in 1. The average *NS*^{prob} in case BC is 0.3 and in case R, 0.32. The average of 95% CI in case BC is 62% and in case R, 60%. This is indicative that model error precludes an improvement in predictions when regionalization improves.

It is also worth noting that while the model is considered adequate for basins X1353, X9257 and C8Z1 based on the observed flow summaries (PCs – Figure 5a and 5b, cases BC_{reg} and BC_M), the model appears to fail to represent flow dynamics at a finer temporal scale in X1353 and X9257. This is exemplified in Figure 8 for catchment X1353 when a smaller sub-set from the model parameter space is utilized for predictions, corresponding to case R (i.e. a higher regionalization quality).

Figure 6 also shows that results improve for case MR with respect to case M in all the catchments – *NS*^{prob} and 95% CI improve in all 16 catchments. The average *NS*^{prob} in case M is 0.68, while in MR it is 0.88. The average of 95% CI in case M is 72%, while in MR it is 95%. Proving that when the model is able to reproduce the PCs, an improvement in hydrological indices regionalization leads to better flow predictions (accuracy, precision, and reliability), in this case, regionalization quality can be targeted to improve model performance.

The following set of analyses of variance (ANOVAs) have been performed on *NS^{prob}* and 95% CI to quantify whether an improvement in the different sources of knowledge makes a significant statistical difference in the quality of the predictions. The analyses have compared (1) BC with R, which reflects whether improving regionalization quality makes a difference when there is error from the model, and (2) M with MR, which reflects whether improving regionalization quality, makes a difference when the model is perfect. ANOVA cannot reject the null hypothesis that improving regionalization (R) with respect to case BC has no effect on *NS^{prob}*, or the null hypothesis that improving regionalization (R) with respect to case BC has no effect on 95% CI. In other words, the differences in *NS^{prob}* and 95%CI when regionalization is improved for the case of non-zero model error are not statistically significant. Meanwhile, ANOVA rejects the null hypothesis that improving regionalization when the model is perfect (going from M to MR) does not make a difference in *NS^{prob}* and 95% CI. Or, improving the regionalization accuracy and precision when the model is perfect makes a statistically significant difference on the accuracy, precision, and reliability of predictions. For MR, the *NS^{prob}* and 95% CI are improved. *NS^{prob}* is over 0.9 in 13 out of the 16 catchments for case MR, while in case M *NS^{prob}* is over 0.9 in 2 out

of the 16 (X9040 and AN433), and 95% CI is over 90% in 13 catchments, while in case M, 95% CI is over 90% in 6 (X9257, X1265, X9221, AN530, AN313, C8Z1).



Figure 5: Adequacy and suitability metrics: (a) p-values, and (b) BFs. The metrics are assessed for both model and regionalization procedures, so that a subscript "reg" denotes metrics relevant for the regionalization procedure, and a subscript "M" denotes metrics are not repeated.



Figure 6: a) NSprob for cases BC and R, b) NSprob for cases M and MR, c) 95% CI for cases BC and R, and d) 95% CI for M and MR.

Figures 7 to 10 illustrate the 95% prediction intervals for cases BC, R, M, and MR in four catchments: Figure 7 shows catchment C8Z1, where the model can reproduce the regionalized PCs with high probability (Figure 5a, case Reg&Mod) and both the regionalization and the model are adequate (Figure 5a and Figure 5b, case BC_{reg} and case BC_M). Figure 8 shows catchment X1353, where the model represents the regionalized PCs with high probability in the ungauged basin (Figure 5a, case Reg&Mod) and the model is a suitable and adequate representation of the PCs based on observations (Figure 5a and Figure 5b, case BC_M); however, the regionalization is inadequate (Figure 5b, case BC_{reg}), although it is suitable (see Figure 5b, case BC_{reg}). Figure 9 shows catchment AN439, where the model is not likely to reproduce the regionalization is a suitable and adequate representation of the PCs based on dequate representation of the PCs based on the observations of flows (Figure 5a, case BC_{reg}), but the model is not (Figure 5a, case Reg&Mod). The regionalization is a suitable and adequate representation of the PCs based on the observations of flows (Figure 5a and 5b, case BC_{reg}), but the model is not (Figure 5a and Figure 5b, case BC_M). Furthermore, Figure 10 shows catchment X9040, where the model is not likely to reproduce the regionalized PCs (Figure 5a, case Reg&Mod), and the regionalization and model are not able to represent the PCs based on observations (Figure 5a and 5b, case BC_{reg} and BC_M).

For catchment C8Z1, Figure 7 shows that accuracy, precision and reliability of the predictions are very good in BC: NS^{prob} is equal to 0.68 and 95% CI is equal to 70%. Also in case R, NS^{prob} is equal to 0.71 and 95% CI to 74%, as well as in case M: NS^{prob} is equal to 0.72 and 95% CI to 87%.

In case MR, NS^{prob} is improved slightly more, up to 0.93, while reliability remains similar (95% CI is equal to 75%).

For catchment X1353, Figure 8 shows that accuracy is quite similar in BC and R, but the performance relative to NS^{prob} is poor for the original regionalization (0.65). When R is implemented, NS^{prob} improves to 0.72. NS^{prob} is not mainly damaged by the variance of the regionalization model, since the additional gain is only 0.07 with respect case BC (NS^{prob} equal to 0.65). Reliability is also similar: 95% CI is equal to 45% in case BC and equal to 52% in case R. Comparing case M with case MR, both accuracy and precision are high in case M – NS^{prob} is equal to 0.76 – and very high in MR – NS^{prob} is equal to 0.999. Reliability is also largely improved: 95% CI is equal to 47% for case M and equal to 100% in case MR.



Figure7: Basin C8Z1



Figure 9: Basin AN439

Credibility bounds for observations and synthetic data in basin AN439 are shown in Figure 9. For this basin, BC shows that model structure is not able to capture many of the observed high flows (some of the peaks are out of the prior prediction intervals), which is reflected in the relatively low performance, as NS^{prob} is 0.51 and 95% CI is 58%. An improvement in regionalization (R) does not lead to a notable improvement in the predictions: NS^{prob} improves to 0.57 and 95%CI to 59%. However, performance of synthetic data is high in case M, when PCs are synthetically generated by the PDM and corrupted by the noise of the original regionalization procedure, which is reflected in NS^{prob} equal to 0.86 and 95% CI to 68%, and a further improvement when the regionalization is additionally improved, which is reflected in NS^{prob} equal to 0.97 and 95% CI equal to 99%.



Figure 10: Basin X9040

Finally, basin X9040 in Figure 10 exemplifies (1) the impact of a model that is not able to reproduce the regionalized PCs and is not a suitable and adequate representation of the observation based PCs on the quality of predictions, and (2) the impact of a regionalization procedure that is not a suitable and adequate representation of the PCs based on observations on the quality of predictions. As expected, in BC the performance of the predictions is very poor (NS^{prob} is -0.67 and 95% CI is 25%), and improving the regionalization alone does not lead to a

high level of improvement, as NS^{prob} is -0.71 and 95% CI is 33%. However, when the model/inputs is able to reproduce the PCs, case M shows that NS^{prob} is very high (NS^{prob} is 0.91), but reliability is still very poor (95% CI is equal to 14%). In this case, an additional improvement in regionalization, case MR, leads to perfect reliability (the 95% CI is 100%), and accuracy and precision are also very high (NS^{prob} is 0.98).

2.6 CONCLUSIONS

In predictions for ungauged basins, a common practice is to use a pre-selected model structure and regionalization technique for a catchment, while there is usually no justification for their suitability (ability to reproduce the available information) and adequacy (the knowledge gained by considering a model and/or regionalization). This study proposes statistical tests to quantify suitability and adequacy, and suggests a path to disaggregate the different sources of uncertainty and set priorities in order to reduce the predictive uncertainty.

This research advances the standard practices for predicting flows in ungauged basins when the information given by multiple regionalized hydrological indices is assimilated into rainfall-runoff models by a Bayesian conditioning process. The work uses Random Forest regionalized principal components that update the information given by a hydrological model via a Bayesian approach. The research also contributes new suitability and adequacy tests:

- A suitability test is based on the simultaneous inference concept and explicitly quantifies the following null hypotheses: (a) a model is able to reproduce regionalized (or observed, if available) information with high reliability, (b) a regionalization procedure is able to reproduce the available information in a catchment.
- An adequacy test is based on Bayes' Factors and explicitly quantifies the knowledge gained
 (a) by including a regionalization procedure, and (b) by including a model.

These suitability and adequacy tests allow identification of the dominant source of uncertainty (between the regionalization and the model) and, hence, where the efforts must be concentrated to reduce it. Application to a set of 16 catchments in northern Spain leads to the following findings:

 The error in the Random Forest regionalized principal components is relatively small and regionalization by Random Forest only relies on a small number of assumptions (it does not require the preselection of a linear/non-linear model form and the number of predictors; it does not make assumptions about the model errors' statistical nature; and it is an efficient – quality and time - data learning algorithm). Demonstrating also that the high-dimensional hydrological indices space extensively reduced via PCA, and the broken stick method provides sufficient information for an efficient regionalization via Random Forest in the case of basins of northern Spain.

- 2. When the model passes the suitability test for the regionalized indices, as well as the suitability and adequacy tests for the available information, improving regionalization quality can be targeted as a means for securing better predictions.
- 3. However, model error precludes performance improvement when regionalization quality improves. Therefore, regionalization quality is not directly related to the prediction quality. Priority should be given to the selection of a representative model structure (which includes model inputs), then followed by regionalization improvement.
- 4. Improving the accuracy and precision of regionalization for an error-free hydrological model provides better flow prediction quality (reliability, accuracy, precision).
- 5. The developed regionalization procedure via the Random Forest extrapolation passes the suitability test (i.e. regionalization is likely to reproduce the hydrological information) in all catchments; and it also passes the adequacy test (i.e. provides a gain in knowledge) in 75% of basins.

2.7 APPENDIX A

In this appendix the equivalence between equation (2.3) and (2.4) is shown. Applying Bayes' Law and using the knowledge the PC^{reg} are independent on the Model, *M*, and Inputs, *I*, the following derivation shows that both equations are equivalent:

$$BF = \frac{P(PC \mid PC^{reg}, M, I)}{P(PC \mid M, I)} = \frac{P(PC_{\theta,M,I}^{sim} \mid PC, PC^{reg}) * P(PC \mid PC^{reg})}{P(PC_{\theta,M,I}^{sim} \mid PC^{reg}) * P(PC \mid PC_{\theta,M,I}^{sim})} = \frac{P(PC_{\theta,M,I}^{sim} \mid PC) * P(PC \mid PC^{reg})}{P(PC_{\theta,M,I}^{sim}) * P(PC \mid PC_{\theta,M,I}^{sim})} = \frac{P(PC \mid PC^{reg})}{P(PC)} \quad A.1.$$

Then the metric can be defined also as a ratio between $P(PC|PC^{reg})$ and P(PC) to quantify if the regionalized PCs are good descriptors of the data-based PCs (no hydrological model involved) and then it helps to evaluate the quality of the regionalization procedure. In equation (A1), numerator is derived from the likelihood function, i.e., by calculating the probability of the $(PC - PC^{reg})$ error and the denominator is approached by a product of uniform distributions. BF over 1 indicates that the regionalization procedure provides information about the data PCs over the prior knowledge, and it is disinformative otherwise.

2.8 APPENDIX B

Similar to Appendix A, here it is shown that equation (2.5) and (2.6) are equivalent. Applying Bayes' Law and using the knowledge that PC^{reg} are independent on the Model, *M*, and Inputs, *I*, the following derivation shows that both equations are equivalent:

$$BF = \frac{P(PC \mid PC^{reg}, M, I)}{P(PC \mid PC^{reg})} = \frac{P(PC^{reg} \mid PC \mid M, I) * P(PC \mid M, I)}{P(PC^{reg} \mid M, I) * P(PC \mid PC^{reg})} = \frac{P(PC^{reg} \mid PC \mid) * P(PC \mid M, I)}{P(PC^{reg}) * P(PC \mid PC^{reg})} = \frac{P(PC \mid M, I)}{P(PC \mid PC \mid M)}$$
B.1.

Thus, the measure can be defined as a ratio between $P(PC \mid PC_{\theta,M,I}^{sim})$ and P(PC) to quantify how much information is added by the model (including its structure, inputs and parameters sampling) over the prior knowledge about the data-PC. If BF is over 1, the model is informative and is dissinformative if BF is lower than 1. The severity of disinformation can be assessed based on the inverse of the BF value. The numerator in equation B1 is approximated using a mixture of N Gaussians fitted to the PCs simulated by the model, *M*, and the denominator is factorized as a product of uniform distributions.

2.9 REFERENCES IN CHAPTER 2.

- Almeida, S. L., Bulygina, N., McIntyre, N., Wagener, T. and Buytaert, W. (2012). Predicting flows in ungauged catchments using correlated information sources. In: British Hydrological Society's Eleventh National Hydrology Symposium, Hydrology for a Changing World, Dundee 2012
- Almeida, S. M. C. L. (2014). The Value of Regionalised Information for Hydrological Modelling. PhD thesis, Imperial College London, London, UK.
- Arnell, N. W. (1999). A simple water balance model for the simulation of streamflow over a large geographic domain. Journal of Hydrology, 217(3), 314-335.
- Arora, V. K. (2002). The use of the aridity index to assess climate change effect on annual runoff. Journal of Hydrology, 265(1), 164-177.
- Beven, K. J. (2000). Uniqueness of place and process representations in hydrological modelling. Hydrology and Earth System Sciences Discussions, 4(2), 203-213.
- Beven, K. and Westerberg, I. (2011). On red herrings and real herrings: disinformation and information in hydrological inference. Hydrol. Process. 25: 1676–1680. Doi: 10.1002/hyp.7963
- Blöschl, G. (Ed.). (2013). Runoff prediction in ungauged basins: synthesis across processes, places and scales. Cambridge University Press.
- Booker, D. J., & Snelder, T. H. (2012). Comparing methods for estimating flow duration curves at ungauged sites. Journal of Hydrology, 434, 78-94.
- Booker, D. J. (2013). Spatial and temporal patterns in the frequency of events exceeding three times the median flow (FRE3) across New Zealand. Journal of Hydrology (New Zealand), 52(1), 15.
- Booker, D. J., & Woods, R. A. (2014). Comparing and combining physically-based and empiricallybased approaches for estimating the hydrology of ungauged catchments. Journal of

Hydrology, 508, 227-239.

- Breiman, L., J. H. Friedman, R. A. Olshen, C. J. Stone, (1984). Classification and Regression Trees, Wadsworth, Belmont, Calif.
- Breiman L. (2001). Random Forests. Machine Learning, 45: 5-32. DOI: 10.1023/A: 1010933404324.
- Bulygina, N., McIntyre, N. and Wheater, H. (2009). Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis.
 Hydrology and Earth System Sciences, 13 (6), 893-904, doi: 10.5194/hess-13-893-2009.
- Bulygina, N., & Gupta, H. (2010). How Bayesian data assimilation can be used to estimate the mathematical structure of a model. Stochastic Environmental Research and Risk Assessment, 24(6), 925-937.
- Bulygina, N., McIntyre, N. and Wheater, H. (2011). Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. Water Resources Research, 47 (2), doi: 10.1029/2010wr009240.
- Bulygina, Nataliya, and Hoshin Gupta. (2011). "Correcting the mathematical structure of a hydrological model via Bayesian data assimilation." Water Resources Research 47.5 (2011).
- Cabus, P. (2008). River flow prediction through rainfall–runoff modelling with a probabilitydistributed model (PDM) in Flanders, Belgium. Agricultural water management, 95(7), 859-868.
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T. and Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44 (12), doi: 10.1029/2007wr006735.
- Clark, M. P., Kavetski, D. and Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research, 47, doi: 10.1029/2010wr009827.
- Doucet, A., Godsill, S., and Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering, Stat Comput, 10, 197-208, 10.1023/a: 1008935410038, 2000.
- Fenicia, F., McDonnell, J. J. and Savenije, H. H. G. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. Water Resources Research, 44 (6), doi: 10.1029/2007wr006386
- Fenicia, F., D. Kavetski, and H. H. G. Savenije (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47, W11510, doi: 10.1029/2010WR010174.

- Friendly M, Monette G, Fox J. Elliptical insights: understanding statistical methods through elliptical geometry. Stat Sci 2013; 28(1):1–39. http:// dx.doi.org/10.1214/12-STS402
- Gallego, G., Cuevas, C., Mohedano, R., & Garcia, N. (2013). On the Mahalanobis distance classification criterion for multidimensional normal distributions. Signal Processing, IEEE Transactions on, 61(17), 4387-4396.
- Gelman, J. B. C. et al. (2013) Bayesian Data Analysis, Third Edition. Andrew Gelman, John B.
 Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin ISBN 9781439840955
 CAT# K11900
- Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology, 2204-2214.
- Kavetski, D., S. W. Franks, and G. Kuczera (2002). Confronting input uncertainty in environmental modelling, in Calibration of Watershed Models, Water Sci. Appl., vol. 6, edited by Q. Duan et al., pp. 49–68, AGU, Washington, D. C.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Calibration of conceptual hydrological models revisited: 1. Eliminating numerical artifacts, J. Hydrol., in press.
- Kollat, J. B., Reed, P. M., & Wagener, T. (2012). When are multiobjective calibration trade-offs in hydrologic models meaningful?. Water Resources Research, 48(3).
- Lee, H., McIntyre, N., Wheater, H., & Young, A. (2005). Selection of conceptual models for regionalization of the rainfall-runoff relationship. Journal of Hydrology, 312(1), 125-147.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest.R news, 2(3), 18-22.
- MacKellar, N. C., Dadson, S. J., New, M., & Wolski, P. (2013). Evaluation of the JULES land surface model in simulating catchment hydrology in Southern Africa. Hydrology and Earth System Sciences Discussions, 10(8), 11093-11128.
- McMillan, H., Krueger, T. and Freer, J. (2012), Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrol. Process. 26: 4078–4111. doi: 10.1002/hyp.9384
- Moore, R. J. (2007). The PDM rainfall-runoff model. Hydrology and Earth System Sciences, 11 (1), 483-499, doi: 10.5194/hess-11-483-2007.
- Muller, P., A. Erkanli, and M. West (1996), Bayesian curve fitting using multivariate normal mixtures, Biometrika, 83(1), 67–79, doi: 10.1093/ biomet/83.1.67
- Olden, J. D. and Poff, N. L. (2003), Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Applic., 19: 101–121. doi: 10.1002/rra.700

Pechlivanidis, I. G., McIntyre, N. R., & Wheater, H. S. (2010). Calibration of the semi-distributed

PDM rainfall-runoff model in the Upper Lee catchment, UK. Journal of Hydrology, 386(1), 198-209.

- Peñas Silva, F. J. (2014a). Clasificación del régimen hidrológico natural y predicción de características hidroecológicas en el tercio norte de la Península Ibérica. Classification of the natural flow regime and prediction of hydroecological characteristics in the Northern third of the Iberian Peninsula.
- Peñas, F. J., Barquín, J., Snelder, T. H., Booker, D. J., and Álvarez, C. (2014b). The influence of methodological procedures on hydrological classification performance, Hydrol. Earth Syst. Sci., 18, 3393-3409, doi: 10.5194/hess-18-3393-2014
- Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis, 49(4), 974-997.
- Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks (2011). Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, Water Resour. Res., 47, W11516, doi:10.1029/2011WR010643.
- Ribeiro, M, I (2014). Gaussian Probability Density Functions: Properties and Error Characterization. Institute for Systems and Robotics.
- Singh, R. (2013). An uncertainty framework for hydrologic projections in gauged and ungauged basins under non-stationary climate conditions (Doctoral dissertation, The Pennsylvania State University).
- Smith, T., L. Marshall, and A. Sharma (2014), Predicting hydrologic response through a hierarchical catchment knowledgebase: A Bayes' empirical Bayes' approach, Water Resour. Res., 50, 1189–1204, doi:10.1002/2013WR015079.
- Snelder, T. H., Lamouroux, N., Leathwick, J. R., Pella, H., Sauquet, E., & Shankar, U. (2009). Predictive mapping of the natural flow regimes of France.Journal of Hydrology, 373(1), 57-67.
- Snelder T. H., Lamouroux N., Pella H. 2011. Empirical modelling of large scale patterns in river bed surface grain size. Geomorphology, 127: 189-197. DOI: 10.1016/j.geomorph.2010.12.015
- Snelder T. H., Barquin Ortiz J., Booker D. J., Lamouroux N., Pella H., Shankar U. 2012. Can bottomup procedures improve the performance of stream classifications? Aquatic Sciences, 74: 45-59. DOI: 10.1007/s00027-011-0194-7.
- Snelder, T. H., Datry, T., Lamouroux, N., Larned, S. T., Sauquet, E., Pella, H., & Catalogne, C. (2013). Regionalization of patterns of flow intermittence from gauging station

records. Hydrology and Earth System Sciences, 17(7), 2685-2699.

- Srikanthan, R., Amirthanathan, G. E., & Kuczera, G. (2007). Real-time flood forecasting using ensemble Kalman filter. In MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand (pp. 1789-1795).
- Thompson, J. R., Green, A. J., Kingston, D. G., & Gosling, S. N. (2013). Assessment of uncertainty in river flow projections for the Mekong River using multiple GCMs and hydrological models. Journal of hydrology, 486, 1-30.
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V. and Sorooshian, S. (2001). A framework for development and application of hydrological models. Hydrology and Earth System Sciences, 5 (1), 13-26, doi: 10.5194/hess-5-13-2001.
- Wagener, T. and Montanari, A. (2011). Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. Water Resources Research, 47 (6), doi: 10.1029/2010wr009469.
- Willems, P., Mora, D., Vansteenkiste, T., Taye, M. T., & Van Steenbergen, N. (2014). Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes–Part 2: Intercomparison of models and calibration approaches. Journal of hydrology, 510, 591-609.
- Westerberg, I. K., and Birkel, C. (2015), Observational uncertainties in hypothesis testing: investigating the hydrological functioning of a tropical catchment. Hydrol. Process. 29: 4863–4879. Doi: 10.1002/hyp.1053
- Westerberg, I. K. and McMillan, H. K. (2015) Uncertainty in hydrological signatures, Hydrol. Earth Syst. Sci., 19, 3951-3968, doi: 10.5194/hess-19-3951-2015, 2015.
- Yadav, M., Wagener, T. and Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources, 30 (8), 1756-1774, doi: 10.1016/j.advwatres.2007.01.005.
- Zhang, Z., Wagener, T., Reed, P., & Bhushan, R. (2008). Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. Water Resources Research, 44(12).
- Team, R. Core. "R: A language and environment for statistical computing. R Found Stat Comput Vienna, Austria." (2012).
- http://servicios2.marm.es/sia/visualizacion/descargas/dma.jsp
- http://info.igme.es/cartografiadigital/geologica/Magna50.aspx

CHAPTER 3: Identification of Dominant Hydrological Mechanisms for Ungauged Basins via Bayesian Approach

Part of the work developed for this chapter has been presented at EGU General Assembly 2016 [*Prieto et al., 2016a; 2016b*] *and AGU 2016* [*Prieto et al., 2016b*]:

Prieto, C., Le-Vine, N., Vitolo, C., García, E., and Medina, R. (2016a). Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Abstract for the European Geosciences Union General Assembly 2016, available online: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-8580-2.pdf

Prieto, C., Le-Vine, N., Vitolo, C., García, E., and Medina, R. (2016b). Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Presentation for the European Geosciences Union General Assembly 2016, available online: http://presentations.copernicus.org/EGU2016-8580_presentation.pdf

Prieto, C., Le-Vine, N., Vitolo, C., and Medina, R. (2016c). Dominant hydrolocial process identification for ungauged basins: Bayesian approach. Presentation for the European Geosciences Union General Assembly 2016, available online: https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/160806

ABSTRACT

This work presents a Bayesian statistical methodology to identify dominant hydrological mechanisms for ungauged catchments using an ensemble of hydrological models. First, information contained in a plethora of hydrological indices is summarized using the Principal Components Analysis. Second, the derived Principal Components are regionalized using the functional-form-free Random Forests technique. And third, dominant hydrological mechanisms are identified in ungauged basins using the regionalized pieces of information via Bayesian approach. The method is demonstrated for basins in northern Spain. The results show that, for some processes, dominant hydrological mechanisms can be identified with only a few pieces of information; and, for the other processes, dominant mechanisms cannot be identified even when the majority of available information is used. The methodology reliably identifies dominant mechanisms when there is no model error, and a large quantity of high quality information is added. Results show that a successful identification of dominant processes can be linked to 1) presence of model error, 2) information quality, 3) available information content and 4) hydrological process sensitivity to the available hydrological indices.

3.1. INTRODUCTION

Identification of dominant hydrological mechanisms shaping catchment behavior is a major research challenge for both gauged and ungauged basins [Clark et al., 2008; Clark et al., 2011a; Coxon et al., 2014], with the largest application in catchment hydrological model identification, or selection. A hydrological model combines different hypotheses for each process governing catchments working [Beven, 2001]; and the hypotheses are to be assigned different degrees of belief [Clark et al., 2011a], and can never be validated, due to the lack of full access to the natural

phenomenon [Wagener et al., 2004]. Because of data scarcity, it is a common practice to use a single pre-selected model structure to represent the processes for both gauged and ungauged catchments [Blöschl, 2013; Wagener and Montanari, 2011; Ley et al., 2016], regardless of the widely acknowledged 'uniqueness of the place' paradigm [Beven, 2000], and inability of a single model structure to be a suitable description of a hydrological system [Clark et al. 2011a; 2011b; 2015a; 2015b; 2015c]. Such ignorance of model structural error may lead to predictive uncertainty underestimation, and prediction prone to bias [Reicherte and Omlin, 1997]. To improve on this, flexible and adaptive approaches have been recently developed (e.g. RRMT [Wagener et al. 2001; 2004], FUSE [Clark et al., 2008; 2011b]; FLEX and SUPERFLEX [Fenizia et al., 2006; 2008a; 2011]; SUMA [Clark et al., 2015a; 2015b; 2015c]) that provide multiple mechanisms for each hydrological process, decomposing a hydrological model into a set of testable hypotheses [Clark et al., 2008; 2011a; 2011b; 2015a; 2015b; 2015c; Fenicia et al. 2008]. Hydrological summaries (or indices) have been used to assist with the hypothesis testing for gauged catchments [Clark et al., 2011b; Coxon et al., 2014], although identifying a test statistic that accounts for uncertainties in model structure/input and hydrological indices still remains a challenge [Clark et al., 2011a; Boyle et al., 2000; Wagener et al., 2003; Wagener et al., 2007; Kirchner et al., 2006]. The approach can be extended to ungauged catchments by including the additional uncertainty to regionalize hydrological information into the analysis.

Among the plethora of possible hydrological summaries, a selection of hydrological indices must characterize the most important components of the flow regime. To address this, Olden and Poff [2003] chose 171 indices that represent five different aspects of the flow regime, and demonstrated the utility of principal components analysis (PCA) to compactly describe variability in the data. More recently, Yadav et al. [2007] estimated 39 indices, which were divided into seven classes by means of linear and Spearman rank correlation coefficients; Coxon et al. [2014], used 3 signatures to evaluate how well the model can capture the predominant flow behavior over decadal, annual and monthly timescales; Westerberg et al. [2016] used 15 hydrological indices (9 indices to describe the flow distribution and 6 to describe flow dynamics), order to quantify the uncertainty coming from the discharge data in 43 UK catchments and propagated these uncertainties into the regionalization of the indices. Then, to regionalize hydrological indices, linear regression relationships are usually fit between the hydrological indices and relevant catchment descriptors [Yadav et al., 2007; Zhang et al., 2008]. As such relationships are proven to be non-linear [Snelder et al., 2009], Random Forests have recently been proposed for the regionalization [Snelder et al., 2013; Peñas et al., 2014a], which extend the concept of a regression tree [Breiman et al., 1984] to a set of trees [Breiman, 2001]. While retaining the advantage of being functional dependence form hypothesis-free, Random Forests offer an improvement in accuracy over a regression tree, and are more robust against random variations in inputs [Snelder et al., 2013]. Further, Random Forests have been demonstrated to be efficient when compared to other machine learning algorithms [Peñas et al., 2014a], or other more physically based approaches [Booker and Woods, 2014].

The contribution of this research is a Bayesian methodology to constrain an ensemble of hydrological models using regionalized information, and to help identifying dominant hydrological processes for ungauged basins. Bayesian methods are particularly well posed for hypothesis testing [Raftery 1993], as well as for describing system nonlinearities and uncertainties [e.g., Kavetski et al., 2002; Vrugt et al., 2008; Renard et al., 2010; Bulygina et al., 2009; Almeida et al., 2014]. First, the information contained in the variety of available hydrological indices is summarized, using the Principal Components Analysis. Second, the information is regionalized using the Random Forests technique, which also provides quantification of the prediction error. Third, rainfall-runoff models in the ensemble FUSE [Clark et al., 2008] are assigned likelihoods using Bayesian conditioning on the regionalized information. And fourth, a statistic is devised that allows hypothesis testing about dominant hydrological processes for ungauged basins. The method is demonstrated for 16 basins in northern Spain with diverse hydrological regimes. The results show that success dominant hydrological process identification can be linked to 1) the size of model error, 2) regionalization quality, 3) information content in regionalized indices, and 4) information required to identify hydrological processes.

3.2 CATCHMENTS AND DATA DESCRIPTION

Hydrological records and properties from a set of 92 small (2 km²) to medium-size (1,038 km²) basins located in the north of Spain are used to regionalize hydrological indices (Figure 11). The basins comprise a subset from the 156 basins used by Peñas et al. [2014a], and are selected to ensure a natural hydrological regime according to the Water Framework Directive (articles 5 and 6, sections characterization of the water bodies and the analysis of pressures and impacts [http://servicios2.marm.es/sia/visualization/descargas/dma.jsp]). Sixty-two basins in the selected set drain into the Cantabric Sea, and the other 30 basins drain into the Mediterranean Sea.

The watersheds cover a wide range of geology, soils, topography, land uses and climatic conditions. The major lithological groups based on the Geological and Mining Institute of Spain (http://www.igme.es/internet/default.asp) are clay, sand and gravel in basins that drain into the

Mediterranean Sea, although there are also siliceous and calcareous rocks in those located in the Pyrenees. In addition, slates predominate in the western basins that drain into the Cantabrian Sea, while calcareous rocks predominate in the eastern basins. The average altitude ranges between 200 and 2200 m, and the slope of the main channel varies between 9 and 63%. Urbanized area covers less than 8% in each basin, and the dominant land use is broadleaf forests, coniferous forests and pastures.

Annual average rainfall varies between 450 and 1800 mm, average temperature varies between 5 and 15°C, annual average PET varies between 600 and 960 mm, and the aridity index ranges between 0.3 and 1.7 [Arora, 2002]. Figure 11 categorises the basins based on 1) the aridity index, and 2) minimum monthly average temperature; so that climatic conditions are considered to be dry if the aridity index is above 1, and humid otherwise [Arora, 2002]; and snow is likely to occur when mean monthly temperature is below 0°C. Annual average flow ranges between 20 and 1520 mm per year, and the runoff coefficient varies between 0.03 and 0.97.

A subset of 16 basins (out of the 92 basins) is further selected and treated as ungauged in hydrological modeling to assess the proposed methodology; each basin in the subset has a minimum of eight years of synchronized daily precipitation, daily flow and monthly potential evapotranspiration [Peñas et al., 2014a] (Figure 11, and Table 5). Daily precipitation data for the basins are provided by the Spanish Meteorological Agency (AEMET).

Chapter 3: Identification of Dominant Hydrological Mechanisms for Ungauged Basins via Bayesian Approach



Figure 11: The selected set of 16 basin

Table 5: Catchment characteristics: (1) calcareous rocks, (2) clay, (3) conglomerates rocks, (4) sand, (5) sedimentary rocks, (6) shale (sedimentary) rocks, (7) siliceous rocks, (8) slates, (9) volcanic rocks. See Peñas et al. [2014b] for permeability and geology.

Name	٩	RRC	Average	Average	Annual	Annual PET			0	seolo	gy (%	Area	occu	ipied)	_	
	(km2)	obs	Altit	Slope	Rainfall mm/year	Mm/Year	Permeability	(1)	(2)	(3)	(4)	(5)	(9)	(7)	(8)	(6)
x1353	529	0.61	1047	0.52	1459	660	low	2	0	10	11	4	0	-	72	0
x1404	293	0.81	1068	0.50	1449	638	very low	0	0	പ	16	11	0	0	68	0
x1303	377	0.64	500	0.39	1342	734	low	19	10	0	7	ŝ	0	35	31	0
x1265	294	0.57	1188	0.53	1041	716	very low	18	0	19	4	∞	0	0	50	0
9257	80	0.33	696	0.24	814	962	wo	Ч	0	7	49	0	48	0	0	0
x9040b	623	0.69	1505	0.52	1356	705	No	17	0	ഹ	47	1	22	9	0	2
9269	75	0.39	1385	0.48	1333	619	low	2	0	0	73	0	20	0	0	0
9197	283	0.24	1147	0.33	681	655	No	17	ъ	ŝ	0	1	73	0	0	0
9221	22	0.79	839	0.24	1134	564	low	65	0	0	0	0	35	0	0	0
AN439	152	0.95	702	0.27	1469	701	low	06	2	0	1	1	ഹ	0	0	-
AN433	554	0.41	789	0.22	1376	619	high	53	0	0	9	11	30	0	0	0
AN520	73	0.35	1071	0.24	1683	609	low	11	0	0	55	0	33	0	0	0
AN530	95	0.39	805	0.36	1371	693	low	0	0	0	80	0	20	0	0	0
AN313	477	0.35	789	0.21	1038	650	high	57	0	1	0	10	21	10	0	0
c8z1	114	0.66	598	0.42	1794	684	high	28	ŝ	0	28	0	37	0	Ч	7
c7z1	28	0.65	483	0.39	1809	731	wo	Ŋ	Ч	0	21	0	18	0	51	4

158

3.3 HYDROLOGIC MODELS

The Framework for Understanding Structural Errors (FUSE) has been developed by Clark et al. [2008] and includes well-established conceptual rainfall-runoff models (i.e. PRMS, SACRAMENTO, TOPMODEL and ARNO/VIC) (Figure 12, Table 6). Each model is characterized by a different architecture for the upper and lower soil layers to represent a variety of dominant processes for evaporation, vertical percolation, interflow, base flow and surface runoff (Figure 12, Table 6). Rainfall and potential evapotranspiration observations are used as inputs to simulate a streamflow discharge time series. The original Fortran code has been re-implemented by Vitolo et al. [2012] in the R programming language and made available on a public repository (GitHub). While FUSE produces 1248 different model structures for a catchment, rainfall error is not included into the inference, as input data uncertainty is not addressed in this study, resulting in 624 different model structures per catchment. For each catchment and each model structure, 1000 parameter sets were sampled from the prior uniform parameter distribution [Clark et al., 2011b] using the Latin Hypercube method; and 9,984,000 (16 basins x 624 models x 1000 parameter samples) simulations, representing at least 8 years each, were run on a High Performance Computing facility. The first year of the simulations was used as a warm-up period and discarded from the analysis.



Figure 12: FUSE

PROCESSES	MECHANISMS
P1:	A single state variable
Process in	Separate state variables for tension storage (below field capacity) and free
the Upper	storage (above field capacity).
Soil	Separate state variables for tension storage (below field capacity) and further
	discretization of upper zone tension storage into two zones.
P2:	A reservoir of fixed size.
Process in	One tension reservoir and two parallel tanks.
the Lower	A reservoir of unlimited size (fractional rate).
Soil	A reservoir of unlimited size (power recession).
P3:	Saturated area is related to storage in the unsaturated zone via a Pareto
Runoff	distribution
Generation	Saturated area is a linear function of tension storage
	The saturated area is related to storage in the saturated zone via the
	topographic index.
P4:	Characterized by water availability from field capacity to saturation
Percolation	Depends on the lower zone percolation demand.
	Characterized by water availability from wilting point to saturation
P5:	Depends on the relative root fractions in each of the soil layers
Evaporation	Sequential evaporation model: The potential evaporative demand is first
	satisfied by evaporation from the upper soil layer, and any residual
	evaporative demand is satisfied by evaporation from the lower soil layer.
P6: Interflow	No interflow
	Linear function of free storage in the upper layer
P7: Routing	No routing
	Routing

Table 6: Processes and mechanisms in the 624 FUSE models

3.4 METHOD

3.4.1 Hydrological indices and catchment descriptors

Following Olden and Poff [2003], flow records from the 92 basins are used to calculate 103 hydrological indices that reflect different aspects of flow regime. The indices represent the mean and standard deviation of: 1) annual and monthly flows, 2) high and low flows, 3) duration and frequency of high flows, 4) rate of change in flows; and 5) time of maximum and minimum flow

events. The principal component analysis is used to transform the hydrological indices into uncorrelated (orthogonal) principal components (PCs); and the broken stick method is utilized to select those PCs that explain the most of variability in the hydrological indices [Jackson, 1993; Peres-Neto et al., 2005].

The analysis considers 16 out of 50 available catchment descriptors (CDs) that are selected by Peñas et al. [2013]; the CDs represent the least correlated subset with the Pearson's correlation coefficient below 0.7. The selected features represent area, climate (mean annual precipitation and PET, ratio of minimum quarterly precipitation to maximum quarterly precipitation), topography (average catchment elevation and gradient), basin geometry (drainage density, number of river confluences), land use (area covered by agricultural land, broadleaf forest, coniferous forest, bare land, pasture and urban areas), and geology (average rock density and permeability). The climatic variables are derived from monthly climate series calculated from a 1 km × 1 km grid map, developed by the Centre for Hydrographic Studies (CEDEX, Ministry of Public works and Ministry of Agriculture and Environment, Spain). Topography and basin geometry are derived using a 25 m digital elevation model. Land use is derived from the Soil Occupancy Information System (in Spanish SIOSE) at a 1:25,000 scale developed by the National Geographic Institute of the Spanish Government. The geological variables are derived from the lithostratigraphic and permeability maps at scale 1:200,000, developed by the Spanish Geologic and Mining Institute (in Spanish IGM) of the Spanish Government.

3.4.2 Regionalization by Random Forests

Regionalization by Random Forests has some desirable theoretical and practical advantages [Breiman, 2001]:

- model structure assumption-free, and do not require preselecting linear/nonlinear model form and the number of predictors;
- no assumptions on the statistical nature of the model errors, i.e. normality, independency, and homoscedasticity (as required in multiple linear regressions);
- time/ quality efficient.

Random Forests are trained using the selected PCs (predictands) and the corresponding 16 catchment descriptors (predictors) for the 92 gauged basins using 'randomForest v4.6.7' R package [Liaw and Wiener, 2002]. The Random Forests algorithm resamples with replacement a number of trees to compose a forest. A tree is grown such that at each node a number of catchment descriptors (CDs) is randomly sampled and those that give the best split (the lowest mean square error) are chosen [Liaw et al., 2002; Snelder et al., 2012]. The resampling and the

use of a random subset of predictors introduce randomness to the way a tree is built in Random Forests [Breiman, 2001] as compared to a single regression tree [Breiman, 1984]. A regionalized value for an ungauged basin is estimated as the mean of all individual predictions from each tree in the forests [Snelder et al., 2011].

In the study, the 16 basins with synchronized hydrometeorological daily data are treated as 'ungauged' sites. For each of the 16 basins, the other 91 basins are used to regionalize the PCs (the "leave-one-out" technique). The regionalized PCs are affected by uncertainty in the regionalization model structure, e.g. uncertainty in the CDs selection, errors in the exact values of the CDs, and the available gauged basins being only a sample from the entire population [Almeida, 2014]). The regionalization uncertainty is evaluated by means of a jack-knife (or "leave-one-out") strategy. Each time, one catchment is left out and the remaining 90 are used to estimate the regionalization models for each PC. Then, the residuals between the regionalized PCs and the PCs based on the observed data are calculated. The process is repeated for all catchments resulting in a set of 91 residuals, to which a joint probability density function (JPDF) is fitted. The JPDF characterizes the regionalization error for Random Forests, and defines regionalization model likelihoods.

3.4.3 Posterior distribution of mechanisms

Each hydrological process p=1, P is represented using one of the mechanisms mec_m^p , m = 1, N^p (Table 6); and a model structure M_k, $\{M_k\}$, k=1, N, is defined by a combination of such mechanisms to represent all P considered hydrological processes. For each process p, the posterior probability of each mechanism mec_m^p , m = 1, N^p in a model ensemble $\{M_k\}$ is a sum of the posterior model probabilities over those model configurations that represent the process p via the considered mechanism mec_m^p (there are N_m^p models with mec_m^p in a model ensemble):

$$p(mec_m^p | PC^{Reg}, I, \{M_k\}) \propto \sum_{Mk:mec_m^p} p(M_k | PC^{Reg}, I)$$
(3.1)

And the posterior probability of a model M_k , given the regionalized principal components PC^{reg} and inputs *I*, can be calculated as:

$$p(M_k | PC^{Reg}, I) \propto L(M_k | PC^{Reg}, I) * p(M_k | I)$$
(3.2)

Where $L(M_k | PC^{Reg}, I)$ is a likelihood of the hydrological model M_k , given the regionalized principal components PC^{reg} and inputs *I*; $p(M_k | I)$ is a prior probability of the model M_k . A key assumption here is that at least one model structure M_k and its inputs *I* are capable of
representing the true underlying PCs. The models are considered equally likely a priori, and are assigned equal prior probabilities 1/N. Then, equation (3.2) can be re-written as (3.3) (the prior model probability of 1/N is included into the proportionality constant):

$$p(M_k | PC^{Reg}, I) \propto \int_{\Theta_k} L(PC^{sim}_{\Theta_k} | PC^{Reg}, I) * p(\Theta_k | I) * d\Theta_k(3.3)$$

Where $L(PC_{\theta_k}^{sim}|PC^{Reg})$ is a likelihood of principal components $PC_{\theta_k}^{sim}$ simulated by the model M_k with inputs *I* and parameters θ_k , and $p(\theta_k)$ is a prior parameter distribution for model M_k . The likelihood function $L(PC_{\theta_k}^{sim}|PC^{Reg})$ is taken proportional to the PC residual distribution from the regionalisation procedure described in section 3.4.2, and the prior $p(\theta_k)$ is taken as a uniform distribution over a hypercube θ_k (defined based on the parameter ranges given in Clark et al. [2011b]).

To account for a non-equal mechanism sampling frequency in a selected model ensemble (as an ensemble may include more mechanisms of one type than of the other type), the mechanism posterior is approximated as the posterior based on the ensemble (given by equation (3.1)) weighed by the inverse of the mechanism sampling frequency in the ensemble $\frac{N_m^p}{N}$ [see Importance Sampling in Doucet et al., 2000]:

$$p(mec_m^p | PC^{Reg}, I) \propto \sum_{M_k:mec_m^p} p(M_k | PC^{Reg}, I) * \frac{1}{N_m^p} \quad (3.4)$$

To account for a statistical spread (due to the ensemble model selection) around the mechanism posterior approximation given by (3.4), model re-sampling with replacement (boot-strapping) is implemented in the next section. For each process p, N_m^p models with a mechanism mec_m^p are resampled from the ensemble of models that represent the process p via the mechanism mec_m^p (there are also N_m^p models with mechanism mec_m^p in the initial model ensemble); and the mechanism mec_m^p posterior is calculated using (3.4). The re-sampling is done a large number of times (the study re-samples 10,000 times for each mechanism), resulting in a statistical distribution for each mechanism posterior probability that is less dependent on the choice of a model ensemble (due to the finiteness of the initial model ensemble, it is not possible to make the probability density function fully independent of the ensemble selection).

3.4.4 Identification of dominant mechanisms

A mechanism mec_m^p representing a process p is defined as dominant for a set of competing mechanisms mec_i^p , i = 1, N^p when the posterior probability for the mechanism mec_m^p is greater

than a sum of the posterior probabilities for the other competing mechanisms to represent the process p, i.e.: $p(mec_m^p | PC^{Reg}, I) > \sum_{i \neq m} p(mec_i^p | PC^{Reg}, I)$. As the exact values of the posterior mechanism probabilities are not known, and can only be approximated via the probability density functions (see the previous section), identification of a dominant mechanism for a process p becomes a multiple hypotheses testing problem (one test for each mechanism for a process p – in total N^p tests for the process).

The Bonferroni multiple-comparison correction [Yosef, 1988] is used to identify the dominant mechanisms for each basin, and for each process. The correction controls the family-wise error rate (FWER) for multiple hypothesis testing, so that an experiment-wide (each experiment consists of N^p tests) critical value does not exceed some α (equals 0.05 in the study). This is achieved by taking the critical value as α/N^p for each individual hypothesis testing in the family, where a family consists of N^p hyposeses for a process p (a family for each process).

For each basin, each process p, and each mechanism mec_m^p , the null hypothesis H₀ is that a mechanism mec_m^p is not dominant, i.e.: $p(mec_m^p | PC^{Reg}, I) \leq \sum_{i \neq m} p(mec_i^p | PC^{Reg}, I)$. As the posterior mechanism probabilities are not known exactly, and are described via re-sampling (see the previous section), the hypothesis testing is based on comparing multiple samples from the probability density functions for the mechanism posterior probabilities, and calculating a fraction of times when the samples from the mec_m^p posterior probability distribution exceed a sum of samples from the other competing mechanism posterior probabilities. So that if the fraction is less or equal to α/N^p , H_0 is accepted (mec_m^p is not dominant), and rejected, otherwise (and mec_m^p is considered as dominant). The results are summarized as the percentage of identified dominant mechanisms, and as the percentage of dominant mechanisms detected correctly from all detected.

3.4.5 Sensitivity analysis: cases

To examine the influence of regionalization quality, model error, and information quantity, the following five cases are analyzed for each of the 16 catchments:

<u>Case BC</u>: The FUSE model ensemble is used to simulate flows driven by the observed daily rainfall and estimated monthly PET (section 3.2). The number of PCs is selected based on the broken stick method to explain the most of variability in the observation-based PCs; and the PCs are regionalized using Random Forests (section 3.4.2).

Case R: The key difference from the base case BC is an improved accuracy and precision of the regionalized PC values. The regionalized PC values are generated as draws from Gaussian

164

distributions centered on the corresponding PC values calculated from the observed flows (assumed not available in the case BC), with a standard deviation fixed at 5% of the corresponding PC value range (the range is derived using PCs for the 92 basins).

<u>Case M</u>: The key difference from the base case BC is that model and input are error-free. This is achieved by substituting the observed flow time series with the synthetic flows generated by a model (and its corresponding parameter set) in the ensemble such that its corresponding PC values are the closest to the corresponding observation-based PC values (a normalized distance is used). The PC values calculated from such synthetic flow are added to the regionalization noise calculated based on PC^{abs} and PC^{Reg} to produce synthetic regionalized PCs for the case, so that the regionalization quality is similar to the base case (BC). Note, the formula relating PCs to hydrological indices are kept the same as for the observed flows, as there it is not meaningful to derive the PC decomposition for the 103-dimensional synthetic hydrological index space based on 16 basins only (the observation-based PC decomposition is based on 92 basins).

<u>Case MR</u>: The case combines the two modifications to the base case implemented in the cases R and M to improve both the regionalization and model quality.

<u>Case MRI</u>: The case is a modification of the case MR and uses at least 90% of information available in the PCs (based on the number of PCs), so that the information quantity is no less than in the case MR.

The percentage of identified dominant mechanisms, and the percentage of dominant mechanisms detected correctly from all detected (not available for the cases BC and R) are reported for all catchments and for all cases (when available).

3.5 RESULTS AND DISCUSSION

3.5.1 Regionalization and likelihood definitions

103 hydrological indices are calculated for each of the 92 basins, and the broken stick method is used to select four principal components (in a 103-dimensional space) that explain 63.3%, 12.4%, 6.9% and 4.8% of variability, respectively, or 87.6% of variability in total.

The PCs are regionalized using the Random Forests method, and Table 7 shows that an average regionalization quality, defined for each PC as a ratio of its distribution scale parameter to its range, ranges between 10% and 16%. Regionalization of PC1 is the most precise, but the least accurate; while regionalization of PC3 is the least precise, and regionalization of PC4 is the most accurate. The obtained regionalization quality is comparable with, or better than that given by

a step-wise regression regionalization in Almeida [2014], where regionalization precision varied between 7% and 27%.

	Location	Scale	Normalized scale
PC1	2.42	3.69	0.1
PC2	0.18	2.28	0.13
PC3	0.05	1.7	0.16
PC4	0	1.9	0.12

Table 7: Regionalization quality for individual PCs for a selected catchment. The normalized scale is derived by dividing the scale by the corresponding PC range.

The likelihood functions in (3.3) are estimated based on residuals distributions for the regionalized PCs (i.e. the differences between the regionalized and observation-based PCs). Pearson's linear correlation test shows that there is a significant correlation between the residuals of PC2 and PC3 only (the *p*-value is 0.03). Therefore, the 4-dimensional likelihood is approximated as a product of marginal residual distributions for PC1 and PC4 and a joint distribution for PC2 and PC3. The marginal distribution for PC1 residuals is chosen as an extreme value type I distribution (EV1) based on the chi-square test; the marginal distribution for PC4 residuals is chosen as Gaussian distribution based on the chi-squared test; and the joint distribution for PC2 and PC3 residuals is chosen as a two-dimensional Gaussian distribution based on the chi-squared test; and the joint distribution for PC2 and PC3 residuals is chosen as a two-dimensional Gaussian distribution based on the chi-squared test and the joint distribution for PC2 and PC3 and Mardia tests (all tests have 5% significance level). And approximations of marginal and joint distributions are illustrated for a selected basin in Figure 13 and 14 (estimated using the leave-one-out technique).





Figure 13: Marginal distributions and empirical histograms (EMP) for regionalised residuals for selected catchment



Figure 14: Regionalized residuals for PC2 and PC3 for a selected catchment: a) a fitted joint Gaussian distribution, and b) a scatterplot.

Figure 15 shows information content in the first four PCs calculated for the observed flows (the cases BC and R) and synthetically generated flows (the cases M and MR). The first four PCs for

the observations have larger information content, and the number of PCs required to represent at least 90% of information (used in the case MRI) varies between 4 and 50 PCs.



Figure 15: Information content in the first four PCs for the observed and synthetic flows

3.5.2 Prediction quality

Figure 16 shows the number of dominant mechanisms identified (A), and correctly identified (B) for each of the seven processes described in Table 6 for the set of 16 basins (treated as ungauged). The mechanisms for routing, evaporation, interflow and runoff are the most identifiable processes for the observation-based case (BC), and are identified for 16, 13, 11, and 9 catchments (out of 16), respectively (Figure 17 subplots BC a) to BC d)). Note, that it is not possible to estimate the number of correctly identified dominant mechanisms for the cases BC and R, as the answer is unknown. Further, dominant mechanisms for the process in the upper soil and percolation are not identifiable for any basin in the case BC. Improving regionalization (case R) leads to mixed results, when the number of identified dominant mechanisms may increase, decrease, or stay the same as in the case BC. This can be attributed to the presence of (the same) model/input error in the two cases, so that improving regionalized information does not necessarily lead to better predictions.

Figure 17 shows the most frequently identified hydrological mechanisms based on the four PCs regionalized via Random Forests from gauged basins (the case BC) – dominant mechanisms for routing, evaporation, interflow and runoff processes. All basins are better described with a runoff routing mechanism (as opposite to runoff delivered directly to the outlet); and sequential

evaporation, no interflow, and topographic-index-based runoff mechanisms are preferred for the majority of the basins. A reservoir of unlimited size with a fractional rate is preferred in all 4 cases where a mechanism for the lower soil layer process is identified; the mechanisms for the upper soil layer process and percolation are not identified for any basin. Due to the small sample size (16 basins), and the lack of variability in identified mechanisms for any given process, it is not possible to draw meaningful conclusions on any correspondence between climatic/ geographic/ geological catchment descriptors and the identified dominant hydrological mechanisms.

When model/input error is excluded (cases M, MR, and MRI), improving the quality of regionalized information tends to identify the dominant mechanisms with higher reliability (correctly identified mechanisms) (Figure 16). Moreover, adding high quality information always increases the number of reliable mechanism identifications (compare M and MR vs. MRI in Figure 16). Further, the mechanism for the upper soil process appears to be the least identifiable (identified only for 25% of considered basins) for the best-case scenario (the case MRI); and mechanisms do not get correctly identified in all 16 basins (up to 10 basins for each process). This might be caused by inability of the 103 static signatures (the PCs are derived from) to capture and represent the details of the considered hydrological dynamics.

The cases M, MR, and MRI allow evaluating accuracy of the mechanism identification (as ground truth is known) - Figure 18 shows that when there is no model/ input error and high quality/ high information content regionalized data is employed (the MRI case) dominant hydrological mechanism identification is reliable, i.e. if the procedure identifies a dominant mechanism, the identification is accurate. There is a larger variability in the dominant mechanisms identified for each process in the case MRI (Figure 18), than in the case BC (Figure 17), but it is still not meaningful to relate the identified mechanisms to the corresponding catchment properties, as the case MRI is based on synthetically generated flows.

169



Figure 16: Identified dominant mechanisms (A) and correctly identified dominant mechanisms (B) for each case.



Figure 17: Case BC - identified mechanisms for the processes: a) routing, b) evaporation, c) interflow, d) runoff and e) process in the lower soil.

Chapter 3: Identification of Dominant Hydrological Mechanisms for Ungauged Basins via Bayesian Approach



Figure 18: Case MRI - identified and correctly identified mechanisms for the processes: a) routing, b) evaporation, c) interflow, d) runoff, e) process in the lower soil, f) percolation and g) process in the upper soil.

3.6 CONCLUSIONS

This paper proposes a statistical methodology to identify the dominant mechanisms in ungauged catchments and poses the path to reduce the identification uncertainty. The study merges the recent developments in hydrological indices selection and regression method with the strengths of the Bayesian inference. The study uses an ensemble of hydrological model structures FUSE [Clark et al., 2008; 2011b] to predict mechanisms in the ungauged basins, where each hydrological process is given by a combination of mechanisms and mechanisms are treated as a sample from a complete (but unobservable) set of mechanisms; and this information is combined with information provided by PCs regionalized by Random Forests, via a Bayesian approach. Further, the work provides a multiple hypothesis comparison test that allows identifying a dominant mechanism for each hydrological process in a catchment. An application to a set of 16 catchments in northern Spain leads to the following findings:

1. When model error exists, improving regionalization doesn't necessarily imply a better mechanism identification. Only when model and input errors are decreased/ eliminated

and a large quantity of information is available, improving regionalization quality enhances the mechanism identification.

- 2. When there is no model/ input error and a large quantity of high quality regionalized information is available, the proposed procedure for dominant hydrological mechanism identification is reliable, i.e. if a mechanism is identified as dominant, then the identification is accurate.
- 3. The information content in the first four PCs allows identifying dominant mechanisms for routing, evaporation, interflow and runoff.
- 4. Commonly considered large set of hydrological indices [Olden and Poff, 2003] might be insufficiently for dominant mechanism identification for a number of hydrological processes (i.e. for the process in the upper soil, and for the percolation process).
- 5. Dominant hydrological mechanism identification is found sensitive to 1) model and input errors; 2) regionalization quality; and 3) information quantity.

3.7 REFERENCES IN CHAPTER 3.

- Almeida, S. M. C. L. (2014). The Value of Regionalised Information for Hydrological Modelling. PhD thesis, Imperial College London, London, UK.
- Arora, V. K. (2002). The use of the aridity index to assess climate change effect on annual runoff. Journal of Hydrology, 265(1), 164-177.
- Beven, K. J. (2000). Uniqueness of place and process representations in hydrological modelling. Hydrology and Earth System Sciences Discussions, 4(2), 203-213.
- Beven, K. J. (2001). How far can we go in distributed hydrological modelling?, Hydrol. Earth Syst. Sci., 5, 1-12, doi:10.5194/hess-5-1-2001, 2001.
- Blöschl, G. (Ed.). (2013). Runoff prediction in ungauged basins: synthesis across processes, places and scales. Cambridge University Press.
- Booker, D. J., & Woods, R. A. (2014). Comparing and combining physically-based and empiricallybased approaches for estimating the hydrology of ungauged catchments. Journal of Hydrology, 508, 227-239.
- Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Research, 36(12), 3663-3674.
- Breiman, L., J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and Regression Trees, Wadsworth, Belmont, Calif., 1984.

Breiman L. 2001. Random Forests. Machine Learning, 45: 5-32. DOI: 10.1023/A:1010933404324.Bulygina, N., McIntyre, N., and Wheater, H. (2009). Conditioning rainfall-runoff model

parameters for ungauged catchments and land management impacts analysis. Hydrology and Earth System Sciences, 13 (6), 893-904. doi: 10.5194/hess-13-893-2009

- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44 (12). doi: 10.1029/2007wr006735
- Clark, M. P., Kavetski, D., and Fenicia, F. (2011a). Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research, 47. doi: 10.1029/2010wr009827
- Clark, M. P., McMillan, H. K., Collins, D. B., Kavetski, D., and Woods, R. A. (2011b). Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses. Hydrological Processes, 25(4), 523-543.
- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M. (2015a): A unified approach for process-based hydrologic modeling: 1. Modeling concept. Water Resources Research, 51, 2498-2514. doi: 10.1002/2015wr017198
- Clark, M. P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., and Marks, D. G. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. Water Resources Research, 51, 2515-2542. doi: 10.1002/2015wr017200
- Clark, M. P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W.,, Brekke, L. D., Arnold, J. R., Gochis, D. J., Rasmussen, R.M., Tarboton, D.G., Mahat, V., Flerchinger, G.N., and Marks, D.G. (2015c). The Structure for Unifying Multiple Modeling Alternatives (SUMMA), version 1: Technical description. NCAR Technical Note NCAR/TN-514+ STR, 54 pp., National Center for Atmospheric Research, Boulder, Colo. doi: 10.5065/D6WQ01TD
- Coxon G., Freer J., Wagener T., Odoni N. A. and Clark M. (2014), Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, Hydrol. Process., 28, pages 6135–6150, doi: 10.1002/hyp.10096
- Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and computing, 10(3), 197-208.
- Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L. (2006). Is the groundwater reservoir linear? Learning from data in hydrological modelling.Hydrology and Earth System Sciences Discussions, 10(1), 139-150.

- Fenicia, F., McDonnell, J. J., and Savenije, H. H. G. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. Water Resources Research, 44 (6). doi: 10.1029/2007wr006386
- Fenicia, F., Kavetski, D., and Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47, W11510. doi: 10.1029/2010WR010174
- Jackson, D. A. (1993). Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology, 2204-2214
- Kavetski, D., Franks, S.W., and Kuczera, G. (2002). Confronting input uncertainty in environmental modelling, in Calibration of Watershed Models, Water Sci. Appl., vol. 6, edited by Duan, Q., et al., pp. 49–68, AGU, Washington, D. C.
- Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, 42(3).
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest.R news, 2(3), 18-22.
- Ley, R., H. Hellebrand, M. C. Casper, and F. Fenicia (2016), Comparing classical performance measures with signature indices derived from flow duration curves to assess model structures as tools for catchment classification, Hydrology Research, 47(1), 1-14, doi:10.2166/nh.2015.221.
- Olden, J. D. and Poff, N. L. (2003), Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Applic., 19: 101–121. doi: 10.1002/rra.700
- Peñas Silva, F. J. (2014a). Clasificación del régimen hidrológico natural y predicción de características hidroecológicas en el tercio norte de la Península Ibérica. Classification of the natural flow regime and prediction of hydroecological characteristics in the Northern third of the Iberian Peninsula. PhD thesis, University of Cantabria, Santander, Spain.
- Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis, 49(4), 974-997.
- Raftery, A.E. (1993). Bayesian model selection in structural equation models. In Testing Stuctural Equation Models (K.A. Bollen and J.S. Long, eds.), Beverly Hills: Sage, pp. 163-180. Earlier version.

Reichert, P., and Omlin, M. (1997). On the usefulness of overparameterized ecological

models. Ecological Modelling, 95, 289-299. http://doi.org/10.1016/S0304-3800(96)00043-9

- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46, W05521. doi:10.1029/2009WR008328
- Snelder, T. H., Lamouroux, N., Leathwick, J. R., Pella, H., Sauquet, E., & Shankar, U. (2009). Predictive mapping of the natural flow regimes of France.Journal of Hydrology, 373(1), 57-67.
- Snelder T. H., Lamouroux N., Pella H. 2011. Empirical modelling of large scale patterns in river bed surface grain size. Geomorphology, 127: 189-197. DOI: 10.1016/j.geomorph.2010.12.015
- Snelder T. H., Barquin Ortiz J., Booker D. J., Lamouroux N., Pella H., Shankar U. 2012. Can bottomup procedures improve the performance of stream classifications? Aquatic Sciences, 74: 45-59. DOI: 10.1007/s00027-011-0194-7.
- Snelder, T. H., Datry, T., Lamouroux, N., Larned, S. T., Sauquet, E., Pella, H., & Catalogne, C. (2013). Regionalization of patterns of flow intermittence from gauging station records. Hydrology and Earth System Sciences, 17(7), 2685-2699.
- Vitolo, C., Buytaert, W. O. U. T. E. R., & Reusser, D. (2012, July). Hydrological Models as Web Services: An Implementation Using OGC Standards. In Proceedings of the 10th International Conference on Hydroinformatics, HIC, Hamburg, Germany.
- Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research, 44(12).
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., and Sorooshian, S. (2001). A framework for development and application of hydrological models. Hydrology and Earth System Sciences, 5 (1), 13-26. doi: 10.5194/hess-5-13-2001
- Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., and Gupta, H.V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrological Processes, 17(2): 455-476.
- Wagener, T., Wheater, H. S., and Gupta, H. V. (2004). Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, London, Imperial College Press.
- Wagener, T., Sivapalan, M., Troch, P. and Woods, R. (2007). Catchment classification and hydrologic similarity. Geography Compass, 1 (4), 901-931. doi: 10.1111/j.1749-8198.2007.00039.x
- Wagener, T. and Montanari, A. (2011). Convergence of approaches toward reducing uncertainty

in predictions in ungauged basins. Water Resources Research, 47 (6). doi: 10.1029/2010wr009469

- Westerberg, I. K., T. Wagener, G. Coxon, H. K. McMillan, A. Castellarin, A. Montanari, and J. Freer (2016), Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resour. Res., 52, 1847–1865, doi:10.1002/2015WR017635.
- Yadav, M., Wagener, T., and Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources, 30 (8), 1756-1774. doi: 10.1016/j.advwatres.2007.01.005
- Yosef, H. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance" (PDF). Biometrika. 75 (4): 800–802. doi:10.1093/biomet/75.4.800.
- Zhang, Z., Wagener, T., Reed, P., and Bhushan, R. (2008). Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization. Water Resources Research, 44(12).

http://servicios2.marm.es/sia/visualizacion/descargas/dma.jsp

http://info.igme.es/cartografiadigital/geologica/Magna50.aspx

Chapter 4: Contributions, Conclusions and Future Research

4.1 CONTRIBUTIONS

The overall aim of this thesis has been to develop methodologies to improve predictions of hydrological variables in ungauged basins, with a focus on addressing model structural uncertainty in PUBs via a Bayesian approach. This objective has been accomplished by fulfilling the following specific goals and the methodological developments have been applied to basins in northern Spain with varied hydro climatological regimes:

- 1. Incorporating new advances in the selection of hydrological indexes and regionalization methods into the regionalization procedure that provide information constrains in ungagued basins in a Bayesian approach. First, an extensive set of available hydrological signatures has been reduced to a compact orthogonal set of information pieces using principal components analysis. This advances the standard practice of utilizing a semi-empirical selection of individual hydrological signatures. Second, Random Forests that are free from assumptions concerning functional form have been used to regionalize the selected information. This allows the traditional assumption related to linear regression between catchment properties and the characteristics of hydrological response to be relaxed. Third, the regionalized information has been utilized to constrain hydrological predictions in the ungauged basins via a Bayesian approach.
- 2. Proposing two metrics for assessing the suitability and adequacy of a selected model (which includes structure, inputs and parameter sampling) and information regionalization procedure, and evaluating the impact of model error and the regionalization procedure on flow predictions via a Bayesian approach. Such metrics are based on a) how well the model reproduces regionalized information and how well the model structure and the regionalization procedure reproduce the hydrological information in the basin; and b) knowledge gain from considering both the model over what is known from regionalization alone and regionalization over what is known from the model alone. These two metrics have been defined based on the concepts of the Mahalanobis distance (see Gallego et al., 2013) and Bayes Factors [Gelman, 2013].
- 3. Devising a statistical test that employs a hydrological model ensemble to identify dominant hydrological mechanisms in ungagued basins that allow defining a minimal necessary structure for a hydrological model, via a Bayesian approach. The approach adopts the vision of the multiple working hypotheses framework [Clark et al., 2011]. The proposed test uses an ensemble of hydrological model structures and is based on the Bonferroni correction [Yosef, 1988] for the multiple comparisons problem.

4. Analyzing the sensitivity of the mechanisms identification process to a) the quality of the regionalization, b) the selected model ensemble error, and c) the information content.

4.2 CONCLUSIONS

4.2.1 STATISTICAL METRICS TO ASSESS REGIONALIZATION AND MODEL ADEQUACY IN PUBS

This subsection summarizes conclusions from Chapter 2. This Chapter has answered objectives 1 and 2 proposed in Chapter 1.

In predictions for ungauged basins, a common practice is to use a pre-selected model structure and regionalization technique for a catchment, while there is usually no justification for their suitability (ability to reproduce the available information) and adequacy (the knowledge gained by considering a model and/or regionalization). This study proposes statistical tests to quantify suitability and adequacy, and suggests a path to disaggregate the different sources of uncertainty and set priorities in order to reduce the predictive uncertainty.

This research advances the standard practices for predicting flows in ungauged basins when the information given by multiple regionalized hydrological indices is assimilated into rainfall-runoff models by a Bayesian conditioning process. The work uses Random Forest regionalized principal components that update the information given by a hydrological model via a Bayesian approach. The research also contributes with new suitability and adequacy tests:

- A suitability test is based on the simultaneous inference concept and explicitly quantifies the following null hypotheses: (a) a model is able to reproduce regionalized (or observed, if available) information with high reliability, (b) a regionalization procedure is able to reproduce the available information in a catchment.
- An adequacy test is based on Bayes Factors [Gelman, 2013] and explicitly quantifies the knowledge gained (a) by including a regionalization procedure, and (b) by including a model.

These suitability and adequacy tests allow identification of the dominant source of uncertainty (between the regionalization and the model) and, hence, where the efforts must be concentrated to reduce it. Application to a set of 16 catchments in northern Spain leads to the following findings:

 The error in the Random Forest regionalized principal components is relatively small with the advantage that regionalization by Random Forest only relies on a small number of assumptions (it does not require the preselection of a linear/non-linear model form and the number of predictors, it does not make assumptions about the model errors' statistical nature and it is an efficient – quality and time - data learning algorithm). Moreover, the high-dimensional hydrological indices space extensively reduced via PCA, and the "broken stick" method provides sufficient information for an efficient regionalization via Random Forest in the case of basins of northern Spain.

- When the model passes the suitability test for the regionalized information and the suitability and adequacy tests for the available information, improving regionalization quality can be targeted as a means for securing better predictions (see basins X1353, X9257 and C8Z1, Figure 6.).
- 3. However, model error precludes performance improvement when regionalization quality improves. Therefore, regionalization quality is not directly related to the predictions quality. Priority should be given to the selection of a representative model structure (which includes model inputs), then followed by regionalization improvement. Not the other way around.
- 4. Improving the accuracy and precision of regionalization for an error-free hydrological model provides better flow predictions quality (reliability, accuracy, precision).
- 5. The developed regionalization procedure via the Random Forest passes the suitability test (i.e. regionalization is likely to reproduce the hydrological information) in all catchments; and it also passes the adequacy test (i.e. provides a gain in knowledge) in 75% of basins.

4.2.2 IDENTIFICATION OF DOMINANT HYDROLOGICAL MECHANISMS FOR UNGAUGED BASINS VIA BAYESIAN APPROACH

This subsection summarizes conclusions from Chapter 3. This Chapter has answered objectives 3 and 4 proposed in Chapter 1.

This work proposes a statistical methodology to identify the dominant mechanisms in ungauged catchments and poses the path to reduce the identification uncertainty. The study merges the recent developments in hydrological indices selection and regression method with the strengths of the Bayesian inference. The study uses an ensemble of hydrological model structures FUSE [Clark et al., 2008; 2011b] to predict mechanisms in the ungauged basins, where each hydrological process is given by a combination of mechanisms and mechanisms are treated as a sample from a complete (but unobservable) set of mechanisms; and this information is combined with information provided by PCs regionalized by Random Forests, via a Bayesian approach. Further, the work provides a multiple hypothesis comparison test that allows identifying a dominant mechanism for each hydrological process in a catchment. An application to a set of 16 catchments in northern Spain leads to the following findings:

- 1. When model error exists, improving regionalization doesn't necessarily imply a better mechanism identification. Only when model and input errors are decreased/ eliminated and a large quantity of information is available, improving regionalization quality enhances the mechanism identification.
- 2. When there is no model/ input error and a large quantity of high quality regionalized information is available, the proposed procedure for dominant hydrological mechanism identification is reliable, i.e. if a mechanism is identified as dominant, then the identification is accurate.
- 3. The information content in the first four PCs allows identifying dominant mechanisms for routing, evaporation, interflow and runoff.
- 4. Commonly considered large set of hydrological indices [Olden and Poff, 2003] might be insufficiently for dominant mechanism identification for a number of hydrological processes (i.e. for the process in the upper soil, and for the percolation process).
- 5. Dominant hydrological mechanism identification is found sensitive to 1) model and input errors; 2) regionalization quality; and 3) information quantity.

4.3 FUTURE RESEARCH

Each chapter of this thesis, and especially Chapters 2 and 3, have led to several research questions that require answers. Namely: Disinformation due to misinterpretation of data, information content and model structure uncertainty.

4.3.1 Disinformation due to misinterpretation of data

Chapter 2 concluded, and chapter 3 corroborated, that in most of the basins, uncertainty due to the model error dominates regionalization uncertainty. Model error includes error due to model structure, inputs error and sampling parameters. This finding has led to the following research questions: is it feasible to disaggregate the uncertainty coming from the data from the uncertainty coming from the model structure? Which are the relative contributions of the information in the observations (post-processed series of observed data, see Nearing et al., 2016) and the information in the model structure? Which is the quantitative impact of the disinformative periods of data on the total predictive uncertainty [e.g., Reichert and Mieleitner, 2009]? (Note: it is worth to notice that this disinformation is introduced by propagating a generated dataset - based on observations of precipitation and flow data series-, through an incorrect model) How the amount of information associated with each storm can be quantified and how this information can be integrated into the likelihood function? These research questions align with the recently philosophical basis for hydrological uncertainty presented by Nearing et al. [2016]. In their paper, Nearing and colleagues claim that disinformation is not a

property of data itself, but rather is due to the ability to decode the data. That is to say, the way that the numerical output from a measurement device is related to the underlying physical state of the system that is attempted to be measured - either due to model uncertainty or to the fact that the measurements lack full specifying information about the evaluation observations. As well as the disinformation is fundamental and unavoidable due to the simulacrum account defined by Cartwright's [1983].

The issues raised have led to the Author of this thesis to design a methodology to identify disinformative periods in the post-processed [Nearing et al., 2016] data series of precipitation and flow, and then quantify the impact of such periods on the total predictive uncertainty. To that end, the proposed methodology relies on an assessment of the convergence of different approaches to detect such periods, and extends Wagener et al. [2003] Dynamic Identifiability Analysis (DYNIA). The Author of this thesis is working in this methodology under the supervision of Professor Dr. Thorsten Wagener (Bristol University) and in collaboration with the researchers Dr. Susana Almeida, and Dr. Christopher Hutton.

In this regard, the hydroclimatological data from catchments in northern Spain (Figure 1.) was analyzed in terms of (1) the distribution of rain gauges and thermometers in the catchments, using the cumulative density function of each catchment's elevation (Figure 19 and 20); (2) parallel coordinate plots of catchment descriptors and hydrological indices (Figures 21 and 22); and (3) flow duration curves. This examination shed light on the errors in the data. Further, MOPEX criteria were only met for both rain gauges and thermometers in a single basin. The findings presented in Chapter 2 and this further analysis led to the question of how to detect disinformative periods in the data and how to separate input uncertainty from uncertainty in the model structure.

SCENARIOS								
Constant Bias	1A	P-30% P	P-20% P	P-10% P	P+10% P	P+20% P	P+30% P	
	1B	Q-30% Q	Q-20% Q	Q-10% Q	Q+10% Q	Q+20% Q	Q+30% Q	
Missing Events	2A	Storm (4/3/06-12/3/06)	Storm (8/11/05-9/11/05)	Storm (27/12/05-5/1/06)	Storm (14/4/06-18/4/06)	Storm (5/1/05- 6/1/05)		
	2B	Flow (14/02/06-2/6/06)	Flow (28/10/05- 26/12/05) U (14/02/06-2/6/06)	Flow (3/3/05-13/3/05)	Flow (5/9/05-24/9/05)	Flow (15/12/04- 15/1/05)		
Untergendantic	ЗA	0.5*P	0.75*P	1.25*P	1.5*P			
Heterocedastic	ЗB	0.5*Q	0.75*Q	1.25*Q	1.5*Q			
Heterocedastic Auto correlated Random Error	4A	heteroscedastic multiplier 0.2	heteroscedastic multiplier 0.4	heteroscedastic multiplier 0.6	heteroscedastic multiplier 0.8			
	4B	heteroscedastic multiplier 0.2	heteroscedastic multiplier 0.4	heteroscedastic multiplier 0.6	heteroscedastic multiplier 0.8			
Random Error in Rainfall	5A	heteroscedastic multiplier 0.2	heteroscedastic multiplier 0.4	heteroscedastic multiplier 0.6	heteroscedastic multiplier 0.8			
Inverse Heterocedastic Auto correlated Random Error	6A	heteroscedastic multiplier 0.2	heteroscedastic multiplier 0.4	heteroscedastic multiplier 0.6	heteroscedastic multiplier 0.8			
	6B	heteroscedastic multiplier 0.2	Heteroscedastic multiplier 0.4	heteroscedastic multiplier 0.6	heteroscedastic multiplier 0.8			
Shift storms	7A	Storm (3/3/05-8/3/05) Shift to (2/3/05- 7/3/05)	Storm (11/8/05) Shift to (10/8/05)	Shift all 1 day	Storm (29/10/05-31/10/05) Shift to (28/10/05- 30/10/05)	St (29/10/05-31/1 18/ Shift to (28/10 (8/11/05	:orm 10/05)& (9/11/05- 11/05) 0/05-30/10/05)& 5-17/11/05)	

Table 8. Introduced	errors	for one	of the	catchments
TUDIE D. IIILI DUULEU	CHUIS	jui une	Uj line	cuttiments.



Figure 19: Elevation of the raingauges in the cumulative distribution function (CDF) of the digital elevation model of the catchments.



Figure 20: Elevation of the thermometers in the cumulative distribution function (CDF) of the digital elevation model of the catchments.

Thus, work has commenced to establish a methodology centered around the following steps: First, classify the most common types of errors in the observed rainfall and flow data series on the basis of, for example, the catalogue provided by McMillan et al. [2012]. Table 8 provides an example of this approach, with the first column indicating the type of error and the second column denoting the rainfall (part A) and the flow (part B). Second, develop a procedure for identifying synthetic errors. This procedure considers the data space, parameter space, and model structure space in light of the assumption that a combined approach can help detect disinformative events. This proposed procedure is as follows:

- 1. Generate a synthetic time-series of discharge Q_T ("true" case) with a hydrological model structure, using a real rainfall (R_T) time series. The methodology has started to be tested with the PDM (Chapter 2, section 4) hydrological model [Moore, 2007], however the procedure is general to be implemented with other model structures.
- 2. Create metrics/graphs of performance to identify potentially erroneous data for the "true" case. That case then acts as a reference for comparison when additional errors are introduced to the data. These metrics/graphs are: (a) an event-based rainfall

runoff coefficient for the data space; (b) DYNIA [Wagener et al., 2003] with PDM [Moore, 2007] allowing only α parameter to vary for the parameter space. The α parameter was selected in first instance because it represents the contribution of quick and slow flows; and (c) the ensemble of behavioral models for the model structure space,

- 3. Plot the 3 metrics/graphs for each specific type of error introduced to the true data $(R_T, Q_T first column in Table 8)$ to compare them to the true-case metrics calculated in step 2. In its horizontal rows, Table 8 indicates the tested scenarios for each type of error.
- 4. From the perspective of comparative hydrology, assess how these same error types affect catchments in different climatic regions, with different catchment descriptors, and different hydrological signatures (such as those highlighted in the parallel coordinate plots in Figures 21 and 22).



Figure 21: Parallel coordinate plots of catchment descriptors for catchments in northern Spain. X-axis: catchments descriptors. Y-axis: value of catchments descriptors. The first set of explored catchments is highlighted.



Figure 22: Parallel coordinate plots of hydrological indices for catchments in northern Spain. X-axis: hydrological indices. Y-axis: value of hydrological indices. The first set of explored catchments is highlighted.

Preliminary results

The four catchments highlighted in Figure 21 and Figure 22 were examined and findings show that the event-based rainfall runoff coefficient plots only captures the erroneous data when the value of that coefficient is above 1. Therefore, a larger rainfall runoff coefficient made it easier to detect disinformative events in the catchments, while drier conditions make this task more challenging. The value of the rainfall runoff coefficient based on events is above 1 when either rainfall is underestimated or discharge is overestimated (e.g., Figure 23). For other cases, an understanding of likely evapo-transpiration and storage is necessary to evaluate whether very small runoff coefficients are plausible. Moreover, the dynamic identifiability analysis seems to be effective at detecting missing events (rain and flow) on the basis of parameter α (e.g., Figure 24). Thus, it is suggested to go further with future research employing the DYNIA [Wagener, 2003] approach to analyze other parameters from the PDM model [Moore, 2007], i.e. how the corrupted data map (or not) in DYNIA for other PDM hydrological model parameters different from α . Further studies should also investigate the effect of the combination of errors on predictive uncertainty.



Figure 20: Data space: example of the event-based rainfall runoff coefficient for a constant bias in the flow series. Red: flow (mm/day). Black: rainfall runoff coefficient based on events. X-axis: time (days). left y-axis: bias in the "true" flow series. Right y-axis: rainfall runoff coefficient.



Figure 24: Parameter space: example of DYNIA [Wagener, 2003] for missing events of flow. X-axis: time (days). Y-axis: contribution of quick flow [Moore, 2007].

This approach to detecting disinformative periods is naturally linked to the above mentioned research question about: How much information is contained in each storm, and how can it be quantified? This question led to the Author of this thesis to propose the research project, "Different ways of assessing storm-by-storm information content". This proposal has been supervised by Emeritus Professor Dr. Keith Beven (Lancaster University) and constitutes work

for the advanced course "Uncertainty in Environmental Modeling" (June 2015, Uppsala University).

In this regard, some suggestions include separating informative and disinformative events (via the above approach), extending Beven and Smith's [2014] framework to evaluate the information content in the storms and defining the likelihood. As well as using Bayes Factors [Kass and Raftery, 1995] to evaluate how much each model is supported by the information in the storms.

4.3.2 Information content

Chapter 3 of this thesis has demonstrated that all available information is not required to obtain reliable predictions. That finding prompted the following research questions: Which PCs contain information restricting the predictions for each basin? What is the relationship between the information contained in these PCs (and the hydrological indices contained therein) and the basins' dominant mechanisms? Which PCs should be selected a priori? How much and which hydrological information (synthesized in the PCs) can be assimilated into hydrological model's structure? Knowing the trade-off between model complexity and the information contained in the indices would identify where more resources should be employed to reduce the predictive uncertainty. The Author of this thesis is already exploring a formal methodology to address these questions. This methodology relies on a sensitivity analysis approach and is thus aligned with recently published work by Markstrom et al. [2016].

Preliminary results

A first analysis has indicated that in some catchments, only 13% of available information is needed to produce reliable predictions. Moreover, a regional sensitivity analysis [Spear and Hornberger, 1980] has demonstrated that a dominant mechanism has not been identified for some catchments, because different model structures with dissimilar mechanisms generate similar PC vectors (in the n-dimensional PC space). In other words, the degree of separation in the n-dimensional PC space allows identifying the dominant mechanism. The routing process (2 competing mechanisms) and the 4PCs depicted in Figure 25 (a basin for which a dominant mechanism has been identified) and Figure 26 (no dominant mechanism has been identified) exemplify this.

Figure 25 illustrates a shift between the posterior distributions of mechanisms 1 and 2 for PC1. Thus, for PC1, the posterior probability of no routing process is, on average, lower than the posterior probability of the presence of a routing process. The distance from the posterior probability of no routing process to the actual PC1 is greater than the distance from the posterior probability of routing process to the actual PC1. Therefore, the posterior probability of no routing process is lower. A regional sensitivity analysis [Spear and Hornberger, 1980] revealed that each PC is sensitive (to a greater or lesser degree) to the routing process (Figure 25). However, when all 4 PCs are simultaneously reproduced in Figure 26, there are model structures both with and without routing mechanisms that produce similar quadruples of PCs. Figure 25 does not depict such behavior. Rather, the degree of separation in the posterior distributions for the 4D vector (i.e., the 4 PCs together) is higher. This likely points to a higher separation in the 4D space in Figure 25, such that the true vector of the PCs is further from the PCs generated by the no-routing structures. Therefore, the degree of separation between each competing mechanism in the 4D PCs space is what matters. In addition, adopting other global sensitivity analysis methods to analyze the interactions and the effect of the secondary factors is recommended.



Figure 25: Sensitivity of the routing process to the information in the individual PCs. A dominant mechanism is identified. X-axis: value of each individual PCi, i=1,4. Y-axis: p(PCi <=X), i=1, 4. Black: Posterior Cumulative Density Function of "no routing". Red: Posterior Cumulative Density Function of "routing". Blue: Corresponding "true" PC.



Figure 26: Sensitivity of the routing process to the information in the individual PCs. A dominant mechanism is not identified. X-axis: value of each individual PCi, i=1,4. Y-axis: p(PCi <=X), i=1, 4. Black: Posterior Cumulative Density Function of "no routing". Red: Posterior Cumulative Density Function of "routing". Blue: Corresponding "true" PC.

Additionally, a methodology to select hydrological indices based on their information content is being devised. This research is in collaboration with the researchers from National Center of Atmospheric Research (NCAR) Dr. Nans Addor and Dr. Martyn Clark. The Author of this thesis is contributing as co-author with the conception of the methodology, intellectual and research effort and results interpretation. Part of this methodology was presented by these researchers and the Author of this thesis at AGU 2016 [Addor et al., 2016] and will also be presented at EGU 2017 [Addor et al., 2017]. The methodology is being developed using 671 catchments in the contiguous United States and can be summarized as follows. A first goal was to investigate whether hydrological indices can be inferred from catchments descriptors alone. To this end, hydrological indices and catchment descriptors have been related via Random Forests datalearning algorithm. The predictions using Random Forests have been further compared with those obtained using a conceptual hydrological model (SACRAMENTO). First findings revealed that climatic attributes have the most significant influence, and strongly condition how well hydrological indices can be predicted by Random Forests and simulated by the hydrological model. In contrast, soil characteristics at the catchment scale are not found to be significant predictors, which raises questions on how to best use soil data for hydrological modeling. Furthermore, indices with high spatial variability are challenging to predict. A ranking of indices based on their information content was proposed, and it is suggested that indices with high information content are best suited for model calibration, model selection and understanding hydrologic similarity. See Addor et al., [2017] for further details.

4.3.3 Model structure uncertainty

A) Model selection and combination

The methodological developments presented in Chapter 3 use the complete ensemble of available model structures. However, utilizing the entire available competing hypothesis may be a computationally expensive endeavor. Exploring a strategy that allows integrating into the ensemble those configurations of model structures and parameters that are most supported by the regionalized information in each ungauged basin is suggested.

The Author of this thesis has initiated investigations in this direction and advances have been made. The employed strategy adopted the vision of the multiple working hypotheses framework [Chamberlain, 1965; Clark et al., 2011] and was based on the following steps. First, for each ungauged basin, the available model structures and parameters configurations were ranked on the basis of how likely they were to reproduce the inner-workings of an ungauged catchment [Beven, 2001], using the regionalized information (more specifically regionalized PCs in the context of this thesis). Second, the model structures and parameters with the most support are employed to create a multi-configuration ensemble for each ungauged basin, with each member assigned a weight that is proportional to the regionalization error (as derived through this thesis), via a conditioning Bayesian approach.

Reliability is lost when the most supported configurations [Jeffreys, 1961] are employed with respect to using an approach that employs the fully available (but incomplete) ensemble; however, at the same time gains in precision and computational efficiency are obtained. It is therefore compared the quality of the predictions achieved through the approach using the selected model structures and parameters [Jeffreys, 1961] with the quality of those attained through the approach that utilizes the full available model structures and parameters [Prieto et al., 2016].

The work on this area to date has followed the methodology described below: First, a selection of configurations has been made based on BFs. The reason for using BFs is that Bayesian methods are well suited for comparing and choosing hypotheses [Raftery, 1993]. Bayesian methods are also an appropriate choice for describing system nonlinearities and uncertainties [e.g., Kavetski et al., 2002; Vrugt et al., 2008; Renard et al., 2010; Bulygina et al., 2009; Almeida et al., 2014]. Furthermore, BFs have many desirable properties: they do not make the strong

193

assumption that one hypothesis is true [Raftery, 1993], they permit numerous hypotheses to be compared without methodological changes, they do not require models to be nested and they protect against model complexity and overfitting (which means they automatically favor simpler models [Marshal et al., 2005]). Finally, they also offer a quantitative scale for categorizing and interpreting the extent to which each hypothesis supports the ungauged basin's regionalized information [Jeffreys, 1961]. Initially, Bayes' Factors [Gelman, 2013] as given in (4.1.) have been employed to rank the hydrological configurations.

$$BF_{k,0} = \frac{L\left(\theta_{i,M_k}, M_k \middle| PC^{Reg}\right)}{L\left(\theta_{j,M_0}, M_0 \middle| PC^{Reg}\right)} (4.1)$$

Equation (4.1.) indicates that given two hydrological configurations, namely ($\theta_{i,Mk}$, M_k) and ($\theta_{j,M0}$, M_0), as well as the residuals distribution of the regionalization procedure and inputs *I*, BFs [Gelman, 2013] are defined as a posterior-to-prior-odds ratio for configuration ($\theta_{i,Mk}$, M_k) that is more effective for reproducing the regionalized information PC^{Reg} than configuration ($\theta_{j,M0}$, M_0) [Kass and Raftery, 1995]. Considering the same ensemble of models, but different configurations that are equally probably a priori; setting also that the hypothesis in the denominator used as benchmark is a uniform prior distribution in the configurations space; and setting that the hypothesis in the numerator is one of the posterior configurations posterior (see deductions in Almeida, 2014), equation (4.1) becomes equation (4.2):

$$BF_{k,0} = \frac{p(\theta_{i,M_k}, M_k \mid PC^{Reg})}{p(\theta_{i,M_0})}$$
(4.2)

The logic behind using the uniform configurations prior distribution as a benchmark is that this prior distribution represents what is known before any hydrological indices or PCs are introduced. In other words, it represents a step in the system's understanding that precedes the point at which the PCs are calculated.

Once the competing configurations are ranked, the second step is to select the configurations. A way that has been explored is selecting those configurations that are at least strongly supported (BFs > 10) by the regionalized information [Jeffreys, 1961] and to combine these configurations via a Bayesian approach. This is in contrast with to the traditional approach, in which predictions are usually made based on the "best" single hypothesis. This traditional approach also leads to an overconfidence of the predictive capacity of the selected configurations and is prone to bias. Moreover, it results in predictions that accurately represent some phenomena at the expense of others. Thus, if there are N_{tb} configurations with BFs > 10 [Jeffreys, 1961], i.e. selected as the most appropriate choice for the ungauged basin, these selected configurations (θ_s , M_s), s={1...Ntb}, are combined using Bayes' law:

$$p(\theta_s, M_s | PC^{Reg}, I) \propto L(\theta_s, M_s | PC^{Reg}) * p(\theta_s, M_s)$$
 (4.3)

where: $L(\theta_s, M_s | PC^{Reg})$ is the likelihood that the configuration (θ_s, M_s) reproduces PC^{Reg} , and $p(\theta_s, M_s)$ is a uniform prior distribution in the selected configurations space.

Preliminary results:

Initial investigations regarding the two approaches (i.e. using both the most supported configurations and the full available configurations) were conducted; moreover, the cases BC, R, M and MR as defined in Chapter 3 were compared. The results indicated that: 1) a small subset of the most supported (i.e. BFs > 10 [Jeffreys, 1961]) configurations of model structures and parameters provides reliable, accurate and precise predictions when the ensemble is able to reproduce the regionalized information and regionalization is of high quality; 2) the reliability, accuracy and precision of predictions made using the two approaches converge in the absence of a model error and when the regionalization quality is high –this was found by a set of ANOVA experiments; and the convergence in terms of reliability might be linked to the regionalization quality; and 4) the convergence in terms of precision might be linked to the model error.

The above results are illustrated in the bar plots in Figures 27–29. In particular, Figure 27 provides reliability in terms of the 95% CI for each case; bar plots in Figure 28 indicate the level of precision (estimated as the mean of the posterior 95% CI); and Figure 29 displays the posterior *NS*^{prob}. Figure 30 then compares the prediction bounds using the two approaches for a high-performing catchment in terms of *NS*^{prob} for case BC; Figure 31 does the same for a poorly performing catchment.

195



Figure 27: 95% CI for cases a) BC, b) R, c) M and d) MR.



Figure 28: Precision for cases a) BC, b) R, c) M and d) MR.



Figure 29: NSprob for cases a) BC, b) R, c) M and d) MR.



Figure 30: 95% CI for a high-performing catchment for cases BC, R, M and MR – see Chapter 3 for definitions of these cases.



Figure 31: 95% CI for a poorly performing catchment for cases BC, R, M and MR – see Chapter 3 for definitions of these cases.

B) Impact of the prior distribution on hypothesis selection

The above analysis led to inquiries about the impact of the prior distribution in the hypothesis selection. When regionalized indices are used for Bayesian conditioning in ungauged basins, it is assumed that a uniform prior distribution on the hydrological model parameter space is an appropriate representation of the lack of knowledge concerning the hydrological indices in the ungauged catchment [Almeida et al., 2013]. Although uniform prior distributions in the hydrological model parameters space have commonly been used to represent the lack of knowledge about a system [Bulygina et al., 2011], the selection of an objective prior distribution that should be utilized in hydrological modeling remains a major challenge for the hydrological community [Renard et al., 2011] (see also Chapter 1 of this thesis). Almeida et al. [2013] have demonstrated that when hydrological indices are employed to constrain hydrological predictions in data-scarce environments, a prior parameters distribution that maps into uniform indices should be used instead of a uniform distribution in the parameters space. This is because the latter leads to preferred behaviors of the system (which are summarized in hydrological indices, or in the context of this thesis PCs).

These reflections have given rise to research questions concerning how much the model structures are supported by the regionalized information when a uniform distribution is used in the hydrological indices space as compared to when a uniform distribution is utilized in the hydrological parameters space. In other words, how much information does each model structure adds about the regionalized information when a uniform distribution is used in the hydrological indices space as opposed to in the hydrological model parameters space.
Preliminary results

The Author of this thesis has begun to explore these issues using BFs as given by equation (4.4.).

$$BF_{uss,ups} = \frac{L\left(M_{k} \mid PC^{Reg}, I\right)_{uss}}{L\left(M_{k} \mid PC^{Reg}, I\right)_{ups}} = \frac{\int_{\Theta_{k}} L\left(PC_{\theta ik}^{sim} \mid PC^{Reg}, I\right)_{*p}\left(PC_{\theta ik}^{sim} \mid I\right)_{uss} d\theta_{i,k}}{\int_{\Theta_{k}} L\left(PC_{\theta ik}^{sim} \mid PC^{Reg}, I\right)_{*p}\left(PC_{\theta ik}^{sim} \mid I\right)_{ups} d\theta_{i,k}}$$
(4.4)

where: $L(PC_{\theta_{ik}}^{sim}|PC^{Reg}, I)$ is the likelihood of $PC_{\theta_{i,k}}^{sim}$ simulated by the model M_k with inputs I and parameters $\theta_{i,k}$, given PC^{Reg} ; $p(PC_{\theta_{i,k}}^{sim}|I)_{uss}$ is a prior distribution of the PCs; and $p(PC_{\theta_{i,k}}^{sim}|I)_{ups}$ is a prior distribution of the hydrological model parameters. The likelihood function $L(PC_{\theta_{i,k}}^{sim}|PC^{Reg}, I)$ is proposed to be proportional to the PCs residuals distribution from the regionalization procedure as used in this thesis, the prior $p(PC_{\theta_{i,k}}^{sim}|I)_{uss}$ is proposed to be a uniform distribution in the space of the PCs [Almeida et al., 2013], and $p(PC_{\theta_{i,k}}^{sim}|I)_{ups}$ is a uniform distribution in the hydrological parameters space. As it is generally not possible to sample from $p(PC_{\theta_{i,k}}^{sim}|I)_{uss}$, this prior distribution can be represented by sampling hydrological parameters $\theta_{i,k}$ uniformly from a hypercube Θ_k and assigning the following weights to the sampled parameter sets [Almeida et al., 2013]:

$$p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{uss} \propto \frac{p\left(PC_{\theta_{i,k}}^{sim}|I\right)_{ups}}{p\left(PC_{\theta_{i,k}}^{sim}|\{M_k\},\{\Theta_k\},I\right)} \quad (4.5)$$

Where $p(PC_{\theta_{i,k}}^{sim}|\{M_k\}, \{\Theta_k\}, I)$ is a joint probability density function of PCs generated by all models $\{M_k\}$ with model parameters sampled uniformly from the corresponding parameter hypercubes $\{\Theta_k\}$ and input *I*. $p(PC_{\theta_{i,k}}^{sim}|\{M_k\}, \{\Theta_k\}, I)$ is proposed to be approximated via a mixture of Gaussians [Muller et al., 1996] or multidimensional histograms. This approach decreases the probability of the sets of parameters that lead to favorite $PC_{\theta_{i,k}}^{sim}$ and implementing it has revealed that all model structures lead to BFs over 100 in the Jeffreys [1961] scale. The models are therefore decisively supported by regionalized information when a uniform prior distribution in the PCs space is used.

The idea behind utilizing uniform prior distribution in the PCs space motivated research question concerning whether the selection and combination of model structures might be performed by BFs [Kass and Raftery, 1995], but using uniform prior distribution in the PCs in the numerator and using a uniform distribution in the hydrological model parameters space as a benchmark in the denominator (4.6). It is suggested to further explore this approach:

$$BF_{uss,ups} = \frac{\int_{\Theta_k} L(PC_{\theta_{i,k}}^{sim} | PC^{Reg}, I)_{*p}(PC_{\theta_{i,k}}^{sim} | I)_{uss}^{d\theta_{i,k}}}{\int_{\Theta_k} p(PC_{\theta_{i,k}}^{sim} | I)_{ups}^{d\theta_{i,k}}} \quad (4.6)$$

Once the competing model structures have been ranked, it is proposed to combine those that add substantial knowledge about the regionalized PCs (i.e. BFs > 10 [Jeffreys, 1961]) using Bayesian model averaging (BMA) [e.g. see Vrugt and Robinson, 2007] as given in equation (4.7):

$$p(Q|PC^{Reg}, I) = \sum_{k=1}^{K} p(Q|M_k, PC^{Reg}, I) * p(M_k|PC^{Reg}, I)$$
(4.7)

Where $p(M_k | PC^{Reg}, I)$ is the posterior probability of a model M_k from a set of hydrological model structures $\{M_k\}$, k=1,N and it is provided by Bayes' law (4.8):

$$p(M_k | PC^{Reg}, I) \propto L(M_k | PC^{Reg}, I) * p(M_k | I)_{uss}$$
(4.8)

 $L(M_k | PC^{Reg}, I)$ is a likelihood of the hydrological model M_k given PC^{reg} and inputs *I*; and $p(M_k | I)_{uss}$ is a prior probability of the model M_k . A key assumption is that at least one model structure M_k and its inputs *I* are capable of representing the true underlying PCs. This leads to equation (4.8) being re-written as:

$$p(M_k | PC^{Reg}, I) \propto \int_{\Theta_k} L\left(PC^{sim}_{\theta_{i,k}} | PC^{Reg}, I\right) * p\left(PC^{sim}_{\theta_{i,k}} | I\right)_{uss} d\theta_{i,k} * \int_{\Theta_k} p\left(PC^{sim}_{\theta_{i,k}} | I\right)_{uss} d\theta_{i,k} \quad (4.9)$$

In equation (4.9.), $p(Q|M_k, PC^{Reg}, I)$ is the posterior predictive distribution of flows. This distribution might be fitted using a mixture of Gaussians, where the covariance is obtained by approaching a multivariate kernel distribution to the simulated flows Q.

In the case that BFs (4.6.) select a high number of model structures, the following strategy is planned to pursue BMA [Vrugt et al., 2011]: randomly sample N models from the ensemble of $\{M_s\}$ selected structures, run BMA, generate a performance metric (e.g., RMSE), and then minimize this metric by selecting the right combination of N models from $\{M_s\}$ such that the best selection of N models with respect to such metric is obtained.

Finally, it is worth noting the following regarding BMA [Jasper Vrugt 2017, personal communication]: *BMA (as a weighting approach) will not in any way help to determine epistemic uncertainty; it is an engineering solution to conceptual uncertainty that does not reveal what exactly is wrong with a model. However, as BMA looks at a combination of models, by definition it might not include the best model. The thinking is that averaging weaker models might produce a better simulation forecast than the best model; as such, the point of averaging is to obtain the best weighted forecast. However, if ensemble spread does not capture the data, BMA does not do anything as weights must lie on simplex (i.e. be positive and add up to one). Bayesian model*

averaging is an interpolation method, therefore, the BMA model is always found between the extreme ends of the forecasts/simulations.

C) Model structure error

Of course, the problems related to model structure uncertainty quantification and reduction are not new, and their importance has been widely claimed in hydrology [e.g. Gupta et al., 2008; Clark et al., 2008, 2011; Wagener et al., 2001; Bulygina and Gupta, 2010; Fenicia et al., 2008, 2011]. Based on this and the above statement [Jasper Vrugt 2017, personal communication], it is suggested to apply the methodologies developed in Chapter 3 with the "Structure to Unify Multiple Modeling Alternatives" (SUMMA; Clark et al., 2015a; 2015b; 2015c). In particular, SUMMA permits a common set of conservation equations to represent different physical processes.

It is also proposed to investigate whether growth equations commonly applied to predict the evolution of different type of complex systems (e.g. in biology [Blumberg, 1968], demography [Verhulst, 1838; Smith, 1963]; ecology [Richards, 1959], etc.) can be transposed to approach hydrological systems behavior, with an operational purpose. In this regard, the framework presented in Prieto et al. [2013] and applied by Alvarez et al. [2015] is suggested to be explored. Prieto et al. [2013] presented a parsimonious (3-5) lumped hydrological model, LEM (Logistic Equilibrium Model), under the idea that a hydrological system tends to reach an equilibrium discharge, given persistent conditions of rainfall, PET and land uses. And this equilibrium discharge might be expressed as a function of a dynamic aridity index. Initially, LEM was proposed by identifying homologies between the growth of a generic system and the evolution of the flow in the outlet of the catchment. To date, LEM [Prieto et al., 2013] has been applied with consultancy purposes in the Nam Ngum river basin, Lao PDR, for a Hydropower Decision Support (World Bank, 2013) [Alvarez et al., 2015]. Furthermore, LEM [Prieto et al., 2013] has been applied to Evaluate Water Supply Risks into the Emerging and Sustainable Cities initiative (Inter-American Development Bank [http://www.iadb.org/en/topics/emerging-andsustainable-cities/emerging-and-sustainable-cities-initiative,6656.html] in Huancayo (Peru), Panama City (Panama), Tegucigalpa (Honduras), Santa Marta (Colombia) and Cuzco (Perú).

Moreover, the variance of the hydrological indices and the stochasticity of the model parameters should also be integrated into the likelihood function. The methodologies developed in this thesis are well suited for that task. Assimilating the temporal variability of the hydrological indices into the likelihood function and using the "trading-space-for-time" approach would also make it feasible to assess the effects of climate change and temporal

201

variability in land use on the predictions chain. This proposition links with the recent *panta rhei* (everything flows) paradigm [Montanari et al. 2013].

4.3.4 Preliminary conclusions of the future research

Disinformation due to misinterpretation of data: Methodology to identify disinformative periods of data

- 1. A larger rainfall runoff coefficient makes it easier to detect disinformative events in the catchments, while drier conditions make this task more challenging. The value of the rainfall runoff coefficient based on events is above 1 when either rainfall is underestimated or discharge is overestimated.
- 2. DYNIA [Wagener et al., 2003] detects missing events (rain and flow) on the basis of parameter α that quantifies the contribution of the quick flow.
- 3. The next step in this research will employ the DYNIA [Wagener, 2003] approach to analyze other parameters from the PDM model [Moore, 2007].
- 4. It is needed to explore a methodology that allows quantifying the information content in the storms.

Information content: Sensitivity of the process to a dominant mechanism

- 1. The identification of dominant mechanisms is sensitive to the degree of separation between each competing mechanism in the 4D PCs space. I.e. not to the separation of individual PCs.
- 2. Adopting other global sensitivity analysis methods (different from RSA) to analyze the interactions and the effect of the secondary factors is recommended.
- 3. Climatic indices have the most significant influence, and strongly condition how well hydrological indices can be predicted by RFs and simulated by the hydrological model.
- 4. Soil characteristics are not found to be significant predictors. This can be attributed to the data resolution to be assimilated into the models (statistical or hydrological).
- 5. Hydrological indices with high spatial variability are challenging to be predicted.
- 6. It is necessary to develop a ranking of hydrological indices based on their information content. Such that, this classification provides basis for constraining model predictions, selecting model structures and understand hydrological similarities.

Model structure uncertainty

A) Model selection and combination: computational problem

 A small subset of the most supported (i.e. BFs > 10 [Jeffreys, 1961]) configurations of model structures and parameters provides reliable, accurate and precise predictions when the ensemble is able to reproduce the regionalized information and regionalization is of high quality.

 Reliability, accuracy and precision of predictions made using the two approaches converge in the absence of a model error and when the regionalization quality is high. Convergence in terms of reliability might be linked to the regionalization quality; and convergence in terms of precision might be linked to the model error.

B) Impact of the prior distribution on hypothesis selection

- All models are decisively supported by regionalized information when a uniform distribution is used in the hydrological indices space as compared to when a uniform distribution is utilized in the hydrological model parameters space.
- 2. Selection of model structures might be performed by BFs [Kass and Raftery, 1995] using uniform prior distribution in the PCs in the numerator and using a uniform distribution in the hydrological model parameters space as a benchmark in the denominator.
- Once the competing model structures have been ranked, it is proposed to combine those that add substantial knowledge about the regionalized PCs (i.e. BFs > 10 [Jeffreys, 1961]) using Bayesian model averaging (BMA) [e.g. see Vrugt and Robinson, 2007].
- 4. In the case that BFs select a high number of model structures, the following strategy is planned to pursue BMA [Vrugt et al., 2011]: randomly sample N models from the ensemble of $\{M_s\}$ selected structures, run BMA, generate a performance metric (e.g., RMSE), and then minimize this metric by selecting the right combination of N models from $\{M_s\}$ such that the best selection of N models with respect to such metric is obtained.

C) Model structure error: Logistic Equilibrium Model

- 1. The Logistic Equilibrium Model (LEM) might be a parsimonious (3-5 parameters) hydrological representation of the analyzed humid catchments.
- LEM has provided accurate results from an operational point of view in the analyzed case studies in U.K., the Nam Ngum river basin (Lao PDR), Huancayo (Peru), Panama City (Panama), Tegucigalpa (Honduras), Santa Marta (Colombia) and Cuzco (Ecuador).
- 3. Therefore, the following homologies between the growth of a generic system and the evolution of the flow at the outlet of a river basin might be used as basis for lumped conceptual models: 1) the concept of the equilibrium discharge that can be expressed as a function of a dynamic aridity index, 2) a rate at which the system approaches such equilibrium discharge, which is constantly changing and generally not attainable, and 3)

a delay between the input (precipitation) and output (discharge) in the system behaviour.

4.4 **REFERENCES IN THE CHAPTER 4.**

- Addor, N., Clark, M., Prieto, P., Andrew, J., Mizukami, N., Nearing, G., Le Vine, N. (2016). On the information content of hydrological signatures and their relationship to catchment attributes, Abstract for the Fall Meeting of the American Geophysical Union, San Fransisco,
 https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/160052
- Almeida, S., Bulygina, N., McIntyre, N., Wagener, T., & Buytaert, W. (2013). Improving parameter priors for data-scarce estimation problems. Water Resources Research, 49(9), 6090-6095.
- Almeida, S. M. C. L. (2014). The Value of Regionalised Information for Hydrological Modelling. PhD thesis, Imperial College London, London, UK.
- Beven, K., & Smith, P. (2014). Concepts of information content and likelihood in parameter calibration for hydrological simulation models. Journal of Hydrologic Engineering, 20(1), A4014010.
- Blumberg, A.: Logistic Growth Rate Functions, J. Theor. Biol, 21 (1968) 42, 1968
- Bulygina, N., McIntyre, N. and Wheater, H. (2011). Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. Water Resources Research, 47 (2), doi: 10.1029/2010wr009240.
- Markstrom, S. L., Hay, L. E., and Clark, M. P.: Towards simplification of hydrologic modeling: identification of dominant processes, Hydrol. Earth Syst. Sci., 20, 4655-4671, doi:10.5194/hess-20-4655-2016, 2016.
- McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality.Hydrological Processes, 26(26), 4078-4111.
- Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T. and Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44 (12), doi: 10.1029/2007wr006735.
- Clark, M. P., Kavetski, D. and Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research, 47, doi: 10.1029/2010wr009827.

- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M. (2015a): A unified approach for process-based hydrologic modeling: 1. Modeling concept. Water Resources Research, 51, 2498-2514. doi: 10.1002/2015wr017198
- Clark, M. P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger, G. N., and Marks, D. G. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. Water Resources Research, 51, 2515-2542. doi: 10.1002/2015wr017200
- Clark, M. P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W.,, Brekke, L. D., Arnold, J. R., Gochis, D. J., Rasmussen, R.M., Tarboton, D.G., Mahat, V., Flerchinger, G.N., and Marks, D.G. (2015c). The Structure for Unifying Multiple Modeling Alternatives (SUMMA), version 1: Technical description. NCAR Technical Note NCAR/TN-514+ STR, 54 pp., National Center for Atmospheric Research, Boulder, Colo. doi: 10.5065/D6WQ01TD
- Montanari, Alberto, G. Young, H. H. G. Savenije, D. Hughes, T. Wagener, L. L. Ren, D. Koutsoyiannis et al. ""Panta Rhei—everything flows": change in hydrology and society—the IAHS scientific decade 2013–2022." Hydrological Sciences Journal 58, no. 6 (2013): 1256-1275.
- Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., & Weijs, S. V. (2016). A philosophical basis for hydrological uncertainty. Hydrological Sciences Journal, 61(9), 1666-1678.
- Olden, J. D. and Poff, N. L. (2003), Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Applic., 19: 101–121. doi: 10.1002/rra.700
- Prieto, C., Le Vine, N., Vitolo, C., García., E, Medina, R. (2016a), Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Abstract for the Meeting of the European Geophysical Union, Vienna, available online: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-8580-2.pdf
- Prieto, C., Le Vine, N., Vitolo, C., García., E, Medina, R. (2016a), Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach. Abstract for the Meeting of the European Geophysical Union, Vienna, available online: http://presentations.copernicus.org/EGU2016-8580_presentation.pdf
- Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in Peel Inlet—II. Identification of critical uncertainties via generalized sensitivity analysis.Water Research, 14(1), 43-49.

- Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, Water Resour. Res., 45, W10402, doi:10.1029/2009WR
- Reichert, P., Langhans, S. D., Lienert, J., & Schuwirth, N. (2015). The conceptual foundation of environmental decision support. Journal of environmental management, 154, 316-332.007814.
- Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks (2011). Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, Water Resour. Res., 47, W11516, doi:10.1029/2011WR010643.
- Richards, F. J.: A flexible growth function for empirical use, Journal of experimental Botany, 10(2): 290-301, 1959.
- Smith, F. E., Population dynamics in Daphnia Magna and a new model for population growth, Ecology 44 (4), 1963
- Verhulst, P. F. (1838): Notice sur la loi que la population suit dans son accroissement, Correspondance Mathematique et Physique Publiee par A, Quetelet 10: 113-121.
- Vrugt, J. A., & Robinson, B. A. (2007). Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research, 43(1).
- Vrugt, J. A., & Ter Braak, C. J. (2011). DREAM (D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. Hydrology and Earth System Sciences, 15(12).
- Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrological Processes, 17(2), 455-476.
- Yosef, H. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance" (PDF). Biometrika. 75 (4): 800–802. doi:10.1093/biomet/75.4.800.

206

Published Work in Conference Proceedings

Geophysical Research Abstracts Vol. 17, EGU2015-9832, 2015 EGU General Assembly 2015 © Author(s) 2015. CC Attribution 3.0 License.



Improving standard practices for prediction in ungauged basins: Bayesian approach

Cristina Prieto (1), Nataliya Le-Vine (2), Eduardo García (1), and Raúl Medina (1) (1) IHCantabria, University of Cantabria, Santander, Spain (prietoc@unican.es), (2) Imperial College London, UK

In hydrological modelling, the prediction of flows in ungauged basins is still a defiance. Among the different alternatives to quantify and reduce the uncertainty in the predictions, a Bayesian framework has proven to be advantageous. This framework allows flow prediction in ungauged basins based on regionalised hydrological indices. Being grounded on probability theory, the procedure requires a number of assumptions and decisions to be made. Among the most important ones are 1) selection of representative hydrological signatures, 2) selection of regionalization model functional form, and 3) a 'perfect' model/ input assumption. The contribution of this research is to address these three assumptions. First, to reduce an extensive set of available hydrological signatures we select a compact orthogonal set of information pieces using Principal Component Analysis. This advances the standard practice of semi-empirical selection of individual hydrological signatures. Second, we use functional form-assumption-free Random Forests to regionalize the selected information. This allows the traditional assumption of linear regression between catchment properties and characteristics of hydrological response to be relaxes. And third, we propose utilizing non-traditional metrics to flag-up possible model/ input errors: Bayes' Factor and a newly-proposed 'Suitability' test. This addresses the typical non-realistic assumption that model is 'perfect' and the input is noise-free. The proposed methodological developments are illustrated for the empirical challenge of flow prediction in rivers in Northern Spain.



Reducing model structural uncertainty in predictions for ungauged basins via Bayesian approach.

Cristina Prieto (1,2), Nataliya Le Vine (2), Claudia Vitolo (2,3), Eduardo García (1), and Raúl Medina (1) (1) Environmental Hydraulics Institute "IH Cantabria". University of Cantabria, Santander, Spain (prietoc@unican.es; c.prieto13@imperial.ac.uk), (2) Department of Civil and Environmental Engineering, Imperial College London, (3) Institute of Environment, Health and Society, Brunel University London

A catchment is a complex system where a multitude of interrelated energy, water and vegetation processes occur at different temporal and spatial scales. A rainfall-runoff model is a simplified representation of the system, and serves as a hypothesis about an inner catchment working. In predictions for ungauged basins, a common practice is to use a pre-selected assumed-to-be-perfect model structure to represent all catchments under analysis. However, it is unlikely that the same model structure is appropriate for diverse catchments due to the 'uniqueness of the place'. At the same time, there is no obvious justification to select a single model structure as a suitable description of the system.

The contribution of this research is a move forward in the 'one size fits all' problem for predicting flows in ungauged basins. We present a statistical methodology, which allows regionalization that considers the information given by different hydrological model structures. First, the information to be regionalised is compactly represented via Principal Component Analysis. Second, the most significant principal components are regionalised using non-linear regionalisation method based on Random Forests. Third, a regionalisation error structure is derived based on the gauged catchments to be used in the Bayesian condition of the rainfall-runoff structures and their parameters. The methodological developments are demonstrated for predicting flows in ungauged basins of Northern Spain; and the results show that the methodology allows improving the flow prediction.

2016 AGU Fall Meeting



H11A-1282: Dominant hydrological process identification for ungauged basins: Bayesian approach

Monday, 12 December 2016 08:00 - 12:20 Q Moscone South - Poster Hall



ePoster - <u>Agu2016_prietoCetAleposter.pdf</u>

The research presents a methodology to address the challenge of dominant hydrologic process identification in ungauged basins. The methodology allows constraining an ensemble of hydrological models using regionalised information that summarises general hydrological behaviour for a catchment. First, the information contained in the plethora of available hydrological indices (signatures) is efficiently summarised, using the Principal Component Analysis. Second, the information is regionalised using the functional-form-assumption-free Random Forests technique, which also provides quantification of the prediction error. Third, the dominant processes are identified in ungauged catchments given the regionalised hydrological signatures via a Bayesian approach .

The method is demonstrated for ungauged basins of Northern Spain with diverse hydrological regimes. The results show that a successful identification of dominant processes can be linked to 1) the size of model error, 2) signature regionalisation quality, 3) information content in regionalised signatures, and 4) hydrological process sensitivity to information available. Further, we show that high quality hydrological predictions can be achieved with only a few pieces of information the dominant processes are the most sensitive to.

Authors

Cristina Prieto *

Environmental Hydraulics Institute "IH Cantabria"

Imperial College London

Nataliya Le Vine Imperial College London

Claudia Vitolo

European Centre for Mediumrange Weather Forecasts (ECMWF)

Imperial College London

<u>Nans Addor</u> National Center for Atmospheric Research

Raul Medina Environmental Hydraulics Institute "IH Cantabria"

Find Similar

View Related Events

Day: Monday, 12 December 2016



H23H-1680: On the information content of hydrological signatures and their relationship to catchment attributes

Tuesday, 13 December 2016 13:40 - 18:00 *Q* Moscone South - Poster Hall

Hydrological signatures, which are indices characterizing hydrologic behavior, are increasingly used for the evaluation, calibration and selection of hydrological models. Their key advantage is to provide more direct insights into specific hydrological processes than aggregated metrics (e.g., the Nash-Sutcliffe efficiency). A plethora of signatures now exists, which enable characterizing a variety of hydrograph features, but also makes the selection of signatures for new studies challenging.

Here we claim that the selection of signatures should be based on their information content, which we estimated using several approaches, all leading to similar conclusions. To explore the relationship between hydrological signatures and catchment attributes, we used a previously published data set of 671 catchments in the contiguous United States, that we expanded by characterizing the climatic conditions, topography, soil and vegetation of each catchment. We then used a data-learning algorithm (random forests) to explore whether hydrological signatures could be inferred from catchment attributes alone. We find that some signatures can be predicted remarkably well by random forests and, interestingly, the same signatures are well captured when simulating discharge using a conceptual hydrological model. We discuss what this result reveals about our understanding of hydrological processes shaping hydrological signatures. We also identify which catchment attributes exert the strongest control on catchment behavior, in particular during extreme hydrological events. Overall, climatic attributes have the most significant influence, and strongly condition how well hydrological signatures can be predicted by random forests and simulated by the hydrological model. In contrast, soil characteristics at the catchment scale are not found to be significant predictors by random forests, which raises questions on how to best use soil data for hydrological modeling, for instance for parameter estimation. We conclude with a ranking of signatures based on their information content, and a discussion of the implications of our findings for model calibration, model selection and understanding hydrologic similarity.

Authors

<u>Nans Addor *</u> National Center for Atmospheric Research

<u>Martyn P Clark</u> National Center for Atmospheric Research

Cristina Prieto University of Cantabria

Imperial College London

<u>Andrew James Newman</u> National Center for Atmospheric Research

Find Similar

Naoki Mizukami National Center for Atmospheric Research

<u>Grey S Nearing</u> National Center for Atmospheric Research

NASA Goddard Space Flight Center

Nataliya Le Vine Imperial College London

Geophysical Research Abstracts Vol. 19, EGU2017-13157, 2017 EGU General Assembly 2017 © Author(s) 2017. CC Attribution 3.0 License.



Addressing model structural uncertainty in PUBs via Bayesian approach

Cristina Prieto (1,2,3), Nataliya Le-Vine (1), Claudia Vitolo (4), and Raúl Medina (3)

(1) Imperial College London, United Kingdom (c.prieto13@alumni.imperial.ac.uk), (2) University of Bristol, Civil Engineering, Bristol, United Kingdom:, (3) Environmental Hydraulics Institute "IHCantabria". University of Cantabria, Santander, Spain, (4) European Centre for Medium-range Weather Forecasts, United Kingdom.

A catchment is a complex system where a multitude of interrelated energy, water and vegetation processes occur at different temporal and spatial scales. A rainfall-runoff model is a simplified representation of the system, and serves as a hypothesis about catchment inner working. In predictions for ungauged basins, a common practice is to use a pre-selected model structure for a catchment, while there is usually no justification for its suitability (due to the lack of observed flows). This research aims moving beyond the 'one size fits all' problem. First, two metrics are proposed to assess suitability and adequacy of a selected model based on a) how well the model reproduces regionalised information, b) knowledge gain from considering the model over what is known from regionalised information via Bayesian approach. And third, available model structures are ranked and weighed based on their skill to support regionalised information, and then used in a multi-model ensemble to provide probabilistic predictions. The methodology is applied to basins in Northern Spain with varied hydroclimatological regimes. The results show that prediction quality is sensitive to model (or ensemble) error, quality of regionalised information content.



On the information content of hydrological signatures and their relationship to catchment attributes

Nans Addor (1), Martyn P. Clark (1), Cristina Prieto (2,3), Andrew J. Newman (1), Naoki Mizukami (1), Grey Nearing (1), and Nataliya Le Vine (3)

 National Center for Atmospheric Research, Research Applications Laboratory, Boulder, United States (naddor@ucar.edu),
 Environmental Hydraulics Institute "IH Cantabria", University of Cantabria, Santander, Spain, (3) Department of Civil and Environmental Engineering, Imperial College, London, UK

Hydrological signatures, which are indices characterizing hydrologic behavior, are increasingly used for the evaluation, calibration and selection of hydrological models. Their key advantage is to provide more direct insights into specific hydrological processes than aggregated metrics (e.g., the Nash-Sutcliffe efficiency). A plethora of signatures now exists, which enable characterizing a variety of hydrograph features, but also makes the selection of signatures for new studies challenging. Here we propose that the selection of signatures should be based on their information content, which we estimated using several approaches, all leading to similar conclusions. To explore the relationship between hydrological signatures and the landscape, we extended a previously published data set of hydrometeorological time series for 671 catchments in the contiguous United States, by characterizing the climatic conditions, topography, soil, vegetation and stream network of each catchment. This new catchment attributes data set will soon be in open access, and we are looking forward to introducing it to the community.

We used this data set in a data-learning algorithm (random forests) to explore whether hydrological signatures could be inferred from catchment attributes alone. We find that some signatures can be predicted remarkably well by random forests and, interestingly, the same signatures are well captured when simulating discharge using a conceptual hydrological model. We discuss what this result reveals about our understanding of hydrological processes shaping hydrological signatures. We also identify which catchment attributes exert the strongest control on catchment behavior, in particular during extreme hydrological events. Overall, climatic attributes have the most significant influence, and strongly condition how well hydrological signatures can be predicted by random forests and simulated by the hydrological model. In contrast, soil characteristics at the catchment scale are not found to be significant predictors by random forests, which raises questions on how to best use soil data for hydrological modeling, for instance for parameter estimation. We finally demonstrate that signatures with high spatial variability are poorly captured by random forests and model simulations, which makes their regionalization delicate. We conclude with a ranking of signatures based on their information content, and propose that the signatures with high information content are best suited for model calibration, model selection and understanding hydrologic similarity.

PROPUESTA DE UN MODELO HIDROLÓGICO AGREGADO BASADO EN ECUACIONES GENERALES DE CRECIMIENTO. APLICACIÓN A LA CUENCA DEL RÍO TEIFI (UK).

Tema .B (primera opción), Tema M (segunda opción).

Cristina Prieto, Eduardo García, Roberto Mínguez, César Álvarez, Raúl Medina

Instituto de Hidráulica Ambiental "IH Cantabria". Universidad de Cantabria.

<u>prietoc@unican.es</u>

Los modelos hidrológicos son hoy en día herramientas muy usadas con diversos fines, como extender las series de caudales en cuencas sin aforos, diseñar frente a avenidas, evaluar estrategias de gestión, predecir la respuesta de las cuencas ante distintos tipos de clima, la calidad del agua, o los parámetros ecológicos, analizar el impacto del cambio climático, caracterizar hábitats, etc. La mayoría de modelos usados con fines prácticos para estimar caudales en cuencas donde se dispone de datos históricos de aforos, son de tipo agregado conceptual, recurriéndose a técnicas de regionalización en aquellos lugares donde no hay medidas [1]. No obstante, la predicción fiable de caudales en cuencas sin instrumentar, sigue siendo una cuestión aún no resuelta de forma satisfactoria [2]. Un problema para desarrollar este tipo de modelos es la dificultad de determinar su estructura y parámetros. En este artículo se resume parte de un trabajo en el que se ha explorado el potencial de las ecuaciones generales de crecimiento [3], como base para construir un modelo hidrológico agregado, conceptual, parsimonioso [4].

Aplicando homologías entre el crecimiento de un sistema genérico y la evolución del caudal en una cuenca, y adoptando otras hipótesis complementarias, se llega a una estructura compacta de 4 parámetros, de los que 1 puede fijarse, sin apenas pérdida de capacidad predictiva. El modelo asume que las cuencas, en condiciones de precipitación, temperatura y usos del suelo constantes, tienden a un caudal de equilibrio (Qeq) que depende de un coeficiente de escorrentía de equilibrio (ceq). Este coeficiente se calcula mediante una extensión de las expresiones del índice de aridez [5], dada por un índice de aridez dinámico, obtenido introduciendo un factor de memoria (λ), que permite tener en cuenta tanto la historia de evapotranspiración potencial y precipitación previas, como los rasgos específicos de la cuenca, por medio de un parámetro P1.

En una cuenca que drena a una sección de cauce, la variable cuya evolución se desea predecir es el caudal instantáneo (Q) en dicha sección, expresado como caudal específico, con unidades de velocidad (mm/día en este trabajo). Los procesos de transformación hidrológica estarán gobernados por unos parámetros fijos y, al menos, dos variables independientes: la precipitación y la evapotranspiración potencial (o, en su defecto, la temperatura máxima diaria). La variable de estado es Q, y el crecimiento del sistema equivale a su variación en el tiempo, dado por el producto de una función no acotada y otra que incluya un factor limitante; estas funciones pueden adoptar formas muy diversas, aunque aquí emplearemos la más básica, la de Verhulst [6]. Partiendo de la analogía entre la capacidad de carga de una cuenca y su Qeq, se propone la siguiente ecuación para formular el nuevo modelo hidrológico conceptual:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = \mathbf{A} \cdot \mathbf{Q}(t) \cdot \left(\mathbf{Q}_{\mathrm{eq}}(t - \tau) - \mathbf{Q}(t) \right)$$

El parámetro **A** simboliza la tasa de respuesta de la cuenca frente al desequilibrio instantáneo impuesto por las variaciones del Qeq, tiene dimensiones del inverso de la longitud (mm⁻¹ en este artículo) y representa un volumen de almacenamiento equivalente. La solución analítica de la ec.1 es la base del esquema numérico empleado y sus parámetros son λ , **P1**, **A** y τ . De estos 4 parámetros, λ puede quedar fijo a efectos prácticos, con lo que la estructura básica tiene 3 parámetros libres. Una variante de este modelo, que también se ha analizado, consistía en suponer que A es diferente cuando el caudal aumenta o disminuye, lo que añadía un parámetro más.

Para probar la aptitud de la estructura se analizó una cuenca húmeda de UK, la del río Teifi (Gales), previamente estudiada ([1], [7], [8]); y sus resultados se contrastaron con los del modelo IHACRES [9], que es un modelo hidrológico agregado, conceptual, parsimonioso, muy usado para simular caudales en continuo. Usando como función objetivo la media entre la unidad menos el coeficiente de eficiencia de Nash-Sutcliffe (NS) y la desviación en volumen (Bias), se comprobó que la estructura fundamental de 3 parámetros reproducía adecuadamente el comportamiento de dicha cuenca, mejorando el obtenido con el IHACRES. Además, los estimadores resultaron estables frente a diferentes periodos de calibración y evaluación. En un análisis de sensibilidad se demostró que ya liberar λ ; ya introducir un parámetro adicional, diferenciando una A de ascenso de una de descenso, apenas mejoraba los resultados. En la cuenca del río Teifi se dispone de un total de 11 años, por consistencia con [1], [7] y [8], y se calibró con sub-periodos de 3 (8 casos, #1 al #8), 10 (#1-8), 1 (#X) y 8 (#1-6) años; todos dieron buenos ajustes de NS, entre 0.905 (#2) y 0.954 (#8). Para evaluar el modelo, cada conjunto de parámetros obtenido en la calibración de los 8 intervalos de 3 años se empleó en los 7 restantes, el menor NS fue 0.886 (min Ev_TEIFI) y el mayor 0.95 (Max Ev_TEIFI). En la imagen de la izquierda se muestran estos resultados y la comparación con el IHACRES. El Bias fue nulo siempre, al aplicar la función objetivo señalada. En la figura de la derecha se ven los caudales estimados con el #8 en un fragmento de su calibración. De la inspección visual de todos los ajustes, se concluyó que el modelo reproducía bien las pautas de la señal medida, en magnitud y tiempo, pero sobreestima las curvas de recesión asociadas a caudales menores de 2 mm/día. La mayor divergencia entre las distribuciones acumuladas de caudales observados y estimados, se dio en la parte baja y fue de un 6% respecto al medio.



El nuevo modelo tiene como propiedades más relevantes el ser compacto, parsimonioso y bien condicionado para la optimización. La estructura se presenta bajo una única ecuación diferencial no lineal y, aunque se aplicó a datos diarios, se puede aplicar con otros incrementos de tiempo. A diferencia de en los modelos tipo al IHACRES, la ecuación de Verhulst genera una rama ascendente en la respuesta de la cuenca, y además el retardo aparece explícitamente (τ), pudiendo adoptar un valor no entero. El análisis visual de los resultados indujo a pensar que el modelo de 3 parámetros refleja una respuesta hidrológica basada en un mecanismo de exceso de saturación; asimismo, hay que señalar que no permite alcanzar un caudal nulo, lo que, en su versión actual, lo hace inadecuado para ríos efímeros de zonas áridas o semiáridas. Por último, si aceptamos la hipótesis de que un sistema hidrológico se rige por una determinada ecuación diferencial a escala infinitesimal, es más fácil una caracterización de los sistemas hidrológicos independiente de la escala de tiempo de análisis, con relaciones conocidas entre los parámetros óptimos y la base de referencia para la que se han obtenido, abriendo las posibilidades de regionalizar modelos y clasificar cuencas de forma más robusta que las propuestas hasta la fecha. En definitiva, se ha comprobado el potencial de la aplicación, en el ámbito hidrológico, de las ecuaciones generales de crecimiento y, en particular, la ecuación logística o de Verhulst [6]. En este marco de trabajo, es clave el concepto de caudal de equilibrio, definido como aquél que la cuenca persigue continuamente sin llegar a alcanzar, y cuyo valor depende de la propia cuenca, de la lluvia instantánea y de la historia meteorológica reciente, expresada mediante un índice de aridez dinámico.

BIBLIOGRAFÍA:[1] Littlewood, I.G., 2003. Improved UI identification for 7 Welsh rivers: implications for estimating continuous streamflow at ungauged sites. Hydrol. Sci. J.48 (5): 743-762.[2] Wagener, T., & A.Montanari (2011). Convergence of approaches toward reducing uncertainty in PUB. Water Resour. Res., 47[3]Tsoularis, A.&Wallace, J., 2002. Analysis of logistic growth models. Math. Biosci.179(1):21-55.[4]Wagener, T., Wheater, H.S.&Gupta, H.V.2004. RRM in gauged and ungauged catchments. Imperial College Press, London, UK, 332pp. . [5] Arora, V. K., 2002. The use of the aridity index to assess climate change effect on annual runoff. J.Hydrol.265:164-177 [6]Verhulst, P.F., 1838.Notice sur la loi que la population suit dans son accroissement. Correspondance Mathematique et Physique Publiee par A. Quetelet 10:113-12. [7] Littlewood, I.G. 2001. Practical aspects of calibrating and selecting UH-based models for continuous river flow simulation. Hydrol. Sci. J. 46(5): 795-811.[8]Littlewood, I.G., 2002b.Improved UH characterisation of the daily flow regime (including low flows) of the River Teifi, Wales: towards better rainfall–streamflow models for regionalisation. Hydrol. Earth Systems Sci. 6(5), 899–911. [9]Jakeman, A.J., Littlewood, I.G.&Whitehead, P. G., 1990.Computation of the IUH and identifiable component flows with application to two small upland catchments. J.Hydrol.117:275-300.

Aplicación de un modelo logístico triparamétrico a la estimación de caudales diarios en la cuenca del río Nam Ngum (Laos)

(Tema B:Hidrología y gestión del agua)

César Álvarez Díaz

Profesor Titular de Universidad. Universidad de Cantabria

alvarezc@unican.es

Eduardo García Alonso

Doctor Ingeniero de Camino. Instituto e Hidráulica Ambienta. Universidad de Cantabria

edelwar@gmail.com

Cristina Prieto Sierra

Ingeniera de Camino. Instituto de Hidráulica Ambiental de la Universidad de Cantabria

prietoc@unican.es

Jorge Rojo Gómez

Master en Ingeniería Ambienta. Instituto de Hidráulica Ambiental de la Universidad de Cantabria

rojoj@unican.es

Beatriz Tejerina Vega

Master en Ingeniería de Costas y Puertos. Instituto de Hidráulica Ambiental de la Universidad de Cantabria

tejerinab@unican.es

La cuenca del río Mekong a su paso por Laos va a sufrir profundas transformaciones en los próximos años debido a la construcción de multitud de embalses y centrales para generación de energía hidroeléctrica. La elevada pluviometría de la zona y su situación estratégica, rodeado de países de fuerte crecimiento económico, hacen que la producción eléctrica constituya uno de los campos prioritarios para el desarrollo de la República Popular de Laos.

La planificación y diseño de estas infraestructuras hidráulicas exige un conocimiento detallado del régimen de caudales en los puntos de toma de las centrales, tanto por lo que esto supone a la hora de optimizar la producción, como para conocer el impacto ambiental de estas actuaciones en los ríos intervenidos.

El río Nam Ngum es uno de los afluentes del Mekong en los que están previstas mayores inversiones en este campo. Con una superficie de cuenca de unos 20.000 km², la información hidrometeorológica disponible es escasa dado su gran extensión, su reducida densidad de población y el bajo nivel de presión actual sobre sus recursos hídricos.

Para paliar esta falta de información, diversas instituciones internacionales (Banco Mundial, Banco Asiático de Desarrollo, organismos de cooperación internacional, etc.) han financiado varios estudios para la aplicación de modelos hidrológicos para el cálculo de caudales en los puntos de toma de las futuras centrales. La calibración y validación de estos modelos se ha realizado con base en la información foronómica disponible, utilizando, principalmente, modelos hidrológicos como el Hec Hms, del Cuerpo de Ingenieros del Ejército de los Estados Unidos y otros expresamente desarrollados para estas cuencas. Todos estos modelos se caracterizan por la gran cantidad de parámetros que incorporan en la resolución de las diferentes ecuaciones con los que representan los procesos hidrológicos. Pese a que con ellos se ha conseguido la estimación de caudales a nivel diario con un

aceptable grado de ajuste (figura 1), la sobreparametrización de este tipo de herramientas complica el propio proceso de calibración y puede restar validez a la extrapolación de los resultados de los mismos.



Figura 1. Caudales observados y calculados con el modelo HecHms en un punto de la cuenca del río NamNgu.

En este artículo se presenta la calibración de un modelo logístico (LEM) en esta cuenca y la comparación de los resultados obtenidos con el modelo HecHms. Los resultados de dicho análisis muestran un adecuado ajuste a los con los datos medidos, similar al obtenido con el modelos de SMA del HecHms (figura 2), lo que le convierte en una herramienta muy potente para el análisis de series históricas de precipitaciones, dada el reducido número de parámetros de los que depende.



Figura 2. Comparación de los coeficientes de Nasch-Shutcliff correspondientes a los resultados de calibración del modelo HecHms y LEM.

Referencias bibliográficas

N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, T. L. Veith, 2007.Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, Vol. 50(3): 885–900.

D. Roy, S. Begam, S. Ghosh and S. Jana, 2013. Calibration and validation of hec-hms model for a river basin in eastern India. ARPN Journal of Engineering and Applied Sciences, VOL. 8, NO. 1.

J. American Water Resources Assoc Arnold, J. G., R. Srinivisan, R. S. Muttiah, and P. M. Allen, 2008. Largearea hydrologic modeling and assessment: Part I. Model development. Journal of the American Water Resources Association, Vol. 34, No.1.

Refsgaard, J. C, 1997. Parameterisation, calibration, and validation of distributed hydrological models. Journal of Hydrology, Vol. 198, Issues 1–4, 69–97.

Certificates

Research Stays and Collaborations with International Research Institutions

Imperial College London

Registry Suite 3, Sherfield Building South Kensington London SW7 2AZ Tel: +44 20 7594 7268 records@imperial.ac.uk

Dean Surtees Senior Assistant Registrar

5 June 2014

Ref: 00903210

TO WHOM IT MAY CONCERN

THIS IS TO CERTIFY THAT THE UNDERNAMED:

Surname:	Prieto
Forenames:	Cristina
Date of Birth:	13 June 1982

Is currently registered on a Full Time Postgraduate course for the academic year 2013/14, details as follows:

Course: Year: Start Date: Expected End Date: Civil Engineering Research (OCC) Year 2 28 September 2013 30 April 2014







Graduate Education Director, Engineering Professor Majid Mirmehdi

Graduate Education Manager, Engineering Ms Gina Stuart Graduate Education Team, Engineering University of Bristol Queens' Building University Walk Clifton Bristol BS8 1TR Tel: +44 (0)117 331 5621 E-mail: gina.stuart@bristol.ac.uk

19 May 2015

TO WHOM IT MAY CONCERN

RE: Miss Cristina Prieto (1458368)

I confirm that Cristina Prieto (date of birth 13 June 1982) was registered as a Full Time postgraduate visiting research student in the Department of Civil Engineering at the University of Bristol. During her visit she carried out research with Professor Thorsten Wagener on prediction flows in ungauged basins.

Miss Prieto commenced her research visit on 1 May 2014 and completed her visit on 30 November 2014.

Yours faithfully

GL Shear

Gina Stuart \bigcirc Graduate Education Manager, Engineering



International Expert Reviewers of the PhD Thesis


National Center for Atmospheric Research

Research Applications Laboratory

3450 Mitchell Lane, Boulder, CO 80301 USA Phone: 303-497-8422 Fax: 303-497-8401 www.ral.ucar.edu

20 March 2017

I am writing to provide a brief review of the PhD Thesis of Cristina Prieto from an international perspective. Ms. Prieto's thesis is titled "Development of methodologies to disaggregate, quantify and reduce uncertainty in Predictions in Ungauged Basins (PUBs)" and was developed under the supervision of Dr. Raúl Santanaría, Dr. Eduardo Alonso, and Dr. Nataliya Le Vine at the Universidad De Cantabria.

It may be useful if I provide some brief background on my capability to judge Ms. Prieto's thesis. I am a Senior Scientist at the National Center for Atmospheric Research where my research focus is on developing process-based hydrologic models. I have a PhD from the University of Colorado (1998), and I have over 18 years of research experience in academic organizations. Currently I lead a vibrant research group in Computational Hydrology. I am widely published (over 130 peer-reviewed journal articles) and I regularly present my research in national and international venues. I was recently elected Fellow of the American Geophysical Union, an honor limited to one out of one thousand members in any given year, and I am the incoming Editor-in-Chief of the leading hydrology journal *Water Resources Research*. I write as someone with a strong reputation in the field of computational hydrology and who understands scholarly contributions to this field.

The general objective of Ms. Prieto's thesis is to advance flow predictions in ungauged basins. This is accomplished by developing new methodologies to identify, quantify and reduce uncertainties in process-based model simulations of hydrologic processes. The main contributions are:

- <u>Chapter 1</u>. Ms. Prieto's insightful discussion on the different ways to quantify model error could be published as a very nice commentary in a leading hydrology journal. Ms. Prieto clearly defines the problem, including the underlying philosophical issues, and she clearly characterizes the capability of competing methods to quantify uncertainty. Ms. Prieto provides a comprehensive and thoughtful critique of existing work, which, in my opinion, goes well beyond the critiques published by the leaders of our field. This first chapter provides good motivation, not only for Ms. Prieto's thesis, but also for the active community of hydrologists who strive to quantify uncertainty in hydrologic models.
- <u>Chapter 2</u>. Ms. Prieto has substantially advanced the methods for Predictions in Ungauged Basins (PUB). Historically, the dominant approach to the PUB problem was the method of parameter transfer: i.e., calibrate models for individual gauged basins, find empirical relations between calibrated parameters and geophysical attributes, and use these empirical relations to transfer parameters to ungauged basins. The parameter transfer method was largely unsuccessful both because calibrated parameter sets reflect the unique (non-predictable) character of individual basins and because of strong parameter

interactions. Such problems mean that relations between calibrated parameters and geophysical attributes are typically rather weak. Ms. Prieto has been exploring a relatively new approach, where, instead of regionalizing model parameters, Ms. Prieto regionalizes hydrologic signatures, and uses the regionalized hydrologic signatures to constrain model parameters.

Ms. Prieto's main advance is her combination of sophisticated regionalization methods (regression forests) with an advanced Bayesian inference scheme, providing a new level of sophistication in both the regionalization and inference steps. A remarkable aspect of Ms. Prieto's work is her development of an adequacy metric, which provides the capability to separate the knowledge gained from the regionalization and from the hydrologic model. I fully expect that this work, once published, will set the new standard in PUB methods, and the hydrologic modeling community will strive to improve upon Ms. Prieto's research.

- <u>Chapter 3</u>. In this chapter Ms. Prieto extends her work on the PUB problem to multiple hydrologic models, providing more insights on the role of model errors in PUB studies. A key contribution here is development of methods to infer the dominant hydrologic processes in ungauged catchments, providing the community with a tool to disentangle results from multiple working hypotheses. In my opinion it could be useful to publish two separate papers from this chapter: The first paper could describe the general method to infer dominant hydrologic processes, and the second paper could describe the application of the method to the PUB problem.
- <u>Chapter 4</u>. In her concluding chapter Ms. Prieto begins to address the problem of disinformation in data. This work shows considerable promise, and I expect that, once complete, Ms. Prieto will round out her contributions to the PUB problem through a comprehensive analysis of data errors.

It has been a pleasure to read and review this extraordinary thesis. The work here substantially advances capabilities for predictions in ungauged basins, to the extent that Ms. Prieto's work will set a new standard for others working in this field. The work presented in Ms. Prieto's thesis is at a very high international standard. Please know that I am always available and delighted to answer any questions you may have, please feel free to reach me via email or phone.

Yours Sincerely,

Dr. Martyn P. Clark Senior Scientist, Research Applications Laboratory, National Center for Atmospheric Research (NCAR) Editor in Chief, Water Resources Research (incoming) Fellow of the American Geophysical Union Email: mclark@ucar.edu



Prof. Dawei HAN Professor of Hydroinformatics Department of Civil Engineering, University of Bristol, BS8 1TR, UK Tel: +44 117 3315739 E-mail: d.han@bristol.ac.uk

8 March 2017

International Perspective Review of Ms Cristina Prieto Sierra's PhD Thesis

Ms Cristina Prieto Sierra's thesis aims to advance flow prediction capabilities for ungauged basins by developing methodologies for identifying, quantifying, and reducing the uncertainty associated with the various related sources of errors. Her thesis moves towards overcoming some of the major issues in traditional hydrological modelling. Her topic is very important in hydrology and at the cutting edge of the research field. The whole thesis contains four substantial chapters.

Chapter 1 describes the motivation for the PhD project. It provides an overview of the problem statement and analyses current regionalization approaches for ungauged basins, the most common sources of uncertainty in ungauged catchments, and probabilistic methods for quantifying uncertainty, with a particular focus on Bayesian approaches. The candidate has carried out an excellent literature review and provided her insightful analysis on the existing research. The writing is comprehensive and of excellent quality.

Chapter 2 focuses on the impact of the decisions to be made when regionalized information is used for model parameter estimation. The study merges new developments about indices selection and regionalization methodologies with the Bayesian prediction procedure. Two statistical metrics are adopted for assessing suitability and adequacy using Bayesian statistics. The study explores the impact of the predictions quality to the choice of the regionalization procedure and hydrological model. The research has been done thoroughly and logically with academic rigor. The work is novel and publishable (already written in the format of an academic article).

In Chapter 3, using formal Bayesian probability theory, an ensemble of hydrological models, and regionalized information, it presents a statistical test for determining the dominant mechanisms in ungauged catchments. The study analyzes the sensitivity of the mechanism identification procedure on the quality of the regionalization procedure, the model ensemble error, and the quantity of information contained in the selected regionalized indices. This chapter is again novel and publishable.

Chapter 4 describes how the thesis contributes to the current state of the art, offers general conclusions, and provides recommendations for future research. It is well written and of excellent quality.

The candidate has presented her work in many key international conferences, and has spent significant times in several top research institutions in the world. Therefore, she has made strong connections with leading academics in the field, and is well honed in her communication skills. Her thesis has made novel contributions to the existing knowledge gaps in hydrology. In conclusion, I believe that there is enough material for her to defend as a PhD candidate from an international point of view.

If any further information is needed, please don't hesitate to contact me.

Yours Sincerely

D.Han

Prof. Dawei HAN