UNIVERSIDAD DE CÓRDOBA

TESIS DOCTORAL

Estudio comparativo completo de varios métodos basados en datos para la gestión de los recursos hídricos en ambientes mediterráneos a través de diferentes escalas temporales.

Comprehensive comparison of several data-driven methods for Mediterranean water resources management across diferent temporal scales.

Doctorando: Gulliver Acevedo, Zacarias
Directores: Herrero Lantarón, Javier; Polo Gómez, María José

Programa de Doctorado en Dinámica de los Flujos Biogeoquímicos y sus Aplicaciones

10 febrero del 2021

TITULO: *Comprehensive comparison of several data-driven methods for Mediterranean water resources management across diferent temporal scales*

AUTOR: *Zacarías Gulliver Acevedo*

Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

TÍTULO DE LA TESIS: *Comprehensive comparison of several data-driven methods for Mediterranean water resources management across diferent temporal scales.*

DOCTORANDO/A: *Zacarías Gulliver Acevedo*

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

Javier Herrero Lantarón y María José Polo Gómez, contratado con cargo a proyecto y catedrática de Ingeniería Hidráulica del Departamento de Agronomía, respectivamente, ambos de la Universidad de Córdoba, como directores de la tesis doctoral del alumno del Programa de Doctorado “Dinámica de Flujos Biogeoquímicos y sus Aplicaciones” Zacarías Gulliver Acevedo

INFORMAN

Que el doctorando ha cubierto y en cierta forma extendido los objetivos iniciales propuestos en la tesis, aplicando y analizando diversas metodologías basadas en datos para hidrología en zonas Mediterráneas a distintas escalas y horizontes temporales. De manera que, tras la tesis y como conclusión de ésta, se dispone de una amplia revisión, análisis y metodología de aplicación de diversos métodos, como las Redes Bayesianas Neuronales, los Procesos Gaussianos, o las Regresiones Lineales Múltiples, para su uso en hidrología. Y lo más importante, una gran cantidad de ejemplos de su aplicación a distintas escalas temporales (hora, día y mes) para diversos problemas de pronóstico hidrológico concretos (sequía, inundación, caudal estacional) en dos cuencas Mediterráneas (Guadalfeo y Guadalhorce) con distintas respuestas hidrológicas por sus diferencias en geografía, incidencia de borrascas, exposición a fenómenos de lluvias torrenciales (DANA) y presencia e importancia de la nieve en el ciclo hidrológico de cada cuenca. El interés particular de este trabajo reside en la aplicación de metodologías basadas en datos en zonas mediterráneas, donde precisamente las series de datos no suelen ser muy extensas, o son incompletas por las dificultades de medir caudales en zonas torrenciales semiáridas. El doctorando propone diversas soluciones para superar estas limitaciones, como la de incluir series de datos procedentes de modelado físico o el uso de índices asociados a oscilaciones atmosféricas como la NAO, WEMO o distintos MOI.

Uno de los puntos clave para la aplicación de los métodos expuestos es la selección de Variables de Entrada IVS, que es discutida con profusión a lo largo de todos los ejemplos incluidos en la tesis, lo cual es de gran interés para futuras aplicaciones de estos métodos para problemas de pronóstico hidrológico en zonas Mediterráneas semejantes. La tesis también analiza el éxito (o no éxito) de otras metodologías, comúnmente asociadas al término Machine Learning, en temas hidrológicos, relacionadas con los problemas de clasificación, las tendencias de aprendizaje de los modelos, o el análisis y visualización de los resultados.

Esta tesis, al poner cierto orden en qué modelos pueden ser más útiles desde el punto de vista hidrológico en un mundo de Modelos Basados en Datos en clara expansión, donde existe tal profusión de métodos y metodologías, marca un camino para futuros trabajos en este sentido. El trabajo también tiene un enfoque final eminentemente práctico que permite vislumbrar una fácil transferencia de los resultados alcanzados a ámbitos privados o públicos. Se han generado rutinas y modelos que podrían ser incorporados en los Sistemas de Ayuda a la Decisión de los organismos encargados de la gestión del agua o su aprovechamiento (energía, suministro, riego o regulación) o en los Sistemas de Alerta de los servicios hidro-meteorológicos.

Cada uno de los capítulos recogidos en este trabajo ha formado parte de una aportación a sendos congresos indexados por revisión por pares, dos de ellos indexados en el JCR. El trabajo también ha sido evaluado por dos revisores extranjeros de alto prestigio que han avalado la calidad de la tesis y han certificado que cumple los requisitos para ser presentada.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 15 de enero de 2021

Firma de los directores

Javier Herrero

Firmado digitalmente por Javier Herrero
Nombre de reconocimiento (DN): cn=Javier Herrero, o, ou, email=javier.herrero@gmail.com, c=ES
Fecha: 2021.01.15 17:30:51 +01'00'

Fdo.:Javier Herrero Lantarón

POLO GOMEZ MARIA JOSE - 30534116Y

Firmado digitalmente por POLO GOMEZ MARIA JOSE - 30534116Y
Nombre de reconocimiento (DN): c=ES, serialNumber=IDCES-30534116Y, givenName=MARIA JOSE, sn=POLO GOMEZ, cn=POLO GOMEZ MARIA JOSE - 30534116Y
Fecha: 2021.01.19 10:14:59 +01'00'

Fdo.: María José Polo Gómez

IISTA

Instituto Interuniversitario de Investigación del Sistema Tierra en Andalucía

University of Cordoba

Doctoral Thesis

Comprehensive comparison of several data-driven methods for Mediterranean water resources management across diferent temporal scales

Author:

Zacarías Gulliver Acevedo

Supervisors:

Dr.María José Polo Gómez

Dr.Javier Herrero Lantarón

January 2021

DFH Dinámica Fluvial e Hidrología

UNIVERSIDAD D CÓRDOBA

Declaration of Authorship

I, Zacarías Gulliver Acevedo, declare that this thesis titled, 'Comprehensive comparison of several data-driven methods for Mediterranean water resources management across diferent temporal scales' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed: _____

Date: _____

Como el río desemboca en el océano,
desemboca en lo que tú eres

∼ Jalaluddin Rumi ∼

UNIVERSITY OF CORDOBA

Abstract

Hydraulic Engineering Faculty
Department of Agronomy

Doctor of Philosophy

Comprehensive comparison of several data-driven methods for Mediterranean water resources management across diferent temporal scales

by Zacarías Gulliver Acevedo

Since the beginning of time, there has been innovation in the knowledge and technology of water and the hydraulic systems, to achieve an efficient and upgrade management of them. In this project, as an opening hypothesis, we will apply computational techniques and Artificial Intelligence concepts. Given that the primary asset of these studies is data, we have preferred to use the term "Data-Driven", as the term Artificial Intelligence can cause confusion in non-experts. This is an expanding field in all aspects of science and life, where the computing and processing powers are increasing periodic, so does the generation of information. There we have 5G technology, or the Internet of things, where the exponential build up in the volume of data utilised, pushes us to set up frameworks for the treatment and analysis of the information.

Data-Driven techniques offers enormous potential to transform our perception to understand, monitor and predict the states of hydro-meteorological variables. Its application provides benefits, however, performing these exercises requires practice and explicit knowledge. Therefore, a deeper understanding of the capabilities and limitations of novel computational techniques within our field of knowledge is needed. Hence, it is essential to carry out "hydro-informatics" experiences under this assumption. For the development of these models, we identify which points are the most relevant and need to be taken into account in regional conditions or frameworks. In consequence, we will work with the time series collected in the different monitoring networks, selecting the hydrological points of interest, in order to further develop hydrological frameworks that are useful for water management and optimisation. Here, we are interested in seeing the practical applicability to hydro-meteorology under Mediterranean conditions, where data are sometimes scarce, by selecting two hydrographic basins in south-east Andalusia: the Guadalhorce river (Málaga) and the Guadalfeo river (Granada).

In chapter 1, an introduction to the doctoral thesis is made. Likewise, we establish the general and the specific objectives, and the motivation of the thesis. Afterwards, we describe the three fundamental exercises to be carried out in the research work: Regression, Classification and Optimisation. Ultimately, we carry out a brief review of previous works under Mediterranean climatic conditions and similar assumptions.

Chapter 2 presents the study areas, analysing the spatial and temporal characteristics of two Andalusian Mediterranean basins in south-east Spain: Guadalhorce (GH) and Guadalfeo (GF). These are hydrographic basins with highly variable/heterogeneous space-time patterns. The first hydrological system, GH, contains an area of socio-economic importance, such is the city of Málaga. The second, GF, to the north has the Sierra Nevada National Park, crowned by the Mulhacén peak and flowing in a few kilometres

into the area of Motril. In this particular water system, we find large gradients of the geophysical agents. Both systems have regulation structures of great interest for the development and study of their optimisation. We also review the monitoring networks available in these basins, and which environmental agents and/or processes should be taken into account to meet the objectives of this work. We carry out a bibliographic review of the most relevant historical floods, listing the factors associated with these extreme events. In the data analysis stage of this chapter, we focus on the spatial-temporal evolution of the risk of flooding in the two mouths of the Guadalhorce and Guadalfeo Rivers into the Alborán Sea. We quantify that had stepped up in recent years, noting that dangerous practices have increased the risk of flooding because of the intrusion of land uses with high-costs. This chapter also analyses collected data within the monitoring networks, to understand the occurrence of floods in the river GH related to upstream discharges. We found that this basin has limitations in regulation and cannot mitigate costs downstream. The results got, were part of the work presented in Egüen et al. (2015). These analyses allow us to identify in which parts of the flood management of this hydrological system need a more precise optimisation. Finally, a summary of another important hydrological risk is carried out, such as droughts, and how these water deficits can be represented by standardised indices, both in rainfall and the flow rates.

The various approaches and methodologies for hydro-meteorological time series modelling are discussed in the chapter 3. The contrasting concepts are exposed antagonistically, to focus on the different design choices that we need to make: black box vs. grey box vs. white box, parametric vs. non-parametric, static vs. dynamic, linear vs. non-linear, frequency vs. Bayesian, single vs. multiple, among others..., detailing the advantages and disadvantages of each approach. We presented some ideas that emerged in this part of the research in Herrero et al. (2014). The partition, management and data transformation steps for the correct application of these experimental methods are also discussed. This is of great importance, since part of the hard work in the application of these methods comes from the transformation of the data. So that, the algorithms and transfer functions work correctly. Finally, we focus on how to test and validate the deterministic and probabilistic behaviours through evaluative coefficients to avoid coefficients that mask the results, and therefore focus on the behaviours of our interest, in our case precision and predictability. We have also taken parsimony into account in models based on neural networks, since they can easily fall into over-parameterisation.

In chapter 4, we present the experimental work, where seven short-term, six daily and one hourly rainfall-runoff regressions are performed. The case studies correspond to various

points of interest within the study areas with important implications for hydrological management. On an hourly scale, we analyse the efficiency and predictive capacities of the MLR and BNN at ten time horizons for the level of the Guadalhorce River in Cártama. We found that, for closer predictive horizons, a simpler approach such as linear (MLR) can outperform other with a priori higher capabilities, such as non-linear (BNN). This finding could simplify greatly its development and application. At a daily scale, we establish a comparative framework between the two previous models and a complete Bayesian method such as the Gaussian Processes. This DD computational technique, allows us to apply different transfer functions under a single model. This is an advantage over the other two DD models, since the results show that they work well in one domain, but do not work well in the other. During the construction of the models, we do the selection of the input variables in a progressive way, through a trial-and-error method, where the significant improvements with respect to the last predictor structure are taken into account preserving the principle of parsimony. Here, we have used different types of data: real data collected in the monitoring networks, and data generated in parallel from physically based hydrological modelling (WiMMed). The results are robust, where the major limitation is the high computational cost by the recurrent and iterative method used. Some results of this chapter, were presented in Gulliver et al. (2014).

In chapter 5 three medium-term time scale prediction experiments are performed. We base the first modelling experiment on a quarterly scale, where a hydrological time scheme determines the cumulative flow for specific time horizons. We start the scheme according to the relevant dates where hydrological planning takes place. It is validated that the forecasts are more prosperous after have been consumed the first six months of the hydrological year. Instead of the three months in which we carry out the evaluations. The observed input variables quantified in the water system are: cumulative stream flow, cumulative rainfall, cumulative snowfall values and atmospheric oscillations (AO). At the level of modelling with DD, this experience has shown the importance of combining mixed regression classification models instead of only regression models within static frameworks. In this manner, we reduce and narrow the space of possible solutions and, therefore, we optimised the predictive behaviour of the DD model. During the development of this exercise, we have also carried out a classification practice comparing three DD classifiers: *Probabilistic Neural Network* (PNN), *K-Nearest Neighbour* (KNN) and *Support Vector Machine* (SVM). We see that the SVM behaves better than the others with our data. However, more research is still needed on classifiers in hydro-meteorological frameworks like ours, because of their variability. We showed this part of the doctoral thesis in Gulliver et al. (2016).

In the second section of this chapter (Sec. 5.3), we carry out a rain forecast exercise on a monthly scale. To do so, we use BNN following the same construction method of the SVI model exposed in the previous chapter (Sec. Ref. Chapter 4), thus validating it in another time scale. However, the results in predictive terms are poor for this hydro-meteorological variable. This confirms the difficulty of predicting this variable from historical data and without the incorporation of dynamic tools. Thus, the need for complex hydrodynamic modelling for the prediction of this important variable is confirmed. On the other hand, this case serves to empirically infer the causality of the most relevant atmospheric oscillations in the points of study. From multiple simulations with the model-based approach it has been possible to establish which indices have a greater influence.

In the last section of this chapter (Section 5.4), an exercise was carried out to predict the deviation or anomaly of rainfall and runoff indices for four time series representative of different locations within the Guadalfeo BR. In this case, we verified the suitability of seven statistical distributions to characterize the anomalies/deviations under Mediterranean conditions. Under this hypothesis, the indices that passed the Shapiro-Wilk normality test were modelled to analyse the capabilities of BNN to predict these indices at various time horizons. Here, predictions of negative phases (droughts or deficit periods) have been poor, and the behaviour of the models for positive phases (wet periods) has been more successful. Regarding the causal inference of IC and its possible influence on the study area, we found out how NAO and WEMO help forecasts for shorter time horizons, while MOI helps for longer cumulative time horizons/times. We have analysed the relevance of these atmospheric variables in each case where sometimes their introduction was convenient and sometimes not, following the rules of construction and detailing them in each case study.

Throughout the work, the usefulness of mixed modelling approaches has been verified, using models based on observed data from the different monitoring networks with physical modelling for the reproduction of essential hydrological processes. With the proposed methodology, a positive influence of atmospheric oscillations has been observed for medium-term prediction within the study regions, finding no evidence for short-term predictions (daily scale).

The final conclusions and the most important points for future work are presented in the chapter 6.

Applications of this type of methods are currently necessary. They help us to establish relationships based on measured hydro-meteorological data and thus "based on real data",

without hypothesizing any assumptions. These data-based experiences are very useful for limiting future uncertainty and optimizing water resources. The establishment of temporal relationships between different environmental agents allows us, through supervised methods, to establish causal relationships. From here a physical inference exercise is necessary to add coherence and establish a robust scientific exercise.

The results obtained in this work, reaffirm the practicality of implementing this Data-Driven frameworks, in both the public and private spheres, being a good starting point for technology transfer. Most of the routines and models provided in this thesis, could be directly applied in Hydro-meteorological Services, or Decision Support Systems for water officials. This includes potential users as varied as public administrations and basin organisations, reservoir managers, energy companies that manage hydroelectric generation, irrigation communities, water bottling plants,... etc. The establishment of iterative and automatic frameworks for data processing and modelling, needs to be implemented, to make the most of the data collected in the water systems.

RESUMEN

Desde el inicio de los tiempos, se innova en el conocimiento y la tecnología de los sistemas hídricos e hidráulicos con el fin de conseguir una eficiente y correcta gestión de los mismos. En este proyecto, como hipótesis de partida, se van a aplicar diversas técnicas computacionales y conceptos de Inteligencia Artificial. Dado que el principal activo de estas aplicaciones son los datos, optamos por el término "*Data-Driven*" (DD), ya que el término de Inteligencia Artificial puede causar confusión en los no expertos. Este es un campo en expansión en todos los aspectos de la ciencia y de la vida, donde al tiempo que se incrementan las capacidades de computación y de procesamiento, se incrementa la generación de datos. Ahí tenemos la tecnología 5G, o el internet de las cosas, donde el incremento exponencial del volumen de datos que se utilizan nos obliga a desarrollar marcos para el tratamiento y el análisis de los mismos.

Los métodos DD tienen un enorme potencial para transformar nuestra habilidad de establecer un seguimiento supervisado y predecir estados de variables hidro-meteorológicas. Su aplicación provee claramente de beneficios, sin embargo realizar estos ejercicios requiere una práctica y un conocimiento específico. Por ello, es necesario un entendimiento más profundo de las capacidades y de las limitaciones de estas técnicas computacionales, dentro de nuestro campo de conocimiento y casos específicos. Por estos motivos, es esencial realizar experiencias "hidro-informáticas" bajo este supuesto, identificando así qué puntos son los más relevantes y a tener en cuenta en el desarrollo y la validación de estos modelos en condiciones o marcos más regionales. Para ello, trabajaremos con las series temporales recogidas en las diferentes redes de monitorización, con series resultantes de modelado hidro-meteorológico y con series de las oscilaciones atmosféricas más relevantes en la zona de estudio.

El objetivo principal de este trabajo, es el desarrollo y la validación de marcos metodológicos basados en datos. Para ello, se seleccionan puntos de interés, con el fin de desarrollar marcos hidro-meteorológicos útiles en la gestión y optimización de los recursos hídricos. En este supuesto, nos interesa ver la aplicabilidad práctica de estas herramientas de aprendizaje automático, *machine learning*, en condiciones mediterráneas y locales, donde los datos a veces son escasos o de baja calidad.

En el primer capítulo (Cap.1) se realiza una introducción a la tesis doctoral, estableciendo los objetivos tanto generales como específicos, y la motivación de la tesis. Seguidamente se realiza a modo introductorio una descripción de los tres ejercicios fundamentales a realizar en el trabajo de investigación: Regresión, Clasificación y Optimización. Finalmente, se

realiza una revisión del estado del arte de trabajos previos bajo condiciones climáticas mediterráneas y similares.

El capítulo 2 presenta las zonas de estudio, analizando las características espacio-temporales de dos cuencas mediterráneas andaluzas situadas en el sureste español: río Guadalhorce (GH) y río Guadalfeo (GF). Son cuencas hidrográficas con unos patrones espacio-temporales altamente variables/heterogéneos. El primer sistema hidrológico, GH, contiene una zona de gran importancia socio-económica como es la ciudad de Málaga. El segundo, GF, al norte tiene situado el Parque Nacional de Sierra Nevada, coronado por el pico Mulhacén y desemboca a pocos kilómetros en la costa de Motril. Esto hace que este sea un sistema con grandes gradientes geo-morfológicos e hidro-meteorológicos. En ambas cuencas existen estructuras de regulación de gran interés para el desarrollo y estudio de su optimización. También se revisan las redes de monitorización disponibles en estas cuencas, y que agentes deben ser tenidos en cuenta para la consecución de los objetivos del presente trabajo. En la etapa de análisis de datos de este capítulo, nos centramos en la evolución espacio temporal del riesgo frente a las inundaciones en las desembocaduras de ambos sistemas hidrológicos al mar de Alborán. Se cuantifica el aumento del riesgo frente a inundaciones ante la intrusión de usos del suelo con altos costes en las zonas potencialmente inundables en estos últimos años, constatando así una mala práctica en la planificación del territorio dentro de la zona de estudio. También, en este capítulo se analizan los datos registrados con el fin de comprender la ocurrencia de avenidas en el río GH y su relación con los desembalses aguas arriba. En este análisis se pudo identificar, como ante algunos eventos pluviométricos extremos ($> 100mm/24h$), esta cuenca tiene limitaciones en la regulación, no pudiendo así mitigar los costes aguas abajo. Parte de los resultados obtenidos formaron parte del trabajo presentado en Egüen et al. (2015). Estos análisis nos permiten identificar la necesidad de una optimización temporal más precisa en la gestión de avenidas en este sistema hidrológico. Finalmente, realizamos un análisis de otro riesgo hidrológico importante como son las sequías, y cómo podemos representar este déficit hídrico mediante índices estandarizados, tanto para la pluviometría como para la escorrentía.

En el capítulo 3 se analizan los diversos enfoques y metodologías para el modelado de series temporales hidro-meteorológicas. Los enfoques se exponen de forma antagonista entre las diferentes opciones de modelado que tenemos: caja negra vs. caja gris vs. caja blanca, paramétricos vs. no-paramétricos, estático vs. dinámico, lineal vs. no-lineal, frecuentista vs. bayesiano, único vs múltiple, entre otros..., enumerando las ventajas e inconvenientes de cada enfoque. Algunas ideas surgidas en esta parte de la investigación,

fueron expuestas en Herrero et al. (2014). Por otro lado, también se discuten los pasos de partición, gestión y transformación de los datos para una correcta aplicación de este tipo de métodos experimentales. Esto es de gran importancia, ya que parte del trabajo duro en la aplicación de este tipo de metodologías, proviene de la transformación de los datos para que los algoritmos y las funciones de transferencia funcionen correctamente. En la parte final de este capítulo, nos centramos en cómo evaluar y validar el comportamiento determinista y probabilístico mediante coeficientes evaluativos. En este punto, prestamos especial atención en evitar la utilización de coeficientes que enmascaren los resultados o muy generalistas, y por lo tanto nos centramos en aquellos que evalúan las capacidades predictivas y de precisión de los modelos. También se ha tenido en cuenta la parsimonia para los modelos basados en redes neuronales, ya que pueden caer fácilmente en una sobre-parametrización.

El capítulo 4 expone trabajo puramente experimental, donde se realizan siete regresiones lluvia escorrentía a corto plazo, seis diarias y una horaria. Los casos de estudio corresponden a diversos puntos de interés dentro de las zonas de estudio, con importantes implicaciones en la gestión hidrológica. A escala horaria se analiza las capacidades de eficiencia y predictivas de la Regresión Lineal Múltiple (MLR) y Redes Neuronales Bayesianas (BNN) a diez horizontes temporales para el nivel del río Guadalhorce en el puente de Cártama. Se encontró que para horizontes predictivos más cercanos, un enfoque más sencillo como puede ser el lineal (MLR), puede superar a uno con mayores capacidades predictivas a priori, como pueden ser uno no lineal (BNN). Simplificando así, el desarrollo y la implementación de este tipo de técnicas computacionales bajo este tipo de marcos hidrológicos. Por otro lado, a escala diaria se establece un marco comparativo entre los dos modelos anteriores, MLR y BNN, y un método bayesiano completo: Procesos Gaussianos (GP). Esta técnica computacional, nos permite aplicar funciones de transferencia de diferente naturaleza bajo un único modelo. Esto es una ventaja con respecto a los otros dos modelos computacionales, ya que los resultados nos indican que a veces funcionan bien en un dominio, pero no funcionan bien en el contrario. Durante la construcción de los modelos, la selección de las variables de entrada se realiza de forma progresiva, mediante un método de prueba y error, donde se tienen en cuenta las mejoras significativas con respecto a la última estructura de predictores preservando el principio de parsimonia. Se han utilizado datos de diferente naturaleza: datos reales recogidos en las redes de monitorización y datos generados paralelamente de modelización hidrológica con base física (WiMMed). Los resultados son robustos donde la principal limitación es el alto coste computacional por el método recurrente e iterativo. Resultados de este capítulo fueron presentados en Gulliver et al. (2014).

En el capítulo 5 se realizan tres experiencias predictivas a escalas temporales de medio plazo. El primer experimento de modelado (Sec.5.2) se basa en una escala de tiempo trimestral, donde se establece un esquema temporal hidrológico para determinar la escorrentía acumulada para horizontes de tiempo específicos. El esquema temporal se define para que estos horizontes coincidan con fechas donde tiene lugar la planificación hidrológica. En este ejercicio, se valida que los pronósticos son más prósperos después de que transcurran los primeros seis meses del año hidrológico, en lugar de los tres primeros en los que realizamos las evaluaciones. Las variables descriptivas analizadas fueron: escorrentía acumulada, precipitación acumulada, valores acumulados de variables de la nieve modelizadas, y oscilaciones atmosféricas (AO). A nivel del modelado con técnicas DD dentro de un marco predictivo estático, esta experiencia ha demostrado la superioridad de los modelos mixtos (clasificación más regresión) frente a la aplicación de modelos regresivos unicamente. De esta manera reducimos y acotamos el espacio de posibles soluciones y así, optimizamos el comportamiento predictivo del modelo DD. Dentro de este ejercicio, se comparan tres clasificadores DD: *Probabilistic Neural Network* (PNN), *K-Nearest Neighbour* (KNN) y *Support Vector Machine* (SVM), donde se valida un mejor comportamiento del clasificador SVM. Sin embargo, aún se requiere más investigación sobre los clasificadores en marcos hidro-meteorológicos como los nuestros debido a su alta variabilidad. Esta parte de la tesis doctoral se mostró en Gulliver et al. (2016).

En el segundo apartado de este capítulo (Sec.5.3), realizamos un ejercicio de predicción de lluvia a escala mensual. Para ello usamos un modelo auto-regresivo BNN, siguiendo el mismo método que en el capítulo anterior (Cap.4) para la selección de las variables de entrada (IVS). De esta manera se ha validado el método IVS a otra escala temporal. Los resultados en términos predictivos fueron pobres para esta importante y compleja variable hidro-meteorológica, confirmando así la dificultad de predecir la lluvia a partir de datos históricos y sin la incorporación de herramientas dinámicas de modelado físico. Se constata así la necesidad de modelado hidrodinámico complejo para su predicción. Por otro lado, este caso sirvió para inferir de manera empírica la relevancia de cuatro oscilaciones atmosféricas sobre la lluvia registrada en los puntos de estudio. A partir de simulaciones múltiples y con un enfoque basado en el modelo, se han establecido qué índices tienen una mayor influencia sobre la lluvia registrada en los puntos de estudio.

En el último apartado (Sec.5.4), con las series temporales más representativas se realizó un ejercicio de predicción de anomalías, tanto pluviométricas (tres puntos de estudio) como de escorrentía (un punto de estudio). En este caso, se verifica la idoneidad de siete distribuciones estadísticas para caracterizar las anomalías en condiciones mediterráneas.

Bajo esta hipótesis, los índices que pasaron el test de normalidad *Shapiro-Wilk*, se modelaron con el fin de analizar las capacidades de BNN para predecir estos índices a varios horizontes temporales. Aquí, las predicciones de las fases negativas (sequías ó periodos de déficit) han sido deficientes, y el comportamiento de los modelos para las fases positivas (periodos húmedos) ha sido más exitoso. Con respecto a la inferencia causal de las AO y su posible influencia en la zona de estudio, descubrimos cómo el *North Atlantic Oscillation* (NAO) y el *Western Mediterranean Oscillation* (WEMO) ayudan en pronósticos de horizontes temporales menores, mientras que los *Mediterranean Oscillation Indices* (MOI) ayudan para la predicción a horizontes temporales mayores. Hemos analizado la relevancia de estas variables atmosféricas en cada caso, donde a veces su introducción fue conveniente y en otros no, siguiendo las reglas de construcción y detallándose en cada caso de estudio. Para ello, nos hemos guiado por las tendencias de aprendizaje (*Learning trends*), que se establecen sobre los resultados de las múltiples simulaciones realizadas.

A lo largo del trabajo se ha validado la utilidad de enfoques mixtos de modelado, utilizando modelos basado en datos observados de las diferentes redes de monitorización con modelado físico para la reproducción de procesos hidrológicos esenciales, como por ejemplo la nieve. Con la metodología propuesta, se ha observado una influencia positiva de oscilaciones atmosféricas para la predicción a medio plazo dentro de las regiones de estudio, no encontrando evidencia para la predicciones a corto plazo (escala diaria).

En el capítulo 6 se exponen las conclusiones finales y futuras líneas de investigación.

Las aplicaciones de este tipo de métodos son necesarios en la actualidad, ya que nos ayudan a establecer relaciones basándonos en datos hidro-meteorológicos medidos y por lo tanto "basados en datos reales", sin hipotetizar ningún supuesto. Este trabajo, nos muestra que son útiles para acotar la incertidumbre de los estados futuros y para la optimización de los recursos hídricos. Mediante métodos supervisados, el establecimiento de relaciones temporales entre diferentes agentes medioambientales nos permite el establecimiento de relaciones causales entre ellos. Donde posteriormente, un ejercicio de inferencia física es necesario con el fin de añadir coherencia y establecer un ejercicio científico integro.

Los resultados obtenidos en este trabajo sirven para reafirmar la utilidad de la implementación de este tipo de marcos metodológicos tanto en el ámbito público como privado, siendo un buen punto de partida de transferencia tecnológica. Gran parte de las rutinas y modelos establecidos en esta tesis podrían ser aplicados directamente en Servicios Hidrometeorológicos o Sistemas de Ayuda a la Decisión para los gestores del agua. Esto

incluye usuarios potenciales tan variados como las administraciones públicas y organismos de cuenca, gestores de embalses, empresas energéticas que gestionen la generación hidroeléctrica, comunidades de regantes, plantas embotelladoras de agua,...etc. El establecimiento de marcos iterativos y automáticos para el procesamiento y modelado de los datos necesita ser implementados en los diferentes sistemas hidrológicos locales y regionales, con el fin de explotar todo el potencial de este tipo de técnicas computacionales.

Communications and Publications derived from this thesis

– Chapter 2–

- Egüen, M., Polo, M. J., Gulliver, Z., Contreras, E., Aguilar, C., and Losada, M. A. 2015. Flood risk trends in coastal watersheds in South Spain: direct and indirect impact of river regulation. Proceedings of the International Association of Hydrological Sciences, 370, 51-56.

– Chapter 3 –

- Herrero, J; Gulliver, Z; and Polo, M.J. 2014. Flood Alert System For Early Warning In Mountainous Coastal Watersheds: Coupling Data-Driven And Physically Based Hydrological Models. CUNY Academic Works. http://academicworks.cuny.edu/cc_conf_hic/426

– Chapter 4 –

- Gulliver, Z., Herrero, J., Polo, M.J. 2014. Streamflow forecasting by a data driven method. Two mediterranean study cases. In Proceedings: International work-conference on Time Series - ITISE. Vol. 2, pp. 1303 -1315. Copicentro Granada, S.L, ISBN 978-84-1581-497-4.

– Chapter 5 –

- Gulliver, Z., Herrero, J., and Polo, M.J. 2016, April. Medium-term predictions of cumulative runoff in a Mediterranean mountain river. In EGU General Assembly Conference Abstracts (Vol. 18, p. 14250).

Acknowledgements

I would like to thank my supervisors Dr. Javier Herrero and Prof. María José Polo for their constant support and patience during the running of this project. In addition I would like to thank all the virtuous people that I have met during my stays in Córdoba, Granada and Exeter. Always Grateful!

$\pi\alpha\nu\tau\alpha \quad \rho\epsilon\iota$

$\sim$ Heraclitus $\sim$

(536–470 BC)

Contents

List of Figures

List of Tables

Abbreviations

AI Artificial Intelligence
ANN Artificial Neural Network
AO Atmospheric Oscillations
AW Average Width
BA Bootstrap Aggregating
BMA Bayesian Model Averaging
BN Bayesian Network
BNN Bayesian Neural Network
BRR `Beznar-Rules Reservoir`
CV Cross-Validation
CW Connection Weight
DD Data-Driven
DSS Decision Support System
DT Decision Tree
EV Evidence period
EWA Equal Weights Averaging
FM Flood Management
FP Flood Protection
GCMs Global Circulation Models
GF GuadalFeo River Basin
GH GuadalHorce River Basin
GP Gaussian Processes
GRS Guadalhorce Reservoir System
HL Neurons within the Hidden Layer
HP Hydrological Plan
IVS Input Variable Selection
KCF Kernel Covariance Function
KNN K-Nearest Neighbours
-lin *lin*ear test (MLR)
LR Learning Rate
MAE Mean Absolute Error
ML Machine Learning
MLP Multi Layer Perceptron
MLR Multiple Linear Regression
MOI Mediterranean Oscillation Index
Mom Momentum
MSE Mean Square Error
NAO North Atlantic Oscillation
Nlin *N*on-*lin*ear test (BNN)
NSE Nash–Sutcliffe Efficiency
PM Persist Method
PNN Probabilistic Neural Network
POC Percentage Of Coverage
RB River Basin
RMSE Root Mean Square Error
RBF Radial Basis Function
RI Relative Improvement
RR Rainfall-Runoff
SCoA Strongest Connection Algorithm
SPI Standard Precipitation Index
SRI Standard Runoff Index
SSE Sum Squared Error
T Training period
V Validation period
WA Water Allocation
WEMO Western Mediterranean Oscillation
WFD Water Framework Directive

Symbols

Symbol	Description
	– – – General – – –
Fig.	Figure
Tab.	Table
Chap.	Chapter
Sec.	Section
App.	Appendix
i.e.,	*id est* (that is, in other words)
e.g.,	*exempli gratia* (for example)
––	**– – – Chapter 4 – – –**
x	Input variable
y	Output variable
n	Observations or time slices
X	Potential causal variable or predictor
h	Time horizon $(t + h)$
$\mathcal{D}$	Dataset
R_j	Rainfall collected at location j $(mm \cdot day^{-1}; mm \cdot hour^{-1})$
$R_{\bar{a}w}$	Averaged rainfall of w stations $(mm \cdot day^{-1})$
ϱ_i	Performance coefficient i
τ_i	Relative improvement threshold for the ϱ_i metric
$\mathsf{WH}_{\mathsf{Ca}}$	Maximum water height at Cártama (m)
Q	Releases $(m^3 \cdot h^{-1}; m^3 \cdot day^{-1})$
R_{xy}	Cross-correlation
R_v	Rainfall collected at location v $(mm \cdot hour^{-1})$
V	Celerity value $(m \cdot s^{-1})$
$\mathcal{O}$	Computational cost

$\mathsf{SF}_{\texttt{Ard}}$	Streamflow at Ardales ($Hm^3 \cdot day^{-1}$)
$\mathsf{SF}_{\texttt{GRS}}$	GRS reservoir inflows ($Hm^3 \cdot day^{-1}$)
$\mathsf{SF}_{\texttt{Ru}}$	Rules reservoir inflows ($m^3 \cdot s^{-1}$)
R_{Wi}	Total precipitation ($mm \cdot day^{-1}$)
R_S	Snowfall ($mm \cdot day^{-1}$)
R_e	Effective Rainfall ($mm \cdot day^{-1}$)
S	Snowmelt ($mm \cdot day^{-1}$)
$\mathsf{SF}_{\texttt{Be}}$	Béznar Reservoir inflows ($m^3 \cdot s^{-1}$)
$\mathsf{SF}_{\texttt{Tr}}$	Streamflow at Trevelez ($m^3 \cdot s^{-1}$)
k	Kernel Covariance Function KCF
β	MLR parameters
T	Temperature (^oC)
f_a	Transfer function
θ	BNN parameters
α_h, β_h	BNN hyper-parameters
t_c	Computation time ($s; min$)
$---$	$---$ **Chapter 5** $---$
$\mathsf{cSF}_{\text{Nar}}$	Cumulated streamflow at Narila gauge station ($Hm^3 \cdot \texttt{hor}^{-1}$)
$\mathcal{A}_0$	Initial conditions
$\texttt{hor}$	Time horizon
$\mathsf{c}R$	Cumulative rainfall ($mm \cdot \texttt{hor}^{-1}$)
$\bar{AO}$	Average value of atmospheric oscillations [1]
$\bar{SC}$	Cumulative snow information ($mm \cdot \texttt{hor}^{-1}$)
I_d	Predictors ($d = 1, \ldots, D$)
R_j	Rainfall at location j ($mm \cdot month^{-1}$)
M	Month numerical value ($1, 2, \ldots, 12$)
$MA(q)$	Rainfall moving average of q order ($mm \cdot month^{-1}$)
AO_o	Atmospheric Oscillations ((1), $o = 1, 2, 3, 4$)
N	NAO (1)
W	WEMO (1)
M1 and M2	MOI-GI and MOI-CA (1)
$P(x)$	Probability distribution

– – – Appendix A – – –

$\mathcal{M}$	Modelling/Computational technique
μ	momentum
θ	GP' model Parameters
θ_1	Weights matrix of the hidden layer
θ_2	Weights matrix of the output layer
α_{hyp}	BNN' hyperparameter
β_{hyp}	BNN' hyperparameter
K	ensemble member
l	KCF length scale of GP

$\infty\infty\infty\infty\infty\infty\infty\infty\infty\infty$

$\square\square\square\square\square\square\square\square\square\square$

God willing,

to Mum^3*, Dad,*

Assia, Alma and

Mariam

$\square\square\square\square\square\square\square\square\square\square$

$\infty\infty\infty\infty\infty\infty\infty\infty\infty\infty$

Truth lifts the Heart,

like Water refreshes Thirst

∼ Jalaluddin Rumi ∼

(1207-1273)

Chapter 1

Introduction & State of Art Review

" *Pensar es olvidar diferencias, es generalizar, abstraer. En el abarrotado mundo de Funes no había sino detalles, casi inmediatos*"

∼Jorge Luis Borges∼
Funes el memorioso

1.1 Introduction

The finite nature of water resources leads to a constant renewal of the methodologies used for its correct management and optimization to achieve unitary systems, an oversight of the water bodies within an integrated view and less segmented control scenarios. In the River Basins, we have several hydrological issues that society has to face. The most important are the torrential rains and floods, droughts, to preserve environmental health (both physical and chemical) and equitable operation and fair planning to benefit all users.

Today we live in a rapidly changing climate where we are attending to quick changes on the extreme events, in both magnitude and duration. Therefore advanced methods for prediction should be applied in order to make better planning for the water resources available. Water resources planning cannot be done without the right tools, and scenarios must be done for forecasting situations, which involves decision making.

The "Data-Driven" tools provide guidelines for decision making. According to the Water Framework Directive (WFD, Directive 2000/60/EC) this planning is not only recommended but also mandatory. The main problems in achieving this is that the administration is usually short on resources, personnel and the tools to achieve it. For that reason we must focus on parsimonious tools, as simple and useful as possible. To take decisions in a more socially-responsible manner it must be based on knowledge, taking into account all the information technologies available nowadays in order to develop smart applications for the society. In this way we make the knowledge of our environment more accessible for non expert users, transferring knowledge in a easy and simple way.

"Data-Driven" techniques will allow us to establish the best information connections among data, establishing hydrological correlated frameworks more or less opaque, depending on the specific conditions. In order to develop this frameworks we can find different data sources: data collected in the field, data estimated by mathematical and statistical methods or data modelled by computation, less or more physically. By emphasizing that these established models must be robust and flexible, allowing us to deal with the complexity of water systems, and simulate possible future changes in this complex geophysical systems.

These changes are often part of the natural evolution of geophysical forces and, in other occasions, forced by the human being interaction. In recent times, terms like "climate change" is showing up in every level of the society not without controversy, and even with contradictory information depending on the source. This term search in Google Scholar returns 1,780000 results, while the term "water " returns 7,270000 and, term "ocean" returns 2,910000.

In a context of changing patterns, strong scale effects and regional specificities, it is a challenge to develop models (or methodological frameworks) that can be adapted to different water systems, dealing with the problem of uniqueness (Beven, 2001) on hydrological modelling applications. For that, the requisite is to focus on fast and dynamic computing tools that can be adapted to different hydrological frameworks, developing jointly approaches of various modelling techniques and ensembles of different modelling sources. This will allow us the simulation of extreme future values and the environmental and socio-economic impacts associated in order to take best decisions on future level changes in the water systems.

1.2 Objectives

The aim of this work is to increase knowledge of different Data-Driven techniques to predict extreme events and optimise the management of reservoirs under Mediterranean meteorological conditions. This will ease to make future events more predictable and deal with the uncertainty associated with Mediterranean environments, which are highly variable. As a result, the decision makers will have intelligence support to resolve with smaller errors and narrower ranges. To develop a reasonable and comprehensive plan, one must have a legitimate intuition into the future.

Specific Objectives

In order to address the general objectives of this thesis, the following specific objectives will be achieved:

- Multiple analysis of hydro-meteorological and climatic time series at different temporal scales: hourly, daily, monthly and quarterly.
- Identification of which computational technique's performance is the best under Mediterranean conditions.
- Floods and Droughts forecasting experiences in two study sites.

1.3 Structure of the Thesis

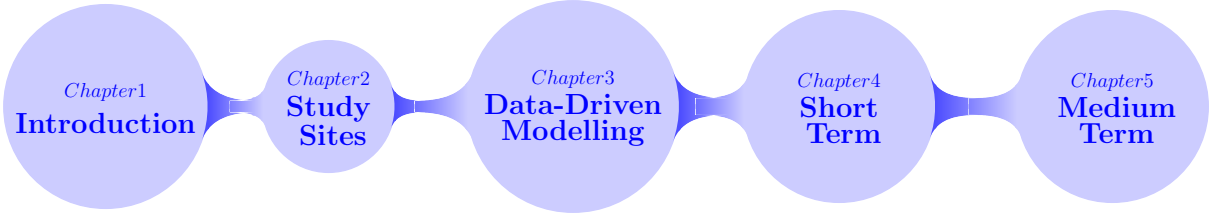


FIGURE 1.1: Novel techniques analysis across different temporal scales

The Thesis is structured under a temporal scale view, from smaller scale to larger scale, starting with time signals more frequent as hourly. This signal can be transformed, and subsequently aggregating data to other time interval scales, as monthly or quarterly.

Chapter 1 exposes the motivations and aims of this thesis with its objectives. Also a review of novel computational techniques for hydrological purposes is carried out, paying special attention to its application under Mediterranean conditions.

Chapter 2 provides a description of the study sites and the data sources available for the water systems. The water systems will be two: the Guadalhorce river basin and the Guadalfeo river basin, giving a detailed description of its physical and hydro-meteorological context. Finally an analysis of climate driving factors, floods and droughts is carried out.

In chapter 3 different data-based methodologies for hydro-meteorological modelling attending to several perspectives are discussed and analysed. Also how to treat and manage the datasets is pointed out followed by the evaluation coefficients to be used for the correct characterisation of models behaviour according to our objectives.

Chapter 4 presents Short-Term Forecasting experiments for different locations at hourly, daily and event window time scale. A deep analysis of several Data Based regression models have been done to forecast water levels and stream-flow values at seven locations.

In chapter 5 we present Medium-Term Forecasting experiments for different locations in the basins under study. Cumulated stream-flow and rainfall are addressed first, finalizing in a Droughts forecasting trial.

Finally in the chapter 6, the Thesis's conclusions are listed and commented with a mention on the future research lines opportunities that arises in each one.

Additional information has been added to the document to describe the Machine Learning techniques used (appendix A) and, the Study Site, Data & Computational Coding used (appendix B), which are listed before the Bibliography.

Mind maps appear throughout the document to visualize the relevant information of each chapter; key words and concepts are represented by nodes from which other concepts can branch out. This diagrams will help to the reader to focus on major points and key issues of this work.

1.4 Modelling and Data-Driven techniques for Hydrology. State of Art

A model is a simplified representation of a complex system; in our case that model will represent river systems in a Mediterranean climate. In the last decade the application of Artificial Intelligence (AI) to deal with complex systems has been verified in all fields of knowledge with considerable advances, with numerous publications and diverse works as is further discussed. That includes applications on diverse fields, as civil and environmental engineering, medicine, biology, economics, etc... since they allow to fit a mathematical or computational model to any domain of knowledge for decision making moving in the margins of uncertainty.

A must study about AI can be found in Russel and Norving (2010), where its establishments are set up. In hydrology, we meet in Savic (2019) a fine explanation of how AI helps in hydrological engineering and how we can benefit from its use. We can note the blend of informatics and hydrology with the term "hydro-informatics", where local actions are indispensable, so that the advances of computerization of data will affect to the entire community.

Day by day the latest advances in AI and Machine Learning (ML) techniques (to reduce terminology let's call them Data Driven (DD) computer techniques) and the storage of massive amounts of data collected pushes us to apply these data-based methodologies, real data whenever possible, for problem-oriented cases of our particular interest. This engineering challenge must produce tools that are able to predict future values for an extended state and quantify the uncertainty in the decision-making phase for a better understanding of the water system.

Many of this DD techniques may be nature-inspired/bio-inspired models of computation (Yang, 2010), meaning that the model is developed with the observation of the nature and transposed to symbolic expressions and finally to computational algorithms. These models deal with forecasting and optimisation and in general with modelling engineering problems. In general, DD techniques offer several frameworks for environmental modelling (Solomatine, 2006; Hsieh, 2009; Islam et al., 2014). In hydrology the application of ML techniques are well known (Hong, 2008; Deka et al., 2014; Remesan and Mathew, 2015) and have been tested with different approaches both on the computer field and on the hydrological environment. More specifically for streamflow forecasting experiences a good review can be found in Yaseen et al. (2015).

Mainly DD techniques provide methodological frameworks for streamflow modelling purposes, where a pre-understanding of the local behaviour patterns of climate and hydrological processes is required, even though their relationships can be complex and notoriously non-linear. Previous studies of DD (e.g., Artificial Neural Networks) over the last decade have established and proved their goodness in hydrological modelling (Govindaraju, 2000; Maier et al., 2010) where there is still room for improvement, as noted in Abrahart et al. (2012). Comparative works of DD techniques under different hydrologic regimes can be found in Adamowski et al. (2012); Rasouli et al. (2012) and Sun et al. (2014).

1.4.1 Regression Problems

The regression cases is a key issue in hydrological applications and within the pattern recognition task in general. The regressions allow us to estimate relationships among data through a mathematical/computational model, where the objective is to minimise or maximise the objective functions (with residuals minimization being one of the most common approaches). For estimation, data can be measured or modelled. Also, the chosen technique can be parametric/non-parametric or linear/non-linear. In either case, the final choice of model must be based on an acceptable accuracy for the intended purpose.

In regression problems a set of input variables $x_1, \ldots, x_d$ consisting of d number of variables has a functional relationship between the d-dimensional predictor $x \in \Re^d$ and the target variable y in the form $y = f(x)$. On the time-line, each observation $y_{(t)}$ can be thought of as related to an underlying function $f_{\mathcal{M}}(x_{(t)})$ through a Data Based computational method or model, let's denote it by $\mathcal{M}$. Normally a model error ϵ or noise model to this relational function $f_{\mathcal{M}}$ between the regressors and the regressands is added, which can be interpreted probabilistically as represents the "plausible/believable" values for the estimation.

Under maths perspective, numerous regressive models can be developed for the intended purposes. The first regression model is an autoregressive model ($\mathrm{AR}_{(p)}$), where the target variable y is related with its antecedent values trough the regression coefficients α in the following form: $y_{(t)} = f(\alpha_1 y_{(t)}, \ldots, \alpha_d y_{(t-p)}) + \epsilon_{(t)}$. Usually the $\epsilon_{(t)}$ represents the stochastic white noise term. This first expression, and the simpler, are convenient for dependent and autocorrelated data modelling, such as many hydro-meteorological patterns, but habitually are insufficient due to the presence of the lag effect. This relationship helps to represent the non-stationary dynamics of the physical systems, conditioning the

target variable estimations to the most recent observable values. The next step may be the inclusion of a statistic metric as the Moving Average of order q ($\text{MA}_{(q)}$). In this sense the mixed Autoregressive Moving Average models ($\text{ARMA}_{(p,q)}$) can model simple time series successfully, even with exogenous inputs within the function relationship (i.e., ARMAX). Finally more complicated models as the Autoregressive Integrated Moving Average (ARIMA) can be developed, which are not recommended for seasonal data. Another version of this model, Seasonal Autoregressive Integrated Moving Average (SARIMA), includes a seasonal term overcoming this limitation.

In that sense, different DD computational techniques are available for regression purposes: Multiple Linear Regression (MLR), Artificial Neural Network (ANN), Support Vector Machines (SVM), Local Polynomial Regression (LASSO) and Partial Least Squares (PLS) among others are good example of them. Hydrometeorological regression applications have been widely developed in the past with varying degrees of success.

1.4.2 Classification Problems

The classification of variables is widely used for the simplification of numerical domains. It is used for river states, hydro-meteorological states, ocean-atmospheric oscillations, geo-spatial modelling, probabilistic models, and in general to the whole process of computational modelling in any field of knowledge. Modelling the classified variables will always be at a reduction of complexity of the solution space compared to the problems of the continuous domains. As a general rule, this translates into a reduction of the numerical dimension of the problem and, therefore, also lower computational requirements.

In a classification problem, a discrete class label C_k, $k = 1, \ldots, K$ of each input vector X_n, $n = 1, \ldots, N$ is known. The classification models is driven by the "classifier" which is the mathematical function that is implemented by a certain algorithm to map the input data into a category. Usually, the computation techniques used in classification coincide with the techniques employed in regression problems, having a supervised learning framework for it. For instance ANNs, SVMs, and GPs are a good example of them. Other techniques as the logistic regression, Probabilistic neural networks (PNN, Wasserman (1993)), Naive Bayes, K-Nearest Neighbour algorithm (KNN) and Classification and Regression Trees (CART, Breiman et al. (1984)) can be implemented and used with classified variables.

A useful machine-learning methods for clustering and classification are often used combined with regression methods (Solomatine et al., 2009) to improve models performance, for instance supporting them with a classification work-flow within the methodological framework. Usually hydrometeorological examples of classification modelling can be found in spatial problems as the domain can be huge and the simplification by classification makes the problem more manageable to be addressed.

1.4.3 Optimisation Problems

In water resources management, the first and clearest example of optimisation problems is dealing with an optimal allocation of water volume for different users at a basin scale. In this situations we have a total quantity of water that need to be allocated to a number of different uses. The objective is to determine the allocation such that the total net benefits for all uses is maximized for a given objective function.

In Labadie (2004) a review of different approaches for reservoir multiobjective optimisation is carried out, with more specific applications being described in Ngo et al. (2007) and Escuder-Bueno et al. (2011). A good compendium of all the foundations and methods for water planning and management of the water systems is shown in Loucks et al. (2005). Here it is outlined the basis of Linear Programming (LP) and Dynamic Programming (DP) as a optimisation method. In hydrological applications a variant, Stochastic Dynamic Programming (SDP, Stedinger et al. (1984)), takes into account the uncertainty of variability through transition matrices, which is especially necessary for optimisation purposes in Mediterranean water systems, which are subject to conditions of high spatial and temporal water variability.

Coupled with the multiobjective optimization problems, the use of Ensemble Streamflow Prediction (ESP) traces is well known in water management. The final product of ESP is a forecasted hydrograph based on various empirical distributions, producing multiple runs of the all possible coming values by different boundary conditions. Among the decision support systems (DSS) for the management and optimization of reservoir systems are the MODSIM-DSS (Colorado, 2011) developed by Colorado State University and the HEC-Rec SIM (from of Engineers-Hydrologic Engineering Center (2013)) developed by the U.S. Army Corps of Engineers.

In this direction, bio-inspired techniques as the Genetic Algorithms (GAs) can deal with very complex water systems due to its non-linear capabilities. In Morley (2008) a robust

framework with GAs for water system distributions optimisation is developed for two big cities: Hanoi (Vietnam) and New York (USA). In that sense, a practical solution is presented in Savić et al. (2011) presenting a Excel shell to achieve various optimal solutions by GAs, avoiding the use of difficult computational environments, putting the view for users with less computational skills. More examples of reservoir optimisation can be found in Malekmohammadi et al. (2009) and Malekmohammadi et al. (2011).

A recurring optimization problem in many engineering applications is dealing with the estimation of parameters for any given model. Also in this modelling challenge, AI algorithms are commonly used nowadays. For instance, in Kisi et al. (2012) the ANNs with artificial bee colony (ABC) is applied to optimise a discharge-sediment model. This type of non-linear optimization techniques are mainly applied during the parametrization process (weights and biases) of the ANNs, replacing the classical optimization algorithms such as the Scaled Conjugate gradient (Moller, 1993). Also we can find applications of the ant colony optimisation (ACO, Dorigo et al. (1996)), where a review of this type of optimisation algorithms for water management can be found in Afshar et al. (2015). Another use of the Evolutionary algorithms is for ANNs' weights optimisation, which transforms the classical ANNs into a new form of machine learning: neuroevolution, which is also applied in this work for regression purposes

Finally, an emphasis must be placed on game theory based methods as a future trend to deal water management problems. This methods are interesting specially for water conflicts among different groups or individuals. In Madani (2010) a extended review is carried out summarizing the applications on water resources conflict resolutions, for instance: on Water or cost/benefit among users or on water allocation among transboundary users. Decisions that should be made as fair and balanced as possible.

1.5 Data Driven modelling under Mediterranean conditions

The Mediterranean water systems are complex. Its climate is characterised by winter rainfall, dry hot summers, and a heterogeneous regime of soil moisture (Poesen and Hooke). Around the globe, we can find Mediterranean climate conditions in areas of South Africa, California, Australia, Chile, and all over the Mediterranean basin. The hydrology patterns have abrupt changes but low rates of occurrence. That translates into

quick and intense pulses of the geophysical processes: rainfall, snowmelt (Herrero et al., 2009), and erosion and its transport rates (Millares et al., 2014a).

(a) Strong contrast effects, snow under semiarid conditions

(b) Reservoirs presence for water optimisation enginering solutions

(c) Very heterogeneous geospatially

(d) Monitoring networks to collect data for tools development

FIGURE 1.2: Mediterranean landscapes for hydro-meteorological modelling challenges. *Source: Z.Gulliver*

In Mediterranean environments, it must be emphasised that specific requirements within the hydrological cycle are necessary to draw a correct representation of the water systems, and properly useful for modelling experiences. The most important features of Mediterranean climates that must be taken into account for modelling experiences are:

1. the temporal variability in climatic magnitudes, which allows for the appearance of high seasonability with torrential storms and drought periods

2. the marked spatial variability because of the geographical conditions, which translates as the appearance of geology/climate/biota mosaics

3. a lack of data availability essential to get accurate results (validated and calibrated) for practical and operational final products (what we should offer to the population and end-users)

These dynamics are characteristic of the Mediterranean environments, having many examples into hydrological modelling (Gallart et al., 1994; Pinol et al., 1997; Moussa et al., 2007; Mayor et al., 2009; Agnese et al., 2014). Also, different works of DB show us their successful application under Mediterranean climate and semiarid conditions (Daliakopoulos et al., 2005; Karran et al., 2014; He et al., 2014), also dealing with intermittent flows (Kisi, 2009; Adamowski and Sun, 2010) which are often in many Mediterranean minor rivers/streams and even in snow affected basins as is pointed out in a work done in Cyprus by Glezakos et al. (2009).

Chapter 2

Study sites & Data

"Friendship is the most necessary thing in life"
$\sim$ Aristotle $\sim$

2.1 Introduction

Middle size coastal river basins around the Mediterranean basin are populated and developed. In these regions, extreme events can cause an unacceptable loss in human, social and economic terms. Here a challenge pops up, developing modelling schemes that can be adapted to different locations with similar hydro-meteorology. So, we have many small-/medium river basins along the Mediterranean see basin. Here emerges the need to set up methodological frameworks to characterise these water systems, often with concentration times of less than 24 hours. To this call, we must add the scarcity of historical data or poor quality of them.

2.2 Study River Basins

In this thesis, we select two Andalusian-Mediterranean Spanish coastal basins: the Guadalhorce river basin (GH) & the Guadalfeo river basin (GF), locations in Fig.2.1. Both of them are medium size, semi-arid, with steep slopes and thin soils having fast travel times of the flow along the valleys. In these ecosystems, drought periods are recurrent, giving rise to supply problems and conflicts of interest, with intense agricultural demand versus urban, recreational and industrial uses.

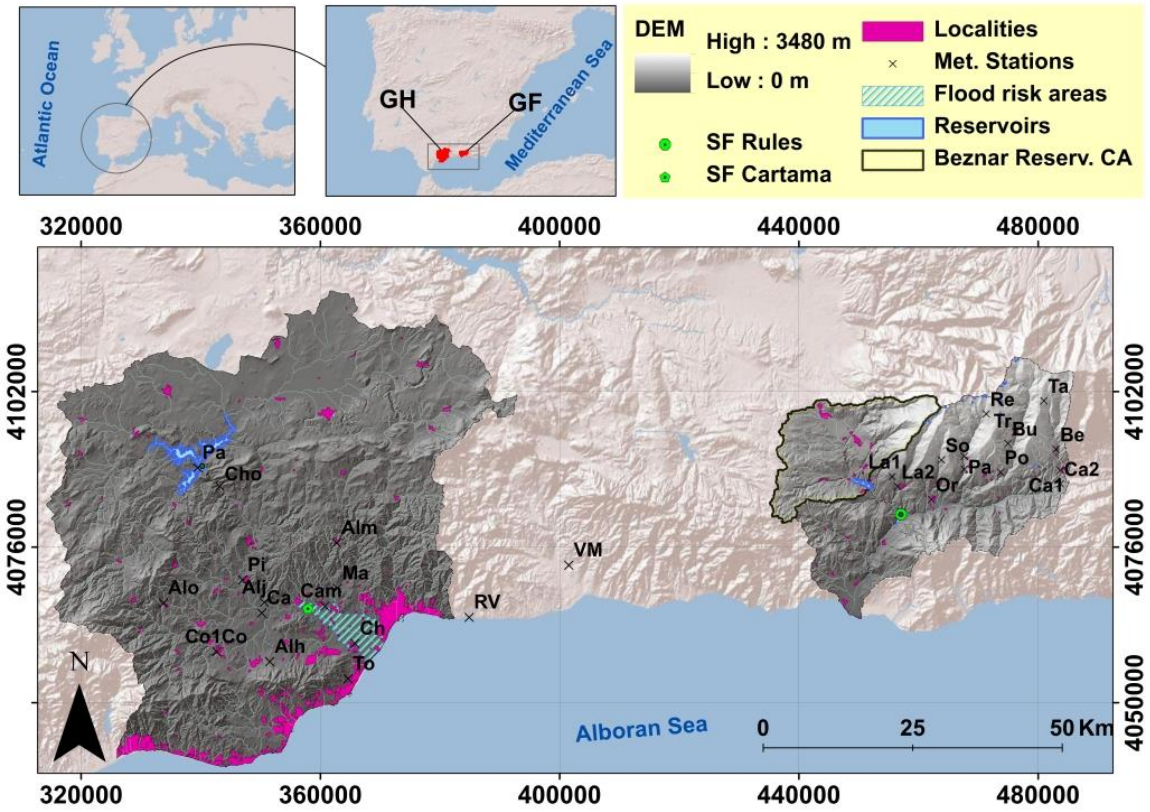


FIGURE 2.1: Guadalhorce River Basin (GH, left) and Guadalfeo River Basin (GF, right) in south Mediterranean Spanish coast. The inhabited areas are highlighted in magenta, while reservoirs are shown in light blue. The gauges stations of Rules and Cártama, a light green and white dot respectively on the map, are the control points selected for this study. The flood risk area downstream Cártama is represented as a striped polygon. The meteorological stations used in this study are located with a cross together with their related abbreviated name.

For its size (Tab.2.1) the basins respond to rapid pulses of the forcing agents. Depending on the physical nature of the terrain and the dominant hydrological processes, that may be either linear and non-linear. The chosen basins allow us to study: flash flood conditions prediction in a regulated basin (GH) and, a heterogeneous snow affected basin (GF) that requires more precise modelling (i.e., Hydrological/process oriented modelling) to capture their water dynamics.

TABLE 2.1: Summary of the geomorphological and hydraulic characteristics of the Guadalhorce (GH) and Guadalfeo (GF) river basins

	Guadalhorce (GH)	Guadalfeo (GF)
Basin area	3157 km^2	1295 km^2
River length	154 km	71 km
Basin max. elevation	1800 m	3480 m
Annual mean of the total runoff	622.71 hm^3	354.63 hm^3
Annual mean precipitation	594.3 mm	608.9 mm
Number of reservoirs	4	2
Storage capacity	369 hm^3	170 hm^3
Storage capacity/unit area	0.12	0.13
Mean Basin Slope	0.2238 m/m	0.3556 m/m
River slope ⋆	0.001/0.008/0.035 m/m	0.012/0.02/0.13 m/m
⋆Floodplains/Middle course/Header streams		

From an administrative point of view, the basins are encompassed within the Andalusian Mediterranean River Basin District (`Med.RBD`, Fig.2.2), declared the subsystem I-4 and the subsystem III-2 for GH and GF. The two basins are managed by the Andalusian Regional Government Administration (*Junta de Andalucía*) since this RBD does not share territory with other regional governments nearby.

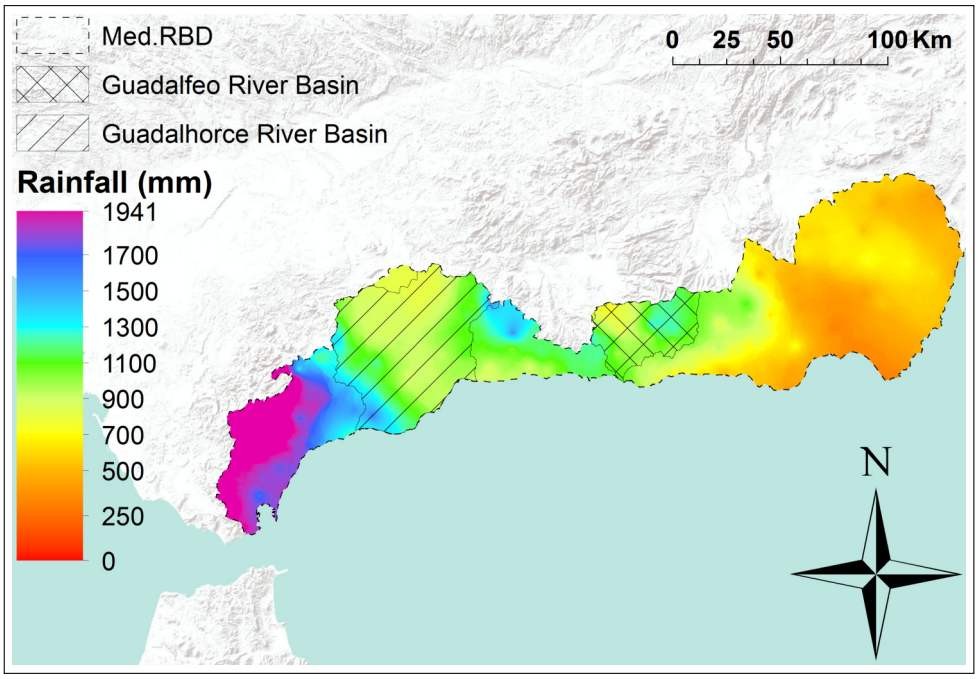


FIGURE 2.2: Mediterranean River Basin District (`Med.RBD`) which encompass the two river basins studied in-depth in this work. The Distribution of Average Annual Rainfall is represented for the 1971-2000 time period. *Source:* REDIAM

2.2.1 Guadalhorce River Basin

The first case research is the Guadalhorce River basin (GH, Tab.2.1), draining an area of approximately 3157 km^2 down to the Alborán Sea, a hilly basin that runs entirety at the province of Málaga. In the middle part of the river course, we find the Guadalhorce Reservoir System, composed of three dams with a total storage capacity of 345.51 hm^3. Traditionally, the GH river springs in *Fuente de los Cien Caños* (Fig.2.3.(A)), where the stream course descends from the western part of the Sierra de Antequera and from the eastern part of the Montes de Málaga (1031 m), Sierra de Antequera, Sierra de las Nieves and Serranía de Ronda to the Mediterranean Sea (Fig.2.1). This geormorpological barriers act as a trigger for orographic rain, which can cause extreme rainfall events and strong regional scale effects. In this fluvial system, the main contribution to the runoff is basically from rainfall, no existing other significant contributions.

For this case investigation, we focus on the floodplains (the mouth of the river), for downstream flood protection purposes of the metropolis area of Málaga, where there are recurrent floods caused by the torrential nature of the precipitation increasing the risk to an important population, on structures of the city and on its industrial area. In Senciales González and Ruiz Sinoga (2013) shows a good spatial analysis of the heavy rainfall events ($>100 mm{\cdot}\text{day}^{-1}$) near the city of Málaga, using 29 stations from 1950 to 2009.

For the subsequent calibration process, the Hydrology Service (SAIH) collected information on capacity stations, which measure water levels. The following figures show the spatial distribution and data from such stations. Concerned to the monitoring networks within the basin, the SAIH net is an automatic system for the collection of data at hourly scale, which has twenty-nine rainfall stations, nine gauge stations (Fig.2.1) and four reservoirs (Tab.2.1). They collect these data in the control centre of the water authorities to provide help in related knowledge delivery.

Within the GuadalHorce basin we can find four dams, three belong to the Guadalhorce Reservoir System (`GRS`): Conde Guadalhorce, Guadalhorce and Guadalteba (Fig.2.3.(B)), and the Cassasola reservoir, not included in this survey. Previous works carried out in this water system are Mediero et al. (2007); Versini et al. (2014); del Mar Gallardo et al. (2017); Flores et al. (2019) attending different objectives.

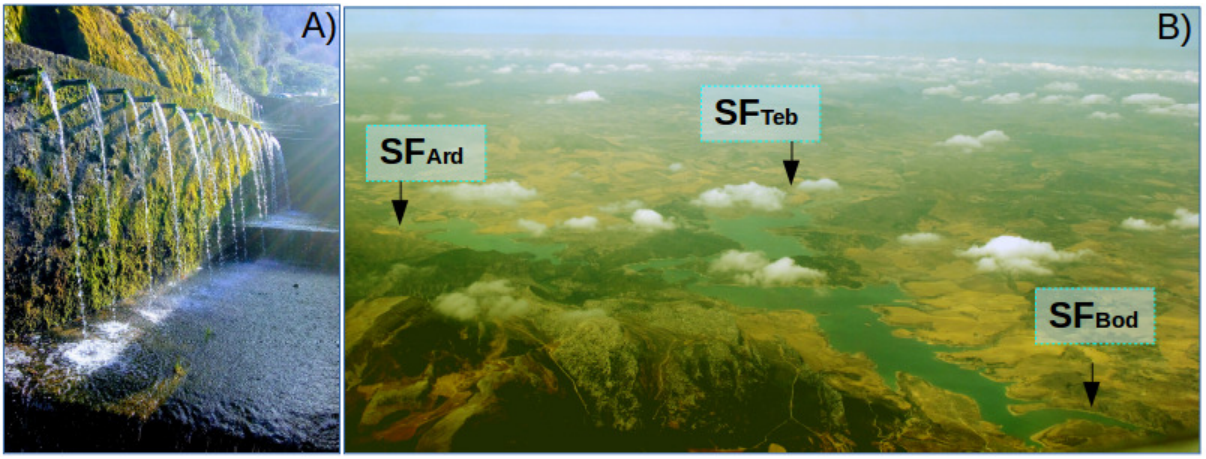


FIGURE 2.3: (A) *Fuente de los Cien Caños* located in Villanueva del Trabuco (Málaga). *Source: Z.Gulliver* (B) Aerial photo of the GuadalHorce Reservoir system. *Source: Z.Gulliver*

2.2.2 Guadalfeo River Basin

The Guadalfeo basin (GF, Tab.2.1) exhibits a singular heterogeneity, with an important ecological value, mainly due to the presence of the Spanish Sierra Nevada range and the National Park with the same name, which is 3479 m above sea level at a very short distance from the coast line (40 km). Under a hydrological view is a snow dominated basin, where several snow melting cycles take place throughout the autumn and winter seasons, followed by a final melting period in spring or early summer. Due to its torrential nature of the climatic events, the semiarid environment and the Mediterranean agricultural activity the erosion rates are considerable.

(a) Aerial view fo the Sierra Nevada peak (b) Beznar-Rules Reservoir system view within the GF range. (A) Southwestern horizon (B) West- basin. Southeastearn horizon *Source: Google maps*
ern horizon. *Source: Gulliver, Z.*

FIGURE 2.4: Guadalfeo River Basin perspectives

In the northern region of the basin (Alpujarras) there is a characteristic and original artificial channel system which has been operational from the Andalusian period. Two are the main purposes of this ancient system; (i) the irrigation of crops downwards and (ii) the artificial recharge of the regional aquifers. It must be emphasised that it constitutes one of the first devices (since XI century) for artificial recharge of aquifers in the Iberian Peninsula (Fernández Escalante et al., 2006).

Intensive horticultural crops (e.g. tomatoes, green beans) have greatly increased in these high-altitude areas, where a cooler climate favours better agro-climatic conditions for growing these products than on the coast or in inland regions, with greenhouse production usually stopping during the summer months. This has greatly altered the natural water cycle in this area, which may cause insufficient flows running for a good ecological status, especially during the end of the spring and summer. A key issue to be addressed by the competent authorities, both from a hydrological and an environmental point of view.

In this River Basin rain-on-snow events also can take place. The last one was on December 29, 2009, caused by a combination of snowmelt and heavy rainfall after 4 consecutive days of rain, measuring up to 300 mm of accumulated rain on those days and increasing rain intensities up to 260 mm in 20 hours.

The water authorities have implemented a monitoring network to collect hydro-meterological information at different locations due to the great interest of this region. It collects the data of the streamflow average flows in different points of the basin. These monitoring points are well located spatially, but some of them are of poor quality: the time series collected are not long enough to satisfy a minimum number of data, and/or the number of missing values is limiting.

In the GF River Basin exists the Beznar-Rules Reservoir System (`BRR`), which consist of two reservoirs (see location in Fig.2.1), Beznar (`Be`) and Rules (`Ru`), inaugurated in 1987 and 2004 and with 56 hm^3 and 110 hm^3 of total capacity respectively. These reservoirs serve as flood control in addition to their role in water management to satisfy different water users. The water allocated in these hydraulic structures is used for: urban demands, irrigation, hydropower and ecological demands, sorted in order of importance. With the construction of the dams, the transport of sediments has been disrupted, especially in the `Ru` reservoir, due to the presence of badlands and high rates of transport (Millares et al., 2014a) on its contributing area (Fig.1.2 (c)). This situation has led to problems caused by the interruption of the natural flow of sediments, due to the silting up of the

reservoir (Millares et al., 2014b), and consequently the rupture of the sediment balance on the coast (Bergillos et al., 2016).

The `Ru` and `Be` reservoirs have available data at daily scale, having enough quality for DB applications and an acceptable time series length to work with. Under an hydrologically point of view, in terms of surface hydrology, the reservoir encloses a closed control volume for the upstream area after the quantification of withdrawals. A more detailed description of the Reservoir system is presented in further paragraphs, attending to the objectives DB modelling experience.

2.3 Data

"Data! Data! Data! I can't make bricks without clay!"

∼ Arthur Conan Doyle ∼

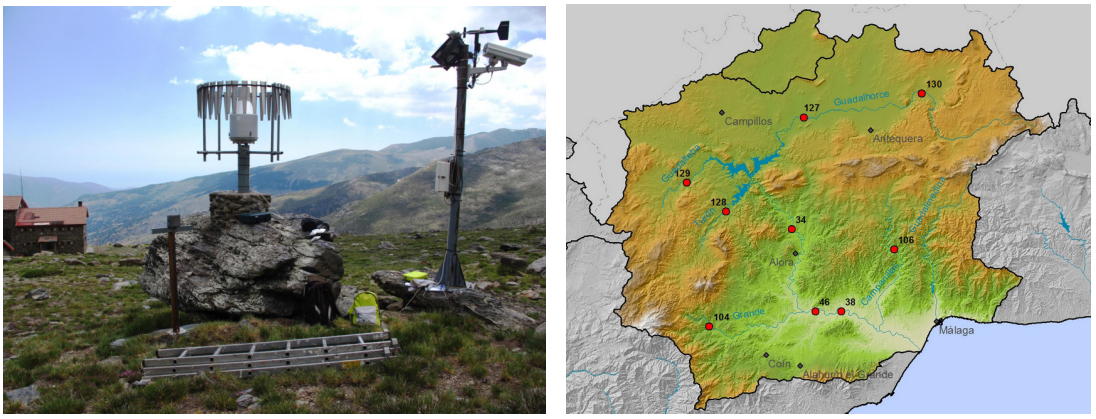
Sherlock Holmes. Adventure XII. The Adventure of the Copper Beeches

2.3.1 Hydro-Meteorology

Different data sources are used to characterise the hydro-meteorological framework of the sites under study: from hydro-meteorological information collected within the monitoring networks, as can be the rivers streamflows, water levels, or the rainfall in various locations, to snow computational modelling information generated by reproducing the hydro-physical processes in the areas under interest.

In the two River Basins under study, GH and GF, we have several monitoring networks established, attending to different objectives. On one hand, we have the monitoring networks established by the *Agencia Estatal de Meteorología* (AEMET), which is the Spain's meteorological agency operating under the Ministry of Environment. These are the oldest, where in the beginning, basic meteorological data such as rainfall and temperature were collected on a daily and monthly basis, where progressively, automated stations have been installed, increasing the number of variables to collect, and also the quality and frequency of measurement. In this sense, the network is very extensive due to the different swings and policies adopted over the years. Anyhow, the maintenance and the quality of the data collected from the stations is very heterogeneous.

On the other hand the monitoring networks established by water organizations where the river stages are registered establishing gauging stations, and more recently automated sensors. Finally additional monitoring networks are established and maintained for research purposes, such as the Guadalfeo network, which aims to characterise high mountain hydro-meteorology.



(a) Installation and Maintenance of Meterological monitoring networks for research purposes: Red Guadalfeo. *Source: Gulliver, Z.*

(b) The SAIH of GH basin which is composed of a network of Water level sensors for river states monitoring purposes, water level heights, at different locations (red point). *Source:* SAIH webpage

FIGURE 2.5: Hydro-meteorological data is collected throughout the monitoring networks within the basins under study

In the basin of the Guadalhorce River, the authorities have implemented an automatic system for monitoring the river courses using water level sensors (Fig.2.5.(b)). For each sensor, three warning levels have been established: I (yellow), II (orange) and III (red). The threshold and interval values are adjusted for each particular sensor. One of the sensors of most interest for this work is the sensor located on the Cartama bridge (sensor n°38 in Fig.2.5.(b)), where the warning levels are: level I for 2.5 meters of water level or higher values, alert II for values of 3.5 meters or higher, and finally level III for 4.5 meters or higher values.

2.3.2 Climatology

Many national agencies have developed their own models where climatic conditions are tested under different benchmarks and for different purposes. Thus, the European Centre for Medium-Range Weather Forecasts (ECMWF) is an independent intergovernmental

organisation supported by 34 European states that offers Medium-range forecasts. These forrecasts are more focused on the precision of the models in the European territory, i.e., regional, through higher resolutions. Each year a report Haiden et al. is shared with the analysis of the forecast data. 40 Year Re-analysis Climate data from ECMWF (ERA-40) have a Full Resolution Pressure Level Analysis Data Set. On the oher hand the American Administration has the National Centers for Environmental Prediction (NCEP), a branch of the NOAA's National Weather Service, with data products as the Global Forecast System (GFS).

These data fits well with the data measured in the meteorological stations at high altitudes maintained by our research group (Red Guadalfeo). For instance, the linear correlation between time series of daily mean pressure (n=1500 days) of Poqueira station at 2510 metres of altitude (Po, Fig.2.6) and the pressure (m.s.l) obtained from GFS for the same time period is R^2=0.91. So, the conditions and climatological data generated in these climatological meshes can be used for modelling. However, it should always be noted that rainfall is the most difficult variable to characterize correctly, since regional and local scale effects always require specific adaptations. These are carried out by means of downscaling techniques, or transfer functions, or any method to be able to couple/link different models at different spatial scales and the same time scale.

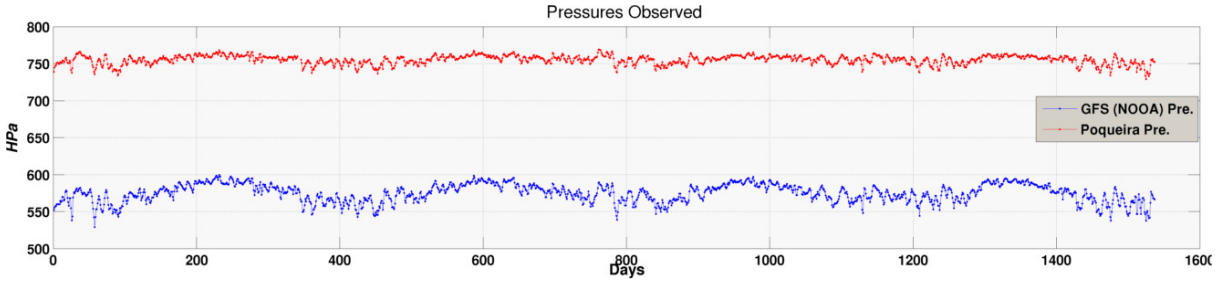


FIGURE 2.6: Representation of the pressure data from the climatic products and the pressure data collected in our monitoring networks (Poqueira, PoC) within the GuadalFeo River Basin

The downscaling methods for using climate products are arranged by the different organisms through their physical models of global circulation. These downscaling methods basically take into account a training period in which the altitude and some orographic and physical characteristics of the study area are taken into account. Downscaling methods or modelling experience remains outside the focus of this work.

2.4 Climate and Water

2.4.1 Climate

This Mediterranean region is characterized by a variable climate: 1) The temperature varies quickly from one day to another specially during fall and the spring season. 2) There are pronounced interannual variations, taking place big differences in the annual and monthly means from one month/year to another. 3) The rainfall events also are highly variable, both in the time of occurrence and in the magnitude of them. In summary the Mediterranean climate is highly variable which always constitutes a challenge the modelling and the prediction of its hydro-meteorological patterns.

The available climatic data are increasing day by day, finding in the global circulation models (CGM) the standard way to work at mesoscale with higher spatio-temporal frequencies. This information is useful for hydro-meteorological forecasting purposes, where we must establish what type of information and climate product we need. In this sense, reanalysis products are available that allow establishing the weather conditions for certain events through statistical or computational analysis. Likewise, climate predictions based on weather model forecast data products are used to develop regional forecasting tools in which downscaling methods are mandatory. Works in that direction can be found in Ghosh and Mujumdar (2008); Rasouli et al. (2012); Kalra et al. (2013); Linares-Rodriguez et al. (2015).

Climate Change: In a way, it is known that the patterns are changing where a key fact is the increasing trends of the temperatures through the globe and over the Mediterranean basin (Pausas, 2004; Harris et al., 2014). This fact adds an intrinsic marked uncertainty to any application and tools based on historical data. Since the trends and the different future scenarios augur an increase in temperatures and changes in water balances known so far, we must take into account this uncertainty for management tools development and to test their adaptation capabilities for changing patterns. These climatic changes are expected to be more pronounced in Mediterranean environments. For example, current research in southern Mediterranean Spain (Pérez-Palazón et al., 2015) shows a downward trend in the number of melting events, both in quantity and number of days per year, which may be a key process in the basins of these latitudes where snow is a major source. Recent works (Mathbout et al., 2017) also show a decrease in extreme events, and it is not very clear what kind of scenarios we are aiming at.

Currently, a very common study is the use of climatic projections established by the different trans-nationals and national bodies for regional or local research purposes. This is done by downscaling this projections according to the characteristics of the study area. The best-known international body is the Intergovernmental Panel on Climate Change (IPCC) composed actually by 195 countries, providing data and future Climate Scenarios. In our opinion, the uncertainty associated with these scenarios is still very high.

2.4.2 Floods

"Floods are acts of God, but flood losses are largely acts of man"
∼ Gilbert F. White ∼

Historically floods are known in southern Spain to cause damage to property and life. Since the sixteenth century there have been records of this type of extreme phenomena, where, due to their catastrophic consequences, this extraordinary episodes remained engraved in the memory of the region's inhabitants with popular expressions, such as as "*la riada del 1907*" in Málaga or "*la avenida del 1973*" in the southern part of Andalusia. In particular, this historical event was analysed in Molina (1974), with very high intensities that produced catastrophic flows, even in dry channels, mostly to areas of Granada and Almería. Even this last dramatic episode produced by natural hazards in the area has been transmitted, and is still remembered by the population through oral tradition for instance. Scaling at national level, it must be highlighted "*las riadas del Válles*" in 1962 (BOTIJA, 1990). In Tab.2.2 a review of the most important recent flood events that affected the basins of our study is presented.

The extreme values are a challenge in any hydrological characterization of a water system for practical purposes. This characterization is always a difficult step in which an appropriate approach is to develop prediction tools capable of establishing different levels of alert in real time. In this sense, DB techniques can be adapted and applied easily. For this we must identify previous states of the forced agents and its classify each synoptic configuration, as they establish the different states.

TABLE 2.2: Historical flood events registered in south-eastern Spain

SC	Region	Date	Type	ST	Rainfall	Source	Moon phase
GH	Ma	21/10/2018	CO+HR	Cártama	-	-	2-3*days*
GH	Ma	03/12/2016	HR	Cártama		Own/ Press	1-4*days*
GH/GF	Al/Gr/Ma	28/12/2009	HR	-	-	Own	2-3*days*
GH	Ma	17/11/2012	CO	-	-	-	1-3*d*
GH	Ma	07/09/2015	CO	-	-	-	2+2 *d*
GH	Ma	14/11/1989	CO	NaN	150mm/3h	Barredo (2007)	2+1
GH	Ma	14/11/1989	CO	Cártama	230mm/24h	Molina (1990)	2+1
GF	Gr	19/10/1973	CO	Murtas	350mm/24h	Molina (1974)	3-1
GF	Al	18-19/10/1973	CO	Zurgena	730mm/48h	Molina (1974)	3
GH	Ma	24/09/1907	CO	NaN	NaN	Press1/Press2	2+3*days*

SC: Study Case, CO: Cold front, HR: Heavy rainfall, ST: Station location

Rainfall: Rainfall intensity. Al: Almería, Ma: Málaga and Gr: Granada.

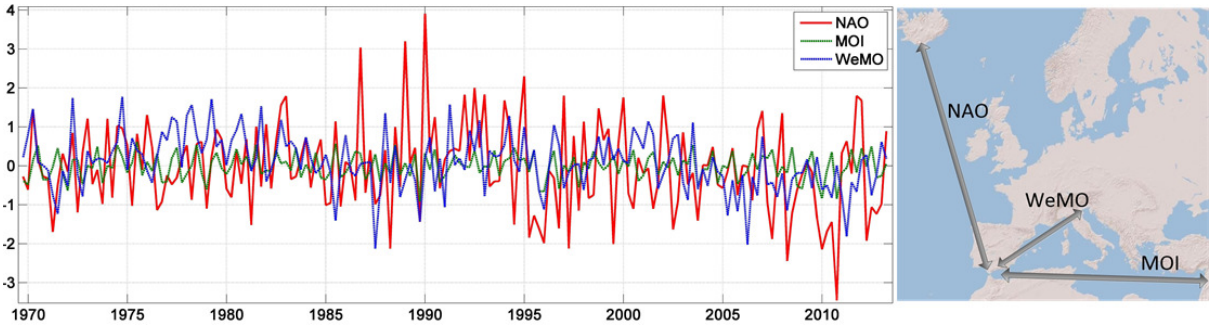


FIGURE 2.7: Quarterly mean values of three atmospheric oscillations

Atmospheric, Geophysical factors and astronomical factors

Across this work, three atmospheric oscillations will be used to test their influence on water resources, either precipitation and water flow, in the water systems under study. The atmospheric oscillations used in this thesis are: North Atlantic Oscillation (NAO), Mediterranean Oscillation (MOI), and the Western Mediterranean Oscillation (WEMO). The version of the NAO used is the normalized pressure difference between Gibraltar

(35°N,5°W) and SW Iceland (65°N, 20°W). The data can be downloaded from the NOAA webpage at monthly and daily timescale. Stull (2015) explains that " *its negative phase cause weaker westerlies that tend to drive the extratropical cyclones toward southern Europe*" thus producing a change on the Atlantic cyclone paths. In the southern part of Spain, the NAO has a clear influence over the occurrence of wet and dry periods. Rodrigo et al. (2000) show that extreme values of the NAO are related to droughts (positive extreme NAO) and floods (negative extreme NAO). Recent works (Egüen et al.) uses NAO to predict hydro-climatological variables (rainfall and streamflow) in a water system that flows into the Atlantic (western) part of Andalusia. Also the MOI (MOI, 2015) has been used as a predictor throughout this work, using its two versions (1. Gribaltar-Israel (30°N, 35°E) [MOI^{GI}], and 2. Cairo-Argiels [MOI^{CA}]). In Martin-Vide and Lopez-Bustins (2006) proposed a general rule for this index: if is negative the depression tends to move towards to the east-west direction and if it is positive to the west-east. Finally, the WEMO index dataset (Martin-Vide and Lopez-Bustins, 2006) is the normalized pressure difference between South of Spain (∼35°N,5°W) and North Italy (Padua, (45°N, 10°E)), which can be used for same purposes. Previous works, study the influence of these indexes on daily extremes of precipitation values (Vicente-Serrano et al., 2009), or about the rainfall variability on a monthly basis as is presented in Trigo et al. (2004). In Gonzalez-Hidalgo et al. (2009) indicate that the NAO has a lesser influence on the precipitation variability at a monthly basis than the MOI and WEMO indexes.

For Atlantic and Mediterranean fronts, the orographic conditions on the shoreline is a key factor in our region, acting as the first barrier for the storms when they come into the continental platform, normally from south-western or eastern horizons. The cold fronts sweep over the mountains where the water vapour and cools rises to higher geopotential layers creating heavy precipitation, extreme rainfalls and consequently floods capable to cause damage to life and property, even in intermittent streams.This also produces strong regional scale effects where flood events occur (i.e., Riada del 1907 in the city of Málaga) that was produced by a intense rainfall upstream in the basin even without having precipitated the rainfall in the downstream areas.

On the other hand, the season of the year is also a key factor with the most at risk season in autumn (Gelabert, 1983). Vallve and Martin-Vide (1998) show that the most frequent flooding season in eastern Spain, is September to November with October as the most frequent month among all the others. Molina (1987) mentions that the autumn term has been the preferred period for this type of events. This temporal occurrence, is well-known as the Guadalhorce River Basin authorities activate alert level when the hydrological

year begins, on September, what differs from the commonly accepted beginning of the hydrological year in the rest of Spain, which is October.

The most typical configuration for the occurrence of extreme rainfall events over this region of Spain corresponds to a cut-off low at high levels (Amengual et al., 2015). These type of events can produce daily values of more than 200 mm of precipitation. Martín et al. (2013) analyses heavy rainfalls in the area for 5 years of data (2006-2010), including some recent events exposed in Tab.2.2. This work concludes that after an atmospheric analysis two variables such as humidity flux at 850 hPa and the Lifted Index (LI) has influence on extreme rainfall events, although they are not decisive since other local conditions and the orographic components can be crucial.

The climatology that exists behind these extreme events is very complex and belongs to atmospheric dynamics and climatic processes. In order to simplify it, the regional extreme events will be classify into two classes: i) *Storms*, The heavy rainfalls are associated with low depressions, and ii) *Cold fronts* where very intense rainfall occurs due to cold isolated depressions that form in Sep-October when the northern air masses meet warmer southern air masses. This means that during risk seasons, warning tools are needed for catastrophic floods caused by this cold front phenomenon. For that we need to characterise the driving factors of the atmospheric dynamics and its meteorological processes.

Flood risk management and assessment

In flood risk governance, the first step in dealing with this natural phenomenon must be the common-sense approach. After that, the protocols to govern flood risk areas are: implementing high resilience development models on the floodplains, and coexisting with this natural threat. This ensures a sustainable anthropogenic activity in these regions. Meanwhile, the Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks (Flood Risk Directive, FRD), gives the guidelines to deal with floods at European level. Article 4 requires Member States to undertake a Preliminary Flood Risk Assessment (PFRA) for each river basin district. The identification of Areas of Potential Significant Flood Risk (APSFR, Art.5) will be based on available or readily derivable information, including the requirements specified in Art.4. The authorities completed the transposition of this Directive into Spanish national legislation the 23 December 2010 by law 903/2010. From that, Spanish water authorities set up flood hazard and risk maps for return periods of 10, 100, and 500 years for high, medium, and low probability scenarios.

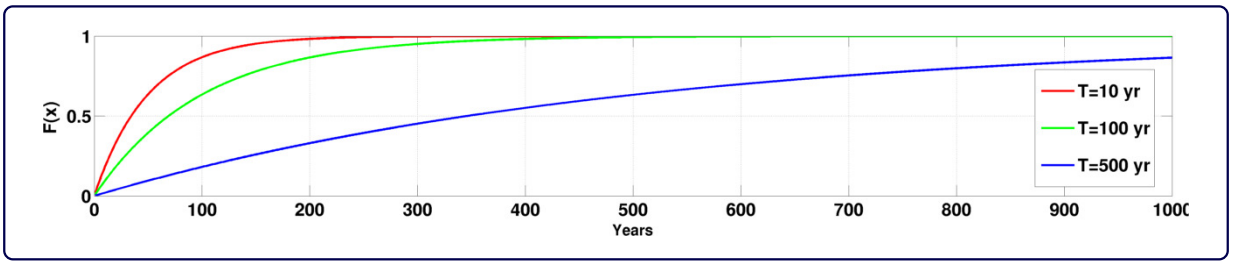


FIGURE 2.8: Cumulated probability F(x)=P(X≤x) for the different returns periods scenarios, established by the Spanish authorities following the European Directives

They base these recurrence intervals on the probability that the event will be equalled or exceeded in any year. For example, the probability for a 500 *yr* return period is F(x)= $1/\mathsf{T} = 1/500 = 0.2\%$, equals to 99.8% of probability that it will not be exceeded. This does not imply that two or more flood events with such intensity can not occur within the same year, since the return period is a statistical concept. If we compute F(x) for a given n (after) consecutive years with $1 - [1 - (1/\mathsf{T})]^n$, for 500 *yr* return period we going to get a 18.14% chance of occurring over 100 consecutive years.

We define the overall risk as the product of the probability of flooding and the cost of the damages, $R = \text{Cost} \cdot P$. Since the possibility of occurrence is random, which can not be avoided, the strategy must focus on the smart development and management of the floodplains under risk. The goal is to minimise a goal function, which are the costs of casualties of a flood event. Depending on the scale analysis, it relates these costs to land use for rural territories. In a more detailed scale, as used for flooding assessment in urban localities, it links the hurt to the building elements of a digital model.

Since ancient times these regions presented demographic pressure and an exploitation of territory for agriculture purposes. For that reason, land use change patterns and the effects on the environment are interesting research topics. To gain context, in the following paragraph we realise a geospatial analysis to asses the temporal evolution of flood vulnerability in the water systems under study.

• Flood risk areas analysis •

The regional quantification of the Land Use/Land Cover (LULC) temporal transformation is always an interesting study. This is relevant in the Andalusian territory since the existence in the past of different property bubbles with a wild urban expansion. This has caused non-conscious anthropogenic intrusion on the riverine areas where floods may

be recurrent. A spatial GIS survey of historical LULC data maps can evaluate this, visualising the occupation within the risk areas.

The zones of interest are the risk maps (e.g., T=10,100,500 and APSFR), maintained by the Andalusian water authorities through the REDIAM platform. These risk areas have been obtained through 2D physical simulations for the river conditions of a certain return period or extreme event frequency to comply with the FRD European requirements. The physical models used to generate the floodplains risk maps are: MIKEFLOOD, HEC-RAS & HEC-HMS and IBER. We can consult more information on the modelling procedure in their respective issues of the FRD.

The LULC GIS data for the characterisation of the land used are still served on the same platform belonging to *Sistema de Información de Ocupación del Suelo de España* (SIOE). Then, we analysed two spatial scales: one with a poorer resolution (1:25000) but with a better temporal record (1956, 1977, 1984, 1999, 2003 and 2007), and another with a higher detail resolution (1:10000) but lower temporal record (2005, 2009, 2011 and 2013). An example of the degree of the detail in this scale is seen in Fig.B.1 and Fig.B.2 at the river mouth of the GH and GF, respectively.

We classify the land uses according to the aim of our investigation. We want to quantify the recent regional development on the floodplains, and to check how the potential risk has evolved under different events scenarios. Finally, the surfaces have been grouped into the following five classes: 1). Urban and civil infrastructures, 2). Wetlands and water bodies, 3). Non-irrigated agriculture, 4). Irrigated agriculture, and 5). Forest and natural areas. This classification system it is, practically the same, as the hierarchical system used by the Andalusian authorities, except for a slight modification: we have separated the irrigated and non-irrigated agriculture. With this, we want to remark the evolution of irrigation agriculture in the lower areas of the valley and therefore with a better potential efficiency for irrigation. We show the class aggregation of the various codes for the two spatial scales analysed in Tab.B.2.

Flood risk analysis results: In this paragraph, we present the result of spatio-temporal variation within the Guadalhorce River Basin of the LULC on the flood risk zones. In Fig.2.9 the data of the lower spatial resolution but a larger time record are represented. We see how the percentage of urban and local infrastructures use has grown with special attention to the variation suffered between 1999 and 2003 for the risk area associated to T10, with an increase of approximately 10%. This type of action goes

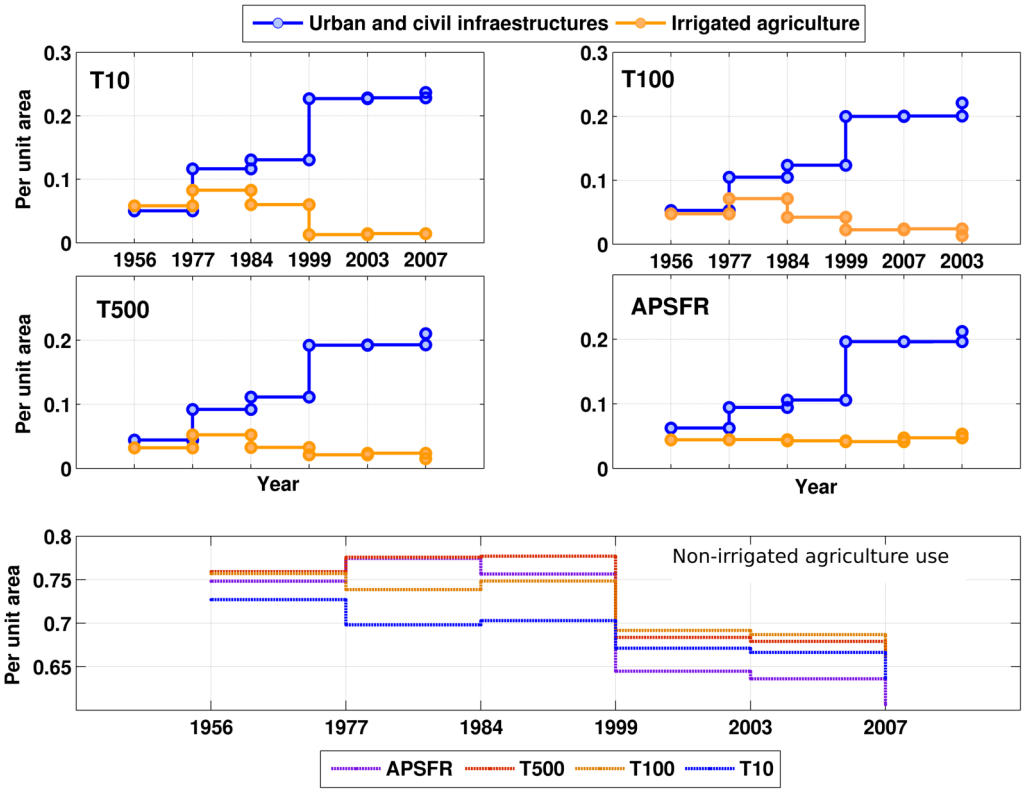


FIGURE 2.9: Representation of the temporal spatial evolution (1956,1977,1984,1999,2003,2007) of the flood risk areas within the Guadalhorce River Basin. The areas analysed are the risk flood areas under different assumptions: T10, T100, and T500 and the APSFR areas at 1:25000 spatial scale

against all logic since this is the area with a higher probability of occurrence. For the other areas, this change point also appears but to a lesser extent. We can also observe how non-irrigated and irrigated agriculture percentages decreases with a considerable change at the same change point. On the other hand, for the data scale in greater detail (Fig.2.10) we can observe a similar trend in the changes of land use within the flood risk areas, where a stabilization is observed starting in 2009. This is especially relevant in urban and civil infrastructures and in non-irrigated agriculture uses.

• Flood occurrence analysis •

The first and the easiest data exploration step is with the use of boxplots for visual summary of batches of data. As it is well known this representations it has statistical values as the extremes (outliers) observations, the upper and lower quartiles, and the median. This first visually give us a quick impression of the groups of data and if are directly related between them.

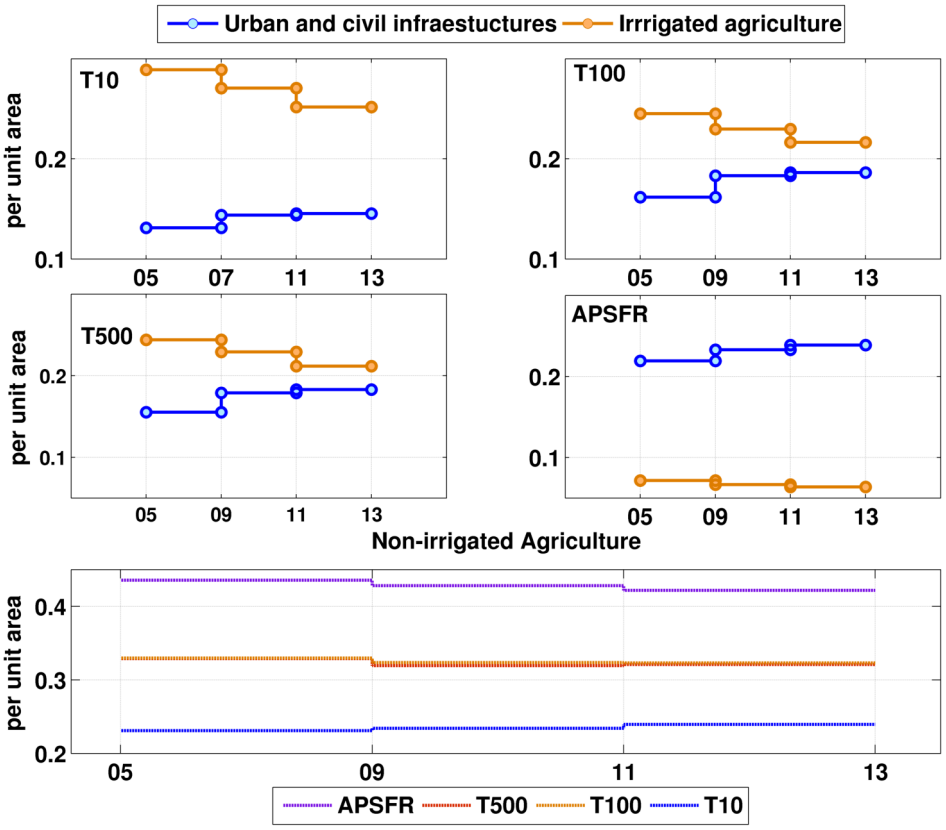


FIGURE 2.10: Representation of the temporal spatial evolution (2005,2009,2011,2013) of the flood risk areas within the Guadalhorce River Basin. The areas analysed are the risk flood areas under different assumptions: T10, T100, and T500 and the APSFR areas at 1:10000 spatial scale

In this section the occurrence of floods in the Guadalhorce Basin is analysed for the period 1996-2014. First, heavy rainfall events were identified as those events with a rainfall higher than 100 mm in 24 hours Senciales González and Ruiz Sinoga (2013). Then, the occurrence of heavy events was compared to the daily maximum water level measured close to the river mouth at Cartama (Fig.2.5.(b) and more detailed in Fig.2.1). The threshold for a flood event was established in 3.5 meters according to Regodon (2013), who studied the last two catastrophic flood events (15/02/2010 and 28/09/2012) through hydrological modelling in the basin (WiMMed, Herrero et al. (2012)). The local water authorities established this water height value as the level II alert (orange) starts, which is between 3.5 to 4.5 metres of water height level. Finally, the occurrence of flood events was similarly compared to the daily maximum discharge from the reservoir system (Fig 2.1), situated in the upstream part of the river course.

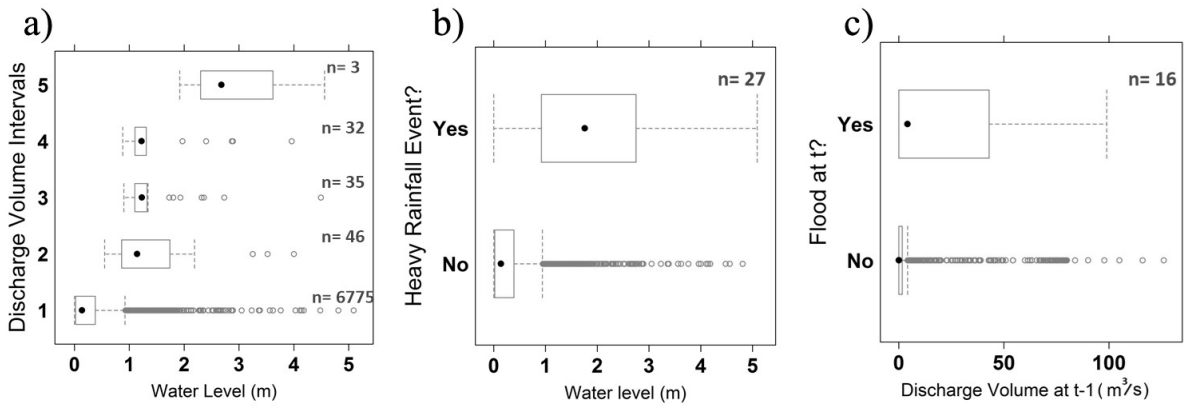


FIGURE 2.11: .a) Relation between discharge intervals and the daily maximum water height (m) measured at Cártama. Range intervals ($m^3 \cdot day$): [0,25.2] (25.2,50.4] (50.4,75.6] (75.6,101] (101,126]. b) Occurrence of heavy rainfall events related with the water level. c) Occurrence of floods related with time-lagged $(t-1)$ discharge values used

Flood occurrence results: From Figure 2.11.a, no correlation was found between the volume discharge intervals and water levels. The three highest discharge events (Fig.2.11.a)) with a mean value of 40 $m^3 \cdot s^{-1}$ (5 times over the annual mean value) corresponded to the flood events occurred on February 2010, an extremely wet year. The annual precipitation of that year nearly doubled the annual mean. At least on 27 occasions there was heavy rainfall and 16 floods in the study period (Fig.2.11.b) and Fig.2.11.c). Thus, some of the heavy rainfalls were attenuated by the system and did not generate significant floods in the basin. The interquartile range of the occurrence of both heavy rainfall and flood events was slightly displaced towards increasing water levels (1-3 m) and discharges (up to 50 $m^3 \cdot s^{-1}$). However, the maximum discharges ($> 100 m^3 \cdot s^{-1}$) were found under the no-flood condition (Fig.2.11.c). Similarly, high water levels were also found for non-heavy rainfalls (Fig.2.11.b).

In the current state of regulation of the GH basin, after episodes of heavy rainfall, floods often occur that the dams cannot retain. Even so, some of the heavy rains were attenuated by the system and did not generate significant flooding. This statement comes from not finding a significant correlation between discharges and floods. However, a deeper analysis with longer data series needs to be carried out. Nevertheless, in the Guadalhorce Basin a proper land use planning within the flooding areas may be a more effective risk reduction measure than the optimization of the hydraulic regulation structures.

2.4.3 Droughts

Hydrological drought is complex in terms of its causing factors and impacts on ecosystems and society (Van Loon, 2015). Droughts in Mediterranean environments are an important factor, as they are particularly sensitive to this phenomenon (Luo et al., 2015), and in their water systems (river basins) it is common to find a deficit in rainfall (PD). This can be computed as: PD = R - ETo, where R is the total rainfall and ETo the total evapotranspiration at monthly scale, the evaporative demand.

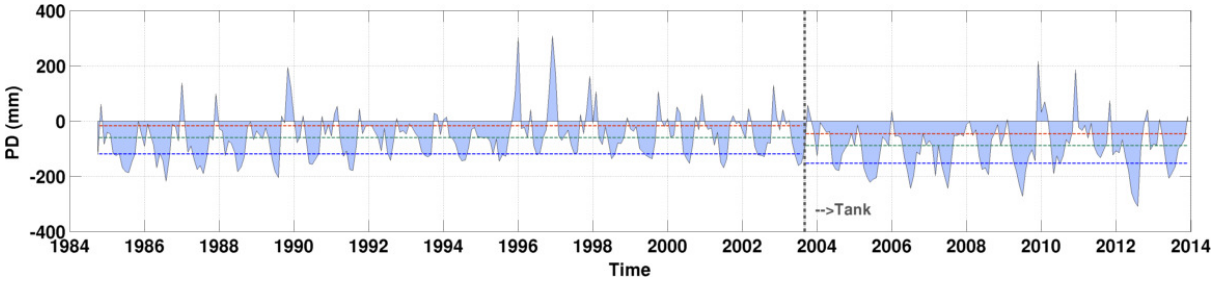
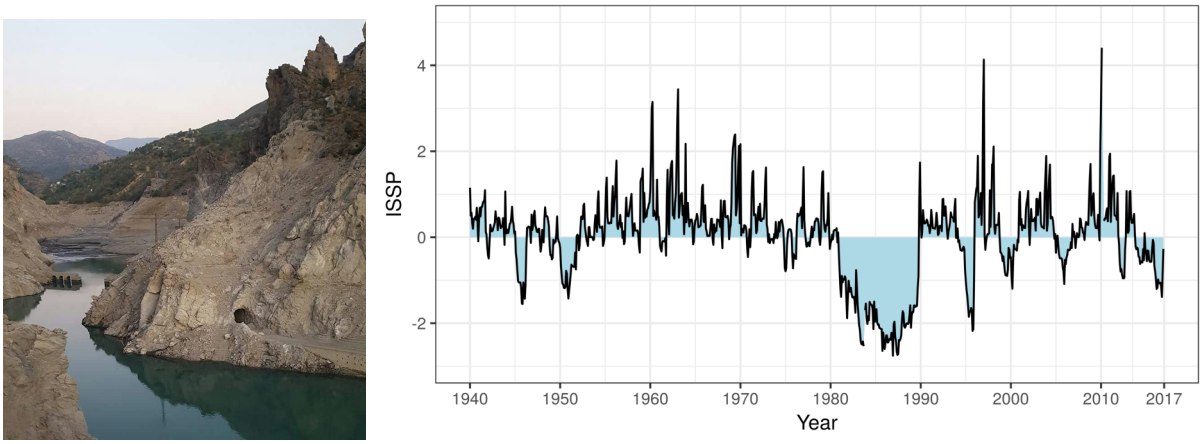


FIGURE 2.12: Monthly Precipitation Deficit (PD) for collected data at Beznar reservoir

In Fig.2.12 data collected at Beznar reservoir is represented, which is situated within the Guadalfeo River Basin. In 09/2003, a change of scale can be observed, marked by a vertical dash line. It indicates a point of scale change because the method of measuring evaporation was changed. Until then, evaporation was measured with a Piche evaporimeter, based on the evaporation of blotting paper, switching to a pan thereafter. This affects the scale and accuracy of the measurements. Some statistics have been drawn to characterize this change, representing the 25th, 50th and 75th percentiles. It is observed that this point is in deficit (below zero) for almost the entire period of time observed, being more intense with the data collected by the tank. The statistics represented has values of -17.1 mm, -59.7 mm and -119.15 mm with the Piche evaporimeter, and -45.7mm, -88.6 mm and -152.6 mm for the tank evaporimeter for the 25th, 50th and 75th percentiles respectively.

Whether because of natural causes or by human action, the droughts episodes seem to have turned into more frequent. In Art.4.6 of the WFD relates to events which could not have been foreseen: "*Temporary deterioration in the status of bodies of water shall not be in breach of the requirements of this Directive if this is the result of circumstances o natural cause or force majeure which are exceptional or could not reasonably have been foreseen, in particular extreme floods and prolonged droughts*". Locally, the water authorities present the Special Plan in Situations of Alert and Temporary Drought (Andalusian Drought

Plan, PES), reporting the archival pluviometric droughts. We feature that the most recent long-term episodes turned up between the hydrological years 1979/80, 1984/85, and longer durations that popped up in the 1990s and completed in November 1995.



(a) An acusated precipitation deficit can cause complications for urban demand.

(b) Anomaly index for the Mediterranean RBD. *Source:REDIAM*

FIGURE 2.13: Droughts are recurrent events in the study sites which can have devastating environmental and social implications.

For drought characterisation, it is indispensable to describe the discrepancies between Aridity Indices and Drought Indices. After the bibliographical review of the descriptors, we post the following terms: *i) Aridity Indices*. These indices give us an idea of the dryness of the area, which is a climate characterisation. The aridity is a function of both precipitation and the potential evapotranspiration rate (ET). From an efficient water management perspective, aridity indices have little importance as is suggested by Maliva and Missimer (2012). Since it is a better qualitative indicator of the climate than a quantitative one, being not an aim in this work. ii) *Drought Indices* (DI). We use these indices for general drought characterisation and drought severity quantification. This characterisation can have into account varied data sources as Rainfall (R), Runoff (SF), Temperature (T),. . . , etc. These indexes are a sensitive drought measure, and standardised in a numerical manner. Real actions analysis on forecasting calculation is an aim of this work (Sec.5.4).

The simpler index is the rainfall deciles (Gibbs, 1967), which establishes a statistical ranking based on the average values observed in a location. In Mishra and Singh (2010) a review of the drought concepts, foundations and several indexes are presented. In this direction we can employ various DI, where the Standardized Precipitation Index (McKee

et al. (1993b), SPI), Standardized Runoff Index (Shraddhanand and Wood (2008), SRI) and Standardized Precipitation Evapotranspiration Index (Vicente-Serrano et al. (2010), SPEI) are the most important. The SPI is used and explained more carefully later in Sec.5.4. We calculate the SRI as the SPI, but using the runoff instead of the precipitation, being more convenient for hydrology applications. The SPEI could be view as an extension of the SPI index, which takes into account both precipitation and potential evapotranspiration (ET) for its determination. Future drought estimations can be done in different ways: we can use techniques as the Bayesian Networks for spatio-temporal drought predictions found in Madadgar and Moradkhani (2014), or study the frequency of a drought index as Santos et al. (2011) did in Portugal. In Sujay and Paresh (2014) we can find an extended review of a DD application for drought modelling. As is exposed in Belayneh et al. (2014), droughts can be modelled with univariate models (ARIMA). We can find examples on regression forecasting with DD in Mishra and Desai (2006); Modarres and Ouarda (2014); Djerbouai and Souag-Gamane (2016), and drought prediction in the Amazonas basin by Lima and AghaKouchak (2017), adopting the Palmer Drought Severity Index (PDSI). To gain the full drought perspective on this matter refers to Mishra and Singh (2011).

Drought Characterisation

The Standardised Precipitation Index (SPI, McKee et al. (1993a)) is a statistical index that takes into account the rainfall collected during a n time, at a monthly scale, with the rainfall cumulative distribution for the period examined. Thus, various probability density functions can fit the frequency distribution of the total rainfall. A key point of this index is that it is truncated to the extremes in a normal distribution, which does not mean that the distribution of monthly rainfall is not following other types of probability functions, as it normally does. Stagge et al. (2015) analyse several candidate probability distributions at European scale, with emphasis on skewed precipitation distributions, i.e. Mediterranean lands, where the Weibull and Gumbel showed lower rejection frequencies in the normality tests. Ayuso et al. (2015) apply the Gamma distribution as the probability density function a region under study in this work, the city of Málaga.

In Tab.2.3 we specify the standardised means of the SPI with its probability. Because we standardise the SPI, we expect these percentages from a normal distribution of the SPI. For that reason, a goodness-of-fit test is always a mandatory step when we work with it (Lloyd-Hughes and Saunders, 2002; Stagge et al., 2015; Svensson et al., 2017).

In the territory under study, the water authorities represents the meteorological droughts by a rainfall anomaly index named Standardized Drought-Precipitation Index (IESP). In Pita López (2001) and Sanchez-Toribio et al. (2010) this index is well described. According to Donaire (2007), this index follows a normal distribution. This same index is also analysed in Gallardo et al. (2016), comparing it with the SPI and SPEI indexes for the 38 grid points of the Andalusian territory. In the succeeding table, we list its restricted values.

TABLE 2.3: SPI values and the corresponding drought classification and event probability (Lloyd-Hughes and Saunders, 2002)

SPI	Classification	Probability (%)
SPI≤ -2.0	extreme drought	2.3
$-2.0 <$SPI≤ -1.5	severe drought	4.4
$-1.5 <$SPI≤ -1.0	moderate drought	9.2
$-1.0 <$SPI≤ 1.0	normal	68.2
$1.0 <$SPI≤ 1.5	moderate wet	9.2
$1.5 <$SPI< 2.0	severe wet	4.4
$2.0 \leq$SPI	extreme wet	2.3

TABLE 2.4: IESP classification values for drought characterisation (Gallardo et al., 2016)

IESP	Classification
IESP ≥ 0	No Drought
$-1 <$IESP≤ 0	Moderate drought
$-1.7 <$IESP≤ -1.0	Severe drought
$-2.3 <$IESP≤ -1.7	Extraordinary drought
IESP ≤ -2.3	Extreme drought

2.5 Conclusions

From this chapter the following conclusions are established:

- The correct hydro-meteorological monitoring through the establishment and conservation of meteorological stations, provides us the opportunity of its subsequent characterisation and modelling from these observations. This is the first step for computational tools development to simulate hydro-meteorological processes of interest on a local/punctual scale based on the historical data.
- The potential flood risks have escalated in recent years because of the urban development intrusion into the floodplains.
- By analysing the extreme events occurrence in the Guadalhorce River Basin, it can be observed that upstream reservoirs have little capacity for regulating floods in some intense rainfall fronts.
- Droughts can be characterised by different Drought Indexes. These indexes are fitted to several probability distributions that need to be tested in each local/site specific drought characterisation.

Chapter 3

Hydrological Data-Driven Modelling

3.1 Introduction

Within the hydrological modelling it is necessary the recognition of hydrological frameworks and the patterns of its hydro-meteorological variables for the application of two fundamental tasks within the hydrological engineering: Forecasting and Optimisation.

In this sense, Data-Driven ($\mathcal{D}$D) modelling offers us various methods to choose from. Our goal is to elucidate what type of computational techniques work best under Mediterranean conditions under a comparative framework. In this understanding, Sec.1.4 shows a review of significant past hydrological DD's modelling actions under similar hydro-meteorological conditions.

To do this, we need to analyse the main differences between the diverse approaches and types of modelling that can be found. In this design analysis we will move through different approaches: under frameworks of different mathematical and/or physical nature, under the parametric nature, under the dynamic nature, under the linear nature, under the probabilistic nature and under the single or the multiple approach, which is carried out in Sec.3.2. Section 3.3 discusses data management, presenting the steps in data partitioning with a focus on improvements in data resampling and pre-processing for successful $\mathcal{D}$D experiences. The quality assessment of the modelling experiences in this activity is made by interpreting several evaluation coefficients, which are shown in Sec.3.4. In Sec.3.5 we spell out the conclusions that have emerged from this chapter of methodological foundations.

3.2 Hydro-metorological Modelling Approaches

"All models are wrong, but some are useful"
~ George E. P. Box ~

The hydrological science uses models to understand the water system under consideration. Hydro-meteorological tools and modelling approaches do not mean necessarily complex mathematics or computer programs. In all cases to develop these models in hydrology one needs to complete four basic methodological stages: 1) Collecting data, 2) Understanding the processes in the water system, 3) Generalising well in new situations and boundary conditions, and 4) Predicting future water system states.

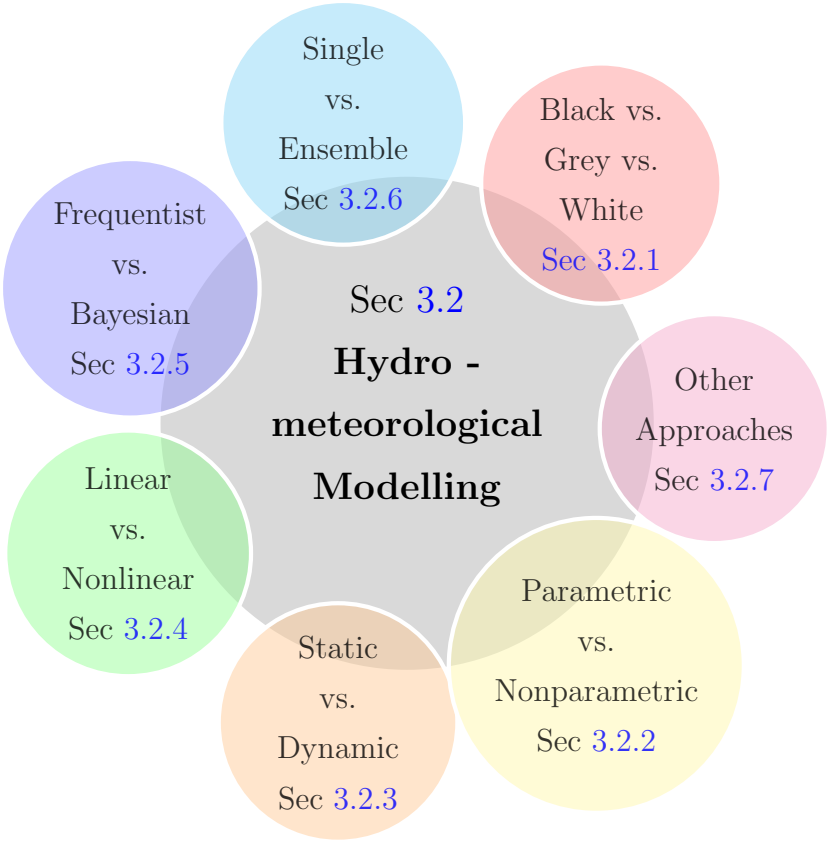


FIGURE 3.1: Hydro-meteorological modelling approaches

The modelling application will conform two types of approximations to face the challenge under consideration: process-oriented, which is more idealised, or problem-oriented, which is more pragmatic. The orientation will depend on the study objectives and the practicality of the model. In this section, we analyse in a comparative way the models, in approaches and computational techniques, that we can apply to describe/identify the

selected water systems. This assessment will point out the interests and disadvantages of particular concepts for the development of such models for subsequent applications.

3.2.1 Black vs. White vs. Grey

■ □ ■ *"Which modelling approach should we apply? "* ■ □ ■

We can classify the hydrological models according to their physical knowledge and the interpretability by their users, as seen in Fig.3.2.(A). The first models are the empirical models or *"black box"* models. One weakness of these applications can be the well-known lack of interpretability in the relationships between inputs and outputs, which are based on a set of parameters (e.g., ARIMA, MLR or simple ANNs). Although these opaque models can perform with precision, it is necessary to focus on frameworks of a *"grey box"* nature under a probabilistic approach.

The *Physically-based* or the *"white box"* approach are process-based models, reproducing the physics of the hydro-meteorological processes in order to emulate the entire hydrological cycle by computation. The physically based models deal with the study of specific water processes in a more detailed way than other models that are more simplified (Fig.3.2 (B)), which commonly have limitations within the spatio-temporal scale. This limitation is a direct consequence of the simplification of the water and hydrological processes. A recognised problem of these tools is that they require more data and, more time for the users-training. The "*grey box*" concept is exposed in Hauth (2008) as "*Grey-box modelling deals with models which are able to integrate the following two kinds of information: qualitative (expert) knowledge and quantitative (data) knowledge, with equal importance*". In fact, this approach involve models in-between the other two approaches: white- and black-box models, where the line of distinction between them is often blurred, being difficult to establish precisely when it is one approach or another (Fig.3.2.(A)).

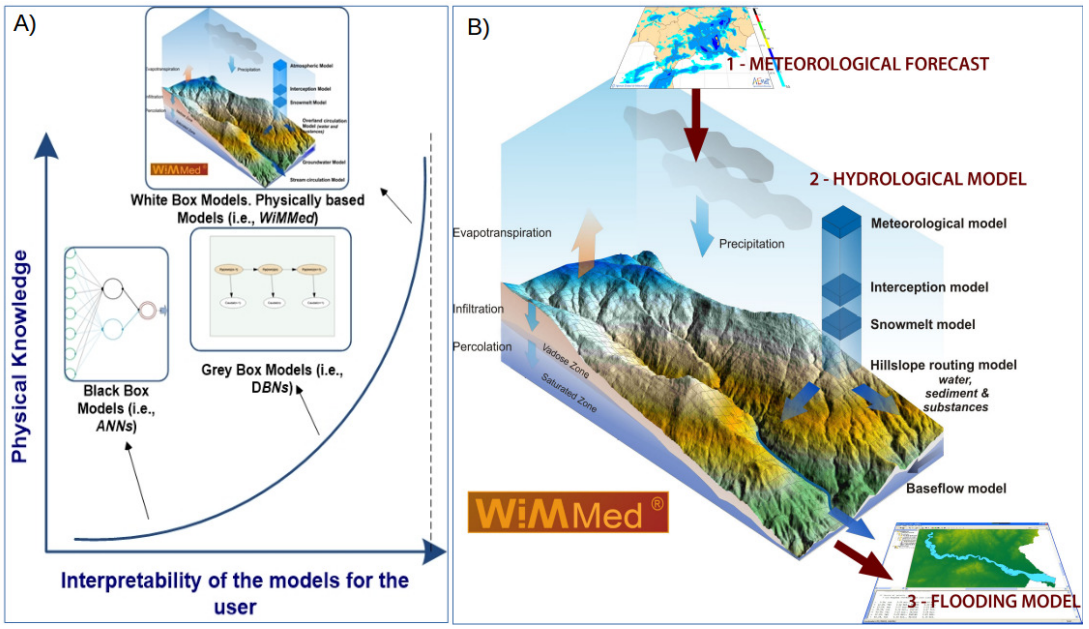


FIGURE 3.2: A) Graphical classification of different modelling approaches, adapted from Giustolisi et al. (2007). B) Flooding forecast as a result of the cascade of physically-based models and a representation of the Physically-based hydrological WiMMed scheme model, from Polo et al. (2009).

For that reason, there is a necessity to migrate towards methods with clear transfer functions (i.e., covariance functions, physical modelling) and which are based on less parametric frameworks. For instance, the joint use of ANNs and the physical models (Abrahart et al., 2012), can offer good options in real applications and allow us to reduce the number of predictors. This also reduces the model complexity and promotes the interface interaction for the non-expert users. For instance, coupling data-driven forecasting with hydrological models may make up a robust tool for alert assessment. We should not rely flood warning alerts on a single model (Schumann et al., 2011), not even on models of a single type. Forecasting with DD models presents the advantage that, once trained with different boundary conditions, they are fast and can give multiple results for ensemble applications (Sec.3.2.6). Another advantage of DD is that a probability associated with predictions can be generated, as under the Bayesian paradigm. This practice is appropriated for risk analysis. Under this approach we can also feed back the model with new data observed through the evidence procedure. On the other hand, the physical models are theoretically prepared to cope with new or more complex situations, and can provide results with a detailed spatial distribution, such as flooded areas.

Within our research group, the Watershed Integrated Model in Mediterranean Environments model (WiMMed) which is a physically-based, distributed hydrological model has

been developed. It uses hourly and daily meteorological data, along with certain physical properties of the soil and subsoil to perform the spatial interpolation and temporal distribution of meteorological variables, rainfall interception, snowmelt, infiltration, runoff, surface slope circulation, calculation of the water in aquifers and basin flow circulation. Thus, it provides the instantaneous value or evolution of the principal flows and state variables, such as water flow volumes, amount of stored water, flooded surfaces and so forth. A detailed description, more info and the theoretical basis of WiMMed is given in Herrero et al. (2012). We can find application examples of this model in the Guadalfeo River Basin in Herrero et al. (2009), Millares et al. (2009) and Aguilar et al. (2010).

In this work, methods employed with a black nature are Multiple Linear Regression MLR (Sec.A.1.1) and simple ANNs (Sec.A.1.2). In the same way, methods applied under the "white nature" approach are WiMMed simulations, which is calibrated and validated for the study site, and the time series of global circulation models (GCMs). Finally, modelling classified under a more "grey nature" are BNN (Sec.A.1.2) and GPs (Sec.A.1.3).

3.2.2 Parametric vs. Non-parametric

Generally *parametric* models are driven by some finite set of parameters (e.g., θ, w), given that the future predictions of a target variable, i.e., Streamflow (SF), are independent of the observed data $\mathcal{D}$. On the other hand, a *non-parametric* (NPar) model assumes that the data distribution cannot be defined in terms of such a finite set of parameters, but they can often be defined by assuming an infinite set. Murphy (2012) defines non-parametric methods as "methods in which the number of parameters grow with the amount of training data". In Russel and Norving (2010) is defined as the model that cannot be characterized by a bounded set of parameters. For instance, in Gaussian Processes, data distribution is driven by a covariance function and a set of parameters θ, and is defined by Ebden (2015) as a "less parametric" model. Extended details of last this computational technique can be found in App.A.

Then, what are the advantages and disadvantages of the respective models? Well, the parametric model are faster to train, simpler and normally needs less data for the parameters estimation. In contrast this models are highly constrained to the specific model form and the mathematical expressions that drives the process. This makes them less flexible to complex data and may have poorer performances than the non-parametric models.

Examples of parametric models are the well-known polynomial regression or Multiple linear regression (MLR, Sec.A.1.1), Hidden Markov Models (HMM), simple ANNs and K-means among others. Among the non-parametric models we find Local Regression (LOESS) for non-linear regressions. On the other hand, computational techniques under a non-parametric nature are Bayesian Neural Networks (BNN, Sec.A.1.2), K-Nearest Neighbour (KNN, Sec.A.1.4), Gaussian Processes (GP, Sec.A.1.3), Lévy Processes (Gardiner, 2009), and Dirichlet Process among others.

A very interesting and representative case is the transformation from a parametric black box model as the simple ANNs to a "non-parametric" as the ANNs under the Bayesian framework exposed by MacKay (1992a). In that paper, the Multilayer Perceptron (MLP) structure, which is 100% a parametric black box model, is transformed to a more grey nature when additional parameters, called hyper-parameters, are added. These hyper-parameters include the bayesian/probabilistic approach through the evidence procedure.

3.2.3 Static vs. Dynamic

The most notable difference between static and dynamic models for a given water system is that while a dynamic model refers to runtime model of the system, static model is the model of the system not during runtime. Most of the dynamic models implicitly contain a static model. Another difference lies in the use of differential equations in dynamic model which are *conspicuous* by their absence in static model. Dynamic models keep changing with reference to time whereas static models are at equilibrium in a steady state. This difference for correlated data, time series, can be expressed in the simpler form as:

$$Y_{(t+h)} = f(X_{(t)}) \tag{3.1}$$

when is a static model which depends only on one state X_t, and a model form as

$$Y_{(t+h)} = f(X_{(t)}, X_{(t-1)}, \ldots, X_{(t-p)}) = f(\{X_{t-p}\}_{p=0}^{+\infty}) \tag{3.2}$$

when is a dynamic model. The main difference is that the future states of the dynamic approaches depend on just one state, let's say t, and one or more previous/antecedent states, $t-1, \ldots t-p$. Static models are more structural than behavioural while dynamic model are a representation of the behaviour of the static components of the system. Static modelling is more rigid than dynamic modelling as it is a time independent view of the

system. It cannot be changed in real time and this is why it is referred to as static modelling.

In probabilistic models, traditional Bayesian Networks usually are under a static framework. The nodes are related through the inference structure by the parameters, which are the conditional probability of the occurrence of the different classes of the variable. On the other hand the dynamic framework (i.e., Dynamic Bayesian Networks) are the same graph idea but under a given (1,2) Markov assumption which is the number of lagged dependencies within the graphical model.

In this work, a comparison of various computational techniques for short term forecasting under a dynamic framework are analysed in Chapter 4. On the other hand, a static framework for volumes forecasting in the form of cumulated streamflow at medium term, is developed in Sec.5.2. Also a dynamic approach is implemented for Rainfall and Droughts medium term forecasting in Sec.5.3 and Sec.5.4 respectively.

3.2.4 Linear vs. Non-Linear

"Non-linear interactions almost always make the behavior of the aggregate more complicated than would be predicted by summing or averaging"

∼ John Henry Holland ∼

In mathematical terms, linear means the use of polynomials of degree one to represent the underlying process (e.g Eq.A.1). Non-linear processes occur when the variations of the target variables are not proportional to the variations of the input variables, by the use of more than one mathematical order, and/or by the use of non-linear transfer functions. Additionally, Data-Driven (DD) approaches can be classified into this two main domains: linear and non-linear modelling (Golestani and Gras, 2014), which can be extrapolated to the three fundamental problems that we may deal in system analysis and engineering: Regression, Classification and Optimisation.

Generally, water and environmental systems are non-linear, although there may be a linear relationship between system variables and related hydrological. For instance, in Basara and Crawford (2002) the root-zone soil moisture is linearly related with atmospheric processes. Neiman et al. (2002) established the linear correlation between upslope flow and rain rate at a coastal basin in California. Another important relationships within the hydro-meteorological time-scales are lineal, for instance the relationship between the

rainfall and the interdecadal or interannual variabilities (Lu, 2003). This underlying time-scale components and forcings, have to be taken into account during the identification procedure.

An important component in the mountainous basins are the baseflow recession curves. In Wittenberg (1999) the storage outflow relationship is computed with an exponential function which implies that the aquifer reacts like a single linear reservoir where storage is proportional to outflow. Also a previous study (Millares et al., 2009) in the GuadalFeo basin (description in Chap.2), analyses the linear and nonlinear behaviours of the baseflow recession, where the Rainfall-Runoff relationship in this basin may have a strong non-linear component produced by the snow processes and the fractured underlying rock.

In optimisation problems, the linear domain is when the local optimum is the global minimum not being necessarily a global optimum for non-linear domains. Lineal optimisation methods as: Gradient Descent (GD), Linear Programing (LP), Quadratic Programing (QP), Dynamic programing (DP) and the Least Square (LS) are some of them. Advanced methods as Genetic Algorythm (GAs) or Particle Swarm Optimisation are also strongly recommended for nonlinear domains.

Understanding the terms "linearity" and "non-linearity" according to the first definition in Sivapalan et al. (2002), linear relationships are commonly simpler and easier to identify than non-linear ones, and therefore involve fewer computational requirements for modelling purposes. In hydrological regression problems, the conversion of rainfall into runoff is one of the most relevant process, which is generally, to a greater or lesser degree, non-linear. This will depend on its geomorphological complexity, the presence of vegetation, the influence of snow, etc.

In Chap.4 linear and nonlinear methods are applied, quantified and compared for forecasting purposes. These are necessary steps to gain knowledge on the dynamical behaviour of the Rainfall-Runoff (RR) response in the water systems under study.

Both in one: In this matter, we must comment that there is a DD' kernel technique that relates inputs and the target vector through covariance terms, called Gaussian Processes (GP, Rasmussen (2004)). It must be highlighted that one of this method's strengths is that it allows the joint use of transfer functions, defined as kernel covariance function (KCF), with different nature but under the same methodological framework. For instance, one GP model can be formed by a linear KCF and by a non-linear KCF. For more details and

its foundations refers to Sec.A.1.3 and practical applications and results refer to Chapter 4.

3.2.5 Frequentist vs. Bayesian

"The subjectivist (i.e., Bayesian) states his judgements, whereas the objectivist sweeps them under the carpet by calling assumptions knowledge, and he basks in the glorious objectivity of science"

~ I.J. Good ~

In complex environments, as water systems, the probabilistic approach gives us the possibility to deal with the uncertainty attached to it. The traditional reasoning, also probability, called "frequentist" refers to past events. This traditional approach is very objective and the underlying parameters remain constant.

In that sense, Bayesian applications has increased over the last decades over different engineering/scientific/technical fields, as it allows to the researchers/engineers to factor expertise and prior knowledge into their computations (Malakoff, 1999). Fundamentally Bayesian means probabilistic with a prior distribution or "belief", where the evidence can be particularly important. Then, as more data is obtained, we can use it to update our belief. The result of this new distribution would be called the posterior distribution, and by repeating the cycle as new data or evidence appears, then the old posterior becomes a new prior; and the cycle is repeated successively.

The main idea of this sequence, in which Bayes' theorem is used for statistical inference, is to update the probability of a hypothesis as more evidence or information becomes available. The Bayes' Rule follow the next expression:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \tag{3.3}$$

where $P(H)$ is defined as the probability of the initial hypothesis and $P(E)$ is defined as the probability of the observed evidence. A good introductory book for Bayes' Rule explanation can be found in Stone (2013). In practice, in Bayesian machine learning we use the Bayes' rule to infer model parameters (θ) from data (D) probability distributions in the following form,

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{3.4}$$

where $P(D)$ is defined as the probability inferred from data and $P(\theta)$ is defined as the probability distribution of a given parameters set. This is the basic approach and the underlying structure behind many Bayesian DD models, which is widely explained and analysed in Bishop (1995) and Nabney (2004).

One of the disadvantages of Bayesian models is the influence of the priori knowledge of the probability of the process to be characterized. Since it has a great influence on the final result, a more expert knowledge in modelling is required than in traditional approaches (i.e. the frequentist inference), in which it can be directly applied.

In this work, the DD techniques used under the Bayesian approach are Gaussian Processes (Gaussian Processes (GP), Sec.A.1.3), Bayesian Neural Networks (Bayesian Neural Network (BNN), Sec.A.1.3). The main difference between this two techniques is how they tackle the bayesian inference: for BNN is approximate while for GP is exact (Rasouli et al., 2012) through the use of covariance matrices.

Under the multiple approaches for assembling various models in one outcome, also a Bayesian framework is applied in this work, called Bayesian Moving Average (BMA), which the perspective approach is exposed in the following section and its foundations in App.C.1.

3.2.6 Single vs. Multiple

"While one's subjective experience is true, it may not be the totality of truth"
$\sim$ Anonymous $\sim$

In regression and classification problems, a single deterministic outcome is always a weak approach and a poor solution from a practical point of view. An effective strategy may be working with multiple models or values to reduce the risk of biased results and conclusions. Having a range of possible values allow us to have a better general idea of the model's behaviour. In other words, we would have a distribution of values.

In hydro-meteorological forecasting and engineering experiences it is also required uncertainty and risk analysis, so is necessary to estimate a range of possible values, an interval, either through the integration of multiple deterministic values or thorough classical probability inference and statistical models. In theory, the majority of novel computational techniques associate an error interval, confidence levels or a distribution of the possible

outcomes to a deterministic or a mean value. Additionally, this interval can be viewed as a quality indicator of the results and the modelling performance, since the higher the errors associated with the predictions, the greater the uncertainty, the lower the quality of the model and the weaker the recognition of the pattern. In this work, the Average Width (Section 3.4) is used as a coefficient to evaluate this quality criterion.

Due to the complexity and the uncertainty into the hydro-meteorological patterns, multiple models/runs with different boundary conditions, input variables or parameter values among others, are always a recommended approach. These variations will produce differences in results, which provides a randomness to these forecasting procedures. In that sense, the collection of runs/simulations/models is called an ensemble. The multiple runs can be assembled in a final result having a probability distribution and deterministic outcomes. These distributions capture the uncertainty associated to the different sources of randomness: multiple predictors sources (e.g. bagging), different training datasets generated by resampling, different boundary conditions (e.g with and without specific forcing agents), and differences on the tuning parameters (e.g ANNs multiple simulations, see App.A.1.2).

3.2.7 More Modelling approaches

Univariate vs. Multivariate. Differences between univariate and multivariate can often misconceived. When the regression is related to multiple variables is a multiple regression, but when is nested by a joint distribution then it is transformed to a multivariate approach. In this work, multivariate regressions are developed in the Gaussian Processes GPs experiences, where the data is fitted to a multivariate Gaussian distribution which can be visually represented by the real functions obtained (more details in Sec.A.1.3). In contrast, multiple regressions have been applied in Chap.4 and Chap.5. On the other hand, univariate means that your data has only one variable.

Deterministic vs. Stochastic. In time series, the deterministic approach are models that generates a single value for a given time step t, while the stochastic approach refers to models that finally generate a list of possible values for t: i.e., a distribution, an interval, an error bar. The inference beyond the interval quantification is associated with hypothetical confidence in which to evaluate whether or not these assumptions are fulfilled. Also applying resampling approaches can transform deterministic models into more stochastic models (Von Storch and Navarra, 2013).

Open Source vs. Commercial Software. Open source is generally free, being more adaptable as you are not locked to a particular approach or to a specific proprietary firm. It also benefits from continuous improvements, bug fixes, and ultimately the richness of a large community of developers. Among the disadvantages is the fact that developers usually pay less attention to the user interface, which is usually less "friendly". This approach also has less support for end-users when setbacks occur during regular use. As for commercial software, it should be noted that it requires less technical knowledge, has a useful user support department usually within the company, and reliability is maintained since there is no unauthenticated modification of the source code.

3.3 Data Management and Preprocessing

"In God we trust. All others must bring data"
~ W. Edwards Deming ~

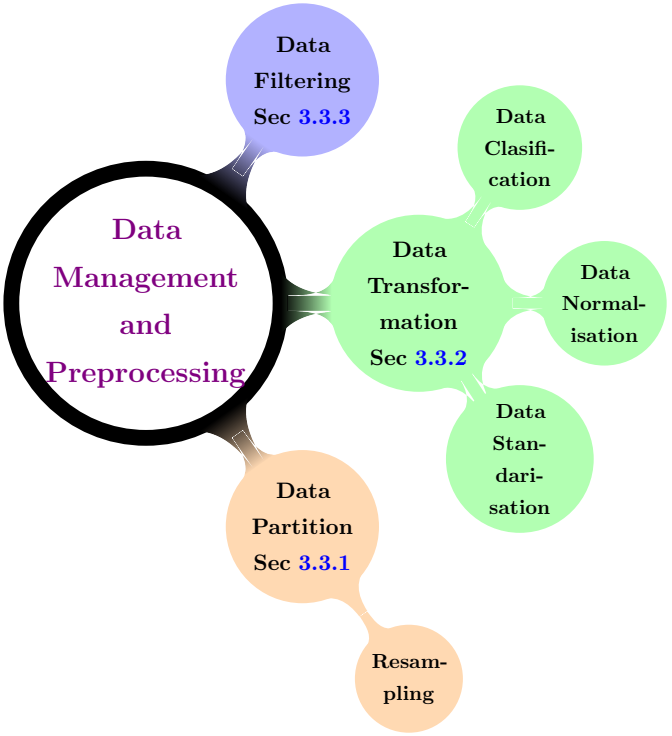


FIGURE 3.3: Data analysis

Data-Driven methods under a supervised framework are very sensitive to how the data are managed and preprocessed. For example, conclusions can not be drawn without quantifying how representative it is the data sample used for the training period. Also

in most of the computational techniques, data transformation is a necessary step for proper functioning, besides being a mandatory methodological step when using data with different units and scales. In this manner we can use and compare data from different sources in a balanced form.

3.3.1 Data Partition

DD methods requires a data division into subsets for the different methodological steps. This division can bias the results and consequently we can draw erroneous conclusions from it. The data partition should allow us to establish batch of data with sufficient quality to perform each methodological step with success: to achieve a proper convergence during the model optimisation period (training), and representative values (more frequent and extreme) of the possible states of the water system for the validation of results (validation). In this work, the time series sets were divided into percentages to provide data for each methodology period depending on the number of the methodological stages needed.

Excessive use of data leads to problems of overfitting due to excessive convergence during training and produces poor generalization for unseen data. In Chapter 4, the 80:20% split ratio of the dataset is applied for cases where there are only two methodological stages: Training and Validation. More Data Based modelling experiences are found in Chap.5. In Sec.5.2 the percentage for the classifiers comparison is 60-40% for the Training (T) and the Validation (V) period respectively. In Sec.5.3 the percentages are again 60-40% for Training (T) and Validation (V) stages respectively.

Apart from that, the inherent uncertainty that exist and the bias produced by the election of one part or another of the data subsets, forces us to apply resampling techniques (Efron, 1979) to be more accurate in the evaluation. Resampling techniques are useful for operational purposes and for the application of the tools to be developed in other water systems. This DD modelling step is done automatically without having to pay attention in each training dataset that we have chosen.

Resampling

"Those who ignore Statistics are condemned to reinvent it"
∼Bradley Efron∼

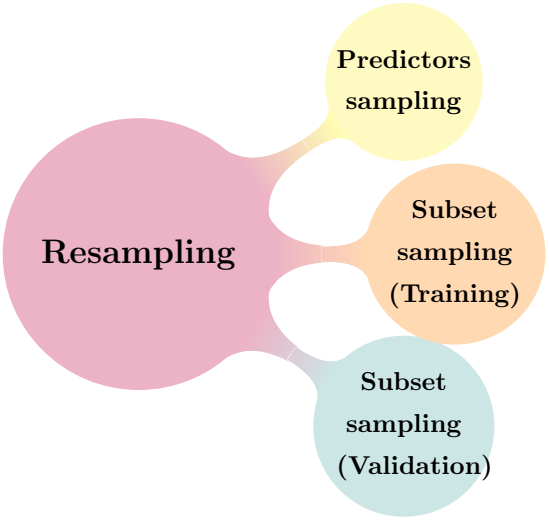


FIGURE 3.4: Resampling methods

Resampling methods consist of creating repeated subsamples from the original data sample, thereby introducing statistical inference on the DB modelling workflow itself. This robustness improvement can be applied on different methodological steps within the modelling experiment/workflow: on the predictors selection, on the subset sampling for training (T) purposes, and on the subset sampling for validation (V) purposes, helping to avoid sampling bias in the methodological context of the hydrological problem-solving. Due to the poor data quality that normally can be found in many Mediterranean basins, specially in our study cases, we should apply this type techniques as a mandatory methodological steps. These resampling improvements help us to avoid biased conclusions by the choice of the predictors, or by the choice of certain parts of the hydrograph during the generation of data subsets for the training and validation periods.

The first resampling application may be to improve the model behaviour during the predictors selection or sampling step, where we look for a sample less biased as possible and more representative and robust as possible. In our case is specially useful when the rainfall predictor is the average value of the collected values from different locations. For that, multiple versions of the rainfall predictor is produced by a method called bagging (a.k.a bootstrap aggregation). This method Breiman (1996) is well known in order to

improve accuracy when the predictors subset selection for training and validation is biased depending on the predictor election. Also it is a recommended manner to generate stochastic noise to the modelling process.

Another improvement can be obtained by resampling the training dataset. For that, we can found the Bootstrap resampling technique (BB), which was introduced by Efron (1979) for IID data. Then, we take a block or random sub-sample of size m from the original sample with replacement. This methodology differs from Montecarlo (MC) as BB is data based, resampling from known samples, while MC generate data given on some parameters inferred previously. MC is more synthetic while BB always works with observed real values.

Finally, for the validation period, the resampling technique is applied through the cross validation method (Browne, 2000). This method is well known in DD applications. Basically cross-validation means to generate various training and validation folders of the entire dataset for averaged validation purposes. The final performance evaluation outcome is computed from various data subsets in a iterative procedure.

3.3.2 Data Transformation

The Standardisation (`std`) of the dataset, is to scale all numeric variables in a numeric range between an inferior limit and an upper limit $[lim_{inf}, lim_{sup}]$. Normally the input data must be standardized into a common range [0, 1]. In Artificial Neural Networks (ANN) the [0.1, 0.9] range is the most used. In this way, it was ensured that the input variables would receive equal attention and weight during the training process (Maier et al., 2010), and also, the training/optimisation algorithms would be more efficient (Dawson and Wilby, 2001). Another aspect to avoid with this data transformation is that, due to the ANN transfer functions used (e.g., logistic function) values close to 0 and 1 produce a slowdown in calculation and therefore in the training period. The lim_{sup} may be modified for hydrograph peak prediction. For example, in Shamseldin (2010) the data is standardised between the [0.1, 0.85] range to ensure the representation of extreme values. This upper limit calibration is selected by the trial and error procedure. The X data set is standardised with the following expression:

$$X_{(i)\,\mathtt{std}} = lim_{inf} + lim_{sup}\frac{X_{(i)} - X_{min}}{X_{max}} \tag{3.5}$$

On the other hand, the *Normalisation* (`norm`) converts a X data set into a normally distributed one $N \sim (0, 1)$ in this way:

$$X_{(i)\,\mathtt{norm}} = \frac{X_{(i)} - \mu(X)}{\sigma(X)} \tag{3.6}$$

The data pre-processing used for ANNs and MLR was just standarization, Eq.3.5, in the range [0.1-0.9]. For the GPs application the mean was assumed zero on the standardized time series, Eq.3.5 and Eq.3.6. Finally, the models outputs are rescaled to be compared and evaluated.

In time series modelling, another popular and important pre-process is the Box-Cox transformation. This Box-Cox transformation is specially important when we work with the assumption of a Gaussian distribution,i.e. Gaussian methods, even if, a priori, it is not. From the practical point of view, in the Box-Cox transformation is important to sum a constant value (λ_2) in order to avoid negative values. The original form of the Box-Cox transformation, as appeared in Box and Cox (1964) and Box and M. (1976), takes the following form:

$$y(\lambda) \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & if\ \lambda \neq 0 \\ \log y, & if\ \lambda = 0 \end{cases} \tag{3.7}$$

and an extended form which could accommodate patterns with negative y's values adding the constant value λ_2:

$$y(\lambda') \begin{cases} \frac{(y-\lambda_2)^{\lambda_1}-1}{\lambda_1}, & if\ \lambda_1 \neq 0 \\ \log(y+\lambda_2), & if\ \lambda_1 = 0 \end{cases} \tag{3.8}$$

where $\lambda' = (\lambda_1, \lambda_2)'$. In practice, we could choose λ_2 such that $y + \lambda_2 > 0$ for any X. In this manner we can view λ_1 as the only parameter.

Finally the last important data transformation is to classify data from a continuous to a discrete domain. The classes assigned to each type depends on the purpose of the applied study. In some situations a binary class problem is a reasonable solution, for instance in a first attempt to classify dry and wet periods in a forecasting method (used in Sec.5.2). For the definition of more than two classes, one of the simple ways of describing the distribution of a random variable is by quantiles. Then the median will be the q_{50}^{th} of the variable, the lower quartile will be denoting the interquartile range for the interval [0-25].

Also the extreme parts of hydrological patterns are values of our interest. The extremes of a distribution are well characterise by the percentiles.

3.3.3 Data Cleaning

Data cleaning (also known as data cleasing) refers to the process of removing invalid data points from a dataset, dealing with gaps and the missing data to make sure that it is well corrected under the expertise of the modeller. The first step is to identify possible outliers, either visually or statistically. Here we must pay close attention as rainfall and streamflow data could be strong skewed under Mediterranean conditions. In this work the outliers are identified in a visual manner as a primary methodological step, taking care not to discard real extreme events that were taken place.

Secondly is to handle data gaps or missing data. For that analysis, gaps of one or two days in duration were filled by interpolation. For gaps longer than two days, no recoding would be done, leaving them as missing data. These days are directly removed from the training and validations data sets.

An alternative option of interpolation to deal with missing data is to fill it with a DB model. For instance the multiple regressions (with exogenous inputs) planned in this work could be used for this purpose. A DD approach to fill the missing hydrological data is to develop a model for that purpose. With the same approach as modelling, we can develop specifically a DD model for that purpose with a similar methodological scheme: (1) Get the best data correlated (a set of good predictors), (2) Develop a DD model, and (3) Validate the model to check if the results are acceptable or not. Then, data can be filled for a specific hydro-meteorological pattern under the study.

It must be commented that for cases where the quality of the record is not sufficient or where the gaps are numerous and long, the high uncertainty associated to the filling is not acceptable for dynamical tools. So it is recommended to separate the signal by events of interest and model it under this assumption, only the events of interest, either under a dynamic or static approach.

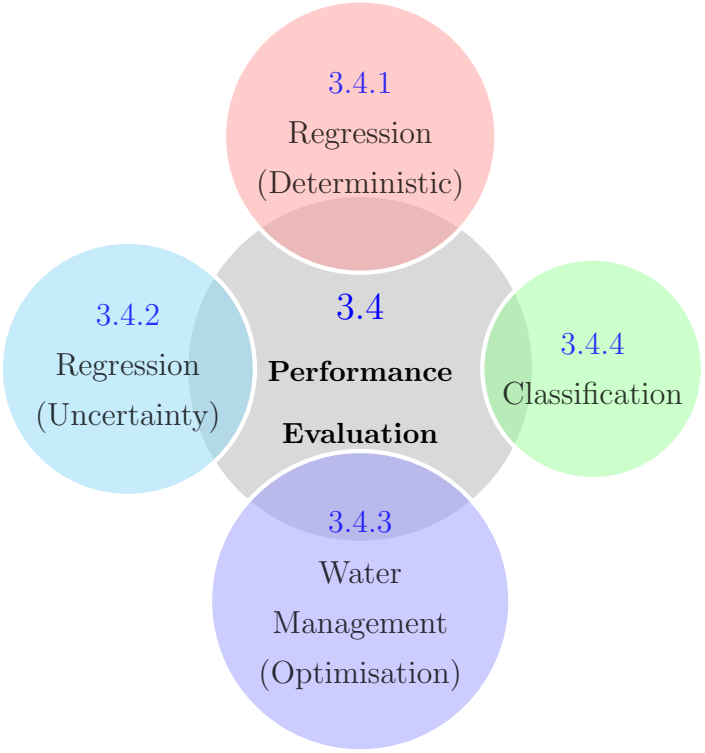


FIGURE 3.5: Performance evaluation

3.4 Performance Evaluation

"All generalizations are false, including this one"
~Mark Twain~

Depending on the methodology, purposes, time scale, objectives and the data used, we will need to use evaluation coefficients of one type or another to evaluate the performance, to analyse the results and to establish conclusions with the proper framework for comparison among different methods. The raw data is split in different data sets during the methodological step that need to be completed within the modelling framework: the parametrisation (optimisation) during the training period (T) and the evaluation of the model generalisation on unseen data during the validation period (V). This analysis is performed with the un-normalised values of the observations and simulations. Then, the coefficients that are presented in this section, will serve us to evaluate the behaviour of the models, techniques and frameworks used throughout this work.

3.4.1 Regression (Deterministic)

In regression cases, with very autocorrelated series, it is important to evaluate the results properly, as the evaluation coefficients can be masked and then, results and analysis can be biased and therefore wrong. For a proper evaluation we need to focus on selecting appropriate objective functions to asses the goodness of fit and the overall rainfall-runoff models performance for example, either by minimizing or maximizing it. In our case, the performance analysis of the results was carried out during the training and validation period using Nash-Sutcliffe Efficiency (NSE), Root Mean Squared Error (RMSE) and Mean Average Error (MAE).

In practical cases, data driven applications don't outperform the simplest form of forecast, the Naïve Method or Persist Method (PM), which is equal to: $y_{(t+h)} = y_{(t)}$ for each horizon prediction h. Therefore, if this empirical threshold is not exceeded, it would not make sense to apply more complex and costly models. For this reason, this is an key point on forecasting performance evaluation since even when the persist behaviour is not exceeded, good performance values can be obtained in terms of efficiency (NSE). Schaefli and Gupta (2007) reflect that "*a convenient and normalized measure of model performance* ($\text{NSE}_{[-\inf \text{ to } 1]}$) *does not provide a reliable basis for comparing the results of different case studies*". Therefore it is necessary to establish a comparative benchmark in this sense, in order to evaluate it correctly with an unmasked criterion for each specific case study, being specially suitable for forecast models (Seibert, 2001). Therefore we must quantify the improvement from the naive forecast criterion and establish it as the baseline. In this work two coefficients are provided for that purpose.

The first one is the Improvement Coefficient (IC) listed by Chen et al. (2014), which is calculated in terms of NSE and RMSE based on the persist baseline ($\mathsf{EC}_{persist}$) by the following expression:

$$\mathsf{IC}(h) = \left(\frac{\mathsf{EC}(h)_{persist} - \mathsf{EC}(h)_{proposed}}{\mathsf{EC}(h)_{persist}} \cdot \mathsf{c} \right) \times 100\% \tag{3.9}$$

The coefficient c takes the value -1 when is NSE ($\mathsf{IC}_{[\text{NSE}]}$) based and 1 when is MAE ($\mathsf{IC}_{[\text{MAE}]}$) based. This empirical threshold for the applications in this thesis are presented in Fig.B.6 and Fig.B.5 for different time scales (hourly and daily), time horizons [$h = 1, \ldots, 5$] and study cases (GH and GF).

The second coefficient aims to quantify the unwanted naive forecast error or phase lag which is normally observed for real-time models in forecasting models. As commented

before, that means that good results with respect to NSE and RMSE coefficients does not guarantee suitable performance to forecast the most important characteristics of the hydrograph, like the beginning of the rising limb, the ending of the falling limb and the timing of the peak. For that, Kitanidis and Bras (1980a) proposed the Persistence Index (PI), also called as the Coefficient of Persistence by posteriors applications (Napolitano et al., 2011). The PI assume that the process is a Wiener process, meaning that the error increases linearly with time, so the best estimate for the future is given by the latest measurement (Kitanidis and Bras, 1980a), that means the PM . Then, if the SF_O is the observed and SF_S the simulated values, the value for each horizon prediction h is calculated as the following expression:

$$\mathsf{PI}_{(t+h)} = 1 - \frac{1 - [\ \sum(\mathsf{SF}_{O(t+h)} - \mathsf{SF}_{S(t+h)})^2\ ]}{[\ \sum(\mathsf{SF}_{O(t+h)} - \mathsf{SF}_{O(t-h)})^2\ ]} \tag{3.10}$$

A different behaviour of our model could be deduced from each evaluation coefficient. For negative PI values mean that the model prediction is worst than the naïve forecast, PI = 0 equal to the naïve forecast and PI = 1 mean a perfect regression on the test values. As an overall rule behaviour of this coefficient, PI ≥ 0.5 are acceptable values to think in further real-time tools development, and values PI < 0.5 indicate a poor behaviour in the hydrograph timing representation. This evaluation is important if we want to obtain real-time tools for floods prevention purposes. For the NSE performance metric, note that NSE = 1 indicates a perfect model, and NSE < 0 are obtained when the model's forecasting is poor. The NSE score is sensitive to the peak flows or extreme values. All these measurements allow us to determine how sure one model can be in making streamflow predictions for the proposed control points. In recent works (Cheng et al., 2017) these two coefficients are treated in a implicit manner.

Related to correlation coefficients that measures the relationship between two variables, the Pearson (R Eq.3.11) and Spearman coefficients (ρ, Eq.3.12) are used in this work. The R correlation coefficient is probably the most widely used measure for linear relationships between two variables. On the other hand the ρ correlation coefficient can be understood as a rank-based version of Pearson's correlation coefficient, which can be used for variables that are not normal-distributed and have a non-linear relationship. If x_i and y_i are the variables of interest to be correlated, $\overline{x}$ $\overline{y}$ its respective mean values, and n is the number of observations,

$$\mathrm{R} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}} \tag{3.11}$$

, and the Spearman's rho

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{3.12}$$

where d_i are the pairwise distances of the ranks of the variables x_i and y_i. This ρ coefficient can be also used in classification cases (Sec.3.4.4).

Model parsimony

"Everything must be made as simple as possible. But not simpler"
~Albert Einstein~

Also, to evaluate the best model it is necessary a criteria to measure the equilibrium among the model behaviour and the its complexity, what means that the number of predictors or parameters and its performance behaviour must be balanced. The parsimonious models not only have the recognition ability, but also have the more important generalization capability (Zhang et al., 1998). For that, the Bayesian Information Criterion (BIC, Rissanen (1978)) is used, which is defined in Dawson and Wilby (2001) as follows:

$$\text{BIC} = n_T \; log(\text{RMSE}) + n(\theta) \; log(n_T) \tag{3.13}$$

where n_T is the number of the observations for the training subset, and $n(\theta) = n(\theta_w) + n(\theta_b)$ is the number of parameters. In this manner we are looking for the optimal parsimony, or just the right amount of predictors or parameters needed to explain the model's purpose in the most balanced way given a cost function, in our case the RMSE.

3.4.2 Regression (Uncertainty)

To assess the uncertainty prediction for a given model $\mathcal{M}$, the Percentage of Coverage (POC, %) and the Average Width (AW) indexes (Zhang et al., 2009; Kasiviswanathan et al., 2016) are used. The POC evaluates the number of times that the target value falls between the uncertainty limits by percentage, see Eq.3.14. Smaller difference between POC and the theoretical coverage level of a simulated confidence interval indicates better performance as is pointed out in Zhang and Zhao (2012).

The AW evaluates de magnitude of the uncertainty estimation of the model, which is calculated by the average of all the uncertainty intervals value during the test period, see Eq.3.15. The AW coefficient is measured in the same units as the target variable. In the

uncertainty intervals analysis it is considered to be, the narrower the prediction interval, that means lower AW, the better the model is. All these metrics allow us to determine how sure one model can be in making streamflow predictions for the proposed control points.

$$\text{POC}(\%) = \left(\frac{1}{n} \sum_{i=1}^{n} c_i \right) \times 100 \tag{3.14}$$

$$\text{AW} = \frac{1}{n} \sum_{i=1}^{n} [y_i^u - y_i^l] \tag{3.15}$$

where n is the number of observations and c_i is equal to 1 if the ith observations falls within the interval y_i^u and y_i^l, which are the upper and lower limits respectively, otherwise c_i is equal to zero. These coefficients to analyse uncertainty estimations in real cases are used in Sec.4.6.

3.4.3 Water Management

These performance metrics are focus on satisfying demands sensitivity. For that cases, the performance criteria in this study are: Squared deficit (D_{sq}), power generation (P_{hp}, $\times 10^8$ kWh), Irrigation deficit ($D_{\texttt{Irr}}$), number of deficit month for irrigation (D_M), and Spill (S_{pll}, Mm^3). For this cases the hydrological framework is focused on the water balance at the reservoir, and whether the theoretical demands are satisfied or not, to a greater or lesser degree.

3.4.4 Classification

To evaluate classified variables, we use the confusion matrix (see Fig.3.6). With this, different accuracy terms are listed in a classification model given. This representation permits to visualise the correct (a.k.a true positives or true negatives) and incorrect percentages (a.k.a false positives or false negatives) in problems where we are working with discrete variables or classes. In this work, this solution is used to evaluate the models behaviour in binary classification schemes, for flood events conditions (Sec.4.8.3), and for wet and dry periods in mid-term predictions (Sec.5.2.3).

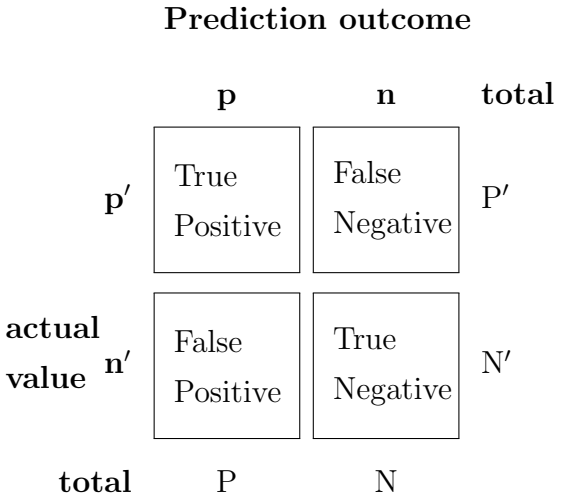


FIGURE 3.6: Confusion Matrix

3.5 Conclusions

This chapter shows a reaching out of the key approaches and conceptions for hydro-meteorological modelling development with Data-Driven models. First, we commented the different modelling approaches under different perspectives. Later, we described a complete picture of the data subsets management, including data division and transformation as well as preprocessing steps to improve the modelling task. Last, we discussed performance evaluation metrics for regression modelling test under an appropriate benchmark, being considered the deterministic and the uncertainty assessments. Finally we described the performance evaluation coefficients for classification problems. From that, to conclude the chapter, the following convictions may be drawn:

- The design of a good Data Based model needs an organised and upstanding methodology. We have shown some key points to pay attention to. The election of the best modelling approach for useful applications, a good data management practices for unbiased inference and an honest performance evaluation for unmasked conclusions are analysed to achieve objectives of this work within a comparison framework.

- For the recognition of the different hydro-meteorological patterns, various approaches with diverse natures are taken to create unique modelling dilemmas within a single compared framework. Adopting different approaches, we obtain other points of view of the same problem. Also, we obtain a greater perspective and therefore a

greater knowledge of the water system under study, and their pulse-response behaviour into the Rainfall-Runoff process. These DD modelling concepts and the experiences with them, add empirical science to the problem where the physical meaning should be inferred.

- From the point of view of the model interpretability, opaque nature methods (i.e., standard ANN explained in App.A.1.2) can migrate to more "grey" nature by adding hyper-parameters under the Bayesian framework. Also this "migration" can be done by generating in a parallel form multiple models members with particular boundary conditions or specific physical meanings.

- We should apply Statistics in Data-Driven methods to avoid biased outcomes and conclusions from our experiences. For that, diverse methodologies that exist for how data is combined (Bagging), sampled (Bootstrap resampling) and managed (Cross Validation) may be applied into the regression model development experience (Sec.3.3.1).

- Growth in computing power has made Bayesian methods for infinite-dimensional models - Bayesian nonparametrics - a suitable framework for inference, finding a practical use for forecasting, where hydro-meteorological regression experiences of computational techniques of this nature (e.g. Gaussian Processes, App.A.1.3) are needed to check its fitness for hydrological purposes.

- We spell out the most pertinent data transformations and evaluation coefficients for a fair performance comparison framework among the models that will be used. We select the more convenient evaluation coefficients to follow the objectives of the work. For the performance analysis for short/mid term forecasting experiences, we use the Nash-Sutcliffe Efficiency, Root Mean Squared Error and Mean Average Error for the deterministic values, and we use Percentage of Coverage and Average Width for uncertainty estimations. For practical real tools advancement purposes, where autocorrelated states are crucial to fit dynamical systems (non-stationary), should be drawn under a performance benchmark where the baseline is the Persist method or Naive method. We assess this through the Persistency Index and the relative improvement over the persist method threshold in terms of NSE and RMSE coefficients.

Chapter 4

Short Term Forecasting

4.1 Introduction

This chapter is focused on the time-scale that belong to the higher sample frequency in this thesis: hourly and daily. Short-term forecasting models have always practical applications in hydraulic engineering and water management. Especially in Mediterranean environments where regulation structures are more frequent than in other latitudes. Given that, one of their main objectives is to insure their water supply during dry and summer months where the concentration time of the coastal basins are usually low. With this conditions, forecast accuracy is specially important for highly variable hydrology and for basins regulated with small reservoirs (Zhao et al., 2011).

So we must store and regulate the water, while at the same time maximize the resource efficiency and minimise impacts of extreme hydro-meteorological events. For flood control in reservoirs, the managers must monitor the water inflows and storage. For that, the runoff forecast plays an important role and should be based on a proper knowledge of the present and past conditions of the water system. Robust forecasting tools are especially needed under flood conditions, when managers have to minimize the costs associated with flooding in floodplains downstream.

In hydrology, the rainfall-runoff (RR) relationship is the most fundamental mathematical/conceptual model that can be established. It produces the runoff hydrograph in response to a hydro-meteorological event, given related variables. Understanding and modelling the RR relationship is therefore important in many water resources problems,

where the key driven processes can be inferred. In this relationship, different hydrological processes may be implicated as forcing agents, for instance the presence of the snow processes in mountainous basins.

In this chapter, seven regression experiences are carried out and analysed. In the Guadalhorce River Basin (GH) three cases at daily time-scale and one at hourly time-scale were selected. In the Guadalfeo River Basin (GF) three regression cases at daily time-scale were selected, because the data available is scarce at hourly basis. This experiences will serve us to analyse the suitability of DD methods to attend different hydrological objectives such as Flood Management (FM), Flood Protection (FP) and Water Allocation (WA).

Sec.4.2 is concerned with the Input Variable Selection (IVS) methodology followed by the election of the predictors and the Predictor Structure (PS) definition. The influence of the rainfall persistence on the runoff for daily patterns (time-scale effect) by comparing different assumptions is examined, and finally the computational techniques used are presented and the comparison framework is defined. In Sec.4.3 and Sec.4.4 the results for GH and GF respectively are presented. In Sec.4.6 and Sec.4.7 the uncertainty quantification and Input relevance for each regression case are analysed. Finally, in Sec.4.8 an analysis of the findings and its discussion are covered, focusing on the particularities in each basin. At the end the chapter conclusions are summarized (Sec.4.9).

4.2 Methodology

In a regression problem, the task is to find a functional representation between output y and input $x = [x_1, \ldots, x_d]$ variables with n observations: $y_{(n)} = f_{\mathcal{M}}(x_{(n)}) \pm \epsilon$ for $n = 1, \ldots, N$, where $f_{\mathcal{M}}$ is the relationship of the model or DB computational model tested and ϵ the noise/variance model. In our hydro-meteorological cases, the dynamic regression method attempts to model a process on the form:

$$\begin{aligned} \mathsf{SF}_{(t+h)} = f_{\mathcal{M}}(\mathsf{PS}) = f_{\mathcal{M}}(\{\mathsf{SF}_{(t-1)}, \ldots, \mathsf{SF}_{(t-p_{\mathsf{SF}})}\}, \{R_{\bar{a}w(t-1)}, \ldots, R_{\bar{a}w(t-p_R)}\}, \ldots \\ \{X_{d(t-1)}, \ldots, X_{d(t-p_d)}\}_{d=1}^{D}) \pm \epsilon_{(t+h)} \end{aligned} \tag{4.1}$$

in which $\mathsf{SF}_{(t+h)}$ is the target variable to forecast for various horizon ahead ($h = 0, 1, \ldots, \mathrm{h}$), $\mathsf{SF}_{(t-p_{\mathsf{SF}})}$ is the antecedent values inputs with $t-1, t-2 \ldots, t-p_{\mathsf{SF}}$, time steps of the target variable, $R_{\bar{a}w(t-p_R)}$ is the Rainfall inputs which can be collected at single one station ($w = 1$) for given location j or the average of several stations ($w \geq 2$) with p_R

antecedent values, and $X_{d(t-p_d)}$ represents any other potential causal variable within the rainfall-runoff process and related to $\mathsf{SF}_{(t+h)}$ having $d = 1, \ldots, D$ candidate variables inputs with its corresponding p_d antecedent values. This constitutes the PS which contains the different regressors that are the inputs' model whose selection is explained below. The ϵ term represents the variance of the model, i.e., the noise model that can be interpreted and inferred in different ways.

In the regression modelling development, the first step to attempt is to define sequentially the predictors during the Input Variable Selection (IVS) procedure, where normally we have a dataset of potential candidate predictors. For that, a wrapper algorithm searches for the best subset of predictors for the target variable estimated at zero time horizon ($h = 0$, see Eq.4.1). Once the best PS has been obtained, an analysis of the Rainfall persistence is carried out (Eq.4.3) for $h = 0$ time horizon. Once the final one is defined, we develop and analyse in a compared way the DD models chosen in this work for the three temporal horizons presented across this chapter ($h = 0, 1, 2$).

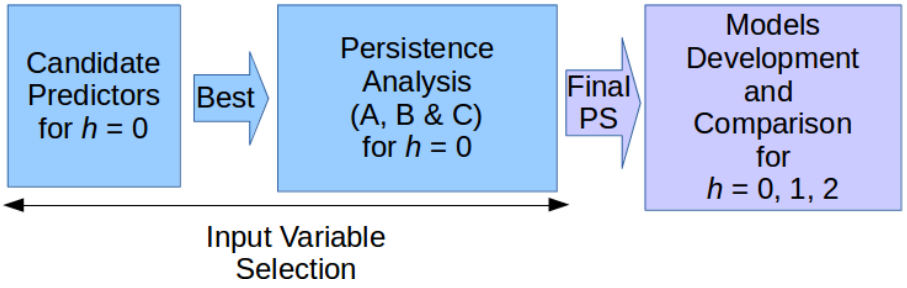


FIGURE 4.1: Methodological steps on forecasting by DB regression

4.2.1 Input Variable Selection

In the Input Variable Selection (IVS) procedure we set the optimal Predictor structure (PS) from each Dataset ($\mathcal{D}$) that will serve us for the modelling experiences in each case of study. Once the optimal PS is defined during the IVS step, the development optimisation and the final comparison among the models are carried out. The dimension of inputs into the PS will be keep as small and representative as possible. The PS should be compact and contain all the relevant information and local specifications within the rainfall-runoff process for our objectives during the regression experience.

In DD applications, the IVS with automated algorithms is a crucial step. During this procedure, the optimal regressors for our objectives must to be chosen carefully in order to avoid a biased selection. These variables selection should be optimal for a proper

optimisation convergence of the model during the training period, and also for a good generalization of it during the validation period. For that purpose, previous works (Sudheer et al., 2002) use the statistical time series properties such as cross-, auto- and partial-auto-correlation in order to define the best antecedent values. More advanced algorithms and methods can be used for the selection of the candidate variables. Recent works also show us the importance of this crucial step in hydrological modelling (Galelli and Castelletti, 2013; Taormina and Chau, 2015) and in general for environmental modelling, e.g., (Galelli et al., 2014) analysed four different algorithms for this purposes in a compared way. The Information Theory foundations (Shannon et al., 1951) can be used for IVS purposes as-well, like in Fernando et al. (2009) who use the partial mutual information and the entropy criterion for this purpose.

At daily basis the number of candidates for the rainfall input is significant, as an extended monitoring network exists in the area. We have to keep in mind that each single location and each combination of locations can be considered as candidates. The proper election of the best rainfall predictor can be tricky. This is specially important for a study of various horizon predictions, adding complexity to the data framework as the rainfall candidate locations may differ for the different horizon predictions. That means that the station with the best performance for one time horizon can differs for two time horizon, for instance. On the other hand, at hourly basis the candidate data that can be used as predictors or explanatory variables of the RR relationship is considerably reduced with respect to daily scale. This is due to the fact that the number of existing meteorological stations collecting data at this frequency is much lower and therefore the exploration of the candidate predictors is faster and more direct.

Description: An iterative model-based algorithm for IVS purposes is developed in this work. It starts with a model-free approach, by analysing the autocorrelation and correlation among the target variable and the input candidate variables (Sudheer et al., 2002). Later it becomes a model-based approach with trial and error methods (Maier and Dandy, 2000) testing the candidate inputs within a DD iterative algorithm framework, which tests different potential and competing predictors. This is achieved by using two different parametrical models: a linear approach: $f_{lin} \rightarrow f_{\text{MLR}}$ (see more of technique foundations in App.A.1.1) for MLR and GP$\sim k_{lin}$, and a non-linear approach: $f_{Nlin} \rightarrow f_{\text{ANN}}$ (thoroughly describe in App.A.1.2) for BNN and the rest of GP kernels.

In this manner we are approximating the underlying function under two different test and fundamental approaches (see more in Sec.3.2.4). The algorithm will be performed

for the entire potential candidate predictors dataset ranking them in order of importance, selecting and evaluating the improvement of the new PS over the old PS, and sequentially defining the predictor structure PS to the final "optimal" one for each model. By running the models (during the training dataset) for each candidate predictors, we will apply the f_{lin} test (that will produce a single value) for each PS. This test is based on a straightforward method (Least Squared, LS) for the parameters optimisation procedure. For the f_{Nlin} test, we will execute several runs in parallel, 50 in our case. This is due to its random weights initialisation ($\theta_{\text{init}} \sim N(0,1)$), which generates different local minima. Then, the model with the maximum performance coefficients are selected as the best non-linear related predictor. The *static approach* for the ANN architecture is used for this f_{Nlin} test during the IVS step. For more information on this refer to App.A.1.2.

Given a $\mathcal{D}$ with n observations of the target variable SF and J candidate locations of Rainfall datasets R_j, $j = 1, \ldots, J$, and a D number of other potential casual variables X_d, $d = 1, \ldots, D$, of importance within the rainfall-runoff process, the IVS step attempts to find the best PS that maximises/minimises a performance coefficient $\varrho_i(\cdot)$ given, $\{\varrho_1(\cdot) = \text{NSE}, \varrho_2(\cdot) = \text{PI}, \varrho_3(\cdot) = \text{BIC}\}$, by two computational models/test: f_{MLR} and f_{BNN}. Then the optimal PS is incrementally built in a sequentially and iterative way (Fig.4.2) testing different inputs by combination as it follows:

1. The autoregressive nature of the model is defined. By partial autocorrelation function (PCAF) the antecedent order p_{SF} of the target variable to be included in the first version of the PS (PS1) is defined, obtaining the symbolic expression: $\mathsf{SF}_{(t)} = f_{\mathcal{M}}(\{\mathsf{SF}_{(t-1)}, \ldots, \mathsf{SF}_{(t-p_{\mathsf{SF}})}\})$.

2. Then a subset of a number of candidate Rainfall Inputs is included as exogenous variable within the autoregressive relationship function, where the best p_R antecedent order values for Rainfall is defined by cross-correlation. The iterative algorithm is run for each location R_j for $h = 0$ time horizon, and the stations are sorted in order of performance according to their NSE and PI coefficients (ϱ_1, ϱ_2). Then the average of the first two $(R_{\bar{a}2})$, three $(R_{\bar{a}3})$, four $(R_{\bar{a}2})$ and five stations is calculated, generating $N(w)$ for $w \geq 2$ of new candidate Rainfall predictors. Then, the algorithm is run, testing this new average rainfall $J + N(w) - 1$ candidate time series, sorting them together with the single-station predictor from highest to lowest based on their performance metrics. The best predictor will be the one that maximises the performance. Finally, once defined the best rainfall predictor, a second version of PS (PS2) is obtained, with the following symbolic expression: $\mathsf{SF}_{(t)} = f_{\mathcal{M}}(\{\mathsf{SF}_{(t-1)}, \ldots, \mathsf{SF}_{(t-p_{\mathsf{SF}})}\}, \{R_{\bar{a}w(t-1)}, \ldots, R_{\bar{a}w(t-p_R)}\})$, where $R_{\bar{a}w}$

is the rainfall collected from a single station $R_{\bar{a}w}$ for $w = 1$, or by averaging of rainfall data collected at various locations $R_{\bar{a}w}$ for $w = 2, \ldots, 5$. The application of the two models in parallel can lead to two situations, specially during this step: that both tests coincide, or that differs during the predictors selection procedure. If the results differ, two optimal precipitation predictors are established by w_{lin} and w_{Nlin} subscripts for further applications in this work. If the two tests coincide, we have unequivocal evidence that this is the strongest connection of the best predictor.

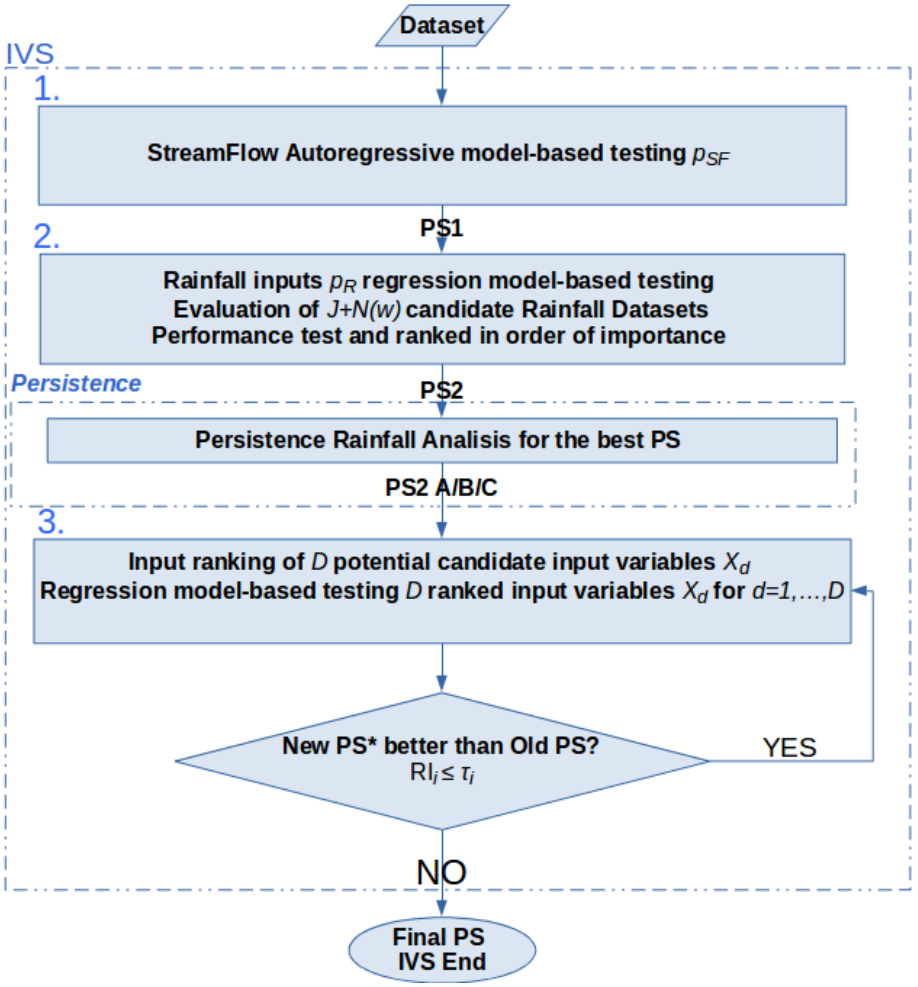


FIGURE 4.2: Flowchart of the Input Variable Selection (IVS)

3. A D number of relevant casual variables (X_d) within the rainfall process may be included as exogenous variables to approximate the underlying function $f_{\mathcal{M}}(\cdot)$ of the hydro-meteorological process. The potential D_x variables to be tested are previously ranked based on the crosscorrelation between each d candidate variable timeserie and the target variable. After a new antecedent order p_d of the candidate variable is defined as in previous stages, the new candidate PS^* is included in the PS, and this is tested again to calculate its new performance by both approaches. This third step

is repeated until the addition of more candidates to the PS^* does not significantly improve the performance metrics of the old PS. This is mathematically determined by a stopping rule procedure based on a threshold. The improvement on each step is computed by its relative performance improvement with respect the previous PS ($\mathrm{RI}_i = (\varrho_i(\mathsf{PS}^*) - \varrho_i(\mathsf{PS}))/\varrho_i(\mathsf{PS})$) with the following stopping rule,

$$\mathrm{RI}_i \leq \tau_i \tag{4.2}$$

where τ_i is a threshold for each $\varrho_i(\cdot)$/performance metric established. In this third step is convenient to add the Bayesian Information Criterion (BIC) as new performance coefficient ϱ_3 to measure the model complexity and to choose the optimal PS in balanced way among accuracy and parsimoniousness. Then, finally the best Predictor Structure (PS) for a given target variable (SF) to further develop modelling experiences is defined with D new input variables having the following form: $\mathsf{SF}_{(t)} = f_{\mathcal{M}}(\{\mathsf{SF}_{(t-1)}, \ldots, \mathsf{SF}_{(t-p_{\mathsf{SF}})}\}, \{R_{\bar{a}w(t-1)}, R_{\bar{a}w(t-p_R)}\}, \ldots$
$\{X_{d(t-1)}, \ldots, X_{d(t-p_{dD})}\}_{d=1}^{D})$.

The procedure is repeated until the algorithm meet the stopping rule in each case study to be analysed. Then the final regression model in the form of Eq.4.1 is obtained. This final PS will be used in the next methodological step.

Persistence Assumptions: At daily scale, the persistence related to the rainfall pulses is examined. For that challenge, the corresponding Predictor Structure (PS) obtained in the step 2 of the previous section is analysed under different assumptions to study the basin time response related with the availability of rainfall predictions at the time step (t) of the prediction. This problem arises from the lack of consistency between the time scale of the data and the hydrological process to be represented. This analysis also minimizes inconsistencies among different monitoring networks, where the starting/ending points of the 24h time period can differs from one to another, swapping the rainfall values collected at lower frequencies to further or antecedent time slices during the upscaling procedure.

In the previous step 2 of the IVS procedure, the variables combinations and their p antecedent values within the PS were chosen based on the correlation coefficients between each time series and the target time series. Once the best $\mathsf{PS2}$ is defined the efficiency and forecasting capabilities of the PS is optimised. For that, two assumptions has been defined: A and B.

$$\begin{aligned} \mathsf{SF}_{(t+h)} &= f_{\mathcal{M}}(\{\mathsf{SF}_{(t-p_{\mathsf{SF}})}\}, \{R_{(t-1)}, \ldots, R_{(t-p_R)}\}) \Rightarrow \text{Assump. A} \\ \mathsf{SF}_{(t+h)} &= f_{\mathcal{M}}(\{\mathsf{SF}_{(t-p_{\mathsf{SF}})}\}, \{R_{(t)}, \ldots, R_{(t-p_R)}\}) \Rightarrow \text{Assump. B} \end{aligned} \tag{4.3}$$

The A assumption means that no predicted Rainfall values are included into the input structure: $R_{(t-1)}, \ldots, R_{(t-p_R)}$ (Eq.4.3). The B assumption includes the predicted values (t) are included into the input structure (Eq.4.3). Here we assume accurate predictions without any uncertainty source, that means the real collected values. This has been done for studying the potential behaviour and its predictive performance with the joint use of meteorological predictions as input variable.

The use of these type of assumptions allows us to infer the physical behaviour of the pulse-response basin dynamics, and the usefulness of climate data trough GCMs models for improvement on forecasting tasks. It is a way of testing the necessity, and to what extent, of exogenous/climate/atmospheric information at global scale to improve our modelling performance behaviour at local scale.

–"Opening the black box"–

Input Relevance Comparison One of the most important drawbacks of ANNs use is the lack of interpretability, being a recurrent argument on experts discussions (more in Sec 3.2.1). In regression experiences the input relevance is always a mandatory step in order to describe the importance of each dependant variable. Different methods have been described to detect the input relevance in variables for ANNs experiences (Olden and Jackson, 2002; Gevrey et al., 2003; Duncan, 2014). The Connection Weight (CW) approach is chosen for the assessment of the input relevance in this work, following (Gevrey et al., 2003). The Connection Weight of an input d, this is, the comparative significance of the input factor d, is estimated with the following equation:

$$\mathsf{CW}_{(d)} = \sum_{x=1}^{\mathtt{HL}} (\theta_{ij} \cdot \theta_{jk}) \tag{4.4}$$

where HL represents the hidden node's number, x represents hidden node's index number, θ_{ij} represents weights values between input i and hidden node j and θ_{jk} represents weights values between hidden node j and the output node k. Then $\mathrm{CW_N}$ is normalised between 0 and 1 for comparison purposes to avoid the influence of local minimum solutions, as

the solution space domain is irregular.

$$\mathsf{CW}_{\mathrm{N}(d)} = \frac{\mathsf{CW}_{(d)} - \min(\mathsf{CW})}{\max(\mathsf{CW}) - \min(\mathsf{CW})} \tag{4.5}$$

where $\mathsf{CW} = (\mathsf{CW}_{(1)}, \ldots, \mathsf{CW}_{(d)})$ and $\mathsf{CW}_{\mathrm{N}(d)}$ is now our d^{th} normalized connection weight, which is the relevance of the different inputs of the neural networks. On the other hand, for the *lin* test the β parameters for each input of the final PS modelled are normalised as well between 0 and 1.

Finally, the input relevance is represented and visualised in each case study for the final PS modelled. For that, the different $\mathrm{CW_N}$ distributions obtained from the multiple simulations are lumped together and plot for each input variable in a boxplot. On the other hand, a single value for the *lin* test ($\beta_{\texttt{MLR}}$) corresponding to the straightforward optimisation method (LS), is plotted as well. In this manner, we can compare the input relevance of both tests together, and rank it in a balanced way, comparing two different fundamental approaches. This analysis is carried out for each case study but, for a clearer discussion, is grouped all together and analysed in a further section (Sec.4.7).

4.2.2 Forecasting regression models comparison

During this Chapter, three computational DD techniques ($\mathcal{M}$) are employed, analysed and compared to asses the potential suitability of each method regarding each case condition: a purely parametrical and linear method as the Multiple Linear Regression (MLR), another less parametrical and non-linear method as Bayesian Neural Networks (BNN) and one non-parametrical and more probabilistic method as Gaussian Processes (GP). A short introduction of these three techniques, with their major limitations, is included below:

$\mathcal{M} \rightarrow$ **MLR** Multiple Linear Regression (MLR) serves as a computational DD technique baseline and a linear regression benchmark. It is recommended for linear relationships between a target variable and independent variables. Should be highlighted that this technique is unable to capture non-stationarity and non-linearity (Yaseen et al., 2015). For supplementary material of this DD technique refer to App.A.1.1.

$\mathcal{M} \rightarrow$ **BNN** The Artificial Neural Networks (ANN) are well known because their flexibility and ability of modelling complicated nonlinear systems. This computational

technique can be applied under a Bayesian framework for uncertainty inference and quantification (Bayesian Neural Networks BNN, MacKay (1992a)). Among its limitations are the uncertainty in the best model election due the local minima (valleys) solutions, and that performance can be inconsistent in some linear problems (Khashei and Bijari, 2010). For supplementary material of this DD technique refer to App.A.1.2.

$\mathcal{M} \rightarrow$ **GP** Gaussian Processes (GP) is full Bayesian computational DD technique. According to Rasouli et al. (2012), the Bayesian inference is more accurate in the GP technique than in the BNN. Among its limitations are: slow inference due the matrix inverse computations, the need to choose a Kernel Covariance Function (KCF) to approximate the underling function, and modelling under Gaussian predictive distribution nature. For supplementary material of this DD technique refer to App.A.1.3.

Once the best PS is defined in the previous IVS step, using a regresion model in the form of Equation 4.1, we start with the model's development procedure. Depending on computational technique used, the development can be more or less time consuming. The single MLR model is straightforward once obtained from the previous step, $f_{lin} \rightarrow f_{\text{MLR}}$, since it does not need further development for same predictors. In the case of the BNN development f_{BNN} the definition of the architecture, learning parameters, and the application of the Bayesian network, among other development steps needs to be carried out, more in App.A.1.2. On the other hand, for the GP case development f_{GP}, the Kernel Covariance Function (KCF) and its parameters need to be defined (see more in App. A.1.3), step that demand intense calculation time.

After the model development procedure, the evaluation step is carried out. From each model $\mathcal{M}$ tested two estimations are obtained. A first estimation is the deterministic value of the target variable. While the probabilistic estimation, the uncertainty quantification, is the second. This variance (error bars) specifies the probabilistic distribution function for the target variable, giving the 95% confidence intervals for each time slice predicted. This value is conditioned by all the information used during the training period. It is the numerical measure of the degree of certainty about the occurrence of an event, which is mathematically computed following different expressions within each technique.

The techniques foundations are well documented in App.A. For that, a descriptive nomenclature is established (Eq.4.6) for each run/test/modelling experience carried out, to simplify the comparative analysis. This nomenclature includes information about each of the

following aspects on the modelling workflow: (1) The study site either Guadalhorce (GH) or Guadalfeo (GF) (details in Chapter 2). (2) Preprocessing steps taken, which may be Wavelet filtering (WV) and resampling methods as Bagging (B) or Bootstrap (BB) (more details in Sec.3.3). (3) The Predictor Structure number z (PSz), which is obtained in the IVS procedure, explained in Sec.4.2.1. (4) The persistence assumption applied (A or B) within the forecasting framework (more in Sec.4.2.1). (5) The computational technique applied for the regression, fitted to the different modelling forms ($\mathcal{M}$) listed previously. The nomenclature structure is showed below,

$$\overbrace{\mathrm{GH/GF}}^{(1)} - \overbrace{\mathrm{WV/B/BB}}^{(2)} - \overbrace{\mathsf{PSz}}^{(3)}\ \overbrace{\mathrm{A/B}}^{(4)} - \overbrace{\mathrm{MLR/BNN/GP}}^{(5)} \tag{4.6}$$

For example, GF1-PS1A-MLR refers to a model structure applied in the Guadalfeo River Basin, case study number one, with the best PS number 1 which is under assumption A modelled with the MLR computational technique. The comparison between models performances allows us to quantify the differences among the approaches tested, and the differences for practical targets under a single framework: performance and forecasting capabilities. Then *Single vs Multiple* (Sec.3.2.6) model approach comparison is carried out through the use of the best performance of a method, a single model and the use of ensemble methods for the use of multiple models. The linearity (*Linear vs Non-Linear*, Sec.3.2.4) of the approach, the probabilistic inference (*Frequentist vs Bayesian*, Sec.3.2.5) and the parametrical (*Parametric vs Non-parametric*) approaches are also compared among them by the definition of the methods nature. The comparison outcomes give us a deeper knowledge of our water systems and the importance and the influence of the hydrological processes.

Model's deterministic performances were assessed on the basis of four performance coefficients: Nash-Sutcliffe Efficiency (NSE), Root Mean Squared Error (RMSE), Persistence Index (PI) and Mean Average Error (MAE). On the other hand, the model's uncertainty skills were assessed on the basis of two performance coefficients: Percentage of Coverage (POC) and Average Width (AW). More on the performance metrics refer to Sec.3.4.1 and Sec.3.4.2 respectively. The following mind map (Figure 4.3) represents the scheme of the sections that follow, with the application of various techniques and approaches in two Mediterranean River Basins at different timescales: GH2 at hourly basis, and GH1, GH3, GH4, GF1, GF2, and GF3 at daily basis.

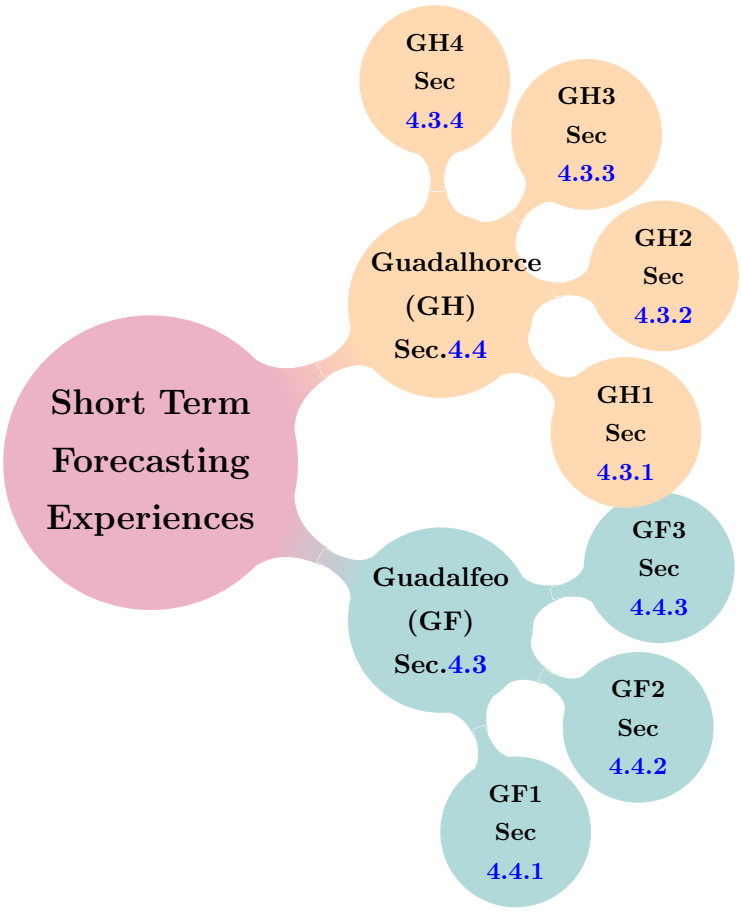


FIGURE 4.3: Short Term forecasting cases

4.3 GH regression experiences for forecasting purposes

The Guadalhorce (GH) River Basin is specially sensitive to the orographic precipitations and the cold front phenomena, which may take place mainly during the autumn season. The torrential character of the precipitation in such a small basin, with a concentration time of less than 12 hours, produces flash flood events with catastrophic effects over the city of Málaga (≈569000 inhabitants in 2017). Despite being regulated in the upstream part of the river course, typical intense Mediterranean rainfall events produce recurrent floods feed from the unregulated contributing area (Sec.2.4.2), where the water managers must take decisions with a very small time window.

From this fact arises the need for specific alert tools which can forecast water stages in the river channel in advance, based on present and antecedent hydro-meteorological states.

For this river basin, we analyse several case studies that include the daily maximum water height (Sec.4.3.1) and the hourly water height (Sec.4.3.2) in Cártama gauge station. This is a key point located 18 km upstream of the river mouth, used by the river basin agency for monitoring and study alert trigger conditions. In this Section we also study cumulated streamflow forecasting for gauge stations located on the head of the reservoirs used for flood control purposes on this hydraulic structures (Sec.4.3.3 & Sec.4.3.4).

For each case the performance of three DB computational techniques, MLR, BNN and GP, for different time horizons ($h = 0, 1, 2, 3, \ldots$) are used. The GP does not apply at hourly basis due its high computational cost, making its application at this time-scale prohibitively slow for more than a few thousand data-points. At daily basis, the best assumption (A or B) is previously selected during the IVS procedure (Sec.4.2.1).

4.3.1 GH1: Daily maximum water height for flash floods forecasting

Case definition In this regression experience, the performance of three DB computational techniques (MLR, BNN and GP) are used to forecast maximum daily water height ($\mathsf{WH}_{\mathsf{Ca}(t+h)}$ in $m \cdot day^{-1}$) at three time horizons ($h = 0, 1, 2$) for the best assumption (A or B) which is selected during each IVS procedure case. This case serve us to analyse the performance of DB techniques in forecasting daily maximum water height for flash flood events predictions. This problem tackles an important issue for water managers: flood control and its mitigation.

In this study, the rainfall data used is collected at 16 meteorological stations, for the time period from 01/10/1995 to 10/10/2010. These daily time series are related with the maximum water heights registered at hourly frequency in Cártama gauge station, Fig.2.1. The daily discharge (Q, daily mean) from hourly time series of the Guadalhorce Reservoir System releases is also used. More information about the hydrological framework and statistical parameters is shown in Tab.B.3.

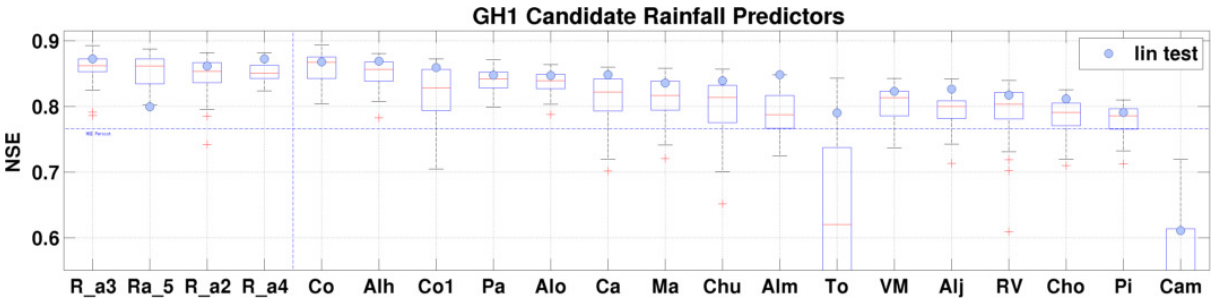


FIGURE 4.5: IVS for rainfall inputs sorted by their maximum value reached by the *Nlin* test outcomes (boxplot). On the left, the results for the combination of 2,3,4 and 5 stations. On the right, the results for single locations. The corresponding *lin* results values are also presented (grey dots)

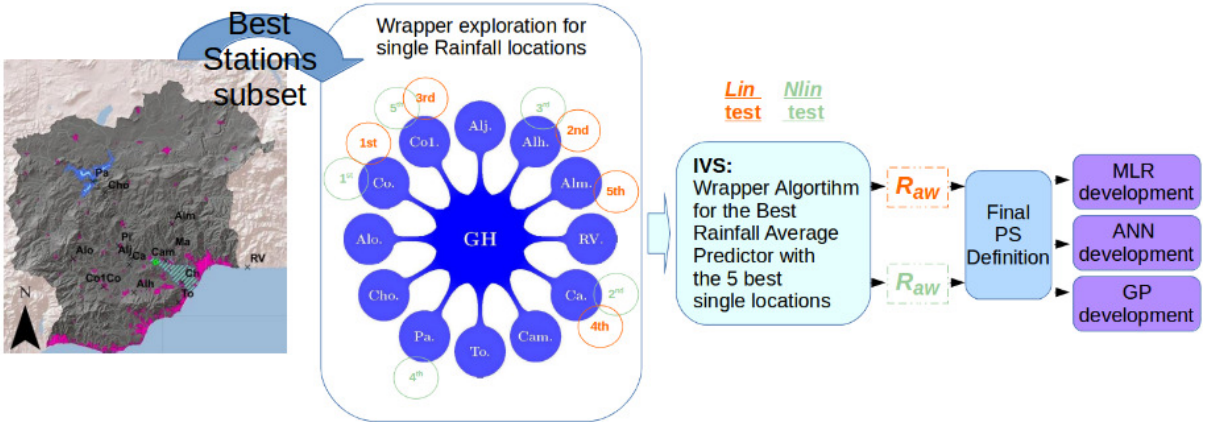


FIGURE 4.4: IVS example for Rainfall variables selection in the GH1 case. This scheme is applicable for the rest of case studies

Input Variable Selection results According to the methodology exposed in Sec 4.2.1 (workflow visualised in Fig 4.4) the best rainfall data collected in a single location obtained were `Co`, `Alh`, `Co1`, `Ca` and `Alm` ranked in order of importance, for the *lin* test. The average of the first best four stations ($R_{\bar{a}4(t)}$) outperformed the single data source. On the other side, the best single rainfall datasets ranked in order of importance for the *Nlin* test was collected at `Co`, `Alh`, `Co1`, `Pa`, and `Alo`. The average of the best three stations ($R_{\bar{a}3(t)}$) outperformed the data collected at a single one. These results are shown graphically in Fig.4.5, where the different rainfall inputs are sorted by their performance. The *lin* and *Nlin* tests match in the best single rainfall station but not with the averages data sets, $w_{lin} = 4$ and $w_{Nlin} = 3$ respectively. In this case, it can be seen from Tab.4.1 that the best PS is number 3 under assumption B (PS3B) taking into account the PI metric for *Nlin* and both NSE and PI for the *lin* test. The PS2B shows a higher NSE metric value for the *Nlin* test, but with a minimal improvement, while the PI is considerably worse

than the one for PS3B. This is why PS3B has finally been established as the final PS. Finally, the following symbolic expression is obtained:

$$\begin{aligned} \mathsf{WH}_{\mathsf{Ca}(t+h)} = f_{\mathcal{M}}(\{\mathsf{WH}_{\mathsf{Ca}(t-1)},\, \mathsf{WH}_{\mathsf{Ca}(t-2)}\}, \ldots \\ \{R_{\bar{a}w(t)},\, R_{\bar{a}w(t-1)},\, R_{\bar{a}w(t-2)}\}, \ldots \\ \{Q_{(t-1)},\, Q_{(t-2)}\}) \end{aligned} \tag{4.7}$$

TABLE 4.1: Relevant PS evaluated with *Nlin* (f_{ANN}) and *lin* (f_{MLR}) test for the GH1 case. *single*: From one station. *multiple*: the average of w stations. Best assumption for each PS and test during IVS procedure is highlighted in bold font

Basin-PS	**Symbolic Model: $\mathsf{SF}_t = f_{\mathcal{M}}(\mathbf{PS})$**	**NSE**	**PI**
	Real data (*single*)	(*Nlin*;*lin*)	(*Nlin*;*lin*)
GH1-PS1A	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{2}; \{R_{a1(t-p)}\}_{p=1}^{3})$	0.897;0.840	0.55;0.31
GH1-PS1B	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{2}; \{R_{a1(t)}, R_{a1(t-p)}\}_{p=1}^{2})$	0.897;0.861	0.55;0.40
	Real data (*multiple*)		
GH1-PS2A	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{2}; \{R_{\bar{a}w(t-p)}\}_{p=1}^{3})$	0.877;0.839	0.39;0.31
GH1-PS2B	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{2}; \{R_{\bar{a}w(t)}, R_{\bar{a}w(t-p)}\}_{p=1}^{2})$	**0.910**;0.865	0.46;0.42
	Real data (*multiple*) + Releases		
GH1-PS3A	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{2}; \{R_{\bar{a}w(t-p)}\}_{p=1}^{3}; \{\bar{Q}_{(t-p)}\}_{p=1}^{2})$	0.881;0.847	0.49;0.34
GH1-PS3B	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{2}; \{R_{\bar{a}w(t)}, R_{\bar{a}w(t-p)}\}_{p=1}^{2}; \{\bar{Q}_{(t-p)}\}_{p=1}^{2})$	0.904;**0.872**	**0.58**;**0.45**

After the input relevance analysis (summarised in Sec.4.7), we obtain the best Predictor Structure (Eq.4.7) for the input model. It is a function of the maximum water level of the current day and the previous day, the average rainfall for two gage stations ($R_{\bar{a}w(t)}$ for $w_{lin} = 4$ and $w_{Nlin} = 3$) for the current day, from the current day and the previous day, and the daily mean of the reservoir releases from the current and the previous day. Also the releases maximum daily values Q_{max} were tested instead of the mean daily values Q, but with poorer simulations results.

Modelling results For 1 day ahead, the MLR model during the validation period has a NSE value of 0.872, and a PI value of 0.45. For 2 day ahead, has a NSE value of 0.835 and a PI value of 0.55. Finally for 3 day ahead has a NSE value of 0.687 and a PI value of

0.30. It can be seen that the PI obtained for $h = 1$ time horizon has an acceptable value (PI$>$ 0.5).

For 1 day ahead, the best BNN model during the validation period has a NSE value of 0.888, and a PI value of 0.52. This model had 6 neurons units in the hidden layer. For 2 day ahead, it has a NSE value of 0.848 and a PI value of 0.59, with 2 neurons units in the hidden layer. Finally, for 3 day ahead, the model has a NSE value of 0.714 and a PI value of 0.36, with 2 neurons units within the hidden layer. It can be seen that the PI values are equal for one and two days ahead, because we are working under the B assumption and the travel times of the flow from the rain gage stations to the stream gauging station are less than one day.

TABLE 4.2: GH1-PS3B models result comparison at training (n=4391) and validation (n=1098) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures (m). Best performance for each horizon and metric during validation is highlighted in bold font

$\mathcal{M}$	h	NSE	RMSE	PI	MAE
MLR	$0 \rightarrow (t)$	0.872;0.872	0.138;0.193	0.24;0.45	0.059;0.072
	$1 \rightarrow (t+1)$	0.804;0.835	0.165;0.220	0.50;0.55	0.076;0.091
	$2 \rightarrow (t+2)$	0.508;0.687	0.235;0.303	-0.20;0.30	0.102;0.128
BNN	$0 \rightarrow (t)$	0.921;**0.888**	0.115;**0.181**	0.64;**0.52**	0.045;**0.062**
	$1 \rightarrow (t+1)$	0.866;**0.848**	0.150;**0.211**	0.64;**0.59**	0.064;**0.082**
	$2 \rightarrow (t+2)$	0.703;**0.714**	0.223;**0.290**	0.40;**0.36**	0.090;**0.120**
GP-k_{lin}	$0 \rightarrow (t)$	0.850;0.876	0.173;0.191	0.35;0.47	0.076;0.074
	$1 \rightarrow (t+1)$	0.801;0.836	0.199;0.219	0.48;0.55	0.094;0.093
	$2 \rightarrow (t+2)$	0.638;0.684	0.269;0.304	0.29;0.29	0.121;0.131

The GP-k_{Lin} model performs better than the other KCF tested for all the lead time forecast with higher NSE, PI and lower RMSE, MAE for the validation period. For the kernel comparison refer to Tab.A.5 in page 229. In addition, it is the only kernel that does not overfit, $\text{NSE}_{(\text{T})} = 0.850$; $\text{NSE}_{(\text{V})} = 0.876$ compared with the other kernels tested, generalising well on unseen data. It can be seen that it has the same efficiency and predictable capabilities than the *lin* test. Finally for 1 day ahead, the GP-k_{lin} model during the validation period has a NSE value of 0.876, and a PI value of 0.47. For 2 day ahead the NSE value decreases to 0.836 and a the PI value increases to 0.55. Finally, for 3 day ahead the NSE and PI values decreases significantly to an uncepptable 0.684 and 0.29 respectively for the validation period.

Discussion Multiple Linear Regression (MLR), Bayesian Neural Networks (BNN) and Gausian Processes (GP) were applied to predict maximum daily fluctuations of water level at the Cártama control point ($\mathsf{SF}_{\mathsf{Ca}}$). Due the basin conditions of this study case, the performance of all models has been negatively affected by its complexities like: high regulation, significant water withdrawals that are not registered, and incomplete time series which decreases the overall performance in the pattern recognition step.

During the first steps of the IVS procedure all the rainfall stations are modelled by the *lin* and the *Nlin* test. The results between both tests are very similar with minimal differences. This separation assures us to maximize the efficiency of each approach but not as an overall. In this case, the first single station of both tests coincides but not the following stations, with minimal differences. The only station that differs among both tests is `Alm`, ranked in fourth position. From a physical perspective that means that a stronger linear relationship with the target variable is taken place in these set of stations. In general, the hydrological relation of rainfall-runoff in this basin is more lineal than other cases (see further examples regarding GF), and that is the reason why the differences among both tests are not too large. Nevertheless, there is a minimal difference among the best rainfall predictor with the average of w stations in each test, 4 for the *lin* and 3 for the *Nlin* test.

For the IVS, only one optimal rainfall input was considered in this and in the subsequent cases. At previous stages, the use of several rainfall independent inputs at different locations was also considered. But with this last option, we never achieved an improvement over the single rainfall input. So the first option with a single rainfall predictor, was established as a standard for most of the study.

During the IVS operation, some additional casual variables X_d were tested. This is the case of the temperature or the atmospheric pressure, collected within the local monitoring networks and from reanalysis respectively. This variable was not finally reflected in the PS as its inclusion did not improve the simulations results any further. The releases from the upper reservoir showed a limited influence as a dependent variable (Table 4.1). This due to the contribution of the water flow coming from the unregulated area of the river basin. This area, closer to the seashore, is driven by fronts coming from the sea. This can be observed by the fact that the maximum rainfall rates are collected in the weather stations by the seashore. When the mean daily values of the releases (Q) are included within the PS we only notice a slight improvement in the peak timing while the efficiency (NSE) of the model decreases.

In this particular case, during the final PS election, the best one is the PS3B for the *lin* test with higher NSE and PI performance metrics. In the other test, the *Nlin*, the best performance metrics do not coincide. The higher NSE efficiency is obtained with PS2B, and the best predictability capacities is obtained with the predictor structure with the releases (PS3B). As the efficiency is quite similar and the timing is much more better reproduced by the model with releases, PS3B was chosen as the final PS for the model development and comparison phase.

The BNN models were found to provide more accurate maximum water levels than the GPs and MLR for the deterministic estimations for all the time horizons and all performance metrics tested. If we compare the errors, NSE, RMSE and MAE, across the three time horizons tested it can be observed as the error increases linearly with time. For that reason the PI performance metric (Tab.4.2) and the relative improvement should be used within the comparison framework. In terms the relative improvement (IC) of NSE compared to persist method values is 15.9%, 34.2% and 29.7% for the three horizons respectively during the validation period. On the other hand, in terms of relative improvement of RMSE is 30.9%, 54.17%, and 11.81% during the validation period. In addition, the GP model with linear KCF (k_{lin}) and MLR has almost identical deterministic results for the validation period. This is apparent comparing the performance coefficients.

Computationally speaking, a single run of MLR is almost instantaneous. The computation time (t_c) for the convergence of the optimal solution for one run during the training phase development, through Least Squared (LS) optimisation, is very fast, $t_c < 1\,seg$. The BNN takes also few seconds for the standard configuration (number of hidden neurons equal to the number of inputs, App.A.1.2), which is the MLP architecture used during the IVS procedure step. For this technique we must take into account learning parameters, such as the learning rate, the momentum, or the number neurons within the hidden layer, since this can slow down or speed up the computation time for a single random run considerably. For instance, as an average for this particular case (GH1), with $n_{\mathrm{T}} = 4359$ observations during the training phase, the parameters optimisation phase trough the Scaled Cojugated Gradient (SCG) algorithm takes 40.3 seconds per run with 85 unit neurons within the hidden layer and three hours and a half for the total 1500 simulations under the dynamic approach (App.A.1.2). Finally, during the GP development for this particular case, a linear kernel takes $t_c \approx 2\,min$ to estimate parameters. On the other hand a neural network kernel takes ten times more ($t_c \approx 20\,min$) and a dot product kernel takes $t_c \approx 25\,min$ during the parameters optimisation. This has limitations when the search for the best kernel and its comparison is realised by the trial and error

procedure. The computer used in this work is a Desktop PC with Intel Core i7 Processor (4x 3.6 GHz) and 16 GB DDR4 RAM, with Linux-based operating system.

4.3.2 GH2: Hourly Water height for flash floods forecasting

Case Definition In this regression experience, the performance of two DD computational techniques (MLR and BNN) are used to forecast hourly water height values ($\mathsf{WH}_{(t+h)}$, collected in $m \cdot hour^{-1}$ units) for ten different time horizons ($h = 0, \ldots, 9$). On the other hand, the GP technique is leaved out due to their high computational cost for the large amount of data treated at this temporal time scale. Hourly basis is always a good example for practical water engineering applications, specially for real-time flood warning tools development purposes, being in consonance the time-scale with the physics of this hydrological phenomenon (Blöschl and Sivapalan, 1995).

In this case study the Bootstrap Resampling technique (BB, Efron (1979)) is applied during the training period to reduce biased or masked results on the election of the training dataset. Then several candidate models are built during both techniques development. The number of bootstrapped runs launched is 50 runs per time horizon.

The hydro-meteorological variables used are: the Water height (WH) registered in Cártama as the target variable, the Rainfall (R) collected at hourly frequency, and the Upstream Releases (Q) as an additional explanatory variable X_d. For this case, the candidate rainfall data set used were collected at hourly basis from five different meteorological stations (locations in Fig.4.6). The time period modelled is from 24/01/2001 to 10/03/2011 with 88740 observations in total, where less than 2% of the collected data was missing or invalid data which is omitted. More information about the hydrological framework and statistical parameters of this modelling experience is listed in Tab.B.4.

In this case study, the methodology and steps take for the final comparison analysis stand as follows:

- The autoregressive order of the target variable (p_{WH}) based on autocorrelation is formulated. As a general rule through this work, this lag is calculated at the point of the correlogram where there is not further change on correlation values. The next step is to test the candidate rainfall data sets by both tests for $h = 0$ time horizon. As at daily basis case, the rainfall datasets are ranked and aggregated sequentially in a bootstrap model-based approach. After that, all this new rainfall

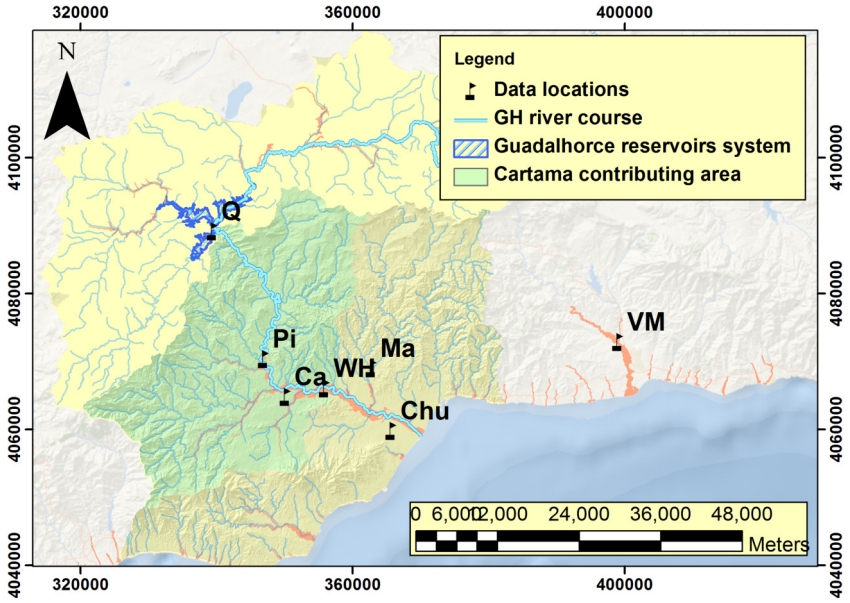


FIGURE 4.6: Map of the stations used for the GH2 case study at hourly basis. The floodplains associated to the most frequent events (T10, more in Sec.2.4.2) in the region is also represented with a orange shape

datasets are tested and ranked in order of importance with the single stations based on its performance (Sec.4.2.1).

- For this case, two different approaches to introduce the rainfall in the final PS are compared during the IVS procedure. A first version (V1), where the wining precipitation from the previous step (j) is introduced as a single input window with its corresponding antecedent values, and a second version (V2) where a new candidate rainfall predictor is introduced as an independent input, having two different rainfall predictors j and v. A higher number of stations to be included independently have not been tested for simplification purposes.

- As in GH1 case, with the same control point, the releases (Q) dataset is tested as a candidate variable (X_1) by a model-based approach (trial and error procedure), analysing the releases dependency by comparison and inferring its relevance within the water system for water height forecasting at hourly basis. Finally the "best" model obtained from BB at each time horizon is presented and compared for the both computational techniques used at this time-scale: MLR and BNN.

Basically the BB technique takes samples with replacement, that means that same data/-values can be repeated. The sample block (or window size) has a length of m consecutive observations. Then, a re-sample dataset with the same observation length as the raw

dataset selected for the training period is generated. Having same statistics metrics (as the mean and standard deviation) than the original sample.

Input Variable Selection results For this time-scale, the number of antecedent values of the target variable included in the final PS was larger ($p_{\mathsf{SF}} = 5$) than at daily basis. This value has been obtained by trial and error procedure. As in previous cases, the candidate rainfall predictors are tested under the model-based approach procedure for both tests. The performances are ranked in order of importance and the different average values are computed sequentially, $R_{a1}, \ldots, R_{a5}$. In Fig.4.7.(*Left*), the cross-correlation values (R_{xy}) among the target variable and different time lags of the stations and its averaged values are shown. A priori, this figure reveals higher values among the water heights measured in Cártama and the Rainfall datasets used as we increase the number of averaged stations based on R_{xy}.

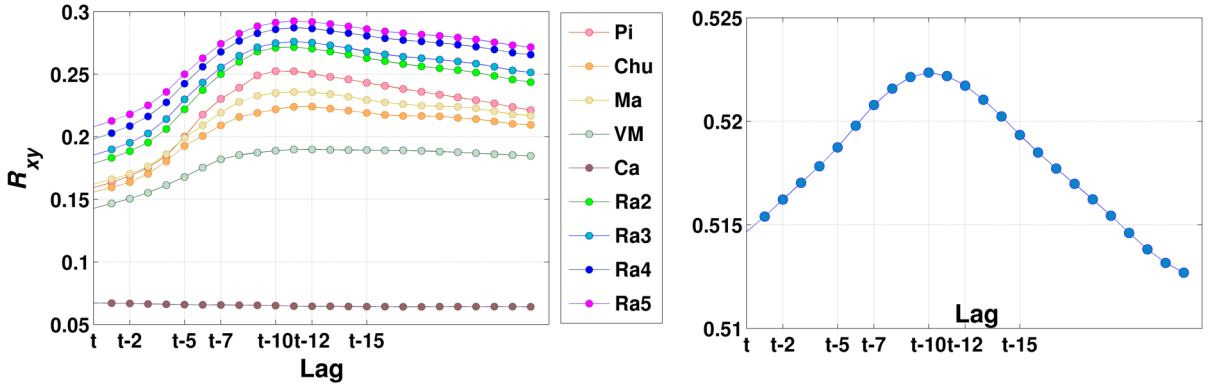


FIGURE 4.7: (*Left*) Cross-correlation R_{xy} between the water-height at Cártama ($\mathsf{WH}_{\mathtt{Ca}}$) and the hourly rainfall collected at five stations: Pizarra (`Pi`), Churriana (`Chu`), Málaga (`Ma`), Vélez-Málaga (`VM`) and Cártama (`Ca`). Also its average values are shown.(*Right*) Cross-correlation R_{xy} between the water-height at Cártama ($\mathsf{WH}_{\mathtt{Ca}}$) and the upstream reservoir releases (Q) dataset

Finally the best values for a single station are obtained by `Pi`=j , `Ma`= v, `Chu`, `VM`, and `Ca` ranked in order of importance. In this occasion, the averaged rainfall values of the best two rainfall locations (`Pi` and `Ma`) outperform the best single data set, but, do not outperform the version where we introduce both datasets independently, being the best configuration for rainfall inputs .

Also the R_{xy} values of reservoir releases (Q) are visualised in Fig.4.7.(*Right*), where finally the best antecedent values are $t-6$ y $t-7$. This result is more in line with the approximate derivative values represented in Fig.4.8. In Tab.4.3 the different configurations and their

performances are listed. With these results, finally the symbolic expression for this case study stands as follows:

$$\begin{aligned}\mathsf{WH}_{(t+h)} = f_{\mathcal{M}}(\{\mathsf{WH}_{(t-1)},\, \mathsf{WH}_{(t-2)}, \mathsf{WH}_{(t-3)}, \mathsf{WH}_{(t-4)},\, \mathsf{WH}_{(t-5)}\}, \dots \\ \{R_{j(t-4)},\, R_{j(t-5)},\, R_{j(t-6)}\}, \dots \\ \{R_{v(t-5)}, R_{v(t-6)}, R_{v(t-7)}\}, \dots \\ \{Q_{(t-6)},\, Q_{(t-7)}\})\end{aligned} \tag{4.8}$$

From this expression it can be seen that the best Predictor Structure for the input model was GH2-BB-PS3-V2, a function of the five antecedent water level values observed, the last three antecedent rainfall values registered in two stations at $t^* = 5$ and $t^* = 4$ starting window points for $j = \mathtt{Pi}$ and $v = \mathtt{Ma}$, and the last two hourly releases measures from $t^* = 6$ starting window point.

TABLE 4.3: Relevant PS evaluated with *Nlin* (ANN) and *lin* (MLR) test for the GH2-BB regression case study at hourly basis. *single*: From one station. *multiple*: the average of w stations. Best assumption for each PS and test during IVS procedure is highlighted in bold font

Basin-PS$_{[p]}$	**Symbolic Model: $\mathsf{WH}_t = f_{\mathcal{M}}(\mathbf{PS})$**	**NSE**	**PI**
	Real data (*single*)	(*Nlin*;*lin*)	(*Nlin*;*lin*)
GH2-BB-PS1$_{V1}$	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{5}, \{R_{i(t-p)}\}_{p=4}^{6})$	0.998;0.998	0.49;0.48
	Real data (*multiple*)		
GH2-BB-PS2$_{V1}$	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{5}, \{R_{\bar{a}w(t-p)}\}_{p=4}^{6})$	0.997;0.998	0.32;0.51
GH2-BB-PS2$_{V2}$	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{5}, \{R_{i(t-p)}\}_{p=4}^{6}, \{R_{j(t-p)}\}_{p=4}^{6})$	0.996;0.998	0.25;0.50
	Real data + Releases (Q)		
GH2-BB-PS3$_{V2}$	$\mathsf{WH}_{(t)} = f(\{\mathsf{WH}_{(t-p)}\}_{p=1}^{5}, \{R_{i(t-p)}\}_{p=4}^{6}, \{R_{j(t-p)}\}_{p=5}^{7}, \{Q_{t-p}\}_{p=6}^{7})$	0.997;**0.998**	0.33;**0.53**

Modelling results In Tab.4.4, the results of the "best" MLR and the "best" BNN model for the validation period are shown for the first five time horizons. The modelling simulations are computed until $h = 9$ time horizon, that have been omitted in the table by simplification purposes, since a clear linear decrease behaviour is seen as the time horizon increases. In any case, the ten temporal horizons are represented graphically in Fig.4.9 for each case, based on the PI coefficient.

During models validation period the best the MLR model found has a NSE value of 0.998 and a PI value of 0.53 for one step ahead, at t. This performance is very good for practical purposes (PI$>$ 0.5). On the other hand, for two hours ahead ($t + 1$) predictions a value

of 0.991 is obtained for the NSE coefficient and 0.35 for the PI. For $t+2$ time horizon the best MLR model has a NSE and PI values of 0.982 and 0.31 respectively, for $t+3$, values of NSE and PI equal to 0.972 and 0.27 are obtained and, finally for $t+4$, NSE and PI are equal to 0.964 and 0.26 respectively.

On the other hand, for the BNN models is found that at t step ahead forecasting the best BNN model has a NSE value of 0.997 and a PI value of 0.33. We find a very poor value for this time horizon with this model clearly exceeded by MLR. Nevertheless, for the $t+1$ time horizon, a NSE value of 0.992 is obtained and the PI value increases to 0.39 for the validation period. For this horizon we see how it works better than the MLR model, surpassing it. Finally, the Water Height forecast for the time horizon $t+2$ offers a NSE and an IP value of 0.983 and 0.32 respectively, a NSE and a PI of 0.973 and 0.28 for $t+3$, and lastly, a NSE and an PI of 0.962 and 0.23 respectively for $t+4$.

TABLE 4.4: MLR and the best ANN model obtained for Cártama at hourly scale for training (n=53244) and validation (n=17748) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures (m)

$\mathcal{M}$	h	NSE	RMSE	PI	MAE
GH2-BB-MLR	$0 \rightarrow (t)$	0.994;**0.998**	0.021;**0.024**	1.00;**0.53**	0.008;**0.009**
	$1 \rightarrow (t+1)$	0.985;0.991	0.034;0.052	0.99;0.35	0.012;0.016
	$2 \rightarrow (t+2)$	0.988;0.982	0.029;0.073	0.99;0.31	0.014;0.020
	$3 \rightarrow (t+3)$	0.977;0.972	0.039;0.092	0.99;0.27	0.017;0.027
	$4 \rightarrow (t+4)$	0.946;**0.964**	0.064;**0.105**	0.97;**0.26**	0.023;**0.034**
GH2-BB-BNN	$0 \rightarrow (t)$	0.994;0.997	0.021;0.028	1.00;0.33	0.008;0.010
	$1 \rightarrow (t+1)$	0.989;**0.992**	0.028;**0.050**	0.99;**0.39**	0.010;**0.015**
	$2 \rightarrow (t+2)$	0.976;**0.983**	0.043;**0.073**	0.99;**0.32**	0.014;**0.023**
	$3 \rightarrow (t+3)$	0.969;**0.973**	0.047;**0.091**	0.98;**0.28**	0.017;**0.026**
	$4 \rightarrow (t+4)$	0.956;0.962	0.058;0.108	0.98;0.23	0.020;0.029

Discussion During the IVS step, two versions of Rainfall inputs are tested for comparison purposes by both models. In this comparison we found that the version with only one station works better with a single location for the PS1 than for the average and the *lin* test with two locations, that is version V2. The rainfall average values PS2 as a predictor have not worked as in the daily case, where as a general rule they behave better. Finally the best behaviour is obtained with $\mathrm{PS3}_{\mathrm{V2}}$, with the inclusion of the releases and the

introduction of two independent stations. Finally comment that the releases inclusion within the final PS improves slightly the behaviour in both techniques.

The vector window of the Rainfall predictor to be introduced in the final PS is a model free based approach analysing its corresponding R_{xy} values. In this sense we have found that a maximum cross-correlation value does not assure us a better performance of the model for timing terms. According to Fig.4.7 the maximum R_{xy} values among target and rainfall variables, are obtained for lag $t-11$. Also during the selection of the optimal window we must choose the value of t^*, which is the starting time step from which we take its consecutive antecedent values. For example, for a regressive order $p=3$ the window has the following symbolic window $\{R_{t-(t^*+1)}, R_{t-(t^*+2)}, R_{t-(t^*+3)}\}$. If our election is based on the R_{xy} values finally a convenient value would be $t^*=10$ or 11 for rainfall, and $t^*=9$ or 10 for releases. This election is driven by the cross-correlation peak value. Then, the final window for $p_R=3$ would be: $\{R_{t-10}, R_{t-11}, R_{t-12}\}$.

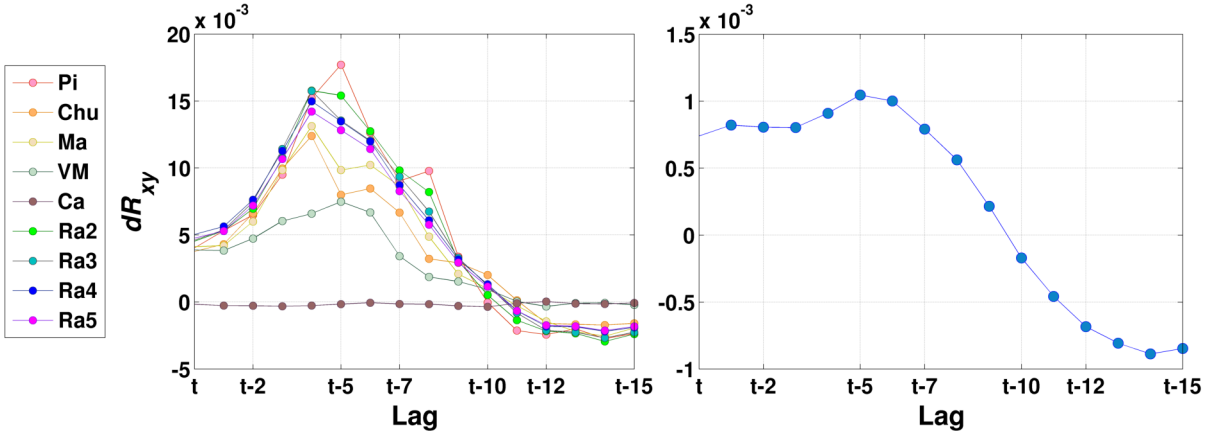


FIGURE 4.8: The approximate derivate of the cross-correlation $\delta(R_{xy})$ between Rainfall (R) values collected at different locations and Water level height WH (*Left*) and WH and upstream reservoir releases Q (*Right*) for different lags at hourly basis

In this manner, we would make sure that the predictors with the highest correlation are included in the final PS, and then to be tested. In reality we have found that this it does not assure us a better behaviour, where other values of t^* offers better performances as an optimal window for predictors and higher model behaviour in practice. In our case, the rate of change or approximate derivative of the cross-correlation values, $d(R_{xy})$, offers a more practical information for the election of the predictors, as it is represented in Fig.4.8. It is clear to see the point at which the derivative cross-correlation decreases revealing the optimal time slices of causal variables. These $d(R_{xy})$ values tell us, in a more accurate way, about the optimal point t^* to approximate the rainfall-runoff underlying function in this case study on a hourly basis.

A reasonable way to compute the concentration time is by using the cross-correlation between the hourly rainfall (R) and the water height (WH) values collected during the time period under study. The visual analysis shown in Fig.4.7 indicates that the station with maximum R_{xy} values is `Pi`, which is located in the contributing area, with the peak situated at $t^* = 11$ hours lag value. This is a reasonable value and in line with previous works (Viesca Álvarez, 2011) in the study area. On the other hand the cross-correlation time domain analysis between the releases Q and the WH also indicates an empirical approximation of the velocity values of moving wave or celerity. In this case the maximum R_{xy} peak value corresponds to a lag value of $t^* = 10$ hours. With a river course distance from the upstream reservoir to the Cartama gauge/control point of 51.49 km approximately, a celerity value of $V = 1.30\, m \cdot s^{-1}$ is obtained, being a reasonable value. If is done with the $t - 5$ value, which corresponds to the $max(d(R_{xy}))$ a celerity value of $V = 2.86\, m \cdot s^{-1}$ results, indicating where the rising limb of the flood begins. In this sense further physical space time characterisation of the wave is convenient.

The persistence values are very good for the MLR model (PI> 0.5) at $h = 0$ time horizon forecast, where appears its potential use for a practical forecasting and management tool. It can be appreciated in Figure 4.9 that the performance differences of the two different models and two PS tested here, MLR and BNN, are minimal during the validation period, from $h = 1$ time horizon. This suggest more linear *lin* relationship for the first time horizon estimation at t, or that the non-linear method fails to capture the relationship for this time horizon. Also in this figure we can observe the evolution with respect to the different temporal horizons of the maximum and the 75th percentile values obtained from the set of candidate MLR and BNN models generated trough bootstrap simulations.

Also the relative improvement over redundant performance metrics of the persist method is quantified at hourly basis. This relative terms (IC) over the NSE and RMSE persist methods for each time horizon are computed only for the validation period. Then, the best MLR model performs for one hour ahead forecasting ($h = 0 \rightarrow t$) over 0.19% and 29.65% ICs for NSE and RMSE respectively. As is presented in Tab.4.4, for the next time horizons (h=1,2,3) the best boostrapped DB method is BNN. In terms of IC of NSE compared to persist method values is 0.45% 0.65% and 0.80% for the three horizons rspectively. On the other hand, in terms of relative improvement of RMSE is 18.27% 13.36% 10.82% for $h = 1, 2, 3$ time horizons respectively. Finally the best MLR model for $t + 4$ time horizon its relative improvement values are 1.0% and 9.65% respectively.

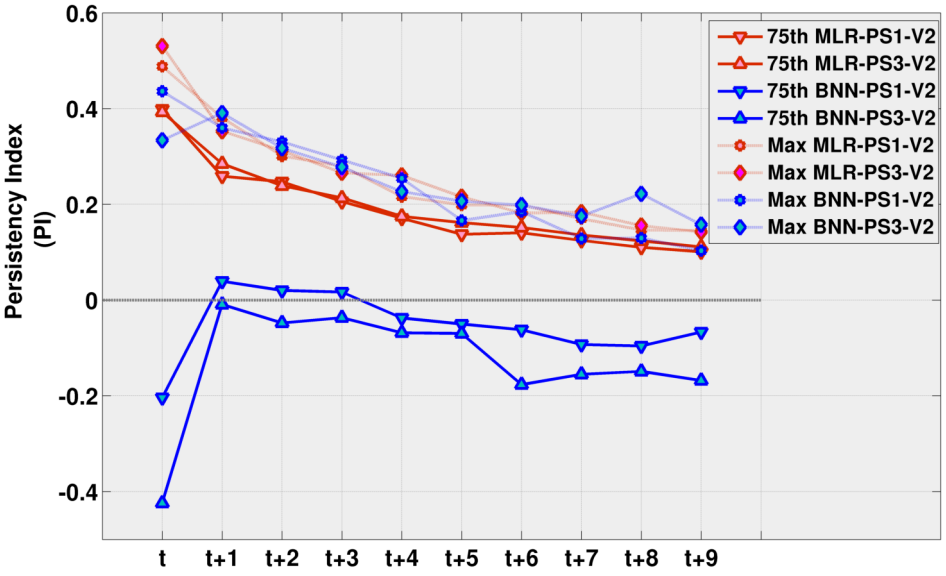


FIGURE 4.9: The maximum and the 75th percentile statistic values of 50 bootstrapped simulations at ten time horizons for both DB models at hourly basis and two different PS: PS1 and PS3 (Tab.4.3)

As it has been commented previously, one of the main disadvantages of the GP is its great computational cost, with computational complexity cost of $\mathcal{O}(n)^3$ and a computational storage cost of $\mathcal{O}(n)^2$ for a n number of training points. For this reason it is impossible to apply at hourly basis under a model-based approach. In this sense, several works (Hensman et al., 2013; Belyaev et al., 2014; Bui et al., 2015) show us different approaches and robust inference frameworks to overcome these limitations, leaving it as potential future work.

Normally, during the flood impact assessment phase, one the objective function is to minimize the cost of the water height-damage relationship of the flooding area (Hammond et al., 2015). Therefore, WH is related more quickly with alerts conditions from the water managers, since there is this direct relationship of the event and the affected areas costs. For further development physical modelling on the floodplains is necessary, being outside the objectives of this work.

4.3.3 GH3: Daily cumulated volumes forecasting experience for reservoir flood protection

Case Definition In this regression experience, the performance of three DB computational techniques (MLR, BNN and GP) are used to forecast cumulated daily streamflow

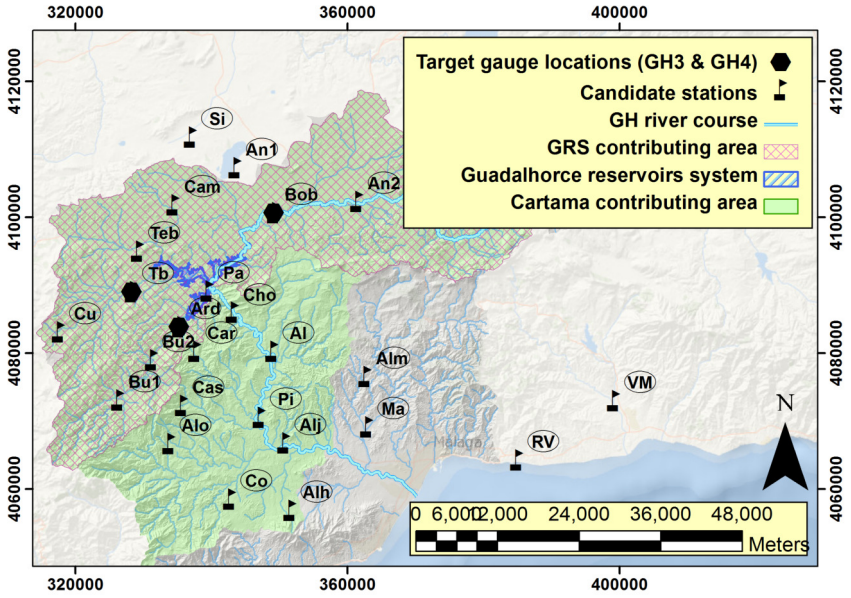


FIGURE 4.10: Map of the GH3 and GH4 case study stations at daily basis

($\mathsf{SF}_{\texttt{Ard}}$ in $Hm^3 \cdot day^{-1}$) for three time horizons ($h = 0, 1, 2$) for the best assumption (A or B) which is selected during the IVS procedure. The target variable is the total streamflow at daily scale collected at Ardales gauge station, Fig.4.10 for data points locations. This particular case is attempting to serve as preliminary forecasting experience for the Conde Guadalhorce `CGR` reservoir inflows.

The `Ard` gauge station is located at the Turón river approximately 4.39 km upstream of the reservoir tail, being the main tributary of this reservoir, so other contributions are disregarded. It has a contributing area of 271 km^2 approximately covering the 8.6% of the total area of the GH basin, with a total capacity of 66 Hm^3. This reservoir structure was inaugurated in the 1921 and have been heightened in 1947.

As predictors, Rainfall (R) at 23 different locations are used for this case. The time period modelled ranges from 01/01/2000 to 30/09/2009. More on data features can be found in Tab.B.5.

Input Variable Selection results According to the methodology exposed in Sec.4.2.1 the best rainfall data collected in a single location is obtained. Ranked in order of importance, `Alh`, `Cas`, `Pi`, `Pa`, and `Si`, are the best locations for the *lin* test. The average of this five stations ($R_{\bar{a}5}$) outperformed the best single data source `Alh`. For the *Nlin* test the best single location are: `Cas`, `Co`, `Alh`, `Si` and `Alm`, with the average of the first two ($R_{\bar{a}2}$) locations obtaining the maximum efficiency. The *lin* and *Nlin* tests do not match neither, in the best single rainfall station, or in the number of data sets to be aggregated,

with $w_{lin} = 5$ and $w_{Nlin} = 2$ respectively. The results are represented in Fig.B.11 in page 248.

The assumptions do not coincide in this case. It can be seen from Tab.4.5 that the best PS is the number 2 under assumption B (PS2B), obtaining the following symbolic expression:

$$\begin{aligned} \mathsf{SF}_{\mathtt{Ard}(t+h)} = f_{\mathcal{M}}(\{\mathsf{SF}_{\mathtt{Ard}(t-1)},\, \mathsf{SF}_{\mathtt{Ard}(t-2)}\}, \dots \\ \{R_{\bar{a}w(t)},\, R_{\bar{a}w(t-1)},\, R_{\bar{a}w(t-2)}\}) \end{aligned} \tag{4.9}$$

TABLE 4.5: Relevant PS evaluated with *Nlin* (ANN) and *lin* (MLR) test for the GH3 regression case. *single*: From one station. *multiple*: the average of w stations. Best assumption for each PS and test during IVS procedure is highlighted in bold font

Basin-PS$_{[p]}$	**Symbolic Model: SF**$_t = f_{\mathcal{M}}$**(PS)**	**NSE**	**PI**
	Real data (*single*)	(*Nlin*;*lin*)	(*Nlin*;*lin*)
GH3-PS1A	$\mathsf{SF}_{(t)} = f(\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{a1(t-p)}\}_{p=1}{}^{3})$	0.857;0.835	0.48;0.39
GH3-PS1B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{a1(t)}, R_{a1(t-p)}\}_{p=1}^{2})$	0.867;0.828	0.51;0.38
	Real data (*multiple*)		
GH3-PS2A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}w(t-p)}\}_{p=1}^{3})$	0.865;**0.848**	0.52;**0.44**
GH3-PS2B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}w(t)}, R_{\bar{a}w(t-p)}\}_{p=1}^{2})$	**0.892**;0.841	**0.61**;0.42

From Eq.4.9 it can be seen that the best Predictor Structure for the input model is a function of the volumes registered the two previous days, the average rainfall for the best five stations ($R_{\bar{a}5(t)}$) for each test for the current day and the two previous days. Alternatively, in Tab.4.5 reveals how the *lin* test works better under A assumption, instead the *Nlin* that works better under B assumption. Finally the B assumption will be used for next study steps for simplification purposes.

Modelling results In Tab.4.5 is shown that the MLR model during the validation period has a NSE value of 0.841, and a PI value of 0.42. NSE decreases to 0.758 while *PI* increases to 0.66 for $h = 1$ time horizon. Finally for $h = 2$ time ahead, NSE worsen to 0.611 and PI remains at 0.48. It can be seen that even for $h = 1$ time horizon forecasting the PI shows that the predictions improve considerably the naive assumption (PI $>$ 0.5).

For 1 day ahead, the "best" single BNN model during the validation period has a NSE value of 0.906 resulting a PI value of 0.66. This model has the maximum number of neurons units within the hidden layer tested in this work: 85. For $h = 1$ time horizon

forecasting, a NSE value of 0.832 and a PI value of 0.72 was achieved, with 50 neurons units in the hidden layer. Finally for $h = 2$ time horizon, a poor NSE value of 0.695 but a PI good value was obtained (0.59) having 30 neurons units within the hidden layer. In this case the PI values for $h = 1$ and $h = 2$ time horizons are remarkable.

TABLE 4.6: GH3-PS2B model results comparison for training (n=2575) and validation (n=644) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($Hm^3 \cdot day^{-1}$). Best performance for each horizon and metric during validation is highlighted in bold font

$\mathcal{M}$	h	NSE	RMSE	PI	MAE
MLR	$0 \rightarrow (t)$	0.826;0.841	0.081;0.060	0.30;0.42	0.023;0.026
	$1 \rightarrow (t+1)$	0.735;0.758	0.100;0.074	0.46;0.60	0.033;0.036
	$2 \rightarrow (t+2)$	0.580;0.611	0.126;0.094	0.47;0.48	0.042;0.047
BNN	$0 \rightarrow (t)$	0.963;**0.906**	0.037;**0.046**	0.85;**0.66**	0.013;**0.017**
	$1 \rightarrow (t+1)$	0.909;**0.832**	0.059;**0.062**	0.81;**0.72**	0.020;**0.025**
	$2 \rightarrow (t+2)$	0.735;**0.695**	0.100;**0.084**	0.66;**0.59**	0.029;**0.034**
$GP_{k_{P2}}$	$0 \rightarrow (t)$	0.999;0.874	0.006;0.054	1.00;0.54	0.002;0.016
	$1 \rightarrow (t+1)$	0.998;0.692	0.010;0.084	1.00;0.49	0.004;0.025
	$0 \rightarrow (t+2)$	0.344;0.507	0.157;0.106	0.16;0.33	0.029;0.035

The best GP model for $h = 0, 1$ time horizon has dot product kernel of a neural network and a squared exponential kernels: $k_P = k_{NN} \cdot k_{SE}$, with acceptable values of NSE, PI and loss functions for the training and validation period (Tab.4.6). For the t forecast, we obtain the performance values of 0.874 in NSE terms, and a PI value of 0.54. For the $t + 1$ forecast, we get a NSE value of 0.692, and a PI value of 0.49. Ultimately, for the $t + 2$ forecast, a NSE value of 0.507 and a PI value of 0.33 are obtained. We find a serious problem of over-fitting in this case.

Discussion Multiple Linear Regression (MLR), Bayesian Neural Networks (BNN) and Gaussian Processes (GP) were applied to daily cumulated streamflow at Ardales (SF_{Ard}, $Hm^3\dot{d}ay^{-1}$). The average of the best locations in each test ($w_{lin} = 5$ and $w_{Nlin} = 2$), Table 4.5, was found as the optimal input variable for Rainfall. The BNN models were found to provide more accurate cumulated streamflow than the GP and MLR for all the models tested.

From the spatial perspective point of view, the locations of three of the best five stations are situated in the runoff generation area, a logical and consequent result a priori. With regard to the other stations, one is located to the north, outside the basin (`Si`), and the other station is located in the reservoir itself (`Pa`). In this way we establish the strongest connections within the rainfall-runoff process. We also should remark the other exception on this, the `Alh` station, that gives good results in the GH1 case as well. It is situated in the downstream part of the reservoirs, nearby the seashore. With this evidence, we can affirm that some stations has better quality data than others, and that are more representative for the fronts that really have influence on the rainfall-runoff process of this sub-basin.

The station `Alh` also has an additional value due to its spatial position. It is located on the other side of the first important mountainous barrier: *la Sierra de Mijas*. This orographic accident has an altitude of up to 1100 meters right in front of the sea. Therefore, it works as a good indicator of the incoming fronts, or orographic rains, that overcome these first physical barriers and discharge into this river basin.

During the first steps of the IVS procedure, we found how the assumption A has a higher performance than B for the *lin* test, and how for the *Nlin* test, the wining assumption is B (Tab.4.5). As the efficiency and forecasting capabilities were higher for the *Nlin* test, the final assumption chosen is B for the model development and comparison phase.

BNN consistently obtained higher NSE and RMSE relative improvements than GP and MLR for both, calibration and validation. With the BNN model, the relative improvement compared to persist method, and based on the NSE values is: 25.3%, 108.4% and 169.3% for the three horizons respectively. On the other hand, the relative improvement compared to persist method, and based on the RMSE values is: 41.7%, 47.1%, and 35.85%. This indicates that the generalisation ability of the best single BNN model is better than the MLR and GP. It must be highlighted the BNN performance for $h = 1$ during the validation period with 0.832 and 0.72 for NSE and PI respectively. This provides us with a very useful tool for prediction for 1 day ahead in real life, and under a practical point of view. That would take advantage of a good rainfall forecast, as the performance for assumption B is reflecting.

The "best" GP model in this action, has a dot product kernel with the following kernels, $k_{P2} = k_{NN} \cdot k_{SE}$ for the $h = 0,1$ time horizons forecasts. For the $h = 2$ time horizon forecast, the dot product kernel is constituted by the following kernels: $k_{P1} = k_{NN} \cdot k_{lin}$. We can discover how the Neural Network is present in both dot product kernels. As

were commented before, the k_{NN} (Eq.A.10) has a non-stationary nature kernel, being always convenient to be evaluated. Our water systems should be modelled by different approaches than can deal with their particularities: Seasonal flows, rare and intense events. For this reason, to try the dot product kernels, looks always to worth it. With this DD computational technique there have been problems of over-fitting, with very good performance during training and failing validation. We applied other types of techniques during the estimation of parameters as Grid or Monte Carlo Integration (Vanhatalo et al., 2013), but we didn't observe any improvement.

The experience for this water system has been good in general terms. The difference of the *lin* and *Nlin* models is considerably greater than the previous case on a daily scale, GH1. In this case we can say that the system responds much better to a non-linear system. This can be seen by comparing the results of MLR and BNN models, or also by the "best" kernel obtained and the linear one. This gives us an idea that these types of basins located further inland have a stronger non-linear response than the areas situated on the coastline.

4.3.4 GH4: Daily cumulated volumes forecasting experience for reservoir flood protection

Case Definition In this regression experience, the performance of three DB computational techniques (MLR, BNN and GP) are used to forecast cumulated daily inputs to Guadalteba-Guadalhorce reservoir system ($\mathsf{SF}_{\mathtt{GGR}}$ in $Hm^3 \cdot day^{-1}$), for three time horizons ($h = 0, 1, 2$) and for the best assumption (A or B) which is selected during the IVS procedure. The Guadalteba-Guadalhorce cumulated reservoir inflow is the target variable established for this case study, serving as a preliminary forecasting experience for Flood Control purposes on reservoir operation at daily basis. This data is calculated by adding up the mean streamflow values collected at daily scale, in two gauge stations: (i) Teba and (ii) Bobadilla (Fig.4.10 for locations, Tab.B.5 for its statistical parameters, and Fig.B.4 for the visualisation of the resulting hydrograph). These gauging points are practically at the tails of the two reservoirs, at which point the assumption is made that the sum of both capacities is equivalent to the total of the inflow to the reservoirs, disregarding other contributions. The double reservoir has a contributing area of 1431 km^2 approximately covering the 45.33 % of the total area of the entire GH River Basin, with a total storage capacity of 306.2 Hm^3.

As predictors, we used the Rainfall (R) collected at 23 different locations. The time period modelled runs from 01/01/2000 to 30/09/2009. More on data features can be found in Tab.B.5.

Input Variable Selection results According to the method exposed in Sec.4.2.1, the best rainfall inputs based on a single location data were `An2`, `Alm`, `Al`, `VM`, and `Pa`, listed in order of importance for the *lin* test. The average of these three stations ($R_{\bar{a}3}$), slightly outperformed a single data source. For the *Nonlin* test, the best single location is: `An2`, `Si`, `Al`, `Alm`, and `Camp`, rank in order of importance. The average of the first four locations ($R_{\bar{a}4}$), outperformed a single one. Here, the w values for each test differs, with $w_{lin} = 3$ and $w_{Nlin} = 4$ values. We represented the results in Fig.B.11 in page 248. Tab.4.7 shows that the best PS is the number 2 under assumption B (PS2B). Then we get the following symbolic expression:

$$\begin{aligned} \mathsf{SF}_{\mathtt{GRS}(t+h)} = f_{\mathcal{M}}(\{\mathsf{SF}_{\mathtt{GRS}(t-1)},\, \mathsf{SF}_{\mathtt{GRS}(t-2)}\}, \ldots \\ \{R_{\bar{a}w(t)},\, R_{\bar{a}w(t-1)},\, R_{\bar{a}w(t-2)}\}) \end{aligned} \tag{4.10}$$

TABLE 4.7: Relevant PS evaluated with *Nlin* (ANN) and *lin* (MLR) test for the GH4 regression case. *single*: From one station. *multiple*: the average of w stations. Best assumption for each PS and test during IVS procedure is highlighted in bold font

Basin-PS$_{[p]}$	**Symbolic Model: SF**$_t = f$**(PS)**	**NSE**	**PI**
	Real data (single)	(*Nlin*;*lin*)	(*Nlin*;*lin*)
GH4-PS1A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}1(t-p)}\}_{p=1}^{3})$	0.773;0.741	0.37;0.27
GH4-PS1B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}1(t)}, R_{\bar{a}1(t-p)}\}_{p=1}^{2})$	0.793;0.761	0.42;0.33
	Real data (multiple)		
GH4-PS2A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}w(t-p)}\}_{p=1}^{3})$	0.777;0.761	0.38;0.33
GH4-PS2B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}w(t)}, R_{\bar{a}w(t-p)}\}_{p=1}^{2})$	**0.800**;0.770	**0.44**;0.36

From Eq.4.10 it can be seen that the best Predictor Structure for the input model is a function of the volumes registered the two previous days and the mean rainfall value of the best, three and four stations for *lin* and *Nlin* test respectively.

Modelling results Table 4.8 shows evaluation coefficients for 1, 2 and 3 days ahead time horizons for three different DB computational techniques. At $h = 0$ time horizon,

the MLR model during the validation period has a NSE value of 0.770, and a PI value of 0.36. For $h = 1$, the NSE value is 0.725 and PI value is 0.48. Finally for $h = 2$ time horizon, the value of NSe decreases to 0.596 and that of PI to 0.33. Poor performance in general for the all the horizons h tested with MLR.

Table 4.8 also shows evaluation coefficients for the best BNN for 1, 2 and 3 days ahead. For 1 day ahead, the best BNN model has a NSE value of 0.841, and a PI value of 0.56 during the validation period. This model has 50 neurons units in the hidden layer. For 2 day ahead, the model has a NSE value of 0.781 and a PI value of 0.59, with 60 neurons units in the hidden layer. Finally for h=2 time ahead, the model has a NSE value of 0.563 and a PI value of 0.42, with 60 neurons units within the hidden layer.

The best GP model for $h = 0$ time horizon forecast has a linear and neural network product kernel: $k_{P1} = k_{lin} \cdot k_{NN}$, with acceptable values of NSE, PI and loss functions for the training and validation period. More on kernels comparison for this case study in pag.228. A NSE value of 0.755 and a PI value of 0.31 is obtained for one day ahead ($h = 0$). For 2 day ahead, NSE is -1.591 and a PI is -3.90. Finally for 3 day ahead ($h = 2$) we obtain a NSE value of 0.616 and a PI value of 0.36. The poor value for h=1 time horizon is due to a huge overestimation of particular event with the selected model. This generates a huge peak and an event outlier. Further analysis in this case is needed.

TABLE 4.8: GH4-PS2B results for training (n=2575) and validation (n=644) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($Hm^3 \cdot day^{-1}$). Best performance for each horizon and metric during validation is highlighted in bold font

$\mathcal{M}$	h	NSE	RMSE	PI	MAE
MLR	$0 \rightarrow (t)$	0.879;0.770	0.151;0.237	0.34;0.36	0.051;0.074
	$1 \rightarrow (t+1)$	0.822;0.725	0.184;0.259	0.44;0.48	0.069;0.095
	$2 \rightarrow (t+2)$	0.724;0.596	0.229;0.314	0.33;0.33	0.091;0.118
BNN	$0 \rightarrow (t)$	0.927;**0.841**	0.117;**0.197**	0.60;**0.56**	0.040;**0.058**
	$1 \rightarrow (t+1)$	0.843;**0.781**	0.172;**0.231**	0.51;**0.59**	0.060;**0.079**
	$2 \rightarrow (t+2)$	0.747;**0.651**	0.219;**0.292**	0.39;**0.42**	0.080;**0.105**
GP_{k_P}	$0 \rightarrow (t)$	0.999;0.755	0.017;0.245	0.99;0.31	0.007;0.071
	$1 \rightarrow (t+1)$	0.995;-1.591	0.032;0.795	0.98;-3.90	0.014;0.109
	$2 \rightarrow (t+2)$	0.924;0.616	0.120;0.306	0.82;0.36	0.043;0.103

Discussion Multiple Linear Regression (MLR), Bayesian Neural Networks (BNN) and Gausian Processes (GP) were applied to daily sum of cumulated streamflow at Bobadilla and Teba ($\mathsf{SF}_{\mathsf{GRS}}$, $Hm^3 \cdot day^{-1}$).

During the rainfall predictors mapping, IVS procedure, the best input variable for Rainfall is the average of the best locations in each test ($w_{lin} = 3$ and $w_{Nlin} = 4$), see Table 4.7. The best "single" station coincides for both tests: `An2`, situated in the runoff generation area. In both tests, the `Al` and `Alm` stations appear in the top five, without being in the runoff generation area. Nevertheless, these stations are situated in convenient locations for predicting the river behaviour. As in the previous experience, it is worth noting the good position of the `Si` station in the $NLin$ test. This station is located at the north of the basin, but outside of the runoff generation area. We see this same behaviour with the `VM` case for the lin test, since a priori is far from the area of influence (Fig.4.10) and belongs to another river basin (Vélez River). This may also be due to the good data quality collection in this station, or because, by its position, it becomes a good indicator for fronts characterisation in the lin test. This station is ranked in eight position in the $Nlin$ case (Fig.B.11). The results show that the locations of the stations does not seem to impact the efficiency of the models.

As in previous case, BNN consistently obtained higher NSE and RMSE relative improvements than GP and MLR for both calibration and validation periods. In terms the relative improvement of NSE compared to persist method, values are 31.0%, 66.1% and 65.1% for the three horizons respectively during the validation period. On the other hand, in terms of relative improvement of RMSE, these are 33.4%, 35.7%, and 24.1% during the validation period. This indicates that the generalisation ability of the best single ANN model is better than the MLR and GP.

4.4 GF regression experiences for forecasting purposes

The Guadafeo River Basin is strongly driven by the snow presence and hydrological processes with high energy pulses related to its steep topographic gradients. It is situated on the southern part of the Sierra Nevada range, which has the highest peak of the Iberian peninsula the Mulhacén (3479 m, Fig.2.1). Extended details about this water system and its hydro-meteorological framework can be found in Chap.2.

The snow processes are crucial for the hydrological behaviour of this basin. To be able to include these snow effects into the rainfall runoff dynamics, and due the absence of direct measurements, we decide to use a hydrological modelling approach. The snow variables are modelled, and used as a predictor to capture the basin's pluvial-nival regime. These snow variables are generated with the WiMMed model, a physically based, fully distributed hydrological model (Polo et al., 2009), especially developed to represent the snow hydrology and dynamics under Mediterranean climate (Herrero et al., 2009, 2012).

In this River Basin, three daily regression experiences at daily basis are carried out. The first two regression experiences are the reservoir inflows modelling for forecasting purposes at Rules (GF1, Sec.4.4.1) and Béznar reservoirs (GF2, Sec.4.4.2). The other regression experience is related to streamflow forecasting in a mountainous subbasin (Trevélez) named GF3 case study (Sec.4.4.3). In this subbasin an important water extraction for urban water use is taking place.

4.4.1 GF1: Daily mean forecasting of reservoir inflows.

Case Definition The performance of three DD computational techniques (MLR, BNN and GP) are used to forecast daily mean Rules reservoir inflows ($\mathsf{SF}_{\mathtt{Ru}}$ in $m^3 \cdot s^{-1}$) for different time horizons ($h = 0, 1, 2$) for the best assumption (A or B), selected during the IVS procedure (Sec.4.2.1). This regression case allows us to predict daily mean of the reservoir inflows at Rules as a control point for water management purposes. The Rules Reservoir is located at the Guadalfeo river. Collects, on one hand, the waters of the Ízbor and Lanjarón tributaries, and on the other hand, the waters of the southern part of Sierra Nevada, which comprises the Trevélez, Poqueira, Cádiar, Chico and Sucio tributaries. Part of the contributing area is regulated by Béznar reservoir which is not taken into account in this case but is modelled separately in the next case study. The Rules reservoir has an unregulated contributing area of 718 km^2 covering the 55.1 % of the total area of the GF basin, with a total storage capacity of 117 Hm^3.

In such water systems, the estimation of the streamflow at different horizons is important for reservoir operation policy (mitigation of floods and reservoir protection). The daily reservoir inflows dataset was facilitated by the reservoir authorities. For this experience, the rainfall data sets used come from 14 meteorological stations and the temperature from 6 meteorological stations, for record period from 01/10/2003 to 31/08/2013.

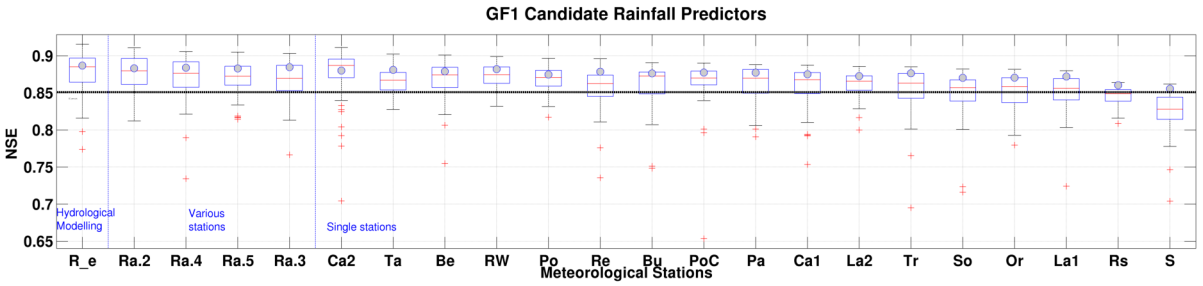


FIGURE 4.11: IVS for rainfall inputs of the GF1 case study (Sec.4.4.1), sorted by their maximum value reached by the *Nlin* test outcomes (boxplot). The corresponding *lin* results values are also presented (grey dots). The 5 boxplots on the left side correspond the R_e and the Rainfall averages values $R_{\bar{a}w}$ of the best single stations. The horizontal dashed black line represents the $\mathsf{EC}_{persist}$ value

In parallel, three variables are modelled with the WiMMed hydrological model: total Precipitation (R_{Wi}, $mm \cdot day^{-1}$), Snowfall (R_S, $mm \cdot day^{-1}$), the effective rainfall (liquid fraction, R_e, $mm \cdot day^{-1}$) and Snowmelt (S, $mm \cdot day^{-1}$) promediated in the whole contributing area, which are used as streamflow predictors as well. All the time series used and their main statistics are listed in Tab.B.6. In addition, a brief analysis of the hydro-meteorology reproduced by the WiMMed model during the study period is carried in App.B.2.3.

Input Variable Selection results According to the methodology exposed in Sec.4.2.1, the best rainfall dataset was collected at stations `Ta`, `Ca2`, `Be`, `Re`, and `PoC`, listed in order of importance. The effective rainfall from the model (R_e) was the best input for the *lin* analysis (MLR, Fig.4.11). On the other test, the best rainfall dataset was collected at `Ca2`, `Ta`, `Be`, `Po`, and `Ca1`, listed in order of importance, where we also found the R_e as the best rainfall predictor for the *Nlin* analysis (BNN). In this particular case, the best Rainfall predictor is a modelled precipitation that takes into account the snow dynamics. Tab.4.9 shows that the best PS is the version 4, under the assumption A (PS4A). Finally, the following symbolic expression is obtained for both approaches:

$$\begin{aligned} \mathsf{SF}_{\mathtt{Ru}(t+h)} = f_{\mathcal{M}}(\{\mathsf{SF}_{\mathtt{Ru}(t-1)}, \mathsf{SF}_{\mathtt{Ru}(t-2)}\}, \dots \\ \{R_{e(t-1)}, R_{e(t-2)}, R_{e(t-3)}\}, \dots \\ \{S_{(t-1)}, S_{(t-2)}\}) \end{aligned} \tag{4.11}$$

TABLE 4.9: Relevant PS evaluated with *Nlin* (f_{ANN}) and *lin* (f_{MLR}) test for the GH1 case. *single*: From one station. *multiple*: the average of w stations. Best assumption for each PS and test during IVS procedure is highlighted in bold font

Basin-PS	**Symbolic Model: $\mathsf{SF}_t = f_{\mathcal{M}}(\mathbf{PS})$**	**NSE**	**PI**
	Real data (*single*)	(*Nlin*;*lin*)	(*Nlin*;*lin*)
GF1-PS1A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{a1(t-p)}\}_{p=1}^{3})$	0.912;0.909	0.41;0.37
GF1-PS1B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{a1(t)}, R_{a1(t-p)}\}_{p=1}^{2})$	0.910;0.907	0.40;0.38
	Real data (*multiple*)		
GF1-PS2A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}w(t-p)}\}_{p=1}^{2})$	0.915;0.906	0.40;0.37
GF1-PS2B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}, \{R_{\bar{a}w(t)}\}_{p=1}^{2}, R_{\bar{a}w(t-p)}\}_{p=1}^{3})$	0.909;0.906	0.39;0.37
	Real data + WiMMed		
GF1-PS3A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{e(t-p)}\}_{p=1}^{3})$	0.919;0.899	0.46;0.32
GF1-PS3B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}, \{R_{e(t)}, R_{e(t-p)}\}_{p=1}^{2})$	0.919;0.898	0.46;0.32
	Real data + WiMMed (Snowmelt)		
GF1-PS4A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{e(t-p)}\}_{p=1}^{3}, \{S_{(t-p)}\}_{p=1}^{2})$	**0.925**;0.908	**0.50**;0.38
GF1-PS4B	$\mathsf{SF}_{(t)} = f[\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{e(t)}, R_{e(t-p)}\}_{p=1}^{2}, \{S_{(t-p)}\}_{p=1}^{2})$	0.922;0.907	0.48;0.38

From Eq.4.11, we can see that the best Predictor Structure for the input model was a function of the streamflow values observed in the past two previous day: the effective Rainfall reproduced for the three previous days, and the simulated Snowmelt of the two antecedent days in the study area.

Modelling results For $h = 0$ lead time forecasting, or one day ahead as we are under assumption A, the MLR model has a NSE value of 0.908, with its corresponding PI value of 0.38. In Tab.4.10 we can observe that, for 2 day ahead, the MLR model has a NSE value of 0.809 and a PI value of 0.29. Finally, for 3 day ahead the MLR DD approximation has a NSE value of 0.676 and a PI value of 0.19. In general, the performance of the MLR in this case is poor.

For 1 day ahead, the "best" BNN model during the validation period has a NSE value of 0.935 and an acceptable PI value of 0.57. This model has 21 neurons units in the hidden layer. This particular case, under assumption A, has good deterministic outcomes for practical purposes. This can be appreciated in Fig.4.12, where the bigger event during the validation period is visualised. It can be seen how the peak timing is modelled successfully for the two horizons represented ($h = 0, 1$). For two days ahead predictions, we can see how it starts to appear the unwanted lag effect. Also the results for two days ahead are represented in this figure, having a NSE value of 0.836 and a PI value of 0.39; this model uses 60 neurons in the hidden layer. Finally, the outcomes for 3 day ahead have a NSE value of 0.715 and a PI value of 0.29, with 85 neurons units in the hidden

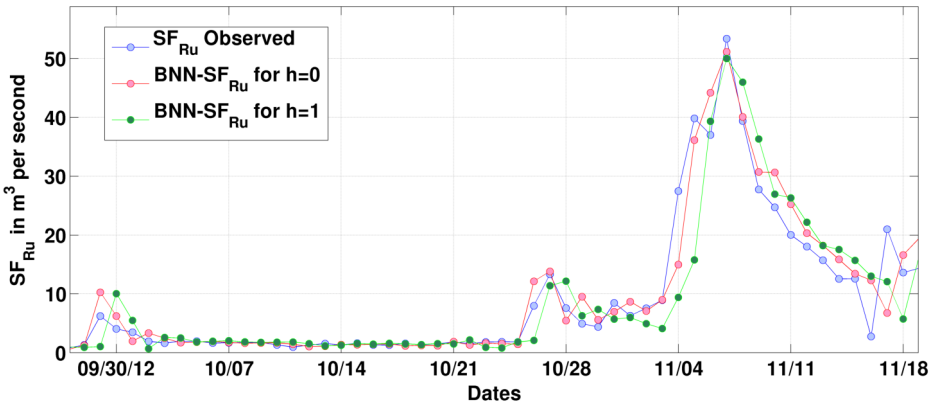


FIGURE 4.12: Hidrogram visualisation detail for the bigger event reproduced during the validation period in the GF1 case study

layer, being the one with the best performance for 3 days ahead among all the models compared. For this particular case, it is interesting to note how the number of hidden units increase with the time horizon.

The GP-k_{P1} model performs better than the others KCF tested for two lead time forecast, with higher NSE, PI and lower RMSE, MAE for the validation period. For the kernel comparison refer to Tab.A.5. GP-k_{lin} has the best performance on efficiency and predictable capabilities for the $h = 2$ time forecasting. Finally, the GP-k_{P1} model has a NSE value of 0.928, and a PI value of 0.52 during the validation period for 1 day ahead. On the other hand, the GP model has a NSE value of 0.811 and a PI value of 0.30 for 2 day ahead. Finally, this GP-k_{lin} has a NSE value of 0.598 and a PI value of 0.00 for 3 day ahead. It must be highlighted that for this last time horizon, the result has the same behaviour as the persist method PI=0, which indicates that the model is not forecasting better than the persist model.

TABLE 4.10: GF1-PS4A results for training (n=2898) and validation (n=973) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($m^3 s^{-1}$). Best performance for each horizon and metric during validation is highlighted in bold font

$\mathcal{M}$	h	NSE	RMSE	PI	MAE
MLR	$0 \rightarrow (t)$	0.913;0.908	2.389;1.622	0.34;0.38	0.810;0.803
	$1 \rightarrow (t+1)$	0.831;0.809	3.324;2.341	0.27;0.29	1.035;1.057
	$2 \rightarrow (t+2)$	0.744;0.676	4.093;3.044	0.21;0.19	1.331;1.343
BNN	$0 \rightarrow (t)$	0.981;**0.935**	1.120;**1.362**	0.85;**0.57**	0.541;**0.672**
	$1 \rightarrow (t+1)$	0.862;**0.836**	3.002;**2.170**	0.40;**0.39**	0.847;**0.936**
	$2 \rightarrow (t+2)$	0.758;**0.715**	3.978;**2.857**	0.25;**0.29**	1.185;**1.292**
GP-k_{P1}	$0 \rightarrow (t)$	0.998;0.928	0.315;1.436	0.99;0.52	0.181;0.685
	$1 \rightarrow (t+1)$	0.999;0.811	0.291;2.328	0.99;0.30	0.157;1.045
	$2 \rightarrow (t+2)$	0.998;0.598	0.339;3.393	0.99;0.00	0.151;1.470

Discussion Multiple Linear Regression (MLR), Bayesian Neural Networks (BNN) and Gausian Processes (GPs) were applied to predict daily mean of the inflow at Rules Reservoir control point. In this case, should be highlighted the good results obtained with the `Ca2` station during the IVS procedure for both tests. This single station (`PS1`) is the remotest one (Fig.2.1), but it captures better the relevance of the fronts that come from the sea and precipitate in this singular coastal basin. At the beginning of the study, the temperature collected at different attitudinal locations, was expected to be a key predictor. We validated several cases with different combinations without finding an improvement.

The inclusion of the WiMMed model simulating the snow agent within the rainfall-runoff process drastically changes the results for the better. This improvement can be observed during IVS procedure, on the superior left side of the panel in Fig.4.11, with the results obtained of the R_e variable as a rainfall predictor. In such basins, the non-linear capabilities of the BNN make it work better compared to the rest once when we apply the physical modelling. That means that the successful assumption is A when the hydrological modelling data is used in the regression problem. If the hydrological modelling is not used, the successful assumption will be B. As the results suggest, with physical modelling we are capable to faithfully reproduce the dynamics of the basin, and its influence, and thus gain predictive skills on the time-line.

Among the different DD compared, the BNN provides more accurate maximum water levels than the GPs and MLR for the deterministic estimations for all the time horizons and all performance metrics tested. In terms the relative improvement (IC) of NSE compared to persist method, values are 9.9%, 14.4% and 19.5% for the three horizons respectively during the validation period. On the other hand, in terms of relative improvement of RMSE, they are 34.1%, 21.9%, and 15.8% during the validation period. In this case the MLR exhibits a strong lag effect under assumption A.

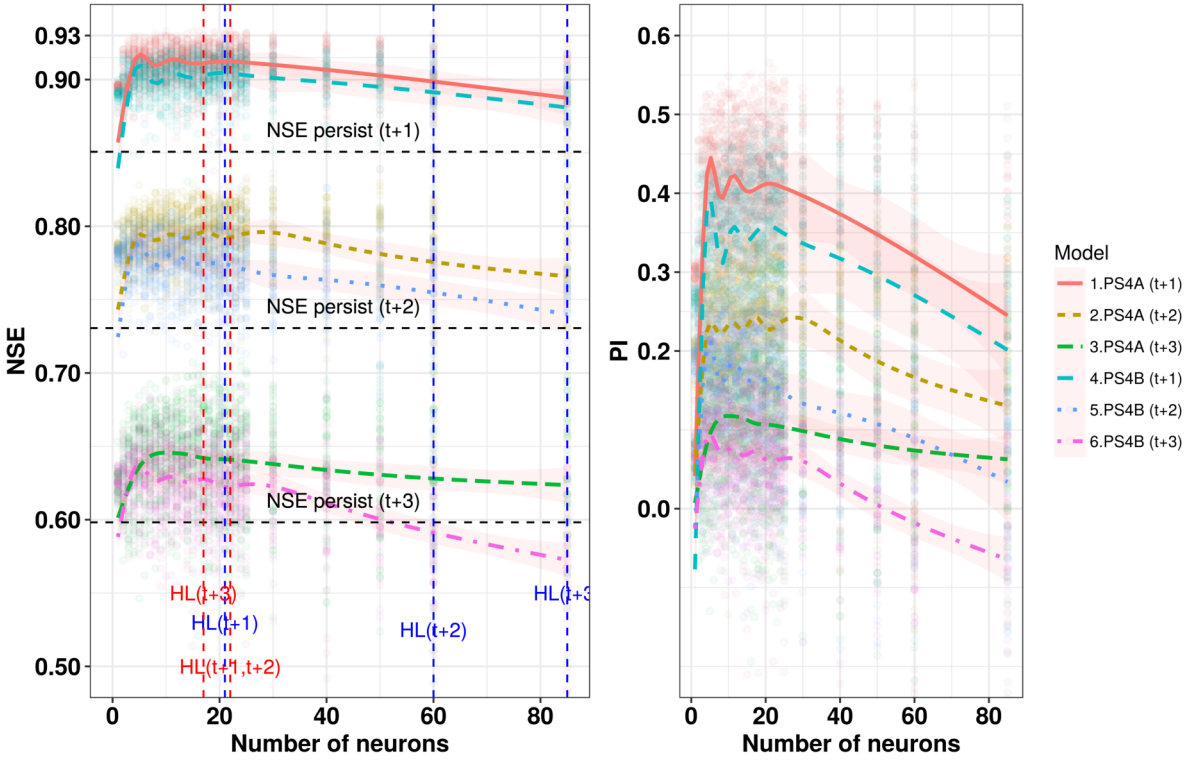


FIGURE 4.13: GF1 BNNs multiple simulations visualisation: regression on NSE and PI values (y axis) versus the number of neurons within the hidden layer (HL, x axis) comparing PS4A and PS4B for this case study under a model-based approach. Black horizontal line represent the NSE persist values for $t+1$, $t+2$ and $t+3$ time horizons predictions. Blue and Red vertical lines are the HL units of the best model and the 75th percentile respectively for the final PSA4 and its three time horizons.

The final results shown by the BNN model in this section, is the "best" single model obtained under the framework of multiple *dynamic* simulations (more refer to pag.215). We tested several models with an increasing number of neuron units. For this case, the values are quite high, 21, 60 and 85 for $h = 0, 1, 2$ horizon respectively. In Fig.4.13 these values are marked by a vertical blue dashed line. It takes higher values as the time horizon h increases. However, taking a more statistical perspective over the general models behaviour from all these simulations, we obtain a more general and clear idea of

its behaviour regarding the "optimal" number of neurons within the hidden layer as a general rule. Therefore, it is more convenient to focus on the 75th percentile values in each set of the 50 random runs per number of neurons units tested. Then, focusing on the 75th percentile, 22, 22 and 17 neurons units are the values of neurons units within the hidden layer for $h = 0, 1, 2$ time horizons respectively. In the figure, they are visualised by a vertical red dashed line.

Within GP applications, we can change the KCF for different horizons. Here, the final selected kernel is k_{P1}, since it offers the best results for the first two temporal horizons. The k_{P1} is a dot product kernel between a linear (Eq.A.9) and a neural network (Eq.A.10) kernel. In Tab.A.6 we mirrored this comparison. This table mirror that for 1 and 2 days lead time forecasting the performance of the product kernel k_P is better compared to the remaining kernels. In Fig.A.13 the real precipitation-runoff function in its domain is displayed. In this figure, we see the visualization of the kernels separately, and also the final product among them. In this way, we can appreciate what each kernel contributes to the final result. But, it is important to remark that, for $h = 2$ lead time forecasting, the performance of the k_{lin} outperformed the remaining non-linear KCFs.

Given the hydrological complexity of the GF water system, this regressive experience has been successful. We counted on good data, measured in a closed control volume such as is a dam, with well quantified inputs and outputs. This reminds us that, to achieve the initial objectives planted in this work, the quality of the real data is crucial.

4.4.2 GF2: Daily mean forecasting of reservoir inflows

Case Definition The performance of three DD computational techniques (MLR, BNN and GP) are used to forecast daily mean Béznar reservoir inflows ($\mathsf{SF}_{\mathtt{Be}}$ in $m^3 \cdot s^{-1}$) for different time horizons ($h = 0, 1, 2$) and for the best assumption (A or B), which is selected during the IVS procedure. The `Be` reservoir is located on the river Ízbor, main tributary of the Guadalfeo (Fig.4.14). It has a contributing area of 352 km^2 approximately, covering the 27.01 % of the total area of the basin. It has a total capacity of 57.2 Hm^3. The regression experience proposed here aims to forecast the inflows at Béznar reservoir ($\mathsf{SF}_{\mathtt{Be}}$), location in Fig.2.1. This reservoir was built to improve the agriculture irrigation requirements of the subtropical and greenhouses crops of the Motril-Salobreña region. It also guarantees urban water supply and water volumes to the Ízbor hydropower plant. Nowadays, it also helps Rules reservoir to accomplish its objectives. In fact, the both work as a single reservoir system and are managed together.

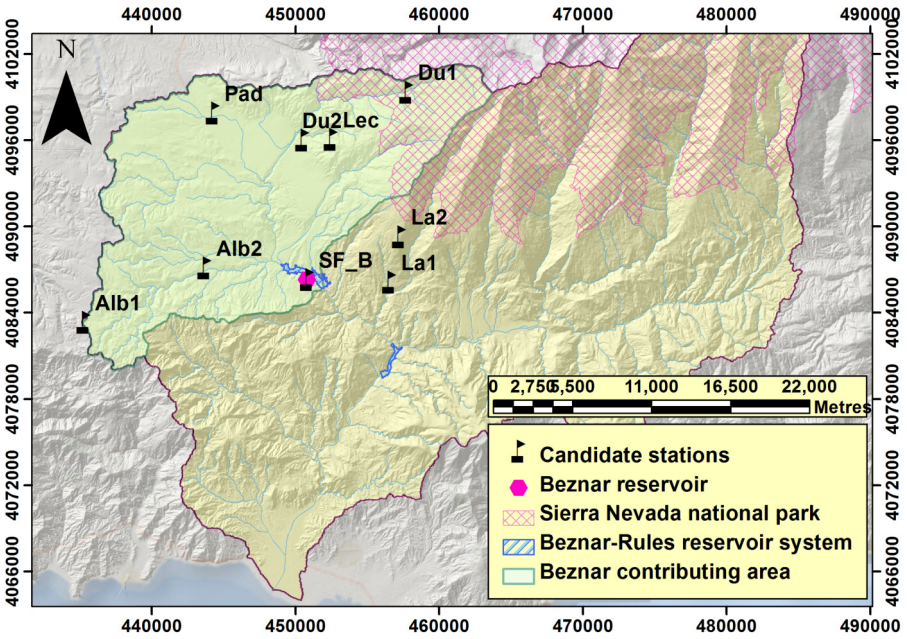


FIGURE 4.14: Northern western part of the Guadalfeo River Basin showing location of different stations at daily basis. The distributed hydrological modelling simulations has been carried out for the entire Béznar contributing area

The daily reservoir inflows dataset was facilitated by the reservoir authorities. Rainfall collected data used comes from 8 meteorological stations: `Alb1`, `Du1`, `Du2`, `Alb2`, `La1`, `La2`, `Pad` and `Lec`, whose locations can be seen in Fig.4.14. In this case, the time period modelled runs from 04/09/2000 to 30/12/2013. More information about the hydrological framework and statistical parameters can be found in Tab.B.7.

The training and validation data sets have n_T=3893 and n_V=973 observations respectively. As the previous case (GF1), hydrological modelling (WiMMed) is used to reproduce relevant snow variables within the basin: total Precipitation (R_{Wi}, mm·day^{-1}), Snowfall (R_S, mm·day^{-1}), the effective Rainfall (liquid fraction, R_e, $mm \cdot day^{-1}$), and Snowmelt (S, $mm \cdot day^{-1}$). These variables are reproduced for the correspondent contributing area at the point of gauging, and same time period. On this case study, the DD modelling attempts to analyse the influence of the hydrological modelling on the experience performance.

Input Variable Selection results According to the methodology exposed in Sec 4.2.1, the best rainfall inputs coming from data collected in a single location were `Lec`, `Alb1`, `La2`, R_S, and S for the *lin* test. In this case, the average of various stations do not outperformed the data from a single station. For the *Nlin* test, the best single rainfall datasets were collected at `Pad`, R_{Wi}, `Du2`, `Lec` and `La2` listed in order of importance. For

this test, the average of the best two stations ($R_{\bar{a}_4(t)}$) outperformed the data collected at a single one. The linear and nonlinear test do not match either in the best single rainfall station or with the averages data sets. For this case, a single value is the best choice for the *lin* case ($w_{lin} = 1$), and the average of the first four for the *Nlin* case ($w_{Nin} = 4$), getting better performances. We find again that the use of a greater number of stations as input variables do not assure us a better representation of the rainfall, and therefore of the results.

Regarding the variables reproduced with the distributed hydrological model, we can see they have a good behaviour in both tests. For instance, the rainfall reproduced by the model (R_{Wi}) is the second best connection for the *Nlin* test. Also, it can be observed that pure snow variables such as R_S and S are the best fourth and fifth for the *lin* test. The IVS results are represented in Fig.B.11. Finally, it can be seen from Tab.4.11 that the best PS is number 3 under the assumption B (PS3B), obtaining the following symbolic expression for both approaches:

$$\begin{aligned} \mathsf{SF}_{\mathtt{Be}(t+h)} = f_{\mathcal{M}}(\{\mathsf{SF}_{\mathtt{Be}(t-1)},\, \mathsf{SF}_{\mathtt{Be}(t-2)}\}, \ldots \\ \{R_{\bar{a}w(t)},\, R_{\bar{a}w(t-1)},\, R_{\bar{a}w(t-2)}\}, \ldots \\ \{S_{(t-1)},\, S_{(t-2)}\}) \end{aligned} \tag{4.12}$$

TABLE 4.11: Relevant PS evaluated with *Nlin* (ANN) and *lin* (MLR) test for the GF2 case. *single*: From one station. *multiple*: the average of w stations. Best assumption for each PS and test during IVS procedure is highlighted in bold font

Basin-PS	**Symbolic Model: $\mathsf{SF}_t = f_{\mathcal{M}}(\mathbf{PS})$**	**NSE**	**PI**
	Real data (*single*)	(*Nlin*;*lin*)	(*Nlin*;*lin*)
GF2-PS1A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{a_1(t-p)}\}_{p=1}^{3})$	0.860;0.803	0.13;-0.22
GF2-PS1B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{a1(t)}, R_{a1(t-p)}\}_{p=1}^{2})$	0.906;0.825	0.41;-0.08
	Real data (*multiple*)		
GF2-PS2A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}2(t-p)}\}_{p=1}^{3})$	0.855;0.809	0.11;-0.19
GF2-PS2B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}2(t)}, R_{\bar{a}2(t-p)}\}_{p=1}^{2})$	0.903;0.765	0.39;-0.48
	Real Data + WiMMed (Snowmelt)		
GF2-PS3A	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}2(t-p)}\}_{p=1}^{3}, \{S_{(t-p)}\}_{p=1}^{2})$	0.851;**0.817**	0.12;**-0.13**
GF2-PS3B	$\mathsf{SF}_{(t)} = f(\{\mathsf{SF}_{(t-p)}\}_{p=1}^{2}, \{R_{\bar{a}2(t)}, R_{\bar{a}2(t-p)}\}_{p=1}^{2}, \{S_{(t-p)}\}_{p=1}^{2})$	**0.908**;0.771	**0.41**;-0.44

Eq.4.12 indicates that the best Predictor Structure for the input model was a function of the streamflow values observed in the past two previous day, the average Rainfall of two

stations for the current day and two previous days, and the simulated Snowmelt values of the two previous day.

During the wrapper methodology procedure followed to build the final PS, we appreciated the poor performance for the *lin* test in general, with a completely different behaviour in the *Nlin* test. As we build the predictor sequentially, PS1 $\rightarrow$ PS2 $\rightarrow$ PS3, we see how the efficiency and predictability for the *lin* test decreases. On the other hand, the efficiency and predictability for the Nlin test increases as we advance steps during the IVS procedure. Finally, we choose the PS3B, due to the values obtained for the *Nlin*, since they have the maximum value for both performance metrics.

Modelling results For the best model, GF2-MLR-PS3B, we obtained values of 0.771 for NSE and -0.44 for PI during the validation period, and for the time horizon h=0. The performance of this specific case is very poor. However, its behaviour improves the persist method for the other two time horizons tested, with NSE values of 0.786 and 0.742 for the NSE coefficient, and 0.07 and 0.08 for the PI coefficient. These values are for $h = 1, 2$ time horizons respectively. In this case, the behaviour of this model under assumption B, has not been always superior to assumption A.

Table 4.12 shows evaluation coefficients for the best BNN model for each time horizon and same PS. For h=0 day ahead, the best BNN model during the validation period has a NSE value of 0.922, and a PI value of 0.51. This model has 20 neurons units in the hidden layer. For $h = 1$ time horizon forecasting the NSE reaches a value of 0.840 and a PI value of 0.30. This BNN model has 4 neurons units within the hidden layer. Finally for $h = 2$ days ahead, the NSE value decreases to 0.795 and a the PI value has a slightly lower value than the previous one (0.27), containing 13 neurons units within the hidden layer.

The "best" KCF for the GP model has a linear (Eq.A.9) and a neural network (Eq.A.10) product kernel: $k_P = k_{lin} \cdot k_{NN}$, with acceptable values of NSE, PI and loss functions for the training and validation period. The GP-k_P model has a good NSE value of 0.925 and a PI value of 0.53 for $h = 0$ time horizon, obtaining the best result for $h = 0$ time horizon compared to the other models for this specific case study. The behaviour of this model for $h = 1$ time horizon is characterised by a NSE value of 0.833 and a PI value of 0.35. Finally, its performance for $h = 2$ decreases to a NSE value of 0.797 and a PI value of 0.28. The behaviour of the GP model is, in this case, the best among all the DD methods compared.

TABLE 4.12: GF2-PS3B results for training (n=3893) and validation (n=973) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($m^3 s^{-1}$). Best performance for each horizon and metric during validation is highlighted in bold font

$\mathcal{M}$	h	NSE	RMSE	PI	MAE
MLR	$0 \rightarrow (t)$	0.636;0.771	0.860;0.463	0.55;-0.44	0.343;0.326
	$1 \rightarrow (t+1)$	0.418;0.786	1.088;0.448	0.33;0.07	0.337;0.297
	$2 \rightarrow (t+2)$	0.386;0.742	1.118;0.491	0.32;0.08	0.367;0.320
BNN	$0 \rightarrow (t)$	0.938;0.922	0.370;0.270	0.92;0.51	0.203;0.190
	$1 \rightarrow (t+1)$	0.506;**0.840**	1.045;0.388	0.44;0.30	0.280;0.250
	$2 \rightarrow (t+2)$	0.482;0.795	1.070;0.439	0.44;0.27	0.302;0.279
GP-k_P	$0 \rightarrow (t)$	0.957;**0.925**	0.295;**0.266**	0.95;**0.53**	0.174;**0.188**
	$1 \rightarrow (t+1)$	0.846;0.826	0.205;**0.208**	0.43;**0.35**	0.130;**0.138**
	$2 \rightarrow (t+2)$	0.815;**0.797**	0.226;**0.224**	0.37;**0.28**	0.143;**0.151**

Discussion Here, the results have differed from the previous case during the IVS procedure, while the physical and data characteristics are practically identical. The time series generated by the physical model, such as R_S or S, give excellent results for the *lin* test, what establishes them as one of the best predictors linearly. The performance of the R_e and R_{Wi} variables as candidate predictors has been poorer for the *lin* test. The *Nlin* test has different behaviour. Potential predictors like S and R_S have the poorer performance for the *Nlin* test, but R_e and R_{Wi} have very good performance. It appears that the behaviours are grouped by pairs. This suggests a more linear and direct behaviour for pure snow variables, i.e., S and R_S, compared to variables containing a larger liquid fraction of water, i.e., R_{Wi} and R_e.From these results, we can affirm that all the processes associated more directly with the snow, i.e., R_S and S. From the results, it is clear that this basin has a more linear, or direct relationship, within the rain-runoff process than in the GF1.

Again, we use the WiMMed hydrological model to help to describe the water system and to improve the forecasting performance, on both efficiency and predictability terms. In these sub-basins, where snow is an important agent in the hydrological cycle, it is important to analyse the approximation of the underlying function of precipitation-runoff with the Snowmelt pulses. In Sec.B.2.3, we test the improvement in the models by including simulated dataseries inputs in the final PS for this specific case.

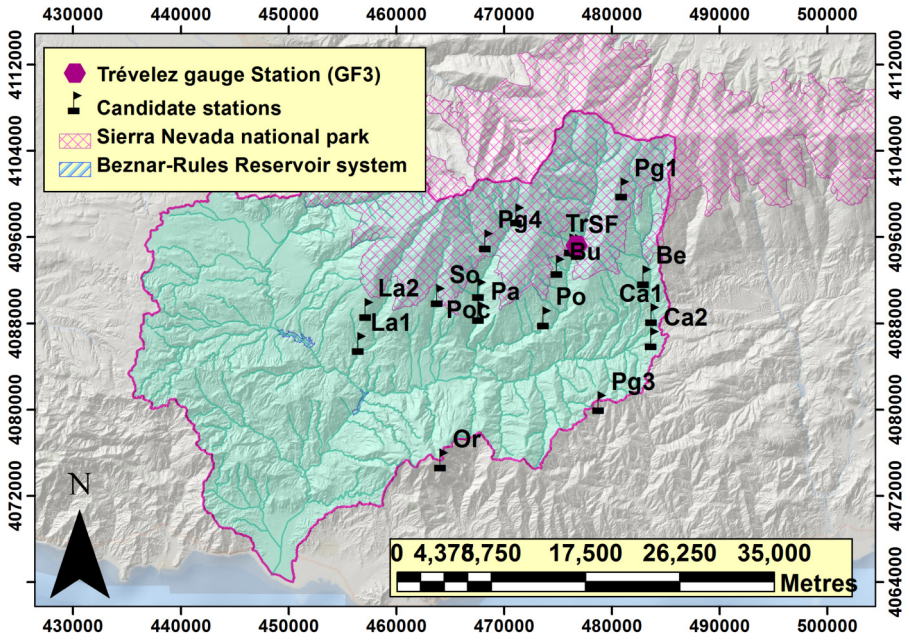


FIGURE 4.15: Northern eastern part of the Guadalfeo River Basin showing location of different stations on daily basis at the GF3 case study

In this case, the best DD computational technique is GP, with a product kernel covariance function: GP$\sim k_P(k_{lin} \cdot k_{NN})$, providing better performance than BNN and MLR for the deterministic estimations in all the time horizons and all the metrics tested. Again, the dot kernel product between a linear kernel (k_{lin}) and a nonlinear kernel (k_{NN}) provides the most balanced results among all the KCFs. In terms of the relative improvement (IC) of NSE compared to persist method, we get values of 11.0%, 13.9% and 13.4% for the three horizons respectively during the validation period. In terms of RMSE, the relative improvement for each horizon is 22.5%, 20.3%, and 17.5% during the validation period.

4.4.3 GF3: Streamflow forecasting for water allocation purposes

Case Definition As previous cases, the performance of three DD computational techniques (MLR, BNN and GP) are used to forecast the streamflow at Trevélez gauge station for different time horizons ($h = 0, 1, 2$) for the best assumption (A or B), which is selected during the IVS procedure (Sec.4.2.1). This experience is relevant as in this sub-basin there is a significant water extraction for the increasing urban demands on the town of Motril and its water distribution network nearby. Between 1.3 and 1.5 Hm^3 of resource volume are extracted annually for this purpose, and the streamflow thresholds for this extraction are 0.027 and 0.12 measured in $m^3 \cdot s^{-1}$.

The Trevélez (Tr) contributing area extends for about 77 km^2 across the southern part of the Sierra Nevada range, covering the 5.9% percentage of the GF total area. The location of the gauge station can be seen in Fig.4.15. This gauge station is at the highest altitude among all case studies (1450 m) in this work, and there the snow processes are significant during the nival season in the area under study. In Tab.B.8 the statistical parameters of the time series are shown. This example has been carried out without using hydrological modeling for snow processes.

Input Variable Selection results According to the method exposed in Sec.4.2.1, the results show that the best rainfall data using a single location were Ca1, Tr, Bu, Be and PoC for the *lin* test. The average of the first best two stations ($R_{\bar{a}_2(t)}$), outperformed the single data source. In the *Nlin* test, we collect that the best single rainfall datasets are at Ca1, Bu, Pa, Tr and Be, listed in order of importance. In this case, the average of the best three stations ($R_{\bar{a}3(t)}$) does not outperform the data collected at a single one, Ca1. The linear and nonlinear test match in the best single rainfall station but not with the averages data sets and the best rainfall predictor. Here, the differences are minimal. We represent these results in Fig.B.11.

$$\begin{aligned} \mathsf{SF}_{\mathtt{Tr}(t+h)} = f_{\mathcal{M}}(\{\mathsf{SF}_{\mathtt{Tr}(t-1)},\, \mathsf{SF}_{\mathtt{Tr}(t-2)}\}, \ldots \\ \{R_{\bar{a}w(t)},\, R_{\bar{a}w(t-1)},\, R_{\bar{a}w(t-2)}\}) \end{aligned} \tag{4.13}$$

TABLE 4.13: Relevant PS evaluated with *Nlin* (ANN) and *lin* (MLR) test for the GF3 case. *single*: From one station. *multiple*: the average of w stations

Basin-PS$_{[p]}$	**Symbolic Model: SF**$_t = f_{\mathcal{M}}$**(PS)**	**NSE**	**PI**
	Real data (single)	(*Nlin*;*lin*)	(*Nlin*;*lin*)
GF3-PS1A	$\mathsf{SF}_{\mathtt{Tr}(t)} = f(\{\mathsf{SF}_{\mathtt{Tr}(t-p)}\}_{p=1}^{2}, \{R_{a1(t-p)}\}_{p=1}^{3})$	0.967;0.966	0.06;0.04
GF3-PS1B	$\mathsf{SF}_{\mathtt{Tr}(t)} = f(\{\mathsf{SF}_{\mathtt{Tr}(t-p)}\}_{p=1}^{2}, \{R_{a1(t)}, R_{a1(t-p)}\}_{p=1}^{2})$	0.968;0.967	0.10;0.07
	Real data (multiple)		
GF3-PS2A	$\mathsf{SF}_{\mathtt{Tr}(t)} = f(\{\mathsf{SF}_{\mathtt{Tr}(t-p)}\}_{p=1}^{2}, \{R_{aw(t-p)}\}_{p=1}^{3})$	0.967;0.966	0.08;0.06
GF3-PS2B	$\mathsf{SF}_{\mathtt{Tr}(t)} = f(\{\mathsf{SF}_{\mathtt{Tr}(t-p)}\}_{p=1}^{2}, \{R_{aw(t)}, R_{aw(t-p)}\}_{p=1}^{2})$	**0.969**;0.966	**0.11**;0.07

From Eq.4.13, it can be seen that the best PS for the input model is a function of the volumes registered the two previous days, the average rainfall for the best first two ($w_{lin} = 2$) and three stations ($w_{Nlin} = 2$) for *lin* and *Nlin* tests, respectively. The

collected values of the current day t, and the two previous days $t-1$, $t-2$ are included as inputs, which is under the assumption B.

Modelling results In Tab.4.14 the results of the MLR, the "best" BNN model, and the GP models are shown. In the MLR case, the performance is just slightly better than the persist model. We obtained during the validation period values of 0.07, 0.01 and 0.01, for the three time horizons respectively ($h = 0, 1, 2$). In this case, the non-linear capabilities of BNN work somewhat better than the linear model. For the time horizon $h = 0$, the NSE value is 0.969 and 0.12 for its corresponding PI value. For the next time horizon modelled ($h = 1$) the NSE decreases to 0.936 and the PI increases to 0.14. For the last time horizon ($h = 2$), the PI value increases again,obtaining a value of 0.15 with a NSE value of 0.896. In any case they are all poor performances. Finally, the GP model performance in this case study was very poor, working similar to the persist method for $h = 0$ time horizon, and worse than the persist method for $h = 1, 2$ time horizons. Their results are not commented for obvious reasons.

TABLE 4.14: GF3-PS3B results for training (n_T=1895) and the mean value of two independent validation periods (n_{V_1}=1338, n_{V_2}=1801) respectively, separated by a semicolon (T;$(V_1 + V_2)/2$). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($m^3 \cdot day^{-1}$). Best performance for each horizon and metric during validation is highlighted in bold font

$\mathcal{M}$	h	NSE	RMSE	PI	MAE
MLR	$0 \rightarrow (t)$	0.952;0.967	0.229;0.418	0.07;0.07	0.092;0.171
	$1 \rightarrow (t+1)$	0.415;0.772	1.091;0.462	0.33;0.01	0.336;0.305
	$2 \rightarrow (t+2)$	0.383;0.724	1.120;0.509	0.32;0.01	0.366;0.331
BNN	$0 \rightarrow (t)$	0.954;**0.969**	0.222;**0.408**	0.13;**0.12**	0.093;**0.184**
	$1 \rightarrow (t+1)$	0.906;**0.936**	0.319;**0.585**	0.18;**0.14**	0.141;**0.279**
	$2 \rightarrow (t+2)$	0.872;**0.896**	0.372;**0.747**	0.21;**0.15**	0.171;**0.368**
GP-k_{P1}	$0 \rightarrow (t)$	0.999;0.966	0.059;0.424	0.98;0.06	0.031;0.202
	$1 \rightarrow (t+1)$	0.995;0.917	0.121;0.666	0.97;-0.10	0.063;0.332
	$2 \rightarrow (t+2)$	0.985;0.826	0.206;0.963	0.93;-0.41	0.095;0.482

Discussion The results in this basin have been poor in general, as they have only slightly exceeded the persistence method for all the DB models tested. This may be by the weak data and hydrological framework in which we have worked. As shown in Tab.4.14 the best DD model is the BNN model, having a better performance for the three steps

ahead tested. In terms of NSE, the improvement of accuracy compared to persist method values is 0.5%, 1.2% and 2.3% for the three horizons. On the other hand, in terms of RMSE, the improvement is 6.8%, 8.1%, and 8.4%. As some previous experiences, during the GP application we have encountered over-fitting problems.

The poor results for this case study may be due to several causes. The first is the absence of physical modelling for the snow agent simulation, and the small size of this particular sub-basin. Another possible cause is bad data as the control point (gauging) is not completely reliable, and therefore that the time-series can not be representative. High flow rates are not quantified due to lack of capacity of the cross section in this particular gauge control point.

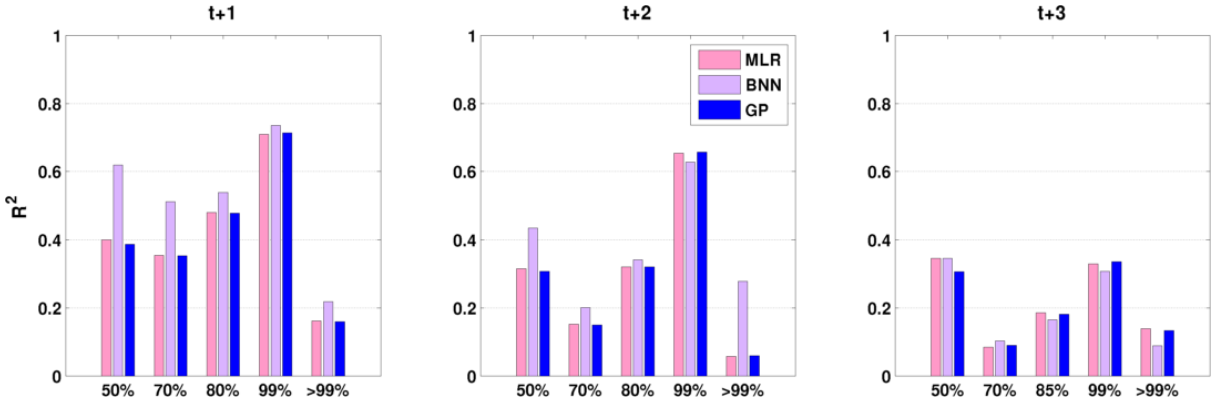
4.5 Extreme values analysis for GH1, GH2 and GF1 study cases:

One of the important points in our work is to model extreme values in the most accurate way. In this manner, we can develop tools for flood alert and the management of water systems. To assess the quality of the representativeness in the high values of the hydrograph, we pay attention to two of the daily basis (GH1 and GF1) cases, and one at hourly basis (GH2). In this section, we present the values classified and sorted, by the sample percentiles, and correlated linearly (R^2) for five classes: $\leq 50\%, \leq 70\%, \leq 80\%, \leq 99\%$ and $> 99\%$.

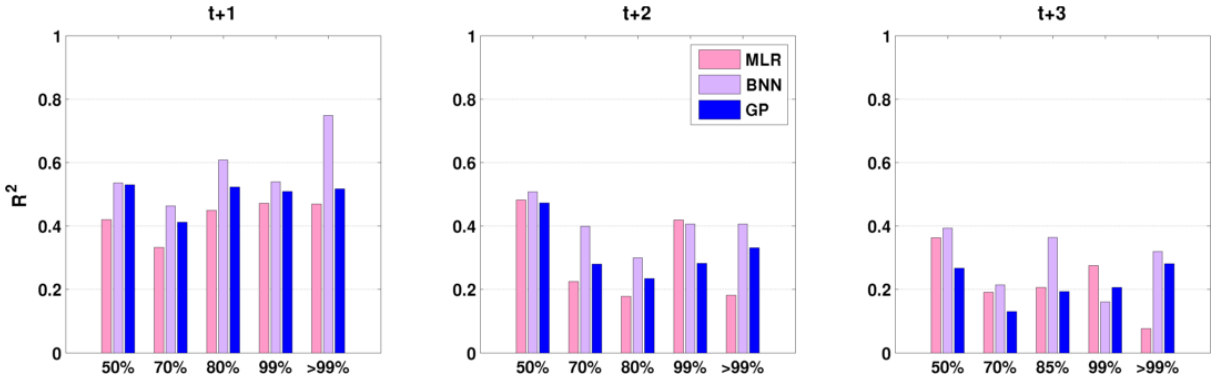
For the GH1-PS3B model case(Fig.4.16.(a)), we can observe that the non-linear model has higher performance than the other techniques for the $h = 0$ time horizon. This superiority over linear models is gradually diluted as we increase the time horizon. If we classify the target values and the simulated values into a discrete form, where $>$ 99th and $\leq$ 99th percentile values are labelled, the BNN method achieves the 67.74% and 48.39 % of a correct representation for $h = 0$ and $h = 1$ respectively. The total of the events that are higher than the 99% percentile are 24. For the GF1-PS4A model (Fig.4.16.(b)), the BNN model exhibits a higher performance than the other techniques.

For the case at hourly basis (GH2), the same analysis is done for the five time horizons studied $(t, t+1, t+2, t+3, t+4)$. This is an interesting case compared to the daily cases as the number of events are sufficient to extract performance conclusions for floods warning tools purposes. Here, the event definition is not established by its 99th percentile

threshold, but by the thresholds of the alert levels (I,II and III, Tab.4.15), directly defined by the water authorities. We change the Alert level III threshold to 4.2 metres of water level height (m), since there are no records for $\mathsf{WH} > 4.5m$.



(a) GH1-PS3B results for $h = 0$, $h = 1$ and $h = 2$ days ahead forecast during the test period (n=1098 *days*)



(b) GF1-PS4A results for $h = 0$, $h = 1$ and $h = 2$ days ahead forecast during the test period (n=725 *days*)

FIGURE 4.16: Results grouped by percentiles of the sample distribution for the validation period

TABLE 4.15: Forecasting overall accuracy (%) at hourly basis for the three Alert levels established by the water authorities evaluated in a binary form. Level I [$2.5m \leq \mathsf{WH}_{(t)}$]. Level II [$3.5m \leq \mathsf{WH}_{(t)}$]. Level III [$\mathsf{WH}_{(t)} > 4.2m$]

Level⇒	I (n_{T}=126)		II (n_{T}=40)		III (n_{T}=3)	
h	MLR	BNN	MLR	BNN	MLR	BNN
$0 \rightarrow t$	90.1	85.2	90.5	90.5	75	33.3
$1 \rightarrow t+1$	73.4	72.7	66.7	54.8	37.5	16.7
$2 \rightarrow t+2$	61.5	75.6	58.8	63.6	50	0
$3 \rightarrow t+3$	55.6	54.7	50.9	23.8	22.2	0
$4 \rightarrow t+4$	51.2	50.8	39.2	0	25	0

In Tab.4.15 we show the results for the '' best" bootstrapped MLR and BNN models under a classification framework at hourly basis. Also, the occurrence observed during the validation period (n_{T}) is listed for each alert level. We can appreciate that the nearest time horizon ($h = 0 \rightarrow t$) has a poorer forecasting behaviour for the BNN model, because of the strong pattern persistence for this assessment. This analysis found evidence to establish a temporal horizon boundary among both fundamental approaches. We can also observe how the MLR model has a better behaviour than BNN, especially for level III where the BNN model does not reproduce it from $t+2$ time horizon. Also, we can see how both models yield identical classification percentages for level II estimations at time horizon t. If we compute the linear correlation values (R^2), for those time steps where the events level II happened, we get 0.72 and 0.52 for MLR and BNN respectively. This gives better results to the MLR approach than the BNN for $\mathsf{WH} > 3.5m$ values. It reflects this in the behaviours for the next Alert level (III), where the MLR model is obviously much superior to BNN in a classified analysis.

To visualise easily the performance of these results, the election of confusion matrices may be a suitable solution (more in Sec.3.4.4). In Fig.4.17, the classifier results for both models tested for time horizon $h = 0$ forecast are visualised within the confusion matrix scheme. The blue cell in the bottom right shows the overall accuracy (in green, same values as listed in Tab.4.15) and misclassification rate (in red) results. The grey column on the right side shows the percentages of all the examples predicted to belong to each alert level thresholds that are correctly and incorrectly classified. For instances, for the Alert level I case, the MLR model forecasted 123 events ($\mathsf{WH}_{(t)} > 2.5m$), with a 95.9% of

true positives and 4.1% of false positives (overprediction). That means that we forecasted alert level I, but the water level doesn't actually rise this level. The bottom grey row shows the percentages of all the examples of each class that are correctly and incorrectly classified. For the same case, MLR($h = 0$) and Alert level I, we can see how the model gets 93.7% of true positives, and 6.3 % of false negatives (underprediction). That means that we forecasted smaller values than the threshold alert level, and where water level actually does reach the Alert level predicted.

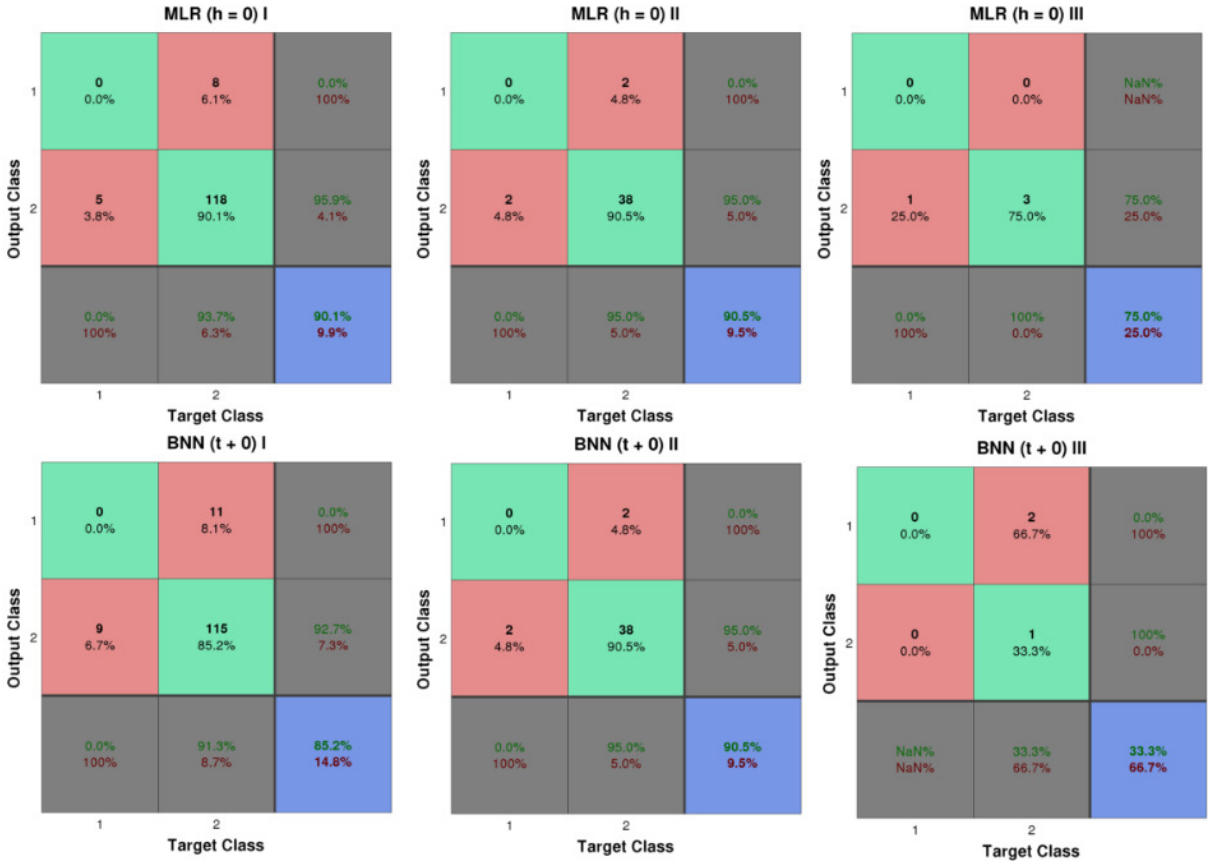


FIGURE 4.17: Visualisation of results through confusion matrices for h=0 time horizon results for the validation period at the three Alert levels established by the water authorities for the Cártama gauge station location

The comes from focusing on the extreme values displayed in this section, indicates how within the Guadalhorce River Basin at daily basis, GH1 study case (Sec.4.3.1), the *Nlin* approaches outperforms the *Lin* approaches for the two first time horizons ($h = 0, 1$) for the >99th percentile water level class. On the other hand, the same study case but at hourly basis, GH2 study case (Sec.4.3.2), the *Lin* approach (MLR) overcomes consistently the *Nlin* approach (BNN) for the three Alert Levels (Tab.4.15) analysed. On the alternative River Basin under study, the GF1 study case within the Guadalfeo River Basin, the preliminary results show us that the *Nlin* approach exceeds the *Lin* approach.

For this last instance, it is difficult to draw coherent inferences from the few extreme data recorded during the period under study, and therefore unrepresentative.

4.6 Uncertainty estimation analysis

We analyse the capacity of deriving confidence intervals at different hypothetical coverage levels in this paragraph. We compare the "best" f_{MLR}, the "best" f_{BNN}, and the "best" f_{GP} model, which is driven by a kernel (KCF). We realise this for each study case of Sec.4.3 and Sec.4.4, and at two time horizons: $h = 0, 1$. For simplification, in this analysis we have omitted the third time horizon if we compare it to the deterministic analysis.

For that, we represent the scatter plots of expected, and the estimated coverage of the selected models at different confidence intervals $0\%, \ldots, 100\%$. As a result, we should pay special attention on the behaviour at the upper tail (near 95% level), which is also represented by a window zoom. As a general guideline for this representation: the scatterplot suggests overestimation of the uncertainty quantification task when the scatterplot is above the 1:1 line, and it is sub-estimated when is below the 1:1 line.

In this way, we test the capacity of each model in the uncertainty's quantification under real data conditions. The capacity to estimate different theoretical coverage levels (i.e., confidence intervals) gives us a fair perspective of the method's noise model behaviour under our conditions. Its coverage performance should be near the theoretical, where the value usually used for decision guide and analysis in hydrological modelling is the 95% confidence intervals level (Loucks et al., 2005), and chosen in this work. So, the further the coverage performance result moves away from the theoretical coverage value, the worse the model's capabilities to quantify uncertainty. Then, we estimated the 95% confidence interval of the deterministic output ($y_{(t)}$) in different ways attending each methodological framework:

$f_{\mathbf{MLR}}$ The noise model $\varepsilon^{\circ}_{(t)}$ for this technique has a $N \sim (0, \sigma)$ distribution. The variance (σ) value changes depending on predictors values. Then, the residuals are standardised as *studentized* (a.k.a., deleted t-residual) residuals. Finally, the intervals got has $100 \cdot (1 - \alpha)\%$ confidence intervals of these t-distributions, which are centred at the residuals. In the Matlab routine used in this work, the value of is defined as 0.05, obtaining directly the 95% confidence intervals.

$f_{\mathbf{BNN}}$ Following Nabney (2004), the variance model of the Bayesian framework ($\varepsilon^*_{(t)}$, more on page 217) has contributions both from the output noise model ($1/\beta_h$), and from the posterior weights distribution $p(\theta|\alpha_h, \beta_h, \mathcal{D})$ from the MLP network. Finally, the confidence interval of interest is estimated by adding and subtracting 2σ for the estimated mean, or deterministic value.

$f_{\mathbf{GP}}$ Following Vanhatalo et al. (2013), the variance model $\epsilon_{(t)}$ has a $N(0, \sigma^2)$ distribution. Ultimately, the confidence interval of interest is determined by adding and subtracting 2σ for the expected mean.

In each model and regression experience, we quantify the uncertainty differently (see more in Appendix A) for a 95% confidence interval estimation. The following Tables (Tab.4.16 and Tab.4.17), show the evaluation coefficients used for the uncertainty performance analysis: Percentage of coverage (POC, Eq.3.14) and Average width (AW, Eq.3.15). First, the results of the six experiments on a daily time-scale are shown, followed by the experience at hourly basis for ten time horizons.

Daily hydrological simulations

TABLE 4.16: Uncertainty evaluation for the GH's experiences at daily scale. Training and Validation values are separated by a semicolon (T;V).

Model \ Horizon	$h = 0$	$h = 1$	$h = 0$	$h = 1$
GH1-PS3B	**POC (%)**		**AW** $[m \cdot day^{-1}]$	
MLR	**91.56;92.35**	**90.53;91.53**	0.28;0.28	0.33;0.34
BNN	94.01;90.44	91.86;87.34	0.23;0.23	0.28;0.28
GP-k_{lin}	26.76;24.04	19.20;16.58	**0.03;0.03**	**0.03;0.03**
GH3-PS2B	**POC (%)**		**AW** $[Hm^3 \cdot day^{-1}]$	
MLR	96.62;95.34	96.08;93.17	0.15;0.16	0.20;0.21
BNN	**97.94;95.03**	**97.13;94.72**	0.15;0.16	0.17;0.18
GP-k_{P2}	98.29;90.99	97.94;88.51	**0.02;0.02**	**0.03;0.03**
GH4-PS2B	**POC (%)**		**AW** $[Hm^3 \cdot day^{-1}]$	
MLR	95.46;93.32	93.86;90.99	0.28;0.31	**0.34;0.36**
BNN	**96.54;93.32**	**96.12;93.48**	0.34;0.36	0.44;0.47
GP-k_P	99.88;98.14	99.88;98.29	**0.25;0.25**	0.43;0.43

GH1-PSB3 Here, the MLR model has the closer POC values (≈95%) to the theoretical for all the periods modelled (T and V), and the two horizons tested ($h = 0$, 1). Under

this assumption, MLR has the most balanced and homogeneous values compared to the others. It was found that the BNN model produced smaller AW values across all the horizons tested, with 0.23 and 0.28 (*metres* of water height) respectively. But their POC values are lower than MLR during the validation period. We appreciate how the BNN model estimation moves away from the theoretical value during the validation period. As is shown in Tab.4.16, the GP-k_{lin} model has a bleak behaviour regarding uncertainty ($\bar{\text{POC}} \approx 20\%$). For this action, the MLR's variance performs better than the BNN and GP$\sim k_{lin}$ models.

In this experience the kernel finally chosen (GP$\sim k_{lin}$), was based only on its deterministic behaviour (Tab.A.5). But, if we take into account the uncertainty quantification capabilities, other kernels yield better results. We can see this in Fig.4.18 (b), it represents the uncertainty estimations of all the kernels compared for this case study for $h = 0$. We can note how the k_{lin} estimations are very poor. The k_{P1} estimations and all the other kernels, show much better behaviours by obtaining values closer to the theoretical. Then, for the GP$\sim k_{P1}$ model case, the POC values got are 89.05;88.16% and 84.60;78.14% and the AW values are 0.08;0.08 and 0.12;0.12, for $h = 0, 1$ time horizons for T and V periods respectively. In this case, this GP$\sim k_{P1}$ model would have the lowest value of the AW coefficient. But not to exceed the behaviour of MLR with consideration to POC values.

GH3-PS3A For this problem, we get the best POC values ($\approx 95\%$) with acceptable values of AW with the BNN model. The calculated values, are 0.16 and 0.18 ($Hm^3 \cdot day^{-1}$) for t and t+1 forecast, respectively. As is shown in Tab.4.16, the GP$\sim k_{P2}$ model produces smaller AW values, but has the poorest POC values ($< 91\%$), for both time horizons and modelling periods (T and V). From Fig.4.18 (c), it can be seen how the BNN's uncertainty estimations overestimates during the training period, but generalises remarkably during the validation period. Individually, it can be viewed how the MLR model underestimates the uncertainty during the validation period, even being closer to the theoretical during the training period. As a general behaviour, the BNN overestimates the uncertainty, but in our theoretical values ($\sim 95\%$), works very well for the two time horizons tested.

GH4-PS2B The best model, based on its POC values, is obtained with the BNN model again. Acceptable AW values have been obtained with the GP-$_{P1}$ and MLR model for h=0,1, respectively. For this particular case, the uncertainty quantification capabilities for all the models tested is poor for practical reservoir development tools, since they all show very high AW values, ≤ 0.25 measured in $Hm^3 \cdot day^{-1}$. Alternatively, analysing the graphical representation of the behaviour of the models in Fig.4.18 (d), follow that the GP$\sim k_{P1}$ overestimates, but its AW value is the best for $h = 0$ step ahead forecast.

On this, we must comment that for instances, an overestimation of the Percentage of Coverage (POC) value with an acceptable Average Width coefficient value (AW, more on Sec.3.4.2), will be always better than an underestimation in real-world applications, since the failure is always worse.

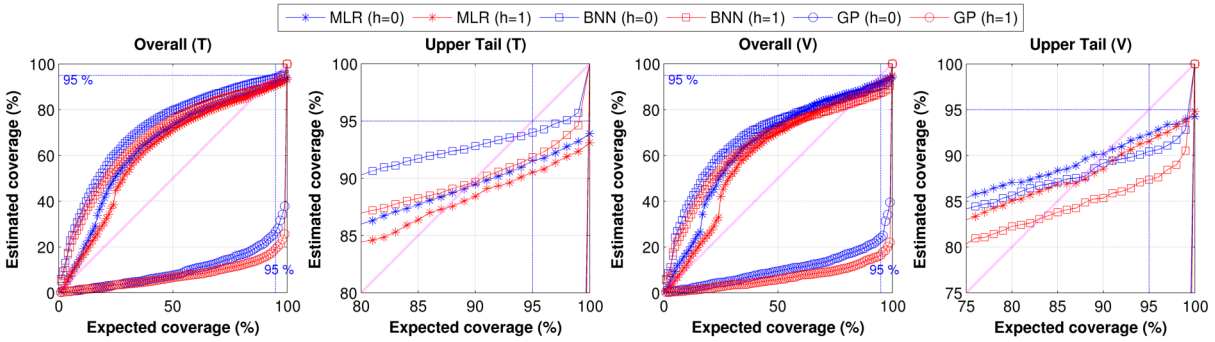
TABLE 4.17: Uncertainty evaluation for the GF's experiences at daily scale. Training and Validation values separated by a semicolon (T;V)

Model \ Horizon	$h = 0$	$h = 1$	$h = 0$	$h = 1$
GF1-PSA4	**POC (%)**		**AW**$[m^3 \cdot s^{-1}]$	
MLR	84.77;72.83	85.64;75.17	3.25;3.34	3.25;3.34
BNN	**98.86;97.93**	**97.90;96.28**	6.27;6.65	7.32;7.73
GP-k_P	96.79;88.83	98.00;87.72	**1.76;2.84**	**1.90;4.30**
GF2-PS3B	**POC (%)**		**AW**$[m^3 \cdot s^{-1}]$	
MLR	**94.21;96.78**	**97.01;97.24**	2.21;2.22	2.52;2.54
BNN	97.85;99.08	97.70;98.01	1.75;1.76	2.27;2.29
GP-k_P	90.10;87.27	89.83;85.89	**0.74;0.74**	**0.84;0.84**
GF3-PS2B	**POC (%)**		**AW**$[m^3 \cdot s^{-1}]$	
MLR	**95.62;94.17**	**95.04;92.83**	0.58;0.82	0.83;1.19
BNN	89.82;76.46	92.88;78.25	**0.33;0.33**	**0.68;0.71**
GP-k_{lin}	97.39;89.46	97.39;89.76	0.34;1.00	0.63;1.55

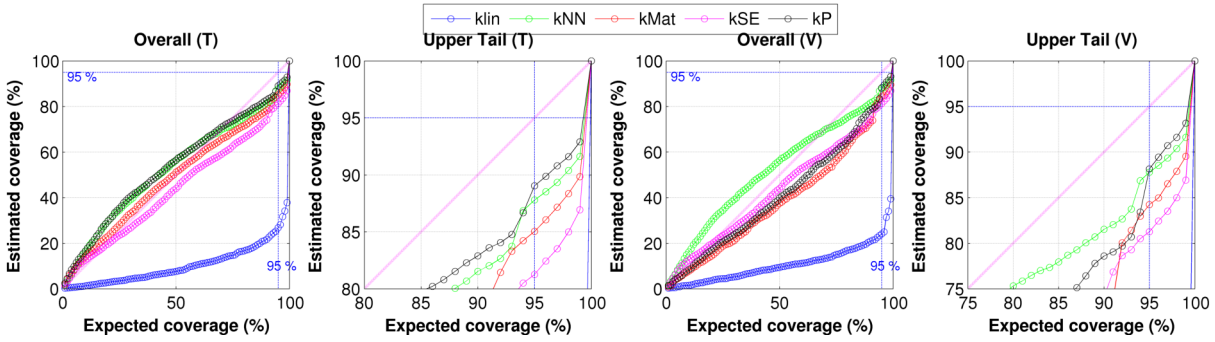
GF1-PSA4 Concerning to the uncertainty evaluation (Tab.4.17), the best POC values for all the horizons is achieved by the BNN model, even though it produces higher values of AW. In contrast, the more balanced uncertainty model is the GP model. We observed that the latter has an acceptable value of POC, and the lowest values of the AW coefficient. From Fig.4.19.(a), we can appreciate how the MLR model underestimates for all datasets and time horizons. This suggests a strong non-linear relationship on the uncertainty estimations that the linear model can't reproduce.

For this, we represented the variance models for five significant kernels in Fig.4.19.(b). The winning kernel in the deterministic assessments was k_{P1}, which has a modest behaviour in the quantification of uncertainty. It has a general tendency to overestimate during training and underestimate during validation. Anyhow, its behaviour is weaker than the k_{SE} GP model, which gives a more balanced noise model, and similar to the k_{Mat} GP model.

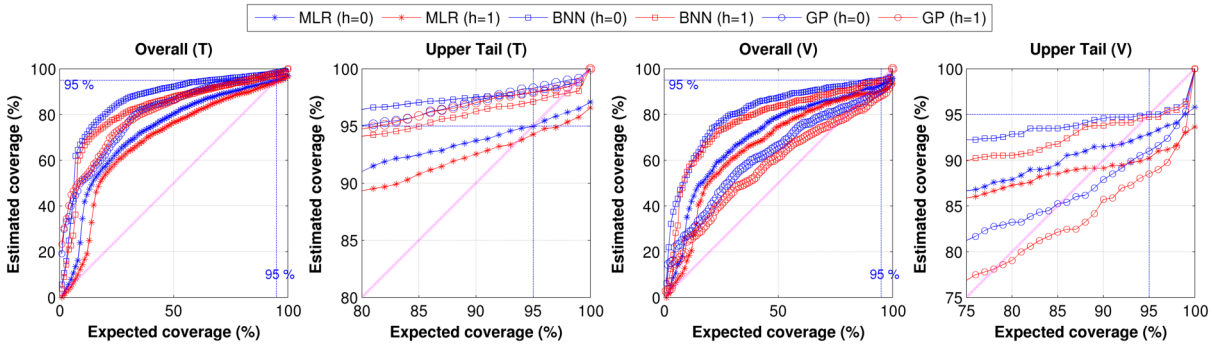
GF2-PS3B Here, the MLR has the closest POC values to the hypothetical, but on the contrary has the higher AW values for all the cases. As the earlier case, the good POC



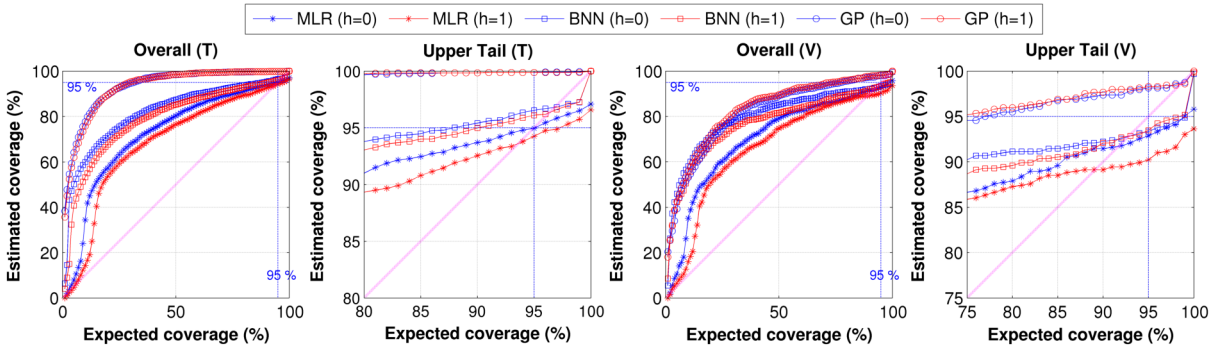
(a) Plot of estimated and expected coverage of different confidence intervals for the "best" MLR, BNN and GP models for both training and validation periods in the GH1 case study



(b) Plot of estimated and expected coverage of hypothetical confidence intervals for GP models with five different kernels: k_{lin}, k_{NN}, k_{Mat}, k_{SE} and k_{p1}. The estimations are for $h = 0$ time horizon during the training (T) and the validation (V) periods for the GH1-PS3B (Tab.4.1) model case

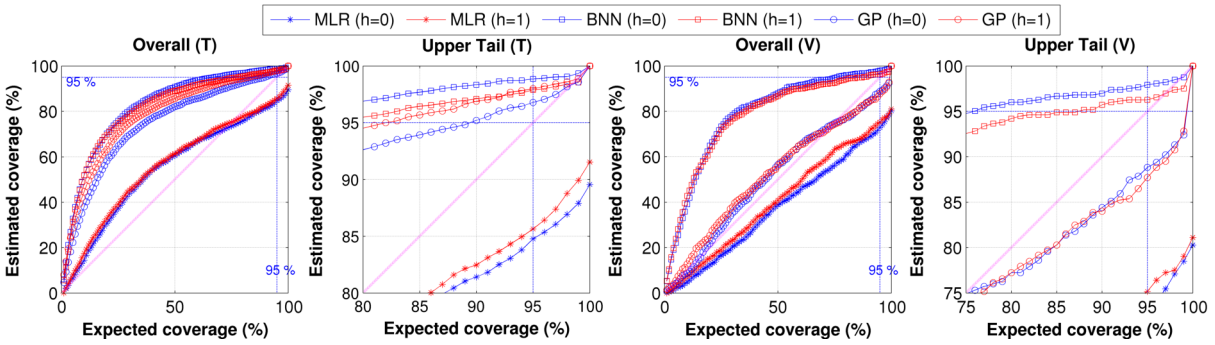


(c) Plot of estimated and expected coverage of different confidence intervals for the "best" MLR, BNN and GP models for both training (T) and validation (V) periods in GH3

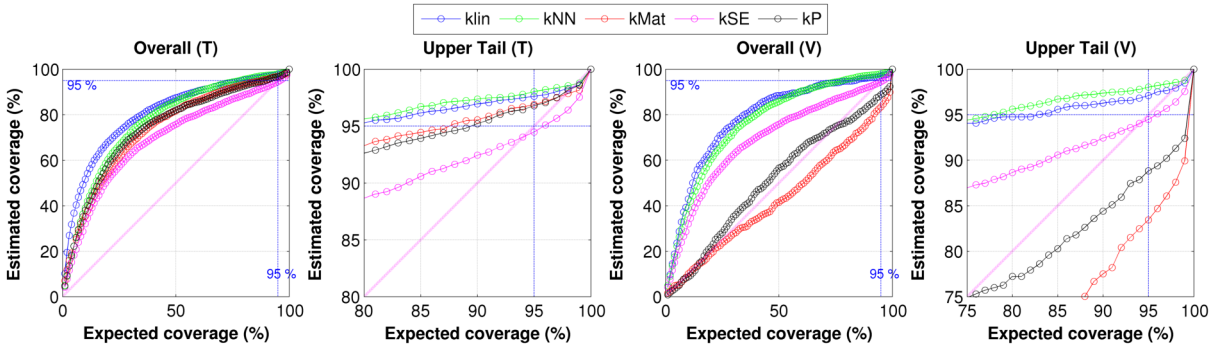


(d) Plot of estimated and expected coverage of different confidence intervals for the "best" MLR, BNN and GP models for both training (T) and validation (V) periods in GH4

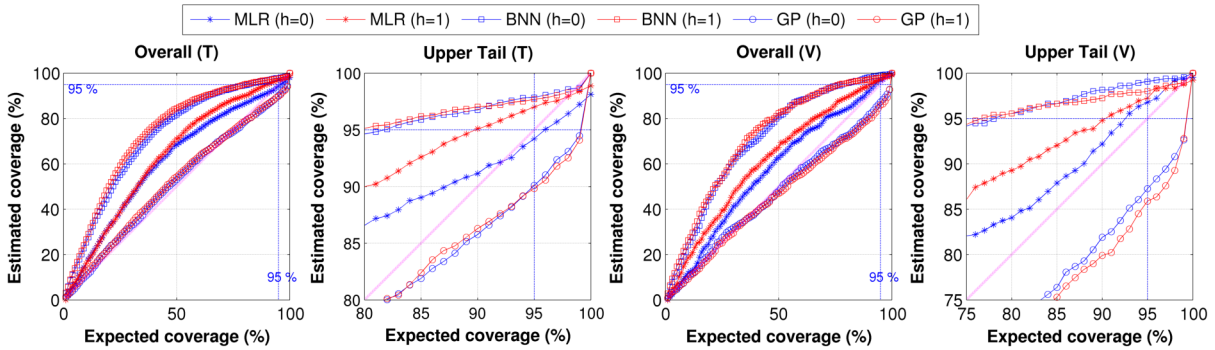
FIGURE 4.18: Plot of estimated and expected coverage of different confidence intervals of three Data Based models during both, training and validation periods, for the regression experiences carried out in the Guadalhorce River Basin



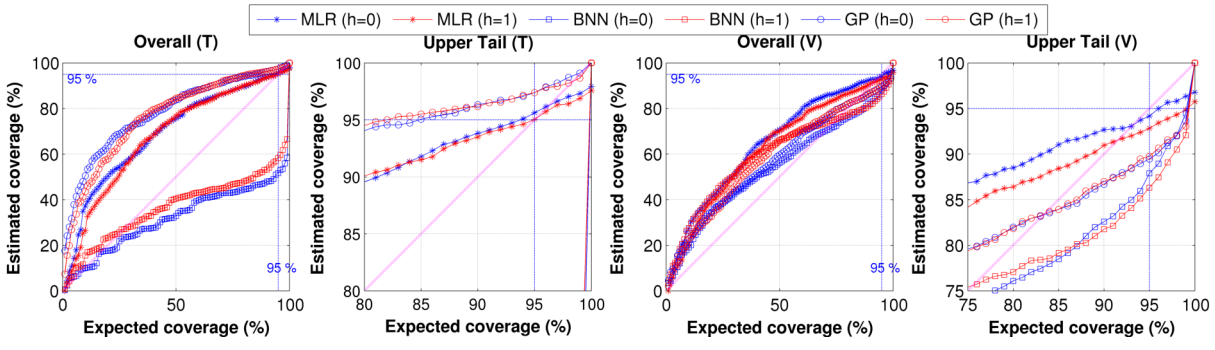
(a) Plot of estimated and expected coverage of different confidence intervals for the "best" MLR, BNN and GP models for both training (T) and validation (V) periods in GF1



(b) Plot of estimated and expected coverage of hypothetical confidence intervals for five different GP models- kernels: k_{lin}, k_{NN}, k_{Mat}, k_{SE} and k_{P1}. The estimations are for $h = 0$ time horizon during the training (T) and the validation (V) periods for the GF1-PS4A model (Tab.4.9) case



(c) Plot of estimated and expected coverage of different confidence intervals for the "best" MLR, BNN and GP models for both training (T) and validation (V) periods in GF2



(d) Plot of estimated and expected coverage of different confidence intervals for the "best" MLR, BNN and GP models for both training (T) and validation (V) periods in GF3

FIGURE 4.19: Plot of estimated and expected coverage of different confidence intervals of three Data Based models during both, training and validation periods, for the regression experiences carried out in the Guadalfeo River Basin.

values got with the BNN model, are masked by its high variance estimation (AW), see Tab.4.17. The GP$\sim k_{P1}$ model has the lowest AW values of 0.74 and 0.84, for the two horizons respectively (measured in $m^3 \cdot s^{-1}$). But it has poorer values than the other DD models. The GP model underestimates as shows Fig.4.19.(c), but this width intervals numerical values are useful for practical applications, since it really limits the estimations in real-world applications. For this case we see that the most balanced is the BNN, after evaluating its behaviour under all the assumptions.

GF3-PS2B Regarding the uncertainty predicted at the GF3 regression case, the MLR model has better POC values (≈95%) for all the periods (T and V), and the two horizons tested (Tab.4.17). On the other hand, it was found that the BNN model produced less AW values across all the horizon tested, with 0.33 and 0.71 (measured in $m^3 \cdot s^{-1}$) respectively, but their POC values are the poorest among all the techniques compared. For this water sub-system, the results indicates that the *lin* approach, on the uncertainty quantification task, overcomes the *Nlin* approaches as is illustrated in Fig.4.19.(d).

Hourly hydrological simulations

TABLE 4.18: Uncertainty evaluation for the GH2 experience. Training and Validation values separated by a semicolon (T;V). Uncertainty error metrics (POC, AW) are expressed in percentage (%) and in their original unit measures (m) respectively

Model	Coefficient	t	$t+1$	$t+2$	$t+3$	$t+4$
GH2-MLR	POC (%)	94.76;93.28	95.04;93.62	95.10;93.69	95.11;93.73	95.08;93.75
GH2-MLR	AW (m)	.105;.116	.160;.177	.201;.223	.232;.259	.257;.288
GH2-BNN	POC (%)	96.77;87.87	97.07;88.53	98.07;89.19	98.17;88.46	8.01;88.66
GH2-BNN	AW (m)	.075;.073	.110;.107	.150;.147	.160;.155	.182;.178

The last uncertainty analysis is the GH2 case study at hourly basis. For it, we analised the uncertainty generated at five time horizons: $t, t+1, \ldots, t+4$. In Tab.4.18, the uncertainty coefficients for the "best" MLR and BNN resulting models are listed. It can be seen for the MLR model, how it performs with the smaller difference between its Percentage of Coverage (POC) values and the 95% hypothesis, for both calibration and validation periods. BNN consistently obtained lower AW coefficient values. The BNN,

got POC values within the range $87 - 89\%$ for the validation period, showing that has poor generalisation ability at this time basis.

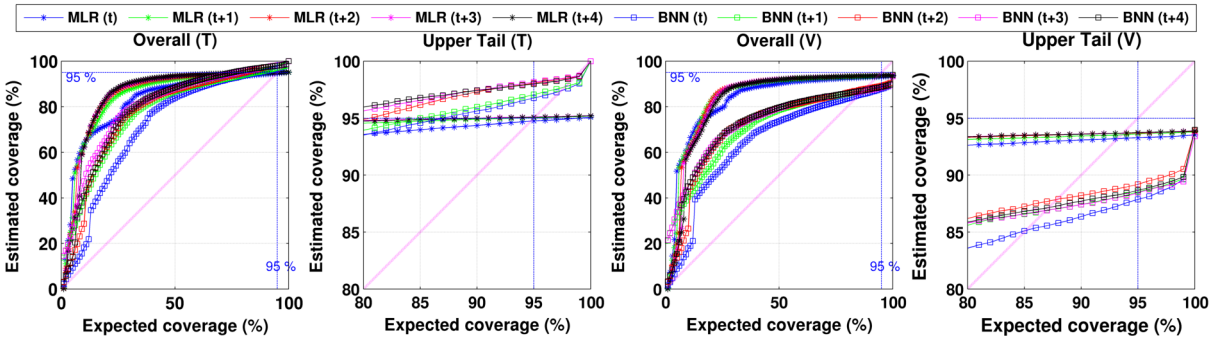


FIGURE 4.20: Plot of estimated and expected coverage of different confidence intervals for MLR and BNN for both calibration and validation periods in GH2

Here, the MLR models perform better than the BNN on coverage estimations. At hourly basis, the uncertainty quantification has been poorer than at daily basis, as it will always be better to overestimate than underestimate. Paying attention in the upper tail of Fig.4.20, we can see how the BNN technique overestimates during the training period, and underestimates during the validation period, with low coverage percentages (POC$\approx$ 88%). We have to take into account the amount of simulated data where an analysis centred on the extreme values is done in a further section, Sec.4.8.3.

The analysis of uncertainty carried out in this section helps us to test the capacity and the utility of these Data-Driven computational models for Decision tool development. Also, we draw a natural probabilistic approach to the modelling experience with the estimation of the confidence intervals, or plausible values. Therefore, we associated an interval to those deterministic model values, having calibrated and conditioned it to historical data. These estimates are useful in practical effects, as long as they have an acceptable interval width (AW) and percentage of success (POC) during the evaluative phases of the modelling. Here we must pay special attention to the most important events of our interest, for example, flood events.

Pappenberger and Beven (2006) discussed brilliantly the idiosyncrasy of the uncertainty estimation, and some guidelines for its use in the hydrological science. They highlight that a common Code of Practice is needed among the hydrological community. This will provide comparative frameworks, that take into account all the modelling aspects that are relevant into the uncertainty quantification task. Juston et al. (2013) state the reasons why it is convenient to estimate uncertainty, remembering that: "*uncertainty estimation*

should not be the end point of a project but should rather be part of a learning process about how to represent the hydrology in a particular place".

4.7 Input Variable Relevance Analysis at daily basis

In regression experiences, a posterior evaluation of the inputs relevance is constantly recommended, where the influence of each input x_d is figure out. This step is necessary in the design of the modelling experiences, to extract meaningful knowledge from them. So we can infer, from data, the physical interpretation of the hydro-meteorological processes modelled and the contribution, and in what degree, of the different inputs and/or agents investigated. Under this angle, the linear (lin) and non-linear ($Nlin$) tests, have coefficients to develop this contribution analysis, and which is done in this section.

The lin test follows the Eq.A.1, where the multiple relationship is driven by the coefficient β_i that multiplies the predictor or input variable d. This results in $\beta = (\beta_0, \beta_1, \ldots, \beta_d)$, which shows the importance of the variable in the relationship, and how it affects to the dependent variable or regressand. The final β coefficients are normalised to the numerical range [0-1] for comparison, leaving apart the independent term β_0. For the input relevance within the $Nlin$ test, the Connection Weight CW_N approach exposed in Olden and Jackson (2002) is used, which is also normalised for comparison. For more details on this step refer to Sec.4.2.1.

The visual representation of the input relevance results of the regressions experiences at daily basis are shown in the next figures and commented below. In these figures (Fig.4.21), the $\mathrm{CW_N}$ distributions of several simulations are lumped together for each input variable of the final PS modelled. In these, the boxes show the inter-quartile range, the whiskers the maximum and minim values. Also, the red horizontal line shows the median, and the grey circle is the normalised coefficients got with the MLR model (β_{MLR}).

As a general practice, we find that the target variable depends firmly on the antecedent values measured. In autoregressive modelling is reasonable, and even desirable within the dynamic methods that seek to represent water systems with a strong non-stationary nature. We must highlight that we are looking for models that should estimate a future state value, from the present or most recent time.

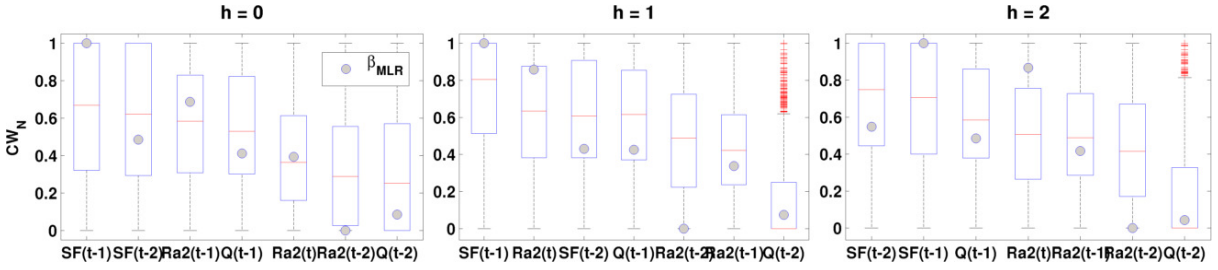
In the first of the Guadalhorce river cases, case GH1 (Fig.4.21.(a)), we can see that the predicted water level depends slightly on the earlier releases $\{Q_{(t-1)}\}$. We can appreciate

this matter in Table 4.1, comparing the PS2B and PS3B models, which was discussed in the previous section. In the GH3 case (Fig.4.21.(b)), we can see how the rainfall influence on the output model, being relevant in all time horizons tested. Also, should be highlighted the high relevance of the $R_{\bar{a}5(t-2)}$ input, showing a strong delay memory within the process, for h=0 time horizon forecast case. It normalises this strange behaviour in $h = 1, 2$ look-ahead time steps. In the GH4 case (Fig.4.21.(c)), the result suggests that the predicted volumes depend on its antecedent values with poor influence of two days previous for $h = 0$, gaining a strong influence as h increases. Here, the rainfall influence is similar for the three previous days being very homogeneous.

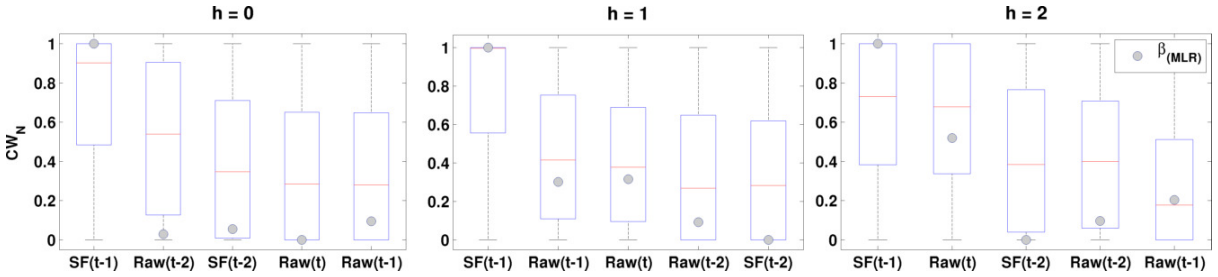
For the Guadalfeo River cases, it can be seen for the GF1 case (Fig.4.21.(d)) that the predicted streamflow level depends on the previous effective rainfall value ($R_{e(t-1)}$), and previous Snowmelt values $\{S_{(t-1)}\}$. These snow variables are reproduced by hydrological modelling. In the *x-axis* of the figure, the inputs are sorted in order of importance. In the GF2 case (Fig.4.21.(e)), we can see how the performance depends less than previous cases on the rainfall values $\{R_{\bar{a}2(t-p)}\}$. Accordingly to this figure, the autoregressive nature of the model becomes stronger. That means that its antecedent values have more relevance (the p order), as we increase the h value. The GF3 is not plotted and analysed since the regression has not been satisfactory, being discarded its discussion.

We observe sizeable differences between the input relevance magnitudes within the comparative paradigm of *lin vs. Nlin*. This is relevant in the cases GH3, GH4, and GF2 as seen in Fig.4.21, when the normalised values between zero and one, differ from one approach to another. In the GH3 case, we can see how that disparity between approaches in the variables $R_{aw(t-2)}$ and $SF_{(t-2)}$ for $h = 0$ time horizon forecast (Fig.4.21.(b)). In case GH4, this disparity appears in the different horizons $h = 0, 1, 2$ for rain, especially for the value registered in $R_{aw(t-2)}$ (Fig.4.21.(c)). Finally, for GF2 it highlights the difference of the antecedent values of the target variable, $SF_{(t-1)}$ and $SF_{(t-2)}$, for $h = 2$ time horizon forecast (Fig.4.21.(d)).

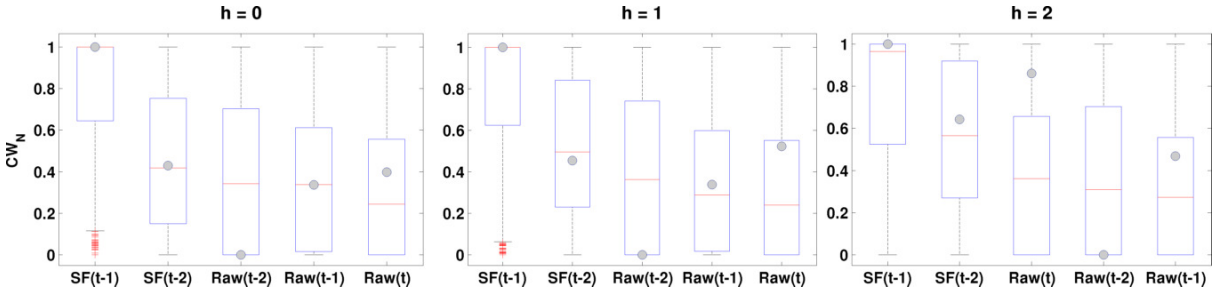
This analysis can identify which variables contribute, more or less, to the prediction of the target variable. Commonly, a standard pruning procedure could be implemented, where a minimum contribution can be established using a threshold. Then, we eliminate, "prune", the input variables that do not exceed it. Finally, we have to compare the new pruned model, with the previous model, to figure out if there is a performance improvement. This is an often methodological step, whereby in this way we drop unnecessary noise, generating a more balanced and parsimonious models.



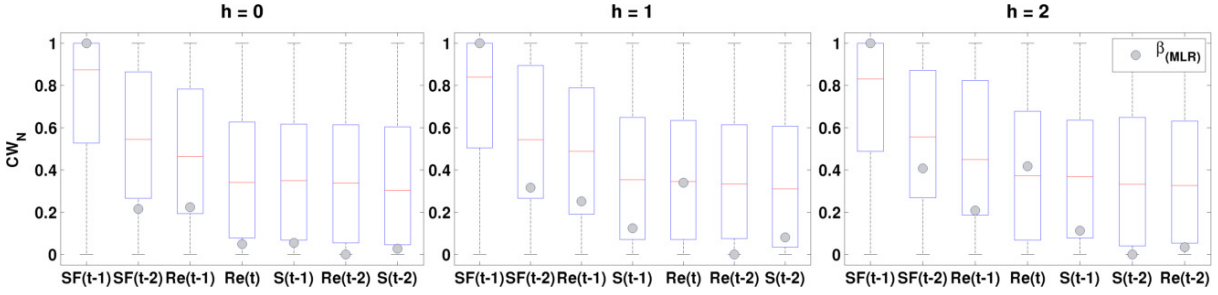
(a) Comparison of *lin* ($\beta_{(\mathrm{MLR})}$) and *Nlin* (CW_N) input relevance results for the GH1-PS3B model.



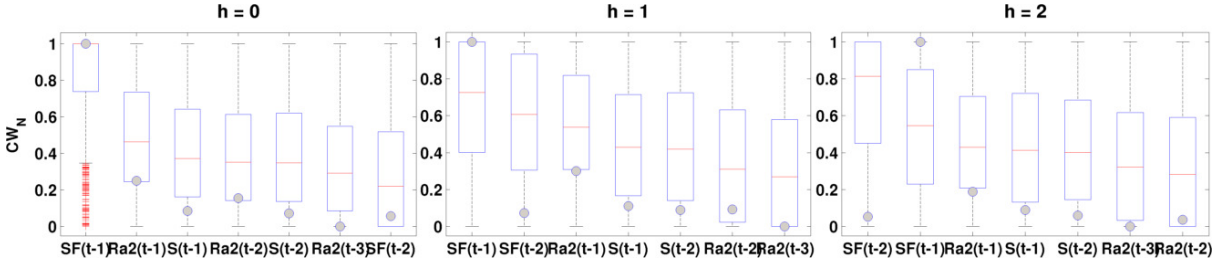
(b) Comparison of *lin* ($\beta_{(\mathrm{MLR})}$) and *Nlin* ($\mathrm{CW_N}$) input relevance results for the GH3-PS3B model



(c) Comparison of *lin* ($\beta_{(\mathrm{MLR})}$) and *Nlin* ($\mathrm{CW_N}$) input relevance results for the GH4-PS2B model



(d) Comparison of *lin* ($\beta_{(\mathrm{MLR})}$) and *Nlin* ($\mathrm{CW_N}$) input relevance results for the GF1-PS4A model



(e) Comparison of *lin* (β_{MLR}) and *Nlin* ($\mathrm{CW_N}$) input relevance results for the GF2-PS4B model

FIGURE 4.21: Input relevance for the regression experiences carried out at daily basis. Input variables are sorted by their non-linear relevance ($\mathrm{CW_N}$) median value

4.8 Chapter Discussion

4.8.1 Input Variable Selection

In this Chapter, we developed the Input Variable Selection (IVS) procedure with a wrapper methodology, under a stepwise view, establishing the best predictor for two different approaches. This is done by running in a parallel mode two different DD test/models: one linear (MLR), and one non-linear (BNN).

For river stages forecasting in a rainfall-runoff process based problem, the principal variable that must be explored is Rainfall. The results of the best rainfall predictors show that they are not defined by a clear physical meaningful, as the location of the station in the runoff generation area, or the station altitude, or the proximity of the collecting/-data points i.e., the distance. During the autoregressive model development, it has been observed that for $p_R > 3$ order values the autocorrelation model is not very sensitive on a daily basis, so initial testing orders of two and three days are good enough as the correlograms show that there are no relevant changes for larger orders. Also, is observed that the inclusion of several stations average values ($R_{\bar{a}w}$ for $w \geq 5$), may decrease the overall performance sometimes, not ensuring a greater model predictability of this fundamental agent within the water system under study.

We mapped the best rainfall predictors fitting to each modelling approach (*lin* and *Nlin*), to maximize their behaviour during training and validation. We have observed that these differences do not provide relevant information during the IVS procedure. We have three options in the building PS process: that the strongest connection is found in a single station, that it is the average value of the best stations, or that we introduce different locations or rainfall datasets as independent input variables. The average almost always exceeds a single value of precipitation, and the behaviour of the models with only one input to represent the rain agent, surpasses the models with several independent inputs. In addition, this last choice considerably increases the size of final PS. Anyway, at hourly basis (GH2) the best PS was composed of two stations that are introduced independently, performing better than the aggregated form.

We considered other hydro-meteorological causal variables as candidate exogenous inputs (X_d, Eq.4.1). In the GH and GF cases, we test the temperature (T) in four modes: average ($\bar{T}$), maximum (T_x), minimum (T_m), and the daily variance (T_Δ). In a first trial, the temperature does not improve as expected, and we may get only slight improvements

with T_{Δ}. This weak improvements don't overcome the minimum threshold, not providing relevant information to the underlying process. For a basin with such dependence in the temperature as the GF, because of its snow-related processes, the solution came with the inclusion of physically based modelling (WiMMed). The model filled the gaps left in the regression model representation considerably, improving the performance of the models performance. The improvement has been more pronounced in GF1 than in GF2, which may be because the GF1 has a higher percentage of the contributing area above 2000 meters.

For rising limbs, Rainfall-Runoff generation processes may be the predominant forcing agent, while baseflow and snowmelt (only for GF) may dominate during the falling limbs. The Guadalhorce basin runoff and peak response to the rainfall pulses become more linear than the Guadalfeo basin, and can be ascribed to that a fundamental source of the resource in the Guadalfeo come from the snow. This may add a strong non-linear component to the Rainfall-runoff process when it is present.

We investigated the basin response and the rainfall persistence effect of the collected data by two different assumptions: A and B (Sec.4.2.1). This timing study has been done only for $h = 0$ lead time forecasting in the experiences presented at daily basis. The test results for one lead time forecasting are provided in the corresponding IVS tables: Tab.4.1, Tab.4.5 and Tab.4.7 for the GH cases, and Tab.4.9, Tab.4.11 and Tab.4.13 for the GF cases. According to rainfall persistence and the performance of the models, we got the best results with the assumption B, for all the cases carried out in the Guadalhorce River Basin: GH1, GH3 and GH4. For the GF1 action, the assumption A has higher performance with including snow simulated data.

Also, we can appreciate a greater linear relationship in the rainfall-runoff process when $R_{(t)}$ is included in the model expression or final PS. This means that the differences between the deterministic behaviours of MLR and BNN, are minor during their application. In the GF case, A is the best assumption, even though assumption B provides similar efficiency and accuracy. This is due to a greater delay and non-linear effect of the rain component on rising river dynamics.

The results show that IVS procedure outcomes are reliable for our purposes, getting good results in hydrological modelling, especially for management of complex input data. This is very common in inputs as rainfall, when we have multiple candidate meteorological stations. By using autocorrelation, daily streamflow values of the last two days, $t - 1$ and $t - 2$, were included as the correlogram do not change for more antecedent values,

based on correlation between the rainfall and other candidate variables. We also checked the lag order by the trial and error procedure, not finding positive change with other combinations.

4.8.2 Data Driven Models development and its particularities

The forecasting experiences in each case have compared the three day-ahead forecast performance of three DD techniques: Multiple Linear Regression (MLR), Bayesian Neural Network (BNN) and Gaussian Processes (GP). The comparison analysis have been focus on deterministic (Sec.4.3 and Sec.4.4) and on uncertainty outcomes (Sec.4.6), with redundant evaluation for our hydrological objectives established in chapter 3.

MLR The first and simplest technique used is MLR. It was expected that its behaviour was the poorest among all the computational techniques used due to the non-linear character of the geophysical processes in general, and especially within the rainfall-runoff (RR) process (Singh, 1964; Chiu and Huang, 1970). This technique has served as a baseline within the forecasting comparative framework between different techniques of mathematical nature because of its simplicity. In similar works (Adamowski et al., 2012; Rasouli et al., 2012; Mekanik et al., 2013; Latt and Wittenberg, 2014; Choubin et al., 2016), we found its inclusion within computational techniques comparison exercises.

However, in our particular case, sometimes the relative difference with the non-linear models is greater than in others, being able to extract which underlying RR relationship behaves more linearly. In the uncertainty quantification task, the theoretical coverage of the 95% has been more stable in all cases approaching the expected. In contrast, the average width (AW) values have been higher in all cases compared to GP and BNN, limiting its use for practical purposes.

At daily basis, the MLR's behaviour has been the poorer among the three techniques tested, for both the deterministic and the uncertainty outcomes. But at hourly basis, the behaviour has been different, generating better outcomes for some time horizons than the BNN model. Under the bootstrap resampling framework, the "best" MLR (lin) model can outperform the "best" BNN ($Nlin$) model. This outcome validates the intuition that not always complex models ensure better results. The MLR method can deal the purposed problem with an acceptable quality. And finally, the interpretability of the model is simpler.

BNN The second DD computational technique used is BNN, which have great potential in the recognition of strongly non-linear relationships among multiple variables, and with higher capacities in terms of predictability than the other techniques used. This technique surpasses the standart ANN model as was previously demonstrated in Khan and Coulibaly (2006); Maiti and Tiwari (2014). This DD technique also has other advantages, such as Bayesian regularization, or as a framework for estimating error intervals (uncertainty quantification). In this work, this computational technique outperformed to the other two (MLR and GP) for the deterministic outcomes for almost all cases.

The main disadvantage of this method is the uncertainty during the parameters election/decisions during development phase, as the solution space domain problem of BNNs are usually complex and high dimensional. This must be addressed in any BNN application, looking for the "best" architecture and learning framework for our datasets and specific conditions. For that reason, always is convenient to try different configurations by trial and error and evaluate which one works best for the objectives planned. In this sense, the most important aspects found during this work, such as the definition of its architecture (hidden neurons units [`HL`], the transfer function [f_a]), the choice of the learning algorithm and its parameters such as the Learning Rate (LR) or the Momentum (μ), are discussed below:

• Regarding to the **optimal BNN architecture**, in between layers, as a general structure for all work, there is a single hidden layer as is recommended previously for Rainfall-Runoff modelling (Govindaraju, 2000). Concerned to its candidate transfer function that integrates the input layer signals (f_a, Fig.A.3), the hyperbolic tangent (`tanh`) were compared to the radial basis function (`rbf`) in our real data cases. Finally, the f_a=`tanh` is established as we get better results based on efficiency, parsimony and computation time. More on this comparison refer to page 211.

On the optimal number of parameters, i.e., number of units within the hidden layer (`HL`), were marked out with statistical significance by running multiple random cases, reducing the uncertainty in the model selection which is always a tricky question and time consuming step. The BNN architecture which reaches the maximum efficiency varies, but if we focus on the 75th percentile statistic the optimal number of neurons within the hidden layer are low in general (`HL`≈14, see Tab.A.1 in App.A) which may increase with greater predictive horizons. This is represented in the following figures: 4.13, B.12, B.13 B.14, B.15, B.16. In these figures, the several runs carried out during the *dynamic* approach, with an increasing number of neurons

within the BNN structure, are represented. In these figures can be appreciated the overall behaviour of different configurations comparing them.

These results are suitable in order to secure the ability of the model to generalise well on new data, which are in line with fundamental recommendations (Dawson and Wilby, 2001). However, we can see how in other works (Taormina and Chau, 2015), finally a high number of hidden neurons are used (100 units) as the most appropriate. Hosseini and Mahjouri (2016) get the best models with a range between 3-13 units within the hidden layer. Others authors as Singh and Borah (2013); Kasiviswanathan and Sudheer (2013) finally left a low number as the optimal, `HL` ≈ 3, after tested under a model approach a range from a minimum of 2-3 hidden units to 10. Almost the same as the range used in Anctil and Rat (2005) between 2-12 neurons units within the hidden layer. In this sense the evaluation of the model parsimony through the BIC coefficient is important, which penalizes the over-parametrized models. If we represent the BIC values obtained from simulations by increasing the number of neurons. For example, in Fig.B.10 different set of predictors in the GF1 and GF2 cases are represented with its correspondent BIC values depending on the number of neurons, or what is the same, depending on the number of parameters (Eq.3.13). We can clearly see how there is a first zone where it drops to a minimum with few units, and starts with a clear tendency to increases, that means less parsimony on models. More information on this refer to page 209.

• Concerned to the θ **parameters optimisation** procedure, a traditional Gradient Descent (GD) backpropagation algorithm as the Scaled Conjugate Gradient (SCG, Moller (1993)) has finally been established as the most convenient for our work (more on page 211). Another aspect during its development phase is that we have to deal with a non-straightforward calibration due to the existence of local minima during the parametrisation or training period. This fact generates different models and behaviours that has to be manipulated during the model development process. At first, this can be a disadvantage, since it complicates its management, having to work with multiples candidate members models, and therefore, more time consuming during the development phase. But on the other hand, it is an advantage since it generates a variability that offers us a more robust model solutions development and opportunities such as the ensembles (further developed in App.C).

Following Nabney (2004), multiple random simulations of his initial weights and biases ($\theta : \omega, b$), which are uniformly distributed over the zero-Gaussian interval $\theta_{init} \sim N(0, \sigma)$ are selected. Another θ_{init} rule depending on the number of inputs and outputs has been tested previously, obtaining a lower performance variability of

the resulting member models, and ultimately a poorer behaviour in efficiency and predictability terms. This comparison can be found in pag.211. In a first attempt, two different values of iterations (500 & 2000) to achieve the loss function goal during the training period were compared, establishing the value of 2000 iterations an appropriate number to reach very good quality of solutions on both methodological datasets, and exceeding the performance values obtained with 500 iterations. This comparison is represented in Fig.A.12 for the GF1 case at daily basis. Also in this sense there is not an over-fitting risk thanks to a weight decay regularisation term within the Bayesian framework used. Finally, the learning rate (LR) and momentum (Mom) terms within the backpropagation neural network algorithm are also investigated through a grid-searching analysis by running multiple cases (pag.214). Finally, the 0.2 numerical value is established as the most appropriated for both learning terms.

Regarding the uncertainty in the optimal member model election among the multiple models generated randomly, 50 candidate models from $\theta_{init} \sim N(0, \sigma)$ were generated per each test, choosing "the best" set of parameters θ that maximises the third quartile (75th) and maximum (max.) statistics, in order of importance, of NSE and PI evaluation coefficients during the validation period. In other works as in Toth and Brath (2007), the best model is the one that reached the minimum loss function (MSE) during training period. In Napolitano et al. (2011) the final model prediction is computed as the average of 50 random runs. The models tested in this work fit well for the training dataset, choosing the best as the model that captures the hydrological nature also for the validation dataset. Considering both periods can be identified which models fits better in both periods while avoiding over-fitting.

• On **Bayesian hyper-parameters**, α_h, β_h, within the Bayesian framework implemented by Nabney (2004). This parameters are optimised during the evidence procedure with the last 20% of the training dataset. We used five repetitions (epochs) during this procedure, to converge to the final posterior values of the dataset under study $P(\mathcal{D}|\alpha_h, \beta_h)$. A sensitivity analysis of the prior values and its influence on the general BNN's behaviour is carried out for four cases (GH3, GH4, GF1 and GF2) at daily basis under a model-based approach. In this manner the behaviour of the models given different values of prior α_h and β_h were analysed. Finally, the prior hyper-parameters established within the Bayesian framework were α_h=0.05 and β_h=50 for all cases. More on hyper-parameters foundations can be found on page

217. More supplementary material on BNN development in general is presented in detail in App.A.1.2.

GP The third DD method used is GP, which has a much higher computational cost than the previous ones. For this reason, its use at an hourly scale has been ruled out. Although there are previous GP works at hourly basis, e.g., Zhang et al. (2016), the number of data observations simulated is low, $n \approx 1500$. In our case, the observed data finally processed has been much higher, $n \approx 50000$, making its application impossible for practical and comparative purposes.

The core of this Bayesian method is the use of the kernel covariance functions (KCF), and its possible combination among them, constructing more complex kernels (multiple). Initially six different candidate KCF has been proven for all cases: four single kernels, and two dot product kernel, which are presented in Sec.A.1.3. For simplification purposes, in each regressive experience at daily scale only the "best" GP model that maximizes the performance coefficients is compared with the other two DD computational techniques (MLR and BNN) applied in this chapter.

Analysing the GP models deterministic results composed with just a single kernel (k_{Lin}, k_{NN}, k_{SE}, k_{Mat}) the k_{lin} behaves better than the others for some case studies, i.e., GH1 and GF3. This results coincides with the cases where a more linear behaviour of the basin has been confirmed, thus confirming a more linear relationship in the rainfall-runoff process for these case studies. Sorted in order of importance the k_{Lin}, k_{NN}, k_{Mat}, k_{SE} kernels have behaved better for our case studies. It should be noted that our results differs from previous works (Rasouli et al., 2012), which uses the k_{SE} successfully to related input variables and the target variable at the same temporal scale, not being the optimal kernel for our cases. As is commented in Rasmussen and Williams (2005) (p.83) "the squared exponential is probably the most widely-used kernel within the kernel machines field", remembering also the argument given by Stein (1999): "The oscillating behaviour of the spectral density would be unrealistic for many physical processes". Anyhow, in previous works (Grbić et al., 2013) a dot product kernel composed by the k_{SE}, a periodic kernel and a constant kernels is used to model the water temperature successfully. Similar forecasting works uses the squared exponential kernel as well, for instances in Karbasi (2018) in order to model evapotranspiration, and in Sun et al. (2014) to develop a monthly streamflow forecasting experiences for 430 basins across all the U.S.

As a general rule, the dot product kernels ($k_{P1} = k_{Lin} \cdot k_{NN}$, $k_{P2} = k_{SE} \cdot k_{NN}$) have better behaviour in four of the six datasets tested at daily basis (Tab.4.19). In this work, only two dot product kernels are tested by the trial and error procedure, combining in this manner different covariance relationships. Although it is expected that through more complex kernel structures, the underlying function beyond the rainfall-runoff process may be approximated in a more precise way and thus improve its predictive behaviour and capabilities. An interesting future work is the composition of more complex kernels for the cases under study. In this sense, recent important advances has been done in this direction with the automatic statistician project, which is introduced in Ghahramani (2015). Also works as presented in Duvenaud (2014) and Schaechtle and Mansinghka (2016) are being made to obtain multiple GP structures in an automated and straightforward way.

In Tab.4.19 the best resulting GP model for all the cases are summarised, where it can be seen that the k_P is the most frequent wining kernel. The GP$\sim k_{P1}$ structure is composed by two kernels, with stationary and non-stationary nature (k_{NN}), modelling correctly the river state under any regime. The stationary k_{lin} can reproduce abrupt changes produced by the predictors under the linear approach. In the case of the GP$\sim k_{P2}$ kernel structure, is driven by a stationary non-lineal kernel (k_{SE}). This can be visualised very well in the real function obtained by GP$\sim k_{P1}$ at the GF1 case, which is represented in Fig.A.13 for its domain.

Comparison analysis problem In Table 4.19, the findings on DD modelling for each short-term forecasting experience at daily basis remembering the main objective (*Obj.*) are summarised from the hydrological point of view. In each cell is exposed the technique that behaves better. For that, it is considered the performance of the *Deterministic* value, the *Uncertainty* quantification, the best *Kernel* for the GP models, and the wining rainfall assumption for the cases at daily basis.

TABLE 4.19: Summary of the regression experiences for different hydrological problems to identify the best computational technique tested and the best modelling approach for daily and hourly basis temporal scale

Study Area	*Deterministic*	*Uncertainty*	*Kernel Covariance Function*	*Lin vs Nlin*	*Obj.*	*Rainfall Assump.*
GH1	BNN	MLR/GP	k_{lin}	*Nlin/lin*	FP/FM	B
GH2	MLR	MLR	NA	*lin*	FP/FM	NA
GH3	BNN	BNN/GP	$k_{P(1,2)}$	*Nlin*	FP	B
GH4	BNN	BNN/GP	k_{P1}	*Nlin*	FP	A≈B
GF1	BNN/GP	BNN/GP	k_{P1}	*Nlin*	FP	A
GF2	GP	MLR/GP	k_{P1}	*Nlin*	FP	B
GF3	BNN	MLR	k_{P1}	*Nlin/lin*	WA	B

Objective (*Obj.*): Flood Management (FM) Flood Protection (FP) and Water allocation (WA)
Assump: Persistence Assumption NA: Does not apply

In general, the behaviour of BNN, GPs and MLR (ranked in order of importance) for a regression case are: medium for warning tools (floods protection) and medium-high for the application as management tools, at daily scale. The strengths found of GPs versus BNN for non-linear pattern recognition of a specific water/environmental system are: (*a*) the variety of transfer functions (by covariance through KCF) give more flexible tools for different phenomena adaptation and, (*b*) the calibration process is straightforward and therefore less uncertainty in the final model selection. In any case, the BNNs deterministic results, or their ensemble results, are better than the GP's deterministic results. Conversely, the bayesian inference for the uncertainty estimation associated to each GP results are better as we can see in the uncertainty discussion and the results summarise listed in Tab.4.19.

Summarising, the deterministic prediction capabilities of the BNN are higher than the GP. This is due to the chaotic nature of the BNNs calibration process where different random initial points, in a high dimensional parameter space, converge to different local minima in each weights optimisation run, offering single models, which maximises performance with superior predictive capabilities to the other models used. This has been carried out by performing multiple simulations, which is more time consuming and has higher computational cost but in contrast offers useful outputs for further modelling development. One of these outputs is the generation of a subset with the best models as a candidates members for ensemble

modelling, App.C.1. One of the main GP setbacks is the slow inference, but if we compare it to the time used for the search of the best model on BNNs development, or adding an extra methodological step as with the ensembles, the computation times are similar.

At the beginning, the non-linear models were found to provide more accurate forecast than the linear. However, Clements et al. (2004) concluded that there is no clear evidence in favour of non-linear over linear models in terms of forecast performance. In this work, we verified each case under different motivations, finding that there is not a fixed pattern due to the high variability on the differences among both approaches and the results in real cases, Tab.4.19. For instance, particularly in the GH case, linear relationships may overcome the non-linear under some conditions, i.e., at hourly basis (GH2). Also may exist more linear relationships among some hydrological processes not being necessary a non-linear approach, which may be more complex. In a manner, the GP model approach has an advantage over the others, which is the kernel configuration being able to choose a linear or a non-linear approaches, or set by the dot product kernel both approaches in one as we have seen in this chapter. The other DD techniques used in this work do not offer us this possibility. In reality, a priori we do not know which is the best approach, although we will always have the intuition that the non-linear approach will outperform the linear one. This is not always the case, as we have seen throughout this work. Then, we should work under a comparative by trial and error procedure, being always a very time consuming step.

Computationally speaking, a single run of MLR is almost instantaneous. The computation time (t_c) for the convergence of the optimal solution for one run during the training phase development, through Least Squared (LS) optimisation, is very fast, $t_c < 1\, seg$. The BNN takes also few seconds for the standard configuration (number of hidden neurons equal to the number of inputs, App.A.1.2), which is the MLP architecture used during the IVS procedure step. For this technique we must take into account learning parameters, such as the learning rate, the momentum, or the number neurons within the hidden layer, since this can slow down or speed up the computation time for a single random run considerably. For instance, as an average for a particular case at daily basis, e.g., GH1, with $n_{\mathrm{T}} = 4359$ observations during the training phase, the parameters optimisation phase trough the Scaled Conjugated Gradient (SCG) algorithm takes 40.3 seconds per run with 85 unit neurons within the hidden layer and three hours and a half for the total 1500 simulations under the dynamic approach (App.A.1.2). Finally, during the GP development for

this particular case, a linear kernel takes $t_c \approx 2\,min$ to estimate parameters. On the other hand, a neural network kernel takes ten times more ($t_c \approx 20\,min$) and a dot product kernel takes $t_c \approx 25\,min$ during the parameters optimisation. This has limitations when the search for the best kernel and its comparison is realised by the trial and error procedure. The computer used in this work is a Desktop PC with Intel Core i7 Processor (4x 3.6 GHz) and 16 GB DDR4 RAM, with Linux-based operating system.

4.8.3 Hydrological & Real flood forecasting tool development suitability analysis

In the following paragraphs, we focus specifically on the feasibility analysis of the regressions experiences for its preliminary use for hydrological purposes, and for the development of real flood forecasting tools. One of the most relevant coefficients for real forecasting is the Persistence Index PI (Kitanidis and Bras, 1980b; Kasiviswanathan et al., 2016), that has already been evaluated and discussed before in each case of study. In addition, we have analysed the floods under a classified framework (Sec.4.5). Alternatively, other attitudes need further analysis and are further studied in greater detail, in order to obtain a broader perspective of their behaviour. For this reason, we discuss two key points that need to be taken into account: (a). Hydrological applications, and (b). Flood management.

(a) Data-Driven Models for Hydrological Modelling: We work under more or less opaque frames. The interesting thing about these Data-Driven experiences is that we can extract information, for instances from the regression experiments, where we need to infer causation from it. The application of data-based models is currently applied in countless fields of knowledge according to all the jobs that appear constantly. This is convenient since, traditionally one of the major disadvantages attributed to DD models, specially to Artificial Neural Networks (ANN), is their lack of physical knowledge or interpretability. In a certain degree that affirmation is true, since the geophysical processes modelled are not driven by physical equations, i.e. "*black box*" approach. But as we have seen in this chapter, we can overcome some limitations as is the input relevance. From this, we can infer physical meaning by analysing it. In this sense, the recent advances as Deep Learning are transforming

the standard basic approach until now, creating a new opportunity on ANN modelling possibilities appears in hydrological sciences Marçais and de Dreuzy (2017); Shen et al. (2018); Shen (2018).

Other of the DD methods used in this work that fall in a more "*grey box*" group, as can be Gaussian Processes, are driven by its Kernel Covariance Functions (KCF). Should be commented that their development in this work has been a little superficial, not exploiting in depth all its potential. It is also worth mentioning that its methodological and inference framework is the most powerful and robust method, compared to the other DD models used from the statistical/probabilistic point of view. As a representative example of the modelling potential of complex time series with GP, refer to the Mauna Loa Atmospheric Carbon Dioxide (CO_2) averaged concentrations at monthly basis. This action is fully given in Rasmussen and Williams (2005) and Vanhatalo et al. (2013) on its Section 5.4.3 and Section 9.1.1 respectively. In these works, a periodic signal with a positive trend is successfully simulated. This practical examples demonstrated how the combination of simple KCF, derive in a GP additive regression model that fits remarkably to compound patterns, allowing us to simulate its future projection. This is effective under climate change scenarios.

(b) Flood Management: For a correct flood management, the existence of predictive tools of these high hydrograph values is of vital importance. Therefore, in Sec.4.5, we have analysed the predictive behaviour of the models at the upper tail ($>$ 99th percentile) on a daily scale. We have carried this analysis out by linear correlation. At hourly basis, we have analysed the extremes values under a classification framework of the different alert levels, which are those established by the local water authorities. For practical purposes, the results show us how the linear approach behaves better at hourly basis, getting acceptable classification rates (Tab.4.15). This shows us its suitability for application in flood management. The percentages rates decrease as the predictive horizon increases. At a daily basis, we see how the *Nlin* approach improves the *Lin*, getting an acceptable behaviour for practical application in the GH1-PS3B case for $h = 1 \rightarrow t+1$ time horizon (Fig.4.16). These results provide ample arguments with evidence to think about implementing these tools in spots of profit.

Also, in the GH1 and GH2 regression experiences, we can develop a decision framework to determinate the best choice of the daily releases of the reservoir system situated upstream. Is mandatory to relate the gauge station with independent variables as the upstream releases, which affect to the downstream river states. In

earlier Sec.2.4.2, we analyse recent wet periods with intense rainfall events in this basin. We discovered that the regulation capacity was limited during wet periods, where the reservoirs were at full of capacity.

The utility of resulting Symbolic expression (Eq.4.7), is that could serve as reservoir management tools, in order to optimise the daily releases through an iterative decision scheme, as is shown in Fig.4.22. At the beginning of the t period, we should know the $\{R_{(t-p)}\}$, $\{Q_{(t-p)}\}$ and $\{\mathsf{WH}_{Ca(t-p)}\}$ values, while only forecast of the $R_{(t)}$ variable will be necessary. This would be indispensable under hypothesis B, since under assumption A not forecast will be needed $R_{(t-1)}$. Then, the releases policies will be a function of the previous water states being easy to implement, since we always know them. The scheme allows us to optimise releases with an iterative approach (e.g. Monte-Carlo) to develop probabilistic risk zones on the floodplains with stochastic nature. The optimisation applications and the development of a decision tool for that purpose is out of the scope of this chapter, where recent works (Ehsani et al., 2016) show us its suitability. This matter provides an outline for future work to address this aim.

At the start of the work we considered to provide a decision framework based on the rainfall-runoff process, for the management and optimisation of floods with the collected data in this water system. To establish and reproduce as precisely as possible this hydrological process by modelling, and introduce a decision-dependent variable such as the releases. The results have shown us a medium-low dependence of the releases, where it is observed that the peak timing improves a little on both scales (daily and hourly). Anyhow, the arises at both time-scales offer a truly good starting point to continue working on the development of these hydro-informatics frames.

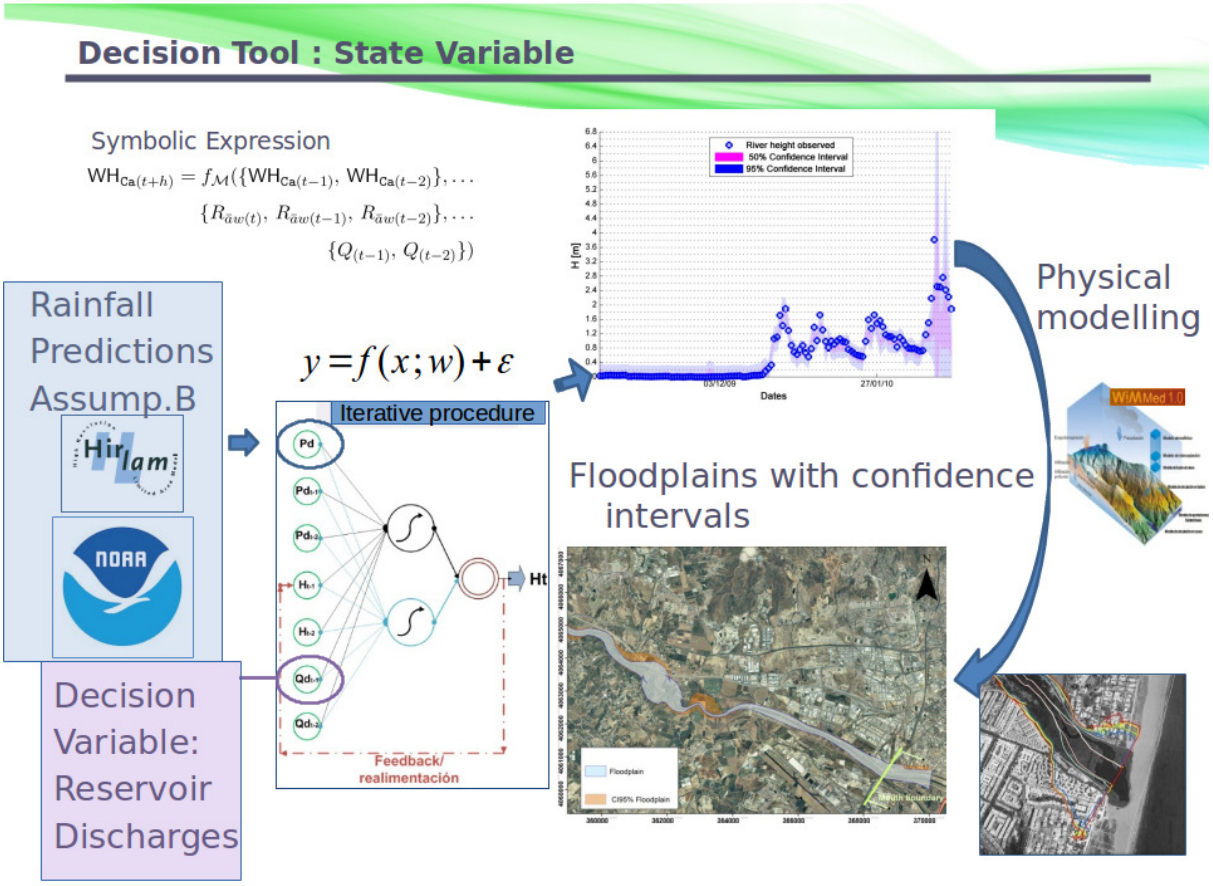


FIGURE 4.22: Optimisation through an iterative decision scheme for flood management with the resulting symbolic expressions of the regression experience carried out at Cártama gauge station

4.9 Conclusions

In this Chapter, we have made the following analysis for short term forecasting purposes:

- The behaviour of three Data-Driven (DD) computational techniques: Multiple Linear regression (MLR), Bayesian Neural Networks (BNN) and Gaussian Processes (GP) with respect their deterministic and their uncertainty outcome for hydrological management purposes.
- The suitability analysis of the DD computational techniques tested for real-time forecasting development with automated supervised learning frameworks.

We gave a compared report of three DD techniques of unique nature for short term forecasting (at hourly and daily basis) of river states by modelling the rainfall-runoff process. In this manner, we tested automatically the capabilities for forecasting development tools, at locations of interest, within the water systems under consideration.

For that, we managed hydro-meteorological data from varied sources: collected data on regional the monitoring networks and hydrological modelling. We decided the best predictors in each experience within a wrapper framework, by two distinct approaches, one linear (*Lin*), and other non-linear (*Nlin*). Likewise, we measured the input relevance under the linear/non-linear test set. Carried out that, the major findings of this study are:

- Whether the modelling approach is *lin* or *Nlin*, the initial step consists in the Input Variable Selection (IVS), to identify the optimal small-scale Predictor Structure that represents our rainfall-runoff process better. This PS should be simple as possible, looking for the optimal parsimony with a balanced amount of predictors needed that explain the model well. Especially with the main predictors, which are the rainfall records, where only the data series that are relevant, are selected. For this purpose, we carried out an iterative backward-wrapper selection method, i.e. model-based approach. In this way, we determined the most relevant input variables within the rainfall-runoff relationship in our Mediterranean water systems.
- At hourly basis, we found how the Multiple Linear Regression (MLR) method with re-sampling, overcomes to the other DD methods employed, on both deterministic and uncertainty performances. This is discernible from the model-based approach

comparison among the *lin* and *Nlin* test, which is carried out for various time horizons.

- During Bayesian Neural Networks (BNN) method development, the model architecture, learning and Bayesian parameters have to be carefully chosen for our specific hydro-meteorological data and conditions. From that, it is observed that the BNN model development is the most time consuming technique, compared with the other methods. Also because the calibration is not straightforward, and the solution space is very large. Anyhow, this computational technique overcomes the other DD methods tested, in both predictability and efficiency terms, in almost all the cases of study at daily basis. In general terms, we turned up that this computational technique is the choicest among the three DD compared.

- During Gaussian Processes (GP) method development, we analysed the kernel covariance function (KCF) decision. For that, single kernels and dot product kernels has been analysed as candidates KCF. Finally, we discovered that the dot product kernel (k_P), among a linear (k_{Lin}) and a non-linear (k_{NN} or k_{SE}) KCF, adapts better to our conditions than the resulting models driven by a single kernel. Also should be emphasized, the appearance of over-fitting problems as a true limitation for real industries. No extensive testing on our implementation of GP have been done. This is justified to the need for a stronger prior study of the Bayesian basics, which involves strong statistical inference ability.

- Under a deterministic point of view, the results at daily basis reveals that the best computational technique is BNN. Under the uncertainty point of view, the GP method produces the most balanced outputs for anticipating in our Mediterranean conditions. Also, it is pointed out the practicality of the joint use of hydrological modelling and DD techniques. This is effective in a basin of complex nature, where the snow is a fundamental hydro-meteorological agent. With our experiences, we showed the sense of the physical meaning within in a DD framework, to represent properly hydrological processes in Mediterranean water systems.

- In predictability terms, we prefer the use of *"black box"* models to more conceptual or physically based ones. This assumption has several reasons. First, we can easily formulate black box models, learning from data under a supervised framework. Second, these types of models are more easily to online applicability, and to develop *"data-based smart applications"*. And, finally, for forecasting purposes, the

prediction capabilities of the BNN technique, due to its chaotic nature, is a valuable strength to take into consideration. The Gradient Descent back-propagation optimisation algorithms find acceptable solutions in our study.

- The method followed in this chapter allows us to find the strongest connections, and to establish the most robust hydro-meteorological relationships within the rainfall-runoff process. The two basins are complex, both because of their size and geomorphological characteristics, and the physical scale effects. We identified which are the best spatial and temporal relationships between all the data collected in different monitoring networks, and at different time

Chapter 5

Medium Term Forecasting

"In the confrontation between the river and the rock,
the river always wins, not by strength, but by perseverance"

∼ Louis Agassiz ∼

5.1 Introduction

Forecasting on a medium/seasonal term horizon, on a monthly and quarterly basis for instance, is mandatory for water resources management purposes, and environments under Mediterranean climate conditions. In our specific study areas, the water managers decide the water volume allocations for each water use for the rest of the hydrological year (following spring and the summer seasons) in certain dates, at the end of winter or at the beginning of spring season.

It is plausible, and useful, to characterise the future hydro-meteorological behaviours as precise as we can, where dynamic tools may help to handle the uncertainty, concretely by inferring past data and conditions and therefore forecasting future hydro-meteorological states and time periods. Just having a glimpse about the hydro-meteorological values "in advance", the decision phase can be improved with high valuable information and tools to choose optimal strategies that avoid water shortage and controversies among the water users at different spatial scales, from regional to global. So, we can mitigate the costs associated with this negative scenario with forecasting through the development of DD tools and appropriate hydrological data frameworks.

To forecast medium-long term hydro-meteorological variables just based on antecedent data with its past values (an univariate case) is complex, because they move in very chaotic patterns. We need to focus on scales of a marked periodicity whenever possible, because that allows us to model an underlying pattern (components) and sense its tendency in advance about future volumes by several methods, e.g., wavelet decomposition. Also, exogenous inputs as atmospheric oscillations, may help. For multivariate cases, ML methods can be a suitable solution under a static modelling framework.

This chapter deals, in the first place, with seasonal predictions of the cumulated streamflow in a mountainous sub-basin (Narila) within the Guadalfeo River Basin under a static regression approach in Sec.5.2. Subsequently, we carry out a rainfall forecasting analysis at a monthly time scale with a Bayesian Neural Networks (Nabney, 2004). The average of the Mediterranean RBD and two relevant locations with collected data from the meteorological stations within the GF Basin are presented in Section 5.3. Section 5.4 deals with droughts forecasting, represented by Drought Indexes (DI), for the same locations than the earlier section. Finally, the chapter conclusions and main findings for medium term forecasting experiences are commented on Section 5.5.

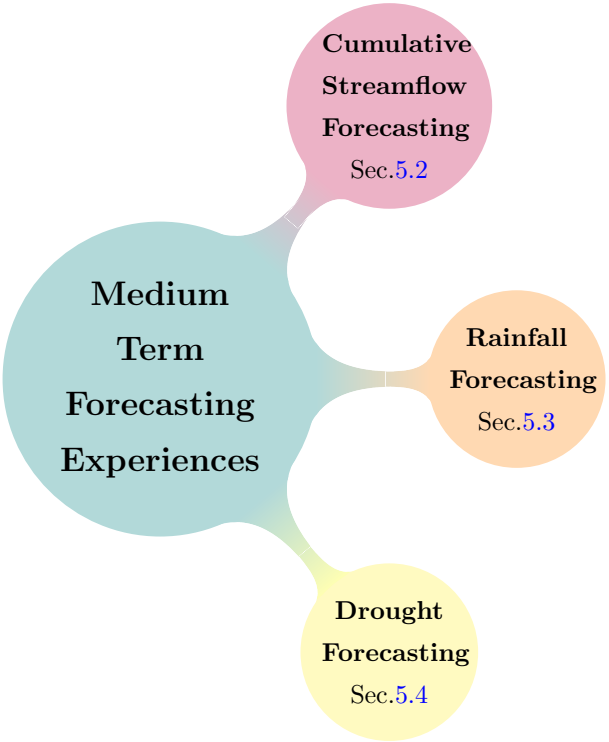


FIGURE 5.1: Medium Term forecasting cases

5.2 Cumulative Streamflow Forecasting

In this section we focus on quarterly-scale predictions within the Guadalfeo basin. A precise prediction on cumulated streamflow is an interesting approach for the water managers, and to gain knowledge about the water systems for management and operational purposes. A plain way to make cumulative streamflow forecasts can be to use historical streamflow time series with additional information as rainfall and other hydro-climatological variables. In this manner we search if there is an useful hidden hydro-meteorological pattern on the time-line for our water management objectives. In this regression experience, the objective is cumulated streamflow ($\mathsf{cSF}_{\mathtt{Nar}}$) forecasting at Narila gauge station within the Guadalfeo River for practical hydrological planning purposes.

The demand to foresee forthcoming hydrometeorological patterns is not a modern social practice. From immemorial times it is investigated, establishing temporal schemes, where specific indicators and hydrometeorological predictors provide us a notion of imminent patterns and/or trends. We note these folk predictions in Spain as "*cabañuelas*" (Blanc and Blanc, 2015), which are weather predictions based atmospheric observations and weather changes observed, for decades or centuries, in the first few days of January and August. In Andalusia it is more commonplace to look at them just in the first days of August. To achieve that, hydrometeorological indicators in partnership with the moon phase are drawn into account. In this empirical understanding, there is plenty to convince. These predictive precedents have remained in Portugal, Latin America, Africa (Roncoli et al., 2002; Changrsquo et al., 2013), India and Australia among some. In varied cultures, they pass these prediction exercises and its empirical knowledge from generation to generation.

In this study, the Gaussian Processes DD computational method is used as a modelling tool within a static framework, and the effect of two steps within the workflow modelling framework are quantified: (1) a binary classification according to the dry/wet character of the year and/or time period of years on the mid-term/seasonal regression forecasting problem, and (2) the inclusion effect of snow information from physical models in the mid-term/seasonal regression forecasting problem.

5.2.1 Study Site and Data

Narila gauge station is located in the upstream part (Fig.5.2) of the Guadalfeo River (Sec.2.2.2), with an associated contributing area of 67 km^2, where there have been no

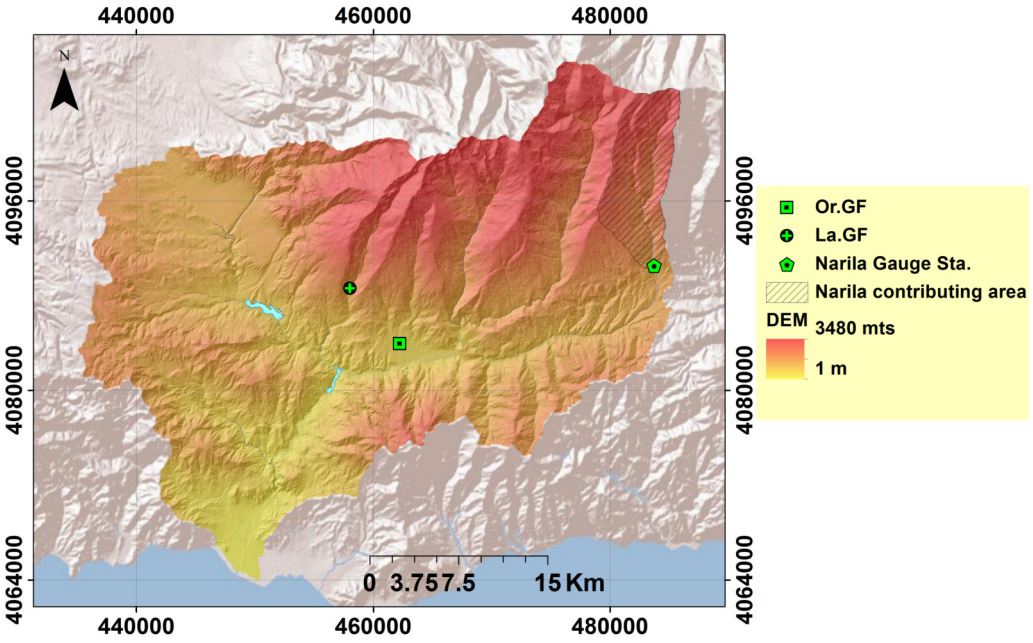


FIGURE 5.2: Data locations for medium-term/sesonal forecasting experiences carried out in this chapter. Narila gauge station and Lánjaron station `Lanj` for the cumulated streamflow regression experience (Sec.5.2). Lánjaron station (`Lanj`) and Órgiva station (`Org`) for Rainfall and Droughts regression experience developed in Sec.5.3 and Sec.5.4 respectively.

recent human alterations to the natural hydrological cycle. In the upper part of this sub-basin, also exits a traditional irrigation system called *acequias de careo* (Velasco and Serrano, 2011), being the oldest managed aquifer recharge system in Europe (Martos-Rosillo et al., 2019). This historical irrigation systems are important as at the height of Bérchules the river dries out frequently influenced by them. But, this water channel system, developed more than 1000 years ago, still acts almost naturally as it is integrated in the hydrological cycle to feed the traditional agro-systems of the region.

Different data sources were used for candidate predictors including previous hydro-meteorological information, from hydrological modelling and atmospheric oscillations (AO). The time series of river discharges cover from 1969 to 2012, a period of 43 years. The rainfall information is collected at the Lánjaron (`Lan`) meteorological station, the longest time series within this water system. Snow information in the form of Snowfall (R_S, quantified in $mm \cdot month^{-1}$ units) and Snowmelt (S, quantified in $mm \cdot month^{-1}$) time-series have been reconstructed by the hydrological physically based WiMMed model (Herrero et al., 2014) to reproduce the snow dynamics for the same modelled time period. As was demonstrated in the previous short-term forecasting experiences (Sec.4.4), the snow agent reproduction may help us with predictions in this water system. Finally the NAO,

the WEMO and the MOI (with its two versions) are the AO tested (Sec.2.4.1) as an exogenous variable to help the estimation capabilities.

5.2.2 Methodology

First, we form specific time frameworks from the hydrological point of view. The predictions are calculated within the same hydrological year (Oct-Sep) with the initial hydrometeorological conditions ($\mathcal{A}_0$) of two management scenarios: (Sce-A) after the first three months have passed, and (Sce-B) after six months of the current hydrological year. Then, we seek to estimate the cumulated streamflow (cSF, target variable) for several future time horizons $\mathcal{H}$ with different duration: at quarterly scale, at six-month scale, and at nine-month scale ($\texttt{hor} = 1, 2, \ldots, \mathcal{H}$), aggregating them sequentially. This temporal scheme and the meaning of the different hor for each scenario, is well described in Fig.5.3. We base the time scheme on a quarterly basis, approximated into the different seasons: winter (W), spring (S) and summer (Su). So at the end, we will have different time horizons under one or another scenario, such that A(1), A(2), ...,A($\mathcal{H}$) having in total eight cases, with $\mathcal{H}$ =5 and $\mathcal{H}$ =3 values for Sce-A and Sce-B respectively.

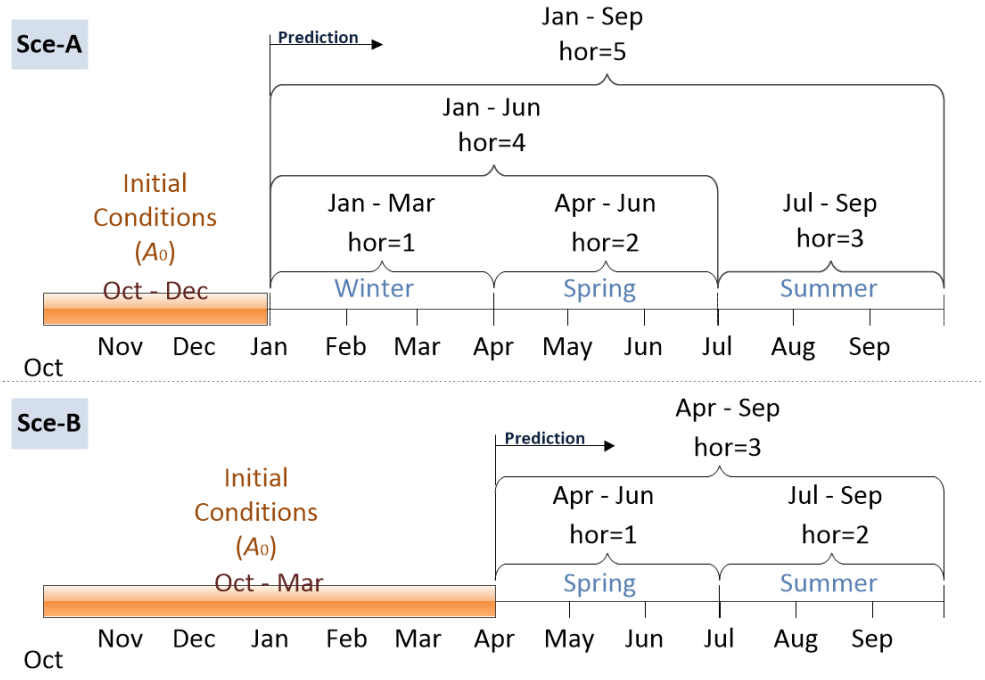


FIGURE 5.3: Temporal chart and time horizons (hor) for scenario A (Sce-A) and scenario B (Sce-B)

The input variables values are observations and simulations carried out during the initial conditions period ($\mathcal{A}_0$), which are quantified in the same metric units per time horizon ($\texttt{hor}$). The input variables used are: $\texttt{c}\bar{\texttt{SF}}$ for the cumulative streamflow measured in $Hm^3 \cdot \texttt{hor}^{-1}$, $\texttt{c}R$ for the cumulative rainfall measured in $mm \cdot \texttt{hor}^{-1}$, $\bar{AO}$ for the average value of atmospheric oscillations which is dimensionless, and $\bar{SC}$ for the cumulative snow information which offer measures in $mm \cdot \texttt{hor}^{-1}$ reproduced physically by the WiMMed model for the Narila contributing area (Fig.5.2).

Then, a binary classification of streamflow data series is carried out to improve the regression performance, as it is represented on a flowchart in Fig.5.4. This is a first attempt to classify the data set into two groups: dry (C_1) and wet (C_2) periods. A wet period means that the forecasted $\texttt{cSF}_{(t)}(\texttt{hor})$ value for a given t year, is higher than the historical observed mean ($> \texttt{c}\bar{\texttt{SF}}(\texttt{hor})$) for the same time horizon $\texttt{hor}$, and dry periods when the $\texttt{cSF}_{(t)}(\texttt{hor})$ observed value is equal or lower than the observed mean:

$$d(\texttt{cSF}_{(t)}) = \begin{cases} C_1 \; if \; \texttt{cSF}_{(t)}(\texttt{hor}) \; \leq \; \texttt{c}\bar{\texttt{SF}}(\texttt{hor}) \\ C_2 \; if \; \texttt{cSF}_{(t)}(\texttt{hor}) \; > \; \texttt{c}\bar{\texttt{SF}}(\texttt{hor}) \end{cases} \quad (5.1)$$

for a T number of observed years ($t = 1, \ldots, T$), that are used during the training period.

For that, different Machine Learning classifiers are tested and compared: Probabilistic Neural Network ($\mathcal{PNN}$), K-Nearest Neighbour ($\mathcal{KNN}$), and Support Vector Machine ($\mathcal{SVM}$). The Probabilistic Neural Network ($\mathcal{PNN}$, Wasserman (1993)) is a classification model with a two hidden layer network architecture, and the radial basis function (Fig.A.3) as a transfer function f_a. The $\mathcal{KNN}$ classifier (Fix and Hodges Jr, 1951) is a well known classification technique, a non-parametric method, being one of the most fundamental and simple method for classification problems. The last technique is $\mathcal{SVM}$ (Suykens and Vandewalle, 1999; Vapnik, 1995), a kernel method, with similar features and behaviours than $\mathcal{ANN}$. More on these classifiers foundations can be found in App.A.1.4.

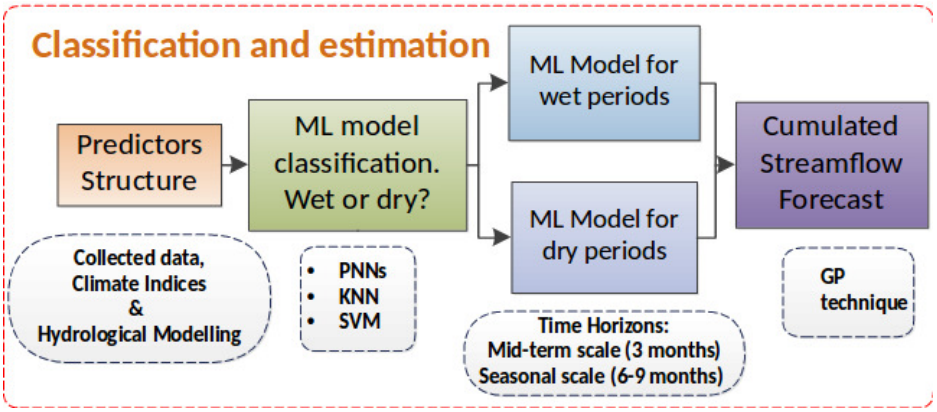


FIGURE 5.4: Flowchart showing the methodological steps of cumulative streamflow forecasting.

To question the classifiers, a cross-validation framework is set. For that, we launch 1000 simulations for each classification method and time horizon with a random choice of the training number of years. In each run, we randomly split the data set D into two folders: Training (25 years) & Validation (18 years). Under each scheme, we provide the final scores (overall accuracy, %), which is computed as the mean of the 1000 subset runs during the validation period to avoid biased results. To test the efficiency of the regression experience, we used the MAE and the RMSE coefficients (Sec.3.4). To compare the classification improvement on the baseline of the mentioned coefficients, we compute its Relative Improvement (RI, %) as follows,

$$\mathrm{RI}_i = \left(\frac{\varrho_i(\texttt{Class})}{\varrho_i(\texttt{NoClass})} - 1 \right) \cdot 100 \tag{5.2}$$

where $\varrho_i(\texttt{Class})$ is the i performance metric value obtained with the classification scheme and $\varrho_i(\texttt{NoClass})$ the outcome obtained with no classification, that means just with the static regression.

Finally, a static regression is applied to forecast cumulative streamflow for all the $\mathcal{H}$ horizons planned under each scenario, separately for each class (dry and wet). For that, a Gaussian Process is applied with squared exponential ($\mathcal{GP} \sim k_{SE}$, Eq.A.12) as the covariance kernel function (KCF). The $\mathcal{GP}$ method is choose for various reasons: the model calibration is straightforward, its non-linear nature, the predictable capabilities are similar to the $\mathcal{BNN}$, and the uncertainty quantification is better. A detailed discussion on this matter refer to Sec.4.8.2. The static regression per each hydrological year for a given set of predictors (I_d), lets say predictor structure (PS), is in the following form:

$$\mathsf{cSF}_{(t)}(\mathtt{hor}) = f_{\mathcal{GP}}(\mathsf{PS}) = f_{\mathcal{GP}}(I_1, I_2, \ldots, I_d) \tag{5.3}$$

On the predictors, for a given t-year, two types of Predictor Structure (PS) are defined (Eq.5.4) for a comparative assessment: PS1, with just the fundamental information and PS2, with the use of hydrological modelling that has the advantage to reproduce the snow agent on the predictions for water planning purposes.

$$\begin{aligned} \mathsf{cSF}_{(t)}(\mathtt{hor}) &= f_{\mathcal{GP}}(\{\mathsf{c\bar{S}F}_{(t)}(\mathcal{A}_0)\}, \{\mathsf{c}R_{(t)}(\mathcal{A}_0)\}, \{\bar{AO}_{(t)}(\mathcal{A}_0)\}) \Rightarrow \mathsf{PS1} \\ \mathsf{cSF}_{(t)}(\mathtt{hor}) &= f_{\mathcal{GP}}(\{\mathsf{c\bar{S}F}_{(t)}(\mathcal{A}_0)\}, \{\mathsf{c}R_{(t)}(\mathcal{A}_0)\}, \{\bar{AO}_{(t)}(\mathcal{A}_0)\}, \{\bar{SC}_{(t)}(\mathcal{A}_0)\}) \Rightarrow \mathsf{PS2} \end{aligned} \tag{5.4}$$

5.2.3 Results

• **Inputs:** Given an accumulated sum of observed precipitation ($\mathsf{c}R_{(t)}$), and an accumulated sum of Streamflow (cSF_t) during a t year, the AO explanatory capabilities (PS+AO) are quantified. The operation of two candidate expressions (Eq.5.4) are based on the structure of the $\mathcal{GP} \sim k_{\text{SE}}$ model. It can be seen how the AO inclusion slightly improves the classification model behaviour.

• **Classifiers Comparison:** Tab.5.1 shows the overall accuracy of the three classifiers ($\mathcal{PNN}$, $\mathcal{KNN}$, $\mathcal{SVM}$) for the wet and dry periods. In each one, we list eight cases that belong to scenarios A and B, with five and three time horizons, respectively. Having in mind the Fig.5.3, we have that the $\mathcal{PNN}$ classifier gets the strongest average behaviour for the first two horizons 1 and 2 of Sce-A. In contrast, it throws the poorest performances for Sce-B horizons. We look at the finest average behaviour for scenario B with the $\mathcal{SVM}$ classifier. It is worth noting, for this technique, the much greater improvement of the average behaviour compared with the other classifiers for A(3) and B(3) cases. Finally, with the $\mathcal{SVM}$ classifier, we bring a more balanced behaviour for all the trials. We base this on its mean value, represented as μ in Tab.5.1.

TABLE 5.1: Cross-Validation results. The overall accuracy (%) for the different scenarios and time horizons in Fig.5.3 comparing three type of classifiers are presented. Best average performance for each horizon and scenario during validation is highlighted in **bold** font. The best single performance is presented below in brackets [ ], and the best result of each time horizon is marked with double brackets [[ ]]. μ is the mean value of all the simulations for each classifier.

Method	Winter **A(1)**	Spring **A(2)**	Summer **A(3)**	W+S **A(4)**	W+S+Su **A(5)**	Spring **B(1)**	Summer **B(2)**	S+Su **B(3)**
$\mathcal{PNN}$	**65.36**	**52.50**	40.44	56.44	48.96	47.00	50.40	40.30
$\mu = 50.20$	[88.89]	[[83.33]]	[66.67]	[83.33]	[[88.89]]	[[77.78]]	[83.33]	[72.22]
$\mathcal{KNN}$	65.09	51.14	42.27	**57.6**	**49.1**	47.80	**50.50**	41.80
$\mu = 50.70$	[[94.44]]	[77.78]	[72.22]	[[88.89]]	[77.78]	[[77.78]]	[[88.89]]	[72.22]
$\mathcal{SVM}$	64.44	51.71	**57.37**	54.03	47.86	**50.20**	48.80	**52.00**
$\mu = 53.30$	[88.89]	[[83.33]]	[[83.33]]	[83.33]	[83.33]	[[77.78]]	[77.78]	[[77.78]]

• **Classification improvement:** In Tab.5.2 the comparison of static regression residuals with classification (`Class`) or without it (`NoClass`) is carried out. We make this step with $\mathcal{SVM}$ classifier. Tab.5.2 shows the GP ($\mathcal{GP} \sim k_{SE}$) regression performance in terms of MAE, SSE and RMSE (measured in $\text{Hm}^3 \cdot \texttt{hor}^{-1}$). As seen from this table, the accuracy of the `Class` method is better than the `NoClass` method for the validation period. The relative MAE, SSE and RMSE between the `Class` and `NoClass` methods for the validation period are presented. For all the scenarios and time horizons, we can see that the classification approaches improves the regression performance on the forecasting task.

TABLE 5.2: Regression evaluation without and with classification for all the models tested. Absolute error metrics (MAE, SSE, RMSE) are expressed in their original unit measures (Hm^3) per time horizon ($\texttt{hor}^{-1}$, according to scenario in Fig.5.3). The relative improvements (RI_i, Eq.5.2) are expressed in percentage (%).

Method	**Error**	Winter **A(1)**	Spring **A(2)**	Summer **A(3)**	W+S **A(4)**	W+S+Su **A(5)**	Spring **B(1)**	Summer **B(2)**	S+Su **B(3)**
NoClass	MAE	2.72	2.71	2.92	4.59	5.54	2.85	2.95	2.90
	SSE	232.97	243.87	254.70	651.61	817.41	247.94	258.19	253.30
	RMSE	3.60	3.68	3.76	6.02	6.73	3.71	3.79	3.75
Class	MAE	2.22	1.57	1.54	2.19	3.02	1.81	1.52	1.56
	SSE	163.44	93.35	115.25	221.18	232.31	131.32	102.77	117.80
	RMSE	3.01	2.27	2.53	3.05	3.59	2.70	2.39	2.56
RI (%)	MAE	18.4	42.1	47.3	52.3	45.5	36.5	48.5	46.2
	SSE	29.8	61.7	54.8	66.1	71.6	47.0	60.2	53.5
	RMSE	16.4	38.3	32.7	49.3	46.7	27.2	36.9	31.7

5.2.4 Discussion and Conclusions

In this section, we analysed cumulated streamflow forecasting for certain time horizons at seasonal time-scale, based on a quarterly basis. In this case study, we have added a classification step to the regression to verify whether or not it helps the predictions and to what degree. As in earlier experiences in this work, it has been quantified if there is improvement through two additional candidate predictors: atmospheric oscillations (AO), or by reproducing the snow component through distributed hydrological modelling.

To do this, we differentiated two scenarios (Sce-A and Sce-B), each one has its own temporal scheme (Fig.5.3) within the water year (t) in which the predictions are made. In the framework proposed, we used the cumulative and average values of the predictors of interest for the early three autumn months (Oct. to Dec.), and for the first six months (Oct. to Mar.) for Sce-A and Sce-B, respectively. Doing this, we set the initial conditions ($\mathcal{A}_0$), to forecast cSF for a time horizon (hor) of interest. In this sense, the results of the Sce-B shows that it is a viable solution for the water managers, because more information is available at the time of the prediction. Also coincides with the date of the year where the water authorities meet to allocate volumes for the rest of the water year, for the end of the spring and the summer period. In this way, we take into account the resource levels and the hydrological year until that moment, i.e., initial conditions.

For the classification step, we investigated three common methods: $\mathcal{PNN}$, $\mathcal{KNN}$ and $\mathcal{SVM}$. The results collected are homogeneous (Tab.5.1) where there is not an obvious "rightest" classifier for all tested cases. For example, we observe it how $\mathcal{PNN}$ has the poorest behaviour under Sce-B. On the contrary, it offers the highest behaviour for the first two horizons under Sce-A (A(1) $\rightarrow$ Winter, and A(2) $\rightarrow$ Spring). As seen, under Sce-B the largest behaviour is obtained with the $\mathcal{SVM}$ method. The best classifier for our hydrological scheme is $\mathcal{SVM}$, $\mathcal{KNN}$ and $\mathcal{PNN}$ sorted in order of importance (based on its μ value), taken the $\mathcal{SVM}$ classifier as the most balanced classifier in our data framework. This is in line with previous studies (Dezfooli et al., 2017). We can find an additional discussion on the classifier parameters in the App.A.1.4.

In the same way, related works were realised for water quality purposes (Modaresi and Araghinejad, 2014), finding that the $\mathcal{SVM}$, $\mathcal{PNN}$ and $\mathcal{KNN}$ classifiers, rated in order of relevance achieved the strongest performances. Dezfooli et al. (2017) compared the same three classifiers for water quality modelling, and concluded that the $\mathcal{PNN}$ is the rightest and the $\mathcal{SVM}$ the worst classifier. Khashei et al. (2012) got a similar pattern of results. They develop an innovative hybridization of artificial neural networks ($\mathcal{ANN}$) classification method and compares it with the $\mathcal{SVM}$ and $\mathcal{KNN}$ classifiers among others. This study concludes that its novel $\mathcal{ANN}$ model surpasses the classification skills of $\mathcal{SVM}$ and $\mathcal{KNN}$ ranked in charge of importance. Within the hydrological field, this comparative assessments are carried out with a greater number for water quality problems (Li et al., 2013; Modaresi and Araghinejad, 2014; Dezfooli et al., 2017). Although are studies on the matter (Modaresi et al., 2018), we have found a lower frequency regarding hydrological physical issues, such as volumes. It is considered convenient to carry out this classification modelling experiences, where a deeper examination is needed.

In our research, we chose the $\mathcal{SVM}$ for the cases into the subsequent methodological step for simplification. Since one of this exercise purpose was to test the classificatory step within the design structure (Fig.5.4), its relative improvement (RI) is quantified for each case study (Tab.5.2). In this sense, we recognize how the classification step in wet and dry periods always improves the posterior regression. Applying this classification scheme, we found a relative gain within the 16.4-71.6 % range. It is notable that the joint/nested modelling approach (classification and regression) outperforms the modelling approach just with the regression. This is important in Mediterranean environments because there is a great variability. Here we have made the simplest situation of classification (a binary framework), where if we migrate to multi-class label frameworks, it is expected that there will be a higher advance due to the solution space is more constrained.

By having an important nival component within the river basin, snow plays a fundamental role in the streamflow's amount that will be available afterwards in the coming time periods. As seen in an earlier experience (Sec.4.4), the snow information reproduced by hydrological modelling (WiMMed) improves the predictive behaviour on a daily basis. Although this water system has a quick response because of its size and slope. The results show that by using one modelled input, quantifying the snow component, the predictions improve. This could be tricky, since we are not taken into account other physical processes within the snow cycle, such as evapotranspiration or evaposublimination. This question is estimated in Herrero et al. (2016), finding that they are not despicable at all. Another limitation of these conceptualisations, is that we have ignored the groundwater runoff component in the conceptual model.

The main sources of uncertainty in informing decision-making depend on the variable being forecasted and on the forecast horizon, but there is also a source associated with the location (Crochemore et al., 2016). The methodological approach tested is generic to be adapted to other study cases. It would be only necessary to aggregate the observed data on a monthly basis. One limitation of this method, is that they need an enough length of observations to get reliable results. This is important when investigating decision support tools that are aimed at giving us an idea of future hydro-meteorological conditions. We want to open these methods into the basin management where the managers must rely on this information to make better planning. Results offer a basis for software search and other smart applications that surround water planning.

We proved in this section that hybrid methodologies of classification and regression are workable approaches to predict variables with hydrological interest under Mediterranean conditions. Here the accumulated volumes for different temporal periods. It needs a deeper probabilistic development to improve it, and to transform it, into a practical tool. The present findings confirm that exist underlying patterns, and/or hidden patterns that can be searched within a Data Driven framework. The quid of the question remains in the temporal planning schemes that apply.

5.3 Rainfall Forecasting

Rainfall cannot be predicted straightforwardly due to its chaotic nature and the complexity of the atmospheric processes involved. Rainfall is one of the most exigent variables to predict accurately. At medium-term time-scale, predicting rain within a water system is always a challenge. Besides, it is the major component within the Mediterranean water cycle. Therefore, its characterization is essential in hydrological planning, where the necessity of rainfall forecasting tools is indisputable at a medium term time scale.

The reality show us that the Globe's climate is a splendid illustration of the Chaos theory. Currently, on the big data era, the predictions are linked to physical based atmospheric simulations, i.e., Global Circulation Models (GCM), generating a huge amount of atmospheric variables at different time scales. To predict the rainfall accurately we need this complex hydrodynamic modelling. Then, a common strategy is to nest the data from global to a regional scale (Regional Circulation Models, RCM) (Giorgi, 1990; Grotch and MacCracken, 1991), using models with finer spatial resolution (Déqué et al., 2007), or adapted to regional conditions through downscaling techniques (Wilby and Wigley, 2000; Schmidli et al., 2006). Tripathi et al. (2006) showed how Data-Driven models also can help on the downscaling task. A more hydrological application is presented in Berg et al. (2018), which merge different data sets to offer near-real-time updated hydrological forcing data of precipitation. The problem with CGM and RCM models is that are complex, slow to calibrate, heavy data, and the uncertainty of the estimates increase with each step to be adapted into a more local scale.

Therefore it is necessary to develop automated tools, where as far as the data quality and observed length permits, allow us to prove time serie frameworks for prediction precipitation at local scales at a point of interest from historical data. Also will serve as a practice for the later development of technological frameworks where Data-Driven procedures are used jointly with GCM models. For example, to improve or reinforce the GCM forecasts at a more localised scale. Previous works showed us how to deal with this challenge under the DD perspective. Silverman and Dracup (2000) show how Artificial Neural Networks (ANN) are used to forecast rainfall in California. Abbot and Marohasy (2014) showed how that their ANN forecasts overcome the official Australian forecasts. Both are regions with similar dynamics on the hydro-meteorology to those found in south-east Spain.

In our site study, the rainfall periods extend from the Autumn to Spring. On September-October, heavy torrential rainfalls can take place due to the cold front phenomena. In

terms of annual volumes, a recent study Pérez-Palazón et al. (2015) shows a slight negative trend in the area for the last 50 years.

The monthly regression experiences are carried out in this section for the average rainfall of the Mediterranean River Basin District (`Med.RBD`, Fig.2.2) and two datasets collected by monitoring networks of the Guadalfeo River Basin (allocated within this District): `Lan`.GF and `Org`.GF (Fig.5.2). This is because there is a special interest on this region, due to the possibility of its coupling to integrated management tools at a basin scale, where its analysis is developed in the subsequent section. With this experience, we analyse the potentiality of the monthly rainfall rate forecast in the study area for decision task purposes.

This section is organised as follows: Sec.5.3.1 presents the data selected within the different study regions and the proposed DD method to forecast rainfall on a monthly basis. Sec.5.3.2 includes the results of the best set of predictors obtained in each case of study, and subsequently the rainfall forecasted values obtained for three time horizons. In Sec.5.3.3 the results are discussed. Sec.5.5 presents the final conclusions, separately and in conjunction, with the other medium-term regression experiences made through this chapter.

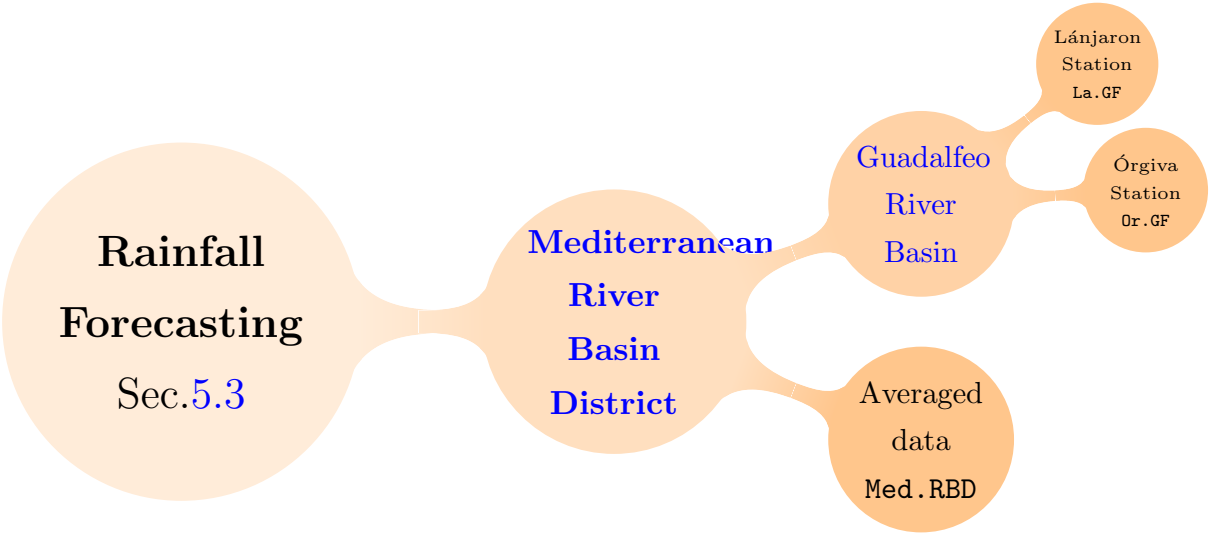


FIGURE 5.5: Rainfall forecasting cases studied in this section according to three different data sources

5.3.1 Data & Methodology

In this study, the average monthly rainfall data of the Mediterranean River Basin District (`Med.RBD`) and data collected at two locations in the Guadalfeo River Basin (`GF.RB`) are used (Fig.5.5):

`Med.RBD` The monthly data is provided by the water public administration through the REDIAM platform, calculated for the overall administrative basin district under its drought plan. Note that these data is for an area delimited by its Atlantic/western border, the Cádiz province, and its Mediterranean/eastern border, the Almería province (Fig.2.2). In geophysical and hydro-climatological terms is a very heterogeneous area since it includes regions very different hydrologically. Sinoga et al. (2011) report that the annual rainfall mean in Almería, the eastern part of this RBD, is 194.5 mm, in Málaga, the middle part, is 477.1 mm and in San Roque (Cádiz), the western part, is 838.2 mm. The `Med.RBD` dataset has 804 months in total (67 years) of rainfall observations (Tab.5.3).

`GF.RB` In this water system two daily rainfall values collected in Lánjaron (`Lan`.GF) and Órgiva (`Org`.GF) stations are used. Locations are shown in Fig.5.2. The daily values are aggregated to the cumulated values on a monthly basis. These stations have been chosen because they are those of the longest record registered, with better data collection quality based on previous experiences. The record period in these locations are large enough to develop regressive DD experiences on a medium-term time scale. In the GF RB the longest rainfall time-series is collected at the `Lanj` station with 736 months in total (≈61 years) of rainfall observations (Tab.5.3).

In Tab.5.3 statistical information of the time series used in the study is presented: the time period, the mean ($\bar{X}$) values, the maximum (X_{max}) values, and the standard deviation S_x values can be consulted for each case. The data subsets are listed for each methodological step with a 80:20 data partition ratio.

TABLE 5.3: Rainfall data properties and statistical parameters of the time series used in this section. Separated by a semicolon: the training and the validation subsets respectively, in brackets the value computed for the entire dataset is showed.

	Locations		
	Mediterranean (`Med`.RBD)	Lanjarón (`Lan`.GF)	Órgiva (`Org`.GF)
Time Period	01/1950-12/2016	09/1947-12/2008	07/1967-12/2008
Months (N)	643;161	589;147	403;84
$\bar{X}$ (mm)	41.5;43[41.8]	40.1;40.7[40.2]	41.4;30.1[37.3]
X_{max} (mm)	304.4;303.1	309.6;301.4	766.2;196.7
S_x (mm)	45.1;46.56[45.37]	48.0;48.8[48.2]	69.4;35.9[66.1]
$\bar{X} \rightarrow$ mean value, $X_{max} \rightarrow$ Maximum value $S_x \rightarrow$ Standard deviation.			

In this regression experience, our main objective is to forecast rainfall collected in a given location (j) within the study areas at different month lead time horizon ($h = 0, 1, 2$, Eq.5.5). This is done by establishing a functional relationship (f) driven by a Data Based model ($\mathcal{M}$) within a supervised learning framework for a given predictor structure (PS) with the following form

$$R_{j(t+h)} = f_{\mathcal{M}}(\mathsf{PS}) \tag{5.5}$$

For that, the PS is carefully constructed in a stepwise manner. The input variables within the PS used to forecast rainfall on a monthly basis are: antecedent values of order p_R of the target variable $R_{j(t-1)}, \ldots, R_{j(t-p)}$, a predictor representing the numerical month value of the last observation ($M = 1, \ldots, 12$), the Moving Average of order q (MA(q)), and antecedent values of order p_{AO} of Atmospheric Oscillations ($\{\text{AO}_{(t-p)}\}_{p=1}^{p_{\text{AO}}}$).

Some input exogenous variables are introduced within the PS to capture the time-scale modelled and the particularities of the rainfall patterns. The marked seasonal periodicity, represented by the month of the year, is one of them. That will introduce useful seasonal information to the modelling process. Also, MA(q) is tested and evaluated as a relevant input for several q order. The AO used are NAO (N), WEMO (W), and MOI with its two versions (M1 and M2 for MOI-GI and MOI-CA respectively) to test their capabilities as monthly rainfall predictors in the study areas. Some AO statistical indicators are listed in Tab.B.10, and more explanatory information of the AO variable can be found in Sec.2.4.1.

Then, given a monthly rainfall time series with N observations of the target variable $R_{j(t)}$, we attempt to develop a forecasting regression experience on a monthly basis. The computational Data-Driven technique in this case is Bayesian Neural Networks (f_{BNN}, Nabney (2004)). For that, a standard BNN configuration with a three layers feedforward architecture (Fig.A.3) is used, with the `tansig` as transfer function f_a within the hidden layer (`HL`) and a linear relational function in the output layer (`Out`). The optimal architecture definition during the BNNs development is also investigated, realising multiple simulations under the dynamic approach for a number of the following candidate values within the hidden layer `HL` : $\{1, 2, 3, \ldots, 25\}$ to search for the best regularised BNN model architecture with different number of unit neurons within the hidden layer. On the hyper-parameters values, the α_{h} value is set to 0.005, and the β_{h} value to 50. According to Nabney (2004), this parameter combination generates sharper results and less smooth model response, which is convenient, even crucial for rainfall applications under our hydro-meteorological conditions. Extreme rainfall events (wet) at a monthly basis are sporadic, and we need to examine during the evaluation a set of parameters that reflects correctly the cases collected so far. For this, we must have into account both, appropriate general behaviour, and positive behaviour during these wetter months.

For the evaluation of the results, the statistical performance metrics of R^2 (Eq.3.11), PI, RMSE and BIC are used (Sec.3.4). For test linear correlation, the Pearson coefficient (R, Eq.3.11) is used. Usually, the R coefficient is obtained via a least squares line. In order to interpret R, a value of +1 represents a perfect positive correlation, -1 a perfect negative correlation, and 0 indicates the total absence of correlation between variables. As in previous DD applications in this work, the methodological steps are the same: 1^{st}). there is a Training period (T, with a 80% data ratio) for parameters estimation, and 2^{nd}). there is a Validation period (V, with a 20% data ratio) on unseen data to check if the BNN candidate model generalises well. The bootstrap resampling is applied for the training period, generating an additional variability on the optimal space solution search, and non-biased outcomes to the sample selection for the training period. Sometimes, it also helps to correct the unwanted lagged effect on the resulting supervised models. For that, 25 bootstrapped models are tested per each configuration. This results in 625 simulations for PS analysed.

As in Chapter 4, the building model process is a mix of model free and model-based approaches. This is carried out in sequential steps under a stepwise approach as described below:

1. A first autoregressive version of the BNN model is defined. By partial autocorrelation function (PCAF), the antecedent order p_R of the target variable $R_{(t+h)}$ is established. Then an univariate autoregressive BNN model (PS1), the simplest model with only antecedent information is formed, obtaining the following symbolic expression form:

$$R_{j(t+h)} = f_{\text{BNN}}(\text{PS1}) = f_{\text{BNN}}(\{R_{(t-p)}\}_{p=1}^{p_R}) \tag{5.6}$$

2. Secondly, the antecedent Moving Average values ($\{MA(q)_{(t-1)}\}$) of order q, and p_M antecedent Month input values ($\{M_{(t-p)}\}_{p=1}^{p_M}$) are introduced and tested in a new PS in order to introduce a seasonal term within the PS. This attempts to help the underlying function to reproduce seasonal information, comparing if there is improvement over the previous PS. Then, the PS2 takes the form of the following expression,

$$R_{j(t+h)} = f_{\text{BNN}}(\text{PS2}) = f_{\text{BNN}}(\{R_{(t-p)}\}_{p=1}^{p_R}; \{M_{(t-p)}\}_{p=1}^{p_M}; \{MA(q)_{(t-1)}\}) \tag{5.7}$$

3. Thirdly, a subset of $O = 4$ candidate Atmospheric Oscillations (AO_o) are included within the relational function as an exogenous input. We use different AOs to test their suitability on the prediction capabilities. Then, if the model improves, a third PS (PS3) is established with the following expression:

$$\begin{aligned} \text{R}_{j(t+h)} = f_{\text{BNN}}(\text{PS3}) = f_{\text{BNN}}(\{R_{(t-p)}\}_{p=1}^{p_R}; \{M_{(t-p)}\}_{p=1}^{p_M}; \{MA(q)\}; ... \\ \{\text{AO}_{o(t-p)}\}_{p=1}^{p_{AO}}) \end{aligned} \tag{5.8}$$

4. The final expression is pruned using the CW_N values obtained from the whole bootstrapped simulations. About the input relevance characterised by CW_N in the BNN method, refer to previous Sec.4.2.1. Then, the median values of each CW_N input are established as its relevancy indicator, and the values under a threshold τ_{CW} are eliminated from the PS. The new optimised PS* is defined only with the more relevant inputs $\text{CW}_\text{N} > \tau_{\text{CW}}$, and compared with the previous PS under the stepwise approach.

Then, three different PS structures of input variables are evaluated for $h = 0$ time horizon: 1) just the univariate approach case (PS1), 2) the second version PS2, formed with the inclusion of the months vector (M) and the Moving average of order q (MA(q)) to the PS, and 3) PS3, formed with the inclusion of the best/optimal form, either single AO or AO

multiple combination. As throughout this work, this analysis is done under a stepwise framework, where we evaluate the model behaviour and complexity through different performance metrics, and sequentially by its relative improvements. This is computed with the following expression:

$$\mathrm{RI}_i = \left(\frac{\varrho_i(\mathsf{PS}^*)}{\varrho_i(\mathsf{PS})} - 1 \right) \cdot 100 \tag{5.9}$$

After the search, and the building procedure of the best PS in terms of R^2, PI, RMSE, and BIC metrics for 1-month ahead, a rainfall forecasting performance for 2-month and 3-months ahead ($h = 1, 2$, Eq.5.5) is evaluated in the same evaluation terms.

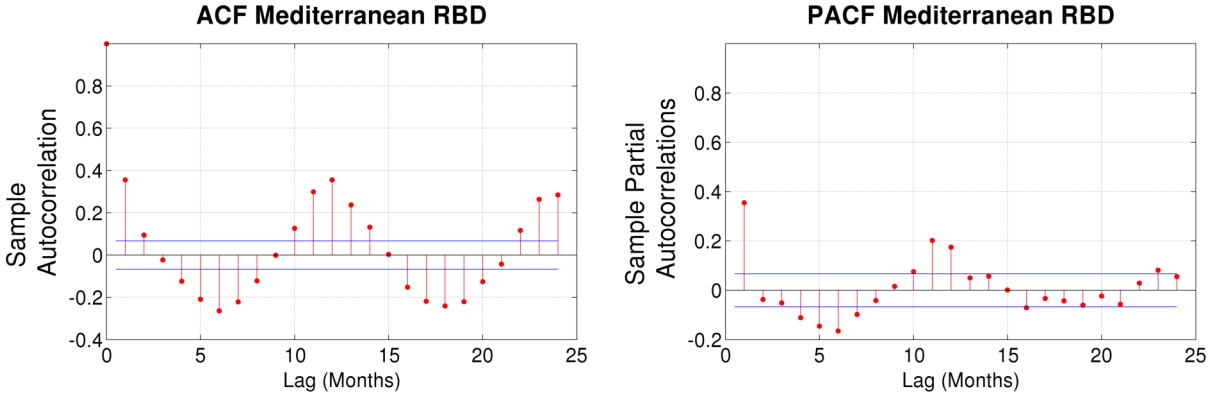
5.3.2 Results

In this section the performance of rainfall forecasting on a monthly basis is presented. First, in a stepwise manner, three combinations of input variables as competing predictors (PS1,PS2,PS3) are tested. Once the best PS is defined, a brief forecasting analysis is made for three time horizons in each case study. To do this, we list the "best" single BNN model from multiple bootstrapped simulations, where we focus on unmasked forecasting capabilities and their practical use in real conditions.

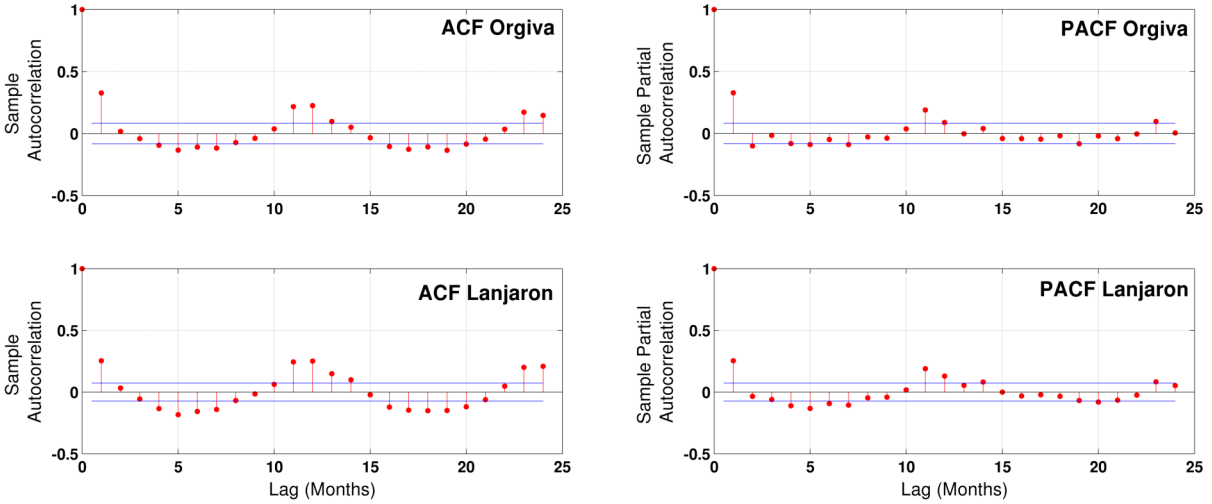
• Inputs Variable Selection results •

Firstly, for the PS1 definition (Eq.5.6), the autocorrelation values of ACF and PACF for the `Med.RBD` case and the GF RB cases are calculated (Fig.5.6). For it, a lag range of 0-24 months of this medium term forecasting experience is represented. It can be seen, for the three cases under study, poor autoregressive dependency for immediate consecutive lags until time slice $t - 13$, where the new period cycle takes place. This representation suggest that until time slices $t - 11/t - 12$ ($\approx$ a solar cycle), is the best observed pattern to include to help the model reproduce the seasonal cycle. The order $p_R = 12$ is a usual auto-correlation value at this time scale basis. So then, we are taken into account a window of the last antecedent year to forecast at three consecutive time horizons.

From the candidate models generated from the multiple simulations to search best parametrisation of PS1, the result of the "best" bootstrapped single model is presented in Tab.5.4. Their scatterplots of training and validation periods, for the three locations, are presented



(a) The autocorrelation function and the partial autocorrelation function of Mediterranean RBD monthly rainfall.



(b) The autocorrelation function and the partial autocorrelation function of collected rainfall time series in Órgiva and Lánjaron stations.

FIGURE 5.6: The autocorrelation function and the partial autocorrelation function of monthly Rainfall values.

in Fig.5.9. For the Med.RBD case, the BNN-PS1 model has a R^2 value of 0.294, a PI value of 0.4 and RMSE value of 39.60 ($mm \cdot month^{-1}$) during the validation period. For the Lan.GF case, it has a R^2 value of 0.356, a PI value of 0.45 and RMSE value of 42.02 ($mm \cdot month^{-1}$). The last example, the Org.GF case, has a R^2 value of 0.208, a PI value of 0.35 and a RMSE value of 35.41 ($mm \cdot month^{-1}$). The simpler PS form performs poor, giving bad results, but under the wrapper perspective, it is always useful and convenient, as a starting point, or baseline model for further comparison with future candidate models and/or predictors (PS).

TABLE 5.4: Best model performance metrics for the relevant PS evaluated at each case for $h = 0$ time step ahead horizon. PS3 values are taken from Tab.B.11. To follow stepwise predictor structure development: PS1 (Eq.5.6), → PS2 (Eq.5.7), → PS3 (Eq.5.8).

Location (z)	PS	R^2	PI	RMSE	BIC
Med.RBD	PS1	0.389;0.294	0.53;0.40	29.14;39.60	2538.6
	PS2	0.357;0.326	0.49;0.43	31.65;38.59	2557.9
	PS3	0.374;0.376	0.56;0.45	31.01;38.59	2299.9
	PS3*	0.387;0.347	0.50;0.44	34.65;38.98	2454.3
Lan.GF	PS1	0.434;0.356	0.62;0.45	39.65;42.02	2250.5
	PS2	0.535;0.419	0.71;0.54	34.30;38.12	2177.9
	PS3	0.363;0.429	0.58;0.59	25.74;26.28	2222.8
	PS3*	0.419;0.459	0.61;0.61	36.17;36.02	2211.1
Org.GF	PS1	0.640;0.208	0.80;0.35	31.91;35.41	1838.5
	PS2	0.548;0.275	0.70;0.40	33.76;34.11	1816.2
	PS3	0.356;0.285	0.62;0.48	36.98;31.93	1847.7
	PS3*	0.421;0.253	0.69;0.41	40.14;33.83	1887.2

During the PS2 definition (Eq.5.7), two new inputs are introduced, representing the antecedent month of the solar year ($\{M_{(t-1)}\}$), and the moving average ($\{MA(q)_{(t-1)}\}$), to evaluate if it improves with respect to the behaviour of the previous PS1. In this step, in order to choose the optimal q order, several values are correlated with the target variable, selecting finally $q = 12$ order value that maximises the linear correlation. The result of the "best" bootstrapped single model is presented in Tab.5.4 and its scatterplot in Fig.5.9 for all the study cases. During its respective validation periods, at the Med.RBD case, the R^2 value is 0.326, the PI value is 0.43 and the RMSE value is 38.59 ($mm \cdot month^{-1}$). Alternatively in the Lan.GF case, R^2 value of 0.419, a PI value of 0.54 and RMSE value of 38.12 ($mm \cdot month^{-1}$) is obtained. Finally for the Org.GF case, a R^2 value of 0.275, a PI value of 0.40 and RMSE value of 34.11 ($mm \cdot month^{-1}$) is found. Superior results are seen for all the cases compared with previous PS1. In Tab.5.5 is quantified numerically by their respective RI_i (Eq.5.9) over previous model (PS1→PS2).

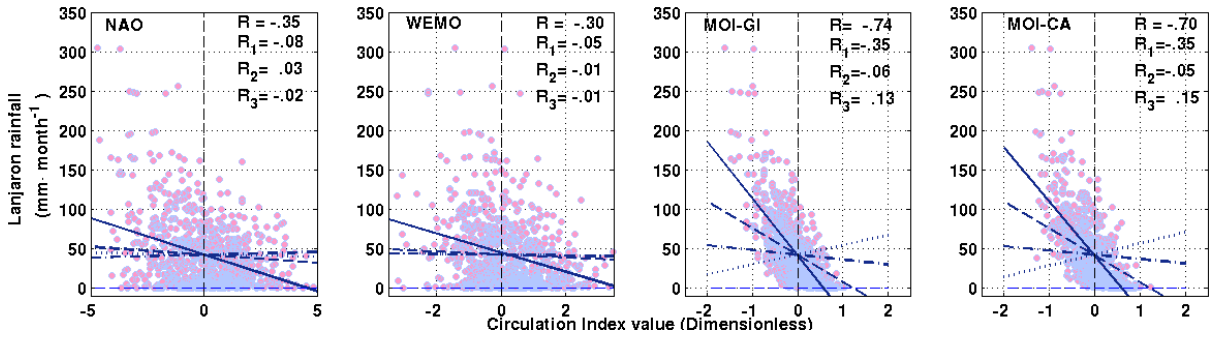
Finally for the PS3 (Eq.5.8) definition, the sensitivity of individual and grouped Atmospheric Oscillations (AO) in forecasting rainfall is evaluated on a data model-based approach. In this manner we are testing its influence and suitability as an exogenous

inputs. This is carried out by correlation, and in parallel by the trial and error procedure, we tested and compared them with each other, analysing which one helps better to forecast monthly rainfall at all the cases selected in this section (Tab.B.18).

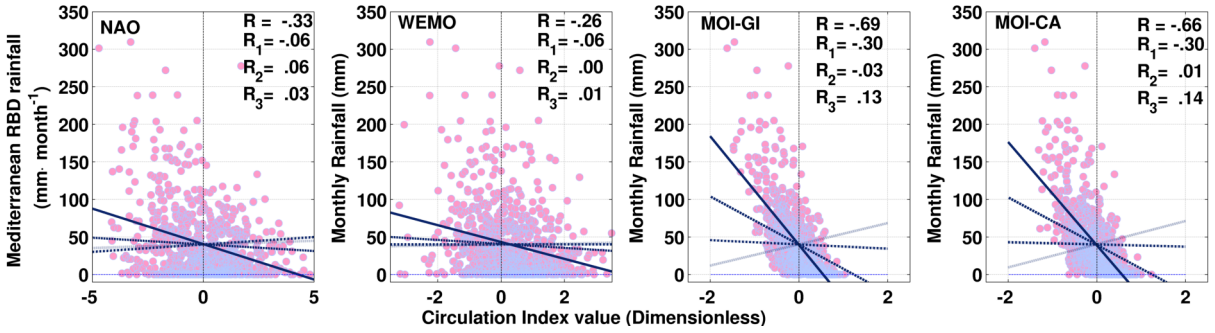
TABLE 5.5: Relative Improvements (RI_i) during the PS configuration over the most significant evaluation coefficients on a monthly basis: R^2, PI and BIC. When the $RI_i\tau >$ 0.05 condition is complied, is highlighted with bold font.

Location (z)	PSx →PSx+1	$RI_1(R^2)$	RI_2(PI)	RI_3(BIC)	RI_4(RMSE)
Med.RBD	**PS1→PS2**	−8.2%;**+10.9%**	-7.5%;**+7.5%**	-0.7%	-8.61%;+2.55%
	PS2→PS3	+4.8%;**+15.3%**	+14.3%;+4.65%	**+10.1%**	+2.02%;±0%
	PS3→PS3*	+11.5%;-7.7%	-10.7%;+2.2%	-6.7%	-11.70%;-1.01%
Lan.GF	**PS1→PS2**	+23.3%;**+17.7%**	+14.5%;**+20.0%**	+3.20%	+13.49%;**+9.28%**
	PS2→PS3	-32.15%;+2.38%	-18.31%;**+9.26%**	-2.06%	+24.95%;**+31.06%**
	PS3→PS3*	+15.42%;**+6.99%**	+5.17%;+3.39%	+0.53%	-40.52%;-37.06%
Org.GF	**PS1→PS2**	-14.4%;**+32.2%**	-12.5%;**+14.3%**	+1.21%	-5.80%;+3.67%
	PS2→PS3	-22.8%;-15.3%	-7.14%;**+10.0%**	-4.9%	-9.53%;**+6.39%**
	PS3→PS3*	+15%-11.2%	+11.3%;-14.6%	-2.2%	-8.54%;-5.95%

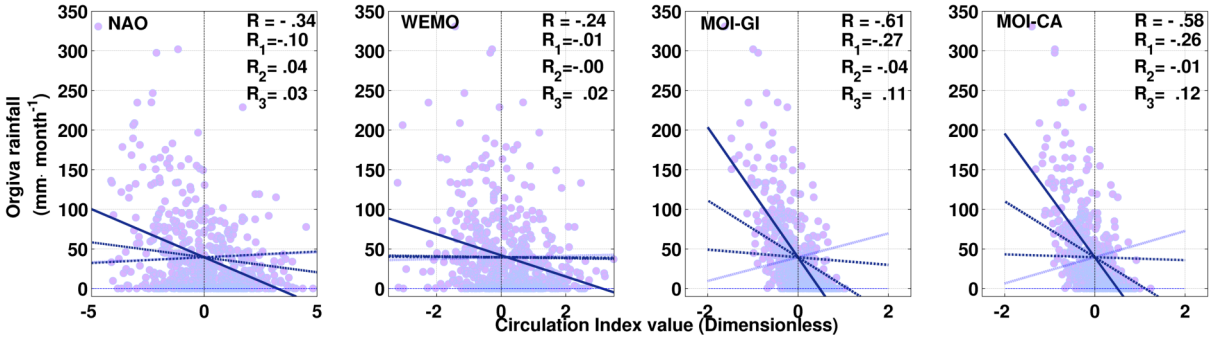
In Fig.5.7 the scatterplot of relationship between the monthly rainfall ($R_{z(t)}$) and four different Atmospheric Oscillations ($AO_{o(t)}$) are represented. The Pearson correlation for t, $t-1$, $t-2$, and $t-3$ are listed (represented by R, R_1, R_2, R_3 respectively), also the least square line (LSL) of this scatterplots are plotted to visualise the true regression line. From the scatterplots is clear that the MOI index, with its GI and CA versions ranked in order of importance, has the highest correlation for the three target rainfall datasets. It shows a strong negative correlation, $R \approx -0.7$. This implies that the negative AO phase (i.e., wet or cold phase), is associated with rainfall events of the three datasets. Alternatively this correlation is much lower for the negative phases of NAO and WEMO. The correlation evaluation of the data presented in Fig.5.7, also revealed a significant a steep decline as we move backwards on the time line, that means cross-correlation among the target ($R_{z(t-1)}$) and AO antecedent values. These are indicated with R_1, R_2 and R_3 for $AO_{o(t-1)}$, $AO_{o(t-2)}$, and $AO_{o(t-3)}$ respectively. From the figure, it is clear that the antecedent MOI values are weakly correlated, $R_1 \approx 0.3$, while there is no correlation at all with NAO and WEMO indexes antecedent values, $R_1 \approx 0$. Ranked in order of importance, this correlation analysis shows that the WEMO and NAO has the poorest linear correlation.



(a) Scatter plots among the Med.RBD rainfall data set and four atmospheric oscillations



(b) Scatter plots among the Lan.GF rainfall data set and four atmospheric oscillations



(c) Scatter plots among the Org.GF rainfall data set and four atmospheric oscillations

FIGURE 5.7: Scatterplots visualising correlation between atmospheric oscillations and rainfall values on a monthly basis used in Sec.5.3 for medium-term forecasting experiences. R_1 corresponds the correlation among $R_{(t)}$ and $AO_{(t)}$ (solid line ——), R_2 corresponds the correlation among $R_{(t)}$ and $AO_{(t-1)}$ (dashed line – – –), and R_3 corresponds to the correlation among $R_{(t)}$ and $AO_{(t-2)}$ (dashed-dot line – · – · –), and R_4 corresponds to the correlation among $R_{(t)}$ and $AO_{(t-3)}$ (dashed-dot line · · ·)

At the same time, several AO predictors combinations are tested in a model-based approach in order to select the best AO combination for its inclusion in the final PS3. For that, just one antecedent value is used $p_{AO} = 1$ as the optimal value, while $p_{AO} = 2$ and $p_{AO} = 3$ produced over-fitting, and therefore a worse generalization in all cases. Then, the AO resulting combinations can be: as a single predictor (N, W, M1 or M2), as a

combination of pairs (N-W,N-M1,N-M2,W-M1,W-M2 or M1-M2), as a combination of triplets (N-W-M1, W-M1-M2,N-W-M2 or N-M1-M2), or using all (N-W-M1-M2) within this PS version (PS3). This results are presented and analysed in a more extend form in App.B. For the Med.RBD case, the best combination is the model M1, which contains the pair N-W. This provides a performance of 0.376,0.45 and 38.59 ($mm \cdot month^{-1}$) values for R^2, PI and RMSE metrics, during the validation period. On the other hand, for the Lan.GF case, the most balanced combination is also a pair (M8), which is formed by the W-M1 combination. For this case, the performance during the validation period is 0.429, 0.59 and 26.28 ($mm \cdot month^{-1}$). Finally, for the Org.RBD case, the best PS3 contains just one AO value, which corresponds to the WEMO index. For the same units and metrics, respectively this model has a performance of 0.285, 0.48 and 31.93. Corresponding to the PS2→PS3 stepwise evaluation, Tab.5.4 shows the performance metrics for each study site at $h = 0$ time horizon forecasting. For the three study cases in this section, it can be seen how their performance significantly improves until the PS3 model, passing the rule of a evidence of a minimum 5% improvement ($\text{RI}_i \geq 5\%$) at least in two performance metrics.

A last optimisation step by pruning the final PS (PS3→PS3*) is realised. In practice, this step only improves performance in the Lan.GF case, but not enough to establish it as the best PS, $\text{RI}_i > 5\%$. Tab.5.5 shows that the PS3* model behaviour has better performance based on the efficiency (ϱ_1), and on predictability (ϱ_2) metrics. On the other hand, on the parsimony (ϱ_3) term it remains almost the same, and on the residual coefficient (ϱ_4, RMSE) it decreases significantly. In this particular case, with a τ value of 0.4, only two variables were established as unimportant: $\{R_{\text{Lan}(t-5)}, R_{\text{Lan}(t-4)}\}$ (Fig.5.8). This is based in its median CW_N value obtained for the 625 runs of the PS3 model which are represented in boxplots.

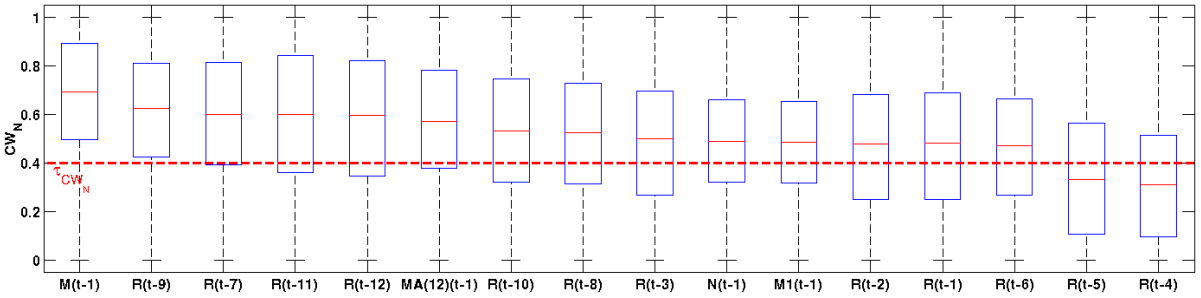
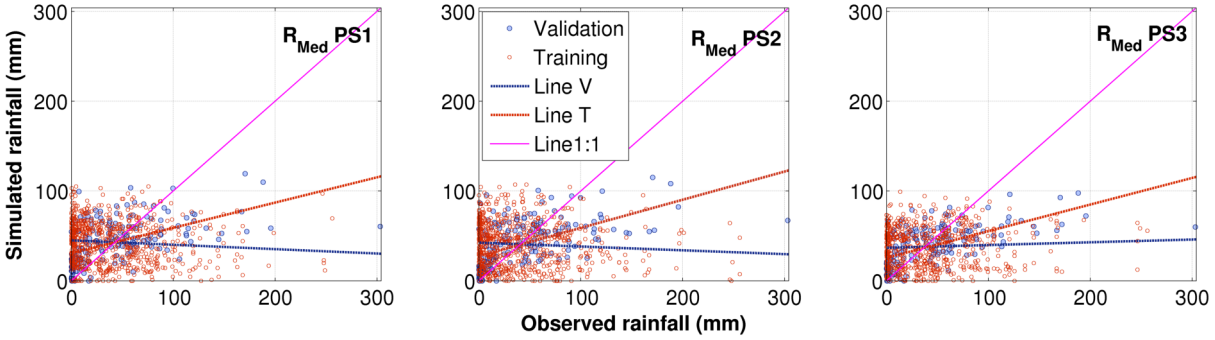


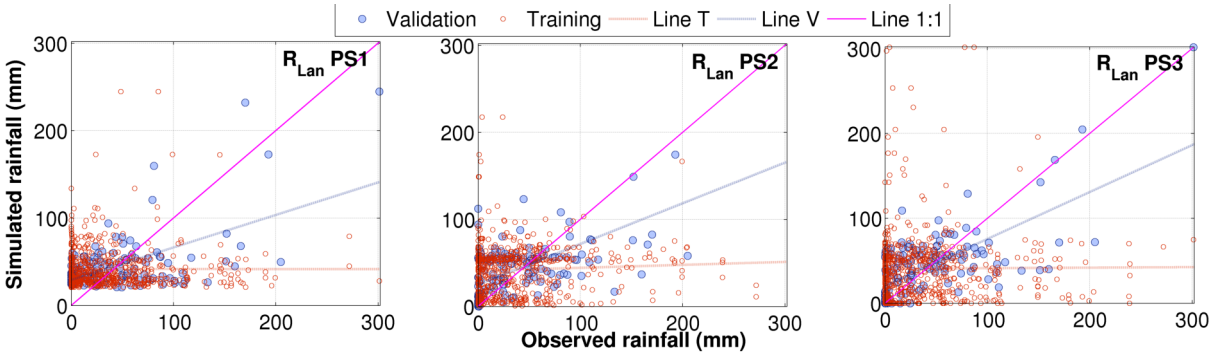
FIGURE 5.8: Inputs relevance representation for the Lan.GF-PS3-BNN model. Boxplots of its Connection Weights (Sec.4.2.1, CW_N) values of a total of 625 bootstrapped runs different hidden units within the hidden layer (HL). With horizontal red dashed line the relevance threshold ($\tau_{\text{CW}} = 0.4$) used for pruning purposes.

In all the cases, the evaluation of its final inclusion is realised if generalises better and if comply the condition $\mathrm{RI}_i > \tau_{\mathrm{RI}} = 0.05$, that means a minimum relative improvement of 5% based on the i performance metric, on at least two ϱ_i for $i = 1, \ldots, 4$ performance metrics. It can be seen (Tab.5.5) how this improvements are homogeneous directing to more balanced models. We must bear in mind, that a poorer performance during the training period accompanied by an improvement during the validation period is an advisable behaviour. It points to a minor overfitting problems, and therefore a better generalisation for unseen modelling cases. This is particularly important when investigating the best PS among various potential candidates.

Finally, in Eq.5.10, Eq.5.11 and Eq.5.12, the final expressions used for the forecasting experience are presented. This is carried out for the three rainfall datasets selected, at three different time horizons ($h = 0, 1, 2$) in the following paragraphs.



(a) PS1, PS2 and PS3 BNN-predicted rainfall values at $h = 0$ lead time in the Med.RBD case.



(b) PS1, PS2 and PS3 BNN-predicted rainfall values at $h = 0$ lead time in the Lan.RBD case

FIGURE 5.9: Scatterplots of simulated values (y axis, $mm \cdot month^{-1}$) versus observed rainfall values (x axis, $mm \cdot month^{-1}$) for training (T) and validation (V) periods for the three different Predictor Structures (PS1, PS2 and PS3) evaluated at $h = 0$ lead time during the Input Variable Selection procedure, (Tab.5.4)

$$R_{\texttt{Med}(t+h)} = f_{\text{BNN}}(\{R_{(t-p)}\}_{p=1}^{12}; \{M_{(t-1)}\}; \{MA(12)_{(t-1)}\}; \{\text{N}_{(t-1)}, \text{W}_{(t-1)}\}) \tag{5.10}$$

$$R_{\texttt{Lan}(t+h)} = f_{\text{BNN}}(\{R_{(t-p)}\}_{p=1}^{12}; \{M_{(t-1)}\}; \{MA(12)_{(t-1)}\}; \{\text{N}_{(t-1)}, \text{M1}_{(t-1)}\}) \tag{5.11}$$

$$R_{\texttt{Org}(t+h)} = f_{\text{BNN}}(\{R_{(t-p)}\}_{p=1}^{12}; \{M_{(t-1)}\}; \{MA(12)_{(t-1)}\}) \{\text{W}_{(t-1)}\} \tag{5.12}$$

• Rainfall forecasting at the Mediterranean River Basin District •

Med.RBD The results of the best single PS3-BNN model (Eq.5.10), in terms of R^2, RMSE, PI and MAE, for 1-3 month lead time are shown in Tab.5.6. In this particular case, the PS3-BNN model performs better for 3 month lead time forecasting horizon with higher R^2 and PI; and lower RMSE and MAE than the 2 month lead time forecasting horizon ($h = 1$) during the validation period. The performance does not deteriorate as we increase the lead time forecasting horizon. We must not forget that target variable is the rainfall mean for a fairly large and heterogeneous area (Fig.2.2).

TABLE 5.6: Med.RBD rainfall results for training (N=643) and validation (N=161) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($mm \cdot month^{-1}$).

h	R^2	RMSE	PI	MAE
$0 \rightarrow (t)$	0.374;0.376	31.01;38.60	0.56;0.45	22.07;25.38
$1 \rightarrow (t+1)$	0.540;0.275	31.68;43.17	0.76;0.48	21.57;28.94
$2 \rightarrow (t+2)$	0.535;0.277	28.02;41.28	0.77;0.62	19.49;26.40

• Rainfall forecasting in the Guadalfeo River Basin •

Lánjaron The results of the best single PS3-BNN model (Eq.5.11), in terms of R^2, RMSE, PI and MAE, for 1-3 month lead time are listed in Tab.5.7. The PS3-BNN model

performs better for 3 month lead time forecasting horizon with higher R^2 and PI, and lower RMSE and MAE than the 2 month lead time forecasting horizon ($h = 1$) during the validation period. In this case, the best performance does not deteriorate as we increase the lead time horizon, on the contrary, it improves the behaviour for almost all performance metrics.

TABLE 5.7: Lan.GF rainfall results for training (N=589) and validation (N=147) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($mm \cdot month^{-1}$).

h	R^2	RMSE	PI	MAE
$0 \rightarrow (t)$	0.363;0.429	36.51;36.91	0.58;0.59	25.74;26.28
$1 \rightarrow (t+1)$	0.792;0.372	23.10;41.96	0.89;0.57	15.47;29.72
$2 \rightarrow (t+2)$	0.679;0.409	28.01;40.06	0.85;0.65	17.49;26.79

Órgiva In Tab.5.8 the results of the best single PS3-BNN model (Eq.5.11) for three time horizons are presented. The results for the Órgiva station are the weaker, compared with the previous two cases, for the 2- and 3- month lead time forecasting horizon. This may be due to its data length added with the orographic effects, that means stronger influence of regional effects. Since the physical location of this meteorological station is more physically enclosed than Lánjaron location. For this case, the efficiency deteriorates as we increase the lead time horizon. From the results, it is clear that there is a lack of representative historical data at a monthly basis for DD applications.

TABLE 5.8: Org.GF rainfall results for training (N=403) and validation (N=84) period respectively, separated by a semicolon (T;V). Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($mm \cdot month^{-1}$).

h	R^2	RMSE	PI	MAE
$0 \rightarrow (t)$	0.356;0.285	36.98;31.93	0.62;0.48	26.67;26.13
$1 \rightarrow (t+1)$	0.219;0.243	55.45;41.26	0.59;0.34	35.07;33.69
$2 \rightarrow (t+2)$	0.378;0.188	47.55;38.24	0.71;0.46	30.43;29.21

5.3.3 Discussion

As previous forecasting experiences carried out throughout this work, the IVS procedure for Rainfall forecasting is conducted with the BNN method under a wrapper manner within a stepwise view. This recurrent procedure follows the basics steps of the generic flowchart of any automatic IVS algorithm (Galelli et al., 2014). According to the non-linear assumption of the BNN's hidden transfer function (hyperbolic tangent, Eq.A.1.2), this approach reports faithfully on non-linear benchmark problems. The main limitation is the lack of the use of non-linear correlations measures, i.e., the coefficients used for the model free approach steps. Future research could examine another kind of measures as the Spearman's rho (Eq.3.12), and the mutual information criterion (Shannon et al., 1951).

To optimize the winning predictor (PS), an additional pruning step was introduced as a last step during the PS building procedure. In practice, it never improves the best previous PS candidate, based on evaluation and parsimony coefficients (Tab.5.5). For that reason, it remains unclear to which degree structure optimisation capabilities are attributed to our pruning method. It is a question of future research to investigate its optimisation capabilities within a comparison assessment framework with other pruning methods. One option could be the Optimal Brain Surgeon method (Hassibi and Stork, 1993), which is implemented in Nørgård et al. (1996), a ANN based toolbox for system identification.

The properties of the atmospheric oscillations datasets related to rainfall on a monthly basis at various locations are explored in depth. The best AO is different per each case, not finding a dominant oscillation for all cases. By the model-based approach procedure, we map interactions and dependencies between the oscillations (Tab.B.11). Due to the local focus of our study, we chose regional atmospheric oscillations such us NAO, WEMO and MOI. The most influential AOs are combination of pairs with NAO and WEMO, while in some cases the MOI-GI inclusion appears more convenient. The combination of pairs outperformed single, combination of triplets, and the complete set of atmospheric oscillations. In Kalra and Ahmad (2009) a similar Data-Driven study is carried out, reporting how the use of AO pairs results in better predictability of streamflow compared with individual oscillations. The use of all combinations is advisable although can be very time-consuming. These oscillations, either atmospheric or ocean, are always interesting to study since they can link all kinds of studies with oscillations at global scales (Stenseth et al., 2003). However, another type of AO such as the East Atlantic

(EA) which can be viewed as a "southward shifted" NAO pattern, and the East Atlantic Western Russia (EAWR) pattern (Barnston and Livezey, 1987), among others, have been left out of the study. Also oceanic oscillations as the Atlantic Multidecadal Oscillation (AMO, Schlesinger and Ramankutty (1994); Delworth and Mann (2000)) can be used as explanatory variable for temperature (Chylek et al., 2014), and for rainfall (Knight et al., 2006; Willems, 2013) in southern Europe. On the other hand, we have not found regional oscillations of the Mediterranean Sea, where the influence of possible eastern modes associated with the Mediterranean Sea region forcings would be resolved and characterised.

In our experiment, the number of resampling runs was twenty-five for each model, considered large enough for catching the main model behaviour. So, there are sets of 625 candidate models for each PS and lead time horizon tested within the trial and error procedure, i.e., for each PS-BNN search, adding an increasing number of neurons within the hidden layer ($\mathtt{HL} = 1, \ldots, 25$). This, as it is commented in the previous chapter, is time consuming, but in contrast we get robust and balanced solutions and therefore relevant conclusions. Other works however, perform 500 simulations (Kasiviswanathan and Sudheer, 2013), 100 random sampling cases (Perrin et al., 2007), or 100 times bootstrapped (Shortridge et al., 2016). While Anctil and Rat (2005) try with only 30 repetitions. There are some interesting research questions for future research that can be derived from the data partition ratio. In this experiment the 80:20 data partition ratio was established. Without a sensitivity analysis for different partition ratios, it may not be apparent its relevance on the forecasting regression experience. This is desirable for future work.

On computational techniques to establish the functional relationship, the use of BNN was considered convenient and satisfactory for the rainfall forecasting experience on a monthly basis, due to the complex and chaotic nature of this type of hydro-meteorological pattern, and the relational potentiality of this DD computational method observed in previous studies under semiarid and Mediterranean climatological conditions (Aksoy and Dahamsheh, 2009; Moustris et al., 2011). Our findings are directly in line with these previous works, and similar behaviours in terms of efficiency as found in this study (Tab.5.6, Tab.5.7 and Tab.5.8) can be noted. Although, the dilemma *lin* vs *Nlin* (Sec.3.2.4) for the forecasting experience at this time scale was not established. Recent studies like Baker et al. (2018) show us successful applications with purely linear methods such as Multiple Linear Regression (MLR). Meanwhile, in previous works (Mekanik et al., 2013) it was shown how the errors during the validation periods for ANN models are generally lower

compared to MLR, for rainfall forecasting, where its application is recommended within the comparative analysis framework as a general rule.

Throughout this work, the comparison of the developed models with the naive persist method is evaluated with the PI metric (Eq.4.3). It is obvious from Tab.5.6, Tab.5.7 and Tab.5.8 that the models obtained performs better than the persist method (PI>0) for all lead time forecast and locations. Our findings on the rainfall forecasting at least hint that the rainfall does not follow a linear forecasting benchmark. This may be the reason why we did not find a lineal deterioration of the performance metrics with a linear increase of the forecasting time horizon (h). The conclusions of this regressive experience are exposed in further Sec.5.5

5.4 Droughts forecasting

Drought forecasting is very relevant in Mediterranean environments, on both frequency and intensity, as this natural hazard appears recurrently. The Mediterranean water systems present a marked deficit among the hydrological inputs (i.e., Rainfall) and outputs (i.e., Evapo-transpiration). We represented it in Fig.2.12 for data collected in the Béznar Reservoir. We can quantify the water deficit conditions in different ways: by pluvial anomalies or by more standardised Drought Indexes (DI), discussed earlier in Sec.2.4.3. Locally, the local water authorities and managers use these representations in their drought plans (PES), through the Standardized Drought-Precipitation Index (IESP) (Pita López, 2001; López, 2001), being a well-established form to represent these extreme events in Mediterranean climatic conditions.

This natural and recurrent hazard can be modelled and forecasted to anticipate these events and their intensity. The most popular manner is to characterise droughts using hydro-meteorological variables, e.g. rainfall, streamflow, evapotranspiration, etc. Once statistically characterised, they are expected to follow a normal distribution. For our study, the Standardised Precipitation Index (SPI, McKee et al. (1993a)) and the Standardised Runoff Index (SRI, Shraddhanand and Wood (2008)) are the drought indexes used. The positive phase of the index is classified as surplus anomaly conditions (wetness), and the negative phase as deficit anomaly conditions (droughtness), see Tab.2.3. We base them on the cumulative probability of a rainfall/streamflow event occurring at a certain location fitted to a probability distribution, where several candidate distributions are available (Sec.2.4.3). These studies are based on a long-term record of measurements, and they require at least 30 years of records. In this sense, the longer the length of the record, the more reliable and representative the SPI/SRI index will be (Wu et al., 2005). One strength of these indexes is that it makes it possible to describe drought on multiple time scales (Belayneh et al., 2014).

Previous works (Rezaeian-Zadeh and Tabari, 2012; Deo and Şahin, 2015) show the modelling suitability of the SPI for forecasting experiences purposes, even though it is developed under different hydro-meteorological conditions than in our experiment. In similar hydro-meteorological conditions, Djerbouai and Souag-Gamane (2016) investigate the accuracy of different models, using DB models, to forecast drought conditions using SPI in North Algeria. Cancelliere et al. (2007) carry out a more grey box modelling approach, i.e., probabilistic, through the application of a Markov chain model for SPI forecasting, revealing the inadequacy of such an approach in Sicily. Recent works with DD methods

(Le et al., 2017), shows how to generate drought projections based on an alternative DI, through Recurrent Neural Networks in California.

In this section, the main purpose is to check the drought estimating capabilities of the Data Based model methods in our data conditions. For that, we checked several SPI versions to characterise droughts from monthly averaged rainfall values of the Mediterranean River Basin District (`Med.RBD`). The same is applied for rainfall measures collected at two independent meteorological stations nearby in the Guadalfeo River Basin (`GF`): Lánjaron (`La`.GF) and Órgiva (`Or`.GF). On the other hand, we test several SRI to characterise droughts from monthly averaged streamflow amounts collected at Narila gauge station (`Nar`). Later, we tested the resulting SPI and SRI time series through a normality test. In this way, we bring a forecasting regression experience based on anomalies indexes to look at the suitability of drought forecasting development tools for water management purposes in the areas under study.

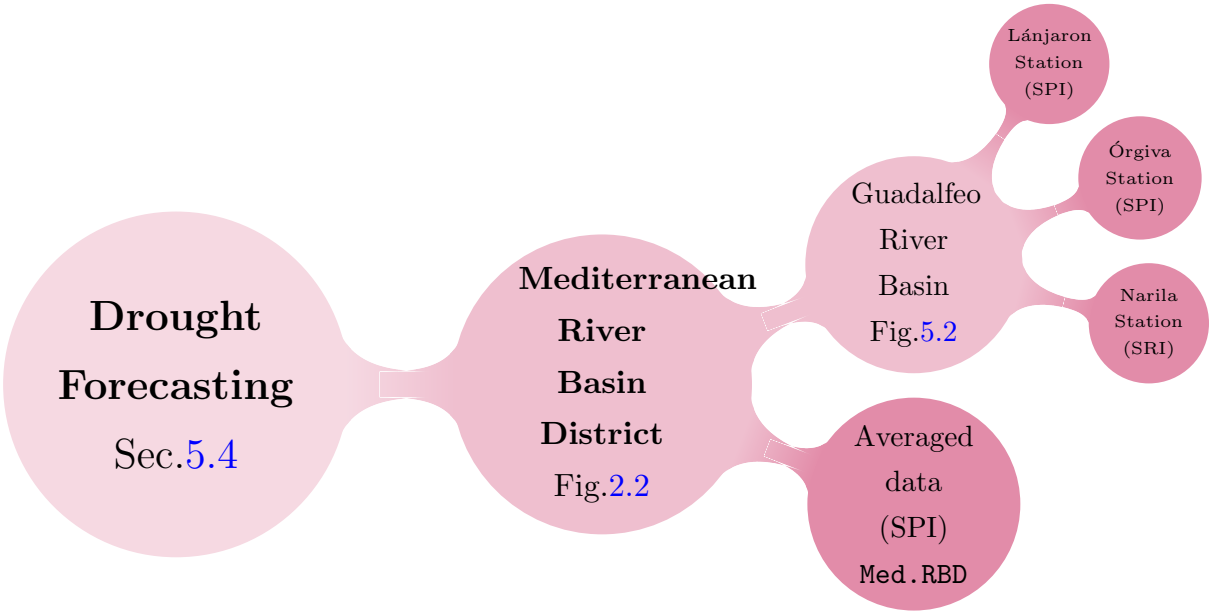


FIGURE 5.10: Drought forecasting cases

5.4.1 Data & Methodology

In this exercise, we tested monthly rainfall time series for drought characterisation suitability using two Drought Indexes (DIs): 1). Standardised Precipitation Index (SPI) and, 2). Standardized Runoff Index (SRI). Then, the suitable indicators that past the normality goodness of fit test are used for forecasting purposes modelling. The flowchart

illustrating this development is shown in Fig.5.11. Here, we realise the examination using the rainfall (R) datasets presented in earlier Sec.5.3.1:

`Med`.RBD We work on the monthly rainfall amounts of the Mediterranean RBD (Fig.2.2) as explained in Sec.5.3.1. In complement, we analyse the drought indicator presented by the water authorities for this case investigation. The statistical nature of this anomaly index differs from the DIs used here (see in Sec.2.4.3).

Guadalfeo RB We use the monthly rainfall values collected in the Lánjaron and Órgiva stations (Sec.5.3.1), and the monthly streamflow values observed at the Narila gauge station (Fig.5.2).

As predictors, we analyse the Atmospheric Oscillations (AO) suitability as a potential explanatory variable. For this reason, it is opportune to test different AOs as potential observations for DIs forecasting. The AOs used are: NAO, WEMO, MOI-GI, and MOI-GA and its combination of pairs (more in Sec.5.3.1). Also, four variables are modelled with the WiMMed hydrological model: total Precipitation (R_{Wi}, $mm \cdot month^{-1}$), Snowfall (R_S, $mm \cdot month^{-1}$), the effective rainfall (liquid fraction, R_e, $mm \cdot month^{-1}$) and Snowmelt (S, $mm \cdot month^{-1}$) promediated in the whole contributing area. All these datasets are adopted for drought characterisation suitability and employed in a forecasting analysis.

For wet and dry periods characterisation through DIs, the data quality should be superb: no data gaps and long enough to draw robust conclusions. Some authors established a minimum dataset length of 30 years. The `Med`.RBD timeserie has rainfall data from 1940 to 2016. We have three gauge stations within the river basin (Fig.5.2), in Lánjaron (`La`.GF, from 1946 to 2008), in Órgiva (`Or`.GF, from 1961 to 2008), and in Narila (from 1971 to 2010), with ≈60, ≈41 and ≈39 years length of record, respectively. Among the time series explored within this river basin, the selected ones ensured the higher quality on collected data: fewer gaps and longer historical registry for a proper DIs characterisation.

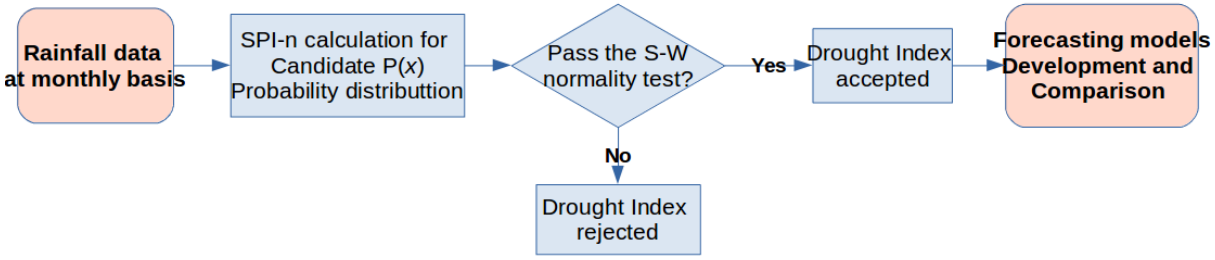


FIGURE 5.11: Flowchart showing the development of the drought indexes forecasting problem carried out in this section

We expect to assess water deficit or surplus at different hydrological levels by using various time scales. Short scales (i.e., 1 to 3 months) give us the information at a more superficial level, such as the humidity of the soil. Also, has a practical use for plants stress assessment. At intermediate scales (i.e., 6 to 9 months) we could infer the state of rivers or reservoirs, the hydrological state of the water system on a medium time scale. From higher scales (i.e., 12 to 24 months), we can infer the overall hydrological status of the basin, and even an idea of the state of water bodies with slower recharge times, such as the aquifers.

For that, we propose different time-scales (n) for medium term forecasting, which are based on five accumulation periods: two months, quarterly, semi-annual, annual basis and biannual basis (n = 2,3,6,12 or 24 months). These accumulation periods, are the window lengths of an backward looking running mean, similar to the Moving Average (MA(q)), where the order q would be equal to the scale n. Therefore, in this case study, unlike the previous one (Sec.5.3), it is not considered convenient to include the MA(q). And it is removed from the PS building process as a candidate input variable. In this way, we remove redundancy and perceptually irrelevant information while preserving parsimony, since it is already implicitly included during the DI calculation.

The SPI and SRIn are computed given a rainfall or a streamflow probability distribution $P(x)$, as described previously in Sec.2.4.3. Following other authors (Stagge et al., 2015), we test seven candidate distributions: the Normal $\mathbf{N}(y)$, the Gamma $\mathbf{G}(y)$, the Generalized extreme value $\mathbf{GEV}(y)$, the Pearson type 3 $\mathbf{P3}(y)$, the Weibull $\mathbf{W}(y)$, the Generalised logistic $\mathbf{GL}(y)$ distributions, and the Gumbel $\mathbf{Gu}(y)$ distribution. For each SPI/SRIn generated, we question the goodness of fit with the Shapiro-Wilk (S-W) normality test for a significance level of 5% (p-value ≥ 0.05). If the condition passed, the sample data set is not significantly different from normal distributed. In a numerical manner, we confirm the possibility that it follows a Normal distribution. If not, there will be evidence to reject the null hypothesis, and so we will discard it as a valid drought index by definition. Once, the accepted SPI/SRIn series are identified, a forecasting regression experience for only one DI per n scale is developed (Fig.5.11). For simplifying purposes, we select only one DI for each n time scale for the modelling phase. If there is more than one, the highest p-value ($max(p)$) will be selected for the posterior forecasting action at several time horizons. We will list the wining DIs in their corresponding tables.

First, we asses the competing Predictors Structure (PS) during the Input Variable Selection (IVS) procedure. We take this step under a wrapper perspective taking into account the parsimony. The first one is just an univariate autoregressive BNN model, where

antecedent values of the SPI/SRI-n has been used to forecast the h step ahead index (PS1).

$$\text{SPI/SRI}n_{(t+h)} = f_{\text{BNN}}(\text{PS1}) = f_{\text{BNN}}(\{\text{DI}n_{(t-p)}\}_{p=1}^{p_{\text{DI}}}) \tag{5.13}$$

From here, it differs depending on the DI used:

- For the SPI cases (`Med`.RBD, `La`.GF and `Or`.GF in Fig.5.2), we test a second version: PS1→PS2. That contains p_{AO} antecedent values of one of the AO (N, W, M1, M2) or their combination of pairs as an exogenous input.

$$\text{SPI}n_{(t+h)} = f_{\text{BNN}}(\text{PS2}) = f_{\text{BNN}}(\{\text{SPI}n_{(t-p)}\}_{p=1}^{p_{\text{DI}}}; \{\text{AO}_{o(t-p)}\}_{p=1}^{p_{\text{AO}}}) \tag{5.14}$$

- In the SRI case, for the streamflow collected at Narila (Fig.5.2), we use the following series obtained with hydrological modelling as candidate input variables (PS1→PS2A): the total precipitation (R_{Wi}), the snowfall (R_S), the snowmelt (S) and the effective rainfall (R_e). We reproduced these time series for the entire contributing sub-basin area, and the same period of study (Eq.5.15). At the same time, we also use the AO in an alternative PS building up attempt (PS1→PS2B), to test additional relevant observations to the model. Again, we tested the antecedent values of AO (N, W, M1, M2) and the combination of pairs among them as an exogenous input.

$$\text{SRI}n_{(t+h)} = \begin{cases} f_{\text{BNN}}(\text{PS2A}) = f_{\text{BNN}}(\{\text{SRI}n_{(t-p)}\}_{p=1}^{p_{\text{DI}}}; \{X_{d(t-p)}\}_{p=1}^{p_{X_d}}) \\ f_{\text{BNN}}(\text{PS2B}) = f_{\text{BNN}}(\{\text{SRI}n_{(t-p)}\}_{p=1}^{p_{\text{DI}}}; \{\text{AO}_{o(t-p)}\}_{p=1}^{p_{\text{AO}}}) \end{cases} \tag{5.15}$$

On the DD modelling matters, the implemented method steps to complete the modelling action are essentially the same as the preceding exercise (Rainfall forecasting in Sec.5.3.1). The only distinction is the target variable to forecast. Here, the target will be the SPI/SRIn to characterise wet and dry periods on pluviometric and hydrological terms. The results shown are the best bootstrapped model-member from the 225 runs per each configuration tested. The accompanying performance metrics are: Correlation coefficient (R^2), RMSE, PI and MAE to test their deterministic behaviour, BIC to quantify parsimony during the model development. The POC and AW coefficients are also used to evaluate the models uncertainty estimations. The Bayesian technique of BNNs produces an error bar for each estimation, which is significant at 95% confidence interval. Again, the closer the coverage rates (POC) obtained with the theoretical assumption, the

better successful the experience will be. On the other hand, lower AW values means more certain estimations and therefore better models. We can find more on these metrics in Sec.3.4.1 and Sec.3.4.2 for deterministic and the uncertainty estimations respectively. In the coming paragraphs, we present the proceeds of the goodness-of-fit, the IVS outcomes and the forecasting performance.

5.4.2 Results

• SPI forecasting experience on the Mediterranean River Basin District •

Goodness-of-fit We compute SPIn with the different candidate probability distributions for the monthly rainfall data provided for the `Med.RBD` case. The candidate index is tested with the Shapiro-Wilk normality test (S-W) to check that it is normally distributed. Tab.5.9 shows the results of the candidate distributions at five different timescales (n= $2, 3, 6, 12, 24$ months), producing 35 SPI time series in total. We can see that only seven SPI time series passed the W-S test. The suitable probability distributions ($P(x)$), that passed the test, are Gamma **G**, for SPI2 and SPI3 time scales, the Generalized extreme value **GEV**, for SPI3 and SPI6 time scales, the Pearson type 3 **P3**, for SPI3, and the Gumbel **Gu** distribution for SPI3 and SPI6. Finally, the W fit of SPI3 & SPI6 in the `Lan` case, and the P3 fit of SPI2 & SPI6 in the `Org` case, are the four selected indexes for further forecasting. It is worth mentioning that the DI proposed by the Andalusian environmental authorities to characterise droughts (IESP, Sec.2.4.3) didn't pass the S-W test, getting a p-value of $1.18 \cdot 10^{-20}$.

TABLE 5.9: Shapiro-Wilk normality test results for the candidate density probability distributions to fit SPIn for the Med.RBD case. Among brackets only the p-values ≥ 0.05

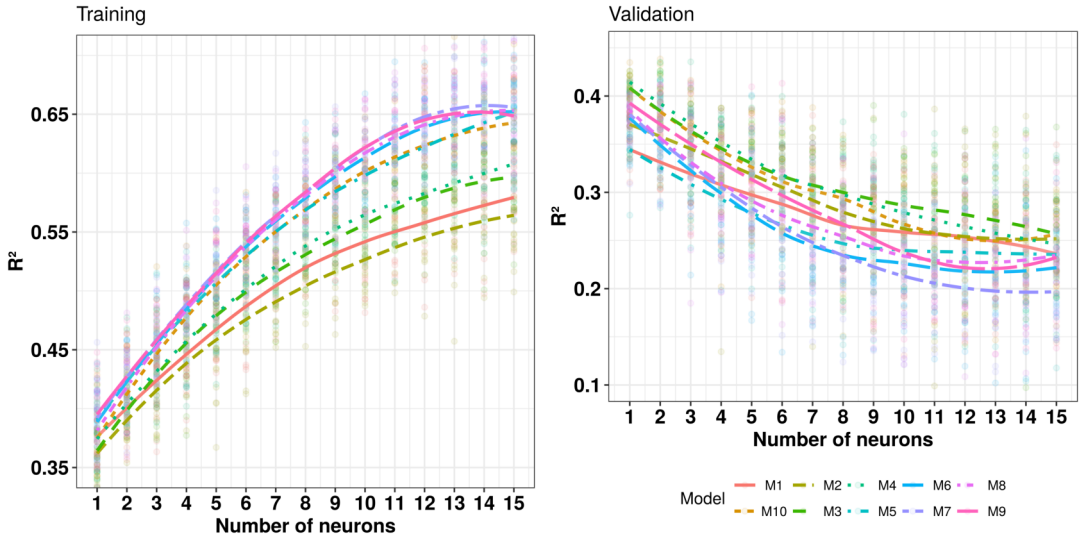
	Mediterranean RBD				
$P(x)$	**SPI2**	**SPI3**	**SPI6**	SPI12	SPI24
N(y)	×	×	×	×	×
G(y)	✓[**0.62**]	✓[**0.84**]	×	×	×
GEV(y)	×	✓[0.18]	✓[**0.42**]	×	×
P3(y)	×	✓[0.61]	×	×	×
W(y)	×	×	×	×	×
GL(y)	×	×	×	×	×
Gu(y)	×	✓[0.32]	✓[0.12]	×	×

TABLE 5.10: Best model performance metrics for the relevant PS evaluated at each SPIn case for $h = 0$ time ahead step horizon at the Mediterranean RBD dataset. Training and Validation values separated by a semicolon (T;V). When the $\mathrm{RI}_i\tau > 0.025$ condition is complied, is highlighted with bold font. HL: Number of hidden units within the BNN structure. AO_o: Atmospheric Oscillation. Model expressions has the $P(x)$-SPIn form

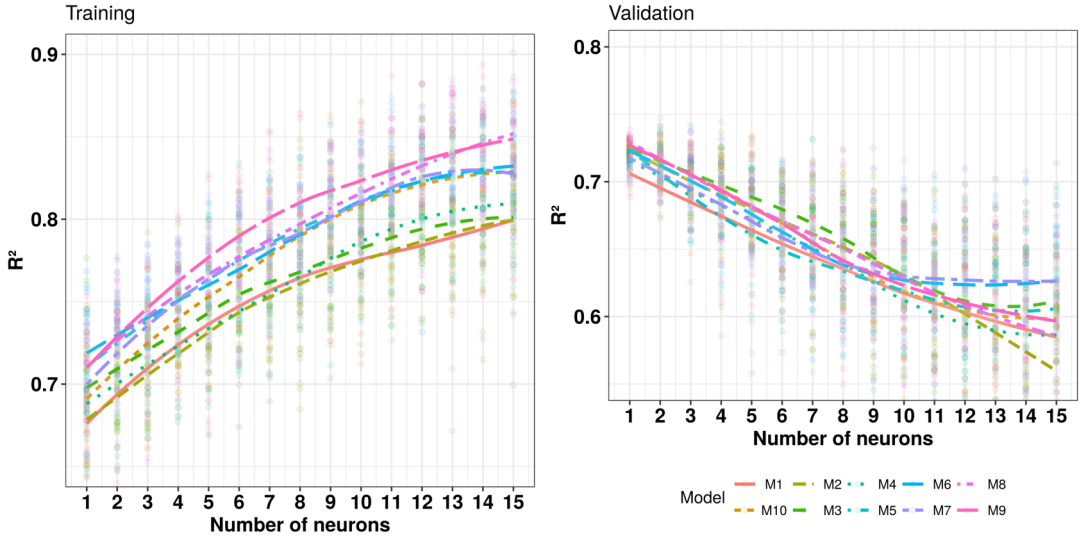
Model	PS	$\mathbf{R}^2$	PI	RMSE	BIC (HL)	AO_o
G-SPI2	**PS1**	0.353;0.447	0.49;0.37	0.490;0.722	-400.7(4)	-
	$\mathrm{RI}_{i(\mathrm{PS1}\rightarrow\mathrm{PS2})}$	+6.2%;2.7%	+8.2%;±0%	+5.3%;+0.1%	**+10.2%**	·
	PS2	0.375;0.435	0.53;0.37	0.464;0.721	-441.5(2)	$\mathrm{AO}_{\mathbf{M3}}$
G-SPI3	PS1	0.658;0.725	0.23;0.11	0.534;0.547	-365.9(2)	-
	$\mathrm{RI}_{i(\mathrm{PS1}\rightarrow\mathrm{PS2})}$	+7.3%;+2.2%	+60.9%;**+45.4%**	-9.7%;**+2.9%**	-19.7%	·
	PS2	0.706;0.741	0.37;0.16	0.586;0.531	-293.6(2)	$\mathrm{AO}_{\mathbf{M3}}$
GEV-SPI6	PS1	0.610;0.668	0.27;0.12	0.553;0.586	-349.5(2)	-
	$\mathrm{RI}_{i(\mathrm{PS1}\rightarrow\mathrm{PS2})}$	+2.6%;**+4.6%**	+33.3%;**+58.3%**	+2.9%;**+4.09%**	**+2.6%**	-
	PS2	0.626;0.699	0.36;0.19	0.537;0.562	-358.5(2)	$\mathrm{AO}_{\mathbf{M3}}$

Input Variable Selection In Tab.5.10 we present the results of the IVS procedure. The analysis of the autocorrelation through ACF and PACF calculation suggests a substantial correlation at 95% confidence level up to 3 and 4 measurement lag. At the beginning, the performance for the three SPIn are homogeneous, where the AOs enhance SPI modelling substantially for the SPI3 and SPI6 cases. The M3(MOI-GI) observations are the most strong AO for pluviometric deviations in this area. The combination of AOs pairs as inputs, can deteriorate the overall operation, being the incorporation of a single input with its corresponding lags the best form to apply it in this case. The behaviour of all the simulations can be seen in Fig.5.12. In this type of figures we visualise different assumptions trough the performance curves where we can appreciate minor differences among competing predictors and also is useful for optimal neural architecture search. For instance, Fig.5.12.(c) shows the M3 model outperforms others AO clearly. The model progress with the AO inclusion within PS is evident, where we even detect increases in all performance metrics, e.g. GEV-SPI6. Subsequently, for further forecasting actions, two PS2 and one PS1 are the most suitable predictors to forecast.

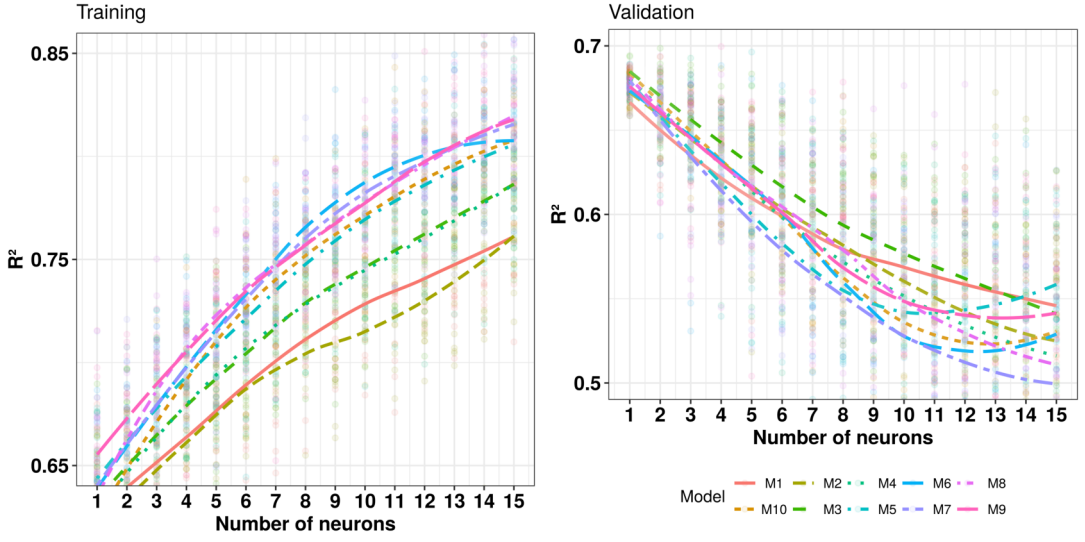
Forecasting In Tab.5.11 we display the results of the finest BNN model for the validation stage for three distinctive time horizons ($h = 0, 1, 2$). We can see how homogeneous



(a) G2-SPI-PS2A



(b) G-SPI3-PS2A



(c) GEV-SPI6-PS2A

FIGURE 5.12: Mediterranean RBD SPI learning trends to test different AOs and their combination of pairs in a BNN model-based approach experiment

the deterministic results are, existing a notorious linear decline, ending in unsatisfactory results for the horizons $h = 1, 2$. The deterministic value of SPI2 offers good values for $h = 0$ in terms of predictability, i.e., PI. But if we look at the uncertainty values we can see how poor quality they are. It is far from the confidence hypothesis of 95%, and the error bar width (AW) is very large. For the other models the deterministic behaviours is worse, in terms of predictability. The uncertainty estimations for these models are successful: POC $\approx$ 95% and the AW values are low, which are good for pragmatic purposes. The SPI3 and SPI6 models comply the DD method confidence intervals hypothesis used, as the values obtained during the validation are close to the theoretician values. Focusing on the time horizons performance in the whole problem, we can see how it breaks down as we step up the time horizon for both, on deterministic and uncertainty estimations.

TABLE 5.11: Performance metrics of the Drought indexes regression forecasting results carried out at the Mediterranean RBD. Training and Validation values separated by a semicolon (T;V). Model$\rightarrow$ $P(x)$-SPIn-PS. Time horizons: $h = 0 \rightarrow (t)$, $h = 1 \rightarrow (t+1)$, $h = 2 \rightarrow (t + 2)$

Model	h	R^2	RMSE	PI	MAE	POC(%)	AW
	0	0.353;0.447	0.490;0.722	0.49;0.37	0.389;0.548	98.2;99.3	1.9;2.0
G-SPI2-PS1	1	0.154;0.095	0.349;0.926	0.58;0.48	0.272;0.751	99.7;100	2.9;2.9
	2	0.111;0.078	0.354;0.934	0.55;0.52	0.279;0.775	99.2;100	2.6;2.6
	0	0.706;0.741	0.586;0.531	0.37;0.16	0.466;0.391	95.3;94.9	1.0;1.2
G-SPI3-PS2	1	0.463;0.484	0.525;0.749	0.36;0.22	0.386;0.564	98.7;98.7	2.1;2.2
	2	0.382;0.289	0.526;0.878	0.43;0.29	0.408;0.675	99.0;97.5	2.5;2.5
	0	0.626;0.699	0.537;0.562	0.36;0.19	0.415;0.419	93.1;95.6	1.3;1.3
GEV-SPI6-PS2	1	0.330;0.415	0.438;0.781	0.36;0.22	0.344;0.610	95.7;96.2	1.6;1.7
	2	0.242;0.230	0.484;0.893	0.43;0.33	0.386;0.700	92.9;94.3	1.9;1.9

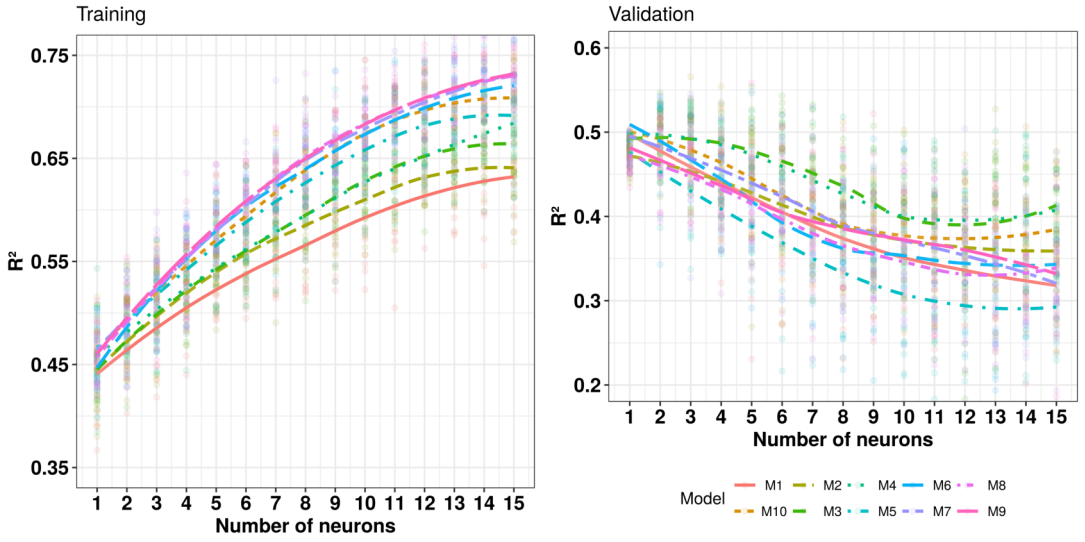
• SPI forecasting experience within the Guadalfeo River Basin •

TABLE 5.12: Shapiro-Wilk normality test results for the candidate density probability distributions to fit SPIn for locations in the Guadalfeo River Basin: `Lan`.GF and `Org`.GF. Among brackets only the p-values ≥ 0.05

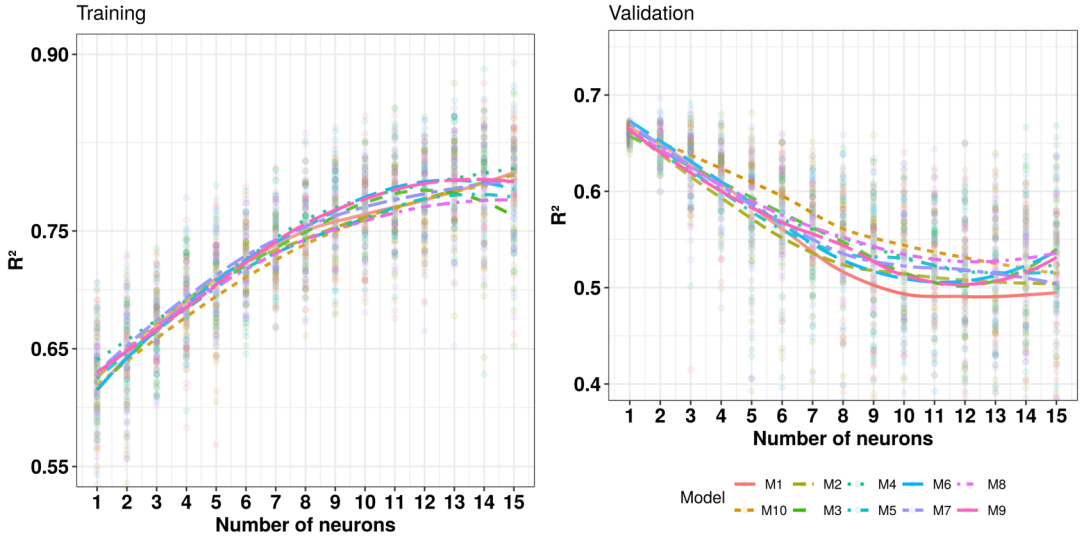
	Lanjaron					Orgiva				
P(x)	SPI2	**SPI3**	**SPI6**	SPI12	SPI24	**SPI2**	SPI3	**SPI6**	SPI12	SPI24
N(y)	×	×	×	×	×	×	×	×	×	×
G(y)	×	×	×	×	×	×	×	×	×	×
GEV(y)	×	×	×	×	×	×	×	×	×	×
P3(y)	×	×	✓[0.08]	×	×	✓**[0.08]**	×	×	×	×
W(y)	×	✓**[0.24]**	✓**[0.16]**	×	×	×	×	×	×	×
GL(y)	×	×	×	×	×	×	×	✓**[0.21]**	×	×
Gu(y)	×	×	×	×	×	×	×	×	×	×

Goodness-of-fit Here, we tested 70 SPI potential time series under the S-W normality test (Tab.5.12) as the candidate DIs in the GF RB case: Lánjaron and Órgiva stations. Only five candidates SPI (in bold) passed it (p-value≥ 0.05). The Pearson type 3 (**P3**) distribution passed for both datasets, the other distributions that passed the test are, the Weibull (**W**) and the Generalised Logistic (**GL**) for Lánjaron and Órgiva respectively. On the n value, it should be commented that long-term scales (e.g., $n = 12, 24$) has not passed the S-W normality test. But nevertheless, short- and medium-term values ($n = 2$ once, $n = 3$ once, and $n = 6$ three times) passed the test. Finally, the W distribution fit of SPI3 and SPI6 in the `Lan` case, and in the `Org` case the SPI2 & SPI6 fitted series with the P3 distribution, are the four selected indexes for further development.

Input Variable Selection In Tab.5.13, we present the rises of the IVS procedure for this study. Our results show how the SPI with lower n values (i.e., $n = 2, 3$) has higher predominant autoregressive nature, as the PCAF correlation values are significant to *lag 3*. The ACF and PCAF values are similar for `Or`.GF case. Then, we use $p_{\text{DI}} = 3$ order value for both cases for PS1, analysing two competing PS: PS1 and PS2. For the `La`.GF case, the best AO input is M3(MOI-GI) and M10(MOI-GI & MOI-CA). This can be appreciated in Fig.5.13. These results are in-line with the preceding case (`Med`.RBD). For the `Or`.GF case, turns out to be the M1(NAO) observations as the strongest AO. Decisively during this procedure (IVS), the models that have surpassed RIτ are three:



(a) G3-PS2A



(b) G6-PS2A

FIGURE 5.13: Lánjaron (La.GF) SPI learning trends to test different atmospheric oscillations and their combination of pairs in a BNN model-based approach experiment

(La.GF-W-SPI3, La.GF-W-SPI6 and Or.GF-P3-SPI2), staying with a PS1 for the Or.GF-GL-SPI6 case.

Forecasting In Tab.5.14 we list the best single model found for each SPIn at three time horizons: $t+1$, $t+2$ and $t+3$. For the La.GF-SPI3-PS2 case, the behaviour of the model has been acceptable for both, the deterministic and the uncertainty part. These models work well for the first two time horizons $h = 0, 1$ but for $h = 2$ breaks. On the other hand, the model La.GF-SPI6-PS2 leaves a decent behaviour for all time boundaries

TABLE 5.13: Best model performance metrics for the relevant PS evaluated at each SPIn case for $h = 0$ time ahead step horizon. Training and Validation values separated by a semicolon (T;V). HL: Number of hidden units within the BNN structure. Model expressions has the Site-$P(x)$-SPIn form

Model	**PS**	**R^2**	**PI**	**RMSE**	**BIC** (HL)	X_d
Lan.GF-W-SPI3	PS1	0.529;0.559	0.39;0.37	0.558;0.672	-262.4(5)	-
	$RI_{i(PS1 \rightarrow PS2)}$	-0.9%;-5.5%	+35.9%;**+2.7%**	+6.9%;+0.6%	**+17.5%**	·
	PS2	0.524;0.566	0.53;0.38	0.519;0.668	-308.5(4)	$AO_{\mathbf{M3}}$
Lan.GF-W-SPI6	PS1	0.660;0.650	0.35;0.12	0.565;0.621	-286.5(2)	-
	$RI_{i(PS1 \rightarrow PS2)}$	+7.4%;**+2.5%**	+31.4%;**+14.3%**	-2.5%;-2.6%	-13.5%	·
	PS2	0.709;0.669	0.46;0.16	0.579;0.605	-247.9(4)	$AO_{\mathbf{M10}}$
Or.GF-P3-SPI2	PS1	0.272;0.331	0.51;0.28	0.473;0.745	-240.9(5)	-
	$RI_{i(PS1 \rightarrow PS2)}$	+44.5%;**+5.1%**	+13.7%;**+14.3%**	+10.4%;**+2.9%**	**+13.1%**	·
	PS2	0.393;0.348	0.58;0.32	0.424;0.723	-272.6(6)	$AO_{\mathbf{M1}}$
Or.GF-GL-SPI6	**PS1**	0.610;0.739	0.41;0.12	0.516;0.504	-252.9(3)	-
	$RI_{i(PS1 \rightarrow PS2)}$	+7.0%;+0.3%	+14.6%;**+16.6%**	0%;+1.1%	-7.2%	-
	PS2	0.653;0.741	0.47;0.14	0.516;0.498	-234.8(4)	$AO_{\mathbf{M1}}$

in uncertainty terms, but not in deterministic. This is because of its moderate precision values and low predictability capacities (PI$<$ 0.5). For the Or.GF cases, the results are poor and the designs fail. We found that the deterministic performance works acceptable in efficiency terms and fails in predictability terms. On the uncertainty concerns, this case showed unsatisfactory results. This is in-line with the past evidence of Rainfall forecasting for the same dataset, that showed unsatisfactory results too.

TABLE 5.14: Performance metrics of the Drought indexes regression forecasting results carried out at the Guadalfeo River Basin. Training and Validation values separated by a semicolon (T;V). The Model has the `Site`-$P(x)$-SPIn-PS form. Time horizons from Eq.5.13: $h = 0 \rightarrow (t)$, $h = 1 \rightarrow (t+1)$, $h = 2 \rightarrow (t+2)$

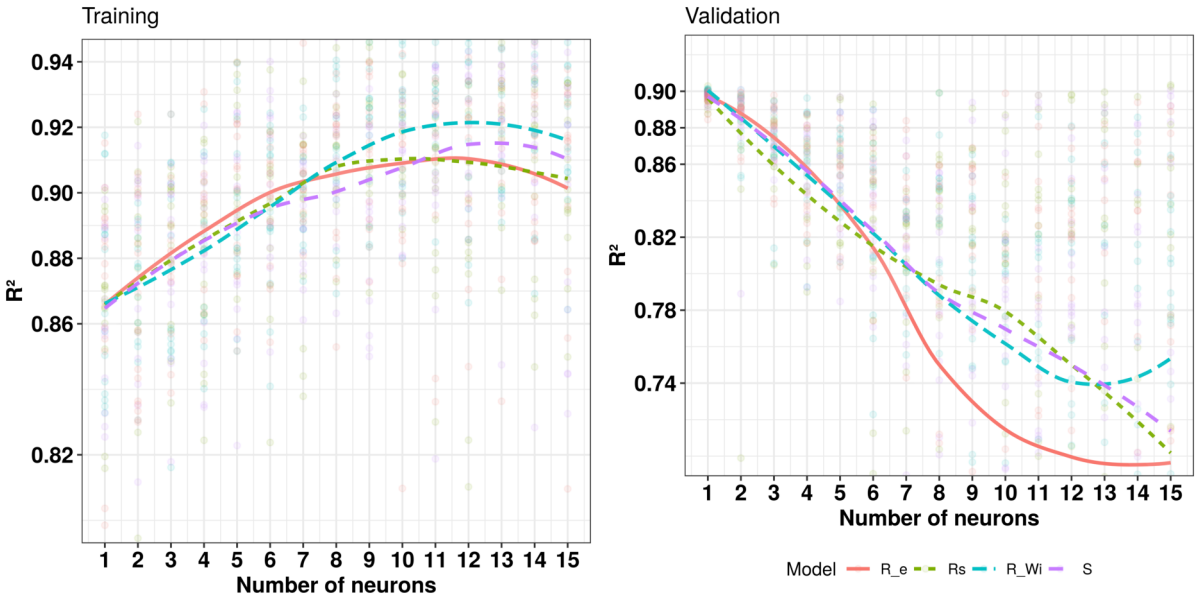
Model	h	R^2	RMSE	PI	MAE	POC(%)	AW
	0	0.524;0.566	0.519;0.668	0.53;0.38	0.401;0.524	97.3;97.1	1.4;1.5
La.GF-W-SPI3-PS2	1	0.233;0.285	0.440;0.878	0.52;0.50	0.344;0.708	94.1;95.0	1.7;1.7
	2	0.021;0.019	0.328;1.021	0.52;0.53	0.268;0.843	99.1;99.3	2.3;2.4
	0	0.709;0.669	0.579;0.605	0.46;0.16	0.467;0.453	96.9;95.0	1.2;1.3
La.GF-W-SPI6-PS2	1	0.458;0.345	0.478;0.854	0.42;0.16	0.371;0.648	92.3;88.5	1.3;1.3
	2	0.256;0.227	0.429;0.926	0.38;0.26	0.327;0.730	98.0;95.7	1.8;1.9
	0	0.393;0.348	0.424;0.723	0.58;0.32	0.324;0.585	99.7;100	2.6;2.4
Or.GF-P3-SPI2-PS2	1	0.041;0.069	0.524;0.885	0.54;0.42	0.433;0.737	100;100	3.8;3.7
	2	0.040;0.001	0.529;0.929	0.52;0.41	0.437;0.784	100;100	3.6;3.6
	0	0.610;0.739	0.516;0.504	0.41;0.12	0.399;0.389	99.1;100	1.8;1.6
Or.GF-GL-SPI6-PS1	1	0.353;0.484	0.497;0.720	0.50;0.14	0.389;0.575	100;100	2.6;2.3
	2	0.227;0.309	0.373;0.822	0.44;0.23	0.299;0.654	100;100	2.7;2.6

• SRI forecasting experience within the Guadalfeo River Basin •

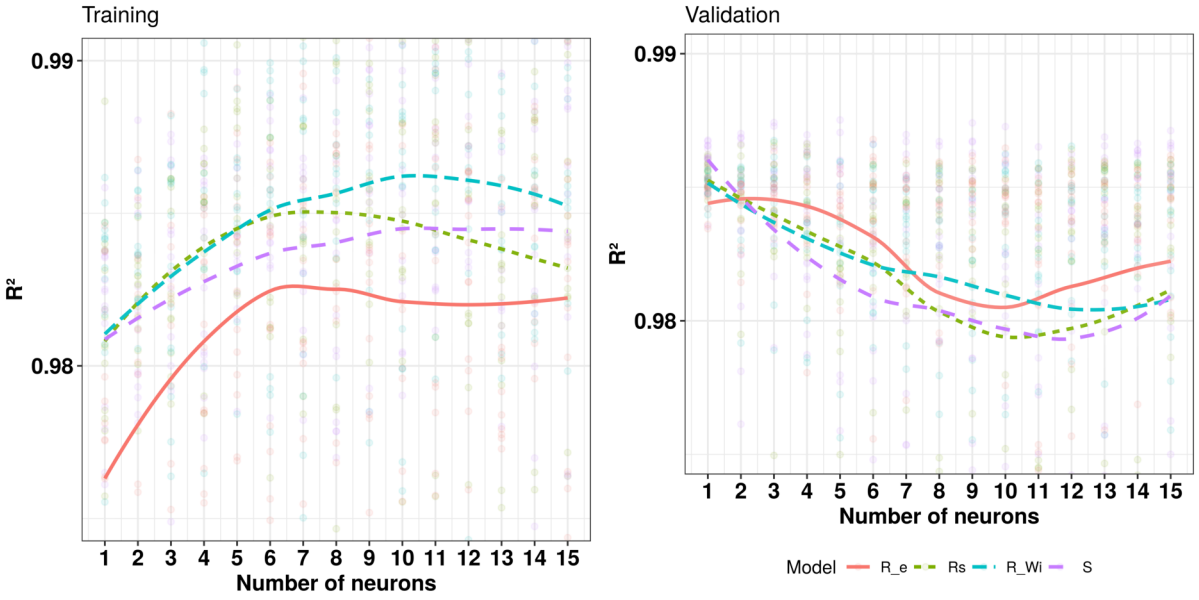
Goodness-of-fit In this new case we used the streamflow observations (SRI) at Narila (Fig.5.2), instead of the rainfall observations (SPI) of the three previous cases. In this case, we tested the goodness-of-fit of 35 SRI candidate indexes with the S-W normality test (Tab.5.15), where only five candidate SRI have passed it (p-value$\geq$ 0.05). Here, the distributions obtained are **GEV** (SRI6 and SRI12) and **GL** (SRI2, SRI3, SRI6), for Generalized extreme value and Generalised Logistic distributions, respectively. The SRI passes the S-W test for four time-scales ($n = 2, 3, 6, 12$) leaving uncharacterised the longer time-scale ($n = 24$). We select the SRI2 and SRI3 time-series had with the GL fit, and the SRI6 and SRI12 time-series got with the GEV fit to develop further forecasting experience.

TABLE 5.15: Shapiro-Wilk normality test results for the candidate density probability distributions to fit SRIn for the Narila gauge station within the Guadalfeo River Basin. Among brackets only the p-values ≥ 0.05

	Streamflow at Narila				
$P(x)$	**SRI2**	**SRI3**	**SRI6**	**SRI12**	SRI24
N(y)	×	×	×	×	×
G(y)	×	×	×	×	×
GEV(y)	×	×	✓**[0.15]**	✓**[0.15]**	×
P3(y)	×	×	×	×	×
W(y)	×	×	×	×	×
GL(y)	✓**[0.15]**	✓**[0.08]**	✓[0.09]	×	×
Gu(y)	×	×	×	×	×



(a) GEV-SRI6-PS2A

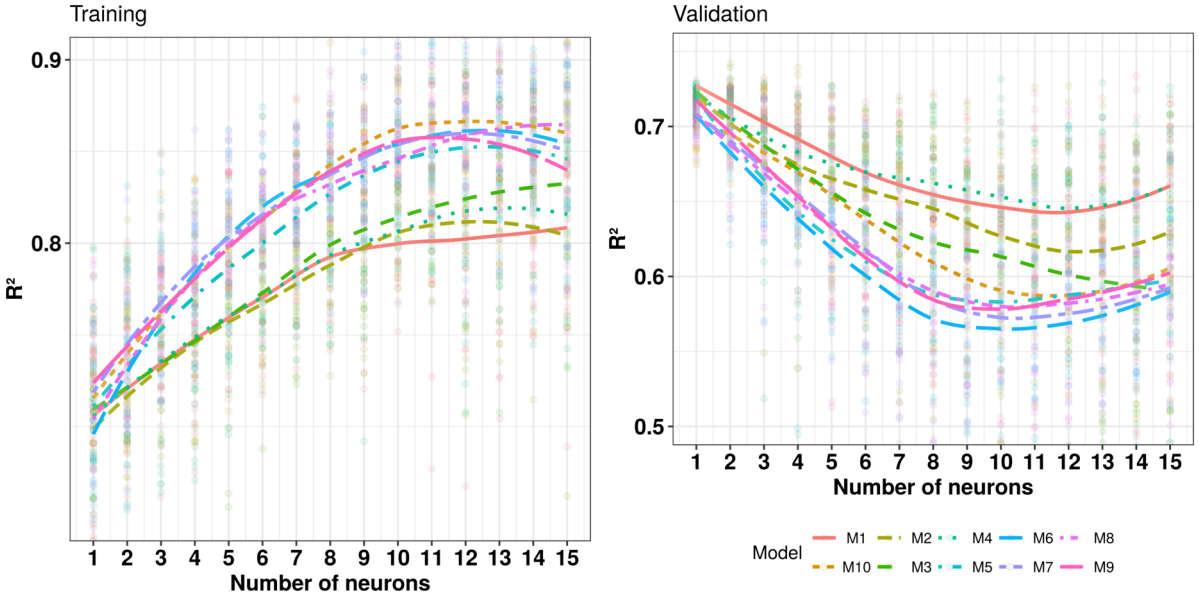


(b) GEV-SRI12-PS2A

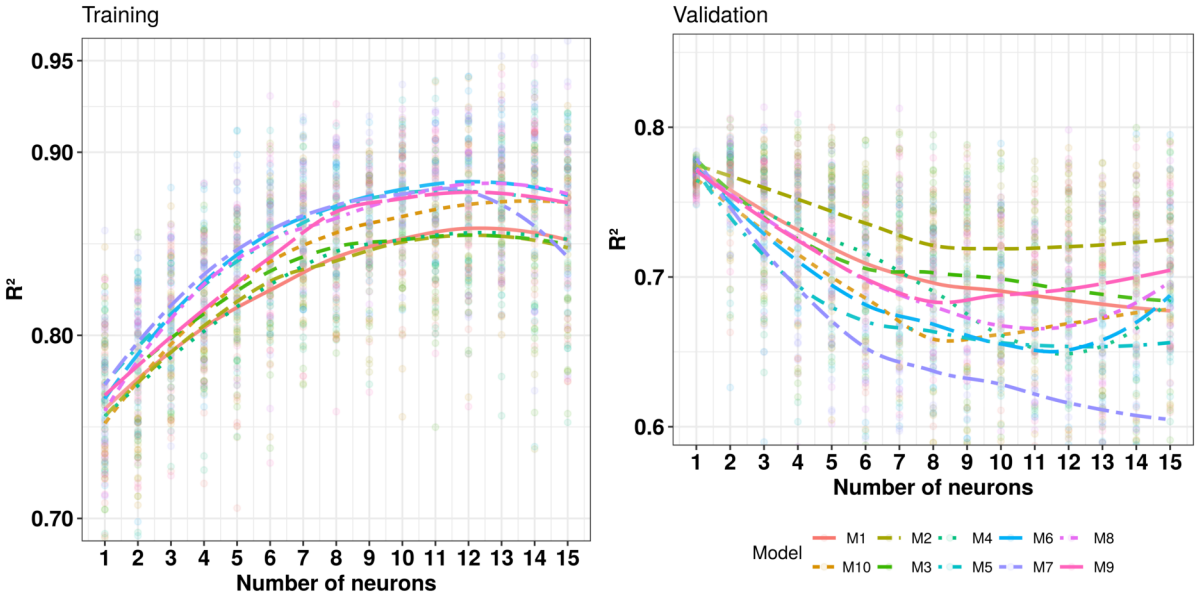
FIGURE 5.14: Narila sub-basin SRI learning trends results to test different hydrological variables in a model-based approach experiment

Input Variable Selection For this examination, Tab.5.16 displays the results of the IVS sequential procedure during the PS definition for the four SRI time-series used. Here, we added supplementary knowledge into PS2 having two possibilities: snow variables reproduced by hydrological modelling (A), and AOs observations (B). The hydrological modelling reports gains for greater time scales ($n = 6, 12$), where the differences between

the candidate variables are minimal. For the SRI6 the ultimate PS is the PS2A, surpassing the PS1 and PS2A behaviours. It has the Snowfall (R_S) reproduced by hydrological modelling as an exogenous inputs. The other case, SRI12, has the Snowmelt (R_S) as the wining predictor, since all maxima are associated with this snowy predictor (Fig.5.14.(b)). On the other hand, we discover how the AOs inclusion reports significant increases at lower time-scales ($n = 2, 3$). In the SRI2 experiment, the M3 single observations and the M10 combination are the best choice among them. Finally, in the SRI3-PS2B, the M8 combination gives the maximum but the best average behaviour is M2, see Fig.5.15. In Fig.5.14 and Fig.5.15 the 225 simulations realized per each case to compare the behaviours of each variable are represented for the PS2A and PS2B respectively.



(a) GL-SRI2-PS2B



(b) GL-SRI3-PS2B

FIGURE 5.15: Narila sub-basin SRI learning trends results to test different AOs and their combination of pairs in a model-based approach experiment

TABLE 5.16: Best model performance metrics for the relevant PS evaluated at each SRI case for $h = 0$ time ahead step horizon. HL: Number of hidden units within the BNN structure. X_d: Explanatory variable. Model expressions has the $P(x)$-SRIn form

Model	PS	R^2	PI	RMSE	BIC (HL)	X_d
GL-SRI2	PS1	0.755;0.741	0.48;0.25	0.435;0.361	-229.0(7)	·
	$RI_{i(PS1\rightarrow PS2A)}$	+2.5%;+0.4%	-4.2%;**+12.0%**	-6.2%;+1.7%	-1.9%	·
	PS2A	0.774;0.744	0.46;0.28	0.462;0.355	-224.5(2)	S
	$RI_{i(PS1\rightarrow PS2B)}$	+4.5;+0.7	-4.2%;**+8.0%**	-2.1;+1.1	**+7.0%**	·
	PS2B	0.789;0.746	0.46;0.27	0.444;0.357	-245.0(3)	$\mathbf{AO_{M1}}$
GL-SRI3	PS1	0.787;0.799	0.49;0.20	0.434;0.330	-229.8(7)	-
	$RI_{i(PS1\rightarrow PS2A)}$	+10.3%;+0.1%	+26.5%;-5%	+27.6%;-0.3%	**+47.9%**	·
	PS2A	0.868;0.800	0.62;0.19	0.314;0.331	-339.9(7)	R_{Wi}
	$RI_{i(PS1\rightarrow PS2B)}$	+0.3%;**+2.5%**	-12.2%;**+35.0%**	+2.7%;**+4.8%**	**+14.9%**	·
	PS2B	0.790;0.819	0.43;0.27	0.422;0.314	-264.2(3)	$\mathbf{AO_{M2}}$
GEV-SRI6	PS1	0.898;0.902	0.66;0.14	0.274;0.247	-390.2(8)	-
	$RI_{i(PS1\rightarrow PS2A)}$	+1.6%;+1.1%	-6.1%;**+71.4%**	-1.5%;**+5.7%**	**+3.3%**	·
	PS2A	0.913;0.912	0.62;0.24	0.278;0.233	-403.2(4)	R_S
	$RI_{i(PS1\rightarrow PS2B)}$	+4.7%;+0.3%	+1.5%;**+35.7%**	+9.8%;+2.4%	**+4.0%**	·
	PS2B	0.940;0.905	0.67;0.19	0.247;0.241	-406.0(10)	AO_{M1}
GEV-SRI12	PS1	0.971;0.986	0.82;0.44	0.113;0.118	-697.7(2)	
	$RI_{i(PS1\rightarrow PS2A)}$	+1.6%;+0.1%	+10.9;**+6.8%**	±0%;**+3.4%**	±0%	
	PS2A	0.987;0.987	0.91;0.47	0.113;0.114	-697.7(2)	S
	$RI_{i(PS1\rightarrow PS2B)}$	+1.3%;+0.1%	+9.8%;+2.3%	-13.3%;±0%	**+3.5%**	·
	PS2B	0.984;0.987	0.90;0.45	0.128;0.118	-722.7(2)	AO_{M1}

Forecasting In Tab.5.17 we present the results of the best single model of four SRI indexes for estimations at three time horizons. In this case we have variety in the specific results of the scale. For the GEV-SRI 12 case, the uncertainty is very low, and coverage values close to 100 %. Here the deterioration is non-linear. These models characterize well, above all on larger time scales.

TABLE 5.17: SRI$-n$ regression forecasting results carried out in this section. Training and Validation values separated by a semicolon (T;V). The Model has the $P(x)$-SPIn form. Time horizons from Eq.5.13: $h = 0 \rightarrow (t)$, $h = 1 \rightarrow (t+1)$, $h = 2 \rightarrow (t+2)$

Model	h	R^2	RMSE	PI	MAE	POC(%)	AW
GL-SRI2-PS2B	0	0.791;0.742	0.463;0.357	0.42;0.27	0.337;0.277	88.8;94.7	0.7;0.8
	1	0.428;0.421	0.580;0.535	0.54;0.30	0.448;0.423	98.9;98.9	1.6;1.7
	2	0.336;0.283	0.709;0.601	0.42;0.34	0.550;0.443	96.5;98.9	1.7;1.7
GL-SR3-PS2B	0	0.831;0.805	0.442;0.326	0.42;0.22	0.301;0.250	96.0;98.9	0.9;0.9
	1	0.650;0.561	0.458;0.488	0.56;0.28	0.319;0.378	98.1;98.9	1.5;1.4
	2	0.331;0.419	0.720;0.571	0.43;0.35	0.553;0.461	64.5;63.8	0.4;0.5
GEV-SRI6-PS2A	0	0.884;0.903	0.323;0.243	0.59;0.18	0.199;0.186	96.3;100	0.7;0.6
	1	0.704;0.754	0.488;0.388	0.50;0.17	0.340;0.298	73.6;76.6	0.4;0.3
	2	0.570;0.624	0.632;0.481	0.47;0.21	0.479;0.369	37.3;45.7	0.0;0.1
GEV-SRI12-PS2A	0	0.983;0.988	0.120;0.112	0.89;0.51	0.087;0.083	99.2;98.9	0.1;0.1
	1	0.919;0.954	0.225;0.215	0.77;0.42	0.167;0.154	99.7;98.9	0.9;0.5
	2	0.893;0.920	0.286;0.288	0.80;0.45	0.211;0.211	99.5;97.9	0.7;0.5

5.4.3 Discussion

In this section, two different candidate Drought/Anomalies Indexes are tested and modelled for forecasting purposes for four different datasets: the Standardised Precipitation Index for a given period of time n (SPIn) for the collected data at `Med.RBD`, `Lan.GF` and `Org.GF` locations, and Standardised Runoff Index (SRIn) for the collected data at `Nar`.GF gauge station. In the corresponding tables (Tab.5.9, Tab.5.12 & Tab.5.15) we show the results got from S-W normality test. In our study, we analyse only four different data sets at five time-scales, resulting 140 candidate DIs in total. Following other works, the ideal for this experiment is to carry out this reiterative analysis on a larger spatial scale (e.g., at autonomic or national scale), from which we could infer arguments with statistical significance. For instances, Lloyd-Hughes and Saunders (2002) was one of the first studies at a continental level, in Stagge et al. (2015) uses a European scale grid resulting in 3950 hydro-meteorological time-series (per cell), in Svensson et al. (2017) develop a study for 121 datasets across the UK territory. Into a local scale, the main limitation in our study is the lack of collected datasets with at least 30 years of observations.

We tested the candidate DIs with the S-W normality test, resulting in a total acceptance rate of 12.1%. This means that our rejection rates are very high (≈88%) compared with

earlier works (Lloyd-Hughes and Saunders, 2002; Stagge et al., 2015; Svensson et al., 2017). This may constitute the object of future studies. We found higher p-values for `Med`.RBD case (Tab.5.9). This implies that higher p-values are associated with the longer time period of observations, showing greater certainty in the hypothesis that cannot be rejected at the 5% significance level. The "official" pluviometric anomaly (IESP) has been analysed. By definition it follows a normal distribution (Donaire, 2007), but it was rejected by the S-W test. We cannot consider this analysis definitive, future research could examine more robust normality test. For instance, driven by three conditions as is shown in Wu et al. (2007) or Naresh Kumar et al. (2009), where they included the S-W test as one of them.

When investigating candidate statistical distributions, the normal distribution (N) is the only one that has not passed the normality test in no case, showing that both variables, precipitation and streamflow, in our conditions, are non-normally distributed data. On the other hand, the G and W show the lower acceptance rates rejection rates ($\approx 10\%$). This does not match with the recommendations for Mediterranean environments of a previous work (Stagge et al., 2015). The well known Gamma distribution only past significant p-value for two cases in the Med.RBD SPI case ($n = 2, 3$), having a null behaviour for the other case studies. This is inline with previous studies in the area (Ayuso et al., 2015). In contrast, the Pearson type 3 (P3) has been a suitable distribution for three cases studied ($\approx 15\%$), all related to the precipitation droughts. The remaining distributions (GEV and GL with $\approx 20\%$) has the higher acceptance rates between all the DIs tested. Then, the most significant distributions are Pearson type 3, Weibull and the Generalised Logistic, where the first two are in tune with other authors for the SPI characterisation (Vicente-Serrano, 2006; Stagge et al., 2015) under analogous climate conditions: Mediterranean mountainous water systems. Previous works (Svensson et al., 2017) show SRI on different hydro-meteorological conditions, comparing twelve candidate $P(x)$ with the conclusion that a three parameter distribution (Tweedie) is the most flexible. But its application can be more time consuming. Later on, this distribution was used successfully in a good drought study for the UK (Barker et al., 2019) and it could be investigated in future studies since it fits well to rain patterns under Mediterranean conditions as is noted in Hasan and Dunn (2011).

On the n (time-scales) values, the most frequent are 3 and 6 months periods, passing the S-W test 6 and 7 times respectively. The long-term time-scale ($n = 24$) are uncharacterised since the proposed time series not passed the S-W test. This result ties well with previous studies wherein this scale simply is not used (Stagge et al., 2015; Svensson et al., 2017).

However, in other study (Belayneh et al., 2014) the n=24 scale is used in a forecasting experience but not performing a normality test within the methodological flowchart. In this effort, for scale 6 we found lag obstacles and over-fitting. The models demonstrate poor predictive capabilities (PI>0.25). In this thought, it would benefit to investigate alternative scales (n =4,5,7 or 9 months) of our concern in medium-term management.

Once the normality test is passed, we model the definitive SPIs, testing different explanatory input predictors. We can infer causality with the comparison of the results from these input variables. For this case, the wrapper PS building followed method is the same as in earlier cases of this work (Sec.4.2.1 and Sec.5.3.1).

During the IVS procedure, the ACF and PCAF function show up that the autoregressive orders are medium ($p \leq 3$). Therefore, we can maintain that our SPIs involves short physical autocorrelation, and consequently their autoregressive power is modest. Once the autoregressive character of the model is specified (PS1), we reach the next frame step, including related observations. In this subject matter, we checked four single AOs, and its partnership among them. For the Med.RBD and Lan.GF problems, the M3(MOI-GI) is the most influential AO. Its involvement gives advanced performance reaching the condition, $\mathrm{RI}_i > 2.5\%$, at least in 50 % of the metrics during the validation stage, and in some cases 100%. Also the variability among the models tested with different AOs intputs decreases as we go further in the time scale n AOs. This can be easily appreciated in Fig.5.12. On the contrary, the Med-SPI2 case not reached this condition. In a trial within the GF, the La.GF, the most influential AO is the MOI-GI and NAO index. In Di Mauro et al. (2008) used the NAO index as an exogenous observation, observing improvements into the SPI forecast result. On the SRI case, we again tested snow variables reproduced by hydrological modelling to describe the SRI pattern. It sounds to serve properly in larger time scales ($n = 6, 12$), and not too good for shorter time scales ($n = 2, 3$). This is for the inclusion into the modelling of intermediate hydro-meteorological processes such as those given by the model: Snowfall, Rainfall, Snowmelt. We have to think that unlike the SPI, which is a rainfall anomaly, the SRI is an anomaly of a pulse-response process, in this case the rainfall-runoff process at basin scale. It is logical since these predictors are resource entries to the basin during the antecedent months. Others (Staudinger et al., 2014) have shown how to use the snowmelt and rainfall to calculate an extension to the SPI. Further work is certainly required to disentangle these complexities in snow influenced river basins.

In our activity, we have encountered on the uncertainty estimates divergent ends. For the Med-GEV-SPI6-PS2, La.GF-W-SPI3-PS2, La.GF-W-SPI6-PS2 and three of the four SRI

cases has reacted correctly, with low AW values and POC values close to the theoretical (95%). However in the other cases, we found AW values remarkably large and POC values far from the theoretical, indicating an ambivalence and overestimations under all conditions, and therefore a modest model. This is because of an impaired interpretation of the problem (inputs-target) or a wrong parametrisation not converging to excellent solutions. In this understanding, we have established a profound trial and error search activity to illustrate these features, such as the BNN's optimisation parameters: Learning rate and momentum, and Bayesian hyper-parameters: α and β, more in App.A.1.2. Although they are different time scales, it is costly to conduct it for each experiment. In any case, the main outcome with these uncertainty results could be interpreted as not good quality, requiring a deeper work in this thought.

For drought forecasting purposes, the models performance in the negative phase of the indices has been insufficient. On the other hand, its behaviour for the positive phase has been tolerable since the models can show predictable capabilities on positive anomalies. The increment of the time horizons increases the error proportionately, having not detected any substantial predictive behaviour for $h > 0$ step ahead values.

5.5 Conclusions

We performed three medium-term forecasting regression actions in two research sectors: the Mediterranean River Basin District and the Guadalfeo Basin. With these ends, we propose to turn out as a first step to set up a prediction practice for decision support. We require further development for a finer fit, with the goal of establishing online information tools for many users. These results offer a useful means to know the total resource at various time horizons on a quarterly basis, rainfall forecasting on a monthly basis, and the "drought characterisation" in advance through two Drought Indexes: Standart Precipitation Index and Standart Runoff Index. With these experiences we can infer physical meanings and particularities of the water systems under study. The model-based approach show us which exogenous observations from a set of candidate are the most relevant variables. Somewhat, we require higher accuracy for agriculture and urban purposes in the medium term planning for the local water managers.

• **Cummulated streamflow forecasting:** Establishing robust relationships in a hydrological framework and time scale such as the one performed here are appropriate for management as it allows us to foresee future volumes in an accumulated manner. Between the two tested scenarios, the scenario B offers good results to continue working in the future. The example showed here, provides us to establish a local train within the development of prediction tools to apply it to the management in a high mountain basin with great influence of snow. Under a static approach, we proved that the hybrid methodologies of classification and regression, work better than for those based only on a regressive framework. Among the classifiers tested, the most balanced for all cases has been the Support Vector Machine computational technique. Also, the results show that both atmospheric oscillations and hydrological modelling help in medium-term estimates. Being a simple practice a priori, the uncertainty associated with the predictions is significant for anyone decides, since it is very variable. For this, it is necessary to relate it to other types of variables or observations based on the physics of the water system and also the atmospheric oscillations.

• **Rainfall Forecasting:** In this study, we apply BNN models to investigate forecasting behaviour at 1, 2 and 3 month ahead rainfall for two Andalusian regions. Our comparative assessment trough a model-based approach suggest that an improvement on rainfall forecasting is found using AOs as exogenous input. The best input form to model it is

trough pairs of AOs, and the more influential AOs are: NAO, WEMO, and MOI ranked in order of importance. It has been viable to represent the general behaviour of rainfall with its marked seasonality. It has been more challenging to identify some predictors that provide us some hope a precise evaluation for the individual events, drier months, or wetlands within the normal cycle. Effort must go on in this direction. This is laborious since the rainfall has a clear chaotic component affected besides regional scale effects. The results showed that this approach can forecast rainfall, looking very promising for medium term forecasting and for water management applications.

• **Drought Forecasting:** This regression practice seeks to consider the capacity of Data Driven models for drought forecasting purposes in the survey areas. The Standard Precipitation Index (SPI-n) characterised the pluviometric anomalies occurrence at varied time scales (n): SPI2, SPI3, SPI6, SPI12 and SPI24. For that, we verified the appropriateness of seven statistical distributions for characterising medium term precipitation deviations under Mediterranean conditions. From the original set of indexes, we take a successor set, which has those that pass the Shapiro-Wilk normality test. We used these DIs for further Bayesian Neural Networks forecasting regression experience development for three time horizons. We used supporting variables as Circulation Indexes and Hydrological modelling, not finding a fix pattern for all the cases under study. The best option can be using just one or a combination of pairs. For the SPIn cases, the most influential AOs are MOI-GI and MOI-CA and its pair combination. On the streamflow anomalies cases (SRI) at Narila, the additional hydrological modelling helps the model forecasting capabilities at larger time-scales ($n = 6, 12$). On the contrary the AOs tested help more at shorter time-scales ($n = 2, 3$) finding that the WEMO and NAO has more relevance.

Chapter 6

Final Conclusions

"It is a capital mistake to theorize before one has data.
Insensibly one begins to twist facts to suit theories,
instead of theories to suit facts"

∼Arthur Conan Doyle∼
Sherlock Holmes. Adventure I. A scandal in Bohemia

This chapter summarises the contributions of this work as a whole, articulates some of the questions raised out, and establishes the following steps to be taken into the future. Each chapter contains its particular discussion section and here the intention is to collect the most important points and bring them all together in the context of the whole thesis.

6.1 Summary of the Research Contributions

The comprehensive aim of this thesis was to investigate several issues referred to Artificial Intelligence (AI) and Machine Learning (ML) for its utilization in Mediterranean water systems. Essentially, the application of AI and ML to water resources is taking data and using algorithms to make predictions and to optimise decisions, and that underlies almost everything that the hydrologist do with hydro-meteorological time series. Since the primary asset of these applications is data, we have preferred to use the term Data-Driven (DD), since the terms AI and ML can cause a lot of confusion for non-experts, where science fiction ideas quickly come to mind.

In a first step, we understood the physical domain to conceptualise with the DD models, with its particularities and its environmental components. In chapter 2, we have conducted an exhaustive study of the spatio-temporal characteristics and the accessible data within the water systems under study: two south-eastern Andalusian basins (Guadalhorce [GH] and Guadalfeo [GF]). Here, we performed a spatial analysis of the flood risks evolution and a temporal analysis of the occurrence of floods events. We found how the potential flood risks have escalated in recent years by the urban development intrusion into the floodplains. In the Guadalhorce RB, a heavily populated basin, we also found from the historical data that the upstream reservoir system has little capacity for regulating floods for some severe rainfall fronts, where the results were part of the work Egüen et al. (2015).

As for the available data, it is true that there are no very long time series available. This is common in these regions, where the culture of measurement has not been as ingrained as in other latitudes. The Mediterranean character of the climate has not helped in the establishment of these measures either: ephemeral channels that are only activated by major rain events and, when they do, carry catastrophic flows capable of ruining installations. As regards rainfall, the torrential nature, with very significant local rainfall concentrations, and the abundance of remote and sparsely inhabited mountain areas, have provided another historical handicap for the achievement of a stable and significant measurement network. Obviously this is a major drawback for the successful application of DD models, which demands vast amounts of data. Re-sampling techniques have helped us to improve behaviour in the presence of these limitations. However, this is the great challenge of this work, since these situations where scarcity of data is the rule, are common in hydrology, especially in drier and less developed countries. For this reason, we have based our work on ancillary data that are easier to obtain on a global scale, such as teleconnection indices and drought-related anomaly indices, which will be important in the results got in the rest of the thesis.

In chapter 3 we compared the key approaches and conceptions for hydro-meteorological modelling development with DD models. Also, this chapter gives a state of art review of modelling paradigms for our purposes. We commented the modelling approaches that can be found under the main conceptual perspectives, trying to comment on all the pros and cons of each approach, and establishing the most appropriate methodological framework for our objectives. We presented this part of the investigation in Herrero et al. (2014). Later, we described a complete picture of the data management for a supervised learning exercise, followed by the data transformation and the pre-processing

steps that optimise the efficiency of the algorithms. We also found how the bootstrapping re-sampling technique helped, in particular, for the Mediterranean typical patterns. Last, we discussed which performance evaluation metrics are under an appropriate benchmark for our purposes. In regression problems, the measures selected for deterministic outcomes were Nash-Sutcliffe efficiency (NSE), coefficient of determination (R^2), root-mean-square error (RMSE) and mean absolute error (MAE), to evaluate the model's efficiency. To test the model's predictability capabilities, we selected the Persistence Index (PI). For the uncertainty outcomes, we selected the Average Width (AW), and Percentage of Coverage (POC). Finally, we selected the Bayesian Information Criterion (BIC) to quantify the model parsimony. On the other hand, we selected the confusion matrix frame, and its associated rates, to evaluate the performance of classification problems.

From a DD modelling point of view, we selected three computational techniques for the regression experiences: Multiple Linear Regression (MLR), Bayesian Neural Networks (BNN) and Gaussian Processes (GP). MLR is a pure linear and parametric model, which is optimised straightforward. MLR is an excellent starting point for any regressive problem because of its simplicity. At an hourly basis (Sec.4.3.2), we found that MLR can overcome the predictive behaviour of BNN for some time horizons. The BNN technique has strong non-linear capabilities, and implemented with the Bayesian approximation framework it can offer us probabilistic estimations. Throughout the work, we visualised different assumptions through the learning trends, that let us appreciate minor differences among competing predictors and the optimal hidden units of the neural network structure in each application. This step also helped on the random initialisation of the parameters, where multiple candidate models are generated, i.e., local minimums, and we need to choose the one that behaves best. In this sense, an opportunity arose with the development of ensembles, although the variability based on the BNN's weights initialisation is less interesting than other sources of variability, for instance based on several DD methods, or based on various training data sets generated by re-sampling. In any case, the BNN method has the Bayesian approximation that allows us to adjust the probabilistic estimates through the evidence procedure, and also providing an input interpretative framework (Automatic Relevance Determination). The GP model is an interpretable, non-parametric, and a fully Bayesian technique which is optimised straightforwardly. The results revealed us the capacity of this technique with the combination of various kernels, for example linear and non-linear, in a single model. One major limitation on its application regarded in the powerful influence of the prior knowledge, where a solid statistics knowledge is a requirement, and therefore its automation is more complex. But as we have mentioned before, various works are progressing on this matter. Among the

three techniques compared in this work, BNN is the most balanced of the three for all the cases used for automated hydro-meteorological patterns regressions. As we discussed previously, it would be interesting to test this forecasting experience for other DD models.

Also, the physical hydrological modelling (WiMMed) was used for the generation of historical time series of snow variables within one basin (GF), that were proved to be an essential agent. The results showed how the identification capabilities of the DD models considerably improve their behaviour with hydrological modelling, both in efficiency and in predictability terms.

In chapter 4 these three DD models are compared in short-term forecasting regressions cases. For that rainfall-runoff modelling through seven real cases at hourly (one) and daily basis (six) were analysed. Within the basins, the points where the estimates have been made are of great hydrological interest: on floodplains of great socio-economic importance (Sec.4.3.1 and Sec.4.3.2), for reservoir inflows (Sec.4.3.3, Sec.4.3.4, Sec.4.4.1 and Sec.4.4.2) and for water allocation purposes (Sec.4.4.3). We discussed the most relevant issues in each application attending to its deterministic and uncertainty estimations. Because of the large number of input predictors available, i.e. meteorological stations, we developed a model-based approach for Input Variable Selection (IVS) purposes, which takes into account the linear non-linear relationships among inputs-outputs within the rainfall-runoff process for each case. We carried out the IVS under a stepwise approach while preserving the parsimony principle to select the most relevant predictors. Initially, also appeared some spatial-temporal inconsistencies within the rainfall-runoff process at a daily basis which was overcome by a persistence analysis. At hourly basis, the behaviour of the single stations exceeds the average value among them, and the behaviour of the models with independent inputs to represent the rain agent, surpasses the models with one input. At daily basis, the behaviour of the averaged models almost always exceeds a single location, and the behaviour of the models with only one input to represent the rain agent, surpasses the models with several independent inputs. The results of the best rainfall predictors show that they are not defined by a clear physical meaningful, as the location of the station in the runoff generation area, or the station altitude, or the distance. Under a deterministic point of view, the results in this chapter shows that the best computational technique is BNN. However, under the uncertainty point of view, the GP method gives the most balanced outputs for forecasting in our Mediterranean conditions. Also, it is pointed out the efficiency of the joint use of hydrological modelling and DD techniques for forecasting purposes. In basins of complex nature, where the snow is a fundamental hydro-meteorological agent, we validated the importance of the physical meaning within

a DD framework to represent hydrological processes in Mediterranean water systems. We presented some chapter proceeds in Gulliver et al. (2014).

In chapter 5, three forecasting experiences for medium term were developed. We base the initial one (Sec.5.2) on a quarterly basis, where we established a hydrological temporal scheme to determine the cumulated streamflow for specific time horizons based on relevant dates where the hydrological planning takes place. We discovered that the forecasts are more prosperous after having consumed the initial six months of the water year, instead of three, in which we conduct the evaluations.The observed input variables quantified in the water system are: the cumulated streamflow, the cumulated rainfall, the cumulated snowfall and atmospheric oscillations (AO) values. At the DD modelling level, this experience has shown the importance of combining mixed regressive classificatory models as opposed to just regressive models within the static frameworks. In this manner we are able to narrow the solution space and thus optimise predictive behaviour of the DD model. We have also carried out a classification exercise comparing three classifiers: Probabilistic Neural Network (PNN), K-Nearest Neighbour (KNN), and Support Vector Machine (SVM). We found that the SVM behaves better than the others. However, more research on classifiers is still required. The results of this DD experience were presented in Gulliver et al. (2016)

In the second experience of the chapter (Sec.5.3), we performed a rainfall prediction exercise using BNN, and following the same model building method (IVS) applied previously (Chap.4). The results have been shown poor in predictive terms ($\mathsf{PI} \approx 0$), confirming the exigency of predicting this variable from historical data. Although complex hydrodynamic modelling is required for the prediction of this important variable, this case serves to infer causality of the most relevant AOs for our study area. Finally, we found that the best input form to included it into the model is trough AOs pairs. The more influential AOs in this experience are: NAO, WEMO, and MOI ranked in order of importance.

In Sec.5.4 a rainfall and runoff anomalies prediction exercise was performed at different locations within the Guadalfeo RB. We verified the appropriateness of seven statistical distributions for characterising the anomalies under Mediterranean conditions. The indices that passed the Shapiro-Wilk normality test were modelled to analyse the BNN model ability to predict these indices at various time horizons. Here, the predictions of the negative phases (droughts) have been poor, and the behaviour of the models for the positive phases has been more successful. Concerning to the AO causality analysis, we found how the NAO and WEMO helps to the forecasts at shorter time horizons, and similarly, MOI helps for longer time horizons. We have analysed the relevance of

these atmospheric variables in each case, concluding that sometimes their inclusion was convenient and sometimes not. For this reason, we detailed this in each case study.

Additionally, the appendix A describes the Data-Driven computational tools used. appendix B analyses the datasets used during this work. Also this appendix provides supplementary material of Chapters 4 and 5.

6.2 A Final Statement

The overall sight of this work, was to analyse a number of methodological options related to Data-Driven hydro-meteorological modelling across different temporal scales that have been emerged from the mood feeling and literature. In particular, the need to rigorously establish frameworks for fair evaluation of the efficiency and the "real" time skills of the method or model tested.

We have developed a computational framework with which to process in a repetitive and creative manner the colossal amount of collected data nowadays. Bayesian approaches allow us to re-estimate the model as new data appears, this being of vital importance at the present moment, as hydro-climatic patterns are changing and the adaptation of the model appears to be a vital feature. We should not conceive the DD modelling as the unique information source during the decision-making process, but should be part of the information sources to be analysed by the decision makers.

The requisite of know-how to apply models is always a good approach in terms research experience and its transfer strategy. This is important since we discuss many points with which we would find ourselves in the practical reality and that can be useful for those who consult it. That is why there is the necessity of this hydro-informatics experiences in complex watersheds under specific climate dynamics. We can extrapolate the results to other regions as we can find similar water systems throughout the Mediterranean basin. This also provides us wisdom to integrate and build online hydrological modelling web services, whether in the public or private sphere.

6.3 Future work

There are many ideas for future directions associated with this work. During the development of this work, numerous future research directions have emerged. This field of computer engineering related to computer data processing is in full development and is offering new solutions in all sectors of knowledge. Hydrology is not left out of this. We now summarise some opportunities for future work based on these ideas.

More sophisticated correlation and evaluation coefficients: On variable correlation matters, more advanced coefficients should be analysed in the future as the linearity limitations of simple correlation coefficients are well known. Although in our case this has been overcome with the model based approach, will be more useful straightforward approaches. More research in other correlation criterion as the entropy (Shannon et al., 1951) is needed. For the assessment of the results comparing the model estimations versus observations, the use of new evaluation frameworks with a statistical basis as the proposed in Ehm et al. (2016) may solve the time scale problems effects in the comparison benchmark exercises.

Use of climatic-atmospherical data: For real hydro-meteorological forecast exercises at regional level, estimates may benefit from satellite products and the coupling complex hydrodynamic modelling, i.e. global and regional circulation models.

Deep learning: Gamboa (2017) shows how to use deep learning for the time series forecast with Convolutional Neural Networks, where a comparison exercise with classic Machine Learning with Artificial Neural Networks appears as one of the next natural steps to be undertaken. Looking forward, further attempts could prove beneficial to the literature.

Filtering: From the literature review, the Wavelet Transform analysis is a pre-processing procedure step for time series modelling that improves the forecasting performance with DD models in geo-sciences. Also permits extracting non-linear trends, making multi-scale correlations and identifying underlying periodicities, which has a remarked influenced under Mediterranean conditions. Because we were interested for real-time forecasting capacities, we only should consider the stationary wavelet transform and the *a'trou* algorithm (Shensa, 1992), although the application of its discrete version is very common

in literature. Therefore, future research should be conducted in more realistic settings to the couple use of wavelet transform and DD methods for hydro-meteorological patterns identification.

Optimisation methods: For parameters optimisation during this work, we have used Gradient Descent algorithms, but later works (Morse and Stanley, 2016) prove us the demand to analyse evolutionary methods. Also, evolutionary methods are beneficial for allocation optimisation in multi-purpose reservoir systems, as this non-linear method can apply to any model or optimisation problem. In future exploration, we need better research to apply and assess genetic algorithms in our reservoir schemes for decision making.

Data Driven models for water management: One aim of this work was the study of DD methods for water resources management, optimisation methods for decision-making, for their application in our particular reservoirs. In this sense, investigating in multi-objective DD methods needs to be undertaken.

6.4 Conclusion

The process has shown that the path is not easy and the results are not spectacular when there is not a large amount of data or it is not of excellent quality (due to problems inherent in the very nature of the physical environment where it takes place). There is no Data-Driven model that best fits all the cases analysed, depending not only on the climatic environment, but also on the particularities of the basin (with or without snow, size, concentration time) and the temporal scale of application. The support of typical data in hydrological studies (flow and precipitation) with other novel data such as global indices and physical model results, has helped to improve the performance of DD models within the forecasting task.

In this regard, the work concludes by arguing that there is no single superior DD model for all cases, where we must apply various methods with different fundamental, and even antagonism approaches. These methods allow us to establish causal and relational hypotheses based on the comparative analysis of the input relevance and on model-based iterative procedures. Using DD approaches, we tested the hypothesis that simple models may be sufficient for tools in hydrological engineering for planning factors. Overall, our results show a relevant effect of atmospheric oscillations as NAO, WEMO, and MOI, related to rainfall and wet and dry periods in the area, especially on larger time-scales. In conclusion, these modelling experiences help to improve our capacity to deal with uncertainty on hydrological matters in complex climate conditions, in our case the Mediterranean climate conditions. The automated and supervised development frameworks proved in this work, can establish, in a quick and robust manner, computational tools for different hydrological engineering tasks.

Employing DD to hydro-meteorological analysis has generated results in the territories under study that serve as a satisfactory background to interpret factors under driven rules, appearing as a promising field for improvement into the understanding of the specificities of unique water systems. Importantly, our results provide evidence for the development and the establishment of smart tools for decision support using the time series collected and generated in the areas under study.

Appendix A

Data-Driven Computational Techniques

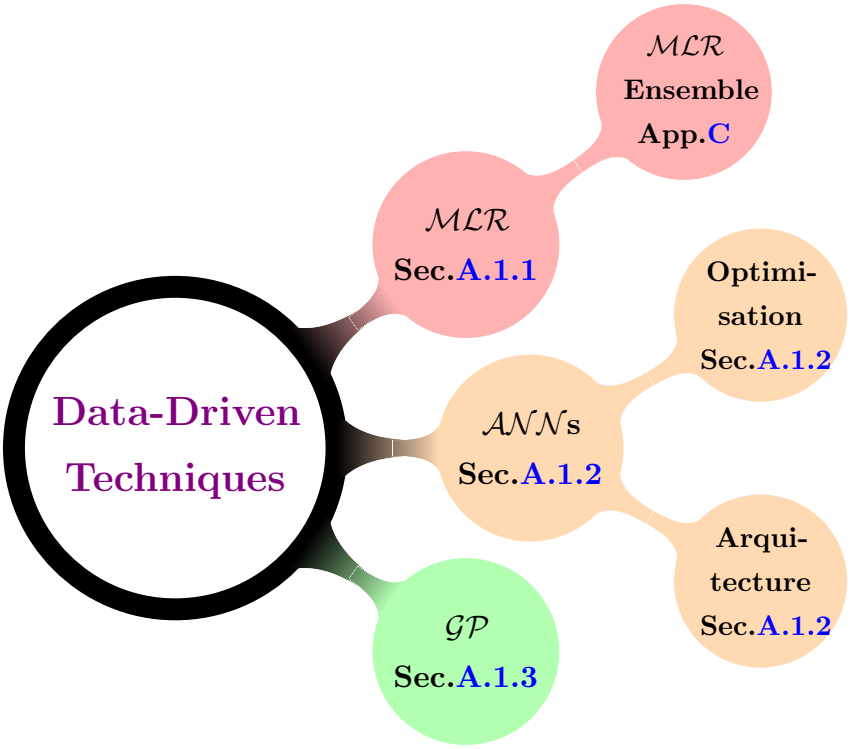


FIGURE A.1: Data-Driven Techniques

The DD computational techniques used through this work, i.e., $\mathcal{M}$, are exposed in this appendix. This appendix does not intend to be an exhaustive description of these techniques, already sufficiently explained in the literature, but a compilation of their main characteristics, together with the problems and solutions adopted throughout this work

in their application to real hydrological problems. In Section A.1.1 the Multiple Linear Regression ($\mathcal{MLR}$) is presented. Section A.1.2 concerns to Artificial Neural Networks ($\mathcal{BNN}$) under the Bayesian framework applied for this technique. In Section A.1.3 we explore Gaussian Processes ($\mathcal{GP}$) for regression problems.

A.1 Machine Learning methods

A.1.1 Multiple Linear Regression

Multiple Linear Regression ($\mathcal{MLR}$) is a linear technique that makes it possible to find the best relation between the variable y and several alternative variables x through the least square (LS) method. $\mathcal{MLR}$ models can be proposed by the accompanying equation:

$$\mathsf{SF}_{(t+h)} = f(\beta_0 + \beta_1 x_{1(t)} + \beta_2 x_{2(t)} + ...\beta_d x_{d(t)}) + \varepsilon^{\circ}_{(t)} \tag{A.1}$$

where $\mathsf{SF}_{(t+h)}$ is the target variable at a time horizon h from time t, and $x_{1(t)}, x_{2(t)}, \ldots, x_{d(t)}$ are the independent variables at t. $\beta_0, \beta_1, \ldots, \beta_d$ are the model parameters and $\varepsilon^{\circ}_{(t)}$ represents the model deviation term. This term is based on random normal distributions with a zero mean and with different variances depending on the predictor values, more in Chatterjee and Hadi (1986). We utilize this variance model for the linear uncertainty assessments. The parameters optimisation process of this technique is straightforward. One strength is that it brings outputs in a probabilistic and explicit manner, while its principal weakness is that the estimated error increases notably with high non-linear relationship between the processes involved. This method is recommended for event modelling and spatial analysis. We can find extended information of this method in specific literature, i.e., Chatterjee and Hadi (1986).

A.1.2 Artificial Neural Networks

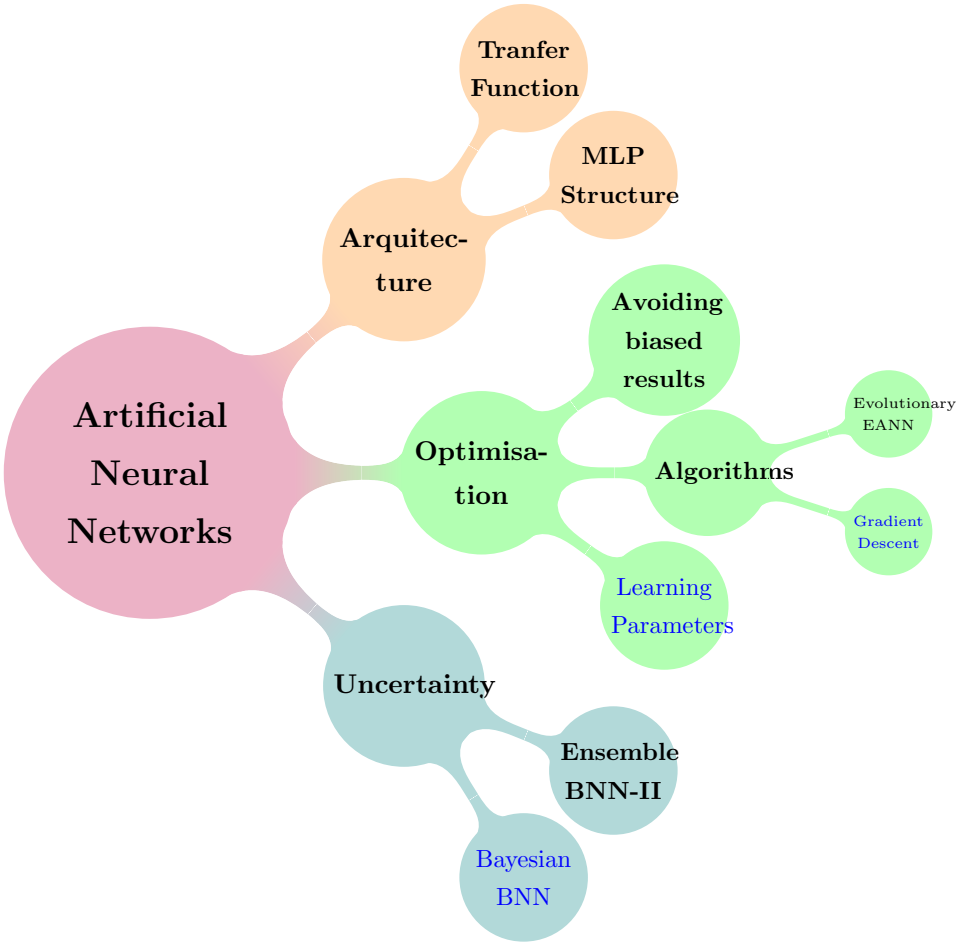
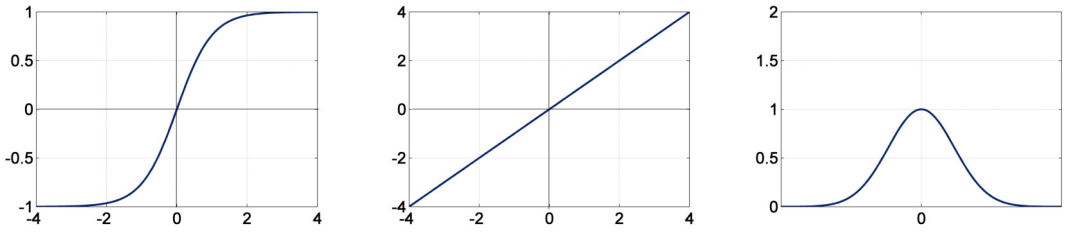


FIGURE A.2: Artificial Neural Networks development

The Artificial Neural Networks ($\mathcal{ANN}$) are flexible when fitting hydrological data. It has been employed on countless times to forecast hydro-meteorological variables in the past with a mandatory review in Maier and Dandy (2000). One strength of this method is that they can set up relationships between data successfully, dealing with complex and non-linear domains. It should be noted that $\mathcal{ANN}$s are a powerful method for recognizing non-linear relationships, this being their main attribute. Below, we explain the foundations of the Mutilayer Perceptron (MLP) structure, which is a type of $\mathcal{ANN}$ configuration, carrying out an analysis of the decisions that we will face during the $\mathcal{ANN}$s application and development.

The Multilayer Perceptron Structure The Multilayer Perceptron (MLP) is a feed-forward artificial neural network ($\mathcal{ANN}$) model, where the basic elements is a single unit/node called a neuron see Fig.A.4. These units integrate the information received

from previous layers and apply a transfer function f_a (see Fig.A.3). The nodes are generally arranged in layers, starting from a first input layer with nodes number same as the number of terms of the functional relationship $f(\cdot)$ represented in Eq.4.1, and ending at the final output layer, in our case, with a single node representing the predicted $\mathsf{SF}_{(t+h)}$ value, y.



(a) Hyperbolic tangent function (`tanh`)
$f_a(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

(b) Lineal function $f_a(x) = x$

(c) Radial basis function (`rbf`)
$f_a(x) = exp(-\frac{\| x - c_N \|^2}{2\sigma_N^2})$

FIGURE A.3: Standard transfer functions (f_a) used in $\mathcal{ANN}$ Multilayer perceptron architecture

Information passes from the input to the output side. It attaches the neurons in one layer to those in the next, but not to those in the same layer. Thus, the output of a neuron in a layer is conditional only on the outputs it receives from previous layers and their corresponding weights and biases. On dynamic time series applications, the input data is trained by supervised learning with a set of predictors as the input data and the corresponding values observed for each time slice t.

The Optimal Architecture As is showed in Fig.A.4, for Rainfall runoff modelling purposes the $\mathcal{ANN}$ structure have three layers in this work: the Input layer (`In`), the hidden layer (`HL`) and the output layer (`Out`). In between layers, there is only a single hidden layer as is recommended previously for Rainfall-Runoff modelling (Govindaraju, 2000; Birikundavyi et al., 2002; Kişi, 2004). During the MLP architecture election procedure, an important parameter that need to be set out is the number of neurons within `HL`. The optimal number of hidden neurons can be complex as it depends on the algorithm of optimisation as well on the data. On the contrary, the number of neurons in the remaining layers, `In` & `Out` layers, are defined by the problem itself. For that, there is not a magic formula for selecting its optimal number of hidden nodes.

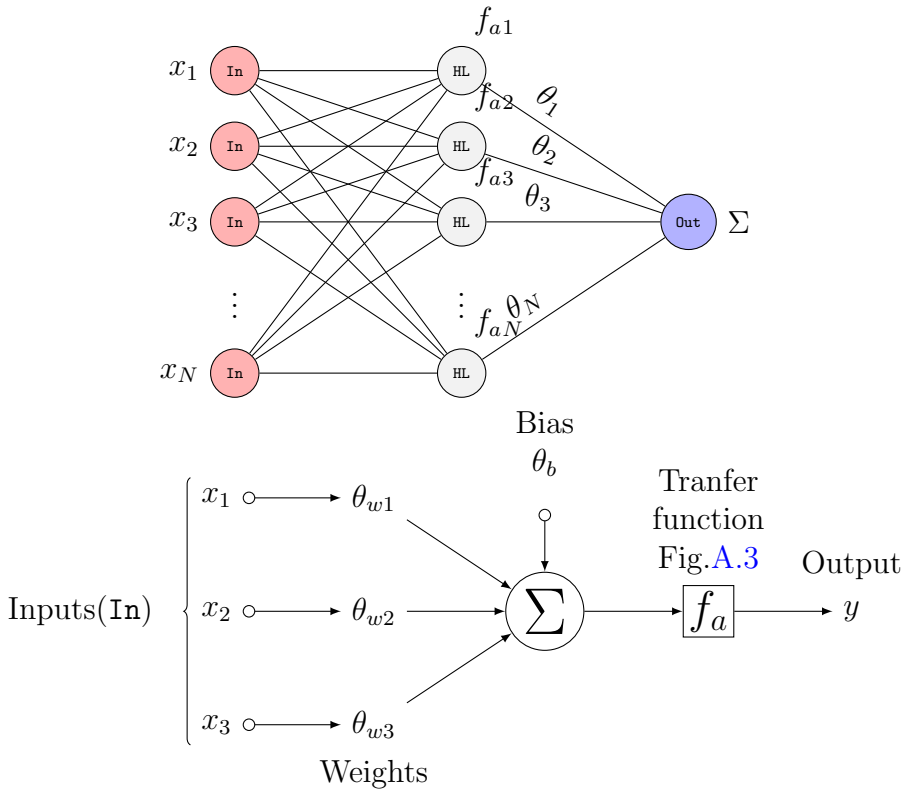


FIGURE A.4: Multilayer Perceptron (MLP) diagram

However, some thumb rules are available for calculating the number of hidden neurons. A rough approximation can be obtained by the geometric pyramid rule proposed by Masters (1993), where for a three layer network with n input and m output neurons, the hidden layer would have $\| \sqrt{(n \cdot m)} \|$ neurons. If this rule is applied in our experiences the optimal HL value is just a few units, $\texttt{HL} \approx 3$. Another good approach was proposed by Fletcher and Goss (1993), establishing the range from $2\sqrt{n+m}$ to $2n+1$ for model-based approaches. In any case, during the development of this work, we could see that testing a range by a trial and error procedure was the best choice. Various works point to it (Coulibaly et al., 2000; Anctil and Rat, 2005; Singh and Borah, 2013; Kasiviswanathan and Sudheer, 2013). During this work, an analysis to determine the optimal number of neurons within the HL in each $\mathcal{ANN}$ configuration is carried out. The regression on the performance metrics with an increasing number of the hidden neurons from the results of the dynamic multiple simulations is visualized in specific graphics. In this way we seek the optimal number of hidden units regions, since we can identify the region that offers the best performances and less variance. The following results, Tab.A.1, are from the case studies of chapter 4. Here we can see how the cases with a more marked linear behaviour, require a smaller number of neurons than those with a non-linear behaviour.

TABLE A.1: Results on the number of hidden neurons units obtained for the best BNN model for the daily study cases shown in chapter 4. Among brackets the 75th percentile value

horizon \ Model	GH1	GH3	GH4	GF1	GF2	GF3
$h = 0$	6[**2**]	85[**85**]	60[**5**]	21[**22**]	20[**7**]	3[**18**]
$h = 1$	2[**2**]	15[**15**]	17[**3**]	60[**22**]	4[**6**]	4[**2**]
$h = 2$	2[**2**]	8[**20**]	15[**10**]	85[**17**]	13[**6**]	2[**2**]

Radial Basis Function Network A Radial Basis Function Network ($\mathcal{RBF}$) is a Multilayer Perceptron (MLP) with the radial basis as a transfer relationship function (f_a, see Fig.A.3.(c)). In the equation, $\|\cdot\|$ represents the euclidean norm, and c_N, σ_N^2 are the training data centre and variance of the hidden neuron N respectively. In such networks, would normally require more neurons in the hidden layer than simple $\mathcal{ANN}$, but requiring lower computation time (Demuth and Beale, 1993). The biggest difference between the $\mathcal{RBF}$ and $\mathcal{ANN}$ is the non-linear relationship in the hidden layers: if in simple $\mathcal{ANN}$s this is achieved by the transfer function, in the $\mathcal{RBF}$ relationship it is based on determining c_N and σ_N^2 of the basis function, which in our case is a Gaussian function. Such networks have been applied successfully to predict flows and floods (Chang et al., 2001; Lin and Chen, 2004). For a comprehensive and detailed study of this type of development of such networks refers to Park and Sandberg (1991). In this work, this network architecture, named `SCG2`, have been compared to a simple $\mathcal{ANN}$ (`SCG`) with the tangent hyperbolic as f_a in Tab.A.2 and Tab.A.3 for some short-term forecasting experiences carried out in Chap.4.

Parameters and Optimisation Algorithms During the training stage, the weights ($\theta : \{\theta_\omega, \theta_b\}$) of each network member $\mathcal{ANN}_\mathsf{K}$ are optimised, and progressively corrected through supervised learning algorithms. The Gradient Descent (GD) back-propagation algorithms, determine the weights of the $\mathcal{ANN}$ model in an iterative procedure by minimising a cost function at every step. The optimization algorithm starts by setting the initial values of the parameters (θ_{init}), in our case a small random value following a $N(0, 1)$ distribution. Then, we evaluate the cost (MSE) and the derivative of the cost (∂MSE). The gradient is calculated in order to know the direction, positive or negative, that will offer us a lower cost in the next iteration. Once the direction is known, we update the parameters according to the learning rate parameter, which controls how much the parameters can vary in each iteration. In theory, this process would end when the cost is

equal to zero. In reality this never happens, so a minimum improvement is established, and when it is not exceeded the iterative optimisation procedure is completed.

A relevant matter during this training step, is that weights are initialise randomly to break symmetry. The most frequent proposal is to initialise it from a normal distribution, $\theta_{init} \sim N(0,1)$. Also, we find rules that take into account the network architecture, for instance, depending on the number of inputs and output units $\theta_{init} \sim f(neu)$ as with the following relationship:

$$\theta_{init} \sim f(neu) = \sqrt{6}/\sqrt{(n(L_{\texttt{In}}) + n(L_{\texttt{Out}}))} \tag{A.2}$$

where $n(L_{\texttt{In}})$ and $n(L_{\texttt{Out}})$ are the number of neurons units in the input and output layer respectively of the contiguous layers. That is equal to the number of inputs, which depends on the problem, and the number of outputs, which in our case will be one.

Both weights initialisation approaches are compared in Fig.A.5 for the GF1-PS4A dataset (Sec.4.4.1). In this way, we can test its final effects on performance with data of our interest. From these learning trends visualisations, we can comprehend that the initialisation with the Normal distribution option leads us to a much greater variability on the models performances, obtaining superior models. We can also see how as we increase the number of neurons in the models obtained with $N(0,1)$ initialisation, they tend to generalize worse. So, based on theory, we are not going to work with models with high number of neurons within `HL`, as recommended by previous works.

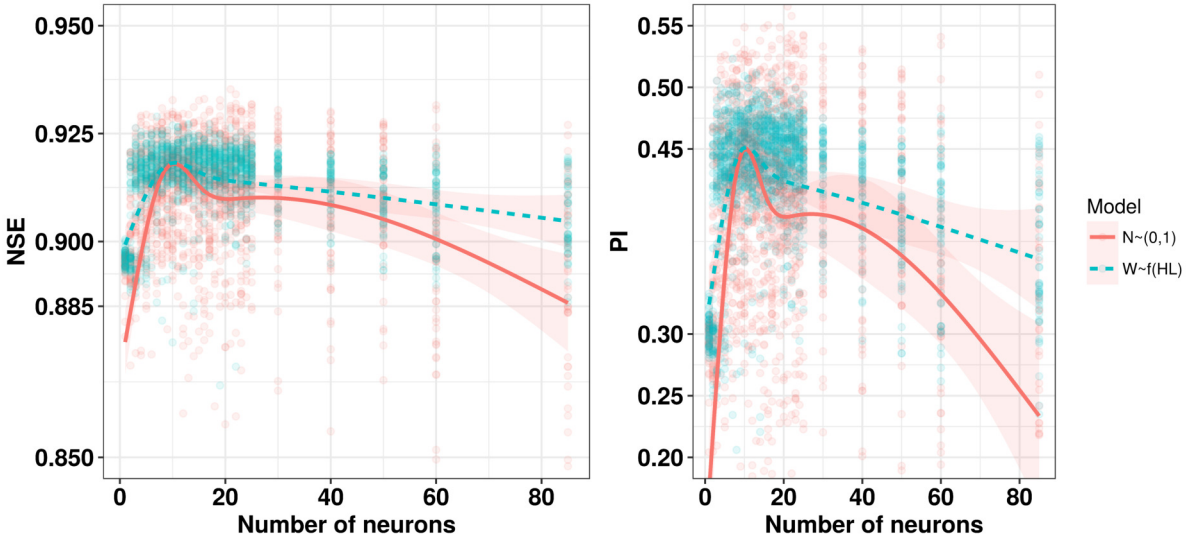


FIGURE A.5: Learning trends comparison for two types of weights initialisation θ_{init}. The model showed is the GF1-PS4A for one day ahead forecasting and for the validation period

In consideration to the GD optimisation algorithms, various variants are usable to work out the $\mathcal{ANN}$s parameters θ. In this activity a comparative survey is realised taking the following six algorithms: Scaled Conjugate Gradient (`SCG`, Moller (1993)), `SCG2` with Radial Basis Function (`SCG2`), Bayesian Regularisation (`BR`, MacKay (1992b)), Levenbert Maquartd (`LM`, Marquardt (1963)), One Step Secant propagation (`OSS`, Battiti (1992)) and the Conjugate Gradient Power-Bale (`SCB`, Powell (1977)). These backpropagation iterative algorithms adjust the weights and biases to minimise the objective function, in our case is the MSE (see in Sec.3.4). Earlier works (Rezaeian-Zadeh and Tabari, 2012) show that the `LM` has been practiced successfully for drought index (DI) forecasting in Iran. In Deo and Şahin (2015) a connected analysis of `OSS` and `SCB` algorithms is carried out deeply.

To contrast the algorithms behaviour with our datasets, repeatedly, we bring the multiple simulations approach with several ascending hidden neurons (Sec A.1.2). We base the decision of the strongest algorithm on the following aspects: 1. on its efficiency with the NSE coefficient, 2. on its predictive capabilities with the PI coefficient, 3. on time by the average computing time ($\overline{\texttt{CT}}$, in seconds s) for the 1500 runs, and 4. on the model complexity characterized by the higher 75th percentile value of neurons within the hidden layer `HL`. We seek to maximize the NSE and PI coefficients and minimize HL and $\overline{\texttt{CT}}$ values. This analysis is computed only for one lead time prediction $(t+1)$ and for the best predictor structure PS resulting from Sec.4.3 to simplify the comparison.

In Tab.A.2 and Tab.A.3 we list the results od six optimisation test for the GH and GF daily cases, respectively. It can be observed that the `LM` optimisation algorithm is the most efficient one in terms of computational costs. On the other hand we see how the SCG is quite balanced in all aspects. The exception of GH3 is noteworthy, since we find a high number of neurons in `HL`. In these results we can also see that the CGB algorithm has a much shorter computation time than all the other algorithms used, about two orders of magnitude less.

TABLE A.2: Performance comparison for six optimisation algorithms (`Algo.`) results for the regression experiences in the Guadalhorce Basin at a daily basis

Case	**GH1-PSB3**				**GH3-PSA3**				**GH4-PSA3**			
`Algo.`	NSE	PI	HL	$\overline{CT}$ (s)	NSE	PI	HL	$\overline{CT}$ (s)	NSE	PI	HL	$\overline{CT}$ (s)
`SCG`	0.888	0.52	2	72	0.906	0.66	85	24	0.841	0.65	5	24
`SCG2`	0.884	0.52	22	84	0.902	0.65	6	22	0.847	0.67	42	23
`BR`	0.889	0.49	25	60	0.909	0.67	9	22	0.837	0.63	43	16
`LM`	0.895	0.52	9	81	0.914	0.69	14	6	0.848	0.68	29	7
`CGB`	0.885	0.49	11	0.8	0.876	0.56	20	0.4	0.791	0.47	19	0.4
`OSS`	0.889	0.51	15	145	0.900	0.64	12	41	0.840	0.64	13	42

TABLE A.3: Performance comparison for six optimisation algorithms (`Algo.`) results for the regression experiences in the Guadalfeo Basin at a daily basis

Case	**GF1-PSA3**				**GF2-PSB3**			
`Algo.`	NSE	PI	HL	$\overline{CT}$ (s)	NSE	PI	HL	$\overline{CT}$ (s)
`SCG`	0.935	0.57	22	69	0.928	0.55	7	29
`SCG2`	0.918	0.56	19	70	0.927	0.55	8	28
`BR`	0.920	0.54	85	13	0.928	0.55	8	12
`LM`	0.918	0.58	4	7	0.929	0.55	11	6
`CGB`	0.875	0.49	12	0.7	0.873	0.24	2	0.6
`OSS`	0.916	0.54	13	129	0.926	0.53	24	59

The Tuning Parameters Learning Rate (LR) and Momentum (`Mom`) are the tuning parameters in Gradient Descent (GD) algorithms. We set their values of these two optimisation parameters in advance. The LR parameter controls how fast or slow we will advance towards the optimal weights solution. If the LR is very large, we can skip an excellent solution. If too small, we may require too many iterations to converge to the finest values. Then, adopting a competent LR is decisive in gaining computational efficiency. The μ term helps to avoid local minima points, so we find a more important global minimum. If is too high, the optimisation is not stable and won't converge. Maier and Dandy (2000) noted that the momentum term must be less than 1.0 for convergence, being between the numeric range $0 < \mu < 1$.

We realise an illustrative grid search approach in this practice, performing multiple simulations for the GF1-PS4A regression experience problem (Sec.4.4.1). For this, we adopted particular values of LR and Mom, assuming the Scaled Conjugated Gradient for the parameters optimisation method. Finally, we determine 8 different values of LR and Mom: 0.001,0.01,0.1,0.185,0.2,0.5,0.75 and 0.9. The resulting simulations were in

total$\mathsf{LR}(8) \cdot \mathsf{Mom}(8) \cdot 1500 = 96000$. Fig.A.6 shows the results per each `LR` and `Mom` pair combination in efficiency terms (NSE). Also, we plot the vertical and horizontal largest (black dashed line), to find the excellent values. We mark the maximun value (max(NSE)) and the NSE 75th percentile of the multiple simulations for each combination. We detect minimal changes and fluctuation from the vast amount of simulations undertaken: 96000. We pay attention to the following numerical ranges: $0.9327 \leq \max(\text{NSE}) \leq 0.9400$ and $0.9223 \leq 75\text{th}(\text{NSE}) \leq 0.9259$. Because of the minor differences got in the simulations, and to identify better the changes in the evaluation coefficient (NSE), the values are standarised between the peak value and the minimal (NSE_{norm}).

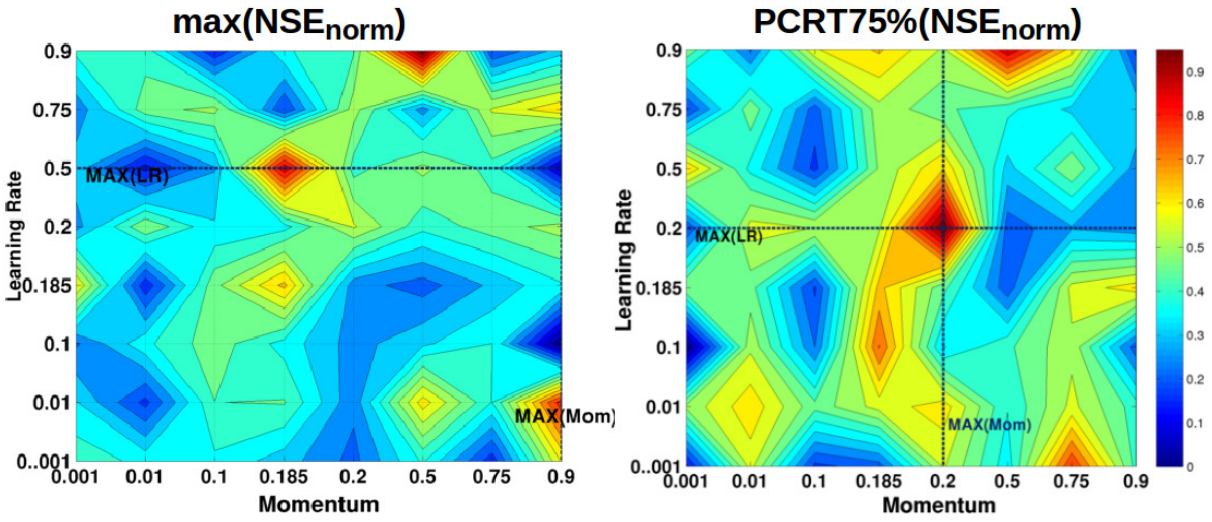


FIGURE A.6: (*Left*): The maximum efficiency value obtained. *Right*): The 75th percentile efficiency value obtained.

The above illustrative example shows that the best learning parameters using the grid search approach for the GF1-PS4A dataset is 0.2, for both, Learning Rate and Momentum. As a general rule, this value is applied throughout the work since the differences found in the above sensitivity analysis are minimal, where balanced values such as those obtained are taken as a best option.

Multiple simulations The optimisation of the $\mathcal{ANN}$s parameters (θ) through a Gradient Descent algorithm is not straightforward due the presence of local minima solutions. For that reason, it is recommended to perform random multiple simulations (Toth and Brath, 2007; Napolitano, 2011). The weights are randomly initialised from a normal distribution $\theta_{init} \sim N(0, 1)$, where the initial weights are obtained through a back-propagation (BP) algorithm, from layer to layer, to minimise the squared error function. This randomness during the θ initialisation, will offer dissimilar paths within the solution space,

and therefore, different parametrisations, models and behaviours. These differences can be a rich source of knowledge used for stochastic purposes. Finally, after having obtained several local minimum solutions, we evaluate the results on untested data, identifying the optimal or global minimum most suitable for our purposes. This search approach help us to identify statistically significant differences between different assumptions: inputs, architectures, parameters, functions, algorithms, etc.

Then, in this thesis, the multiple runs are used in two different ways:

1. A *static approach* on the number of neurons within the hidden layer (HL), or fixed values of parameters which are defined a priori. For regression cases, a standard for $\mathcal{ANN}$s' application is: $\{N - N(\mathtt{HL}) - 1\}$, where N is the units of the input layer (In), which is equal to the number of inputs or the elements of the PS, and N(HL) is the number of hidden units in the hidden layer, which is variable. The output layer will always have one neuron $m = 1$ for the output layer with the same dimensions as the target variable/vector. During the Input Variable Selection (IVS) procedure, the HL value for the *Nlin* test is equal to the number of inputs ($N = \mathtt{HL}$). More details of this method refer to Sec.4.2.1. In this manner, we standardized the model in order to eliminate extra variability during IVS.

2. A *dynamic approach*, where the HL value increases in order to find the " best" model or set of parameters. The data has been tested x times for each trial to prevent local minimum during the learning process, with an increasing number of neurons within the hidden layer (HL=$\{1, 2, \ldots, 25, 30, 40, 50, 60, 85\}$) for the best PS, launching in total 1500 runs with training and validation periods. After that, in order to visualise the overall behaviour related with the hidden neurons, a Generalise Additive Model GAM, (Wood, 2008) is fitted to the evaluation coefficients criterion obtained, generating the **learning trend** plot. These learning trends are plotted in figures 4.13, A.5, A.8, A.9, A.10, A.12, B.10, B.12, B.13 B.14, B.15, B.16 and B.19 under different conditions. By visualising these learning trends for each specific case, we can identify easily the optimal number of neurons region, models behaviours and minor differences among competitive predictors by data processing.

Once we have the results for a set of predictions, we have two choices depending on the objectives:

- The predictor that maximises/minimises the performance metrics based on the best values,

- The predictor that maximises/minimises the performance metrics based on the 25th/75th percentile statistic. This is option is used for the selection of the predictors during this work.

In this step, we do not recommend to use statistics such as the median. We propose to focus on statistics of extremes, either higher (maximising performance) or lower (minimising performance), depending on the performance metric.

The Bayesian Framework We define Bayesian Neural Networks ($\mathcal{BNN}$) as an Artificial Neural Network under the Bayesian regularisation framework expounded by MacKay (1992a). This practical application of Bayes' theorem transforms the "black box" nature of traditional $\mathcal{ANN}$s, into a more probabilistic or "grey nature" approach by implementing statistical information. This allow us to generate error bars for each output, i.e., the uncertainty quantification. This value is from a posterior distribution density $P(\theta \mid \mathcal{D})$ of the MLP' weights θ, see in Sec.A.1.2. In hydrology, Kingston et al. (2005); Zhang et al. (2009) analysed this improvement in a deep way, and in Khan and Coulibaly (2006) and Ticlavilca et al. (2011) tested successfully for different purposes.

$$f_{\mathcal{BNN}} \rightarrow y_{(t)} \;=\; f(\, x_{(t)}\, ; \theta\,) + \varepsilon^*_{(t)} \tag{A.3}$$

It is assumed that the target values are generated by the Eq.A.3 where $f(x, \theta)$ is an unknown function and ε^* is an independent Gaussian noise (MacKay, 1992b; Bishop, 2006; Khan and Coulibaly, 2006), $y_{(t)}$ is the vector of model output, $x_{(t)}$ the vector of inputs, and θ is the vector of parameters (weights and bias). All related to $f(\cdot)$ plus the vector of model error bars $\varepsilon^*_{(t)}$ for each prediction at the period time t. We use the code developed by Nabney (2004) in our research.

As commented, in $\mathcal{BNN}$ we have two types of parameters θ: the weights θ_w of each connection between the neurons, and an independent bias θ_b associated to each neuron, see Fig.A.4. However, within the $\mathcal{BNN}$ framework two more parameters are defined, α_h y β_h, that by convention are called hyper-parameters (MacKay, 1992b; Bishop, 1995; Nabney, 2004) since they control the distribution of the other network's parameters. Thus, $P(\theta)$ is the prior distribution of the parameters for the data observed during the training period, being a fixed set. We use Bayes' rule to calculate the conditional posterior probability ($P(\theta)$, see Eq.3.4) of the parameters for a given dataset D. Then, we have the $P(D|\theta)$ which is the probability of the dataset given this parameters, and $P(D)$ is the evidence term, which ensures that the posterior integrates to 1.

$$P(y|x^*, D) = \int P(y|x^*, \theta)P(\theta|D)\ \partial\theta \tag{A.4}$$

We calculate the posterior calculation values by the evidence approximation. This procedure is only an approximation as we are not integrating the hyper-parameters, and is done through an iterative optimisation procedure. The alternative Bayesian method practiced in this work, Gaussian Processes, integrates the hyper-parameters by involving weight posterior distribution. As is consistent, this presents higher computational costs. Then, during the evidence procedure (EV), we adjust the hyper-parameters taking into account the training data. $P(\theta|D)$ is the posterior distribution of the network weights, and following Nabney (2004) can be written as,

$$\begin{aligned} P(\theta|D) &= \int\int P(\theta, \alpha_h, \beta_h|D)\ \partial\alpha_h\ \partial\beta_h \\ &= \int\int P(\theta|\alpha_h, \beta_h, D)P(\alpha_h, \beta_h|D)\ \partial\alpha_h\ \partial\beta_h \end{aligned} \tag{A.5}$$

For a practical use, if $\alpha_h << \beta_h$ the training algorithm will lead to minor errors. If $\alpha_h >> \beta_h$, the model will emphasize the reduction of the weights magnitude, that means we get greater errors, and the BNN will have a smoother response. The hyper-parameters represents the inverse value of the weights variance. Using small values means that we avoid large weights magnitudes by giving more relevance to the quality of the input data. On the contrary, high values means that the weights will be close to zero. To visualise this, in Fig.A.7 a graphical example of the $\mathcal{BNN}$'s hyper-parameters convergence through the evidence procedure is represented. Initial values of $\alpha_{h0} = 0.05$ and $\beta_{h0} = 50$ are set, then after 7 iterations it is stopped. As can be seen, the hyper-parameters reached the convergence at the fourth epoch with posterior values of $\alpha_h \approx 0.862$ and $\beta_h \approx 872.2$ remaining constant for the next epochs (iterations). Finally we used five iterations for the evidence procedure during this work.

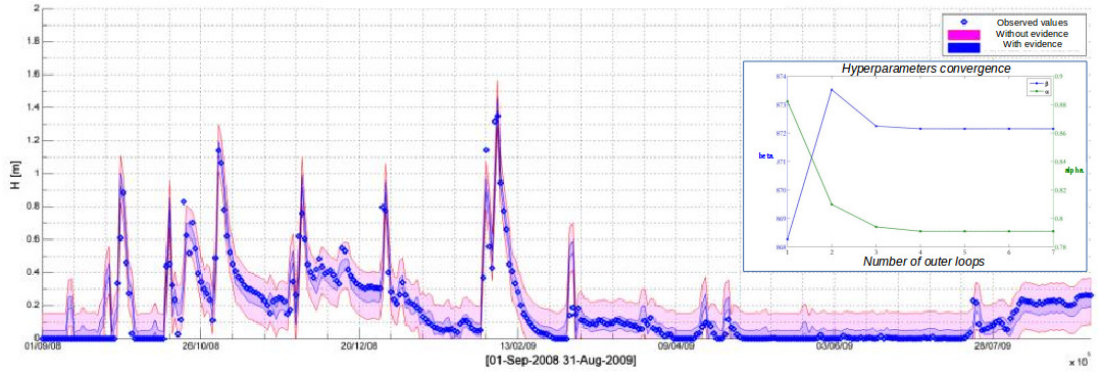


FIGURE A.7: $\mathcal{BNN}$ uncertainty comparison for the GH1 case without evidence (pink) and with the evidence procedure (blue)

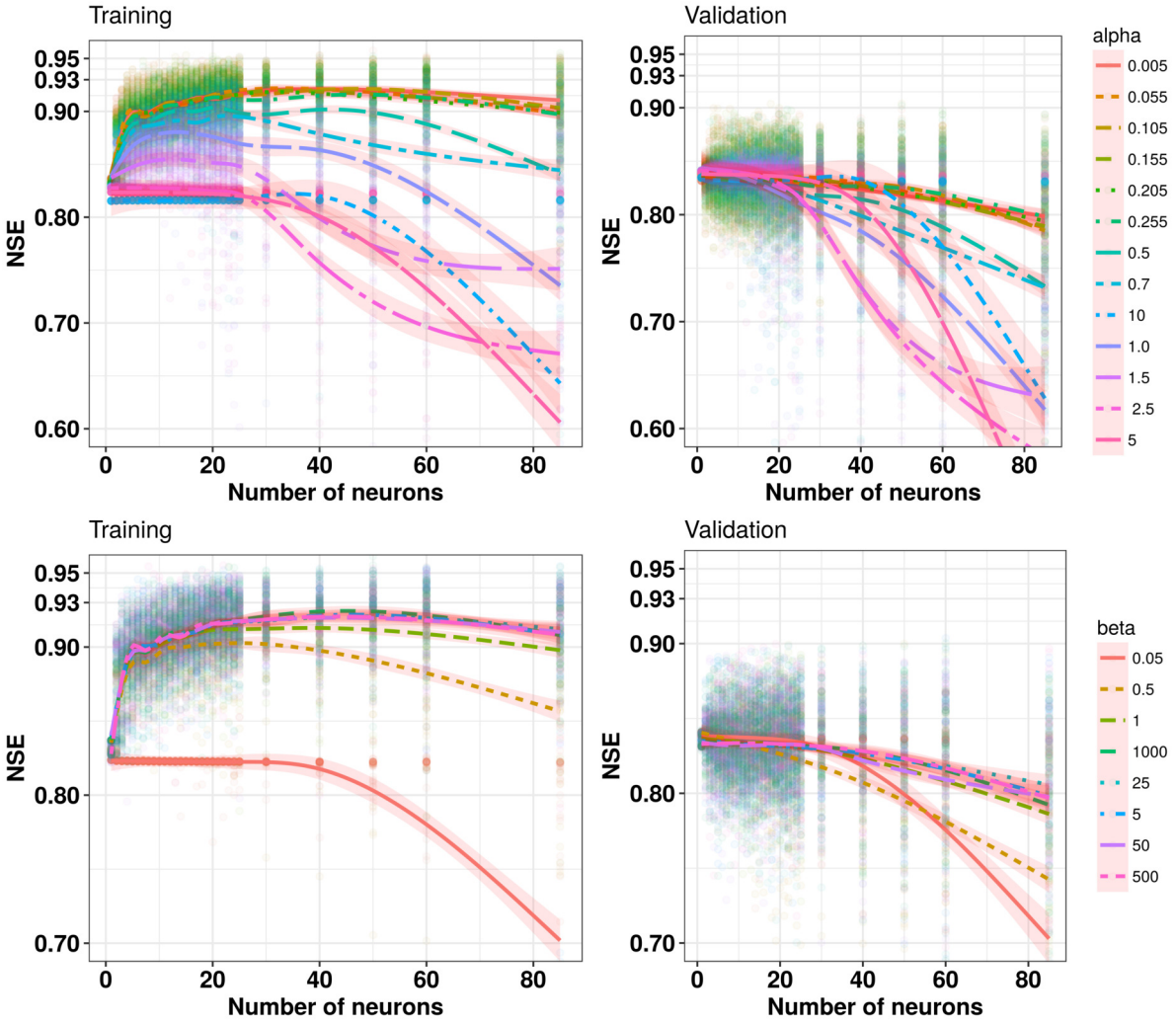
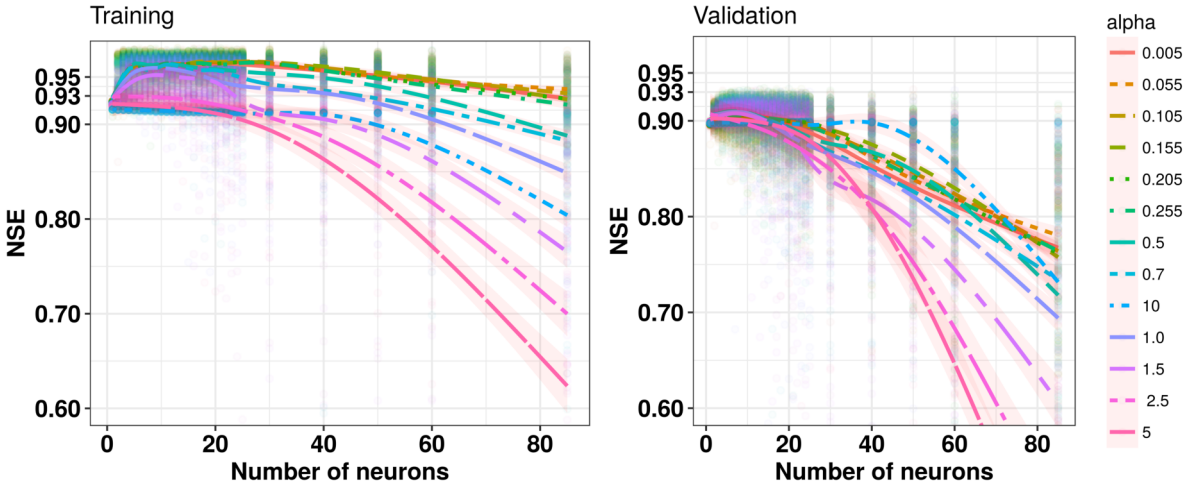
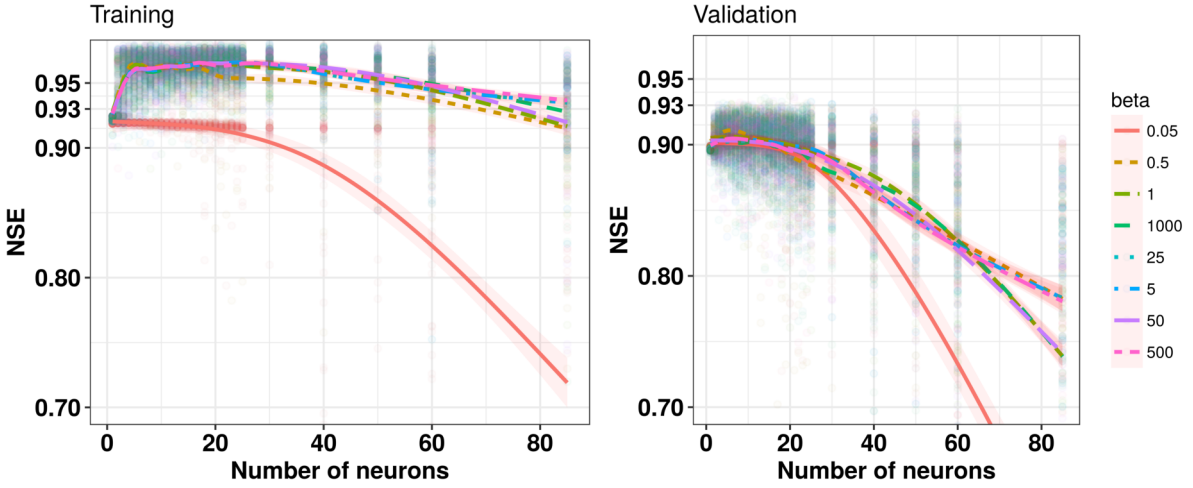


FIGURE A.8: $\mathcal{BNN}$ hyper-parameters α_h and β_h sensitivity analysis carried out in the GH3 case study.

The optimal posterior hyper-parameters estimation is realised during the **evidence** procedure, which is an iterative step. The computation cost of this approximation proposed by Nabney (2004), is lower than one purely Bayesian method as the Monte Carlo integration. In this work, we applied it to find the optimum hyper-parameters where we will reduce the uncertainty of the predictions from new data observed, and so the width interval value (AW coefficient). This may sometimes lead to a decrease in the value of the POC coefficient, where we must seek balanced outcomes between these two coefficients. We can find more on these coefficients in Sec.3.4. In the following figures (from Fig.A.8 to Fig.A.10), we show the grid search results and the learning trends obtained from it at a daily basis (chapter 4). Through these analyses, we confirm the general rule given in Nabney (2004): " It is a good practice to start with relatively small α values so that the model is allowed sufficient flexibility to fit data ".

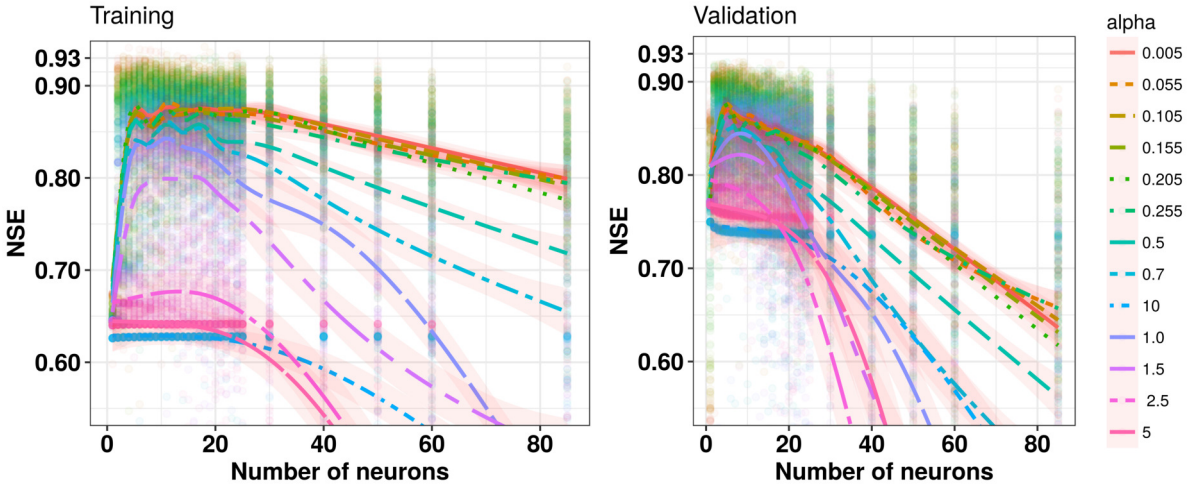


(a) Results for $h = 0$, $h = 1$ and $h = 2$ days ahead forecast during the test period (n=1098 *days*).

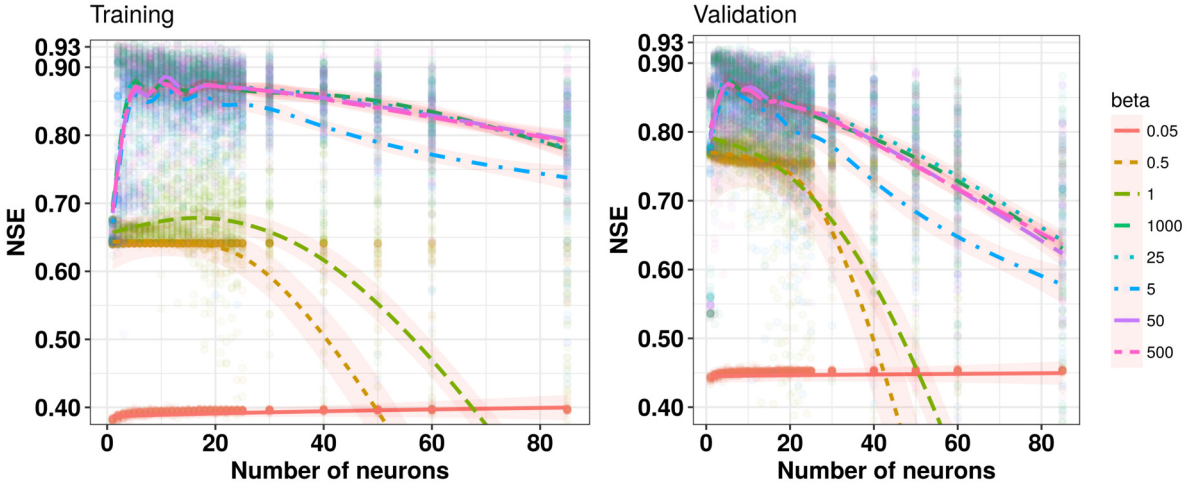


(b) Results for $h = 0$, $h = 1$ and $h = 2$ days ahead forecast during the test period (n=725 *days*).

FIGURE A.9: Learning trends from the BNN hyper-parameters (α_h,β_h) grid search carried out for the GF1 case experience (Sec.4.4.1)



(a) Results for $h = 0$, $h = 1$ and $h = 2$ days ahead forecast during the test period (n=1098 *days*).



(b) Results for $h = 0$ day ahead forecast during the test period (n=725 *days*)

FIGURE A.10: Learning trends from the BNN hyper-parameters (α_h,β_h) grid search carried out in the GF2 case experience (Sec.4.4.2)

• **Improvements to avoid biased results** Usually in $\mathcal{ANN}$ applications, problems of over-fitting and local minima solutions appear during the optimisation of the [θ_w, θ_b] parameters through the Gradient Descent algorithms. The traditional manner to cope this, is by adding a single regularisation term (α). In this work we adopted the $\mathcal{BNN}$ framework (Nabney, 2004), which has two regularisation terms called the hyper-parameters. Again, under the Bayesian regularisation of the $\mathcal{BNN}$ technique over-fitting is avoided. Other common approach to avoid over-fitting problems in $\mathcal{ANN}$s, is by constraining the parameters optimisation procedure by early stopping (Sarle, 1996; Dawson and Wilby, 2001; Maier et al., 2010). This approach can be found in practical RR applications in

several works (Luk et al., 2000; Coulibaly et al., 2000; Lin and Chen, 2004). The early stopping approach is based in the theoretical behaviour of the learning curve presented in Fig.A.11, where the prediction accuracy and the generalisation properties of the model may start to decrease for the validation period, as epochs of the training process go further. This is the optimum number of epochs with which the model training period should stop in order to prevent over-fitting problems.

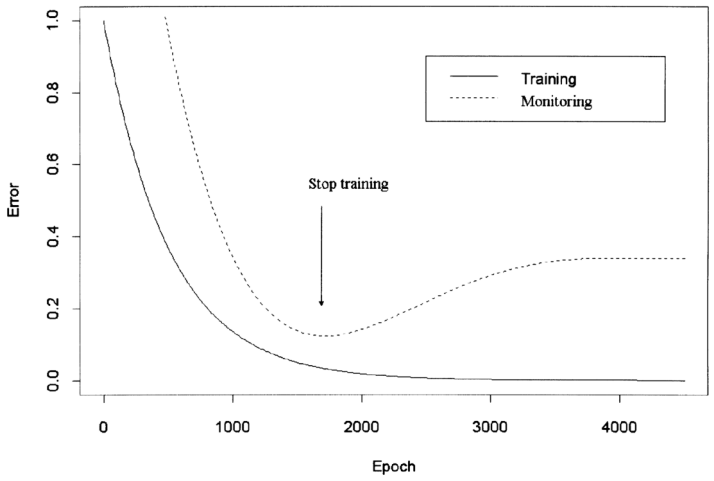


FIGURE A.11: Theoretical behaviour of the curve in ANNs applications. *Source:* Luk et al. (2000)

In Fig.A.12 we plot training and validation evaluation coefficients by NSE and RMSE for two time horizons for the GF1-PS4A model (Sec.4.4.1). Note that the $x - axis$ is sorter in ascending order: higher values epochs produce less flexible $\mathcal{ANN}$ models and lower training errors. Then, we can visualise the learning curve for the GF1 study case for two time horizons and 2 different number of epochs for an increasing number of neurons. In this visualization we can appreciate how the models with early stopping, have a poorer generalisation skills. This is because this technique $\mathcal{BNN}$ is already regulated, and the models are robust to a high number of iterations

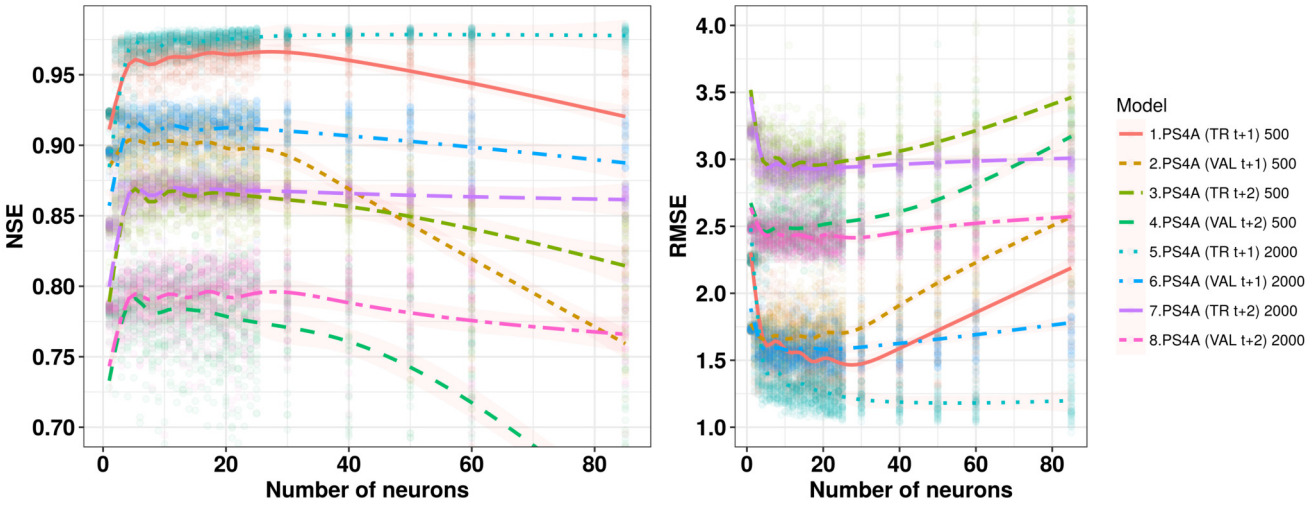


FIGURE A.12: NSE and RMSE learning trends comparison for different number of iterations (500, 2000) in the GF1 case. TR and VAL, training and validation period respectively

In model-based studies, we do an exhaustive search to avoid biased outcomes comparing the different configurations by trial and error. As explained before, we can avoid local minima solutions by generating multiple models and runs with the multiple $\mathcal{ANN}$s simulations approach. The several random candidate members generated through this approach are used to build (construct) other modelling methods as the ensembles, which are explained in a following section (App.C.1).

In principle, over-fitting is not a problem as the networks are regularised to prevent weights becoming too large by the Bayesian hyper-parameters. For that reason, making extra epochs during the training process is a problem, where, as we have visualized, the network remains robust to over-fitting. As we saw in this chapter, models optimised with less epochs and a higher number of units within `HL`, trend to generalises poorer than with more epochs.

A.1.3 Gaussian Processes

In Gaussian Processes ($\mathcal{GP}$) models, the observations are related to changing inputs variables (covariates), considering they are conditionally independent given a predictor and a latent function $f(\cdot)$. Dealing with a simple regression problem for a given $f(x)$ (Eq.4.1) a $\mathcal{GP}$ has an error term ϵ, which is normally distributed with a zero mean and σ^2 variance: $N \sim (0, \sigma^2)$. The key strengths contrasted to the $\mathcal{BNN}$s (Sec.) are: 1) the probability for each state are factored even if is not reflected in the training stage,

and 2) the Bayesian inference is explicit and non-parametrical (by covariance functions). Formerly, a $\mathcal{GP}$ model is driven by the accompanying expression,

$$f(x) \mid \theta_{\mathcal{GP}} \sim \mathcal{GP}(m(x), k(x_i, x_j \mid \theta_{\mathcal{GP}})) \tag{A.6}$$

where $m(x)$ and $k(x_i, x_j)$ are the mean and the kernel covariance function (KCF) of f respectively, with its hyper-parameters $\theta_{\mathcal{GP}}$ encoding our assumptions about the relationship between the observations and the outputs that generated the data. Then the $\mathcal{GP}$ prior has a multivariate Gaussian distribution as below,

$$p(f \mid X, \theta_{\mathcal{GP}}) \sim N(f \mid 0, [K_{f,f}]_{i,j}) \tag{A.7}$$

where the covariance matrix is shaped by the covariance function, $[K_{f,f}]_{i,j} = k(x_i, x_j|\theta_{\mathcal{GP}})$, which define the correlation between different inputs in the process. The hyper-parameters, $\theta_{\mathcal{GP}} = [\sigma^2, l_1, \ldots, l_d]$ has two terms basically: the term l_d which is the length scale that governs the correlation scale in input dimension d, and σ^2 which represents the overall variability of the process (Vanhatalo et al., 2013). The inverse of l_d tells how significant an input is (Rasmussen, 2006), which means large values of l, less influence on the model. After determining the posterior $(f*)$ of f through the training period, we generate the predictive distribution of any new data $(x*)$,

$$p(f* \mid x*, y, \theta_{\mathcal{GP}}, \sigma^2) \tag{A.8}$$

The key point in the $\mathcal{GP}$s is the use of the covariance function. In doing so, we ensure that the $\mathcal{GP}$s will model properly the variance between the observations and the target value through a Gaussian distribution. Different kernel covariance functions are available (see further ideas) with which we can set more complex models (i.e., kernel structures) for any particular case, for instances assigning covariance kernels to a specific array of predictors, or by the addition of single kernels, $k_A(x_i, x_j) = k_1(x_i, x_j) + k_2(x_i, x_j)$, or using a dot product kernel, $k_P(x_i, x_j) = k_1(x_i, x_j) \cdot k_2(x_i, x_j)$.

An essential motive to combine kernels from diverse natures, is the fundamental that just one cannot represent the underlying function. In future work, it be worthwhile to explore multiple KCF by adding and multiplying kernels for the $\mathcal{GP}$ technique. For an extended

explanation of the $\mathcal{GP}$s foundations refer to Rasmussen (2006), and its operation on time-series studies go to Roberts et al. (2012). In our performance, we used the code promoted by Vanhatalo et al. (2013) for the $\mathcal{GP}$s application.

Covariance kernel functions In the $\mathcal{GP}$ models we set the Gaussian distribution as the likelihood function, and the maximum a posterior (MAP) estimate to approximate the θ parameters values (Eq.A.1.2), which regulates the covariance between the inputs and the outputs (Eq.A.6). The high computational cost of this technique force us to have prior knowledge of the different Kernels Covariance Functions (KCF) behaviour. We can discover a comprehensive theoretical information and comparatives in Rasmussen (2006) and Vanhatalo et al. (2013).

***Single* KCF**: The most common base KCFs available are: linear (k_{Lin}, Eq. A.9), the Neural Network (k_{NN}, Eq. A.10), Squared Exponential (k_{SE}, Eq. A.12), and Matern (k_{Mat}, Eq. A.13) which are tested in this work. Among them, the k_{SE} is probably the most used KCF nowadays (Rasouli et al., 2012; Grbić et al., 2013; Sun et al., 2014) for hydrological applications. On the other hand, the k_{NN} is a useful kernel to introduce the non-stationary component into the modelling process (Bishop, 2006; Rasmussen, 2006). The expressions of the KCFs are listed below:

$$k(x_i, x_j)_{Lin} = x_i^T \sum x_j \tag{A.9}$$

$$k(x_i, x_j)_{NN} = \frac{2}{\pi} \sin^{-1} \left(\frac{2\tilde{x}_i^T \sum \tilde{x}_j}{\sqrt{(1 + 2\tilde{x}_i^T \sum \tilde{x}_i)(1 + 2\tilde{x}_j^T \sum \tilde{x}_j)}} \right) \tag{A.10}$$

where $\tilde{x} = (1, \ldots, x_d)^T$ is an augmented input vector. The following KCF are driven by P, defined as

$$P = \sum_{k=1}^{d} \frac{(x_{i,k} - x_{j,k})^2}{l_d^2} \tag{A.11}$$

which is the euclidean distance governed by the length scale, used in the Squared Exponential and Matern KCF, which are listed below,

$$k(x_i, x_j)_{SE} = \sigma_{s\exp}^2 \exp\left(-\frac{1}{2} P\right) \tag{A.12}$$

TABLE A.4: Single Kernel Covariance Functions (KCF) used in this work

Name	Symbol	Expresion
Linear	k_{Lin}	Eq A.9
Neural Network	k_{NN}	Eq A.10
Squared Exponential	k_{SE}	Eq A.12
Matern	k_{Mat}	Eq A.13, A.14

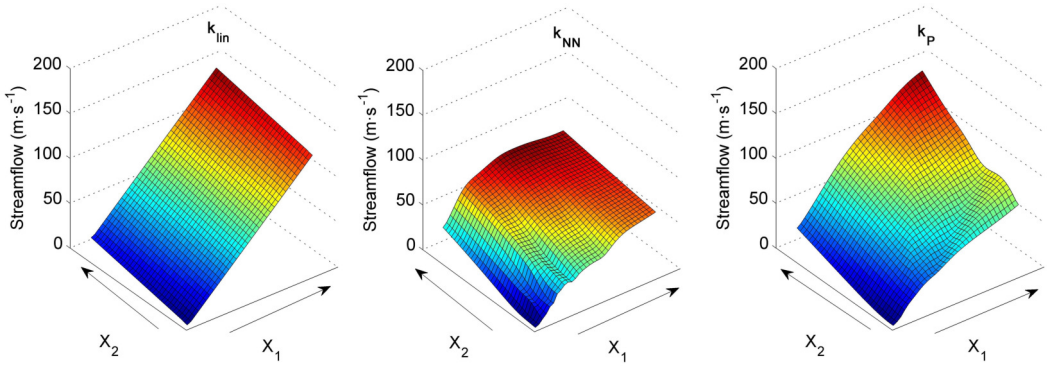
$$k\left(x_i, x_j\right)_{Mat} = \sigma_m^2 \frac{2^1}{\Gamma(\nu)} (\sqrt{2\nu P})^{\nu} K_v(\sqrt{2\nu P}) \tag{A.13}$$

For the Matern KCF, K_ν is a modified Bessel function and the paremeter ν for machine learning normally is chosen as $\nu = 3/2$, obtaining k_{Mat} in a simpler form:

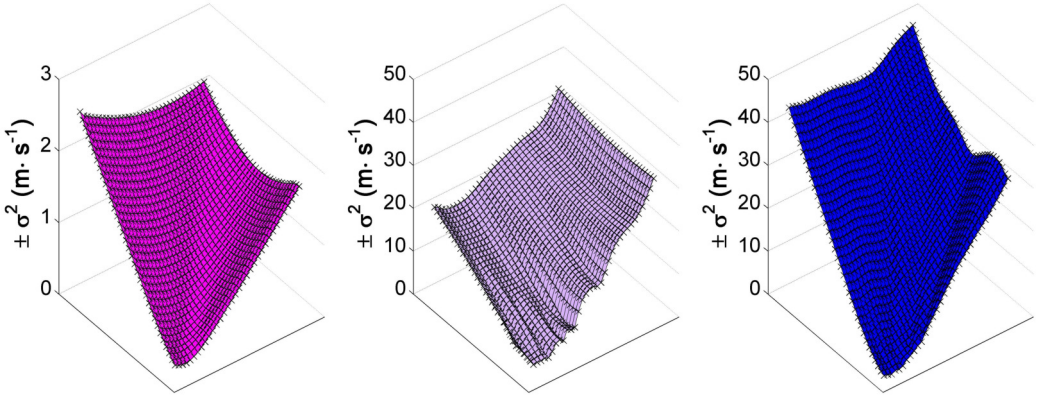
$$k\left(x_i, x_j\right)_{Mat\frac{3}{2}} = (1 + \sqrt{3P}) \exp(-\sqrt{3P}) \tag{A.14}$$

***Multiple* KCF** $\mathcal{GP}$ is a very versatile computational technique due the possibility to build multiple kernels composed from various single kernels. This will help us to deal with complex phenomena and fit better the underlying function. In that sense, the product ($k_{fU} = k_1 \cdot k_2 \cdot \ldots k_U$) and additive kernels ($k_f = k_1 + k_2 + \ldots + k_U$) are the basic operations that can be done among kernels. Also specific kernels can be applied only to selected variables or to a given variable window $\{X_{(t-1)}, \ldots, X_{(t-p)}\}$ as a form of an additive model.

For a deeper analysis on the resulting models of this technique, the regression experiences of the $\mathcal{GP}$ resulting models within the target variable domain are graphically represented below. For this purpose, the real functions finally obtained will be displayed in order to see the final multivariate real domain. For that, the resulting mean and its variance ($\pm\sigma^2$) are represented. In order to visualise higher-dimensional spaces and drawn the real functions for the resulting $\mathcal{GP}$ models, each variable is grouped with the number of element equal to p, which is the maximum lag value. Then, we set a certain delta or gradient in each window. In this way we reproduce the model response to events with increasing variations equal to the gradient established. The initial PS has the following form,



(a) The real functions of the GP-P $\sim \mathcal{GP}\ (0, k_{Lin} \times k_{NN})$ with two input windows for the GF1-PS4A experience. Left and Middle the kernels used in the product, right the resulting kernel



(b) Variance (σ^2) from real function for each kernel

FIGURE A.13: Resulting $\mathcal{GP}$ real function for the GF1-PS4A regression experience.

$$\mathsf{SF}_{(t+h)} = f(\{\overbrace{\mathsf{SF}_{(t-1)}, \ldots, \mathsf{SF}_{(t-p)}}^{\text{Window X}_1}\}, \{\overbrace{R_{a_w(t-1)}, \ldots, R_{a_w(t-p)}}^{\text{Window X}_2}\}) \tag{A.15}$$

where the window X_1 and X_2 represents the antecedent values of the target variable (SF) and the rainfall values. As a result, the behaviour of each kernel used and its dot product can be visualised in its real domain during the observed period of the dataset. We can see in the figure, how the resulting kernel (k_p) has a linear component (k_{lin}), plus a non-stationary and a non-linear kernel (k_{NN}) for dynamic events.

Kernel Covariance Comparison In the following tables, we list the results of 5 candidate Kernel Covariance Functions (KCF) comparison for GF1 and GH1 experiences, developed in Sec.4.3.1 and Sec.4.4.1 respectively.

In Tab.A.5, the GH1, we see how the linear kernel offers the most balanced results, and no over-fitting problems. In all the others, we can observe clearly over-fitting issues, to a greater or lesser extent. Only in some of them, the value obtained during the validation period is acceptable in forecasting terms, since it is above the persist method. It should be highlighted the bad behaviour of k_{SE} in terms of persistency (PI), showing the need for a more profound study during the training step for this kernel. Finally for this case, GH1, the k_{lin} has been chosen for comparison with the other techniques. On the other hand, in Tab.A.6 we observe a similar behaviour. However, in this case the validation values are higher, being quite acceptable values from a forecasting perspective, for most of the kernels and for the first time horizon. If we increase the time horizon, the over-fitting difficulties increase considerably, decreasing the efficiency during the validation period. Finally for this case study, the GF1, k_P has been chosen as the best kernel for comparison with the other computational techniques.

TABLE A.5: List of GH1-PS3B-$\mathcal{GP}$ tested with different Covariance Kernel Function (CKF→ k) evaluated training (n=4391) and validation (n=1098) period respectively → T; V. Absolute error metrics (RMSE, MAE) are expressed in their original unit measures (m). The best KCF result in each time horizon is highlighted in bold font.

CKF	h	NSE	RMSE	PI	MAE
k_{lin}	0	0.850;**0.876**	0.173;**0.191**	0.35;**0.47**	0.076;**0.074**
	1	0.801;**0.836**	0.199;**0.219**	0.48;**0.55**	0.094;**0.093**
	2	0.638;**0.684**	0.269;**0.304**	0.29;**0.29**	0.121;**0.131**
k_{NN}	0	0.979;0.848	0.064;0.211	0.91;0.35	0.023;0.074
	1	0.967;0.807	0.081;0.238	0.91;0.47	0.033;0.094
	2	0.960;0.622	0.090;0.333	0.92;0.16	0.039;0.141
k_{SE}	0	0.971;0.503	0.076;0.382	0.87;-1.13	0.031;0.117
	1	0.938;0.460	0.112;0.398	0.84;-0.47	0.049;0.128
	2	0.921;0.239	0.126;0.472	0.84;-0.70	0.057;0.188
k_{Mat}	0	0.976;0.787	0.070;0.249	0.89;0.09	0.026;0.081
	1	0.958;0.696	0.092;0.298	0.89;0.17	0.039;0.110
	2	0.945;0.359	0.105;0.433	0.89;-0.43	0.046;0.173
k_P	0	0.979;0.853	0.065;0.207	0.91;0.37	0.023;0.070
	1	0.964;0.805	0.085;0.239	0.91;0.47	0.035;0.093
	2	0.949;0.613	0.101;0.336	0.90;0.14	0.043;0.144

TABLE A.6: List of GF1-PS4B-GP tested with different Kernel Covariance Function (k) evaluated at training (n=2898) and validation (n=725) period respectively → T; V. Absolute error metrics (RMSE, MAE) are expressed in their original unit measures ($m^3 s^{-1}$). The best KCF result in each time horizon is highlighted in bold font.

KCF	h	NSE	RMSE	PI	MAE
k_{lin}	0	0.915;0.901	2.352;1.684	0.36;0.34	0.809;0.825
	1	0.839;0.804	3.247;2.368	0.30;0.27	1.048;1.076
	2	0.753;**0.670**	4.016;**3.076**	0.24;**0.18**	1.358;**1.391**
k_{NN}	0	0.999;0.899	0.188;1.703	1.00;0.32	0.109;0.770
	1	0.999;0.754	0.196;2.655	1.00;0.09	0.110;1.108
	2	0.966;0.404	1.481;4.130	0.90;-0.48	0.612;1.473
k_{SE}	0	0.996;0.826	0.497;2.235	0.97;-0.17	0.287;0.815
	1	0.997;0.491	0.456;3.817	0.99;-0.89	0.260;1.371
	2	0.966;0.491	1.491;3.819	0.90;-0.27	0.604;1.491
k_{Mat}	0	0.999;0.896	0.196;1.726	1.00;0.30	0.111;0.779
	1	1.000;0.697	0.179;2.944	1.00;-0.12	0.098;1.159
	2	0.999;0.508	0.304;3.752	1.00;-0.22	0.147;1.489
k_P	0	0.998;**0.928**	0.315;**1.436**	0.99;**0.52**	0.181;**0.685**
	1	0.999;**0.811**	0.291;**2.328**	0.99;**0.30**	0.157;**1.045**
	2	0.998;0.598	0.339;3.393	0.99;0.00	0.151;1.470

A.1.4 Classifiers

In this section we describe the fundamentals of the three classification methods (classifiers) in its binary form, being used in Sec.5.2:

- **Probabilistic Neural Network ($\mathcal{PNN}$)** The $\mathcal{PNN}$ (Wasserman, 1993) is a particular MLP which is composed of four layers (see FigA.14). The input layer contains the predictors, having as many nodes as inputs. The output layer contains the target class, having just one single node. Between layers, there are two hidden layers: a pattern layer, that has many neurons as observations during the training period ($T = 25$). Here, we compute for each node the Gaussian of the Euclidean distance between the given input vector and one training input vector. And finally the Summation layer, that has many neurons as there are input classes (K), in our case two: C_1 and C_2. In each node we sum the outputs of the pattern layer that belongs to each class k.

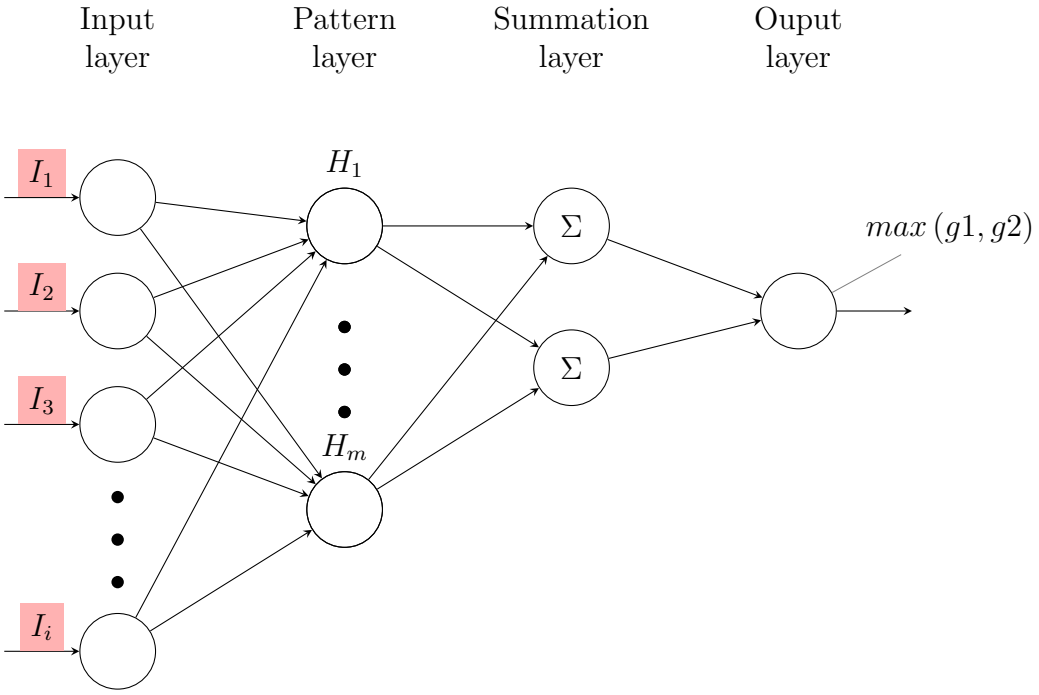


FIGURE A.14: Probabilistic Neural Network architecture used for estimating if the cumulative streamflow value of further time horizons are (or not) higher than the historical observed mean

In this technique, unlike the one used for regression, no parameters are assigned to the connections, but rather probability functions (*pdf*). For this reason, we can classify it as a non-parametric technique (Sec.3.2.2). One of its strengths is that it brings classification robustness in domains with noisy data. Basically, when we apply this classifier, the main setting is to define the optimum value of the δ parameter. The optimum value should be estimated by trial and error or by using heuristic techniques (Yi et al., 2016). A small value of δ, i.e., 0.1, makes this technique behave just like as KNN. In order to simplify its adjustment and application, we will set 0.5 as a convenient value.

- **K-Nearest Neighbour (KNN)** This usual non-parametric classifier establishes new class cases based on a similarity measure, i.e., distances. Different distances can be used, where the generalised form is the Minkowski distance,

$$\left(\sum_{i=1}^{n} |x_i - y_i|^q\right)^{1/q} \tag{A.16}$$

which is the Euclidean distance if $q = 2$, and the Manhattan distance if $q = 1$.

Thinking mathematically, is the most simple method tested in this work, thus establishing a baseline in the comparative framework with the other more complex methods. As we have seen before, simple methods can achieve better results than more complex computational methods. In this work the standardized version of the Euclidean distance is used. Each coordinate difference between x_i and a query point y_i is divided by the standard deviation computed from x ($\sigma(x)$), omitting missing values.

For this DD classifier, the most important parameter that need to set out is K, which are the number of nearest points to take into account. In Enas and Choi (1986) some rough guidelines for the optimal choice of the K value are given. Based on the rules given in this paper, we have selected a mid-point, establishing the optimal k to be $n_T^{(2.5/8)}$. Where n_T is the number of observations used for the parametrisation during the training period.

A.2 Computational Resources

In this work different computational environments are used for data representation, data exploration, programming DB methods for our specific case studies and tools development in order to achieve the objectives set out previously in Sec 1.2. Concerning to Geographical Information Systems (GIS) environments we have used QGIS, an open source software, and ArcGIS as commercial software. In connection with the mathematical/scientific computation environments, mainly Matlab and R are used in this work, which are commercial and open source software approaches respectively. Currently the preferred option is to migrate to an environment like Phyton. The following list shows the specific toolboxes, packages and resources that are used through this work:

- The Netlab toolbox (Nabney, 2004) for Neural Networks and its Bayesian framework for regularisation and uncertainty quantification in Chap.4 and Chap.5.
- GPstuff for Gaussian Processes in Chap.4 and Chap.5.
- The Reservoir R package for DP and SDP in Chap.6.
- The spi R package for drought indexes computation: SPI and SRI.
- The Bayesian Model Averaging (BMA) toolbox (Vrugta, 2015), for data-driven ensemble implementation purposes (App.C).

Appendix B

Supplementary Material

B.1 Supplementary material of chapter 2

In Sec.2.4.2 a spatio-temporal analysis is performed using GIS technology. The Flood Risk spatial data were provided by the water authorities and is open data, as was required by the Flood Risk Directive (FRD). At a visual level, we have represented the spatial data for 2013 at the mouths of the two rivers under study: Fig.B.1 and Fig.B.2. In this way we can appreciate the level of detail for this spatial scale 1:10000. In table B.1 are listed some characteristics. Besides, the land use/land cover classes according to five classes are summarized in Tab.B.2.

TABLE B.1: Features of the floodplains simulations analysed in the lower part of the Guadalfeo River course

Assumption	Q($m^3 \cdot s^{-1}$)	L(Km)	Model	S(Km^2)	Study's purpose
T50	1390	19.4	MIKEFLOOD	39.9446	Reservoir operating rules
T100	1730	19.4	MIKEFLOOD	68.0661	Reservoir operating rules
T500	NA	11.6	HEC-RAS & HEC-HMS	109.3521	River Basin Management Plan
ARPSI				566.7720	Flood Risk Directive

TABLE B.2: Features of the risk floodplains analysed on the entire Guadalhorce River Basin. LULC: Land use/land cover, C_x: Classes and its corresponding land uses that has been grouped under this class.

C_x	**LULC (1:25000) codes**	**LULC (1:10000) codes**
1.Urban and civil infrastructure	111:193	101:194, 346, 936, 1700:2007
2.Water bodies	211:345	200:345, 2008
3.Irrigated agriculture	411:419, 441:449, 471:489	405, 416, 417, 428, 445, 460
4.Non-irrigated agriculture	421:139, 451:469	403, 404, 421, 431, 435, 961:997
5.Forest and scrub. Natural areas	510:935	510:935, 941:950, 1006, 1007

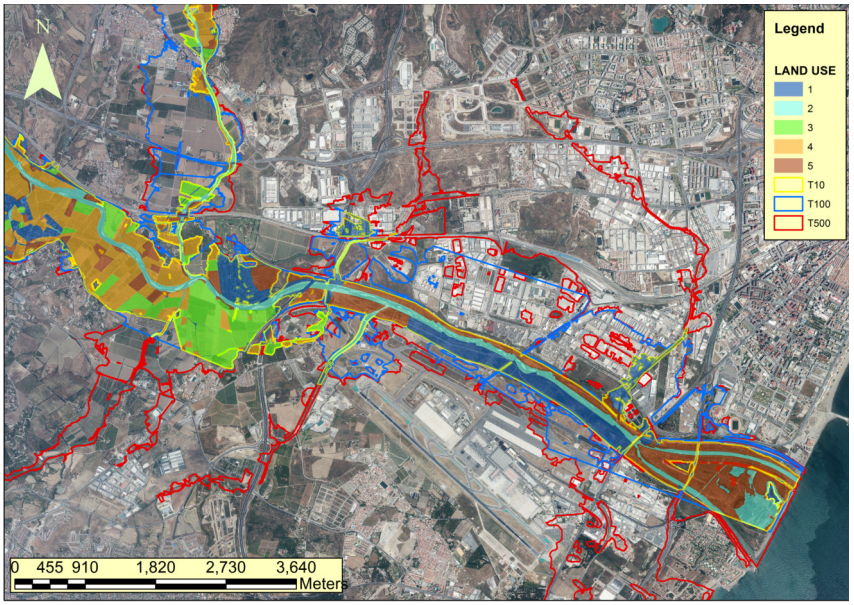


FIGURE B.1: Land Use and Land Cover (2013) visualisation at the Guadalhorce River mouth with the T50, T100 and T500 floodplains scenarios

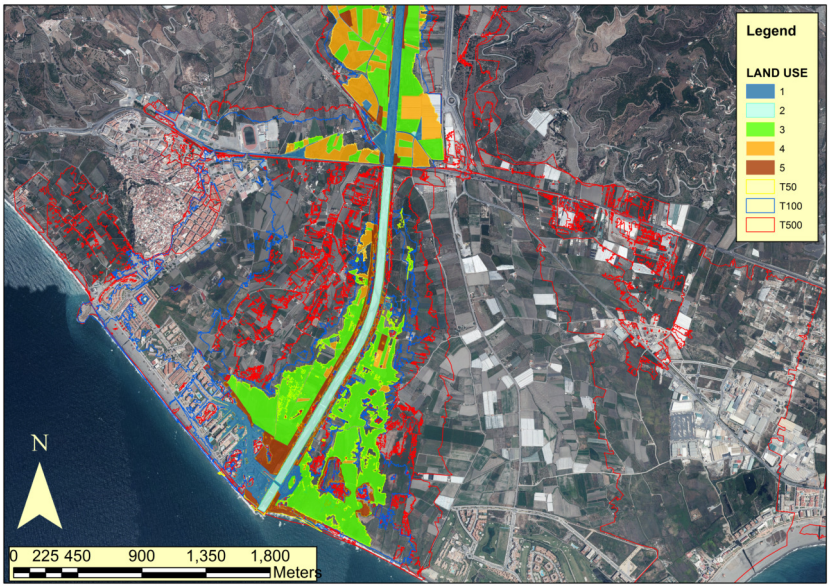


FIGURE B.2: Land Use and Land Cover (2013) visualisation at the Guadalfeo River mouth with the T10, T100 and T500 floodplains scenarios

B.2 Supplementary material of chapter 4

B.2.1 Datasets for Forecasting Experiences

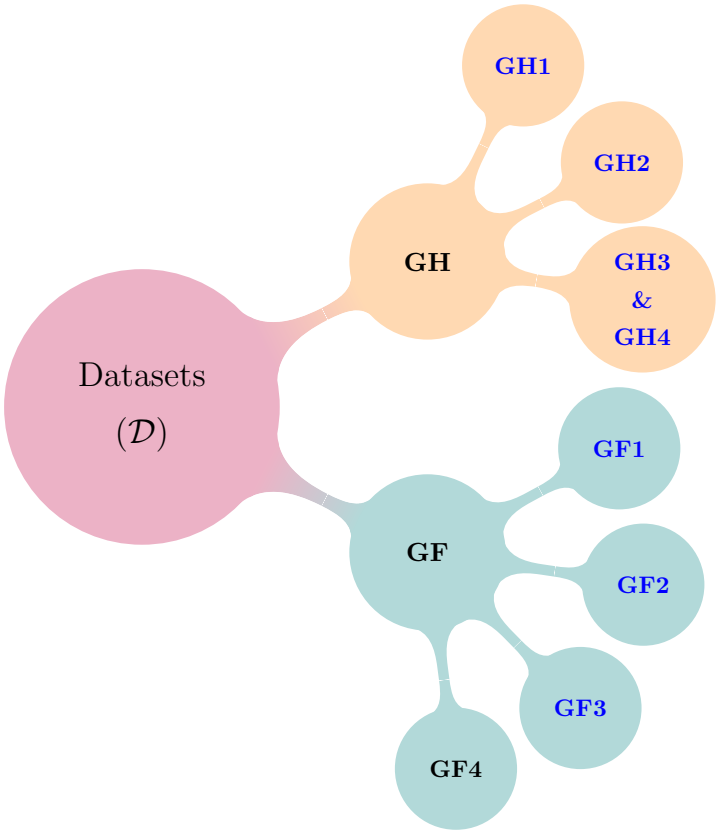


FIGURE B.3: Datasets used for short term forecasting

One of the main objectives of this work (Sec.1.2) is to gain knowledge on regression experiences on the study sites by several cases and locations with different conditions (purposes). For this, various Datasets ($\mathcal{D}$, see Fig.B.3) have been configured for each case depending on the application objective. They are composed of a number of time series of the candidate predictors and the time series of the target variable.

• **GH1** The target variable is water level height at Cartama ($\mathsf{SF}_{\mathsf{Ca}}$), which is located at the head of the floodplains (Fig.2.1). Hydrological, meteorological and Climatological data is used for this regression experience carried out in Sec.4.3.1. In table B.3, the statistical properties of the time series are listed.

TABLE B.3: Statistical parameters of the time series used for the GH1 case are: $Var.$ shows the variable abbreviation or symbol, $\bar{X}$ mean value, S_x Standart deviation, J shows number of candidate subsets

Time Serie	$\bar{X}$	S_x	Range	Unit	$Var.$	J
Water Height‡†	0.66	0.79	6.40	m	$\mathsf{SF}_{\mathtt{Ca}}$	1
Reservoir Releases ‡	5.48	12.07	164.35	m^3s^{-1}	Q	1
Rainfall$^{\ddagger}_{\bar{a}}$	1.65	6.54	97.67	mm	R	16
$\ddagger \rightarrow$ Data collected in the monitoring networks $\dagger \rightarrow$ Target variable						

• **GH2** The target variable is the water level height at Cartama ($\mathsf{WH}_{\mathtt{Ca}}$, m) at hourly scale, which is on the head of the floodplains in the lower part of the Guadalhorce river course (Fig.4.6). For this case (Sec.4.3.2), hydro-meteorological data is used in order to achieve the objectives planned.

TABLE B.4: Statistical parameters of the time series used for the GH2 case are: $Var.$ shows the variable abbreviation or symbol, $\bar{X}$ mean value, S_x Standart deviation, J shows number of candidate subsets

Time Serie	$\bar{X}$	S_x	Range	Unit	$Var.$	J
Water Height‡†	0.66	0.79	6.40	m	$\mathsf{WH}_{\mathtt{Ca}}$	1
Reservoir Releases ‡	5.48	12.07	0-164.35	m^3s^{-1}	Q	1
Rainfall$^{\ddagger}_{\bar{a}}$	1.65	6.54	0-38.2	mm	R	5
$\ddagger \rightarrow$ Data collected in the monitoring networks $\dagger \rightarrow$ Target Value						

• **GH3 & GH4** The target variables for this case are the Guadalhorce Reservoir System inflows (GRS), which is composed by the daily inflows of the Conde Guadalhorce reservoir $\mathsf{SF}_{\mathtt{Ard}}$ and on the other hand the sum of the inflows of the Guadalhorce-Guadalteba reservoir $\mathsf{SF}_{\mathtt{GRS-1/2}} = \mathsf{SF}_{\mathtt{Bob}} + \mathsf{SF}_{\mathtt{Teb}}$, expressed in $Hm^3 \cdot day^{-1}$. The time period is from 01/01/2000 to 30/09/2009. Also Pressure at m.s.l from reanalysis, and atmospheric oscillations (AO) time series are used as candidate predictors. The data locations are shown in Fig.4.10, and this modelling experiences results are presented in Sec.4.3.3 and Sec.4.3.4 for GH3 and GH4 cases respectively.

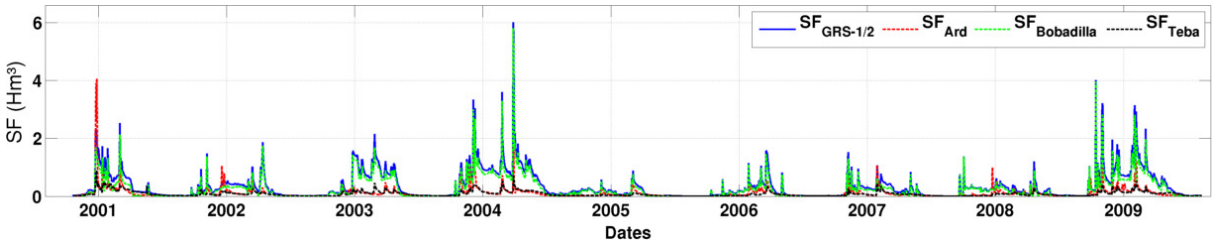


FIGURE B.4: The representation of the GRS inflows ($Hm^3 \cdot day^{-1}$). $\mathsf{SF}_{\mathtt{GRS}} = \mathsf{SF}_{\mathtt{Bob}} + \mathsf{SF}_{\mathtt{Teb}}$

TABLE B.5: Statistical parameters of the time series used for the GH3 and GH4 cases are: $Var.$ shows the variable abbreviation or symbol, $\bar{X}$ mean value, S_x Standart deviation, J shows number of candidate subsets

Time Serie	$\bar{X}$	S_x	Range	Units	$Var.$	J
$\mathtt{GRS}_{1/2}$ Inflow‡†	0.3321	0.4493	0-6.014	Hm^3day^{-1}	$\mathsf{SF}_{\mathtt{GRS}}$	1
$\mathtt{GRS}_{2/2}$ Inflow‡†	0.1012	0.1864	0-4.071	Hm^3day^{-1}	$\mathsf{SF}_{\mathtt{Ard}}$	1
Rainfall$^{\ddagger}_{\bar{a}}$	1.30	6.71	0-229.1	$mm \cdot day^{-1}$	R	23
‡ →Data collected in the monitoring networks, † → Target Value						

• **GF1** At the GF basin, the rainfall data sets are collected at 14 meteorological stations and the temperature at 6 meteorological stations, from 01/10/2003 to 31/08/2013. These time series are related with the reservoir inflow ($\mathsf{SF}_{\mathtt{Ru}}$) values observed at Rules control point (Fig.2.1) in Sec.4.4.1. The three variables modelled were total Precipitation (R_{Wi}, mm), Snowmelt (S, mm) and Snowfall (R_S, mm), which are used as streamflow predictors. From these variables the effective rainfall (R_e, mm) is estimated (Eq.B.1).

$$R_{e(t)} = R_{W_i(t)} - R_{S(t)} \tag{B.1}$$

TABLE B.6: Statistical parameters of the time series used for the GF1 case are: $Var.$ shows the variable abbreviation or symbol, $\bar{X}$ mean value, S_x Standart deviation, J shows number of candidate subsets

Time Serie	$Var.$	$\bar{X}$	S_x	Range	Units	J
Reservoir Inflow‡†	$\mathsf{SF}_{\mathtt{Ru}}$	5.28	7.62	0-145.14	m^3s^{-1}	1
Rainfall$^{\ddagger}_{\bar{a}}$	R	1.81	5.77	0-95	mm	16
Snowmelt^{W_i}	S	0.41	0.86	0-10.52	mm	1
Snowfall^{W_i}	R_S	0.58	1.98	0-25.30	mm	1
Precipitation^{W_i}	R_W	1.97	6.35	0-94.65	mm	1
$\ddagger \rightarrow$ Data collected in the monitoring networks $\dagger \rightarrow$ Target Value						
$W_i \rightarrow$ Time series reproduced by hydrological modelling (WiMMed)						

• **GF2** The target variable are the daily inflows at Beznar reservoir ($\mathsf{SF}_{\mathtt{Be}}$). The development and results of this experience is showed in Section 4.4.2.

TABLE B.7: Statistical parameters of the time series used for the GF2 case are: $Var.$ shows the variable abbreviation or symbol, $\bar{X}$ mean value, S_x Standart deviation, J shows number of candidate subsets

Time Serie	$\bar{X}$	S_x	Range	Units	$Var.$	J
Be Res. Inflow‡†	1.81	1.35	0-43.38	m^3s^{-1}	$\mathsf{SF}_{\mathtt{Be}}$	1
Rainfall$^{\ddagger}_{\bar{a}}$	1.51	5.04	0-181.90	mm	R	8
Snowmelt^{W_i}	0.19	0.51	0-7.86	mm	S	1
Snowfall^{W_i}	0.24	1.04	0-19.12	mm	R_S	1
Precipitation^{W_i}	1.47	4.87	0-101.30	mm	R_W	1
$\ddagger \rightarrow$ Data collected in the monitoring networks $\dagger \rightarrow$ Target Value						
$W_i \rightarrow$ Time series reproduced by hydrological modelling (WiMMed)						

• **GF3** The target variable is the Streamflow at Trevélez gauge station ($\mathsf{SF}_{\mathtt{Tr}}$), see Fig.4.15 for location. In Sec.4.4.3 results are shown and commented.

TABLE B.8: Statistical parameters of the time series used for the GF3 case are: $Var.$ shows the variable abbreviation or symbol, $\bar{X}$ mean value, S_x Standart deviation, J shows number of candidate subsets

Time Serie	$\bar{X}$	S_x	Range	Units	$Var.$	J
Trevelez Str.‡†	1.43	1.71	0-13.06	m^3s^{-1}	$\mathsf{SF}_{\mathtt{Tr}}$	1
Rainfall‡	2.02	6.92	0-305	mm	R	16
‡ → Data collected in the monitoring networks † → Target Value						

B.2.2 Empirical Thresholds

In each regression experience, the deterministic model performance is evaluated based on simplest model: the Persist Method (Krajewski et al., 2020), more details in Sec.3.4. In this work we define them as empirical thresholds ($\mathsf{EC}_{persist}$), as it is a fixed value that depends on the data set we are working on, either during training or validation periods. Doing that, we establish a baseline to define a proper benchmark in the forecasting comparison experience. We check if the model improves the persist method at least. In Fig B.5, the NSE and RMSE based persist values are plotted for six daily cases and five ahead time horizons. In the corresponding results we show the relative improvement over these values. On the other hand, in Fig.B.6, the NSE and RMSE based persist values for five time horizons at hourly basis are plotted.

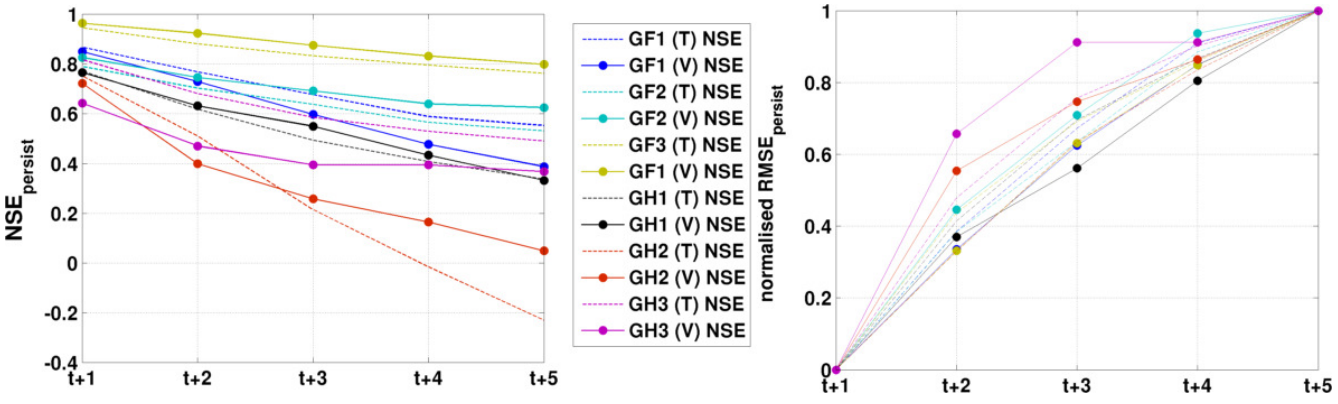


FIGURE B.5: The empirical threshold values based on the NSE coefficient *(Left)* and on the RMSE coefficient *(Right)* for five time horizons at daily basis

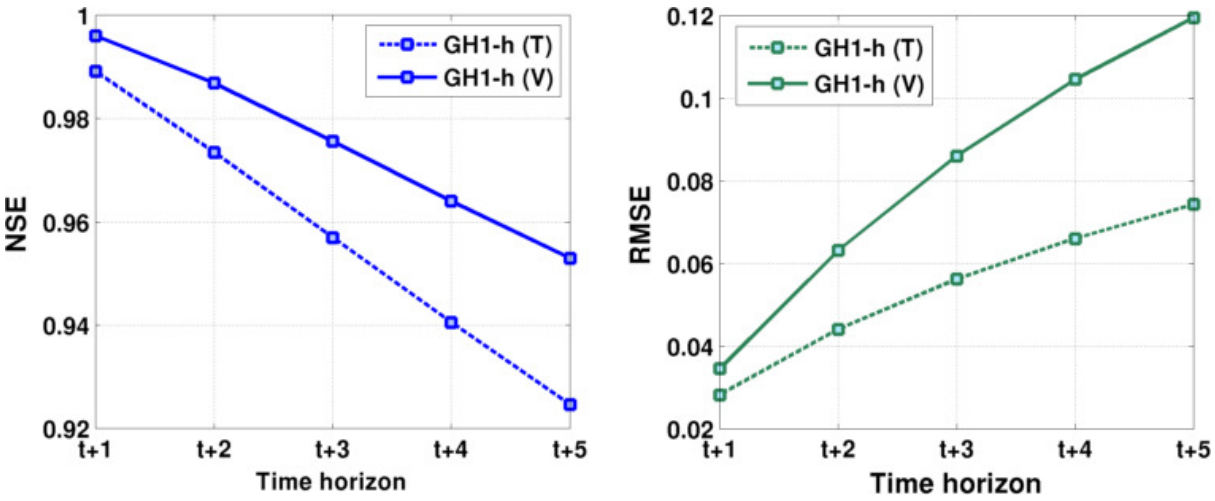


FIGURE B.6: The empirical threshold values based on the NSE coefficient *(Left)* and on the RMSE coefficient *(Right)* for five time horizons at hourly basis (Sec.4.3.2)

B.2.3 Snow variables from hydro-meteorological modelling

The snowy nature of a mountainous basin and the absolute lack of measured data, forces us to reproduce the past hydro-meteorology of the time period studied through physical modelling. Thus we obtain the temporal series of the hydrological processes of interest for two particular cases for the analysis of short term forecasting made in chapter 4: GF1 and GF2. We are particularly interested in those variables related to snow processes, to catch the snow dynamics of the area. The first step is to validate the model with the observed data, that means comparing the values in terms of total and net precipitation of the time-series generated by the physically distributed hydrological model. They are compared with the real values collected in the monitoring network and it is evaluated how good the model has been reproduced the hydro-meteorology of the study area. We have to remember that the real values come from rainfalls that only measure the liquid fraction, without differentiating the snow fraction.

For the GF1 study case the daily values have a correlation (in R^2 terms) of 0.97, 0.97, 0.98 and 0.99 for $R_{\bar{a}_2(t)}$, $R_{\bar{a}_3(t)}$, $R_{\bar{a}_4(t)}$ and $R_{\bar{a}_5(t)}$ respectively. For the GF2 study case, correlation values of 0.97, 0.97 and 0.99 for $R_{\bar{a}_2(t)}$, $R_{\bar{a}_3(t)}$ and $R_{\bar{a}_4(t)}$ respectively are obtained. It is observed that the meteorological module of the WiMMed model represents very accurately the hydro-meteorology of the both study areas. In this sense, figure B.7 shows the scatter plots of simulated total precipitation (R_W, mm) values versus observed daily mean rainfall ($R_{\bar{a}w}$, mm) for an increasing number of stations $w = 2, 3, 4, 5$. It

can be seen how, for more stations, correlation tends to increase. Therefore, with this analysis we can corroborate its correct functioning.

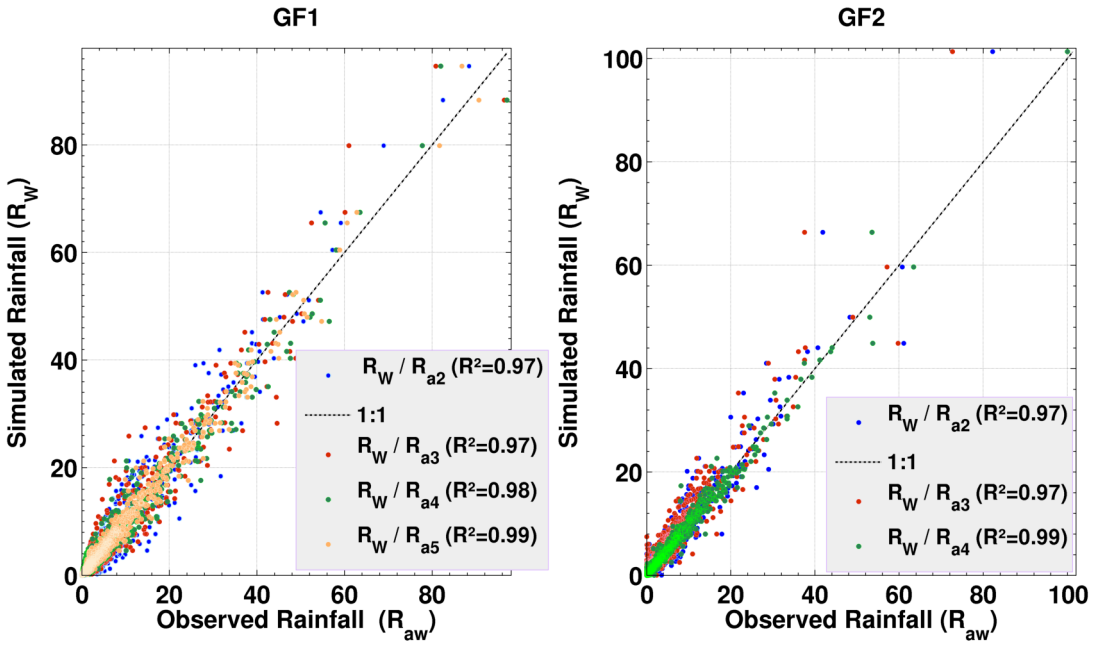


FIGURE B.7: Scatterplot of simulated daily rainfall by WiMMed (y axis, $mm \cdot day^{-1}$) versus the average observed daily rainfall (x axis, $mm \cdot day^{-1}$) for GF1 and GF2 regression experiences

In order to deepen the analysis of the snow contribution from the hydrological WiMMed model results, we plot the reproduced daily snowfall ratio (R_S/R_W) and daily snowmelt (S, $mm \cdot day^{-1}$) in figure B.8. The days to be plotted where chosen if $R_W > 0$ or $S > 0$, obtaining a total number of days that complies these conditions of $n = 2467$ and $n = 1973$ for GF1 and GF2 respectively. The values of GF1 ($31 \pm 33\%$) and GF2 ($22 \pm 30\%$) are similar with a larger percentage in the GF1 case as includes a greater proportion of area at a higher altitude.

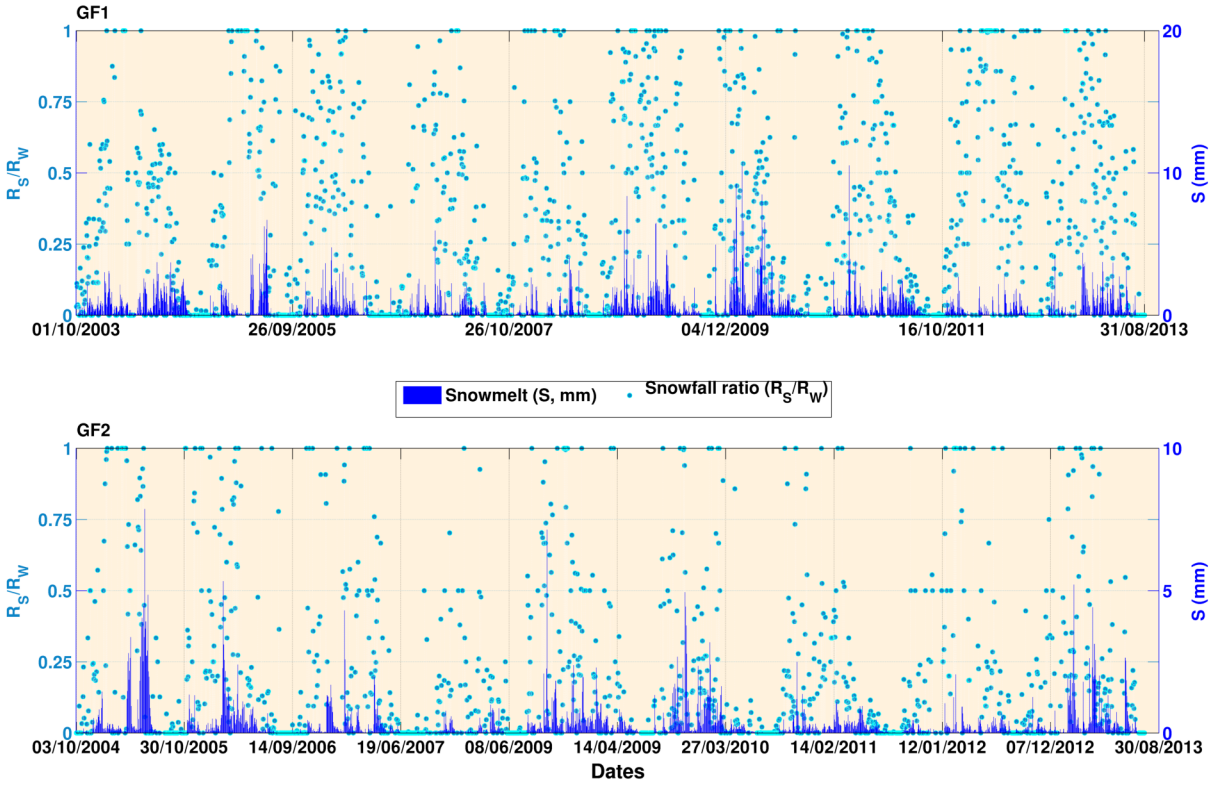


FIGURE B.8: GF1 and GF2 plots of simulated values of daily snowfall ratio (R_S/R_W, left y axis) and daily snowmelt (S, right y axis). This time series were reproduced with the WiMMed hydrological model

Then, once reproduced the real conditions, we can analyse and quantify its influence on the rainfall-runoff process for both, efficiency and timing, by comparing the simulations. In this paragraph we quantify and analyse the necessity of coupling physically based hydrological modelling, within the rainfall-runoff process, and specially focusing on the snowmelt (S) contribution and its time persistence, A and B.

For this analysis, we use again the BNN computational technique within the framework of multiple simulations (1500 runs) and compare with statistical significance of the different deterministic model outcomes. Once the best model was defined previously, in Sec.4.4.1 and Sec.4.4.2, we compare it with the best model of two new PS candidate models: the model based on real data only PS2A with $R_{\bar{a}2}$, and modified version of the final model obtained applying the Assumption B for the Snowmelt variable $S_{(t)}$, named GF1-PS4A-S_B and GF2-PS3B-S_B (see Eq.B.2). Then is quantified the degree of improvement and the most significant differences are highlighted.

$$\begin{aligned}
&\bullet\text{GF1-PS4A-S}_\text{B} \rightarrow \mathsf{SF}_{\mathtt{Ru}(t)} = f_{\mathcal{M}}(\{\mathsf{SF}_{\mathtt{Ru}(t-1)},\, \mathsf{SF}_{\mathtt{Ru}(t-2)}\},\\
&\qquad \{R_{e(t-1)},\, R_{e(t-2)},\, R_{e(t-3)}\}, \{S_{(t)}, S_{(t-1)}\}) \pm \epsilon_{(t)}\\
&\bullet\text{GF2-PS3B-S}_\text{B} \rightarrow \mathsf{SF}_{\mathtt{Be}(t)} = f_{\mathcal{M}}(\{\mathsf{SF}_{\mathtt{Be}(t-1)},\, \mathsf{SF}_{\mathtt{Be}(t-2)}\},\\
&\qquad \{R_{\bar{a}2(t)},\, R_{\bar{a}2(t-1)},\, R_{\bar{a}2(t-2)}\}, \{S_{(t)}, S_{(t-1)}\}) \pm \epsilon_{(t)}
\end{aligned} \tag{B.2}$$

In Tab.B.9 the results of the best model from 1500 random runs are compared for training and validation periods, evaluated by the efficiency coefficient (NSE), persistence index (PI), the BIC values to take into account the model parsimony. Then, a delta rule can be developed for an increasing number of parameters to check the improvement, taking into account the inputs and neurons. This is computed as follows:

$$\left|\frac{BIC(Old) - BIC(New)}{BIC(Old)}\right| \leq 0.05 \tag{B.3}$$

Then this delta rule will give us the chance to evaluate different models on the basis of only a marginal improvement in a criterion. For this case, let's just consider the BIC value, and if passed the $\tau_3 = 0.05$ threshold value of the delta rule (ΔBIC). The values highlight now how there is a clear superiority for the GF1 case but not in the GF2 case, where the GF2-PS3B-S_B model don't improve the previous one GF2-PS3B during the validation period as may tends to overfit. This analysis shows how the snowmelt S values, reproduced for the GF1 and GF2 case studies by hydrological modelling, provide valuable information as an agent to approximate underlying non-linear relationship of the rainfall-runoff process in this Mediterranean river systems.

TABLE B.9: The best $\mathcal{BNN}$ model comparison evaluated by NSE and PI coefficients during the Training and Validation periods separated by semicolon (T;V) are shown. The BIC values and its Relative Improvement (ΔBIC) are also shown

Model (PS)	NSE	PI	BIC	ΔBIC
GF1-PS2A	0.980;0.936	0.85;0.57	716.8	-
GF1-PS4A	0.981;0.935	0.85;0.57	664.2	**0.073**
GF1-PS4A-S_B	0.985;0.940	0.88;0.60	657.2	0.010
GF2-PS2B	0.924;0.920	0.91;0.50	-2159.0	-
GF2-PS3B	0.938;0.922	0.92;0.51	-2278.5	**0.055**
GF2-PS3B-S_B	0.939;0.920	0.93;0.49	-2364.5	0.038

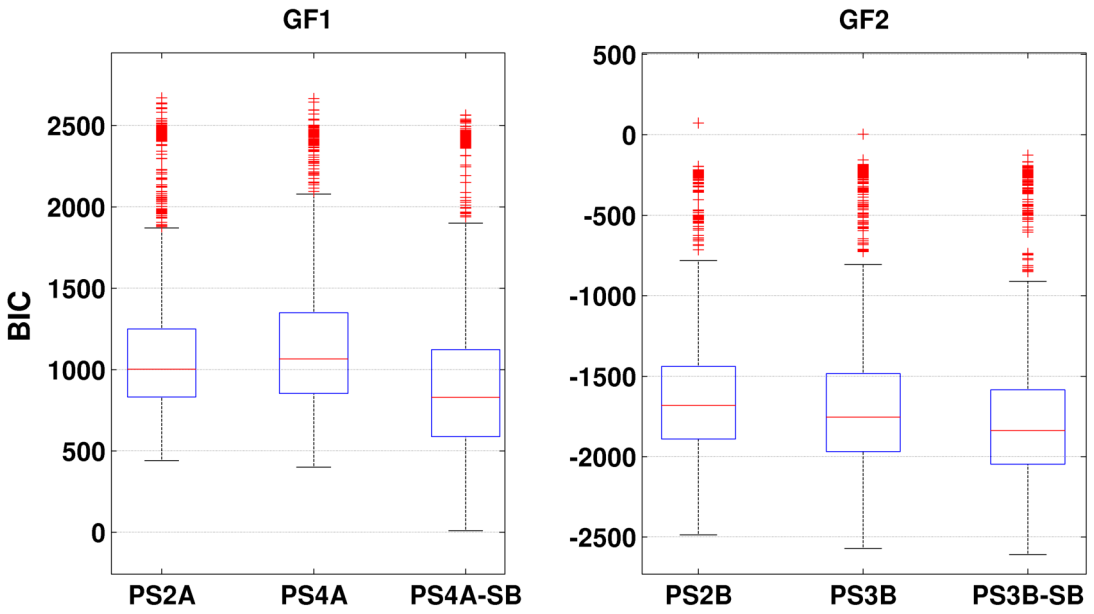


FIGURE B.9: Box-plots of the BIC values described in the text referring to the models comparison in the GF1 and GF2 case studies. Each plot summarises the multiple simulations (1500 runs) computed on the ANNs deterministic values.

In Fig.B.9 we show the BIC values for the GF1 (*left*) and GF2 (*right*) cases. For the GF1 case study, the BIC shows that the PS4A model runs produced better performance values than those of PS2A, which has higher variance and median. In the next step, PS4A→PS4A-S_B, can be seen how it changes drastically to a more pronounced difference when the Snowmelt under assumption B is included. On the other hand for GF2 is slightly improving its behaviour as we add hydrological modelling sequentially to the regression experience (PS2B→PS3B). Later on, if we change the snowmelt under assumption B (PS3B→PS3B-S_B) also improves, having a general downward trend in BIC values. This is normal since the snowmelt has an immediate effect on the rainfall-runoff process that reaches the two hydrological structures (reservoirs) analysed.

Fig.B.10 shows the BIC rates during the recognition time. Representing the BIC learning trend, we can visualise the optimal number of neurons as this metric penalises an escalating number of neurons having into account the behavior (Eq.3.13). Finally, the inclusion of Snowmelt under the assumption B (S_B) has not surpassed the delta rule, and therefore is not included in the final models (more in Chap.4). It is also worth mentioning that for real applications of assumption B for the snowmelt (S_B), would require the forecasted values simulated with the hydrological model, thereby adding a new source of additional uncertainty to the whole modelling procedure.

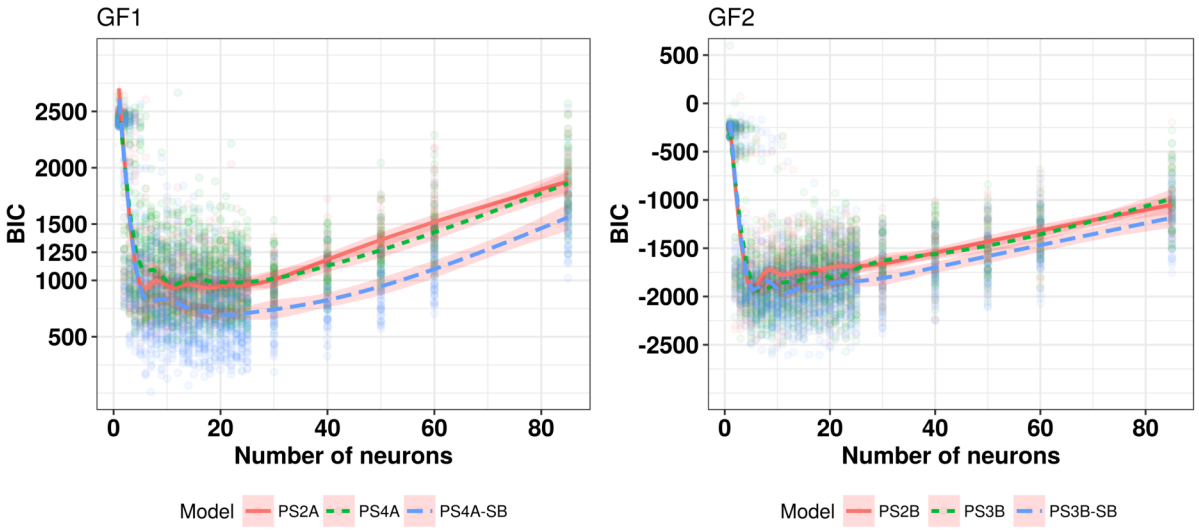


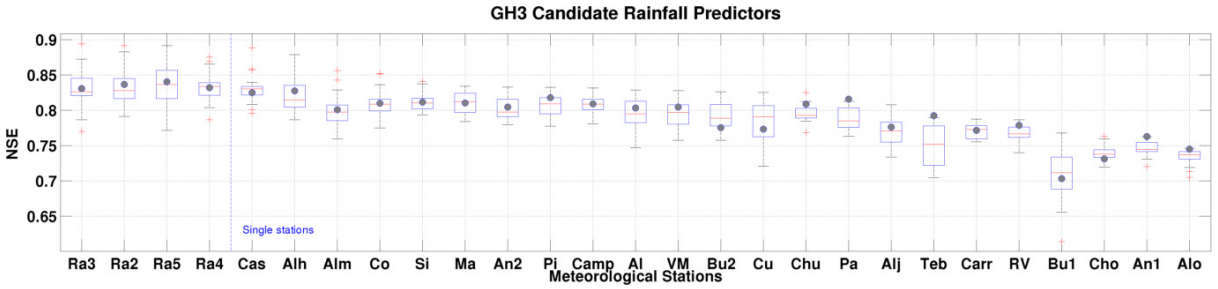
FIGURE B.10: Learning trends for three Predictors Structure (PS) in the GF1 and GF2 cases

B.2.4 Input Variable Selection Results

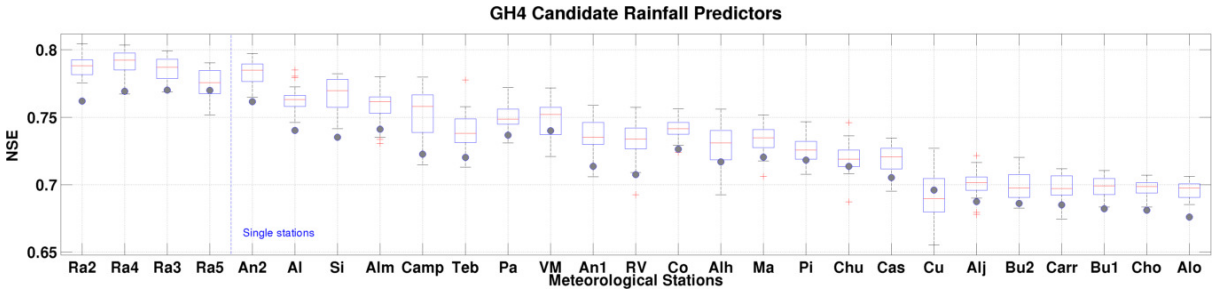
In Fig.B.11, we present the results of the step 2 of the IVS procedure for each case study on a daily basis (Chap.4). All the potential candidate rainfall variables, in several locations, are sorted in order of importance, with the multiple outcomes obtained from the *lin* and *Nlin* test.

B.2.5 Learning trends visualisation

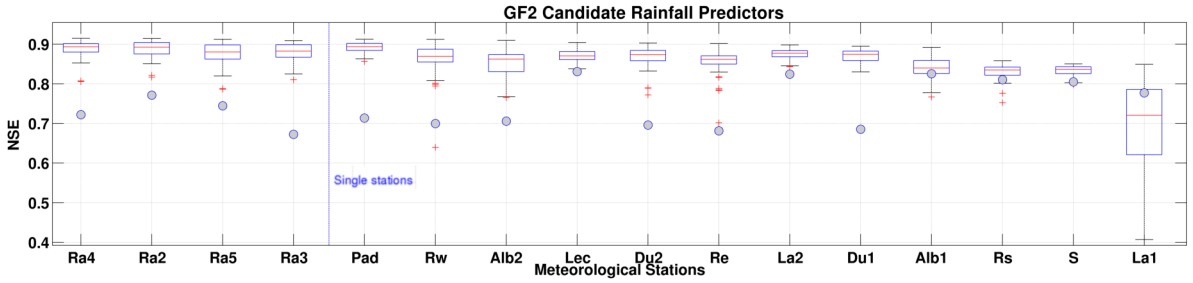
In the applications, the learning trends are the representation of i performance metric (ϱ_i) for a given predictor structure (PS) and boundary conditions during a model phase development (T and/or V). In this manner, we can see its performance for an increasing number of neurons within the BNN structure. This is useful to visualise important minor differences and learning features. It gives good perspective on the best PS, on over-parametrised models regions, on the optimal number of neurons, on the optimal learning parameters values, on the optimal hyper-parameters values and so on and so forth, for a specific given problem. Normally, the models will tend to a better adjustment during training and a worse generalisation on unseen data. All these aspects can be appreciated from the figures. It should be mentioned its iterative and exhaustive search requirements as a disadvantage.



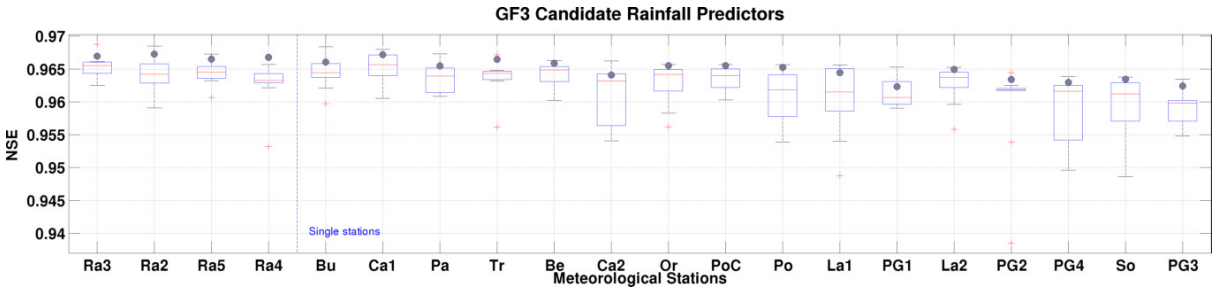
(a) IVS for rainfall inputs of the GH3 case study (Sec.4.3.3), sorted by their maximum value reached by the *Nlin* test outcomes (boxplot). The corresponding *lin* results values are also presented (grey dots).



(b) IVS for rainfall inputs of the GH4 case study (Sec.4.3.4), sorted by their maximum value reached by the *Nlin* test outcomes (boxplot). The corresponding *lin* results values are also presented (grey dots).



(c) IVS for rainfall inputs of the GF2 case study (Sec.4.4.2), sorted by their maximum value reached by the *Nlin* test outcomes (boxplot). The corresponding *lin* results values are also presented (grey dots).



(d) IVS for rainfall inputs of the GF3 case study (Sec.4.4.2), sorted by their maximum value reached by the *Nlin* test outcomes (boxplot). The corresponding *lin* results values are also presented (grey dots).

FIGURE B.11: IVS for rainfall inputs results of study cases at daily basis.

The learning trends representation in the following figures, are used to choose the optimum number of units within the hidden layer, for a given BNNs conditions: Scaled Conjugate Gradient and the hyperbolic tangent as transfer function Fig.A.3.(a), and the linear as output function (Fig.A.3.(b)). In these graphics we plot for each study case the evolution of the error and the persistence index with the number of neurons. Under 25 neurons, the simulations are replicated in increments of 1 neuron. From this value, the increments between neurons increase for each subsequent simulation. The generalized additive model (GAM) fit helping to visualize the trend. This ANN configuration has been applied for the regression cases (Chapter 4 & Chapter 5).

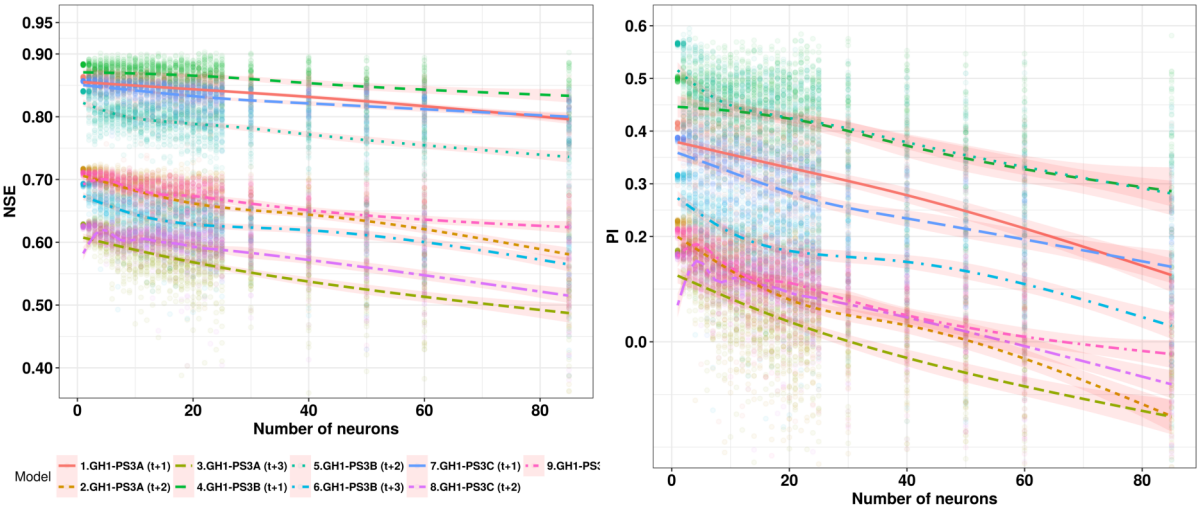


FIGURE B.12: BNNs learning trends results: regression on NSE and PI values versus the number of neurons within the hidden (HL) layer comparing relevant PS Tab 4.1

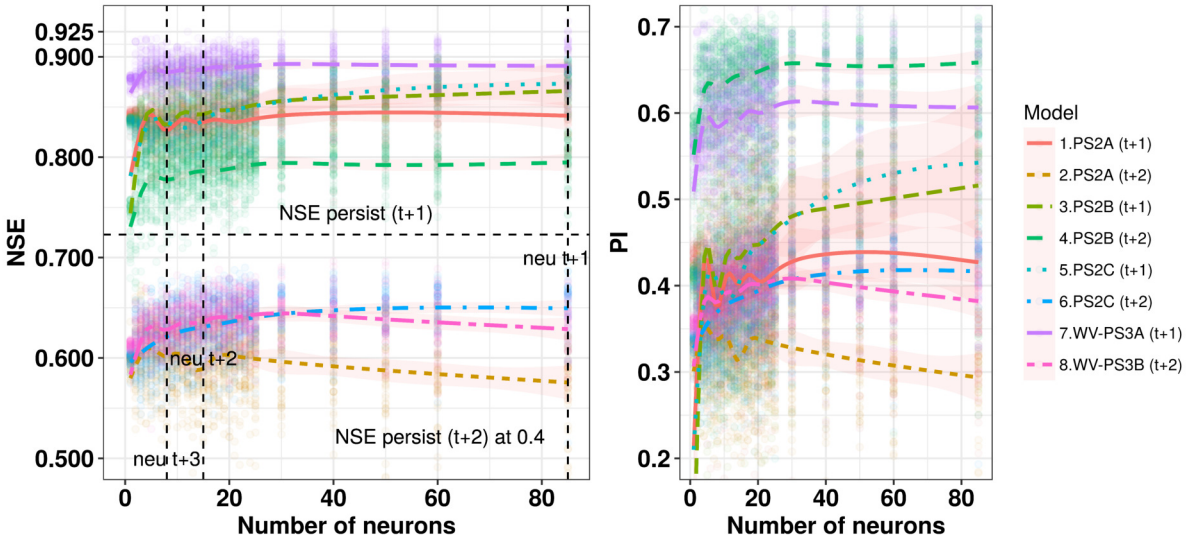


FIGURE B.13: GAM regression on ANNs multiple simulations results comparing relevant PS Tab.4.5 for the GH2 experience

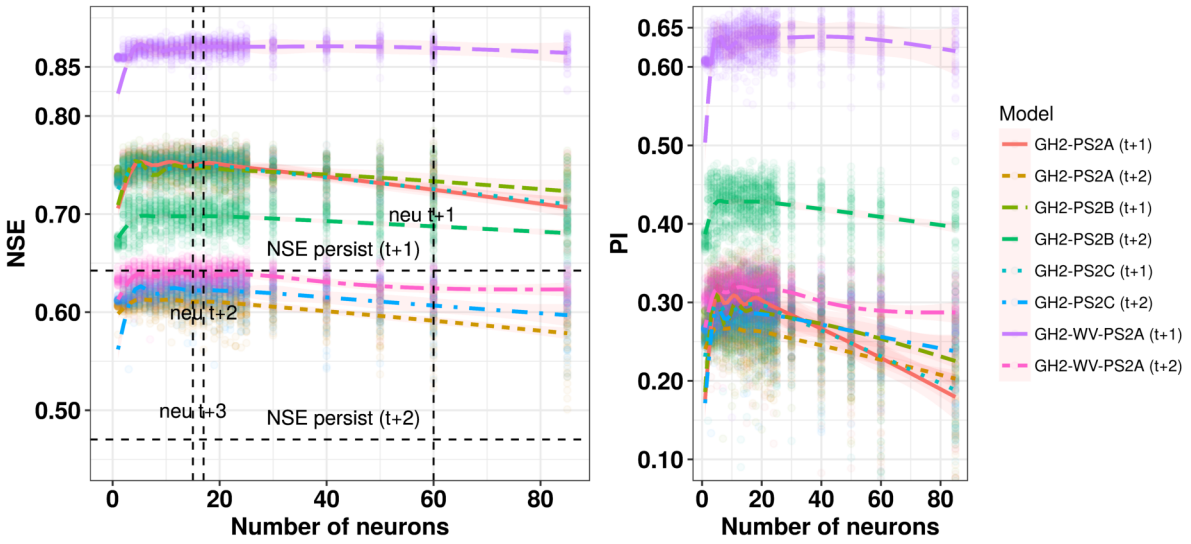


FIGURE B.14: ANNs multiple simulations results: regression on NSE and PI values versus the number of neurons within the hidden (HL) layer comparing relevant PS Tab.4.7

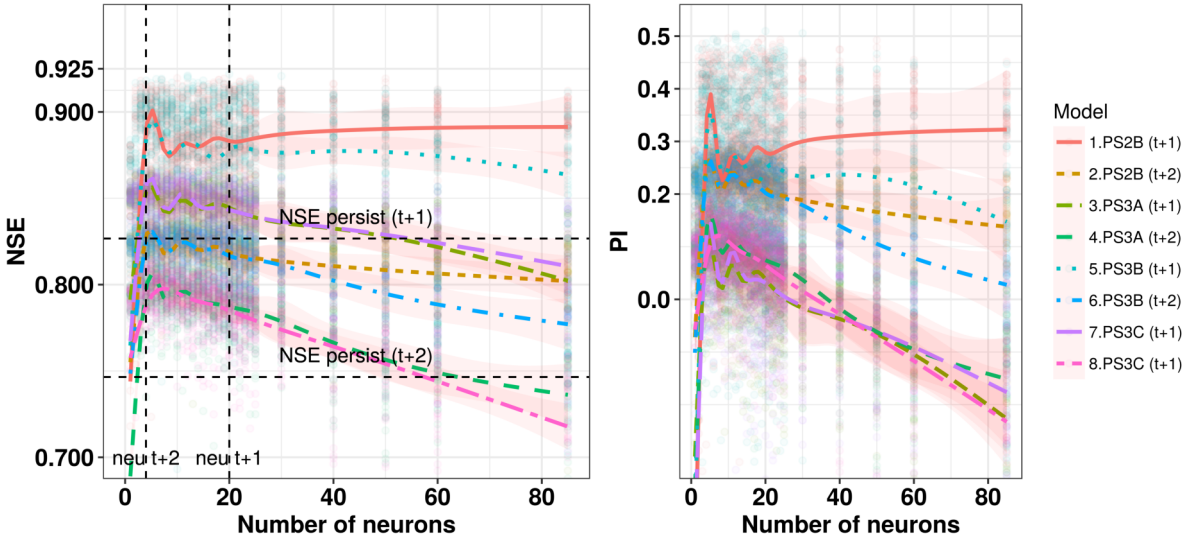


FIGURE B.15: GF2 ANNs multiple simulations results: regression on NSE and PI values versus the number of neurons within the hidden (HL) layer comparing relevant PS Tab 4.11

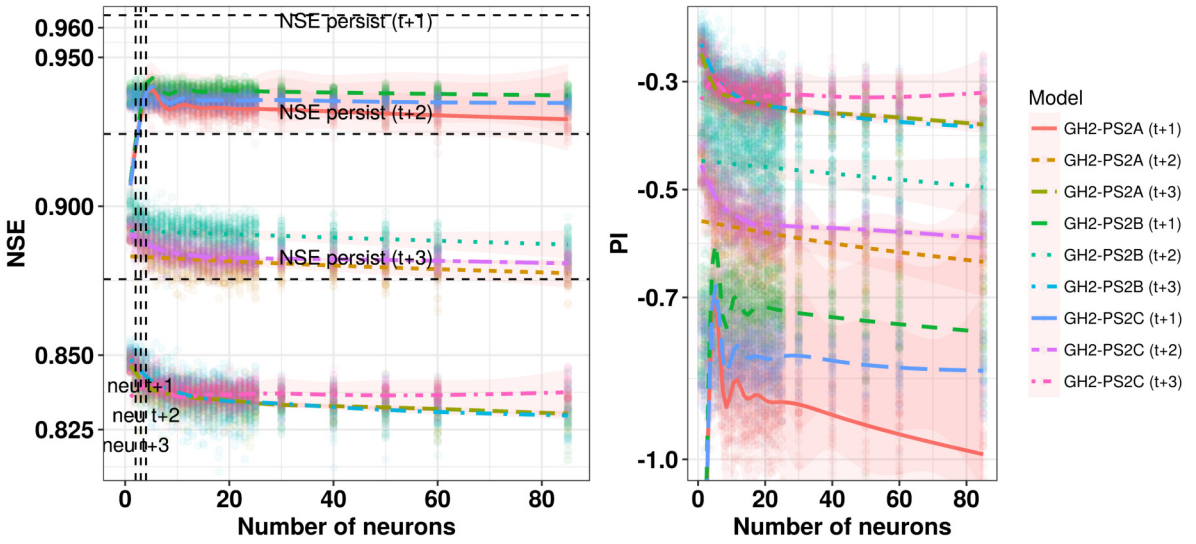


FIGURE B.16: GF3 ANNs multiple simulations results: regression on NSE and PI values versus the number of neurons within the hidden (HL) layer comparing relevant PS Tab 4.13

B.3 Supplementary material of chapter 5

Four Atmospheric Oscillations (Fig.B.17) are used to help the forecasting capabilities of the models in the three regressive experiences carried out within the chapter. In table B.10, its statistical properties are listed.

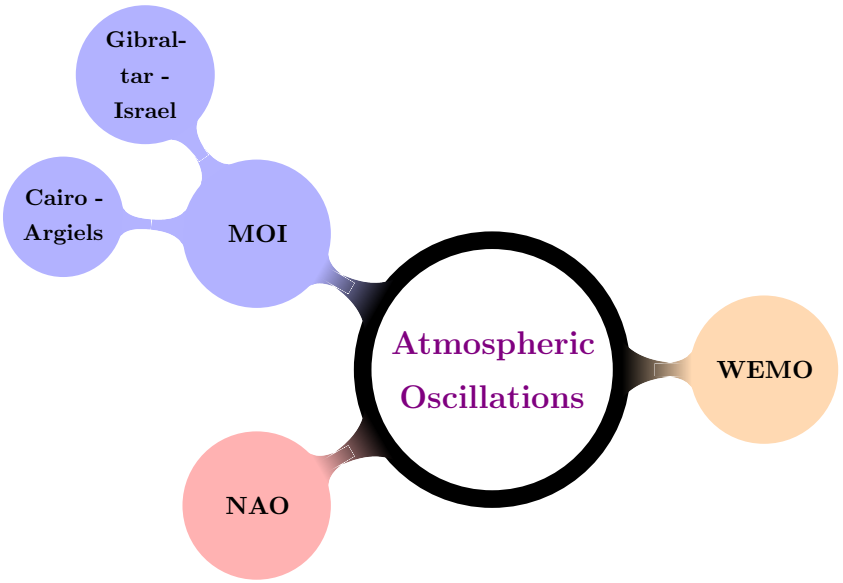


FIGURE B.17: Atmospheric Oscillations (AO)

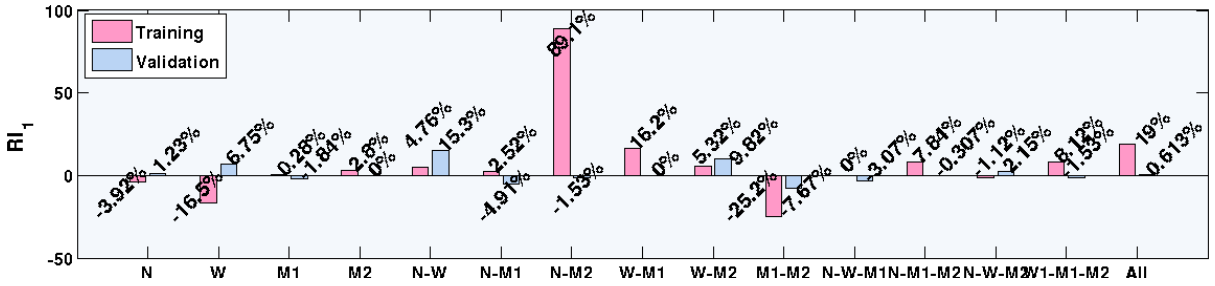
TABLE B.10: Statistical parameters of the Circulation Indices used at monthly basis are: $Var.$ shows the variable abbreviation or symbol, $\bar{X}$ mean value, S_x Standart deviation, J shows number of candidate subsets

Time Serie	$\bar{X}$	S_x	Range	Units	$Var.$	J
NAO‡	-0.02	1.01	-3.18−3.04	[1]	NAO	1
WEMO†	0.03	1.25	-4.29−4.68	[1]	WEMO	1
MOI-GI$^{\diamond}$	-0.01	0.46	-1.62−1.05	[1]	$\mathbf{MOI_{GI}}$	1
MOI-CA$^{\diamond}$	-0.02	0.46	-1.40−1.24	[1]	$\mathbf{MOI_{CA}}$	1
‡ →Data downloaded from the NOAA website						
† →Data downloaded from the UB website						
◇ →Data downloaded from the UEA website						

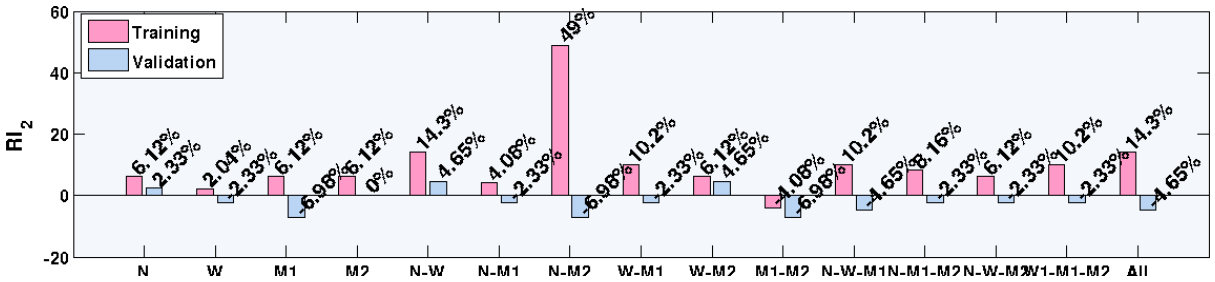
Here we submit supplementary material of the forecasting experiences carried out in chapter 5. Tab.B.11 shows the results of the inclusion of all AOs combinations for the rainfall forecasting exercise carried out in Sec.5.3. On the other hand, in Fig.B.18 we see the results of the relative improvements (RI) based on NSE (RI_1), on PI (RI_2), and on BIC (RI_3). This improvement is with respect to the previous predictor structure, i.e., PS2, during the training and validation periods. In this case, we see how the pair formed by N-W (M5) is the one that gives the best values for both methodological periods. It should be noted that the improvement of the N-M2 pair (M7) is very significant during the training period, however it leads the model to a worse generalization during the validation. Finally, Fig.B.20 show the results of the GEV-SPI6-PS2-M9 model during the validation period (Sec.5.4.2). We can also see in this figure, the error, and the dispersion graphs for both training and validation, showing a correct generalization.

TABLE B.11: Relevant Atmospheric Oscillations combinations evaluated at each case for $h = 0$ time ahead step horizon. Between brackets the relative improvement (RI_i) to respect PS2 performance to follow the stepwise approach development. The AOs are represented as follows: NAO→N, WEMO→W, MOI-GI→M1, MOI-CA→M2

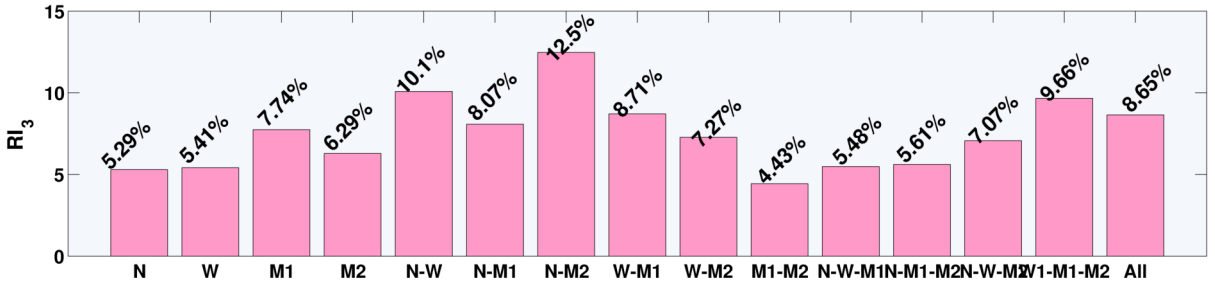
Location (z)	PS3	R^2	PI	RMSE	BIC
	M1(N)	0.343;0.330	0.52;0.44	38.40;38.96	2422.6
	M2(W)	0.298;0.348	0.50;0.42	34.17;39.34	2419.5
	M3(M1)	0.358;0.320	0.52;0.40	30.79;40.10	2360.1
	M4(M2)	0.367;0.326	0.52;0.43	36.89;39.13	2397.1
	M5(N-W)	**0.374;0.376**	**0.56;0.45**	**31.01;38.59**	**2299.9**
Med.RBD	M6(N-M1)	0.366;0.310	0.51;0.42	33.97;39.61	2351.3
	M7(N-M2)	0.675;0.321	0.73;0.40	26.49;40.21	2238.9
	M8(W-M1)	0.415;0.326	0.54;0.42	30.52;39.35	2335.1
	M9(W-M2)	0.376;0.358	0.52;0.45	31.70;38.68	2372.0
	M10(M1-M2)	0.267;0.301	0.47;0.40	34.13;40.29	2444.5
	M11(N-W-M1)	0.357;0.316	0.54;0.41	33.72;39.79	2417.6
	M12(W-M1-M2)	0.385;0.325	0.53;0.42	33.21;39.60	2414.4
	M13(M2-W-N)	0.353;0.333	0.52;0.42	32.28;39.60	2377.1
	M14(N-M1-M2)	0.386;0.321	0.54;0.42	31.54;39.72	2310.7
	M15(N-W-M1-M2)	0.425;0.328	0.56;0.41	29.37;39.71	2336.6
	M1(N)	0.421;0.422	0.63;0.55	37.53;37.61	2309.2
	M2(W)	0.370;0.445	0.57;0.57	36.53;36.61	2223.2
	M3(M1)	0.483;0.448	0.67;0.59	32.79;36.35	2261.9
	M4(M2)	0.317;0.407	0.60;0.57	39.84;37.65	2267.7
	M5(N-W)	0.421;0.409	0.59;0.51	38.48;38.95	2272.8
Lan.GF	**M6(N-M1)**	**0.363;0.429**	**0.58;0.59**	**36.51;36.91**	**2222.8**
	M7(N-M2)	0.513;0.393	0.67;0.52	36.90;38.47	2235.4
	M8(W-M1)	0.418;0.417	0.61;0.52	40.41;39.08	2288.7
	M9(W-M2)	0.530;0.431	0.65;0.54	35.50;37.70	2219.1
	M10(M1-M2)	0.391;0.435	0.59;0.59	39.12;36.90	2263.3
	M11(N-W-M1)	0.412;0.433	0.61;0.55	38.63;37.31	2287.8
	M12(W-M1-M2)	0.698;0.472	0.78;0.58	28.14;36.02	2121.3
	M13(M2-W-N)	0.289;0.404	0.51;0.54	42.15;38.02	2319.8
	M14(N-M1-M2)	0.626;0.449	0.79;0.57	32.17;36.48	2174.3
	M15(N-W-M1-M2)	0.463;0.437	0.66;0.55	37.63;37.30	2266.0
	M1(N)	0.343;0.330	0.52;0.44	38.40;38.96	2422.6
	M2(W)	0.298;0.348	0.50;0.42	34.17;39.34	2419.5
	M3(M1)	0.358;0.320	0.52;0.40	30.79;40.10	2360.1
	M4(M2)	0.367;0.326	0.52;0.43	36.89;39.13	2397.1
	M5(N-W)	**0.374;0.376**	**0.56;0.45**	**31.01;38.59**	**2299.9**
Org.RBD	M6(N-M1)	0.366;0.310	0.51;0.42	33.97;39.61	2351.3
	M7(N-M2)	0.675;0.321	0.73;0.40	26.49;40.21	2238.9
	M8(W-M1)	0.415;0.326	0.54;0.42	30.52;39.35	2335.1
	M9(W-M2)	0.376;0.358	0.52;0.45	31.702;38.676	2372.0
	M10(M1-M2)	0.267;0.301	0.47;0.40	34.13;40.29	2444.5
	M11(N-W-M1)	0.357;0.316	0.54;0.41	33.722;39.789	2417.6
	M12(W-M1-M2)	0.385;0.325	0.53;0.42	33.21;39.60	2414.4
	M13(M2-W-N)	0.353;0.333	0.52;0.42	32.28;39.60	2377.1
	M14(N-M1-M2)	0.386;0.321	0.54;0.42	31.54;39.72	2310.7
	M15(N-W-M1-M2)	0.425;0.328	0.56;0.41	29.37;39.71	2336.6



(a) RI_1 values during the IVS procedure for AOs inclusion



(b) RI_2 values during the IVS procedure for AOs inclusion



(c) RI_3 values during the IVS procedure for AOs inclusion

FIGURE B.18: Relative Improvements (RI) resulted of atmospheric oscillations inclusion for rainfall forecasting at the Mediterranean RBD

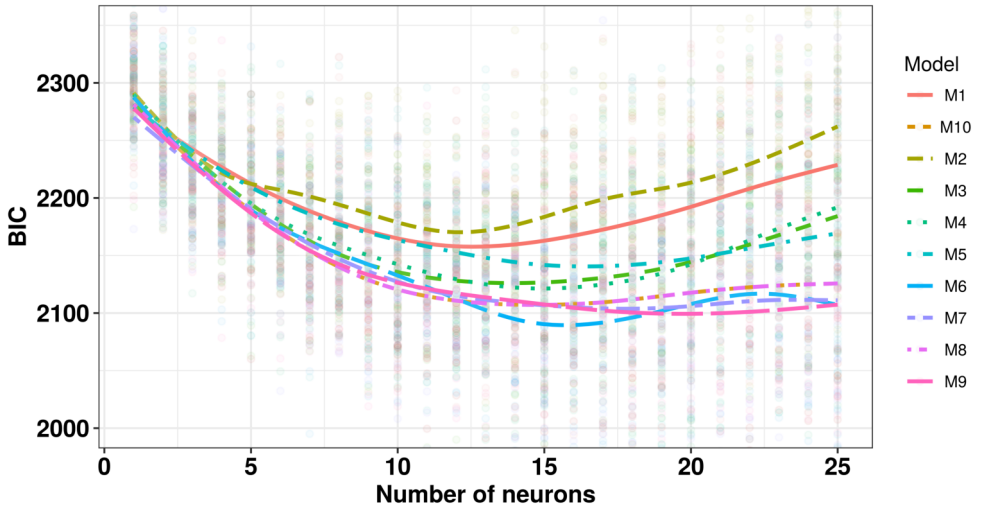


FIGURE B.19: BIC learning trends for `Lan`.GF-PS3-BNN models, testing single and pairs combination of the Atmospheric Oscillations. The corresponding models and its combinations, are listed in Tab.B.11. The triplet and the quadruplet models are leaving apart for simplification purposes

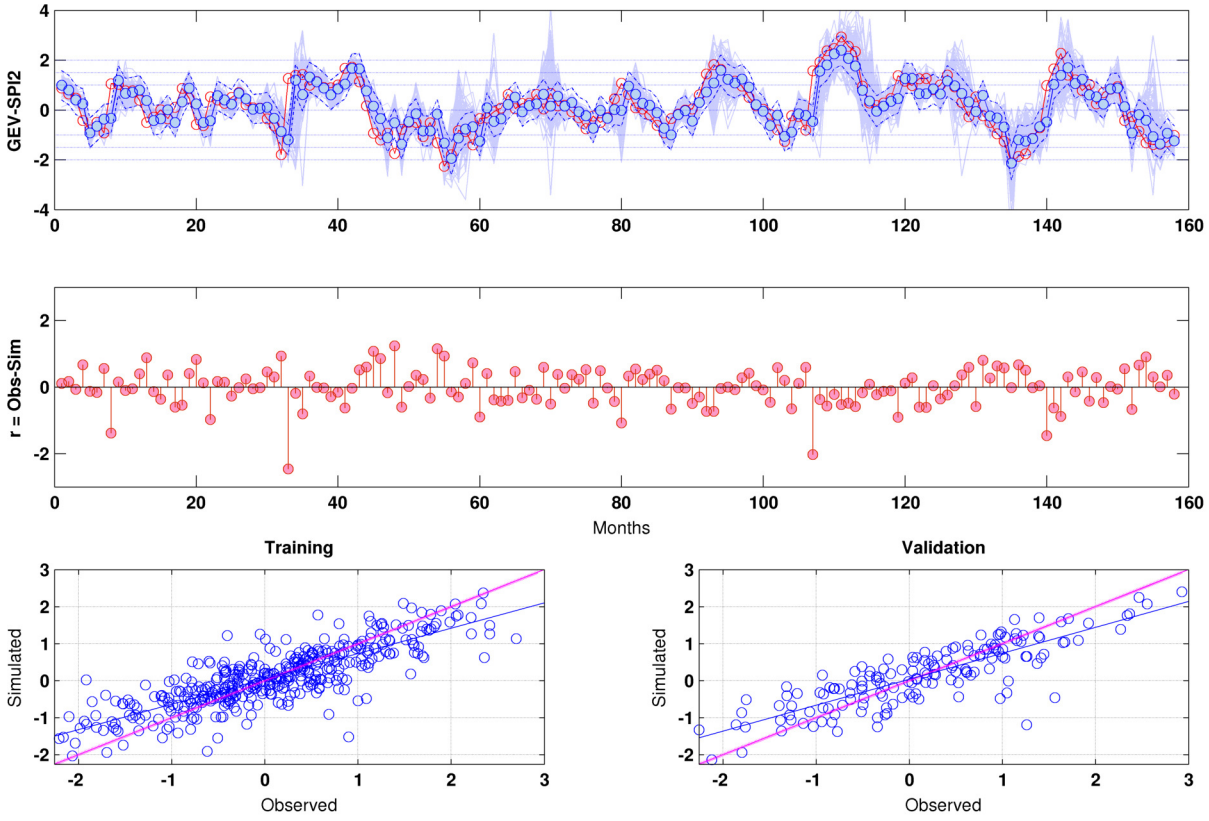


FIGURE B.20: Estimated SPI values, errors r and scatterplots of the best model in the GEV-SPI6-PS2-M9

B.4 Panoramas

(a) Guadalfeo River Basin view

(b) Recreational uses of the Rules (`Ru`) Reservoir because of the particular wind dynamics

(c) Northern Sierra Nevada Panorama with the Canales Reservoir at the left down corner. As can be appreciated Mediterranean water environments has many contrast and strong gradients, where actually the water management must play a mayor role

(d) South-western Granada Panorama from the Andalusian Inter-University Institute for Earth System Research (IISTA-CEAMA). The orographic precipitation in Mediterranean environment has strong importance on the regional scale effects

FIGURE B.21: Panoramas views

B.5 Data Providers

I would like to thank the **hydrological data** providers: from UTE SAIH Sur to Jesus Regodón for information related with the GuadalHorce River, from *La Junta de Andalucía*: thanks to Jose Antonio Caffareña and Javier Villar for the streamflow and water levels height data. Thanks to Francisco Jose Calvo by the provided reservoir information of Beznar and Rules Reservoir System (BRR). I would like to thanks to Javier Herrero and Maria José Perez Palazón for its help on WiMMed model and simulations. Also, I would like to thanks to Darío García Contreras, who helps with the **climatic data** download and use for the GH experiences cases case. For the Atmospheric Oscillation data, I would like to express my gratitude to Joan Albert López-Bustins and Laia Arbiol for providing the WEMO dataset at daily basis. Finally, to the REDIAM (belongs to the Junta de Andalucia) for the **hydro-meteorological data**, from which its web-page the data is downloaded, and AEMET for supplying meteo-data trough this service.

Appendix C

Data-Driven Ensemble

C.1 Introduction

In this work we have not used the data-driven ensembles, which are nevertheless a tool worth testing for the case studies presented. In this appendix, a test exercise of these ensembles is included for the case presented in section 4.3.3 and section 4.3.4.

Ensemble is like executing multiple models and combining several weak models to achieve a stronger one. This idea brings to mind the parable of the blind men and an elephant originated in ancient Indian subcontinent (Schellnhuber et al., 2014; Zimek and Vreeken, 2015). In this parable, each blind man feels a different part of the elephant, in our case different parts or dynamic relationships within the water system. Initially, with multiple models we can approximate real behaviours in a more exact manner. The ensemble implies an extra modelling step that is added into the regression experience: a methodological framework to integrate multiple models into a unique one.

In climate applications this approach is very common. Several climate organisations applied it to estimate the uncertainty associated under different assumptions, or to generate a single value from different models or boundary conditions. Climate organisations such as NCEP or ECMWF works with it by having its own ensemble prediction systems, Global Ensemble Forecast System (GEFS) and the Ensemble system of the ECMWF. These multiple conditions are downscaled to hydrological variables for prediction, forecasting management and so on. In this manner, we force hydrometeorological variables to the regionalities, and overcome the scale effects.

In water management applications, the Ensemble Streamflow Prediction (ESP) technique (Day, 1985) involves a hydrological modelling with present basin initial conditions. Then historical climate patterns are applied to simulate future conditions generating multiple hydrographs. These approaches are important to deal with the uncertainty inherent in the water systems under Mediterranean conditions, which are variable (due its season variability) and has intense pattern changes. Jeong and Kim (2005a) combine ESP with ANN, that means climate services with DB methods, to generate multiple candidate hydrographs predictions for a water allocation problem. For flood forecasting applications, Cloke and Pappenberger (2009) list the virtues of building ensembles from various analytical/numerical weather predictions.

The procedures to set up the ensemble can be the most diverse. In meteo-hydrology problems, ensembles of DD techniques, such as ANNs or MLR, are frequent. Zhou et al. (2002) display how to use ANN ensembles. Tiwari and Chatterjee (2010) develop a short-term rainfall runoff ensemble model for flood forecasting purposes, Jeong and Kim (2005b) present how to work for rainfall-runoff modelling ensembles, and DeWeber and Wagner (2014) use the neural network ensemble to predict daily river temperature. We can use nonlinear DD methods as Genetic algorithms to generate ensembles, as shown in Zameer et al. (2017). Maqsood et al. (2004) demonstrated his appropriate behaviour in weather forecasting.

Under Mediterranean climatic conditions there are few experiences (Anctil et al., 2006). On the other hand, Karvelis et al. (2017) develop a model under this assumption for the wind. Therefore applications for Mediterranean environments are needed and by comparing them with the single models we will assess if the use of several member models is really a breakthrough. For this purpose we will take advantage of the multiple simulations made in the previous chapters and develop ensembles from them, offering some preliminary results in this appendix. As discussed in Sec.3.2.6, the natural complexity of the water systems, the application of bootstrap resampling, and the non-straightforward parametrisation of some of the DD models used, i.e., BNN, leads us to manage multiple models or/and simulations. This gives us the opportunity to get, under a statistical (grey) nature, more robust and workable results.

C.2 Methodology

In this work, we will work with two modes for combining multiple candidates models in an ensemble: (I) Equal Weights Average (EWA) with an uniform weights distribution for $\mathtt{K}$ candidate models, and (II) Bayesian Model Averaging (BMA) with weights for each candidate model conforming to its statistical evidence for $\mathtt{K}$ candidate models. In the ensemble modelling framework, let be denoted by $\mathsf{SF} = \{\mathsf{SF}_{(1)}, \ldots, \mathsf{SF}_{(n)}\}$ the vector of the target variable of interest with n observations. So, it is assumed that exists a subset of $\mathtt{K}$ candidate models for forecasting purposes of vector SF. Then, let be D_{tk} a matrix where each model to ensemble has its forecasted points on the line time, where $k = \{1, \ldots, \mathtt{K}\}$ and $t = \{1, \ldots, n\}$ respectively. Thus, the weighted average ensemble outcome y^{ξ} can be computed as,

$$y_{(t)}^{\mathrm{II}} = \sum_{k=1}^{\mathtt{K}} D_{tk}\beta_k + \varepsilon_{(t)}^{\mathrm{II}} \tag{C.1}$$

where $\varepsilon^{\mathrm{II}}$ is the deviation model variance. The first one is the directest form of an ensemble: EWA. Producing the same weights (Eq.C.2) for all the individuals that makes up the ensemble model. The EWA output average $\bar{y}_{(t)}^{\mathrm{II}}$ will be the mean value of all the outputs (Eq.C.3).

$$\beta_{\mathrm{EWA}} = (\frac{1}{\mathtt{K}}, \ldots, \frac{1}{\mathtt{K}}) \tag{C.2}$$

$$y_{(t)}^{\mathrm{II}} = \frac{1}{\mathtt{K}} \sum_{k=1}^{\mathtt{K}} D_{tk} \tag{C.3}$$

For this instance, there is independence on the training data set, with the exception if a selection of the best candidate members subsets are made. We must do this during the training period. In distinction to the uniform assumption (EWA method), where the weights over the final model are uniform, the Bayesian Model Averaging (BMA, Hoeting et al. (1999)) method based in the weights on posterior model probabilities. The BMA method assumes that the uncertainty forecast of each applicant member can be described with a forecast distribution function $f_k(\cdot)$. It assumes that the probability ($\mathrm{P}(y_{(t)}^{\mathrm{II}})$) of an observation of the target variable at time t is the weighted sum over a k number of probability distributions,

$$P(y_{(t)}) = \sum_{k=1}^{\mathtt{K}} \beta_k f_k(\mathsf{SF}_{(t)}) \tag{C.4}$$

which is assumed to have a normal distribution with variance σ_k,

$$P(y_{(t)}) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(\frac{(y_{(t)}-)^2}{2\sigma_k^2}\right) \tag{C.5}$$

For that, the BMA method calls for the estimation of weights $\beta_{\text{BMA}} = \{\beta_{1\text{BMA}}, \ldots, \beta_{k\text{BMA}}\}$, and standard deviations, $\sigma_{\text{BMA}} = \{\sigma_{1\text{BMA}}, \ldots, \sigma_{k\text{BMA}}\}$. To find the optimal values of these parameters for each candidate model, the Expectation-Maximisation algorithm is used (Raftery et al., 2005). This algorithm alternates between an expectation and a maximisation step until it achieves the convergence, representing the probability that the model will produce the correct forecast. For a legitimate expectation distribution function, the sum of the weights must be up to one and strictly positive. Then, each candidate member has a relative contribution to the forecasting ensemble model over the training period, which must be validated for an unseen subset of data during the validation period.

The uncertainty estimation got from the BMA ensemble models has promised results. This simple procedure can outperform more consuming and complicated Bayesian advanced methods, as is pointed out in Zhang and Zhao (2012) with the MCMC-based method. In our task, we test only the deterministic model values of the EWA and BMA ensembles, leaving apart the comparative assessment of the uncertainty estimations for a future practice.

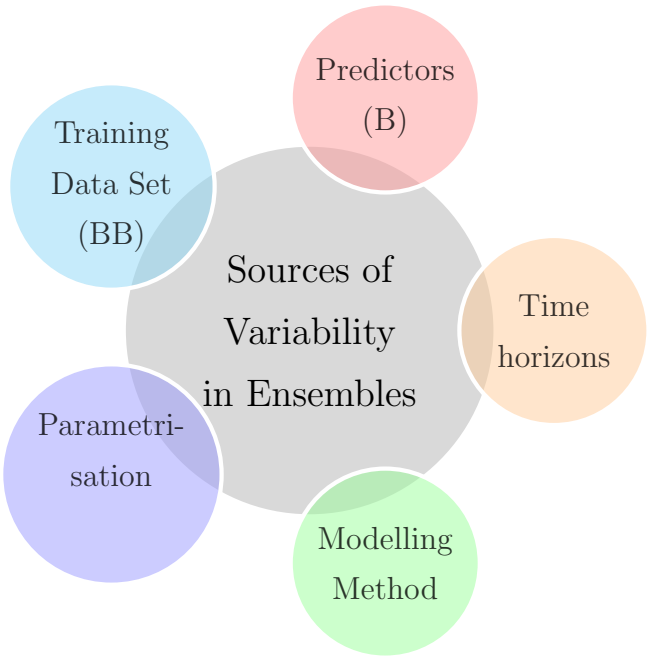


FIGURE C.1: Variability sources in Data Based ensembles models

And, how do we generate the model variability?, and, what is the best source of variability to generate it? We generate that variability by establishing different boundary conditions that can come from various sources. In Fig.C.1 we show a diagram where we propose different sources in which we can modify the conditions in order to generate different models. Next we describe each one of them:

Predictors (B) Combining, modifying and varying the predictors generates different models with similar behaviour. In Rainfall-Runoff modelling problem the simplest option is to work with different rainfall stations. In DD techniques the bagging method (Breiman, 1996) is also a good form to create multiple different models.

Training data set (BB) Various candidate members are generated by Bootstrap re-sampling techniques (see previous Sec.3.3.1). Generating different training data sets, under conditional dynamical and non-stationary rules, generates different models with similar behaviour.

Parametrisation (θ) In Bayesian Neural Networks, the random initialisation of the parameters for a $N \sim (0, \sigma)$ distribution, where different parametrisation solutions is obtained due to the local minima during weights optimisation (*multiple runs*, refer to App.A.1.2). From the candidate members the best models are used for ensemble purposes.

Modelling Method ($\mathcal{M}$) Using different methods to use a single value and then, use the output to develop an ensemble of various techniques is always a good approach. In this type of models, we can use various paradigms in a single model, i.e., linear *vs.* nonlinear, parametric *vs.* non-parametric,

Time horizons (Δt) Making an ensemble from various time horizons (h) models. In our case, a set from three different time horizons are ensemble in a single model for one step ahead forecasting.

The MLR Ensemble ($\mathcal{MLRE}$) In this work, we will develop a Multiple Linear Regression Ensemble based on various/multiple locations `K` rainfall locations (B), plus the variability provided by the samples generated by bootstrap resampling (BB). In this manner, for each run a different parametrisation through the least square optimisation method is found. The candidate models post-processing and the final ensemble result are computed through two methods: Equal Weights Averaging (EWA) and Bayesian model Averaging (BMA, Raftery et al. (2005)). For that, an iterative procedure is developed as is shown in Fig.C.2.

The GH3 and GH4 cases applicattion by the "best" single MLR model are shown in each cases study, Sec.4.3.3 and Sec.4.3.4 respectively. In these experiences, a total number of 270 single models (BB-S, from 27 different rainfall stations and 10 boostrapped models per each station) are generated and evaluated. Then, the ensemble outcomes by the EWA method (Eq.C.3) and the BMA method (Eq.C.4) are computed.

Finally, it should be noted that the number of members `K` is crucial and it is important to properly define the optimal number of them. For the EWA case a sensitivity analyses is carried out to determine the optimum percentile over all set of the candidate members.

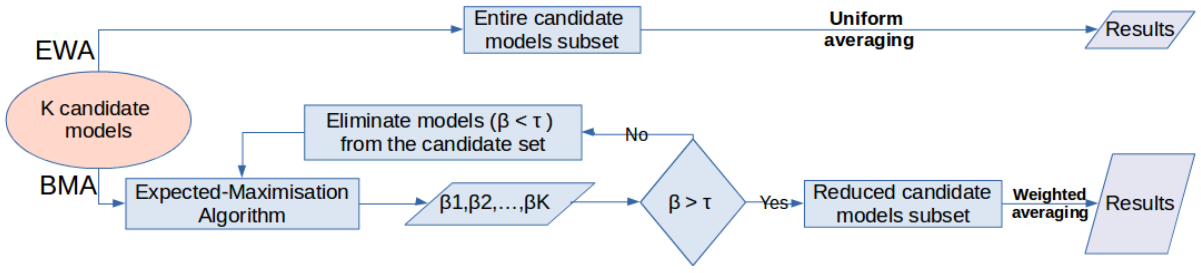


FIGURE C.2: Flowchart showing the development of the two types of Ensemble models

On the BMA application an improvement is added for the selection of the ensemble members forecast, where only those forecast that have certain relative contribution were

selected to be part of the final set of the ensemble. In this work we establish $\beta = 0.05$ as the threshold, which means that the model should have a minimum contribution of 5%. After that, the EM algorithm is run once more to compute the final values of β for the new set of ensemble members. In the BMA ensemble algorithm the β and σ parameters are computed for the training results. Then, the weighted sum for the selected models are realised for the validation period.

C.3 Preliminary Results and Discussion

Here, in Fig.C.3 we show the results of the single models and the two types of MLR ensembles developed for the GH3 and GH4 cases. The variability source that generates different models, comes from the various training samples generated with Bootstrap re-sampling and the different stations or locations used. In this manner we generated several candidate models to post-process into a single model.

From Fig.C.3, the general improvement of the BMA over the other approaches can be appreciated for almost all cases. We can also see how the most convenient choice is a single model, e.g., GH4-MLR for $t + 1$ forecasts, and for other cases the ensemble is superior, e.g., GH3 for $t + 1$, $t + 2$ and GH4 for $t + 2$.

In the results, we see how the BMA model certainly has improvements over the others. These models have a lower dispersion and higher values in terms of efficiency (NSE), both for the median and the maximum statistics than the other methods compared.

In this appendix, we have made a brief analysis for short-term forecasting purposes. Here we compared the behaviour of the best single models with respect their deterministic outcome vs. the behaviour of multiple models, by two types of ensembles, with respect their deterministic outcome.

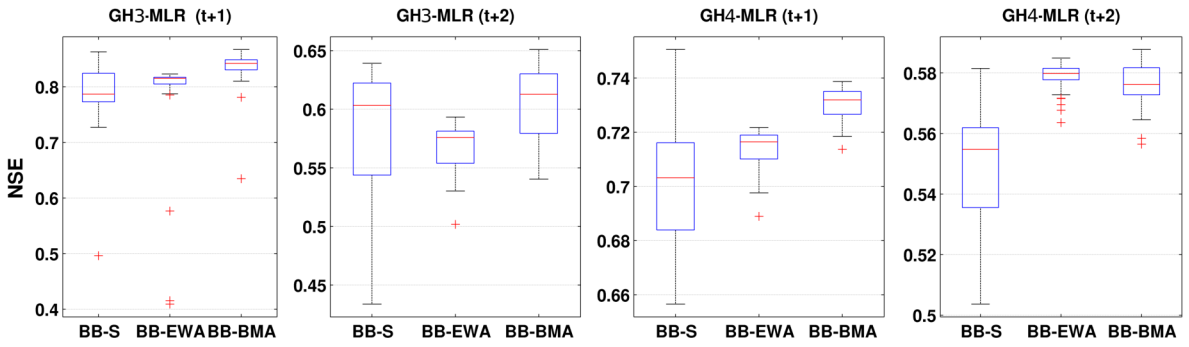


FIGURE C.3: Efficiency results for the validation period for the GH3 and the GH4 case of the three bootstrapped MLR models compared in this section.

.

"Si en tu espíritu hace asiento el deseo del estudio
y de huir de las sombras de la ignorancia,
hallarás en ella el hermoso árbol del honor.
Hace el estudio brillar como estrellas a los grandes,
y a los que no lo son los eleva a igual lucimiento"

La Madraza de Granada **s.XIV**

Bibliography

Abbot, J. and Marohasy, J. (2014). Input selection and optimisation for monthly rainfall forecasting in queensland, australia, using artificial neural networks. *Atmospheric Research*, 138:166–178.

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E., and Wilby, R. L. (2012). Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. *Progress in Physical Geography*, 36(4):480–513.

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., and Sliusarieva, A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*, 48(1):W01528.

Adamowski, J. and Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, 390(1–2):85 – 91.

Afshar, A., Massoumi, F., Afshar, A., and Mariño, M. (2015). State of the art review of ant colony optimization applications in water resource management. *Water Resources Management*, 29(11):3891–3904.

Agnese, C., Baiamonte, G., and Cammalleri, C. (2014). Modelling the occurrence of rainy days under a typical Mediterranean climate. *Advances in Water Resources*, 64:62–76.

Aguilar, C., Herrero, J., and Polo, M. (2010). Topographic effects on solar radiation distribution in mountainous watersheds and their influence on reference evapotranspiration estimates at watershed scale. *Hydrology and Earth System Sciences*, 14(12):2479.

Aksoy, H. and Dahamsheh, A. (2009). Artificial neural network models for forecasting monthly precipitation in jordan. *Stochastic Environmental Research and Risk Assessment*, 23(7):917–931.

Amengual, A., Homar, V., and Jaume, O. (2015). Potential of a probabilistic hydrometeorological forecasting approach for the 28 september 2012 extreme flash flood in

murcia, spain. *Atmospheric Research*, 166:10–23.

Anctil, F., Lauzon, N., Andréassian, V., Oudin, L., and Perrin, C. (2006). Improvement of rainfall-runoff forecasts through mean areal rainfall optimization. *Journal of hydrology*, 328(3-4):717–725.

Anctil, F. and Rat, A. (2005). Evaluation of neural network streamflow forecasting on 47 watersheds. *Journal of Hydrologic Engineering*, 10(1):85–88.

Ayuso, J., Ayuso-Ruiz, P., García-Marín, A., Estévez, J., and Taguas, E. (2015). Local analysis of the characteristics and frequency of extreme droughts in málaga using the spi (standardized precipitation index). In *Project Management and Engineering*, pages 167–179. Springer.

Baker, L., Shaffrey, L., and Scaife, A. A. (2018). Improved seasonal prediction of uk regional precipitation using atmospheric circulation. *International Journal of Climatology*, 38:e437–e453.

Barker, L. J., Hannaford, J., Parry, S., Smith, K. A., Tanguy, M., and Prudhomme, C. (2019). Historic hydrological droughts 1891–2015: systematic characterisation for a diverse set of catchments across the uk. *Hydrology and Earth System Sciences Discussions*.

Barnston, A. G. and Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly weather review*, 115(6):1083–1126.

Barredo, J. I. (2007). Major flood disasters in europe: 1950–2005. *Natural Hazards*, 42(1):125–148.

Basara, J. B. and Crawford, K. C. (2002). Linear relationships between root-zone soil moisture and atmospheric processes in the planetary boundary layer. *Journal of Geophysical Research: Atmospheres*, 107(D15).

Battiti, R. (1992). First-and second-order methods for learning: between steepest descent and Newton's method. *Neural computation*, 4(2):141–166.

Belayneh, A., Adamowski, J., Khalil, B., and Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the awash river basin in ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508:418–429.

Belyaev, M., Burnaev, E., and Kapushev, Y. (2014). Exact inference for gaussian process regression in case of big data with the cartesian product structure. *arXiv preprint arXiv:1403.6573*.

Berg, P., Donnelly, C., and Gustafsson, D. (2018). Near-real-time adjusted reanalysis forcing data for hydrology. *Hydrology and Earth System Sciences*, 22(2):989–1000.

Bergillos, R. J., López-Ruiz, A., Ortega-Sánchez, M., Masselink, G., and Losada, M. A. (2016). Implications of delta retreat on wave propagation and longshore sediment transport-guadalfeo case study (southern spain). *Marine Geology*, 382:1–16.

Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences Discussions*, 5(1):1–12.

Birikundavyi, S., Labib, R., Trung, H. T., and Rousselle, J. (2002). Performance of Neural Networks in Daily Streamflow Forecasting. *Journal of Hydrologic Engineering*, 7(5):392.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, Singapore.

Blanc, J. F. and Blanc, Á. F. (2015). Las cabañuelas 0 la prediccion del tiempo en el saber popular. *NIMBUS nº 11-12*, page 151.

Blöschl, G. and Sivapalan, M. (1995). Scale issues in hydrological modelling: a review. *Hydrological processes*, 9(3-4):251–290.

BOTIJA, M. L. (1990). Influencia de la orografía y de la inestabilidad convectiva en la distribución espacial de lluvias extremas en cataluña. *Acta geológica hispánica*, 25(3):197–208.

Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252.

Box, G. E. P. and M., J. G. (1976). *Time series analysis forecasting and control*. revised edn. Holden-Day San Francisco.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, 44(1):108 – 132.

Bui, D.-M., Huynh-The, T., Lee, S., and Yoon, Y. (2015). Complexity reduction for gaussian process regression in spatio-temporal prediction. In *2015 International Conference on Advanced Technologies for Communications (ATC)*, pages 326–331. IEEE.

Cancelliere, A., Di Mauro, G., Bonaccorso, B., and Rossi, G. (2007). Drought forecasting using the standardized precipitation index. *Water resources management*, 21(5):801–819.

Chang, F.-J., Liang, J.-M., and Chen, Y.-C. (2001). Flood forecasting using radial basis function neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 31(4):530–535.

Changrsquo, L. B., Ngana, J., et al. (2013). Indigenous knowledge in seasonal rainfall prediction in tanzania: A case of the south-western highland of tanzania. *Journal of Geography and Regional Planning*, 3(4):66–72.

Chatterjee, S. and Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, pages 379–393.

Chen, N., Qian, Z., Nabney, I., and Meng, X. (2014). Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction. *Power Systems, IEEE Transactions on*, 29(2):656–665.

Cheng, K.-S., Lien, Y.-T., Wu, Y.-C., and Su, Y.-F. (2017). On the criteria of model performance evaluation for real-time flood forecasting. *Stochastic Environmental Research and Risk Assessment*, 31(5):1123–1146.

Chiu, C.-L. and Huang, J. (1970). Nonlinear time varying model of rainfall-runoff relation. *Water Resources Research*, 6(5):1277–1286.

Choubin, B., Khalighi-Sigaroodi, S., Malekian, A., and Kişi, Ö. (2016). Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological Sciences Journal*, 61(6):1001–1009.

Chylek, P., Klett, J. D., Lesins, G., Dubey, M. K., and Hengartner, N. (2014). The atlantic multidecadal oscillation as a dominant factor of oceanic influence on climate. *Geophysical research letters*, 41(5):1689–1697.

Clements, M. P., Franses, P. H., and Swanson, N. R. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20(2):169–183.

Cloke, H. and Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of hydrology*, 375(3-4):613–626.

Colorado, State, U. (2011). MODSIM-DSS. http://modsim.engr.colostate.edu/index.shtml. [Online; accessed 22-September-2015].

Coulibaly, P., Anctil, F., and Bobee, B. (2000). Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology*, 230(3):244–257.

Crochemore, L., Ramos, M.-H., Pappenberger, F., and Perrin, C. (2016). Seasonal streamflow forecasting by conditioning climatology with precipitation indices. *Hydrology and Earth System Sciences*, 21(3):1573–1591.

Daliakopoulos, I. N., Coulibaly, P., and Tsanis, I. K. (2005). Groundwater level forecasting using artificial neural networks. *Journal of Hydrology*, 309(1):229–240.

Dawson, C. W. and Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, 25(1):80–108.

Day, G. N. (1985). Extended streamflow forecasting using nwsrfs. *Journal of Water Resources Planning and Management*, 111(2):157–170.

Deka, P. C. et al. (2014). Support vector machine applications in the field of hydrology: a review. *Applied Soft Computing*, 19:372–386.

del Mar Gallardo, M., Merino, P., Panizo, L., and Salmerón, A. (2017). Integrating river basin dsss with model checking. *International Journal on Software Tools for Technology Transfer*, pages 1–16.

Delworth, T. L. and Mann, M. E. (2000). Observed and simulated multidecadal variability in the northern hemisphere. *Climate Dynamics*, 16(9):661–676.

Demuth, H. and Beale, M. (1993). Neural network toolbox for use with matlab–user's guide verion 3.0.

Deo, R. C. and Şahin, M. (2015). Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern australia. *Atmospheric research*, 161:65–81.

Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J., Rockel, B., Jacob, D., Kjellström, E., De Castro, M., and van den Hurk, B. (2007). An intercomparison of regional climate simulations for europe: assessing uncertainties in model projections. *Climatic Change*, 81(1):53–70.

DeWeber, J. T. and Wagner, T. (2014). A regional neural network ensemble for predicting mean daily river water temperature. *Journal of Hydrology*, 517:187–200.

Dezfooli, D., Hosseini-Moghari, S.-M., Ebrahimi, K., and Araghinejad, S. (2017). Classification of water quality status based on minimum quality parameters: application of machine learning techniques. *Modeling Earth Systems and Environment*, pages 1–14.

Di Mauro, G., Bonaccorso, G., Cancelliere, A., Rossi, G., et al. (2008). Use of nao index to improve drought forecasting in the mediterranean area: Application to sicily region. *Options Méditerranéennes. Série A: Séminaires Méditerranéens*, 80.

Djerbouai, S. and Souag-Gamane, D. (2016). Drought forecasting using neural networks, wavelet neural networks, and stochastic models: case of the algerois basin in north algeria. *Water resources management*, 30(7):2445–2464.

Donaire, J. J. S. (2007). El índice xerocórico. un indicador geográfico de la sequía pluviométrica (españa y polonia). *Estudios Geográficos*, 68(263):679–708.

Dorigo, M., Maniezzo, V., and Colorni, A. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*

(Cybernetics), 26(1):29–41.

Duncan, A. P. (2014). *The Analysis and Application of Artificial Neural Networks for Early Warning Systems in Hydrology and the Environment*. PhD thesis, University of Exeter.

Duvenaud, D. (2014). *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge.

Ebden, M. (2015). Gaussian processes: A quick introduction. *arXiv preprint arXiv:1505.02965*.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, pages 1–26.

Egüen, M., Aguilar, C., Solari, S., and Losada, M. Non-stationary rainfall and natural flows modeling at the watershed scale. *Journal of Hydrology*.

Egüen, M., Polo, M. J., Gulliver, Z., Contreras, E., Aguilar, C., and Losada, M. A. (2015). Flood risk trends in coastal watersheds in south spain: direct and indirect impact of river regulation. *Proceedings of the International Association of Hydrological Sciences*, 370:51–56.

Ehm, W., Gneiting, T., Jordan, A., and Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, choquet representations and forecast rankings. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(3):505–562.

Ehsani, N., Fekete, B. M., Vörösmarty, C. J., and Tessler, Z. D. (2016). A neural network based general reservoir operation scheme. *Stochastic environmental research and risk assessment*, 30(4):1151–1166.

Enas, G. G. and Choi, S. C. (1986). Choice of the smoothing parameter and efficiency of k-nearest neighbor classification. In *Statistical Methods of Discrimination and Classification*, pages 235–244. Elsevier.

Escuder-Bueno, I., Matheu, E., Altarejos-García, L., and Castillo-Rodríguez, J. T. (2011). *Risk Analysis, Dam Safety, Dam Security and Critical Infrastructure Management*. CRC Press. A decision support framework for the optimization of dam operations and hydropower production related to the Añarbe multipurpose reservoir.

Fernández Escalante, E., García Rodríguez, M., and Villarroya, F. (2006). Las acequias de careo, un dispositivo pionero de recarga artificial de acuíferos en sierra nevada, españa. caracterización e inventario. *Tecnología y desarrollo*, (4):4–33.

Fernando, T., Maier, H., and Dandy, G. (2009). Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach. *Journal of Hydrology*, 367(3):165–176.

Fix, E. and Hodges Jr, J. L. (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. Technical report, California Univ Berkeley.

Fletcher, D. and Goss, E. (1993). Forecasting with neural networks: an application using bankruptcy data. *Information & Management*, 24(3):159–167.

Flores, M. J., Ropero, R. F., and Rumí, R. (2019). Assessment of flood risk in mediterranean catchments: an approach based on bayesian networks. *Stochastic Environmental Research and Risk Assessment*, 33(11-12):1991–2005.

Galelli, S. and Castelletti, A. (2013). Tree-based iterative input variable selection for hydrological modeling. *Water Resources Research*, 49(7):4295–4310.

Galelli, S., Humphrey, G. B., Maier, H. R., Castelletti, A., Dandy, G. C., and Gibbs, M. S. (2014). An evaluation framework for input variable selection algorithms for environmental data-driven models. *Environmental Modelling & Software*, 62:33–51.

Gallardo, M. P., Fortis, S. G., Diez, Y. C., and Esteban-Parra, M. J. (2016). Análisis comparativo de índices de sequía en andalucía para el periodo 1901-2012. *Cuadernos de investigación geográfica*, (42):67–88.

Gallart, F., Llorens, P., and Latron, J. (1994). Studying the role of old agricultural terraces on runoff generation in a small Mediterranean mountainous basin. *Journal of Hydrology*, 159(1):291–303.

Gamboa, J. C. B. (2017). Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*.

Gardiner, C. (2009). *Stochastic methods*, volume 4. springer Berlin.

Gelabert, J. M.-G. (1983). Consideraciones generales sobre la meteorología de las riadas en el levante español. *Estudios Geográficos*, 44(170):31.

Gevrey, M., Dimopoulos, I., and Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling*, 160(3):249–264.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452.

Ghosh, S. and Mujumdar, P. P. (2008). Statistical downscaling of gcm simulations to streamflow using relevance vector machine. *Advances in water resources*, 31(1):132–146.

Gibbs, W. J. (1967). Rainfall deciles as drought indicators.

Giorgi, F. (1990). Simulation of regional climate using a limited area model nested in a general circulation model. *Journal of Climate*, 3(9):941–963.

Giustolisi, O., Doglioni, A., Savic, D., and Webb, B. (2007). A multi-model approach to analysis of environmental phenomena. *Environmental Modelling & Software*, 22(5):674–682.

Glezakos, T. J., Tsiligiridis, T. A., Iliadis, L. S., Yialouris, C. P., Maris, F. P., and Ferentinos, K. P. (2009). Feature extraction for time-series data: An artificial neural network evolutionary training model for the management of mountainous watersheds. *Neurocomputing*, 73(1–3):49 – 59.

Golestani, A. and Gras, R. (2014). Can we predict the unpredictable? *Scientific reports*, 4(6834).

Gonzalez-Hidalgo, J. C., Lopez-Bustins, J.-A., Štepánek, P., Martin-Vide, J., and de Luis, M. (2009). Monthly precipitation trends on the mediterranean fringe of the iberian peninsula during the second-half of the twentieth century (1951–2000). *International Journal of Climatology*, 29(10):1415–1429.

Govindaraju, R. S. (2000). Artificial neural networks in hydrology. By the ASCE task committee on application of artificial neural networks in hydrology.1. *Journal of Hydrologic Engineering*, 5(2):115–123.

Grbić, R., Kurtagić, D., and Slišković, D. (2013). Stream water temperature prediction based on gaussian process regression. *Expert Systems with Applications*, 40(18):7407–7414.

Grbić, R., Kurtagić, D., and Slišković, D. (2013). Stream water temperature prediction based on Gaussian process regression. *Expert Systems with Applications*, 40(18):7407–7414.

Grotch, S. L. and MacCracken, M. C. (1991). The use of general circulation models to predict regional climatic change. *Journal of Climate*, 4(3):286–303.

Gulliver, Z., Herrero, J., and Polo, M. (2014). Streamflow forecasting by a data driven method. two mediterranean study cases. *ITISE Proceedings*, 2:1303–1313.

Gulliver, Z., Herrero, J., and Polo, M. (2016). Medium-term predictions of cumulative runoff in a mediterranean mountain river. In *Geophysical Research Abstracts*, volume 18.

Haiden, T., Janousek, M., Bidlot, J., Buizza, R., Ferranti, L., Prates, F., and Vitart, F. Evaluation of ecmwf forecasts, including the 2018 upgrade.

Hammond, M. J., Chen, A. S., Djordjević, S., Butler, D., and Mark, O. (2015). Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*, 12(1):14–29.

Harris, I., Jones, P., Osborn, T., and Lister, D. (2014). Updated high-resolution grids of monthly climatic observations–the cru ts3. 10 dataset. *International Journal of Climatology*, 34(3):623–642.

Hasan, M. M. and Dunn, P. K. (2011). Two tweedie distributions that are near-optimal for modelling monthly rainfall in australia. *International Journal of Climatology*, 31(9):1389–1397.

Hassibi, B. and Stork, D. G. (1993). Second order derivatives for network pruning: Optimal brain surgeon. In *Advances in neural information processing systems*, pages 164–171.

Hauth, J. (2008). *Grey-Box Modelling for Nonlinear Systems*. PhD thesis, Universität Kaiserslautern.

He, Z., Wen, X., Liu, H., and Du, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, 509:379–386.

Hensman, J., Fusi, N., and Lawrence, N. D. (2013). Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*.

Herrero, J., Gulliver, Z., and Polo, M. J. (2014). Flood alert system for early warning in mountainous coastal watersheds: Coupling data-driven and physically based hydrological models. In *Proceedings of the 11th International Conference on Hydroinfromatics, New York City, USA*.

Herrero, J., Millares, A., Aguilar, C., Díaz, A., Polo, M., and Losada, M. (2012). WiMMed, theoretical basis. *Grupo de Dinámica de Flujos Ambientales (University of Granada) and Grupo de Dinámica Fluvial e Hidrología (University of Córdoba)*, 26.

Herrero, J., Polo, M., Moñino, A., and Losada, M. (2009). An energy balance snowmelt model in a Mediterranean site. *Journal of hydrology*, 371(1):98–107.

Herrero, J., Polo, M. J., and Eugster, W. (2016). Evaposublimation from the snow in the mediterranean mountains of sierra nevada (spain). *Cryosphere*, 10(6).

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401.

Hong, W.-C. (2008). Rainfall forecasting by technological machine learning models. *Applied Mathematics and Computation*, 200(1):41–57.

Hosseini, S. M. and Mahjouri, N. (2016). Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling. *Applied Soft Computing*, 38:329–345.

Hsieh, W. W. (2009). *MaMachine learning methods in the environmental sciences: neural networks and kernelschine learning methods in the environmental sciences: neural networks and kernels*. Cambridge university press.

Islam, T., Srivastava, P. K., Gupta, M., Zhu, X., and Mukherjee, S. (2014). *Computational Intelligence Techniques in Earth and Environmental Sciences*. Springer, London.

Jeong, D.-I. and Kim, Y.-O. (2005a). Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. *Hydrological Processes*, 19(19):3819–3835.

Jeong, D.-I. and Kim, Y.-O. (2005b). Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. *Hydrological Processes: An International Journal*, 19(19):3819–3835.

Juston, J. M., Kauffeldt, A., Montano, B. Q., Seibert, J., Beven, K. J., and Westerberg, I. K. (2013). Smiling in the rain: Seven reasons to be positive about uncertainty in hydrological modelling. *Hydrological Processes*, 27(7):1117–1122.

Kalra, A. and Ahmad, S. (2009). Using oceanic-atmospheric oscillations for long lead time streamflow forecasting. *Water Resources Research*, 45(3).

Kalra, A., Ahmad, S., and Nayak, A. (2013). Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns. *Advances in Water Resources*, 53:150–162.

Karbasi, M. (2018). Forecasting of multi-step ahead reference evapotranspiration using wavelet-gaussian process regression model. *Water Resources Management*, 32(3):1035–1052.

Karran, D. J., Morin, E., and Adamowski, J. (2014). Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes. *J. Hydroinf*, 16(3):671–689.

Karvelis, P., Kolios, S., Georgoulas, G., and Stylios, C. (2017). Ensemble learning for forecasting main meteorological parameters. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3711–3714. IEEE.

Kasiviswanathan, K., He, J., Sudheer, K., and Tay, J.-H. (2016). Potential application of wavelet neural network ensemble to forecast streamflow for flood management. *Journal of Hydrology*, 536:161–173.

Kasiviswanathan, K. and Sudheer, K. (2013). Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. *Stochastic environmental research and risk assessment*, 27(1):137–146.

Khan, M. S. and Coulibaly, P. (2006). Bayesian neural network for rainfall-runoff modeling. *Water Resources Research*, 42(7):1–18.

Khashei, M. and Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with Applications*, 37(1):479–489. cited By 204.

Khashei, M., Hamadani, A. Z., and Bijari, M. (2012). A novel hybrid classification model of artificial neural networks and multiple linear regression models. *Expert Systems with Applications*, 39(3):2606–2620.

Kingston, G. B., Lambert, M. F., and Maier, H. R. (2005). Bayesian training of artificial neural networks used for water resources modeling. *Water Resources Research*, 41(12):1–11.

Kişi, Ö. (2004). River flow modeling using artificial neural networks. *Journal of Hydrologic Engineering*, 9(1):60–63.

Kisi, O. (2009). Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *Journal of Hydrologic Engineering*, 14(8):773–782.

Kisi, O., Ozkan, C., and Akay, B. (2012). Modeling discharge–sediment relationship using neural networks with artificial bee colony algorithm. *Journal of Hydrology*, 428–429:94 – 103.

Kitanidis, P. K. and Bras, R. L. (1980a). Real-time forecasting with a conceptual hydrologic model: 2. applications and results. *Water Resources Research*, 16(6):1034–1044.

Kitanidis, P. K. and Bras, R. L. (1980b). Real-time forecasting with a conceptual hydrologic model: 2. applications and results. *Water Resources Research*, 16(6):1034–1044.

Knight, J. R., Folland, C. K., and Scaife, A. A. (2006). Climate impacts of the atlantic multidecadal oscillation. *Geophysical Research Letters*, 33(17).

Krajewski, W. F., Ghimire, G. R., and Quintero, F. (2020). Streamflow forecasting without models. *Journal of Hydrometeorology*, 21(8):1689–1704.

Labadie, J. W. (2004). Optimal operation of multireservoir systems: State-of-the-art review. *Journal of Water Resources Planning and Management*, 130(2):93–111.

Latt, Z. Z. and Wittenberg, H. (2014). Improving flood forecasting in a developing country: a comparative study of stepwise multiple linear regression and artificial neural network. *Water resources management*, 28(8):2109–2128.

Le, J., El-Askary, H., Allali, M., and Struppa, D. C. (2017). Application of recurrent neural networks for drought projections in california. *Atmospheric Research*, 188:100–106.

Li, W., Yang, M., Liang, Z., Zhu, Y., Mao, W., Shi, J., and Chen, Y. (2013). Assessment for surface water quality in lake taihu tiaoxi river basin china based on support vector machine. *Stochastic environmental research and risk assessment*, 27(8):1861–1870.

Lima, C. H. and AghaKouchak, A. (2017). Droughts in amazonia: Spatiotemporal variability, teleconnections, and seasonal predictions. *Water Resources Research*, 53(12):10824–10840.

Lin, G.-F. and Chen, L.-H. (2004). A non-linear rainfall-runoff model using radial basis function network. *Journal of Hydrology*, 289(1):1–8.

Linares-Rodriguez, A., Lara-Fanego, V., Pozo-Vazquez, D., and Tovar-Pescador, J. (2015). One-day-ahead streamflow forecasting using artificial neural networks and a

meteorological mesoscale model. *Journal of Hydrologic Engineering*, 20(9):05015001.

Lloyd-Hughes, B. and Saunders, M. A. (2002). A drought climatology for europe. *International journal of climatology*, 22(13):1571–1592.

López, M. F. P. (2001). Sequías de la cuenca del guadalquivir. In *Causas y consecuencias de las sequías en España*, page 303. Instituto Interuniversitario de Geografía.

Loucks, D. P., Van Beek, E., Stedinger, J. R., Dijkman, J. P., and Villars, M. T. (2005). *Water resources systems planning and management: an introduction to methods, models and applications*. Paris: Unesco.

Lu, R. (2003). Linear relationship between the interdecadal and interannual variabilities of north china rainfall in rainy season. *Chinese Science Bulletin*, 48(10):1040–1044.

Luk, K., Ball, J. E., and Sharma, A. (2000). A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *Journal of Hydrology*, 227(1):56–65.

Luo, T., Young, R., and Reig, P. (2015). Aqueduct projected water stress country rankings. *Washington, D.C.: World Resources Institute.*

MacKay, D. J. C. (1992a). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*, 4(3):448–472.

MacKay, D. J. C. (1992b). Bayesian Interpolation. *Neural Computation*, 4(3):415–447.

Madadgar, S. and Moradkhani, H. (2014). Spatio-temporal drought forecasting within Bayesian networks. *Journal of Hydrology*, 512:134–146.

Madani, K. (2010). Game theory and water resources. *Journal of Hydrology*, 381(3):225–238.

Maier, H. R. and Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software*, 15(1):101 – 124.

Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*, 25(8):891–909.

Maiti, S. and Tiwari, R. (2014). A comparative study of artificial neural networks, bayesian neural networks and adaptive neuro-fuzzy inference system in groundwater level prediction. *Environmental earth sciences*, 71(7):3147–3160.

Malakoff, D. (1999). Bayes offers a 'new' way to make sense of numbers. *Science*, 286(5444):1460–1464.

Malekmohammadi, B., Kerachian, R., and Zahraie, B. (2009). Developing monthly operating rules for a cascade system of reservoirs: Application of bayesian networks.

Environmental Modelling & Software, 24(12):1420 – 1432.

Malekmohammadi, B., Zahraie, B., and Kerachian, R. (2011). Ranking solutions of multi-objective reservoir operation optimization models using multi-criteria decision analysis. *Expert Systems with Applications*, 38(6):7851 – 7863.

Maliva, R. G. and Missimer, T. M. (2012). *Arid lands water evaluation and management*. Springer Science & Business Media.

Maqsood, I., Khan, M. R., and Abraham, A. (2004). An ensemble of neural networks for weather forecasting. *Neural Computing & Applications*, 13(2):112–122.

Marçais, J. and de Dreuzy, J.-R. (2017). Prospective interest of deep learning for hydrological inference. *Groundwater*, 55(5):688–692.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441.

Martín, J. R., García, M. M., de Pablo Dávila, F., and Soriano, L. R. (2013). Severe rainfall events over the western mediterranean sea: A case study. *Atmospheric Research*, 127:47–63.

Martin-Vide, J. and Lopez-Bustins, J.-A. (2006). The western mediterranean oscillation and rainfall in the iberian peninsula. *International Journal of Climatology*, 26(11):1455–1475.

Martos-Rosillo, S., Ruiz-Constán, A., González-Ramón, A., Mediavilla, R., Martín-Civantos, J., Martínez-Moreno, F., Jódar, J., Marín-Lechado, C., Medialdea, A., Galindo-Zaldívar, J., et al. (2019). The oldest managed aquifer recharge system in europe: New insights from the espino recharge channel (sierra nevada, southern spain). *Journal of Hydrology*, 578:124047.

Masters, T. (1993). *Practical neural network recipes in C++*. Morgan Kaufmann.

Mathbout, S., Lopez-Bustins, J., Royé, D., Martin-Vide, J., Bech, J., and Rodrigo, F. (2017). Observed changes in daily precipitation extremes at annual timescale over the eastern mediterranean during 1961–2012. *Pure and Applied Geophysics*, pages 1–16.

Mayor, Á. G., Bautista, S., and Bellot, J. (2009). Factors and interactions controlling infiltration, runoff, and soil loss at the microscale in a patchy Mediterranean semiarid landscape. *Earth Surface Processes and Landforms*, 34(12):1702–1711.

McKee, T. B., Doesken, N. J., and Kleist, J. (1993a). The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, 17(22).

McKee, T. B., Doesken, N. J., Kleist, J., et al. (1993b). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. American Meteorological Society Boston, MA.

Mediero, L., Garrote, L., and Martin-Carrasco, F. (2007). A probabilistic model to support reservoir operation decisions during flash floods. *Hydrological sciences journal*, 52(3):523–537.

Mekanik, F., Imteaz, M., Gato-Trinidad, S., and Elmahdi, A. (2013). Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology*, 503:11–21.

Millares, A., Polo, M., and Losada, M. (2009). The hydrological response of baseflow in fractured mountain areas. *Hydrology and Earth System Sciences*, 13(7):1261.

Millares, A., Polo, M., Moñino, A., Herrero, J., and Losada, M. (2014a). Bedload dynamics and associated snowmelt influence in mountainous and semiarid alluvial rivers. *Geomorphology*, 206:330–342.

Millares, A., Polo, M., Moñino, A., Herrero, J., and Losada, M. (2014b). Reservoir sedimentation and erosion processes in a snow-influenced basin in southern spain. *Reservoir Sedimentation*, pages 91–98.

Mishra, A. and Desai, V. (2006). Drought forecasting using feed-forward recursive neural network. *ecological modelling*, 198(1-2):127–138.

Mishra, A. K. and Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, 391(1):202–216.

Mishra, A. K. and Singh, V. P. (2011). Drought modeling–a review. *Journal of Hydrology*, 403(1-2):157–175.

Modaresi, F. and Araghinejad, S. (2014). A comparative assessment of support vector machines, probabilistic neural networks, and k-nearest neighbor algorithms for water quality classification. *Water resources management*, 28(12):4095–4111.

Modaresi, F., Araghinejad, S., and Ebrahimi, K. (2018). A comparative assessment of artificial neural network, generalized regression neural network, least-square support vector regression, and k-nearest neighbor regression for monthly streamflow forecasting in linear and nonlinear conditions. *Water Resources Management*, 32(1):243–258.

Modarres, R. and Ouarda, T. B. M. J. (2014). Modeling the relationship between climate oscillations and drought by a multivariate garch model. *Water Resources Research*, 50(1):601–618.

MOI (2015). Mediterranean Oscillation Index.

Molina, J. J. C. (1974). Génesis de las inundaciones de octubre de 1973 en el sureste de la península ibérica. *Cuadernos geográficos de la Universidad de Granada*, (4):149–166.

Molina, J. J. C. (1987). Inundación y avenidas de los ríos del sureste español. *Papeles de Geografía*, (13).

Molina, J. J. C. (1990). Ciclogénesis violenta en el mediterráneo: la inundación de málaga de noviembre de 1989. *Papeles de geografía*, (16):9–34.

Moller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533.

Morley, M. S. (2008). *A Framework for Evolutionary Optimization Applications in Water Distribution Systems*. PhD thesis.

Morse, G. and Stanley, K. O. (2016). Simple evolutionary optimization can rival stochastic gradient descent in neural networks. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, pages 477–484.

Moussa, R., Chahinian, N., and Bocquillon, C. (2007). Distributed hydrological modelling of a Mediterranean mountainous catchment–Model construction and multi-site validation. *Journal of Hydrology*, 337(1):35–51.

Moustris, K. P., Larissi, I. K., Nastos, P. T., and Paliatsos, A. G. (2011). Precipitation forecast using artificial neural networks in specific regions of greece. *Water resources management*, 25(8):1979–1993.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*.

Nabney, I. T. (2004). *NETLAB. Algorithms for Pattern Recognition.* Springer, London.

Napolitano, G. (2011). *An exploration of neural networks for real-time flood forecasting*. PhD thesis, University of Leeds.

Napolitano, G., Serinaldi, F., and See, L. (2011). Impact of emd decomposition and random initialisation of weights in ann hindcasting of daily stream flow series: an empirical examination. *Journal of Hydrology*, 406(3):199–214.

Naresh Kumar, M., Murthy, C., Sesha Sai, M., and Roy, P. (2009). On the use of standardized precipitation index (spi) for drought intensity assessment. *Meteorological applications*, 16(3):381–389.

Neiman, P. J., Ralph, F. M., White, A., Kingsmill, D., and Persson, P. (2002). The statistical relationship between upslope flow and rainfall in california's coastal mountains: Observations during caljet. *Monthly Weather Review*, 130(6):1468–1492.

Ngo, L. L., Madsen, H., and Rosbjerg, D. (2007). Simulation and optimisation modelling approach for operation of the Hoa Binh reservoir, Vietnam. *Journal of Hydrology*, 336(3-4):269–281.

Nørgård, P., Ravn, O., Hansen, L., and Poulsen, N. (1996). The nnsysid toolbox - a matlab toolbox for system identification with neural networks. In *Proceedings of the 1996 IEEE Symposium on Computer-Aided Control System Design*, pages 374–379. IEEE.

of Engineers-Hydrologic Engineering Center, U. A. C. (2013). HEC-ResSim. http://www.hec.usace.army.mil/software/hec-ressim/. [Online; accessed 22-October-2015].

Olden, J. D. and Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological modelling*, 154(1):135–150.

Pappenberger, F. and Beven, K. J. (2006). Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Water resources research*, 42(5).

Park, J. and Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural computation*, 3(2):246–257.

Pausas, J. G. (2004). Changes in fire and climate in the eastern iberian peninsula (mediterranean basin). *Climatic change*, 63(3):337–350.

Pérez-Palazón, M., Pimentel, R., Herrero, J., Aguilar, C., Perales, J., and Polo, M. (2015). Extreme values of snow-related variables in mediterranean regions: trends and long-term forecasting in sierra nevada (spain). *Proceedings of the International Association of Hydrological Sciences*, 369:157.

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., and Mathevet, T. (2007). Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. *Hydrological sciences journal*, 52(1):131–151.

Pinol, J., Beven, K., and Freer, J. (1997). Modelling the hydrological response of Mediterranean catchments, Prades, Catalonia. The use of distributed models as aids to hypothesis formulation. *Hydrological Processes*, 11(9):1287–1306.

Pita López, M. F. (2001). Un nouvel indice de sécheresse pour les domaines méditerranéens: application au bassin du guadalquivir (sudo-uest de l'espagne). In *13e colloque de l'Association internationale de climatologie, Nice, 6-9 septembre 2000, 225-233*. Association Internationale de Climatologie.

Poesen, J. and Hooke, J. Erosion, flooding and channel management in mediterranean environments of southern europe. *Progress in Physical Geography*.

Polo, M. J., Herrero, J., Aguilar, C., Millares, A., Moñino, A., Nieto, S., and Losada, M. (2009). WiMMed, a distributed physically-based watershed model (I): Description and validation. *Environmental Hydraulics: Theoretical, Experimental & Computational Solutions, CRC Press/Balkema*, pages 225–228.

Powell, M. J. (1977). Restart procedures for the conjugate gradient method. *Mathematical programming*, 12(1):241–254.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using bayesian

model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5):1155–1174.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer.

Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT Press, Cambridge, Massachusetts.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press.

Rasouli, K., Hsieh, W. W., and Cannon, A. J. (2012). Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, 414:284–293.

Regodon, J. (2013). *Informe sobre los episodios de mayor relevancia acaecidos en la cuenca del Guadalhorce (Málaga) por parte de la Red Hidrosur. Uso del Sistema de ayuda a la decisión (SAD) y del modelo hidrológico WiMMed*.

Remesan, R. and Mathew, J. (2015). *Machine Learning and Artificial Intelligence-Based Approaches*. Earth Systems Data and Models. Springer International Publishing.

Rezaeian-Zadeh, M. and Tabari, H. (2012). Mlp-based drought forecasting in different climatic regions. *Theoretical and applied climatology*, 109(3-4):407–414.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5):465–471.

Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., and Aigrain, S. (2012). Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371(1984).

Rodrigo, F., Esteban-Parra, M., Pozo-Vázquez, D., and Castro-Díez, Y. (2000). Rainfall variability in southern spain on decadal to centennial time scales. *International Journal of Climatology*, 20(7):721–732.

Roncoli, C., Ingram, K., and Kirshen, P. (2002). Reading the rains: local knowledge and rainfall forecasting in burkina faso. *Society & Natural Resources*, 15(5):409–427.

Russel, S. and Norving, P. (2010). *Artificial Intelligence a Modern Approach.* Pearson Education, Inc., New Jersey.

Sanchez-Toribio, M., Garcia-Marin, R., Conesa-Garcia, C., and Lopez-Bermudez, F. (2010). Evaporative demand and water requirements of the principal crops of the guadalentín valley (se spain) in drought periods. *Spanish Journal of Agricultural Research*, 8(S2):66–75.

Santos, J. F., Portela, M. M., and Pulido-Calvo, I. (2011). Regional frequency analysis of droughts in portugal. *Water resources management*, 25.14:3537–3558.

Sarle, W. S. (1996). Stopped training and other remedies for overfitting. *Computing science and statistics*, pages 352–360.

Savic, D. (2019). What is artificial intelligence and how can water planning and management benefit from it? *IAHR White Papers*, pages 1–3.

Savić, D. a., Bicik, J., and Morley, M. S. (2011). A DSS generator for multiobjective optimisation of spreadsheet-based models. *Environmental Modelling & Software*, 26(5):551–561.

Schaechtle, U. and Mansinghka, V. K. (2016). Gaussian process structure learning via probabilistic inverse compilation. *arXiv preprint arXiv:1611.07051*.

Schaefli, B. and Gupta, H. V. (2007). Do nash values have value? *Hydrological Processes*, 21(15):2075–2080.

Schellnhuber, H. J., Frieler, K., and Kabat, P. (2014). The elephant, the blind, and the intersectoral intercomparison of climate impacts.

Schlesinger, M. E. and Ramankutty, N. (1994). An oscillation in the global climate system of period 65–70 years. *Nature*, 367(6465):723.

Schmidli, J., Frei, C., and Vidale, P. L. (2006). Downscaling from gcm precipitation: a benchmark for dynamical and statistical downscaling methods. *International journal of climatology*, 26(5):679–689.

Schumann, A. H., Wang, Y., and Dietrich, J. (2011). Framing uncertainties in flood forecasting with ensembles. In *Flood risk assessment and management*, pages 53–76. Springer.

Seibert, J. (2001). On the need for benchmarks in hydrological modelling. *Hydrological Processes*, 15(6):1063–1064.

Senciales González, J. M. and Ruiz Sinoga, J. D. (2013). Spatial Analysis of Heavy Rain in Malaga city (Spain). *Boletín de la Asociación de Geógrafos Españoles*, 61:353–360.

Shamseldin, A. Y. (2010). Artificial neural network model for river flow forecasting in a developing country. *Journal of Hydroinformatics*, 12(1):22.

Shannon, C. E., Weaver, W., and Burks, A. W. (1951). The mathematical theory of communication.

Shen, C. (2018). A trans-disciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., and Tsai, W.-P. (2018). Hess opinions: Incubating deep-learning-powered hydrologic science advances as a community. *Hydrology and Earth System Sciences*, 22(11):5639–5656.

Shensa, M. J. (1992). The discrete wavelet transform: wedding the a trous and mallat algorithms. *IEEE Transactions on signal processing*, 40(10):2464–2482.

Shortridge, J. E., Guikema, S. D., and Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, 20(7):2611–2628.

Shraddhanand, S. and Wood, A. W. (2008). Use of a standardized runoff index for characterizing hydrologic drought. *Geophysical Research Letters*, 35(2):512–525.

Silverman, D. and Dracup, J. A. (2000). Artificial neural networks and long-range precipitation prediction in california. *Journal of applied meteorology*, 39(1):57–66.

Singh, K. P. (1964). Nonlinear instantaneous unit-hydrograph theory. *Journal of the Hydraulics Division*, 90(2):313–350.

Singh, P. and Borah, B. (2013). Indian summer monsoon rainfall prediction using artificial neural network. *Stochastic environmental research and risk assessment*, 27(7):1585–1599.

Sinoga, J. D. R., Marin, R. G., Murillo, J. F. M., and Galeote, M. A. G. (2011). Precipitation dynamics in southern spain: trends and cycles. *International Journal of Climatology*, 31(15):2281–2289.

Sivapalan, M., Jothityangkoon, C., and Menabde, M. (2002). Linearity and nonlinearity of basin response as a function of scale: discussion of alternative definitions. *Water Resources Research*, 38(2).

Solomatine, D., See, L. M., and Abrahart, R. (2009). Data-driven modelling: concepts, approaches and experiences. In *Practical hydroinformatics*, pages 17–30. Springer.

Solomatine, D. P. (2006). *Data-Driven Modeling and Computational Intelligence Methods in Hydrology*. John Wiley and Sons, Ltd.

Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., and Stahl, K. (2015). Candidate distributions for climatological drought indices (spi and spei). *International Journal of Climatology*, pages n/a–n/a.

Staudinger, M., Stahl, K., and Seibert, J. (2014). A drought index accounting for snow. *Water Resources Research*, 50(10):7861–7872.

Stedinger, J. R., Sule, B. F., and Loucks, D. P. (1984). Stochastic dynamic programming models for reservoir operation optimization. *Water resources research*, 20(11):1499–1505.

Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

Stenseth, N. C., Ottersen, G., Hurrell, J. W., Mysterud, A., Lima, M., Chan, K.-S., Yoccoz, N. G., and Ådlandsvik, B. (2003). Studying climate effects on ecology through the use of climate indices: the north atlantic oscillation, el nino southern oscillation and beyond. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1529):2087–2096.

Stone, J. V. (2013). *Bayes' Rule: A Tutorial Introduction to Bayesian Analysis*. Sebtel Press.

Stull, R. (2015). Practical meteorology. *An algebra-based survey of atmospheric science. University of British Columbia, Columbia*.

Sudheer, K., Gosain, A., and Ramasastri, K. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological processes*, 16(6):1325–1330.

Sujay, R. and Paresh, Chandra, D. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, 19:372–386.

Sun, A. Y., Wang, D., and Xu, X. (2014). Monthly streamflow forecasting using gaussian process regression. *Journal of Hydrology*, 511(0):72 – 81.

Suykens, J. A. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300.

Svensson, C., Hannaford, J., and Prosdocimi, I. (2017). Statistical distributions for monthly aggregations of precipitation and streamflow in drought indicator applications. *Water Resources Research*, 53(2):999–1018.

Taormina, R. and Chau, K.-W. (2015). Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and extreme learning machines. *Journal of hydrology*, 529:1617–1632.

Ticlavilca, A. M., McKee, M., and Walker, W. R. (2011). Real-time forecasting of short-term irrigation canal demands using a robust multivariate Bayesian learning model. *Irrigation Science*.

Tiwari, M. K. and Chatterjee, C. (2010). Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (banns). *Journal of Hydrology*, 382(1-4):20–33.

Toth, E. and Brath, A. (2007). Multistep ahead streamflow forecasting: Role of calibration data in conceptual and neural network modeling. *Water Resources Research*, 43(11).

Trigo, R. M., Pozo-Vázquez, D., Osborn, T. J., Castro-Díez, Y., Gámiz-Fortis, S., and Esteban-Parra, M. J. (2004). North atlantic oscillation influence on precipitation, river

flow and water resources in the iberian peninsula. *International Journal of Climatology*, 24(8):925–944.

Tripathi, S., Srinivas, V., and Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of hydrology*, 330(3-4):621–640.

Vallve, M. B. and Martin-Vide, J. (1998). Secular climatic oscillations as indicated by catastrophic floods in the spanish mediterranean coastal area (14th–19th centuries). *Climatic change*, 38(4):473–491.

Van Loon, A. F. (2015). Hydrological drought explained. *Wiley Interdisciplinary Reviews: Water*, 2(4):359–392.

Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen, V., and Vehtari, A. (2013). GPstuff: Bayesian Modeling with Gaussian Processes. *J. Mach. Learn. Res.*, 14(1):1175–1179.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer science & business media.

Velasco, M. J. P. and Serrano, A. V. (2011). El diseño de itinerarios turísticos para la puesta en valor del patrimonio territorial. las acequias de careo en el parque nacional de sierra nevada. *Cuadernos de Turismo*, (27):785–809.

Versini, P.-A., Berenguer, M., Corral, C., and Sempere-Torres, D. (2014). An operational flood warning system for poorly gauged basins: demonstration in the guadalhorce basin (spain). *Natural hazards*, 71(3):1355–1378.

Vicente-Serrano, S. M. (2006). Differences in spatial patterns of drought on different time scales: an analysis of the iberian peninsula. *Water Resources Management*, 20(1):37–60.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7):1696–1718.

Vicente-Serrano, S. M., Beguería, S., López-Moreno, J. I., El Kenawy, A. M., and Angulo-Martínez, M. (2009). Daily atmospheric circulation events and extreme precipitation risk in northeast spain: Role of the north atlantic oscillation, the western mediterranean oscillation, and the mediterranean oscillation. *Journal of Geophysical Research: Atmospheres*, 114(D8).

Viesca Álvarez, C. (2011). *Estructura bayesiana de variables hidrometeorológicas para sistemas de alerta en tiempo real.* University of Cordoba.

Von Storch, H. and Navarra, A. (2013). *Analysis of climate variability: Applications of statistical techniques proceedings of an autumn school organized by the Commission of*

the European Community on Elba from October 30 to November 6, 1993. Springer Science & Business Media.

Vrugta, J. A. (2015). Modelavg: A matlab toolbox for postprocessing of model ensembles. *Department of Civil and Environmental Engineering, University of California Irvine*.

Wasserman, P. D. (1993). *Advanced methods in neural computing*. John Wiley & Sons, Inc.

Wilby, R. L. and Wigley, T. (2000). Precipitation predictors for downscaling: observed and general circulation model relationships. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 20(6):641–661.

Willems, P. (2013). Multidecadal oscillatory behaviour of rainfall extremes in europe. *Climatic Change*, 120(4):931–944.

Wittenberg, H. (1999). Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes*, 13(5):715–726.

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):495–518.

Wu, H., Hayes, M. J., Wilhite, D. A., and Svoboda, M. D. (2005). The effect of the length of record on the standardized precipitation index calculation. *International Journal of Climatology*, 25(4):505–520.

Wu, H., Svoboda, M. D., Hayes, M. J., Wilhite, D. A., and Wen, F. (2007). Appropriate application of the standardized precipitation index in arid locations and dry seasons. *International Journal of Climatology*, 27(1):65–79.

Yang, X.-S. (2010). *Nature-inspired metaheuristic algorithms*. Luniver press.

Yaseen, Z. M., El-Shafie, A., Jaafar, O., Afan, H. A., and Sayl, K. N. (2015). Artificial intelligence based models for stream-flow forecasting: 2000–2015. *Journal of Hydrology*, 530:829–844.

Yi, J.-H., Wang, J., and Wang, G.-G. (2016). Improved probabilistic neural networks with self-adaptive strategies for transformer fault diagnosis problem. *Advances in Mechanical Engineering*, 8(1):1687814015624832.

Zameer, A., Arshad, J., Khan, A., and Raja, M. A. Z. (2017). Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks. *Energy conversion and management*, 134:361–372.

Zhang, C., Wei, H., Zhao, X., Liu, T., and Zhang, K. (2016). A gaussian process regression based hybrid approach for short-term wind speed prediction. *Energy Conversion and Management*, 126:1084–1092.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62.

Zhang, X., Liang, F., Srinivasan, R., and Van Liew, M. (2009). Estimating uncertainty of streamflow simulation using bayesian neural networks. *Water resources research*, 45(2).

Zhang, X. and Zhao, K. (2012). Bayesian neural networks for uncertainty analysis of hydrologic modeling: a comparison of two schemes. *Water resources management*, 26(8):2365–2382.

Zhao, T., Cai, X., and Yang, D. (2011). Effect of streamflow forecast uncertainty on real-time reservoir operation. *Advances in water resources*, 34(4):495–504.

Zhou, Z.-H., Wu, J., and Tang, W. (2002). Ensembling neural networks: many could be better than all. *Artificial intelligence*, 137(1):239–263.

Zimek, A. and Vreeken, J. (2015). The blind men and the elephant: On meeting the problem of multiple truths in data from clustering and pattern mining perspectives. *Machine Learning*, 98(1-2):121.

FIGURE C.4: Waterfall. M. C. Escher. 1961. *Lithograph*

Thank you for your attention